

The National Cohort

A prospective epidemiologic study resource for health and disease research in Germany



The National Cohort

A prospective epidemiologic study resource for health and disease research in Germany

Scientific coordination

Prof. Dr. Wolfgang Hoffmann
University of Greifswald
D-17487 Greifswald
wolfgang.hoffmann@uni-greifswald.de

Prof. Dr. Karl-Heinz Jöckel
University Hospital Essen
D-45122 Essen
k-h.joeckel@uk-essen.de

Prof. Dr. Rudolf Kaaks
German Cancer Research Center
D-69120 Heidelberg
r.kaaks@dkfz.de

Prof. Dr. Dr. H.-Erich Wichmann
Helmholtz-Center Munich
D-85764 Neuherberg
wichmann@helmholtz-muenchen.de

Scientific project management

Dr. Karin Halina Greiser, MPH
German Cancer Research Center
D-69120 Heidelberg
h.greiser@dkfz.de

PD Dr. Jakob Linseisen
Helmholtz-Center Munich
D-85764 Neuherberg
j.linseisen@helmholtz-muenchen.de

Administrative coordination

Dr. Susanne Weg-Remers, Mag. rer. pub.
German Cancer Research Center
D-69120 Heidelberg
s.weg-remers@dkfz.de

Applicants

Cluster Bavaria		
Study center Augsburg	Prof. Dr. Dr. H.-Erich Wichmann	Helmholtz Center Munich (HMGU)
	PD Dr. Jakob Linseisen	
Study center Regensburg	Prof. Dr. Dr. Michael Leitzmann	Regensburg University (RU)
	Prof. Dr. Iris M. Heid	
Cluster Baden-Wuerttemberg/Saarland		
Study center Heidelberg/Mannheim	Prof. Dr. Rudolf Kaaks	German Cancer Research Center (DKFZ), Heidelberg
	Dr. Karin Halina Greiser	
	Prof. Dr. Heiko Becher	University of Heidelberg
Study center Freiburg	Prof. Dr. Dr. Karin Michels	University of Freiburg
	Dr. Claudia Schmoor	
Study center Saarbruecken	Prof. Dr. Hermann Brenner	German Cancer Research Center (DKFZ), Heidelberg
	Christa Stegmaier	Saarland Cancer Registry, Saarbruecken
Cluster North Rhine-Westphalia		
Study center Essen	Prof. Dr. Karl-Heinz Jöckel	University Hospital Essen, University of Duisburg-Essen
	PD. Dr. Susanne Moebus	
Study center Muenster	Prof. Dr. Klaus Berger	University of Muenster
	Prof. Dr. Hans-Werner Hense	
Study center Duesseldorf	Prof. Dr. Guido Giani	German Diabetes Center
	Prof. Dr. Ursula Krämer	Leibniz Research Institute for Environmental Medicine (IUF)
Cluster Saxony/Saxony-Anhalt		
Study center Halle	Prof. Dr. Andreas Stang	Martin-Luther-University of Halle-Wittenberg
	Prof. Dr. Johannes Haerting	
Study center Leipzig	Prof. Dr. Markus Löffler	University of Leipzig and University Hospital Leipzig
	Prof. Dr. Joachim Thiery	

Cluster Berlin-Brandenburg		
Study center Berlin-North	Prof. Dr. Tobias Pischon	Max Delbrueck Center for Molecular Medicine (MDC) Berlin-Buch
Study center Berlin-Center	Prof. Dr. Stefan Willich	Charité -- University Medical Center Berlin
	Prof. Dr. Peter Heuschmann	
Study center Berlin-South/Brandenburg	Prof. Dr. Heiner Boeing	German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE)
Study support center	Dr. Bärbel-Maria Kurth	Robert Koch Institut (RKI) Berlin
Cluster Lower Saxony/Hamburg/Bremen		
Study center Brunswick/Hanover	PD Dr. Dr. Frank Pessler (temporary)	Helmholtz Center for Infection Research (HZI)
	Prof. Dr. Lothar Kreienbrock	University of Veterinary Medicine Hanover, Foundation (TiHo)
Study center Hamburg	Prof. Dr. Dieter Flesch-Janys	University Medical Center Hamburg-Eppendorf (UKE)
	Dr. Nadia Obi	
Study center Bremen	Prof. Dr. Wolfgang Ahrens	Bremen Institute of Prevention Research and Social Medicine (BIPS)
	PD Dr. Thomas Behrens	
Cluster Schleswig-Holstein		
Study center Kiel	Prof. Dr. Ute Nöthlings	Christian-Albrechts-University Kiel
	Prof. Dr. Alexander Katalinic	University of Luebeck
Cluster Mecklenburg-Western Pomerania		
Study center Neubrandenburg	Prof. Dr. Henry Völzke	Ernst Moritz Arndt University, Greifswald
	Prof. Dr. Wolfgang Hoffmann	

Executive Summary

1. The National Cohort – A prospective epidemiologic study resource for health and disease research in Germany

Since early 2007, a growing network of German research institutions has jointly been developing plans for a large prospective cohort that should be used as a common, national resource for studies on the risk factors and etiologic mechanisms of major diseases in the German population. The institutions include several of the publically funded Helmholtz institutes, a large number of universities, and other non-Helmholtz public research institutions. A first outline of these plans was formally evaluated in April 2008 by an international review panel, in the context of a five-yearly scientific peer review of the Helmholtz Association. For the biomedical Helmholtz institutes, coordinated by the German Cancer Research Center [DKFZ] and Helmholtz Research Center München [HMGU], start-up funds at the level of 20 Mio Euro for the period 2009-2013 have been provided.

To include the expertise from universities and non-Helmholtz public research institutions as from the earliest planning phase, a national Epidemiological Planning Committee (“EPC”) was established in February 2009 with representatives from the biomedical Helmholtz institutes, the universities, two public research center (Robert Koch Institute [“RKI”] Berlin and the German Institute for Nutrition Research [“DIfE”], Potsdam) and the three German medical scientific societies primarily engaged in epidemiology. It was decided that the National Cohort should have three major goals:

- (i) to serve as a research platform for future epidemiologic population-based research in Germany,
- (ii) to make valuable contributions to answer innovative research questions in the field of epidemiology, basic science, and their intersection and
- (iii) to cooperate internationally with other large-scale prospective studies.

As many epidemiologic research institutions had expressed their interest in participating in subject recruitment for the cohort, in autumn 2009 a formal call for tender was published and international peer review of the applications was organized. The selection of partner centers was based on a pre-defined qualification profile. During the application process, the applicant centers formed regional clusters for scientific and logistic cooperation. The review led to the selection of 24 research institutions, in total, including 14 Universities, 5 Helmholtz Research Centers, three Leibniz centers, and the Robert-Koch-Institute that met the qualification criteria for taking part in cohort recruitment.

Under the auspices of the EPC, a total of 17 thematic working groups* were established early in 2009 to identify the most challenging research questions to be addressed with the National Cohort, and to discuss questions of study design and content. These working groups involved over 200 scientists (among them also many clinical scientists) and covered the major disease areas, as well as a number of specific topics related to risk factor assessment and study logistics. By December 2009, each of the thematic working groups prepared a report with proposals for specific topics to be studied, and recommendations for instruments to be used for data collection or medical examinations. These reports provided the basis of the scientific protocol that is described in part A of the present research application, and that was developed in 2010 by a series of writing teams guided by a small editorial board of EPC members.

* Reports of the thematic working groups can be found on the CD ROM.

2. The National Cohort: Overview of study design

The National Cohort will include a total of 100,000 women and 100,000 men aged 20 to 69 years (40,000 individuals younger than 40 years, and 160,000 individuals aged at least 40 years), who will be recruited through a network of 18 study centers, organized in eight geographic clusters throughout Germany, representing the population of almost all federal states and covering metropolitan, urban and rural regions. At each of the centers, a random sample of the general population will be drawn from local municipal population registries within defined strata of age and sex. An overall study participation proportion of at least 50% will be targeted. All participants will be invited to the study centers to take part in physical and medical examinations, the collection of biomaterials, a computer-assisted personal interview, and to fill in computer-aided or web-based questionnaires. Information will be collected at two different levels of intensity:

Baseline Assessment

- ▶ At the **basic study level ("Level 1")**, applying to all 200,000 study participants, baseline interviews, questionnaires and examinations will be performed, by using a 2.5-h interview and examination protocol. Physical and medical examinations will include measurements of anthropometric indices, a dual-energy x-ray assessment (DXA) of body composition and bone density, blood pressure measurements, measurements of arterial stiffness, an electrocardiogram (ECG), three-dimensional echocardiography, spirometry, measurements of physical activity through accelerometry (7 days), a basic test of physical fitness, an assessment of cognitive function, and a tooth count. A further key element of the core study protocol at Level 1 is the collection of biomaterials from all study participants. The protocols for blood sampling and storage of aliquots of plasma, serum, DMSO-preserved viable white blood cells (in a subgroup), intact lymphocytes, erythrocytes, urine, saliva, nasal swabs and stool samples in a central bio-repository with decentralized back-up storage will be designed to permit the greatest possible choice of future laboratory analyses.
- ▶ At a **second study level ("Level 2")**, applying to a 20% representative sub-sample of the cohort (40,000 subjects) proportionally spread over all participating study centers and age-strata, an intensified and extended 4-hour examination protocol will be employed that, in addition to the general examinations of Level 1, include an oral glucose tolerance test, carotid sonography with measurement of intima-media-thickness, long-term ECG, sleep apnoe assessment, enhanced motoric, sensoric and cognitive function tests, measurement of airway inflammation, assessment of oral health, and additional questionnaires.
- ▶ A further intensified study level is formed by a **whole-body MRI program (including whole body, cardiac, and brain MRI protocol)**, which will also involve 40,000 subjects in selected study centers. Whole-body MRI will generate a comprehensive morphologic and functional data base to be used to estimate the prevalence and incidence of MRI based findings and their regional distribution their changes over time. This data base will be used to identify novel morphologic MRI-based markers of risk for various disease states and subclinical morphologic and functional changes.
- ▶ In addition, add-on **"Level 3"** studies that require further in-depth phenotyping or disease ascertainment may be planned at some or all of the study centers, provided that the additional examinations do not interfere with the Level 1 and Level 2 programs, respectively. Level 3 studies are further conditional on external funding.

Full-scale re-assessment after five years

All participants of the National Cohort will be re-invited for a second examination five years after their baseline recruitment. The full program of the respective study levels, including the MRI program will then be repeated. At the 5-year assessment, intra-individual, medium-term changes in risk factors and prospective changes in quantifiable preclinical morbidity characteristics and incident clinical disease can be investigated.

Short-term calibration/reliability substudies

In addition to the medium-term (5-year) repetition of measurements, short-term (1-year) replication studies will be embedded in the cohort, using a 6000-subject sub-sample of participants of the first (baseline) visit, and a 4000-subject subsample of participants in the second visit. These reliability/calibration studies will serve to estimate within-subject “random” variations in risk factor measurements, so as to allow corrections of attenuation effects resulting from random misclassification.

Methods for follow-up

As from their first enrolment into the National Cohort, and in addition to the five-year repeat visit, all cohort participants will be re-contacted every 2-3 years and asked to fill in short questionnaires about changes in lifestyle and other characteristics (e.g., use of medications, smoking, menopausal status, selected disease symptoms) and about the occurrence of selected, major disease (“active” follow-up). In parallel, a mortality follow-up will be performed every 2-3 years. For the follow-up on incident cancer, it is anticipated that increasing use can be made of systematic record linkage with existing cancer registries (“passive” follow-up), and also the mortality follow-up will be largely based on record linkage.

Biobanking

The major part of the biomaterials collected during recruitment and re-assessment will be stored in a centralized biobank of the National Cohort which will be established at the Helmholtz Zentrum München, and will include a fully automated bio-repository. Decentralized back-up storage is planned for one third of the biomaterials at the recruitment clusters. For all incident cases of major (comparatively frequent) forms of cancer, it is foreseen to systematically collect tumour samples, which will be stored in a centralized tumour bank.

Ethical aspects and data protection

The National Cohort will be conducted in accordance with the Federal Data Protection Act and all other applicable legislation and will take into account advice from the German Ethics Council and other ethically relevant guidelines, to maximally ensure the confidentiality of collected data and samples, and to carefully design informed consent procedures. The concept on ethics and data protection of the National Cohort has been developed and discussed with the corresponding advisory boards and authorities. Before the start of the recruitment, the study protocol will undergo a final full review by the respective responsible data protection commissioners and ethics committees.

Time Line

The projected time schedule for the National Cohort covers a period of 25-30 years. However, the present application refers only to the first 10 years in which 5 years of baseline assessment will be followed by 5 years of re-assessment of the full cohort. Active follow-up by means of questionnaires is scheduled every 2-3 years. Use of data and material for epidemiological studies can start as soon as the baseline recruitment has been completed.

3. Scientific aims of the National Cohort

The National Cohort is designed to address research questions concerning a wide range of possible causes of major chronic diseases. The collection of biomaterials in combination with extensive information from questionnaires and medical examinations represents one of the central components. Using combinations of such information – repeated over time - will make it possible to specifically address pathways of disease development providing clues to the biological mechanisms that may explain observed relationships. The biomarker measurements will cover domains as diverse as nutrition and metabolism, hormonal and other physiological effects of chronic psychological stress, viral and bacterial infections, immune status, genetics and epigenetics. The overarching objective of the National Cohort will be to provide a sound knowledge base for improved and more targeted measures for the primary and secondary prevention of major diseases, tailored to the German population.

Although a wide range of diseases can be prospectively ascertained in the National Cohort, special emphasis will be given to diseases that are responsible for the major burden of disease in Germany, both in terms of frequency and economic consequences.

Overall, **four general objectives** can be defined for the National Cohort, and within their general scope a number of more specific research questions will be of particular focus:

1. Identification of etiological pathways from life-style and environmental risk factors to major chronic diseases and functional impairments

Within the scope of this first general objective, one set of more specific aims is to increase understanding of the role of particular risk factors in the development of the following major forms of chronic diseases:

- ▶ Cardiovascular diseases
- ▶ Diabetes mellitus
- ▶ Cancer
- ▶ Neurologic and psychiatric diseases
- ▶ Respiratory diseases
- ▶ Infectious diseases

The high incidence of these diseases is largely the result of *lifestyle factors* such as smoking, alcohol consumption, nutrition, or lack of physical activity, as well as *other risk factors* such as chronic infections, occupational and environmental influences, child-bearing patterns, and hormone use for contraception and for treatment of postmenopausal symptoms. However, many questions are still open concerning the specificity with which particular risk factors –and their interactions – cause disease, the pathomechanisms, and the quantitative importance of risk factors in terms of etiological and population-attributable fractions. In the National Cohort, therefore, a major emphasis will be placed on the impact of lifestyle and environmental factors as possible causes of disease, with particular interest in the following domains:

- ▶ Body composition, physical fitness, and related metabolic factors
- ▶ Physical activity
- ▶ Diet and nutrition
- ▶ Psychosocial factors
- ▶ Viral and bacterial infections
- ▶ Immune status and immune senescence

Finally, a major and overall methodological objective for the National Cohort in relation to risk factor identification will be to improve quantitative estimates of relative and attribut-

able risks, by collecting repeated measurements over time. In many ongoing studies, the strength of the relationships of key risk factors with disease may have been substantially biased by not having sufficiently accounted for within-subject variations over time. A less than full appreciation of the quantitative importance of specific risk factors may cause erroneous priority setting for prevention. In the National Cohort, the repeated collection of measurements and biomaterials both over the short term, in calibration sub-studies, and over the medium term through the 5-year repeat visit of all study participants, will enable us to quantitatively estimate relative risks more accurately.

2. Studies of the geographic and socio-economic disparities in health status and disease risks in Germany and possible causes and explanations

Within Germany health status varies significantly across socio-economic strata as defined by level of formal education, employment status, income, or by ethnicity (i.e., belonging to immigrant minority groups). Moreover, substantial geographic variations in health status can be observed. A major general objective for the National Cohort will be to contribute detailed descriptions of these variations and to improve understanding of their causes. Besides socio-economic descriptors, environmental, psychosocial and behavioural factors will be studied as key determinants of social health disparities.

3. Development of risk prediction models for identifying individuals at increased risk of developing major chronic diseases, so as to allow personalized prevention strategies

Risk prediction models and algorithms for stratification of individuals into categories of lower and higher risk of developing major chronic diseases can be important tools for personalized medicine and personalized prevention strategies. The National Cohort will provide an outstanding resource for studies aiming to develop and/or validate comprehensive risk models that integrate risk factor information obtained by questionnaires, as well from clinical measurements and assessments of genetic and other biological markers in blood, urine and other bio-specimens.

4. Evaluation of markers for early detection of disease and pre-disease phenotypes, so as to develop effective tools for disease prevention

Detection of chronic disease or subclinical phenotypes at an earlier stage of development may increase chances of an effective cure, and/or minimize side-effects of treatment. Stimulated by recent technological advances in the fields of genomics, transcriptomics, proteomics, metabolomics, and epigenomics, research on bio-markers for early detection of chronic disease or pre-disease states is becoming a major theme for prevention. The stored biospecimen collection of the National Cohort will provide a precious resource for the discovery and validation of novel biomarkers for early disease detection, ensuring a rapid connection between basic discovery research and its confirmation and validation in appropriately designed human population studies.

For the National Cohort, strong emphasis is given to methods to optimize specimen collection, pre-analytical processing, transport and storage, so as to minimize pre-analytical artifacts and to allow future studies on a large variety of markers.

In addition to bio-markers, the large-scale MR imaging component of National Cohort may also provide unique opportunities for evaluation and discovery of specific morphological characteristics that may indicate early stages of disease development.

The sample sizes of the entire cohort and the different sub-cohorts are based on thoroughly conducted statistical calculations that show that these numbers are sufficient, but also nec-

essary to address the above described research objectives and to implement an epidemiologic research tool with adequate statistical power.

4. The National Cohort in the context of existing German and European studies

Added value in comparison to existing German cohorts

Several medium-sized epidemiologic cohort studies are currently ongoing in Germany, covering a total of 100,000 – 120,000 individuals. Many of these studies, however, were planned independently of each other and have different study elements. Thus, only for some of the existing German cohorts could pooled data analyses be conducted, and then only for part of the data collected, addressing a variety, but still limited number of research questions. The National Cohort will have a homogeneous study protocol for a large sample size, which will permit analyses of many research questions for which the existing German cohorts are statistically under-powered.

A further major argument for setting up a new prospective cohort study is that most existing cohorts in Germany already have a median age above 50 years (in some of the larger cohorts [e.g., EPIC] even above 60 years) and cannot be used to study the development of disease at younger ages. In addition, existing cohorts were initiated up to 25 years ago, and lag-times between exposure assessments and including blood sample collection till disease diagnosis are becoming very long. Moreover, for several of the larger German cohorts, such as the two German EPIC cohorts, biobank resources will be largely used up in the coming 10-12 years. The National Cohort aims to ensure the continuing availability of up-to-date prospective biobank resources after 2020.

Finally, the National Cohort will provide a population-based, highly standardized and comprehensive database that covers much of the heterogeneity both with respect to risk factors and major diseases in the general population across Germany.

Added value of the National Cohort in comparison to other European cohorts

Several large epidemiologic cohort studies exist or are being planned in other European countries (e.g., the UK, Netherlands, France, and Sweden). The National Cohort will have several advantages as compared with other large European cohorts, including

- (i) High quality of bio-samples by highly standardized pre-analytical processing and automated storage;
- (ii) Collection and storage of a diverse range of specialized bio-samples, such as living lymphocytes, saliva, nasal swabs, and faeces;
- (iii) Prospective re-examination after 5 years for all 200,000 participants
- (iv) Detailed functional characterisation and examination of intermediate phenotypes and subclinical disease (body composition, cardiovascular examinations, pre-clinical diabetes, neurocognitive functions, specific lung function measurements, musculoskeletal examinations, oral health)
- (v) Implementation of MRI techniques (e.g., MRI of brain, heart, whole body).

In addition, compatibility of questionnaires and medical examinations with those used in other European studies is planned to the greatest extent possible. For example, elements from the questionnaires used in the UK Biobank, CONSTANCES and LifeGene cohorts will also be used in the National Cohort, and examinations such as step test, eye and ear test modules and imaging protocols for MRI from UK Biobank will be adapted. This high compatibility of questionnaires and examinations will enhance future data pooling for studies

that may require cohorts of even larger size, e.g., to identify risk factors and causes of less frequent disease outcomes.

5. Governance structure

As a large and long-term scientific project, with participation of a large number of institutions, the National Cohort will require special organisational and governance structures. The National Cohort will be organized as a registered association (eingetragener Verein) in which universities, Leibniz institutes and Helmholtz centers are members. The general assembly/epidemiologic steering committee (Mitgliederversammlung) of the association elects the board of directors (Vorstand) as governing/managing body (geschäftsführendes Organ) and the head of the central executive office (Geschäftsstelle). Members of the commission of funding partners (Kommission der Zuwendungsgeber) will be BMBF, representatives of the federal states, and Helmholtz. Further elements of the governance structure include the scientific advisory board (wissenschaftlicher Beirat) and the ethics advisory board (Ethikbeirat). Internal boards authorized by the general assembly and the board of directors, as the working group scientific management or the use and access committee will be responsible for particular aspects of the scientific management (details are provided in **Part B**).

6. Outlook – Capacity building for epidemiology in Germany

Modern chronic disease epidemiology increasingly integrates information directly obtained from study participants through questionnaires and interviews with the knowledge and measurement tools from clinical and biological sciences. In Germany, the data and bio-bank resources of the National Cohort will further foster such integration of basic biologic sciences into public-health oriented research.

In addition, and equally importantly, the National Cohort will also provide a major platform for cooperation between epidemiologic research centers at universities and at other research institutions throughout Germany. The latter is crucial, given the federal structure of Germany's research landscape, in which universities are funded mostly by their respective federal states of residence, while the Helmholtz institutes are mainly funded by the German Federal Ministry of Education and Research [BMBF]. In recognition of the needs to strengthen research at the universities in general and to strengthen epidemiologic research capacities on a national level, the German federal states, BMBF and the Helmholtz Association have all expressed interest in contributing funds to the National Cohort.

Concurrent to the planning of the National Cohort, BMBF has initiated German Centers for Health Research, with the purpose to assemble and unify disease specific scientific competence, to close research gaps, and to improve research on early diagnosis, therapy and prevention. In addition to the German Centers for Neurodegenerative Diseases and for Diabetes Research, which have already been implemented (2008-2009), national Centers for Lung Research, Cardiovascular Research, Translational Cancer Research, and Infectious Disease Research are currently being established (2011). The interaction between the National Cohort and the German Centers for Health Research will strengthen research in the respective disease-oriented areas and will provide important synergistic effects. The fact that many of the universities and other research institutions that have been selected as partners of these new, German research centers are also a partner in the National Cohort can be seen as an important basis for this interaction.

7. Examples for specific research questions

To illustrate the scientific potential of the National Cohort, **Table 1.1** of our main application text (**Sect. A.1.4**) lists 37 examples of specific research questions related to major disease outcomes that will be addressed by the National Cohort in ways that are not possible with currently ongoing prospective studies in Germany.

Contents

A.1 INTRODUCTION AND OVERVIEW 1

A.1.1 Societal and economic burden of chronic disease in Germany 1

A.1.2 Prospective cohorts as an indispensable tool for health research 1

A.1.3 The National Cohort: Overview of study design 2

A.1.4 Scientific aims of the National Cohort 6

A.1.5 Arguments for a large cohort in Germany 12

A.1.6 Justification of sample size requirements 15

A.1.7 Outlook: The National Cohort as basis for collaborations
and further research projects 16

A.1.8 Preparatory work for the National Cohort 19

**A.2 SCIENTIFIC BACKGROUND AND RATIONALE FOR STUDY
ELEMENTS 21**

A.2.1 Introduction 21

A.2.2 Major disease outcomes of core focus 23

A.2.2.1 Cardiovascular diseases 24

A.2.2.2 Diabetes mellitus 26

A.2.2.3 Cancer 27

A.2.2.4 Neurologic and psychiatric diseases 28

A.2.2.5 Respiratory diseases 32

A.2.2.6 Infectious diseases 33

A.2.3 Intermediate (preclinical) phenotypes and function measurements ... 35

A.2.3.1 Cardiovascular functions and preclinical phenotypes 35

A.2.3.2 Diabetes-related functional measurements and
preclinical phenotypes 36

A.2.3.3 Cancer-related precursor stages 38

A.2.3.4 Intermediate stages of neurologic and psychiatric diseases 39

A.2.3.5 Respiratory function and preclinical phenotype 39

A.2.3.6 Musculoskeletal functions and phenotypes 40

A.2.4 Major areas of exposure and risk factors	41
A.2.4.1 Physical activity, body composition, and physical fitness	41
A.2.4.2 Diet	44
A.2.4.3 Smoking and alcohol consumption	45
A.2.4.4 Psychosocial factors	45
A.2.4.5 Socioeconomic status	46
A.2.4.6 Sleep-related characteristics	48
A.2.4.7 Chronic infections, immune factors, and microflora	48
A.2.4.8 Occupational and environmental exposures	51
A.3 STUDY DESIGN.....	55
A.3.1 Population sampling and recruitment	55
A.3.1.1 Study regions and study population	55
A.3.1.2 Population sampling	61
A.3.1.3 Participation rates	62
A.3.1.4 Recruitment of study participants	63
A.3.1.5 Contact results, response status, and nonresponder information ...	65
A.3.1.6 Equipment and organization of local study centers.....	65
A.3.2 Baseline interview and questionnaires	68
A.3.2.1 Core interview and questionnaires	68
A.3.2.2 Specific interview and questionnaire modules	73
A.3.3 Baseline physical measurements	84
A.3.3.1 Overview	84
A.3.3.2 Cardiovascular examinations	87
A.3.3.3 Diabetes-related measurements	91
A.3.3.4 Cognitive functioning tests	92
A.3.3.5 Respiratory disease-related measures	93
A.3.3.6 Examination of the musculoskeletal system	94
A.3.3.7 Oral health	94
A.3.3.8 Sense organ functions and diseases	95
A.3.3.9 Anthropometry	97
A.3.3.10 Physical activity and physical fitness	98

A.3.3.11 Feasibility studies for optimization of the final study program	100
A.3.4 Assessments through medical imaging.....	103
A.3.4.1 Rationale.....	103
A.3.4.2 Background.....	103
A.3.4.3 Objectives.....	104
A.3.4.4 Study population, imaging methods, and planned study logistics	107
A.3.4.5 International scientific advisory and data safety monitoring board....	113
A.3.5 Collection, preanalytical processing, storage, and retrieval of biomaterials.....	114
A.3.5.2 Types of biomaterials considered.....	115
A.3.5.3 Concept for collection and preanalytic processing of biological samples	118
A.3.5.4 Concept for transport, storage, and retrieval of biomaterials.....	120
A.3.5.5 Development of a tumor biobank for histologic and molecular tumor subclassification	123
A.3.6 Reassessment of intermediate (preclinical) phenotypes, measurements of function, and risk factors	125
A.3.6.1 Assessments of longitudinal changes in intermediate (preclinical) phenotypes and measurements of function	125
A.3.6.2 Assessment of changes in risk factors for disease (including risk factors that may be determined at a later date in collected biomaterials).	127
A.3.6.3 Reliability / calibration substudies	128
A.3.7 Procedures for prospective follow-up of vital status and disease occurrences	128
A.3.7.1 Introduction.....	128
A.3.7.2 Recontacting and tracking of study participants	129
A.3.7.3 Follow-up procedures for ascertainment of vital status and incident diseases	129
A.3.7.4 Prospective assessment of changes in exposures/risk factors	132
A.3.7.5 Endpoint committees.....	132
A.3.8 Use of secondary data sources.....	133
A.3.8.1 Rationale.....	133

A.3.8.2 Social security data	133
A.3.8.3 Environmental data	135
A.3.8.4 Strengths and possible limitations of secondary data use for the National Cohort	135
A.3.8.5 Experiences with data linkage of secondary data in epidemiologic studies in Germany	137
A.3.8.6 Data quality and quality assurance of secondary data	137
A.3.9 The National Cohort in the context of other cohorts of adults in Germany and internationally	137
A.4 INTEGRATED DATA MANAGEMENT	143
A.4.1 Aims and requirements	143
A.4.2 Organizational model	145
A.4.2.1 Study centers	147
A.4.2.2 Integration centers	148
A.4.2.3 Competence units	149
A.4.2.4 Trust center	150
A.4.2.5 Data transfer and project documentation unit	151
A.4.3 Technical aspects of data collection, processing, storage, and analysis	151
A.4.3.1 Software infrastructure	151
A.4.3.2 Electronic data capture	151
A.4.3.3 Data from medical devices	152
A.4.3.4 Handling of paper forms	152
A.4.3.5 Processing of study data by competence units	153
A.4.3.6 Integration and storage of study data	153
A.4.3.7 Analysis and use	154
A.5 METHODS FOR QUALITY ASSURANCE AND QUALITY CONTROL	155
A.5.1 Organizational structure for quality assurance and quality control .	155
A.5.1.1 Internal quality management	155
A.5.1.2 External quality management	156
A.5.2 Principles for quality assurance and quality control in the National Cohort	156

A.5.3 Standardization of sampling methods and assurance of high participation	158
A.5.4 Quality assurance and quality control of questionnaires	159
A.5.4.1 Quality assurance of questionnaires and interviews	159
A.5.4.2 Quality control of questionnaires and interviews	159
A.5.5 Quality assurance and quality control of medical examinations	159
A.5.5.1 Quality assurance of medical examinations	159
A.5.5.2 Quality control of medical examinations.....	160
A.5.6 Quality assurance and quality control of the collection and processing of biological specimens.....	160
A.5.6.1 Quality assurance of biological specimens.....	160
A.5.6.2 Quality control of biological specimens.....	161
A.5.7 Measures to assure high and consistent image quality in the MRI substudy	162
A.5.8 Quality assurance and quality control during follow-up	163
A.5.9 Quality assurance and quality control of data management.....	164
A.5.9.1 Quality assurance of data management	164
A.5.9.2 Quality control of data management	166
A.6 PLANNED STATISTICAL ANALYSES AND STATISTICAL POWER CONSIDERATIONS	169
A.6.1 General issues and structure of the section	169
A.6.2 Statistical analyses for the National Cohort	170
A.6.2.1 General methods for analyzing different outcomes.....	170
A.6.2.2 Approaches to multicenter analysis	171
A.6.2.3 Special statistical methods	171
A.6.2.4 Risk prediction models.....	175
A.6.2.5 Statistical analyses for studies embedded into the National Cohort ..	175
A.6.3 Basic population profile and expected numbers of incident chronic disease cases in the National Cohort.....	176
A.6.4 Statistical power considerations of the National Cohort	180
A.6.4.1 General remarks on power calculations in cohort studies	180

A.6.4.2 Minimally detectable odds ratios in main effects models	181
A.6.4.3 Interaction effects between genetic and nongenetic ("environmental") risk factors	184
A.6.4.4 Effects of random measurement errors and use of repeat measurements	185
A.6.4.5 Statistical power for studies of sensitivity and specificity of diagnostic markers for early detection	187
A.7 ETHICAL ASPECTS	191
A.7.1 Invitation letters and informed consent	191
A.7.1.1 Information for potential participants	191
A.7.1.2 Informed consent	191
A.7.1.3 Voluntariness and withdrawal	193
A.7.1.4 Risks	193
A.7.1.5 Expense allowance	195
A.7.1.6 Participant insurance	195
A.7.2 Procedures for ethical clearance of National Cohort proposals	196
A.7.2.1 Regulatory framework	196
A.7.2.2 Data protection	197
A.7.2.3 Data handling	197
A.7.2.4 Data access	200
A.7.2.5 Provision of health information to participants	203
A.7.2.6 Communication of study results	204
A.7.2.7 Value for the public and strategy to involve the public	205
A.7.3 Pre-evaluation of the proposed study from an ethical perspective ...	205
B MANAGEMENT OF THE NATIONAL COHORT	207
B.1 Background	207
B.2 Scientific and administrative management	208
B.2.1 Scientific management	209
B.2.2 Legal entity providing the managerial backbone for the National Cohort	214
B.3 Data use and access policies	218

B.4	Career development.....	219
B.5	Project schedule for the National Cohort.....	219
B.6	Costs.....	220
C.1	DESCRIPTION OF THE STUDY CENTERS.....	225
C.1.1	Cluster Bavaria	225
	C.1.1.1 Cluster coordinator	226
	C.1.1.2 Study centers.....	227
C.1.2	Cluster Baden-Wuerttemberg/Saarland.....	231
	C.1.2.1 Cluster coordinator	232
	C.1.2.2 Study centers.....	233
C.1.3	Cluster North Rhine-Westphalia.....	239
	C.1.3.1 Cluster coordinator	240
	C.1.3.2 Study centers.....	240
C.1.4	Cluster Saxony/Saxony-Anhalt	245
	C.1.4.1 Cluster coordinator	246
	C.1.4.2 Study centers.....	246
C.1.5	Cluster Berlin-Brandenburg.....	249
	C.1.5.1 Cluster coordinator	250
	C.1.5.2 Study centers.....	250
C.1.6	Cluster Lower Saxony/Hamburg/Bremen	255
	C.1.6.1 Cluster coordinator	256
	C.1.6.2 Study centers.....	256
C.1.7	Cluster Schleswig-Holstein	263
	C.1.7.1 Cluster coordinator	264
	C.1.7.2 Study center	264
C.1.8	Cluster Mecklenburg-Western Pomerania.....	267
	C.1.8.1 Cluster coordinator	268
	C.1.8.2 Study center	269

C.2 METHODS FOR STATISTICAL POWER AND SAMPLE SIZE CALCULATIONS; SUPPLEMENTARY TABLES AND FIGURES.....	273
C.2.1 Possible effects of self-selection on cumulative disease incidence	273
C.2.2 Minimally detectable odds ratios in main effects models.....	274
C.2.3 Interaction effects between genetic and non-genetic (“environmental”) risk factors.....	280
C.2.4 Sample size requirements for validation/calibration sub-studies	283
C.3 REFERENCES	287

Abbreviations/acronyms

<i>Abbreviation/ acronym</i>	<i>Definition</i>
3D	Three-dimensional
ABI	Ankle-brachial index
ACD	Acid citrate dextrose
AF	(Skin) Autofluorescence
ASI	Anxiety sensitivity index
AGE	Advanced glycation end products
ATC	Anatomical therapeutic chemical classification system (of drugs)
BD 'CPT'	Cell preparation tube (from Becton, Dickinson and Company)
BFI	Big Five Inventory
BfS	Federal Office for Radiation Protection (Bundesamt für Strahlenschutz)
BIPS	Bremen Institute for Prevention Research and Social Medicine (Bremer Institut für Präventionsforschung und Sozialmedizin)
BMBF	Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung)
BMI	Body mass index
BMG	Federal Ministry of Health (Bundesministerium für Gesundheit)
CDC	Centers for Disease Control and Prevention
CES-D	Center for Epidemiologic Studies Depression Scale
CHD	Coronary heart disease
COPD	Chronic obstructive pulmonary disease
COPSOQ	The Copenhagen Psychosocial Questionnaire
CPAR-24	Computer-based physical activity 24-hour recall
CPI	Community periodontal index
CT	Computed tomography
CTA	CT angiography
CVD	Cardiovascular disease
DALY	Disability-adjusted life year
DAS	Disease activity score
DDZ	German Diabetes Center (Deutsches Diabetes-Zentrum)
DFG	German Research Foundation (Deutsche Forschungsgemeinschaft)
DIfE	German Institute for Human Nutrition (Deutsches Institut für Ernährungsforschung)
DKFZ	German Cancer Research Center (Deutsches Krebsforschungszentrum)
DMSO	Dimethylsulfoxide
DSM	Diagnostic and Statistical Manual of Mental Disorders

Abbreviations

DXA	Dual-energy x-ray absorptiometry
DZNE	German Center for Neurodegenerative Diseases
ECG	Electrocardiogram/ electrocardiography
EDCF	Electronic data capture forms
EDTA	Ethylenediaminetetraacetic acid
EEG	Electroencephalogram
EPC	Epidemiological Planning Committee
ERI	Effort reward imbalance
EU	European Union
FEV1	Forced expiratory volume in one second
FPR	False positive rate
FVC	Forced vital capacity
GDR	German Democratic Republic
GEE	Generalized estimating equations
GIF	Generalized impact fraction
GIS	Geographic information system
GWAS	Genome-wide associations studies
HbA1c	Hemoglobin A1c
HFPEF	Heart failure with preserved ejection fraction
HFREF	Heart failure with reduced ejection fraction
HIV	Human immunodeficiency virus
HMGU	Helmholtz Center Munich for Health and the Environment (Helmholtz Zentrum München – Deutsches Forschungszentrum für Gesundheit und Umwelt)
HZI	Helmholtz Center for Infection Research (Helmholtz-Zentrum für Infektionsforschung)
IAB	Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung)
IADL	Instrumental activities of daily living
ICD	International classification of diseases
IDOM	Instrument for database-assisted online recording for medication
IPA	Institute for Prevention and Occupational Medicine (Institut für Prävention und Arbeitsmedizin der Deutschen Gesetzlichen Unfallversicherung)
IUF	Leibniz Research Institute for Environmental Medicine
IT	Information technology
JEM	Job exposure matrix
LADA	Latent autoimmune diabetes in adults
LIMS	Laboratory Information Management System
MDC	Max-Delbrück Center for Molecular Medicine (Max-Delbrück-Centrum für Molekulare Medizin)

MDOR	Minimally detectable odds ratio
MI	Myocardial infarction
MNC	Mononuclear cells
MRA	Magnetic resonance angiography
MRI	Magnetic resonance imaging
MRS	Magnetic resonance spectroscopy
NCT	National Center for Tumor Diseases, Heidelberg (Nationales Centrum für Tumorerkrankungen, Heidelberg)
NO	Nitric oxide
NIH	US National Institutes of Health
OGTT	Oral glucose tolerance test
OHIP	Oral Health Impact Profile questionnaire
PAD	Peripheral arterial disease
PAR	Population attributable risk
PBMC	Peripheral blood mononuclear cells
PR	Public relations
PST	Plasma separation tube
PTCA	Percutaneous transluminal coronary angioplasty
QALY	Quality-adjusted life year
QM	Quality management
RKI	Robert Koch Institute
RLS	Restless legs syndrome
RNA	Ribonucleic acid
RR	Relative risk
SAQ	Self-assessment questionnaire
SES	Socioeconomic status
SOPs	Standard operating procedures
SSK	German Radiation Protection Commission (Deutsche Strahlenschutzkommission)
SST	Serum separation tube
TICS	Trier Inventory for Chronic Stress
TMF	Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.
TPR	True positive rate
UBA	The Federal Environment Agency (Umweltbundesamt)
VIF	Variance inflation factor

Study Acronyms

ARIC study	Atherosclerosis Risk in Communities study
CARLA study	Cardiovascular Disease, Living and Ageing in Halle study

Abbreviations

CHS	Cardiovascular Health Study
EPIC study	European Prospective Investigation into Cancer and Nutrition
FINRISK study	FINRISK study
Heinz-Nixdorf Recall Study	Risk Factors, Evaluation of Coronary Calcification, and Lifestyle Study sponsored by the Heinz-Nixdorf Foundation
HAPIEE study	Health, Alcohol, and Psychosocial Factors in Eastern Europe study
KORA study	Cooperative Health Research in the Region Augsburg
LIFE	Leipzig Interdisciplinary Research Cluster of Genetic Factors, Clinical Phenotypes and Environments
MONICA	Monitoring of trends and determinants in cardiovascular disease project
NHANES	National Health and Nutrition Examination Survey
PLCO cohort	Prostate, Lung, Colorectal, and Ovarian Cancer - PLCO Trial Cohort
SCORE Project	Systematic Coronary Risk Evaluation project
SHIP	Study of Health in Pomerania

Magnetic resonance imaging abbreviations/acronyms

DCE	Delayed contrast enhancement
DTI	Diffusion tensor imaging
DWI	Diffusion-weighted imaging
FLAIR	Fluid-attenuated inversion recovery
fMRI	Functional magnetic resonance imaging
HASTE	Half-Fourier acquisition single-shot turbo spin-echo
MPRAGE	Magnetization prepared rapid gradient echo
MRA	Magnetic resonance angiography
PSIR	Phase-sensitive inversion recovery
SENSE	Sensitivity encoding
STIR	Short tau inversion recovery
SWI	Susceptibility-weighted imaging
T1-w	T1-weighted
T2-w	T2-weighted
TOF	Time-of-flight
TrueFISP	True fast imaging with steady precession
TSE	Turbo spin echo
VIBE	Volumetric interpolated breath-hold examination
WB	Whole-body

A SCIENTIFIC CONCEPT

A.1 Introduction and overview

A.1

A.1.1 Societal and economic burden of chronic disease in Germany

Over the last 30 years, life expectancy in Germany has continuously increased. Cancer and cardiovascular disease (CVD) remain by far the most important causes of death, accounting for approximately 70% of all deaths. Due to the demographic changes in Germany and other Western civilized populations, the importance of chronic diseases will considerably increase over the next 30 years. Major chronic diseases for which the numbers of patients are growing include diabetes mellitus, stroke, chronic respiratory disease, cancer, and neurologic or psychiatric disorders such as dementia and depression¹. These major chronic diseases constitute a large burden on the national health care system, increasing costs for curative care as well as for nursing care due to disease-related disability. Recent estimates indicate for Germany annual direct costs for curative care of 35 billion Euros for CVD, 5.6 billion Euros for diabetes, 15 billion Euros for cancer, and similar expenditures for the treatment of neurologic and psychiatric diseases. Together, these costs represent a major proportion of the overall health care expenditures in Germany, which are approximately 240 billion Euros per year. In addition, these diseases represent major causes of losses in working capacity and days of work lost, causing considerable further indirect costs to the German economy (Gesundheitsberichtserstattung des Bundes <http://www.gbe-bund.de/>).

While life-prolonging prevention strategies do not necessarily reduce cumulative lifetime costs for health care, health economic analyses do indicate that disease prevention rather than curative care represents a more cost-effective means to reduce morbidity and mortality from most common chronic diseases^{2, 3}.

An additionally important societal and economic aspect of chronic disease in Germany is that health status varies significantly across socioeconomic strata as defined by level of formal education, employment status, income, membership to immigrant groups, and number of health care providers per inhabitant⁴. As a consequence, there is substantial geographic variation in health status. An important trend from a societal perspective in Germany and in many other Western European countries is an increasing social disparity, with increasing poverty, on the one hand, and increasing prosperity, on the other.

A.1.2 Prospective cohorts as an indispensable tool for health research

Effective strategies for the prevention of chronic diseases require accurate data about the causes of these diseases and about the potential magnitude of decrease in chronic disease occurrence by avoiding major risk factors. Prospective cohorts represent the optimal design for epidemiologic studies on the causes of chronic diseases. By employing prospective studies the combined effects of lifestyle, environment, and genetic predisposition can be comprehensively, validly, and reliably quantified for a variety of health outcomes or combinations of diseases ("multimorbidity"). They avoid the recall and selection biases that can seriously distort findings in case-control studies, and they make use of unbiased assessments of risk factors based on biomarker concentrations or medical examinations before these are affected by later stages of disease development and treatments, and thus guard against reverse causation biases. Furthermore, through repeated interviews, blood collections, and medical examinations, information can be collected cumulatively over time from all study participants with regard to lifestyle and other risk factors, metabolic status,

and early pathophysiological responses and chronic disease outcomes with such a study design, Finally, prospective studies are indispensable for investigating risk determinants for overall and cause-specific mortality and other conditions that are either associated with a high fatality risk, such as MI, or with memory losses and poor responses to questionnaires or personal interviews, such as cognitive impairment and dementia.

A.1.3 The National Cohort: Overview of study design

The National Cohort will include a total of 200,000 women and men in the age range of 20 to 69 years (40,000 individuals under age 40 and 160,000 individuals over age 40 years), who will be recruited through a network of 18 local study centers (Figure 1.1) spread all over North, East, South, West, and central Germany (Level 1). In a representative subcohort of 40,000 women and men (i.e., a 20% subsample, distributed equally over all 18 study centers), an intensified study protocol is foreseen, which will include a series of additional, more in-depth medical examinations. (Level 2). In a further subcohort of 40,000 women and men, comprehensive whole-body, heart and brain magnetic resonance imaging (MRI) will be applied (Figure 1.2).



Figure 1.1: Recruitment clusters and study centers of the National Cohort

The local study centers are organized in eight geographic clusters throughout Germany, covering the population of almost all federal states. At each of the centers, study participants will be drawn from local municipal population registries within defined strata of age and gender, and covering metropolitan, urban, and rural regions. Eligible subjects will be invited by letter, with written reminders and, when needed, further contacts by telephone. An overall study participation proportion of at least 50% will be targeted.

As described in detail in **Sect. A.3.1**, study participants will be invited to a local study center; data collection will include a computer-assisted personal interview, computer-aided questionnaires, several physical and medical examinations, and biosampling. Information will be collected at different levels of intensity:

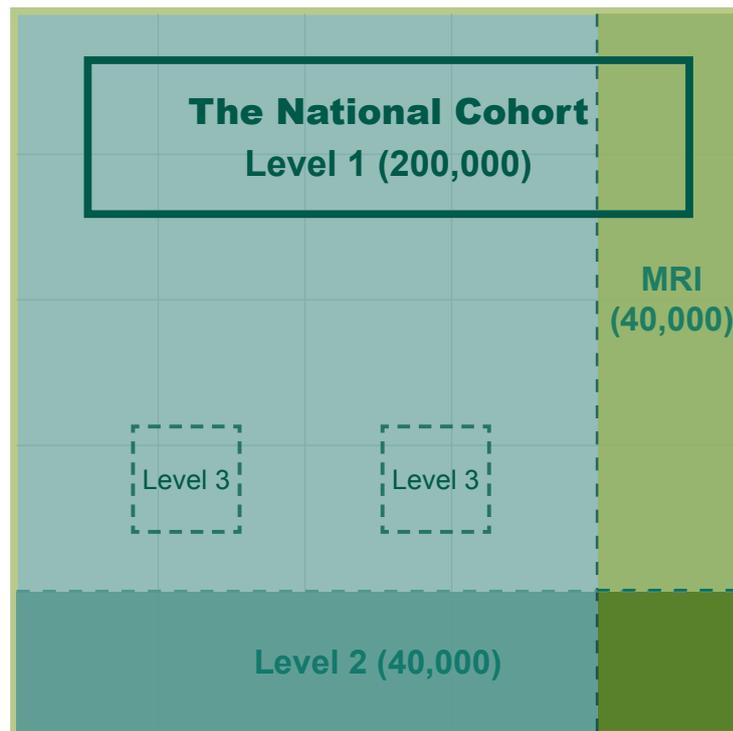


Figure 1.2: *The National Cohort will be organized on several levels*

Level 1: 200,000 participants in all study centers; Level 2: 40,000 participants in all study centers; MRI: 40,000 participants in 4 study centers; Level 3: possible other activities with separate funding

Baseline Assessment

- ▶ At the **basic study level (“Level 1”)**, applying to all 200,000 study participants, baseline interviews, questionnaire and examinations will be performed by using a 2.5-h examination protocol. Physical and medical examinations will include measurements of anthropometric indices (height, weight, waist, and hip circumferences), a dual-energy x-ray absorptiometry (DXA) of body composition and bone density, blood pressure measurements, measurements of arterial stiffness, an electrocardiogram (ECG), three-dimensional echocardiography (3D-ECG), spirometry, measurements of physical activity through accelerometry (7 days), a basic test of physical fitness, and an assessment of cognitive function.
- ▶ A further key element of the core study protocol (Level 1) is the collection of biomaterials from all study participants. A blood sampling protocol will be used that leaves open the greatest possible choice of future laboratory analyses, and includes the collection and storage of intact mononuclear leukocytes. For each study subject, aliquots of plasma, serum, dimethylsulfoxide (DMSO)-preserved viable white blood cells (in a

subgroup), erythrocytes, urine, saliva, nasal swabs, and stool samples will be stored in a central biorepository with decentralized back-up storage. For further details see **Sect. A.3.5, Table 3.12**).

- ▶ At a **second study level (“Level 2”)**, applying to a 20% representative subsample of the cohort (40,000 subjects) proportionally spread over all participating study centers and age strata, an intensified and extended interview and examination protocol will be employed for in total 4 h that, in addition to the general examinations of Level 1, includes an oral glucose tolerance test (OGTT), carotid sonography with measurement of intima–media thickness, 24-h ECG, sleep apnea assessments, enhanced motoric, sensoric and cognitive function tests, and additional questionnaires.
- ▶ A whole-body **MRI program** constitutes another intensified study level, which will also involve 40,000 subjects in four selected study centers. Whole-body MRI will generate a comprehensive morphologic and functional database to be used to estimate the prevalence and incidence of MRI-based findings, regional distribution, and changes over time. This database will be used to identify novel morphologic MRI-based markers of risk for various disease states and subclinical morphologic and functional changes. It should be noted that the participants of “Level 2” (which are examined in all 18 centers) will only partly overlap with the participants of the MRI program, who for technical and financial reasons can only be examined in 4 centers. This means that only 20% of the MRI participants (i.e., 8,000) will be “Level 2” participants at the same time.
- ▶ In addition, add-on **“Level 3”** studies that require further in-depth phenotyping or disease ascertainment may be planned at some or all of the study centers, provided that the additional examinations do not interfere with the Level 1 and Level 2 programs, respectively. Level 3 studies are further conditional on external funding.

Full-scale reassessment after 5 years

All participants of the National Cohort will be reinvited for a second assessment 5 years after recruitment and baseline assessment. The full program of the respective study levels, including the MRI program will then be repeated. At the 5-year reassessment, intraindividual, medium-term changes in risk factors and prospective changes in quantifiable preclinical morbidity characteristics and incident clinical disease can be investigated.

Short-term calibration/reliability substudies

In addition to the medium-term (5-year) repetition of measurements, short-term (1-year) replication studies will be embedded in the cohort, using a 6,000-subject subsample of participants of the first (baseline) visit, and a 4,000-subject subsample of participants in the second visit. These calibration studies will serve to estimate within-subject “random” variations in risk factor measurements in quantitative terms in order to correct for attenuation effects resulting from random misclassification.

Methods for follow-up

As from their first enrollment into the National Cohort, and in addition to the 5-year repeat visit, all cohort participants will be recontacted every 2–3 years and asked to fill in short questionnaires about changes in lifestyle and other characteristics (e.g., use of medications, smoking, menopausal status, and selected disease symptoms) and about the occurrence of selected, major disease (“active” follow-up). In parallel, a mortality follow-up will be performed every 2 to 3 years.

For the follow-up on cancer occurrence, it is also anticipated that increasing use can be made of systematic record linkage with existing cancer registries, which are currently being developed and extended to state-wide coverage in Germany (“passive” follow-up). For all incident cases of major (comparatively frequent) forms of cancer, it is foreseen that tumor samples will be systematically collected and stored in a centralized tumor bank.

Biobanking

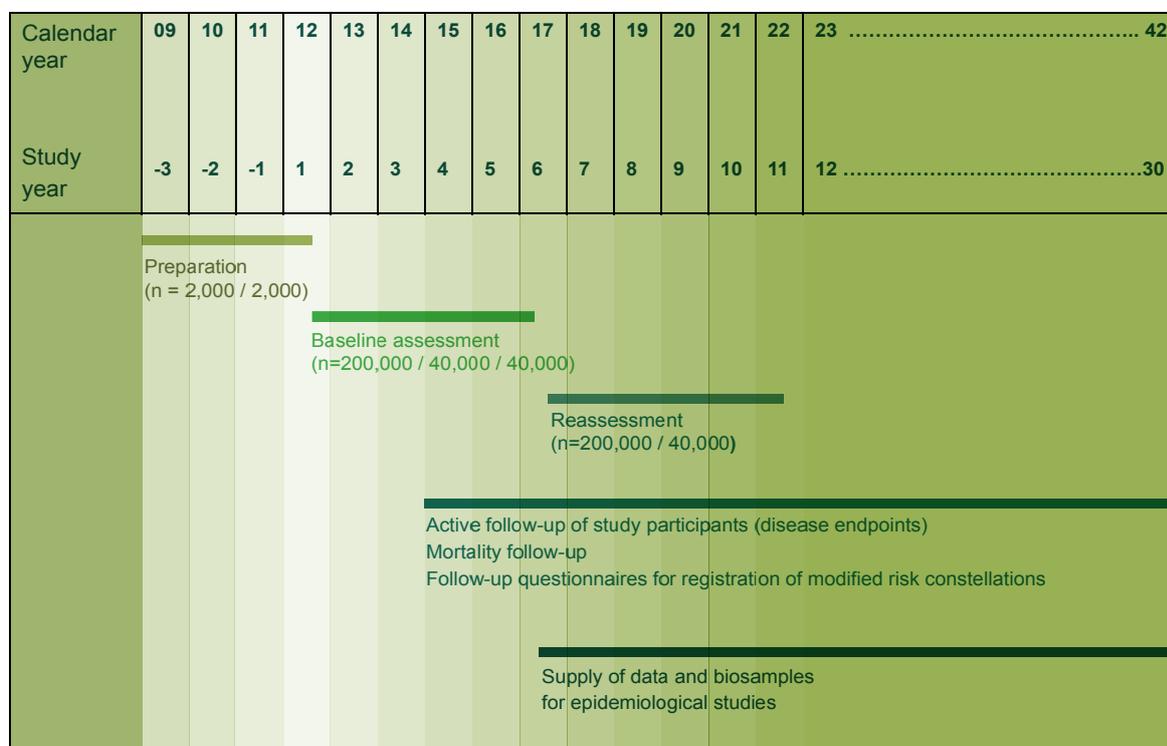
The major part of the biomaterials collected during recruitment and re-assessment will be stored in a centralized biobank of the National Cohort which will be established at the Helmholtz Zentrum München, and will include a fully automated bio-repository. Decentralized back-up storage is planned for one third of the biosamples at the recruitment clusters. For all incident cases of major (comparatively frequent) forms of cancer, it is foreseen to systematically collect tumour samples, which will be stored in a centralized tumour bank.

Ethical aspects and data protection

The National Cohort will be conducted in accordance with the Federal Data Protection Act and all other applicable legislation and will take into account advice from the German Ethics Council and other ethically relevant guidelines, to maximally ensure the confidentiality of collected data and samples, and to carefully design informed consent procedures. The concept on ethics and data protection of the National Cohort has been developed and discussed with the corresponding advisory boards and authorities. Before the start of the recruitment, the study protocol will undergo a final full review by the respective responsible data protection commissioners and ethics committees.

Finally, as a supplement, the cohort will take advantage of the increasing availability of secondary data sources. For example, data from health insurance databases will be accessed to document the use of medical care and to obtain other information about disease status.

Figure 1.3: National Cohort – time schedule: preparation (planning, pretest and pilot phase), 3 years (already funded); recruitment/baseline assessment, reassessment, active follow-up, 10 years (this application); continued active follow-up, supply of data and biosamples for epidemiologic studies for up to 20 years (funding needed after 2022)



Time schedule:

- ▶ The preparatory phase is ongoing, which includes planning, the feasibility study, and the pilot study (see **Figure 1.3**).
- ▶ This application refers to the period of 10 years from mid-2012 to mid-2022 (Recruitment, reexamination, and active follow-up).
- ▶ Thereafter, we anticipate a period of up to 20 years in which data and biosamples will be made available for research projects.
- ▶ Research is possible after recruitment/baseline assessment is finished (except smaller methodological studies which may start earlier). However, research on the cohort-related questions will be possible only after the reassessment and the first active follow-ups are finished.

Funding:

- ▶ Funding for the preparatory phase is covered by Helmholtz association and BMBF.
- ▶ For the current application funding of 210 million Euros is needed for 10 years (for details see **Sect. B.6**).
- ▶ For follow-up of study participants and for the supply of data and biosamples after 2022, additional funding of approx. 5 million Euros per year is needed (for details see **Sect. B.6**).
- ▶ Funding of research: The National Cohort is a resource for epidemiologic and clinical research. This means that for research projects based on this resource separate funding must be found. Here, the state and governmental funding of universities and Helmholtz Centers and standard funding organizations such as the Deutsche Forschungsgemeinschaft (DFG), Federal Ministry of Research (BMBF), Federal Ministry of Health (BMG), the Federal States within Germany, and international funders such as the European Union (EU), US National Institutes of Health (NIH), etc, and industry shall be mentioned. Experience with existing cohorts shows that there are realistic opportunities to receive funding from these organizations for research as soon as the National Cohort is established.

A.1.4 Scientific aims of the National Cohort

The National Cohort is designed to address research questions concerning a wide range of possible causes of major chronic diseases. To this end, the collection of biomaterials in combination with extensive information from questionnaires and medical examinations represents one of the central components. Biomarker measurements will cover domains as diverse as nutrition and metabolism, hormonal and other physiological effects of chronic psychological stress, viral and bacterial infections, immune status, and genetics and epigenetics.

The objectives spanning the studies to be conducted within the National Cohort will provide a sound basis for establishing improved and more targeted measures for the primary and secondary prevention of major diseases, tailored to the German population. In the following, we define general objectives of the National Cohort, and within their general scope we emphasize more specific areas of research that will be of particular focus.

General objective 1: Identification of etiological pathways from lifestyle and environmental risk factors to major chronic diseases and functional impairments.

Within the scope of this first general objective, one set of more specific aims is to increase understanding of the role of particular risk factors in the development of the following major forms of chronic diseases:

- ▶ Cardiovascular diseases
- ▶ Diabetes mellitus
- ▶ Cancer
- ▶ Neurologic and psychiatric diseases
- ▶ Respiratory diseases
- ▶ Infectious diseases

The high incidence of these diseases is largely the result of lifestyle factors such as smoking, alcohol consumption, nutrition, or lack of physical activity, as well as other risk factors such as chronic infections, occupational and environmental influences, child-bearing patterns, and hormone use for contraception and for treatment of postmenopausal symptoms. However, many questions are still open concerning the specificity with which particular risk factors – and their interactions – cause disease, the pathomechanisms, and the quantitative importance of risk factors in terms of etiological and population-attributable fractions. In the National Cohort, therefore, a major emphasis will be placed on the impact of lifestyle and environmental factors as possible causes of disease, with particular interest in the following domains:

- ▶ Body composition, physical fitness, and related metabolic factors
- ▶ Physical activity
- ▶ Diet and nutrition
- ▶ Psychosocial factors
- ▶ Viral and bacterial infections
- ▶ Immune status and immune senescence

For the major disease outcomes, **table 1.1** provides a series of specific study aims which are based on our current knowledge. Many of these can be addressed by the National Cohort in ways that are not possible with currently ongoing prospective studies.

Table 1.1: Selected specific study aims, related to major disease outcomes of interest. These are examples – based on our current knowledge – demonstrating the scientific potential of the National Cohort.

Cardiovascular disease (CVD)

- ▶ To assess the relations of physical activity, physical fitness, and cardiorespiratory fitness to subclinical and clinical CVD; to provide more precise estimates of the relative risk (RR) and attributable fraction of physical activity and physical fitness in relation to CVD; and to better quantify the relative contributions of physical activity and cardiorespiratory fitness to understudied cardiovascular outcomes, such as cerebrovascular diseases, subtypes of stroke, or peripheral arterial disease
- ▶ To assess the burden of subclinical left ventricular diastolic and systolic dysfunction in the general population in Germany; to evaluate the role of subclinical left ventricular dysfunction for incident cardiac disease; and to identify determinants of early left ventricular dysfunction and its progression to overt clinical disease
- ▶ To assess the associations between cardiac dysfunction and subclinical vascular disease, cognitive function, and neurodegenerative disease
- ▶ To improve risk predictions for subclinical and clinical CVD by identifying novel

markers of exposure and by adding new factors to existing risk scores, including genetic and nongenetic exposures and markers of subclinical disease

- ▶ To advance our knowledge concerning the determinants of regional and ethnic differences in CVD incidence
- ▶ To investigate the relations of genetic and epigenetic factors and gene–environment interactions to subclinical and clinical CVD

Diabetes mellitus

- ▶ To advance our knowledge concerning the mechanistic links between adiposity, dyslipidemia, chronic inflammation, and glucose and insulin metabolism disorders, taking into account the modifying effects of genetic, behavioral (diet, exercise, chronic stress, sleep, etc.) and sociocultural factors
- ▶ To study the association and time sequence between pathophysiological changes in glucose and insulin metabolism, development of overt type 2 diabetes and the subsequent occurrence of comorbidity, in particular mental health/cognitive function and susceptibility to infectious diseases
- ▶ To study age-cohort effects of type 2 diabetes incidence, type 2 diabetes progression, and the development of complications
- ▶ To develop simple clinical markers and risk scores for type 2 diabetes for Germany with an emphasis on specific subpopulations beyond age and gender (e.g., birth cohorts and subjects with distinct cultural/ethnic backgrounds)
- ▶ To advance our knowledge concerning the determinants (including metabolic profiles, social/physical environment, and health-related behavior) of regional and ethnic differences in diabetes incidence
- ▶ To gain further insight into the sociodemographic and cultural determinants of diabetes-relevant/diabetes-specific medical management and health care services utilization

Cancer

- ▶ To quantify cancer risks associated with objectively assessed physical activity patterns, including sedentary behavior
- ▶ To examine biological mechanisms and causal pathways through which physical inactivity, adiposity, and reduced physical fitness may favor cancer development
- ▶ To examine the relationships of diet (nutritional biomarkers) with cancer risks
- ▶ To examine relationships of infectious agents, microflora, and immune factors with cancer risk
- ▶ To examine the relationships of established and novel cancer risk factors, in general, with histologic and molecular subtypes of frequent cancers
- ▶ To explore the incidence and persistence of precursor conditions for hematolymphoid malignancies; to explore early epigenetic alterations in specific cell populations; and to investigate whether those alterations contribute to premalignant stages of cancer
- ▶ To develop multi-variable statistical models for prediction of cancer risks

Neurologic and psychiatric diseases

- ▶ To study the incidence of depression and anxiety and to examine biological mechanisms and causal pathways through which physical activity, physical fitness, and other lifestyle factors may favor onset of the two conditions

- ▶ To determine the incidence of subjective memory complaints, mild cognitive impairment and cognitive decline and to examine the relations between physical activity, physical fitness and other lifestyle factors with change in cognitive function
- ▶ To analyze the influence of personality, perceived stress, and social factors on cognitive and emotional functions and on change in these functions over time
- ▶ To investigate the relations between genetic and epigenetic factors, personality, and cognitive and emotional function
- ▶ To use MRI to assess the associations between total and regional brain volumes, alterations in microstructural integrity, vascular brain lesions, subclinical atherosclerosis, and functional connectivity networks and cognitive and emotional functioning, and dementia
- ▶ To determine the incidence of migraine and restless legs syndrome (RLS), to analyze the relations of physical activity, physical fitness, and cardiorespiratory fitness to these two diseases and to examine potential common biological mechanisms and causal pathways for the onset of RLS, migraine, and depression

Respiratory diseases

- ▶ To investigate spirometric lung function in relation to lung disorders (diagnosed or nondiagnosed), nonpulmonary disorders, and overall mortality
- ▶ To evaluate the impact of the environmental burden from air pollution and occupational exposures on remodeling and repair of lung function
- ▶ To quantify the incidence and severity of allergic airway inflammation as found in allergic asthma and allergic rhinitis by measuring exhaled nitric oxide
- ▶ To assess the relations between genetic and epigenetic factors and gene–environment interactions for early, subclinical, and clinical stages of respiratory diseases, particularly for chronic obstructive pulmonary disease, differentiate between the final disease outcome, i.e., chronic bronchitis and emphysema, and identify associated biomarkers of disease
- ▶ To study the predictive value of loss of lung function for subclinical and clinical respiratory diseases, biological aging, and longevity and to determine their relations to established and novel disease markers and genetic and nongenetic susceptibility factors
- ▶ To determine the association between the time sequence of lung function decline, cardiovascular disease, and other comorbidity with disease progression
- ▶ To assess the interdependence of physical activity and physical fitness, mental health, lifestyle factors, and the development of respiratory function with aging and with progression of lung diseases

Infectious diseases

- ▶ To determine the incidence and burden of selected infections and their impact on chronic diseases, emphasizing chronic viral infections (e.g., human cytomegalovirus)
- ▶ To determine the prevalence and incidence of selected infections, including respiratory, gastrointestinal, and zoonotic infections, and their risk factors in the general population
- ▶ To determine risk factors and the dynamics of acquisition and spread of multidrug-resistant bacteria (methicillin-resistant *S. aureus*, extended-spectrum beta-lactamase-producing bacteria, and vancomycin-resistant enterococci) in the general population

- ▶ To examine the effects of microbial populations (“microbiomes”) of the anterior nares, saliva, and stool on the occurrence of acute and chronic diseases
- ▶ To identify lifestyle, nutritional, environmental, and genetic factors that modify the risk of immune dysfunction and immune senescence that can be used for recommendations towards maintaining a healthy immune system and preventing immune senescence

Finally, a major and overall methodological objective for the National Cohort in relation to risk factor identification will be to *improve quantitative estimates* of relative and attributable risks, by collecting *repeated measurements over time*^{5, 6}. In many ongoing studies, the strength of the relationships of key risk factors with disease may have been substantially biased by not having sufficiently accounted for within-subject variations over time. A less than full appreciation of the quantitative importance of specific risk factors may cause erroneous priority setting for prevention. In the National Cohort, the repeated collection of measurements and biomaterials both over the short term, in calibration substudies, and over the medium term through the 5-year repeat visit of all study participants, will enable us to quantitatively estimate RRs more accurately (see also **Sect. A.6**, on statistical methods).

General objective 2: Studies of the geographic and socioeconomic disparities in health status and disease risks in Germany and possible causes and explanations.

Within Germany health status varies significantly across socioeconomic strata as defined by level of formal education, employment status, income, or belonging to immigrant groups, and, in part related to this, substantial geographic variations in health status can also be observed. For example, the incidence of MI is substantially higher and, especially among men, life expectancy lower in East than in West Germany^{7, 8}. However, on a smaller geographic scale, too, there can be strong (sub)regional differences in chronic disease burden. This is the case, for example, in Saarland compared to its neighboring West German Federal States, or within smaller regions of the Ruhr area^{9, 10}. The geographic variations in health status and life expectancy correlate significantly with socioeconomic indicators, such as regional unemployment rates, average per capita income, or numbers of medical doctors per inhabitants¹¹. The disparity in Germany between population subgroups suffering from deprivation or living in relative prosperity is still increasing. Against this general background, a major general objective for the National Cohort will be to contribute detailed descriptions of differences in health status between socioeconomically and geographically defined population strata and also to improve our understanding of these differences, considering age and gender as important effect modifiers. In addition to socioeconomic descriptors, the environmental, psychosocial, and behavioral characteristics will be studied as key determinants for social disparities in health. In addition, data will be collected on the use of general health services and on specialized care, hospital care, and medical interventions.

Table 1.2: Selected study aims for socioeconomic and geographic disparities

Socioeconomic, gender, and geographic disparities

- ▶ To examine how and why important disease risk factors are differentially distributed in East and West Germany
- ▶ To identify how important gender differences are with respect to risk factors and incidence of the target diseases
- ▶ To assess the risk profiles, health status, medical management, and health care in socioeconomically disadvantaged subgroups (e.g., unemployed)
- ▶ To understand whether the difference in life expectancy between East and West Germany is due to differences in lifestyle – and if so, which – or if the health care system is responsible and to what extent

- ▶ To assess what the consequences for the burden of lung disease in the population of some counties in Lower Saxony are (these counties have the highest density of livestock in the world)
- ▶ To quantify the influence and the occurrence of respiratory diseases and CVD close to high-traffic roads (Germany has a very high density of diesel-powered vehicles)

General objective 3: Development of risk prediction models for identifying individuals at increased risk of developing major chronic diseases, so as to allow personalized prevention strategies

Risk prediction models and algorithms for stratifying individuals into categories of lower and higher risk of developing major chronic diseases can be important tools for establishing personalized medicine and personalized prevention strategies¹²⁻¹⁴. Such models can be based on risk factor information obtained by questionnaires or from clinical measurements and assessments of genetic and other biological markers in blood and urine samples. Well-known examples of risk prediction models include the Framingham risk score or the scoring system from the “SCORE” Project¹⁵⁻¹⁷ for CVD, the FINRISK and German Diabetes Risk and Scores for diabetes¹⁸⁻²⁰ or the model by Gail²¹ for breast cancer risk prediction, extended with genetic data²². In addition, although these prediction models work acceptably in identifying subgroups with different risks, they do not predict individual risk well. Current risk scoring systems, however, still leave substantial uncertainties, in particular for persons with intermediate risk²³. These scoring systems could be improved by including more differentiated diagnoses of disease (e.g., for more specific types of CVD outcomes or for different molecular subtypes of breast cancer), by including novel clinical and preclinical risk factors (e.g., based on modern medical imaging), and by increasingly including biochemical and genetic risk factors^{13, 14, 24, 25}. Recent studies on the inclusion of genetic information into risk prediction models have shown considerable promise for diabetes, CVD, and cancer²⁶⁻²⁸. The contribution of genetic information to risk scoring systems will very likely further improve in view of the rapidly growing list of susceptibility genes/alleles that are being identified through genome-wide association studies. The National Cohort will provide an outstanding resource for studies aiming to develop and/or validate comprehensive risk models that integrate polygenic risk scores with nongenetic risk factor information. The substantial sample size of the National Cohort with its large anticipated numbers of endpoints for common diseases will make it possible to evaluate major additive and interactive effects of multiple risk factors on disease outcomes.

General objective 4: Evaluation of candidate biomarkers for early detection of disease and predisease phenotypes, so as to develop effective tools for disease prevention.

Stimulated by recent technological advances in the fields of genomics, transcriptomics, proteomics, metabolomics, and epigenomics, research on biomarkers for early detection of chronic disease or predisease states is becoming a major theme for prevention. Detecting chronic disease or subclinical phenotypes at an earlier stage of development may increase chances of an effective cure, and/or minimize side-effects of treatment. Early disease detection may thus decrease mortality, significantly improve the quality of life of affected patients, and reduce the costs and burdens to the national health care systems. At this stage, it is vital to ensure that the discoveries made from basic research are efficiently and rapidly connected and validated in appropriately designed human studies.

Large-scale prospective cohort studies with highly standardized and valid phenotyping and a broader spectrum of biomaterials of high quality can provide a valuable resource for evaluating candidate biomarkers for the early detection of disease. Major recent examples of blood- and urine-based biomarkers tested through prospective biobank studies have been found for a variety of chronic diseases^{7, 29}. By performing validation studies within pro-

spective cohorts we can examine, using the existing biobank resources, whether a given biomarker (or combination of markers) can predict disease occurrence ahead of its “natural” date of diagnosis (estimation of “lead time”).

To guarantee a high quality of phenotyping at the molecular level for future studies, a comprehensive array of biomaterials must be available in optimal quality. To achieve this, preanalytical artifacts that could incur during specimen collection, primary processing, transport, and storage of the samples need to be minimized. Therefore, all particular components of full blood must be promptly and completely separated, ideally within 1 h of collection, there should be no delay in preparing and freezing aliquots, and volumes should be small to guarantee single use as opposed to repeat thaw-freeze cycles. The high quality of the National Cohort Biobank will be assured by processing the biomaterials locally, not centrally, in a highly standardized manner and automating all steps in preparation, storage, and retrieval of stored materials. For details, see **Sect. A.3.5**.

A.1.5 Arguments for a large cohort in Germany

Several smaller prospective cohort studies already exist in Germany, and in other European countries several large-scale prospective cohorts are currently being set up or have been implemented. It thus may be useful to explain the rationale for planning a new, large cohort study in this country.

Added value in comparison to existing German cohorts

Several medium-sized epidemiologic cohort studies are currently ongoing in Germany, covering a total of 100,000–120,000 individuals (see **Sect. A.3.9**). Not all of these are representative for the general population, however. Many of the studies were planned independently of each other and have different study elements. Thus, only for some of the existing German cohorts could pooled data analyses be conducted, and then for only part of the data collected, addressing rather basic research questions.

- ▶ One first added value of the National Cohort in comparison to existing German cohorts is that it will have a homogeneous study protocol for a very large sample size; thus, research questions can be addressed for which existing German cohorts are statistically underpowered.
- ▶ The participants in most existing cohorts in Germany already have a median age above 50 years (in some of the larger cohorts [e.g., EPIC] even above 60 years) and cannot be used to study the early development of disease.
- ▶ The existing cohorts were initiated up to 25 years ago and the requirements for exposure assessment, including blood sample ascertainment, are now new.
- ▶ The National Cohort will provide a population-representative, highly standardized, comprehensive database that covers much of the heterogeneity both with respect to risk factors and major diseases in all of Germany.

Added value of the National Cohort in comparison to other European cohorts

Several large epidemiologic cohort studies exist or are being planned in other European countries (e.g., the UK, Netherlands, France, Sweden). As summarized in more detail in **Sect. A.3.9** and **Tables 3.16** and **3.17**, the National Cohort will have several advantageous characteristics as compared with other large European cohorts:

- ▶ High quality of biosamples by highly standardized preanalytical processing and automated storage

- ▶ Collection and storage of a diverse range of additional biosamples, such as living cells, saliva, nasal swabs, and faeces
- ▶ Truly prospective design with reassessment after 5 years for all 200,000 participants
- ▶ Very detailed functional characterization and examination of intermediate phenotypes (body composition, cardiovascular examinations, preclinical diabetes, specific lung function measurements, musculoskeletal examinations, and oral health)
- ▶ Implementation of MRI techniques (e.g., MRI of brain, heart, and whole body)
- ▶ Compatibility with existing cohort studies in Germany since the National Cohort questionnaires are derived to a large extent from validated instruments developed in the existing cohorts (e.g., the MONICA-derived cohorts, SHIP and KORA)

In addition, compatibility of questionnaires and medical examinations with those employed in other European studies is planned to the greatest extent possible. For example, elements from the questionnaires used in the UK Biobank and the Swedish LifeGene study will also be used in the National Cohort. Examinations such as step test, eye and ear test modules, and imaging protocols for MRI from UK Biobank will be adapted. Furthermore, the concept and technical details of our biorepository have been developed in close contact with UK Biobank and LifeGene, and are also discussed with the French CONSTANCES study.

Capacity building for epidemiology in Germany

Compared to other Western countries, the epidemiologic infrastructure and capacity in Germany is relatively sparse for a country of 80 million inhabitants. There are historical reasons for this and the situation has only gradually improved in recent years. The National Cohort will play an important role as an epidemiologic resource to improve this situation. However, the specific organization of research in Germany has to be taken into account.

- ▶ **Universities:** In Germany, epidemiologic research is conducted to a major extent at universities, which are typically state-funded. The German Federal States have expressed interest in contributing funds to the National Cohort with the intention of strengthening epidemiologic and biomedical research capacities at their respective universities. In addition, in a current call for applications, BMBF is offering support to expand research in the fields of clinical and population-based epidemiologic research and health services research at German universities. This call directly refers to establishment and use of the National Cohort (<http://www.bmbf.de/foerderungen/15192.php>).
- ▶ **Helmholtz Centers:** Epidemiologic research is also carried out at Helmholtz Centers (which are funded mainly by BMBF). In 2009, the Helmholtz Association established an epidemiologic research program with direct links to the National Cohort at three centers (Max-Delbrück Center for Molecular Medicine, MDC; German Center for Neurodegenerative Diseases, DZNE; and Helmholtz Center for Infection Biology, HZI) in addition to two centers (German Cancer Research Center, DKFZ, and Helmholtz Center Munich for Health and the Environment, HMGU) where epidemiologic research was established 25 years ago.
- ▶ **Leibniz Institutes:** The Leibniz Institutes of the life science section are compared to the Helmholtz centers small state and federal funded internationally competitive research institutes devoted to a specific topic. Epidemiologic expertise has been accumulated in some of these institutes like DIfE (German Institute for Nutrition Research), IUF (Institute for Environmental Medicine, and DDZ (German Institute for Diabetes Research) since the 1990s, and they participate in the field work of the National Cohort.

- ▶ German Centers for Health Research: BMBF has recently initiated German Centers for Health Research. Their purpose is to assemble and unify scientific competence, to close research gaps, and to improve diagnosis, therapy, and – ultimately – prevention. One further task of these centers is to conduct research in epidemiology, health economics, and health services. German centers for neurodegenerative diseases and diabetes research have already been implemented; other centers for lung research, cardiovascular research, translational cancer research, and infectious disease research are being established in 2011. Typically, one coordinating site and four to seven partner sites will collaborate in the German Centers for Health Research. Most designated recruitment sites of the National Cohort are situated at the same locations as the German Centers of Health Research (see **Table 1.3**). Therefore, important synergisms between the National Cohort and the German Centers are to be expected and represent an important strategic advantage for capacity building for epidemiologic research in Germany.

Table 1.3: Recruitment clusters and study centers of the National Cohort (left) and the sites of the German Centers for Health Research (right)

Clusters/partner sites of the National Cohort	Sites of German Centers for Health Research*
Muenchen, Augsburg, Regensburg	Muenchen (incl. Augsburg, Regensburg)
Heidelberg, Freiburg, Mannheim, Saarbruecken	Heidelberg, Freiburg, Tuebingen
Essen, Duesseldorf, Muenster	Essen, Düsseldorf, Witten
Leipzig, Halle	Dresden
Berlin, Potsdam	Berlin, Potsdam
Brunswick, Hanover, Hamburg, Goettingen, Bremen	Brunswick, Hanover, Hamburg, Goettingen
Luebeck, Kiel	Luebeck, Borstel, Kiel
Greifswald, Neubrandenburg	Greifswald, Rostock

*Only sites matching with the partner sites of the National Cohort are listed.

This scientific infrastructure has specific implications for the organization and governance of the National Cohort (see **Part B** for details)

- ▶ Recruitment: We propose to establish 18 study centers for the National Cohort that will serve as local/regional epidemiologic resources and which all provide infrastructure and expertise for specific types of biomedical research.
- ▶ Important organizational infrastructures of the study, such as the biorepository, data management, central executive office (Geschäftsstelle), and internal and external quality control are affiliated with university or Helmholtz sites.
- ▶ Governance: The National Cohort will be organized as a registered association (eingetragener Verein) in which universities, Leibniz institutes and Helmholtz centers are members. The general assembly/epidemiologic steering committee (Mitgliederversammlung) of the association elects the board of directors (Vorstand) as governing/management body (geschäftsführendes Organ) and the head of the central executive office (Geschäftsstelle). Members of the commission of funding partners (Kommission der Zuwendungsgeber) will be BMBF, representatives of the federal states, and Helmholtz. Further elements of the governance structure include the scientific advisory board (wissenschaftlicher Beirat) and the ethics advisory board (Ethikbeirat).

Internal boards authorized by the general assembly and the board of directors, as the working group scientific management or the use and access committee will be responsible for particular aspects of the scientific management (details are provided in **Part B**).

A.1.6 Justification of sample size requirements

Whenever an epidemiologic study has one single outcome and one single risk factor, the sample size necessary to establish a statistically stable association between the two is usually based on statistical power considerations, involving subsequent sample size calculations for detectable effect size, given type I, II errors, and making reasonable assumptions on the incidence of the outcome and the distribution of the risk factor. Even in such highly simplified context, other, not trivial aspects usually have to be considered, including the relevance of the detectable effect size and various underlying assumptions with respect to adequacy of statistical models³⁰⁻³².

For a project of the size and breadth of scope of the National Cohort, developing a notion of the necessary study size becomes a much more complex exercise. Clearly, considering the large number of possible disease outcomes and the great variety of risk factors and related study hypotheses with respect to these outcomes, the problem cannot be defined simply and thus does not lead to a single and straightforward solution. In addition, financial costs must also be taken into consideration, for example, with respect to the embedded study levels with intensified phenotyping.

In **Sect. A.6** of this application, issues related to selecting appropriate forms of statistical modeling are being discussed, and detailed statistical power estimations are presented. The arguments and estimates laid out in that section indicate that N=200,000 for level 1, N=40,000 for level 2, and N=40,000 for the MRI program, and N=8,000 for the combined measurements of level 2 and imaging (as justified in **Sect. A.3.4**) are reasonable sample sizes for the National Cohort. Here, we illustrate this in only simplified form by highlighting the sensitivity of the study in different incidence scenarios. This simplicity results from the simple model we consider here, namely, that of a quantitative trait that is divided into quartiles, where the measure of sensitivity is expressed as the RR that may be detected by contrasting the 4th versus the 1st quartile of the trait for a significance level of 5%, power of 80%, two-sided test, given a certain number of events in the full cohort (Level 1).

Table 1.4 gives an impression of how the detectable RR's are interrelated with the number in the different subcohorts of the National Cohort.

Table 1.4: Minimum detectable odds ratio (statistical power 0.80, significance level 0.05) for comparison of top to bottom quartiles of a quantitative trait, in full cohort setting.

Study level (number of subjects)	Expected number of cases at level 1						
	100	500	1,000	2,500	3,500	5,000	10,000
Level 1 (N=200,000)	2.1	1.4	1.3	1.17	1.14	1.11	1.08
Level 2 or imaging (N=40,000)	3.6	1.9	1.7	1.5	1.3	1.3	1.17
Level 2 plus imaging (N=8000)	9.5	3.6	2.7	1.9	1.7	1.6	1.4

Practical examples for such associations to be examined are:

- ▶ The handgrip strength (Level 1) and the risk of subsequent MI (approximately 2,000 cases at Level 1 within 5 years)
- ▶ The outcome of the step test (Level 2) and MI (approximately 400 cases at Level 2 within 5 years)
- ▶ The volume of the pancreas (imaging) and subsequent diabetes mellitus type 2 (approximately 6,800 cases at Level 1 within 5 years)
- ▶ A combined index of volume of the pancreas and the OGTT (both Level 2 and imaging) and diabetes (approximately 16,000 cases at Level 1 within 10 years)

Table 1.4 shows that, at least for the study examples mentioned above, the intended study size is sufficient, but also necessary, to detect excess risks in the range of 20 to 40% in the full cohort and the different subcohorts. More elaborate estimates of statistical power are presented in **Sect. A.6.4**.

Aside from these examples, **Table 1.4** gives a robust impression of what may be achieved in terms of epidemiologic findings for increasing time under observation. This is highlighted by the fact that after 20 years of follow-up, approximately 23% (47,000) of the original cohort participants will have died.

A.1.7 Outlook: The National Cohort as basis for collaborations and further research projects

The planning phase of the National Cohort has already stimulated several related research proposals and collaborations. They are based on, or developed in parallel with the National Cohort but will be funded and/or governed independently (**Figure 1.4**).

Planned studies, based on the National Cohort: For the following additional cohorts planning has been started and/or feasibility studies are ongoing:

- ▶ **Migrant Cohort:** Germany is among the European countries with the highest proportions of recent immigrants/migrants. The two largest immigrant groups in Germany are the Turks (about 3.5 million) and migrants from the former Soviet Union (more than 2 million individuals). The concept of a Migrant Cohort of 25,000 migrants in selected recruitment areas of the National Cohort (Berlin, Ruhr Area, Mannheim/Ludwigshafen) has already been developed and currently is being tested in a feasibility study funded by BMBF. Here, different recruitment schemes (using social and religious groups, migrant organizations, local meeting places, social services, etc.) are being assessed and existing tools in Turkish and Russian languages are used. Since it was not possible to include this Migrant Cohort in the National Cohort due to financial restrictions, we plan to apply for separate funding. There is strong support for the idea of a Migrant Cohort from the Federal Ministry of Health (BMG) and several German Federal States.
- ▶ **Occupational Cohort:** The National Cohort provides a unique opportunity to study occupational health effects. For all study participants, complete (but crude) occupational histories will be obtained by data retrieval from the national Institute for Employment Research (Federal Employment Agency). These data will be linked to job exposure matrices (JEM), to estimate possible past exposures to occupational risk factors (see **Sect. A.2.4.8**). However, these data are not detailed enough to gain a more precise understanding of the role of working conditions for specific health problems. Therefore, an additional subcohort of the National Cohort shall be established in

which these crude occupational histories will be used to screen for working conditions of specific interest. The selected individuals will be contacted separately and a detailed questionnaire or telephone interview will be applied. The funding of the Occupational Cohort is planned by the official occupational health organizations in Germany (Deutsche Gesetzliche Unfallversicherung and Bundesanstalt für Arbeitsschutz und Arbeitsmedizin). A feasibility study of the Institute for Prevention and Occupational Medicine (Institut für Prävention und Arbeitsmedizin der Deutschen Gesetzlichen Unfallversicherung, IPA) and partners from the National Cohort in cooperation with the Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung, IAB) is in preparation.

Cross-sectional research questions based on the National Cohort: We anticipate a high response rate of more than 50% for the recruitment phase and of more than 75% for baseline assessment and re-assessment (see **Sect. A.3.6**) Furthermore, a nonresponder analysis is planned. In addition, we have the opportunity to calibrate the National Cohort according to representative surveys, see below. In total, this will enable us to address many cross-sectional research questions. Two activities are already being tested in feasibility studies:

- ▶ **Use of Health Monitoring of the Robert Koch Institute to improve representativeness for Germany:** The Robert Koch Institute (RKI) carries out representative surveys for the population of Germany at regular intervals based on their own funding. These data can be used for comparison and especially for calibrating the National Cohort according to this representative sample. To utilize fully this possibility, important elements of the questionnaires of RKI will be integrated in the questionnaires of the National Cohort.
- ▶ **Ionizing radiation:** The National Cohort provides a unique opportunity to determine personal exposure to ionizing radiation in Germany. Therefore, the German Radiation Protection Commission (SSK) supports the idea of using the National Cohort to better characterize the spatial distribution of exposure. To assess cumulative medical exposure, data from health insurance companies could be used to get information on the type and frequency of applied diagnostic and therapeutic devices using ionizing radiation. To test the feasibility of this approach, a pilot study funded by the Federal Office for Radiation Protection (BfS) is ongoing. Furthermore, in a second feasibility study bracelet dosimeters will be worn by 100-200 participants for 1–2 months to measure background radiation (see **Sect. A.3.3.11**).
- ▶ **Zoonoses:** Approximately 20% of the Cohort participants are expected to own household pets. The goal is to study the impact of pathogens that are shared by humans and animals on human health. Pets and farm animals may represent community reservoirs of antimicrobial-resistant bacteria. All participants of the National Cohort will answer screening questions regarding past and present exposure to household pets and farm animals. Individuals reporting a significant exposure history will then be selected for targeted substudies. The ability of the owners to obtain nasal swabs and stool samples from their cats or dogs, and their willingness to have a veterinarian draw blood from their pets will be tested in 2011 in a feasibility study headed by the University of Veterinary Medicine, Hannover. (see **Sect. A.3.3.11**)

Other German population studies, developed in parallel with the National Cohort

- ▶ **Neuro Cohort Bonn:** The German Center for Neurodegenerative Diseases (DZNE) and the University of Bonn have started to establish a prospective cohort to understand the causes and preclinical signs of age-dependent changes in the brain and

their clinical consequences (n=30,000, 30–80 years). This Neuro Cohort Bonn and the National Cohort will be complementary both in design and concerning their scientific questions. To facilitate future collaboration, Bonn aims to harmonize the data standards and standard operating procedures (SOP) with those of the National Cohort and other national and international ongoing cohort studies. Future data and biosamples of the Neuro Cohort will be made available for collaborative research. The DZNE was recently established in the Helmholtz Association with population research into neurodegenerative diseases as one of its core tasks. In this regard, the Neuro Cohort Bonn and the National Cohort are independent but closely interacting partners who will form a strategic alliance for population research in Germany.

- ▶ **German Environmental Health Birth Cohort:** A birth/pregnancy cohort of up to 200,000 newborns is currently planned on behalf of the Federal Environmental Agency (Umweltbundesamt, UBA). One attractive approach under discussion is to use participants of the National Cohort at reproductive age to recruit mothers during pregnancy and to include their newborns into the birth cohort. This recruitment could take place continuously during the 10-year fieldwork period of the National Cohort. The numbers of recruited pregnant women/newborns could be increased by using the infrastructure of the National Cohort to draw a larger random sample in the appropriate age range. Furthermore, it would widen the scope from pregnancy/birth to women who want to become pregnant. In 2009 a planning project was contracted by the UBA. The final report is presently being prepared.
- ▶ **LIFE study in Leipzig:** Objective of LIFE is the investigation of molecular causes of life style and environment associated diseases to develop new processes, products and services for diagnostics and prevention of common life style diseases. LIFE includes population-based studies in adults (n=10,000, 40–80 years) and children (n=5,000, 0–18 years) as well as seven patient cohorts (total n=17,000). The adult cohort is being established in close interaction with the National Cohort.
- ▶ **Clinical research group Goettingen:** At Goettingen university, a prospective cohort will be established, based on a DFG-funded clinical research group. It includes a population cohort using the baseline protocol of the National Cohort (n=3,000, 20–69 years), and a patient cohort (n=3,000, psychotic patients, 20–69 years).
- ▶ **Patient cohorts of German Centers for Health Research:** German Centers for Health Research are being established for lung research, cardiovascular research, translational cancer research, and infectious disease research and have already been implemented for diabetes research and neurodegenerative disorders (see above). Important synergism between the National Cohort and the German Centers (i) should improve our understanding of early development of the diseases of interest, and (ii) the National Cohort can be used as a control cohort for patient cohorts which are being established.

International collaborations and coordination:

- ▶ **Large European Cohorts:** As mentioned above in **Sect. A.1.5** and described in detail in **Sect. A.3.9**, compatibility of questionnaires, medical examinations and other study elements of the National Cohort (n=200,000, 20–69 years), with other large European Cohorts is planned, especially with UK Biobank (n=500,000, 40–69 years), the Swedish LifeGene study (n=500,000, 0–45 years), and the French CONSTANCES study (n=200,000, 18–69 years).
- ▶ **The National Cohort as blueprint for satellite cohort studies in neighbouring German speaking regions:** for example, regions of Northern Italy (South Tyrol), regions in Austria, Luxemburg, and Switzerland. It is planned or being discussed to establish new cohorts (n around 100,000) with study protocols that to a very large extent will be standardized with those of the National Cohort (see **Sect. A.3.8**).

A.1.8 Preparatory work for the National Cohort

Since 2009, the planning and preparation of the National Cohort has been underway. The most important activities are summarized in **Table 1.5**, for details see Part B.

Table 1.5: *Preparatory activities for the National Cohort*

Activity	Time
General	
Funding of pilot phase within Helmholtz (after international review)	2008-13
Preparatory workshop with universities in Heidelberg (220 participants, 180 abstracts)	2008
Epidemiologic planning committee (EPC) established (members from Helmholtz and universities, guests)	2009
Project management team (PMT) established (at Helmholtz)	2009
Thematic working groups established (17 topics, over 230 members, 3/4 from universities)	2009
Decision on clusters and study centers for field work (after international review)	2009
Ad hoc working group on governance and financing (between BMBF and Federal States)	2010
Funding of feasibility study for universities by BMBF (complementary to funding within Helmholtz)	2010-12
Ethical clearance for feasibility study from ethics committee Munich	2010
Discussion of draft of data safety concept with federal and state data protection officers (in Wiesbaden)	2010
Level 1, level 2	
Several planning visits of UK Biobank (at Imperial College, field work in London)	2010
Planning visit of LifeGene (at Karolinska Stockholm)	2009
Planning visit from CONSTANCES (at HMGU Munich)	2010
Planning visits for biorepository (at UK Biobank Manchester, Karolinska Stockholm, several providers)	2009-10
MRI program	
Several planning visits of UK Biobank (at Imperial College London)	2010
International planning workshop (at Helmholtz Berlin)	2010

A.1

Activity	Time
Level 3 and other potential future activities	
Planning meetings for occupational subcohort (with occupational health administrations and BAUA in St. Augustin and in Munich)	2010
Contacts for subcohort of migrants (potential funding agencies)	2010
Planning visits for potential international satellites of the National Cohort (Bozen, Luxembourg, Zürich, Salzburg, Tartu)	2009-11
Several planning meetings for environmental birth cohort (at UBA Berlin, Essen)	2010
Planning contacts with German Centers for Health Research	2010-11

A.2 Scientific background and rationale for study elements

A.2.1 Introduction

For the planning of the National Cohort, several different perspectives have been taken into consideration in deciding upon its specific research emphases and choice of study elements, regarding both the specific diseases to be studied and risk factor information to be collected.

From a **public health perspective**, clearly the National Cohort should focus on diseases and conditions that are major causes of death and healthy years of life lost and that constitute a high burden on the German health care systems in terms of cost. Likewise, on the risk determinant side, the National Cohort should naturally focus on identifying factors that are likely to explain large proportions of these various major diseases and that in large part could represent common risk factors for diseases in different major categories. From a **societal perspective**, the cohort should focus on diseases and health problems that will increase as a result of the aging population structure. In addition, the National Cohort should address the topic of social disparities as a possible causal factor of health inequalities, in view of the increasing gap between individuals suffering from deprivation, on the one hand, and individuals living in relative prosperity, on the other.

From a person's own **individual perspective**, an important consideration for identifying the disease groups to be studied in the National Cohort is that there is substantial evidence that disease occurrence can potentially be influenced, i.e., disease risk increased or decreased or disease occurrence accelerated or delayed, through one's behavior. Many prior studies have summarized such health-related behaviors within the term "lifestyle", which includes physical activity, alcohol consumption, smoking, and diet. This classic concept of lifestyle, however, does not fully and sufficiently reflect the reality of life today; in particular, psychosocial factors, such as personality, social networking, and perceived stress at work or at home need to be added to this classic concept.

From an **epidemiologic perspective**, research should focus on questions that truly require a long-term, prospective approach in order to be answered properly. This excludes diseases and conditions which are rare and could be examined better in a register or case-control study design, or research questions that could be addressed by cross-sectional studies. Finally, from both an **epidemiologic and clinical perspective**, such a long-term project must allow for innovations that can be transferred to prevention and/or to clinical settings.

Integrating these different perspectives the following general decisions were reached:

- ▶ For disease outcomes, the core focus of the study will be on *CVD, diabetes, cancer, neurologic and psychiatric diseases, respiratory diseases, infectious diseases, and musculoskeletal diseases*, as these disease groups best reflect the defined perspectives.
- ▶ Intensive use will be made of diagnostic work-up methods known from clinical settings and scores previously validated for other purposes to assess preclinical and clinical phenotypes, with an emphasis on methods that can be used over a long, prospective study period.
- ▶ For risk determinants, particular emphasis will be placed on areas in which major, common risk factors can be identified for diverse types of chronic disease and on modern tools for determining risk factors; these main areas include *lifestyle-related risk determinants* (diet, physical activity, excess body weight, and related metabolic and physical fitness conditions), *psychosocial factors, chronic infections, immune factors, and bacterial microflora*. In addition, the cohort will address questions in relation to *occupational and environmental exposures*.

Table 2.1 gives an overview of the major disease entities, intermediate phenotypes and measurements of function, and exposures and risk factors that will be in the focus of the National Cohort.

Throughout this text, the term “risk factor” will be used to designate a factor that can potentially cause disease and can clearly be identified to come first in the time sequence of disease development; thus, it is a factor distant from disease. The term risk factor will be extended by specific qualifiers, such as “socioeconomic”, “psychosocial”, “environmental”, “behavioral”, “metabolic”, or “genetic” to distinguish between categories of risk factors. A factor more proximal from disease, known to be caused by other risk factors while itself causing disease, will be termed “*intermediate factor*” since it represents an intermediate state in the time sequence and biological model of the development of the given disease. Disease states within one group must be clearly defined according to clinical symptoms or the results of diagnostic work-up since they might comprise diverse pathological entities of different or only partially overlapping etiologies. Of particular interest in this context are pre-clinical or subclinical phenotypes, which usually can be considered as intermediate factors that ultimately cause manifest diseases. In the multicausal disease model, various conditions have the potential to be a risk factor, an intermediate factor, or a disease state, e.g., hypertension or hypercholesterolemia, depending upon the association of interest. In such case it will be clearly stated whether the specific factor is considered an intermediate factor or a disease state in the analysis.

Finally, we introduce the term “function” to the multicausal model of disease development to describe the ability of an individual, e.g., cognitive, psychosocial, or emotional function, that is important for his or her living reality and related to health behaviors. Functions reflect the whole continuum from normal to pathologic state, including different severity levels or a disease status based on new findings.

Table 2.1: *The National Cohort: Study design in brief, major disease entities, functional intermediate phenotypes, and exposures and risk factors*

Study design in brief:

- ▶ Population-based prospective cohort (men and women)
- ▶ Age range 20–69 years
- ▶ Random sample of inhabitants of defined geographical areas
- ▶ 18 study centers
- ▶ Level 1, n=200,000; Level 2, n=40,000; Level 3, n=variable (additional research questions with additional funding possible; not part of this application); whole-body MRI program, n=40,000
- ▶ Baseline assessment including face-to-face interview, questionnaire modules, physical examinations, testing of cognitive functions, and sampling of biomaterials (blood, urine, saliva, nasal swabs, stool)
- ▶ 2.5-h assessment program at Level 1
- ▶ 4-h intensified assessment program at Level 2
- ▶ 5 years of recruitment, followed by 5 years of reassessment
- ▶ Combination of active follow-up (mailed questionnaires, each 2–3 years) and passive follow-up (linkage with registries)

Major diseases:

- ▶ CVD
- ▶ Diabetes mellitus
- ▶ Cancer
- ▶ Neurologic and psychiatric diseases
- ▶ Respiratory diseases
- ▶ Infectious diseases

Intermediate phenotypes and measurements of function:

- ▶ CVD: Subclinical atherosclerosis (arterial stiffness, ankle–brachial index, intima–media thickness of the carotid artery, brain MRI), cardiac dysfunction (ECG, 3D-Echocardiography, MRI), elevated blood pressure
- ▶ Diabetes: Impaired fasting glucose, impaired glucose tolerance (fasting glucose, oral glucose tolerance test), accumulation of advanced glycation end products (skin autofluorescence), retinopathy (retinal photographs)
- ▶ Cancer: precursor stages of haematologic malignancies
- ▶ Neurologic and psychiatric diseases: mild cognitive impairment (cognitive functioning tests), olfactory function (smell test), brain MRI
- ▶ Respiratory diseases: lung function (spirometry), airway inflammation (exhaled nitric oxide FeNO), lung volume (MRI)
- ▶ Musculoskeletal functions and phenotypes: osteoarthritis (MRI) and rheumatoid arthritis (clinical examination, biomarker measurement), musculoskeletal pain disorders (pain mannequin), osteoporosis (DXA)
- ▶ MRI-based measurements of intermediate/subclinical phenotypes for diseases of large organs and metabolic diseases

Major exposures and risk factors:

- ▶ Body composition
- ▶ Physical activity
- ▶ Physical fitness
- ▶ Diet
- ▶ Smoking and alcohol consumption
- ▶ Psychosocial factors
- ▶ Socioeconomic status
- ▶ Sleep-related characteristics
- ▶ Chronic infections, immune factors, and microflora
- ▶ Occupational and environmental exposures

A.2.2 Major disease outcomes of core focus

In the following, a brief overview is given of the major diseases to be studied in the National Cohort, and the methods for prospective investigation and characterization (examinations, questionnaires, active follow-up, and passive follow-up) are outlined for each of the disease entities. In case of death, death certificates will be used to identify the underlying cause of death. Unless otherwise specified, the term "questionnaire" used in the box refers to the assessment of physician-diagnosed disease (at baseline or during follow-up) as reported by the study participant. For some diseases, however, questionnaires are used to assess symptoms. Self-reported diseases during follow-up will be medically verified using different sources of medical information (for further information, see **Sect. A.3.7**). Analytical data from biosamples are also required but are not mentioned in full here.

A.2.2.1 Cardiovascular diseases

CVD are among the most common chronic diseases in developed countries, and they are the leading cause of death worldwide, accounting for approximately one third of all deaths in men and women³³. In Germany, they account largely for the 20 most frequent hospital diagnoses³⁴ and are responsible for the largest proportion of direct health care costs (35 billion Euros yearly)³⁵.

Among CVD, **coronary heart disease** (CHD) is the single leading cause of death in Germany and worldwide^{34, 36, 37}. In the year 2009, 73,899 persons died from chronic ischemic heart disease (8.6% of all deaths in Germany), and an additional 56,226 died from acute MI (6.6% of all deaths)³⁴. The age-standardized incidence of MI in Germany in the years 2001–2003 was estimated at 356 events for men and 109 events for women per 100,000 persons per year³⁸. CHD is a multifactorial disease resulting from atherosclerotic narrowing of the coronary arteries and the formation of an occlusive thrombus after plaque rupture. Although a number of genetic and nongenetic risk factors for CHD have been identified, the quantitative interplay of these factors remains poorly understood, and the prediction of CHD is still relatively imprecise.

Assessment of CHD:

Examinations and questionnaires (at baseline and during reassessment)

- Level 1: Questionnaire
- 10-s, 12-lead ECG

Active follow-up (and medical verification of self-reports):

- Self-report of physician-diagnosed MI, angina pectoris, coronary artery bypass graft (CABG), percutaneous transluminal angioplasty (PTCA)

Passive follow-up: MI registries, where available

Heart failure is one of the most prevalent forms of CVD, with prevalence and incidence increasing in the aging populations of industrialized nations. The prevalence of heart failure was estimated to increase from 1% in 55- to 64-year-olds, to 7% in 75- to 84-year-olds, and to >10% in >85-year-old subjects (Rotterdam study)³⁹. Incidence in industrialized nations increase from about 0.2-2.5/1000 person years in the 45- to 55-year-old group to 12.4-44/1000 person years in >85-year-old persons (Hillingdon study and Rotterdam study)³⁹, and the lifetime risk of heart failure in >40-year-old persons is >20%³⁹. Data from the Framingham study show 5-year survival rates of 25-40%^{40, 41}.

While MI, hypertension, valvular heart disease, and cardiomyopathy are among the main underlying causes of heart failure, little is known about the role of inflammatory factors and autonomic dysfunction and about the exact role of risk factors such as physical activity, obesity, or dietary factors. Moreover, heart failure can be divided into heart failure with preserved ejection fraction (HFPEF), i.e., impaired left ventricular diastolic function with normal systolic left ventricular function (previously denominated diastolic heart failure), and heart failure with reduced ejection fraction (HFREF, systolic heart failure only⁴²), i. e., with impaired systolic function. HFPEF currently accounts for about 50% of all cases of heart failure⁴³; it is associated with the same morbidity and mortality and is even responsible for a larger percentage of hospitalizations than is HFREF⁴⁴. However, only few population-based studies have described the burden of diastolic dysfunction and HFPEF^{45, 46}. The distinction between HFPEF and HFREF seems especially important since they seem to be partially different entities potentially requiring different approaches to early detection and prevention⁴⁷. However, little is known about the determinants of diastolic dysfunction and its progression to HFPEF. Further, the relationship between other measures of subclinical CVD (such as arterial stiffness) and left ventricular diastolic function has not been described in detail. Like-

wise, knowledge of the possible effects of cardiac dysfunction on cognitive function and risk of neurodegenerative disease is scarce.

There is a need to better understand the determinants underlying development and progression of diastolic dysfunction in HFPEF in order to address the epidemic of heart failure in the population. Identifying modifiable risk factors of HFPEF (in order to improve prevention) and improved models for predicting progression from subclinical stages of left ventricular cardiac diastolic dysfunction to clinically overt heart failure are important aims of the National Cohort. It is also important to help disentangle possible different pathophysiological mechanisms underlying HFPEF and HFREF and to investigate specific causes of death in subjects with HFPEF versus subjects with HFREF (e.g., sudden cardiac death, arrhythmia-related deaths, and other causes of death). In the National Cohort, heart failure and cardiac (dys)function at baseline will be assessed by symptoms, questionnaires, and 3D-Echocardiography, supplemented by cardiac MRI, according to scientific recommendations / guidelines⁴³, ESC Task Force⁴⁸.

Assessment of heart failure:

Examinations and questionnaires (at baseline and during reassessment)

- Level 1: Questionnaires,
3D-Echocardiography
- MRI program: Cardiac MRI

Active follow-up (and medical verification of self-reports):

- Self-report of physician-diagnosed heart failure

Atrial fibrillation is the most common cardiac arrhythmia in developed countries, with an overall prevalence of 1-2% in the general population and a marked age-dependent increase, with a remaining life-time risk of about 25% in 40-year-old subjects⁴⁹. The age-adjusted incidence and prevalence of atrial fibrillation seem to be increasing in Europe and North America⁵⁰, and the prevalence of atrial fibrillation is expected to more than triple in the next 50 years⁵¹. Atrial fibrillation is associated with significant morbidity and mortality; it is a known cause of “cryptogenic” stroke and is associated with cerebral white matter lesions and cognitive dysfunction^{52, 53}. A number of clinical risk factors for atrial fibrillation have been identified; however, characterization of lifestyle factors, genes, and novel biomarkers for risk of atrial fibrillation is still in its infancy. Furthermore, the information about risk factors for asymptomatic atrial fibrillation and regarding progression from paroxysmal to persistent and permanent atrial fibrillation is insufficient⁵⁰.

Assessment of atrial fibrillation:

Examinations (at baseline and during reassessment)

- Level 1: Questionnaire
10-s, 12-lead ECG
- Level 2: Long-term ECG

Active follow-up (and medical verification of self-reports):

- Self-report of physician-diagnosed arrhythmia (and use of antiarrhythmic drugs)

Cerebrovascular diseases such as ischemic or hemorrhagic stroke pertain both to the group of cardiovascular and neurologic diseases. Owing to their close link to cognitive impairments originating from vascular changes (including vascular dementia and vascular depression), the cerebrovascular diseases will be detailed further in **Sect. A.2.2.4** on neurologic and psychiatric diseases, below. In the National Cohort, we will regard the research

area of cerebrovascular diseases as an interdisciplinary field between the cardiovascular and neurologic-psychiatric disorders.

A.2.2.2 Diabetes mellitus

Diabetes mellitus is a metabolic disorder characterized by chronic hyperglycemia with anomalies in carbohydrate, fat, and protein metabolism caused by defects in insulin secretion, insulin action, or a combination of the two. Metabolic syndrome and type 2 diabetes are important risk factors for the development of vascular complications. Life expectancy is substantially reduced in patients with impaired glucose tolerance and type 2 diabetes, and more than 75% die from macrovascular complications such as MI or stroke (see the section on CVD). Peripheral arterial disease (PAD) and/or diabetic neuropathy are common causes of diabetic gangrene, resulting in more than 35,000 amputations per year in patients with type 2 diabetes in Germany. Diabetes is a major cause of severe microvascular complications such as end-stage renal disease or diabetic retinopathy with the risk of blindness. In addition, diabetes and prediabetic states are also established risk factors for cancers of different sites, including colon, endometrium, and pancreas⁵⁴. The worldwide prevalence of diabetes was estimated to be 220.5 million in 2004⁵⁵ and numbers are projected to increase to 366 million by 2030. In Germany, diabetes mellitus is among the most common diseases with about 6 million patients and an equal number of undiagnosed cases or persons at high risk probably exist^{56, 57}.

Type 2 Diabetes: The by far most frequent form of diabetes is type 2, and this type is the major component of the worldwide diabetes epidemic⁵⁸. Therefore, it is crucial to understand the etiology of type 2 diabetes, and the development of preventive strategies for this disease is a key public health target. Among the major, established causes of type 2 diabetes are excess body weight, lack of physical activity, and an unhealthy diet; intervention trials have shown that in high-risk individuals with impaired glucose tolerance the risk of progression to diabetes can be drastically reduced through lifestyle interventions such as weight reduction, dietary changes, and increased physical activity, or by drug therapy with metformin or acarbose⁵⁹⁻⁶⁴. One of the major challenges that still must be resolved, however, is to determine the best strategies for identifying high-risk individuals and those subjects who would benefit most from specific intervention strategies.

Often, type-2 diabetes takes several decades to develop, including a process with different stages of prediabetes, eventually resulting in a clinical diagnosis. The decisive event at the end of this process is the progressive failure of beta-cell function. The pathological mechanisms leading to diabetes may vary between individuals. According to current knowledge, different pronounced pathomechanisms are relevant and different organs (e.g., liver, fat tissue, heart, or brain) may take a leading role. Based on current research findings in the field of diabetes genetics and the expected results arising from other –omics technologies, especially metabolomics, proteomics, and epigenomics, we will be able to refine the definition of subphenotypes of type 2 diabetes. By identifying such subphenotypes, new ways will be opened to understand the interactions between genetic background, metabolism, and lifestyle factors in the development of type 2 diabetes. In combination with the large sample size and precise data on physical activity, body composition, etc., as well as information on the utilization of health care, the National Cohort will offer unique opportunities for studies on diabetes etiology, individualized diabetes prediction and prevention, treatment of diabetes and diabetes complications, also considering quality-of-life aspects among diabetic patients and economic consequences for the health care system.

Assessment of diabetes mellitus:**Examinations and questionnaires** (at baseline and during reassessment)

- Level 1: Questionnaire
 Level 2: OGTT

Active follow-up (and medical verification of self-reports):

- Self-reports of physician-diagnosed diabetes mellitus (and use of diabetes medication)

A.2

Due to the low incidence of type 1 diabetes (including latent autoimmune diabetes in adults, LADA) and gestational diabetes in the given age range of the National Cohort, we do not plan to analyze them separately. However, it might be important to identify prevalent cases. For example, through repeated blood sampling we can determine serum autoantibodies to islet cell antigens and distinguish between LADA and type 2 diabetes.

A.2.2.3 Cancer

“Cancer” is a collective term for many different neoplastic diseases, for which incidence varies widely by organ site and specific molecular-histologic subtypes. In Germany, cancer is the second major cause of mortality after CVD and is responsible for about one third of all deaths. The cancers with the highest incidence in Germany are tumors of the female breast, colorectum, prostate, and lung, which together account for about half of all tumor occurrences among adults. The other half of cancer occurrences is due to a great diversity of tumor types, of which about seven or eight have an intermediate-level incidence, accounting for about 20% of tumor occurrences overall. For the 10–12 most frequent cancers, by organ site, expected cumulative incidence over the first 10 years of follow-up varies from about 200 cases for tumors of the brain and central nervous system to over 1,800 for cancers of the female breast, prostate, or large bowel.

Over the last two decades, great progress has been made in characterizing tumors from a histologic, molecular, and clinical (prognostic) perspective, and it is increasingly being recognized that many tumors exhibit different molecular subtypes that may be attributed to different etiologies⁶⁵⁻⁷⁶. For example, by using gene expression profiles and immunohistochemical characteristics **breast cancer** can be classified into at least four subtypes, according to hormone receptor expression (negative or positive) or presumed cellular origin (basal or luminal)⁷⁰ [“luminal” type, “HER2-positive”, “basal-like”, “normal-like”]. Likewise, **colorectal cancer** can be classified into at least four tumor subtypes, depending on the presence of widespread methylation of CpG islands that cause inactivation of tumor suppressor genes by promoter methylation [“CpG Island Methylator Phenotype”; CIMP], of genomic “microsatellite instability”, and of BRAF mutations^{68, 69}. **Endometrial tumors** occur in at least two basic forms, generally referred to as type I and type II. Type I carcinomas are associated with mutations in the *ras* proto-oncogene, and in the PTEN tumor suppressor gene and often show microsatellite instability, but do not usually show mutations in the p53 tumor suppressor gene. By contrast, a majority of type II tumors have p53 mutations, but almost never have microsatellite instability or *ras* or PTEN mutations. Although most epidemiologic studies so far have not distinguished between these two tumor types, and in those studies that did the numbers of type II tumors were usually small, there is some evidence that endocrine and nutritional lifestyle factors differentially affect the risks of type I and type II tumors⁷⁷. **Ovarian tumors** occur in at least five distinct histologic forms; however, independently of histology, they can be also be classified into “type I” and “type II” tumors, following molecular evidence showing that ovarian cancer may progress both through a step-wise mutation process (low-grade pathway, type I) and a separate pathway with high genetic instability, leading to rapid metastasis without an identifiable precursor lesion (high-grade

pathway, type II)^{71, 73}. Further important examples of tumor heterogeneity are *endometrial cancers*, which exist in at least two major forms (“type I” and “type II”)^{66, 67}, and *lung cancers*^{65, 72}. For the cancers at each of these organ sites a growing body of evidence indicates that different epidemiologic risk factors may exist for the different molecular subtypes⁷⁸⁻⁸³. For *prostate cancer*, evidence showing that epidemiologic risk factors may vary strongly between low-grade and high-grade tumors is also increasing⁸⁴.

A.2

In addition to lung cancer, which is predominantly caused by smoking, a much higher incidence is observed in Western Europe (including Germany) for most of the other “top twelve” cancer types than in economically less developed parts of the world. This is the case, for example, for cancers of the colon and rectum, pancreas, kidney (renal cell tumors), breast, endometrium, ovary, and prostate. Migrant studies and studies of time trends of incidence in Western Europe clearly indicate that environmental and lifestyle (nongenetic) factors must be key determinants in the development of these and several other tumor types. Interestingly, there are also strong international correlations of the incidence rates of these various cancer types with those of type 2 diabetes, which have also been very much on the rise over the last several decades, and CVD. The latter observations suggest that common environmental and lifestyle factors may underlie various forms of chronic disease.

Established risk factors for different cancer types are manifold and range from reproductive behavior, nutrition-related lifestyle factors, smoking and alcohol consumption, use of hormonal medications, and chronic viral and bacterial infections and vary greatly by type of tumor. Epidemiologic findings so far, however, have only partially identified the specific lifestyle factors and other causes that explain the high risks of the aforementioned cancer types. In the National Cohort, we plan to conduct extensive studies especially in the areas of physical activity, excess body weight and metabolism, diet and nutrition, and chronic infections and immune function, and we also plan to investigate the effects of intestinal and oral microflora on tumor development. Account will be taken of other, established risk factors such as use of exogenous hormones for contraception or postmenopausal replacement therapy, use of other medications, family history of cancer, and smoking and alcohol consumption. Furthermore, a point for special attention in the National Cohort is that of socioeconomic and geographic variations in overall cancer risk, taking account of possible differences in the use made of health care (e.g., screening participation)⁸⁵⁻⁸⁸.

There is increasing evidence that different molecular tumor subtypes may develop along different etiological pathways and may be related to different risk factors. It is therefore of utmost importance to collect detailed *data on tumor stage, grade, histology, and molecular subtypes*. The first data source will be cancer registries; however, due to the rapidly moving field, we will be able to provide additional morphologic and molecular subcharacterization of tumor types by establishing a *national, quality-assured bank of tumor samples* (see **Sect. A.3.5.5**)

Assessment of cancer:

Examinations and questionnaires (at baseline and during reassessment)

Level 1: Questionnaire

Active follow-up (and medical verification of self-reports):

Self-report of physician-diagnosed tumors (in case of incomplete cancer registries)

Insofar as possible in combination with passive follow-up through: Cancer registries

Tumor tissue bank: Systematic collection of tumor materials for all incident cases of cancer so as to allow further molecular and refined histopathologic characterization of tumors.

A.2.2.4 Neurologic and psychiatric diseases

Many neurologic diseases are related to age. Thus, even with incidence remaining stable, they will become more frequent in absolute numbers due to the increasing life expectancy.

Dementia and stroke are examples for this and illustrate two contrasting features of many of these age-related neurologic diseases. They are either chronic conditions with slowly progressing impairments over time, such as dementia, or can also have a sudden onset with severe impairment, such as stroke.

While some neurologic or psychiatric conditions occur too rarely to study them even in a large population such as the National Cohort (e.g., schizophrenia or Huntington’s disease), there are conditions that primarily affect younger or middle-aged adults and cause considerable health care costs, and health impairments. This group of conditions includes migraine and other types of headache, restless legs syndrome, depression, and anxiety.

Cerebrovascular diseases are among the leading causes of death worldwide⁵⁵, and survivors often suffer from lifelong disabilities. The clinical syndrome “cerebrovascular disease” is a heterogeneous term for different conditions to be further classified based on the duration of clinical symptoms, underlying pathology, and underlying cause of ischemia⁸⁹. The spectrum of these conditions ranges from small diffuse changes in the fiber tracts of the white matter, visible lesions in the white and gray matter, microbleeds, and lacunar infarcts to ischemic stroke, hemorrhage, and post-stroke atrophy. The common denominator for the whole spectrum is vascular brain changes. Therefore, cerebrovascular diseases will serve as an ideal link pertaining to both research areas, “cerebrovascular diseases” and “neurologic and psychiatric diseases”. Whereas ischemic stroke subtypes share a number of common risk factors, including nonmodifiable (e.g., age, sex, genetic predisposition, lower socioeconomic status, SES) and modifiable ones (e.g., hypertension, smoking, diabetes, smoking and high alcohol consumption), other factors only increase the risk of specific pathological subtypes, such as atrial fibrillation, carotid artery stenosis, or LDL cholesterol, and some might have a divergent impact on some stroke subtypes, such as moderate alcohol consumption. Thus, the impact of several risk factors on the occurrence of distinct pathological and etiological subtypes still needs to be established.

The phenotyping within the National Cohort, e.g., characterization of the arterial system (ankle–brachial index, arterial stiffness, left and right ventricular function, intima–media thickness [CIMT]) combined with brain MRI data in 40,000 subjects will significantly contribute to our understanding of risk factors for cerebrovascular diseases. In particular, by using MRI, very early-stage vascular changes in the brain can be detected before the onset of any clinical symptoms. In combination with the assessment of cognitive and emotional functions (depression and anxiety scores), replicated over time, the natural course and progression of subclinical cerebrovascular diseases can be explored based on prospective data.

Assessment of CVD:

Examinations and questionnaires (at baseline and during reassessment)

Level 1: Questionnaire

MRI program: Brain MRI

Active follow-up (and medical verification of self-reports):

Self-report or proxy report of physician-diagnosed stroke

Passive follow-up: Stroke registries, where available

Cognitive function, cognitive impairment, and dementia: Cognitive function represents a continuum from normal function on one end through different stages of impairment to the stage of dementia on the other end. For the latter it is estimated that in Germany currently around 2 million people, especially women, suffer from Alzheimer’s disease or other forms of dementia, and it is projected that this number will double over the coming two decades. Current evidence suggests that vascular factors, such as midlife hypertension, diabetes, dietary fat intake, high cholesterol, excessive smoking, alcohol misuse, and obesity contribute

significantly to cognitive impairment and subsequently to the development of dementia, and that active engagement in mental, physical, and social activities may postpone the decline in function and the onset of dementia by providing cognitive reserve⁹⁰. Vascular brain changes are thought to reduce cerebral perfusion and hypoperfusion and have been found to cause oxidative stress, neurodegeneration, and cognitive decline⁹¹. Mild cognitive impairment is defined as an intermediate state between normal cognitive function and dementia and is further described in **Sect. A.2.3.4**.

Assessment of major cognitive impairment and dementia:

Examinations and questionnaires (at baseline and during reassessment)

- Level 1: Questionnaires
Neuropsychological test battery
- MRI program: Brain MRI

Active follow-up (and medical verification of self-reports):

- Self-reports, proxy reports, or reports by the participant's general physician of physician-diagnosed Alzheimer's disease, or dementia

Depression and anxiety: Mood and anxiety disorders represent the most frequent mental illnesses in Germany. They generally develop more often in women than in men. During lifetime, about one of four women and one in ten men are affected at least once by **depression**. Major depression is known to have a high comorbidity with CVD^{92, 93}, metabolic syndrome, and diabetes⁹⁴⁻⁹⁶ and atopic, neurologic, and infectious diseases, where it is judged to be likely that depression is not only a frequent consequence of a somatic disease, but can also be an important, independent risk factor. The pathophysiology of major depression is complex and not well understood. Previous studies suggested that systemic, low-grade inflammation may play a role in the onset and course of disease⁹⁷. These studies showed that major depression is accompanied by biochemical and immune changes, suggesting the presence of either an acute or a chronic inflammatory process⁹⁸⁻¹⁰¹. Important determinants of the inflammatory process are the release of cytokines and alterations in iron metabolism, characterized by decreased serum iron and transferrin levels associated with normal or increased ferritin levels^{102, 103}. With regard to **anxiety disorders**, the 12-month prevalence is as high as 14.5%¹⁰⁴, and anxiety is the most common comorbidity for depression, accounting for about 19% of all occurring comorbidities among mental disorders. Marital status (living alone, being divorced, or widowed) and unemployment seem to be risk factors for anxiety disorders in both genders. Current research in anxiety disorders is also exploring the neurobiology and genetic basis for these conditions, especially of panic attacks.

Assessment of depression and anxiety:

Examinations and questionnaires (at baseline and during reassessment)

- Level 1: Questionnaires (incl. depression and anxiety scores)

Active follow-up:

- Self-report of physician-diagnosed depression or anxiety disorders

Headache: Primary headache disorders, including migraine and tension-type headache, affect millions of people worldwide¹⁰⁵. The burden of headache has consequences for the individual and for society due to a reduced quality of life, absenteeism and lower performance at work, and increased disease-related costs¹⁰⁶. Although acute and prophylactic treatments are available, many frequent headache sufferers either do not seek treatment or do not respond satisfactorily to these treatment strategies¹⁰⁷. Behavior-dependent, so-called lifestyle factors have been of interest in studies that have previously investigated the

onset of migraine and tension-type headache. These studies, however, report inconsistent findings. Certain triggering factors for migraine and for tension-type headache have been identified^{108, 109} and include alcohol consumption and physical activity^{110, 111}. Modifications of lifestyle habits might provide a sufficient, feasible, and cost-effective preventive strategy.

Migraine is a common, primary headache disorder characterized by episodic, severe headache and, in about 25% of cases, transient neurologic symptoms known as migraine aura. It is a disorder of young and mid-age adulthood, and women are affected three to four times as often as men. Migraine, particularly migraine with aura, is associated with an unfavorable cardiovascular risk profile and with an increased risk of major CVD. This risk is different according to aura status. Active migraine with aura is associated with a significantly increased risk of subsequent major cardiovascular events, ischemic stroke, MI, coronary revascularization, angina, and death due to ischemic CVD. In Germany, only few studies have investigated the frequency and impact of different headache types on an affected individual¹¹⁰⁻¹¹³. Tension-type headache is even more common, affecting about one in four to one in three adults in 12 months without any difference by gender. For both headache disorders, migraine and tension-type headache, prospective population-based studies are very rare and analytic risk factor analyses based on incident cases are lacking.

Assessment of headache:

Examinations and questionnaires (at baseline and during reassessment)

Level 1: Questionnaire

Active follow-up:

Self-report of physician-diagnosed migraine and tension-type headache

Restless legs syndrome: Standard criteria for the definition and diagnosis of restless legs syndrome (RLS) were published by the International Restless Legs Syndrome Study Group in 1995¹¹⁴. These criteria include the desire to move the limbs usually associated with paresthesia or dysesthesia of the legs, a motor restlessness, a worsening or exclusive presence of symptoms at rest (lying or sitting) with at least partial and temporary relief by activity and a worsening of the symptoms in the evening or during the night. RLS is a common disease with a prevalence of 8–15% in the general population¹¹⁵, affecting women twice as often as men. RLS is one of the few neurologic disorders that can be assessed by posing specific questions to participants in population-based studies. The etiology of RLS is still not known. Parity is a major factor in explaining the difference between the two genders. Few data on risk factors for RLS in the general population are known and all currently existing data come from cross-sectional or case-control studies. RLS prevalence is related to age and cases have a higher prevalence of self-reported diabetes, a higher body mass index (BMI), perform less exercise, and consume less alcohol than non-cases. Cases with RLS report considerably reduced mental health and more depressed mood, either self-reported or professionally assessed, as well as more social isolation. Further research, especially about the impact of health-related behaviors, e.g., physical activity, alcohol consumption, smoking, and diet and psychosocial factors, on the occurrence of RLS is needed since any potential association would be subject to preventive strategies.

Assessment of RLS:

Examinations and questionnaires (at baseline and during reassessment)

Level 1: Questionnaire

Active follow-up (and medical verification of self-reports):

Self-report of physician-diagnosed RLS

Parkinson's disease: Prevalence and incidence estimations of Parkinson's disease vary widely over European populations. Disease symptoms including tremor, fatigue, pain, depression, and cognitive impairment cause a substantial functional decrease in daily life. The prevalence is estimated to be between 65.6 and 125.0 per 100,000, and annual incidence ranges between 5 and 346 per 100,000¹¹⁶.

A.2

Determinants of Parkinson's disease are largely unknown, but both genetic and environmental factors contribute. Recent genome-wide association studies in European and Asian populations uncovered hitherto unknown genetic loci associated with prevalent Parkinson's disease¹¹⁷. Other risk factors include the onset of an atherosclerotic disease, side-effects of drugs, intoxication, and head injuries. The complex interplay between psychiatric disorders and Parkinson's disease is reflected by a Swedish nation-wide study demonstrating that previous hospitalization due to a psychiatric disorder represents a risk factor for future Parkinson's disease, particularly in patients under the age of 50¹¹⁸.

Assessment of Parkinson's disease:

Examinations and questionnaires (at baseline and during reassessment)

Level 1: Questionnaire

Active follow-up (and medical verification of self-reports):

Self-report of physician-diagnosed Parkinson's disease

A.2.2.5 Respiratory diseases

Chronic lung diseases, including chronic obstructive pulmonary disease (COPD), asthma, pulmonary fibrosis, or lung cancer are the second leading cause of death in the world. Lower respiratory infections and COPD are the third and fourth most common causes of death, respectively, and together cause 7.2 million deaths⁵⁵. Of concern is that death rates due to chronic lung disease continue to increase, while mortality due to other leading causes of death such as heart disease, cancer, or stroke are declining. Predictions based on the demographic change and the age composition of the population in Germany forecast an increasing prevalence of all major lung diseases, with substantial impacts on the health care system¹¹⁹. Currently, only limited therapeutic strategies exist to treat chronic lung disease. In this section we focus on COPD and asthma; respiratory tract infections and lung cancer are included in **Sects. A.2.2.6** (infectious diseases) and **A.2.2.3** (cancer), respectively.

COPD is a devastating disease characterized by irreversible airflow obstruction, airway inflammation, and loss of functional pulmonary tissue. While no causal therapy for COPD is available to date, the avoidance of cigarette smoke exposure, its major risk factor, represents one important strategy in the fight against COPD¹²⁰. However, COPD also develops in former and never-smokers and has been associated with a variety of endogenous and exogenous risk factors, including exposure to dust and air pollution¹²¹. COPD has recently been recognized as a systemic disease with several manifestations, including CVD, diabetes, or skeletal muscle dysfunction¹²². Two main pathological features of COPD are small airway disease and emphysema. Small airway disease, also termed obstructive bronchiolitis, includes airway inflammation with increased mucus production, activation of immune cells, airway wall remodeling, and peribronchiolar fibrosis¹²³. Emphysema is defined as destruction of the alveolar architecture due to distal airspace enlargement. In particular, emphysema destroys functional alveolar epithelium and impairs lung function. The inability of the lung to activate self-repair mechanisms in COPD further contributes to the progressive destruction of functional lung tissue in emphysema. Etiological mechanisms that have been associated with COPD include oxidative stress, inflammation, protease-antiprotease imbalance and apoptosis, impaired growth factor signaling, and (auto)immune mechanisms^{124, 125}.

Assessment of COPD:**Examinations and questionnaires** (at baseline and during reassessment)

Level 1: Questionnaire (incl. symptoms);
Lung function (spirometry)

Active follow-up (and medical verification of self-reports):

Self-report of physician-diagnosed COPD (chronic bronchitis and/or emphysema)

A.2

Asthma is an inflammatory disease of the pulmonary airways, characterized by intermittent airflow limitation and symptoms of wheeze and shortness of breath. Significant numbers of children with asthma have persistent symptoms throughout life, significantly reducing quality of life, but asthma frequently also presents in later life and then is often refractory to treatment. The prevalence of asthma is high, with a chronic, relapsing course, and represents a severe global health problem¹²⁶. In 2004, 300 million people worldwide were estimated to be affected by asthma, and this is projected to rise to 400 million by 2025¹²⁶. Occupational asthma due to workplace exposure to dusts or chemicals is the most prevalent occupational lung disease in the European Union. Furthermore, asthma is a multifaceted disease, and various environmental exposures and genetic backgrounds have been implicated in the pathogenesis of this condition¹²⁷. There is considerable interest in identifying genes related to asthma susceptibility, but progress has been hampered by the complexity and heterogeneity of the disease¹²⁸. The sharp increase in the prevalence of asthma in the last several decades indicates the presence of strong environmental factors underlying disease susceptibility, although risk factors, such as air pollution, smoking, or occupational exposures, have already been identified.

The National Cohort will provide an ideal context for studying the incidence of respiratory diseases in Germany, and for studies of the complex interplay between genetic background and environmental burden for the development and progression of lung diseases. Of special interest, in this regard, are the well-established differences in respiratory health (and allergic diseases) between East and West Germany in both children¹²⁹ and adults¹³⁰.

Assessment of asthma:**Examinations and questionnaires** (at baseline and during reassessment)

Level 1: Questionnaire
Airway obstruction (spirometry)

Level 2: Airway inflammation (FeNO)

Active follow-up (and medical verification of self-reports):

Self-report of physician-diagnosed asthma and allergic asthma

A.2.2.6 Infectious diseases

Incident acute infections will be detected using internationally validated, symptom-based measurement instruments, coupled with modern communication methods as further indicated below ("active surveillance"). Due to the high logistic demands, detection of pathogens during acute infections is only planned for targeted studies at Level 3 with separate funding. Exposures to a variety of pathogens will be determined by future serologic analyses of stored biomaterials, e.g., using array-based approaches to detect immunoglobulins against a large number of pathogens of interest. Such "serologic fingerprints" of the participants, as well as prospective changes in these, can then also be linked with a variety of relevant clinical endpoints. However, funding of serologic analyses is not financed within this application.

Hepatitis B and C virus infection, HIV infection, and tuberculosis will not be studied in the National Cohort due to low and declining incidence.

Recurrent respiratory infections are diseases characterized by high frequency of occurrence and high associated economic costs. An adult on average may suffer from two symptomatic respiratory tract infections per year¹³¹. Risk factors for an increased frequency of respiratory infections include sex¹³², older age¹³³, occupational exposures¹³⁴, animal contact, number and type of social contacts, including children in the household¹³⁵, or lifestyle factors (smoking, physical inactivity) and nutritional status¹³⁶. Genetic factors have been identified in the context of the relatively rare, defined immunodeficiency syndromes of adulthood (e.g., common variable immunodeficiency), but factors determining milder phenotypes in the general population remain largely unknown. We will determine the frequency, type (upper vs. lower respiratory tract), and severity (antibiotic use, work-related absenteeism, physician visits, hospitalization, complications, and need for intensive care) of respiratory infections (not of specific pathogens) and also include these endpoints in the models for changes in immune and pulmonary function over time. A combination of these items will be used to compute a respiratory infection risk score. By applying the score at each follow-up we can prospectively assess susceptibility to respiratory infections. Endpoints will be assessed through a combination of interview and active surveillance with modern communication methods. In the interview, respiratory infections will be defined according to symptom-based, validated measurement instruments.

Assessment of recurrent respiratory infections:

Examinations and questionnaires (at baseline and during reassessment)

- Level 1: Questionnaire; respiratory infection risk score
- Level 3: Active surveillance

Recurrent gastrointestinal infections: A wide range of bacterial (e.g., Campylobacter, Salmonella, Yersinia enterocolitica), viral (e.g., norovirus, rotavirus), and parasitic (e.g., Giardia) pathogens cause acute gastrointestinal infections. These infections are still very common in Germany. For example, about 235,000 cases of acute gastrointestinal infections due to Campylobacter, Salmonella, rotavirus and norovirus were reported to the Robert Koch Institute in 2009¹³⁷, and these notified cases represent a minor portion of the true disease burden as demonstrated in the Netherlands¹³⁸, the UK¹³⁹, and Denmark¹⁴⁰. Specifically, in a Dutch cohort study, the ratio of cases observed by active follow-up as compared to notifications to a national surveillance system was 3:1 for Salmonella, 8:1 for Campylobacter, 35:1 for rotavirus, and 1500:1 for norovirus¹³⁸. Known risk factors for acute gastrointestinal infections include sex, age, urban/rural place of residence, SES, and behavioral factors such as dietary habits or contact with animals. However, factors that determine the course and severity of acute gastrointestinal infections are still largely unknown. There is evidence that genetic host factors determine susceptibility for norovirus infection¹⁴¹. Definitions for syndromic screening questions for acute gastrointestinal infections are well established¹⁴² and will be combined with modern communications tools (e.g., text message recall) for further follow-up. At Level 3 of the cohort, those who report episodes of acute gastrointestinal infections by modern communication tools will be asked to return stool samples by mail. This method has been used successfully in prospective studies in Germany¹⁴³, the UK¹⁴², and the Netherlands¹³⁸.

Assessment of recurrent gastrointestinal infections:

Examinations and questionnaires (at baseline and during reassessment)

- Level 1: Questionnaire
- Level 3: Active surveillance; collection of stool samples during acute infection

Periodontal disease is one of the most prevalent infectious disorders in adults. A probing depth of >4 mm, indicating a current periodontal infection, was found in 76.9% of German adults aged 35–44, and an attachment loss of >3 mm (a consequence of previous infection) in 95% of the subjects, with 69% of teeth being affected. Applying the Center for Disease Control (CDC) definition, the prevalence in this age group was 70.9% and 87.4% in adults aged 65–74¹⁴⁴. Indeed, dental treatment costs due to this condition amount to 16 billion Euros, i.e., ~6% of the total expenditures of the German statutory health insurance funds³⁵. Periodontitis, an opportunistic bacterial infectious disease, causes chronic inflammation of the periodontium, ultimately causing tooth loss (as a clinical disease endpoint). The onset, severity, and distinct clinical phenotype of periodontitis are mainly influenced by individual susceptibility and heritability. Besides age and low SES, major risk factors are male gender, oral hygiene behavior, smoking, obesity, and a genetic predisposition. In addition to tooth loss as its clinical endpoint, periodontitis is associated with systemic diseases such as diabetes, CVD, metabolic syndrome, and arthritis^{145–153}. Therefore, periodontitis should be understood as a chronic infectious disease with considerable impact on overall health.

Assessment of periodontal disease:

Examinations and questionnaires (at baseline and during reassessment)

Level 1:	Questionnaire (periodontitis)
	Tooth count
	Collection of saliva samples
Level 2:	Oral examination

A.2.3 Intermediate (preclinical) phenotypes and function measurements

As outlined above, most of the aforementioned major chronic diseases can be considered as distinct clinical endpoints of a disease continuum that begins earlier in life and advances during a person's life in conjunction with subclinical alterations. These subclinical alterations represent intermediate states in the time sequence and biological model of the development of disease. It is important to assess these intermediate states in a large study such as the National Cohort for several reasons: First, such an assessment may help to better predict hard clinical endpoints, and, thus, to identify persons at risk for major disease outcomes, and some of these factors, particularly those that are downstream on the disease pathway, may also be taken as symptoms for early disease. Secondly, it is likely that the effect of nongenetic and genetic risk factors on future disease outcomes may depend on a person's existing phenotype (effect modification). Here, again, proper assessment of intermediate phenotypes may improve risk prediction towards more personalized prevention. Thirdly, by measuring these intermediate phenotypes repeatedly – as it is planned in the National Cohort – we can study the nature of subclinical disease progression as well as its predictors and its impact on major diseases.

A.2.3.1 Cardiovascular functions and preclinical phenotypes

Subclinical atherosclerosis: Evidence indicates that atherosclerosis begins already at an early age with the accumulation of lipid in the intima of arteries to form fatty streaks¹⁵⁴. This atherosclerotic process results rapidly in (subclinical) changes in the structure and function of the arteries¹⁵⁵. Noninvasive measurements of subclinical atherosclerotic disease increasingly are being used for detection disease at an early stage, thereby allowing early stage treatment and prevention of target-organ damage. Thus, the National Cohort will assess subclinical atherosclerosis and its precursors by measuring arterial stiffness, ankle–brachial

index, pulse wave velocity, intima–media thickness of the carotid artery (CIMT), and other parameters, as outlined earlier. In addition, subclinical atherosclerosis will be identified by non-contrast-enhanced MR angiography, brain MRI, and MR-based carotid imaging (see **Sect. A.3.4**).

Cardiac dysfunction: Asymptomatic alterations in cardiac dimensions, texture, or function typically reflect early signs of organ damage. For example, left ventricular hypertrophy is typically considered as an early sign of hypertensive heart disease in individuals with elevated blood pressure. As outlined above (**Sect. A2.2.1**, heart failure), systolic and diastolic dysfunction are not only observed in patients with heart failure but frequently also in individuals not showing overt clinical signs or symptoms of heart failure, and are associated with a higher risk of heart failure and higher mortality¹⁵⁶. In the National Cohort we will therefore assess cardiac dimensions, texture, and function using 3D-echocardiography and cardiac MRI. In addition to left ventricular function and mass, MRI will also provide data on right ventricular function.

Elevated blood pressure: Hypertension is a major risk factor for stroke and MI and contributes considerably to overall mortality. Sympathetic and parasympathetic reflexes are crucial for both short- and long-term control of blood pressure and heart rate. Sympathetic nervous system activation with concomitant parasympathetic dysfunction predisposes to arterial hypertension, insulin resistance, and cardiovascular organ damage¹⁵⁷⁻¹⁵⁹. Excessive sympathetic tone is present in other conditions known to be cardiovascular risk factors, too, such as obesity, metabolic syndrome, and diabetes^{160, 161}, and there is evidence that sympathetic overactivity precedes the development of both hypertension and hyperinsulinemia¹⁶². In the National Cohort, blood pressure will be measured in a standardized fashion. In addition, parameters of autonomic cardiovascular control (e.g., heart rate variability) can be assessed from 10-s ECGs^{163, 164}.

Assessment of cardiovascular functions and intermediate (preclinical) phenotypes:

Examinations (at baseline and during reassessment)

- Level 1: Combined measurement (e.g., by means of Vascular Explorer) of arterial stiffness, ankle–brachial index, pulse wave velocity, and other parameters
3D-Echocardiography
ECG
Blood pressure measurement
- Level 2: Intima–media thickness of the carotid artery (ultrasonography)
- MRI program: Cardiac, carotid, and brain MRI

A.2.3.2 Diabetes-related functional measurements and preclinical phenotypes

Impaired fasting glucose and impaired glucose tolerance are conditions that predispose to type 2 diabetes, which in about 7% of people with these problems progresses to overt diabetes every year¹⁶⁵. Furthermore, impaired glucose tolerance itself carries an increased risk of macrovascular disease¹¹⁸. The OGTT is the international gold standard for assessment of diabetes, and allows classification of study subjects into groups of normal glucose tolerance, impaired fasting glucose, impaired glucose tolerance and overt diabetes. In addition, indices derived from OGTT measurements are used in epidemiologic studies as proxy-measurements for beta-cell function and insulin sensitivity, which are key factors in the pathogenesis of type 2 diabetes^{166, 167} (see also **Sect. A.3.3.3**).

Increased formation and accumulation of **advanced glycation end-products** is one of the pathogenetic mechanisms underlying accelerated atherosclerosis in type 2 diabetes¹⁶⁸ and plays a role in inflammation and immune responses¹⁶⁹. Both crosslinking of proteins by advanced glycation end-products and receptor-mediated cellular activation contribute to loss of vascular elasticity and to propagation and maintenance of inflammation, contributing to the development of microvascular and macrovascular disease. Moreover, cross-linking by advanced glycation end-products results in vascular and myocardial stiffening, which are hallmarks in the pathogenesis of heart failure⁸²³. Several studies have shown an association of skin autofluorescence (AF) with levels of C-reactive protein, suggesting that skin AF may represent both inflammatory and hyperglycemic episodes¹⁷⁰. Levels of advanced glycation end-products in skin tissue (measured in skin biopsies) statistically significantly predict long-term diabetic microvascular complications, even after adjustment for hemoglobin A1c (HbA1c)¹⁷¹⁻¹⁷⁴. Skin AF was better able to predict fatal and nonfatal CVD events and all-cause mortality than was HbA1c¹⁷⁵. The superiority of risk prediction by advanced glycation end-products in skin has been explained by the concept of “metabolic memory”, i.e., the idea that stable advanced glycation end-products bound to long-lived protein such as skin collagen provide long-term memory of episodes of hyperglycemia and oxidative stress, whereas HbA1c only “remembers” this for several weeks. Further diseases for which an association with advanced glycation end-products or the soluble advanced glycation end-products’ receptor has been reported are mild cognitive impairment and Alzheimer’s disease^{176, 177}.

Retinopathy (e.g., microaneurysms, retinal hemorrhages, and exudates) is the most specific microvascular complication of diabetes. It eventually affects up to 40% of patients with type 2 diabetes¹⁷⁸, but is also seen in 2–15% of the nondiabetic general population¹⁷⁹. So far, no population-based data are available for Germany. The threshold for defining diabetes by fasting plasma glucose levels of ≥ 7.0 mmol/l is based on the assumption that it separates persons at high and low risk of developing diabetic microvascular complications. However, there seems to be a linear relationship, rather than a threshold level, between fasting glucose levels and the occurrence of retinopathy¹⁸⁰. The association of hyperglycemia and elevated blood pressure with the occurrence of retinopathy indicates that different etiological pathways contribute to the development of microvascular changes in the retinal vessels^{178, 181} and most likely also in the vascular beds of other organs. Consequently, retinal vascular characteristics have been described as risk indicators for a number of chronic diseases (including diabetes, hypertension, stroke, and ischemic heart disease) and all-cause mortality¹⁸²⁻¹⁸⁴. In addition, retinopathy is also associated with lower cognitive test scores, thus also providing a link to neurocognitive functions^{182, 184}. Computer-assisted grading systems can be used to rapidly assess qualitative and quantitative signs and extent of retinopathy based on retinal photographs¹⁷⁸. The combination of diverse functional measurements of the cardiovascular system, measures of metabolic impairments (e.g., impaired glucose tolerance), and neurologic and psychiatric functional assessment offers the unique opportunity to better understand the mechanisms underlying early pathological changes that ultimately progress to clinical stages of the major diseases in the focus of the National Cohort, and to identify their determinants.

Assessment of diabetes-related functions and intermediate (preclinical) phenotypes:

Examinations (at baseline and during reassessment)

Level 2:	OGTT
	Advanced glycation end products (AGE) reader / skin AF
	Retinal photography

A.2.3.3 Cancer-related precursor stages

Precursor stages of hematologic malignancies: For most tumor types (organ sites) no noninvasive methods are available for diagnosing premalignant phenotypes that can be applied easily within very large cohort studies based on representative population samples. One exception to this, however, is hematologic malignancies, for which premalignant states can be diagnosed in blood samples using cellular and genetic and epigenetic markers. One example is chronic lymphocytic leukemia, one of the most common types of adult leukemia worldwide, which is usually diagnosed on the basis of the malignant B-cell phenotype, similar to normal B1 lymphocytes that coexpress CD19/CD5 with dim surface immunoglobulin. Recent studies have identified small B-cell clones using six-color flow cytometry with antibodies against CD45, CD19, CD5, CD10, and kappa and lambda light chains in addition to immunoglobulin heavy chain gene rearrangement in healthy individuals with no signs of a lymphoproliferative disorder. This condition was termed *monoclonal B-cell lymphocytosis* and has a population prevalence ranging from 3 to 5% in the general population¹⁸⁵. A recent study within the prospective PLCO cohort in the USA showed that malignant B-cell lymphoma is a systematic precursor for chronic lymphocytic leukemia¹⁸⁶ and that malignant B-cell lymphoma existed up to 77 months before chronic lymphocytic leukemia was diagnosed. Another example is monoclonal gammopathy of undetermined significance, (MGUS), a premalignant plasma cell proliferative disorder which is commonly found in older people (>50years), with an average population prevalence of 3%. Again, a recent study within the PLCO cohort showed that *monoclonal gammopathy of undetermined significance* is a precursor stage for multiple myeloma¹⁸⁶. The definition for MGUS requires the presence of serum M protein at a level of less than 3 g/dl with fewer than 10% monoclonal plasma cells in the bone marrow. Affected people have a lifelong elevated risk for the progression to multiple myeloma; however, in only 1% of individuals with monoclonal gammopathy of undetermined significance does this progress to multiple myeloma. There is now evidence in the literature that genetic factors or possibly environmental factors may alter the risk of developing MGUS. Thus, understanding these interactions of genetic and environmental factors will improve early detection so that therapeutic regimens or even prevention strategies that will block the progression of MGUS to the malignant stage can be developed and initiated early on. A third example of a blood-based marker related to development of hematologic malignancies involves specific *chromosomal translocations*, such as the t(14;18)(q32;q21) translocation¹⁸⁷. It is likely that further precursor stages for hematologic malignancies can be identified and confirmed through prospective cohort studies.

A prerequisite for studies on precursor stages of hematologic malignancies is that intact peripheral blood lymphocytes are collected, which is indeed one component of the blood collection protocol for the National Cohort. It thus will also be possible to explore the determinants of progression of precursor stages to malignancy. Key hypotheses for such determinants include respiratory infections and other environmental factors, which may drive chronic antigenic stimulation¹⁸⁸⁻¹⁹⁰, nutrition-related lifestyle factors, and epigenetic alterations in different blood cell populations.

Assessment of cancer-related precursor stages:

Level 3: Collection of intact peripheral (blood) lymphocytes, for future ascertainment of hematologic precursor stages (e.g., malignant B-cell lymphoma, MGUS, chromosomal translocations)

A.2.3.4 Intermediate stages of neurologic and psychiatric diseases

Mild cognitive impairment represents an intermediate state between normal cognitive function and dementia, with the conversion rate to dementia being about three times higher than in subjects without cognitive impairment. Older age, subjective memory impairment, impairment in instrumental activities of daily living, and antecedent lower cognitive performance have been found to be significantly associated with the development of future mild cognitive impairment¹⁹¹. For the National Cohort, neuropsychological tests for cognitive impairment will be used that include tests of episodic memory and executive function, as both have been shown to predict time to progression from normal cognition to mild cognitive impairment as well as from mild impairment to dementia¹⁹². Five-year follow-up testing, at the participants' second visit, will then be used to assess intraindividual changes in cognitive function as a major prospective study outcome. In addition to the cognitive tests, MRI-based examination of the brain will be performed. It will thus be possible to couple specific morphologic changes of brain structures associated with neurodegenerative processes to the assessments of prospective change in cognitive function. Specific details about the methods and logistic organization for MRI imaging are given in **Sect. A.3.4**.

The prevalence of **olfactory dysfunction** in the general population is a matter of debate. Many authors reported frequencies of 1-5% of anosmia within the groups studied¹⁹³, but much higher prevalences can be found among older subjects^{194, 195}. Causes of smell dysfunction are manifold, the most important ones being aging, sinonasal disease, head trauma, or infections of the upper respiratory tract. Loss of olfactory function is often accompanied by depression¹⁹⁶⁻¹⁹⁸ and is also related to neurodegenerative disease, as it has been shown to be an early sign of Parkinson's disease^{197, 199} and Alzheimer's disease^{196, 198}, respectively. In the National Cohort, a brief, standardized smell test will be applied.

Assessment of neurologic and psychiatric intermediate phenotypes:

Examinations and questionnaires (at baseline and during reassessment)

Level 1:	Neuropsychological tests for cognitive impairment
	Smell test
MRI program:	Brain MRI

A.2.3.5 Respiratory function and preclinical phenotype

Lung function embodies significant information regarding the state of the respiratory system and can be used for diagnosis and staging of all major ventilatory disorders. Respiratory symptoms and shortness of breath occur only when up to 50% of the lung capacity may already be lost. According to data obtained in a large cohort study²⁰⁰ it may be possible to detect subclinical, mild to moderate airflow obstruction in up to 30,000 subjects (15%) in the National Cohort. Apart from pulmonary conditions in COPD, asthma, and fibrotic lung disease, reduced lung function has been found to be associated with many chronic medical conditions such as ischemic and congestive heart failure, hypertension, obesity, and diabetes^{201, 202}. In addition, lung function is also a predictor of overall morbidity and mortality²⁰³ and considered to be one of the best noninvasive functional predictors of biological aging and longevity^{204, 205}. Even in the absence of lung diseases, reduced forced expiratory volume in one second (FEV1) is a marker of cardiovascular mortality, independent of smoking history²⁰⁶. Spirometry, as the best established pulmonary function test, will be carried out in the National Cohort to identify subclinical stages. It provides an integrative signal covering mechanistically relevant aspects of lung volume loss and airway obstruction and can detect lung function decline or deterioration over time in the longitudinal setting.

Airway inflammation and deteriorated repair and remodeling processes are major pathophysiological mechanisms underlying chronic lung diseases. In contrast to respiratory function, only few noninvasive techniques can provide information about inflammatory processes in the lung. Among them, only exhaled nitric oxide (NO), FeNO, has gained clinical significance because upregulation of NO by inflammatory cytokines and mediators in central and peripheral airways can be monitored in exhaled air. Elevated levels of exhaled NO in asthmatic individuals are indicative of an ongoing allergic, i.e., mostly eosinophilic inflammation within central and/or peripheral airways. Therefore, exhaled NO is currently considered as a supplement to standard clinical asthma care guidelines. When quantified in terms of sensitivity/specificity and ROC curves, the amount of diagnostic information regarding allergic airway disorders is comparable to that of nonspecific bronchial challenges or sputum analyses. Acute respiratory tract infections are also associated with elevated NO concentrations. Within epidemiologic settings, it is also of interest to identify an allergic component of airway inflammation in subjects with COPD²⁰⁷⁻²⁰⁹.

Assessment of lung function and preclinical phenotype:

Examinations (at baseline and during reassessment)

- Level 1: Spirometry
- Level 2: Exhaled NO
- MRI program: Lung volume

A.2.3.6 Musculoskeletal functions and phenotypes

Osteoarthritis affects about 15-25% of the population over the age of 60, with knees, hands and hips being the most frequently involved sites²¹⁰⁻²¹². The disorder is often highly disabling and progressive and, indeed, osteoarthritis accounts for most of the total hip and knee replacements in Germany. Little is known about factors that provoke the manifestation of clinical symptoms and trigger the patient's decision to seek medical care. Recent studies indicated a substantial genetic influence which is site-specific for hip, knee, and hand osteoarthritis and affects not only the development but also the progression of the disease^{212, 213}. The National Cohort will offer the unique opportunity to investigate the complex interplay among genetic, environmental, and lifestyle factors in the pathogenesis of osteoarthritis and associated comorbidity in a prospective manner. Using blood and urine samples from presymptomatic individuals who develop osteoarthritis later on, particular attention will be paid to identifying biomarkers for early risk prediction. By combining functional examinations and MRI of hip and knee joints we can follow changes in function and structure prospectively to search for subtle MRI markers to predict hip osteoarthritis in the presymptomatic phase and to identify incident cases of hip and knee osteoarthritis by imaging.

Rheumatoid arthritis has a prevalence of about 0.8% of the adult population²¹⁴⁻²¹⁷, and biological agents to treat rheumatoid arthritis and related inflammatory diseases are on positions 1 and 2 of the 30 drugs with the highest sales volume in Germany²¹⁸. Complex interactions between environmental (e.g., viral infections) and behavioral factors (e.g., nutrition, smoking) and the contribution of comorbidity (e.g., metabolic syndrome, CVD) to disease onset and outcome have been recognized. Chronic inflammation itself, in part accompanied by accelerated aging of the immune system, is associated with an increased risk for endothelial damage, CVD, lymphomas, diabetes, and an unfavorable body composition²¹⁹⁻²²¹. Thus, inflammatory arthropathy constitutes a confounding factor in studies of chronic disease. Utilizing the longitudinal design, the National Cohort will evaluate the potential impact of known and novel biomarkers to predict the development of rheumatoid arthritis during the presymptomatic stage. Incident cases will be identified mainly through active follow-up. Importantly, a musculoskeletal examination, including the examination of joints, is planned to be part of the program (see **Sect. A.3.3**).

Musculoskeletal pain disorders belong to the most frequent and costly health complaints in Western societies^{222, 223} and are estimated to affect almost 20% of the adult population in Germany²²⁴. About 10% of the adult population suffers from disabling back pain²²⁵ or chronic widespread pain according to the American College of Rheumatology criteria²²⁶. Musculoskeletal pain is a key symptom of many degenerative and inflammatory musculoskeletal disorders²¹⁶ and a strong predictor for individual functional impairments and for health care utilization²²⁷. There is limited longitudinal evidence regarding risk factors for development and course of musculoskeletal pain problems. We will assess pain locations across the body, as pain in any one region is much more likely to occur in the presence of pain in another region, and systemic patterns prevail²²⁸. A detailed assessment of pain is important in order to investigate cross-sectional and longitudinal associations with the results obtained by MRI. Musculoskeletal pain will be assessed as part of the clinical examination, emphasizing locations (electronic pain mannequin, based on the pain mannequin in the German Pain Questionnaire), intensity, duration, and pain-related disability and functional impairment.

Osteoporosis is characterized by low bone mass and microarchitectural deterioration of bone tissue, resulting in an increased susceptibility to fracture. In Western societies, approximately 1 in 2 women and 1 in 4 men aged >50 years will have an osteoporosis-related fracture in their lifetime. Common sites for osteoporotic fractures are the spine, hip, distal forearm, and proximal humerus. The likelihood of a fracture at any of these sites in developed countries is close to the the risk of incident CVD ²²⁹. Apart from the socioeconomic burden, fragility fractures are responsible for significant morbidity and decreased quality of life because of pain, and loss of mobility and independence. In addition, hip and vertebral fractures are associated with an increased CVD and mortality risk^{230, 231}; a relationship between diabetes and osteoporosis has also been described²³². Although several risk factors for osteoporosis and osteoporosis-related fractures have been identified, including peak bone mass, physical activity, smoking, alcohol, diet, or use of glucocorticoids^{229, 233}, several research questions have not yet been answered, e.g., interactive effects of different exposure factors (e.g., vitamin D supply, physical activity) and the genetic background²³⁴.

Dual energy X-ray absorptiometry (DXA) is used to determine bone mineral density and represents a standard procedure in the diagnosis and management of osteoporosis²³³. The results of bone mineral density measurements are also considered in risk scores developed to predict fracture risk²³⁵.

Assessment of musculo-skeletal functions and phenotypes:

Examinations and questionnaires (at baseline and during reassessment)

- Level 1: Questionnaire,
DXA (lumbar spine, proximal femur)
- Level 2: Medical examination,
Electronic pain mannequin
- Level 3: Biomarkers in blood samples
- MRI programme: Whole-body MRI

A.2.4 Major areas of exposure and risk factors

A.2.4.1 Physical activity, body composition, and physical fitness

Physical activity: In industrially developed countries, increasing epidemiologic evidence indicates that lack of regular physical activity is a major risk factor for common chronic diseases, including heart disease, type 2 diabetes, and several common forms of cancer. This has prompted insistent recommendations by national and international expert panels

that regular physical activity should be promoted as a major approach to reduce the chronic disease burden. The American College of Sports Medicine, the American Heart Association, and the CDC, for example, formulated recommendations for a combined regimen of moderate- and vigorous-intensity exercise for the prevention of heart disease and the promotion of overall health. With regard to cancer, international expert panels judged that increasing physical activity levels, together with tobacco control and body weight control, could be the most promising approach to alleviate the burden of cancer in industrialized countries^{236, 237} although precise estimates of the magnitude of this preventive potential are still lacking.

Expert panels, however, also recognized several important limitations of the epidemiologic evidence so far. Most studies to date have included measures of the total amount of physical activity (used to characterize participants as “active”, “moderately active”, or “inactive”, for example), and only few studies addressed the effects of different physical activity profiles in terms of intensity, duration, or frequency, independently of the overall energy expenditure due to physical activity. Thus, up to now, it is still relatively unclear which type, intensity, and temporal pattern of activity is most beneficial for health.

A second limitation of most studies on physical activity and health has been that activity measurements were based exclusively on questionnaire assessments, which can contain substantial measurement error^{238, 239}. Finally, in most of the large-scale prospective cohort studies so far, physical activity assessments were made only at a single point in time, and intraindividual variations over time in activity levels were not sufficiently accounted for in RR estimates. Taken together, we anticipate that substantially stronger and more valid estimates of both relative and attributable risks for chronic disease will be obtained if more accurate methods are used to assess activity and assessments are repeated over time²⁴⁰.

Assessment of physical activity:

Examinations and questionnaires (at baseline and during reassessment)

Level 1:	Questionnaire
	7-day triaxial accelerometry
Level 2:	24-h physical activity recall

Excess body weight and specific measures of body composition: Excess body weight is one of the best established risk factors for type 2 diabetes and heart disease and is increasingly being recognized as a major risk factor for many forms of cancer⁸⁵. However, a major limitation of most prospective cohort studies is that these have relied primarily on the measurement of BMI to assess the impact of excess weight on risk of chronic disease. BMI does not take body fat distribution into account, nor does it distinguish between fat mass and lean body mass (e.g., skeletal muscle). BMI provides only a very approximate measure of visceral fat mass – the metabolically more active body fat compartment that secretes, for example, a large part of cytokines and hormones implicated in the development of type 2 diabetes²⁴¹. Waist circumference measurements, although correlating more strongly with the amount of visceral fat, still also provide only an approximate measure of intra-abdominal body fat, in comparison to more accurate imaging measurements^{242, 243}. A number of observations indicate that a substantial part of “hidden” obesity is not properly accounted for by these indices and that even among subjects with comparatively low BMI intra-abdominal fat may considerably increase risks of chronic diseases and overall mortality. In the USA and Europe the “metabolic” syndrome has a non-negligible prevalence among subjects classified as normal weight according to their BMI. Furthermore, waist circumference has been shown to be more strongly associated with mortality risk among subjects with a normal BMI (<25 kg/m²) than among those whose BMI is higher. The latter observations suggest that, at least for a certain portion of chronic disease and mortality, low skeletal muscle mass in combination with a relative excess of abdominal fat (“sarcopenic” obesity) could be a major

risk factor for chronic disease^{244, 245}. Thus, most precise measurement of body composition – and change in body composition during follow-up – is mandatory.

Assessment of body composition:

Examinations (at baseline and during reassessment)

- Level 1: Body weight and height
 Waist and hip circumference
 DXA
- MRI program: Whole-body MRI

A.2

Physical fitness: Physical fitness refers to a full set of different attributes that people have or achieve and that relates to the ability to perform physical activity. Among these attributes, cardiorespiratory fitness, which is the ability of the body’s circulatory and respiratory systems to supply oxygen during sustained physical activity, has repeatedly proven to have an important impact on CHD and other CVD²⁴⁶. A recent meta-analysis suggests that cardiorespiratory fitness and physical activity have significantly different relationships to CVD, showing reductions in RR nearly twice as great for cardiorespiratory fitness as for physical activity levels²⁴⁷, although physical activity is the most important determinant of cardiorespiratory fitness. A few prospective cohort studies, so far, have also reported associations between cardiorespiratory fitness and reduced risk of diabetes. Relationships with risks of cancers or dementia, however, have not been addressed by well-designed cohort studies.

In addition to cardiorespiratory fitness, measures of reduced muscle strength have been associated both with increased risks of death from CVD in men and with cancer^{248, 249} independently of other, established risk factors.

Assessment of physical fitness:

Examinations (at baseline and during reassessment)

- Level 1: Hand grip strength
- Level 2: Step test
 One-leg stand / posturometry

Metabolic parameters related to body composition, physical activity, and fitness: The availability of prediagnostic blood samples in the National Cohort will provide the unique opportunity to shed light on physiologic mechanisms that may relate excess body weight, physical inactivity, and reduced physical fitness to risks of CVD, cancer, and mortality by measuring biomarkers from pertinent pathways²⁵⁰⁻²⁵².

Prior studies have shown, for example, that obesity-related markers such as adiponectin, C-reactive protein, or resistin predict risk of CVD beyond established risk factors (including anthropometric measures of adiposity and fat distribution) and that a substantial proportion of CVD incidence may be accounted for by such factors²⁵³. Other adiposity-related metabolic alterations, such as hyperinsulinemia, increases in bioavailable sex steroids (androgens and estrogens), and chronic low-grade inflammation, have been related to risks of developing tumors of the colorectum and breast, for example. Further metabolic factors that have been studied in this context include indicators of glucose tolerance, blood lipid profile, and serum levels of specific growth factors, coagulation factors, inflammation factors, or changes in immune function. The advances in science and technology over the past few years, especially the development of high-throughput laboratory methods, offer the opportunity to identify novel biomarkers for cardiovascular risk factors using discovery-oriented approaches (e.g., metabolomics, proteomics).

The identification of biomarkers that quantify metabolically active adipose tissue beyond anthropometric parameters represents a complementary approach to define an "obesity phenotype" that is relevant for chronic disease risk²⁵¹, and this approach can also provide essential insight into physiologic mechanisms of disease development. In addition, metabolically determined markers of chronic disease risks can constitute an important component in prediction models for chronic disease risks and may also represent direct targets for intervention through means of diet, lifestyle, or drug treatment. Results from prior studies suggest that there may be a substantial overlap in metabolic pathways leading to diabetes, CVD, various cancers, neurodegenerative diseases, and early death.

Assessment of metabolic parameters:

Level 3: Analyses of biomaterials

A.2.4.2 Diet

Since prospective cohort investigations into the causes on chronic disease were started up, as early as in the 1970s, diet has been a central focus of interest, and from the 1980s, many large-scale prospective studies were initiated worldwide to address relationships of diet with cancer, CVD, diabetes, and overall mortality^{254, 255}. In these larger studies, habitual dietary intake was generally assessed using food frequency-type questionnaires, in some cohorts (e.g., the EPIC cohort) combined with 24-h diet recalls in calibration substudies.

Comprehensive summaries of this research by international expert panels, such as that of the World Cancer Research Fund, showed conclusive evidence for increased risks of certain cancer types in association with higher reported consumption levels of alcohol, red meat, and processed meats, and a possible reduction in certain cancer risks at high intake levels of dietary fiber and whole-grain foods, although generally the associations with risk were of a relatively modest magnitude. For CVD, prospective studies and intervention trials have shown reduced risks in relation to higher consumption levels of fruits, vegetables, and n-3 polyunsaturated fatty acids²⁵⁶⁻²⁵⁸, and increased risks in relation to consumption of saturated fatty acids and trans-fatty acids relative to monounsaturated and polyunsaturated fatty acids and red meat. Surprisingly, only a few studies reported specifically about the role of diet for heart failure, atrial fibrillation, and subclinical arteriosclerosis.

In spite of several consistent findings relating diet to risks of cancer and CVD, evidence has also been increasing that there are limits to the validity of dietary intake assessments based on food frequency questionnaires. For example, case-control studies nested within the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort have shown clear inverse relationships of serum vitamin C with gastric cancer risk²⁵⁹ or several serum B vitamins with cancers of the lung and colorectum^{260, 261}, and serum lycopene with risk of prostate cancer²⁶², where parallel evaluations on nutrient intake estimates calculated from questionnaires did not. Furthermore, in one of the regional subcohorts of EPIC, in Norfolk (UK), increased breast cancer risk was observed in association with higher intakes of dietary saturated fats as assessed by replicate, weighed food records, but not with dietary fat assessments by food frequency questionnaires, and a similar contrast was observed more recently in the US Women's Health Initiative cohort²⁶³⁻²⁶⁵. This indicates that measurement error included in data derived from food frequency questionnaires may obscure existing diet-disease association, even associations of substantial magnitude. In-depth dietary validation studies that used recovery-based biomarkers for total dietary energy and protein as reference measurements were more valid for 24-h diet recalls than for food frequency questionnaires^{239, 266}. Thus, increasing precision of dietary assessment in the National Cohort is mandatory; this aim will be achieved through a combination of repeated 24-h recalls, a short

food list (to assess habitual nonconsumers), and the use of nutritional biomarkers, also including new markers that will emerge, e.g., from studies using metabolomics techniques.

Assessment of diet:

Examinations and questionnaires (at baseline and during reassessment)

- Level 1: Repeated simple 24-h diet recalls (web-based application)
Food list
Collection of biomaterials for nutritional biomarker analysis
- Level 2: Exact 24-h diet recall (by telephone)

A.2

A.2.4.3 Smoking and alcohol consumption

Alcohol is an important risk determinant for cancer and other chronic diseases which has previously been studied in great depth and with a relatively high accuracy in many studies²⁶⁷. An assessment of alcohol intake will be undertaken to account for these known effects, in analyses addressing the relationships of disease risks with other risk factors.

Smoking is a well-established risk factor²⁶⁸, for which no major new findings (relationships with chronic disease risk) are expected; lifetime smoking history and current smoking habits, however, must be accurately assessed as major possible confounders in analyses of relationships with other risk factors or as a possible variable to explain associations, e.g., with socioeconomic health disparities.

Assessment of alcohol consumption and smoking behavior:

Examinations and questionnaires (at baseline and during reassessment)

- Level 1: Questionnaire

A.2.4.4 Psychosocial factors

Adverse psychosocial factors, such as work-related stress, social isolation, lack of social support, and negative life events, have been shown to be associated with an increased risk for CVD²⁶⁹⁻²⁷¹, diabetes²⁷²⁻²⁷⁴, obesity²⁷⁵, and depression^{276, 277}. Personality traits, on the other hand, have been shown to be associated with adverse health behavior²⁷⁸⁻²⁸². Several potential causal mechanisms underlying the adverse health effect of psychosocial factors have been suggested, involving the hypothalamic-pituitary-adrenal axis, hypertension and blood pressure reactivity to stress, heart rate variability, endothelial and vascular function, inflammation and immunity, platelets and coagulation factors, fibrinogen, lipids, and the metabolic syndrome^{270, 283-286}. With regard to cancer, it has been hypothesized that psychosocial factors may influence incidence directly through altering immune and endocrine functions or indirectly through health related behaviors, although evidence for a direct causal association of psychosocial factors with cancer is still scarce²⁸⁷⁻²⁸⁹.

Personality: Five factors of human personality, called the “Big Five”, constitute the most widely fundamental human personality factors studied, and include “extraversion”, “agreeableness”, “conscientiousness”, “neuroticism”, and “openness”, although there is considerable variation to both the exact labeling and meaning given to each of these five dimensions²⁹⁰. It has been shown that the “Big Five” factors are relatively stable across human life cycles, starting in young adulthood²⁹¹. They are at least in part heritable^{292, 293}, have been identified in very different cultures²⁹⁴, and probably implied an adaptive value over the course of human phylogenesis²⁹⁵. Research on these five factors of human personality has highlighted the importance of personality to predict health-relevant behaviors. For example, extraversion has been linked to smoking, alcohol consumption, and risky sexual

behavior²⁷⁸⁻²⁸², whereas conscientiousness seems to be inversely associated with smoking, alcohol consumption, risky sexual behavior, and overweight^{278-282, 296, 297} and has been found to be accompanied by healthy habits such as exercising, healthy nutrition, or enough sleep.

Work-related stress is an important psychosocial factor that can be captured by different models, such as the effort–reward imbalance (ERI) model²⁹⁸, or the demand–control model²⁹⁹. The ERI model, an internationally applied model to analyze psychosocial work stress that has been developed in Germany, highlights distress arising from inadequately low rewards in spite of high external efforts. ERI has been found to be associated with poor self-rated health across Western and Eastern European countries³⁰⁰ and to be predictive of chronic disease such as CVD, depression, and other adverse health outcomes^{301, 302}. Differences in the association of work-related psychosocial factors with CVD³⁰³⁻³⁰⁷ and diabetes^{272, 273} have been observed by gender, mostly with stronger effects in men than women. Within the National Cohort, the ERI model will be used to analyze psychosocial work stress (see **Sect. A.3.2.2**).

Social networks and social support: The social network of a person refers mainly to the quantitative aspects of social contacts. It was shown that persons who are well connected with others and were engaged, for example, in political parties, sport clubs, or unions are healthier than people with smaller networks³⁰⁸ and have a lower mortality^{309, 310}. Social support refers to a social network’s provision of psychological and material resources in instrumental, informational, and emotional domains. It can help individuals to cope with stress³⁰⁸, which may reduce disease risk. Social support has been found to be associated with cardiovascular outcomes^{269, 270}, but results differ among gender or CVD outcome^{304, 306, 307}. Information about the social network and social support will be obtained in the National Cohort using standardized instruments developed by Berkman and Syme³⁰⁹ that have been modified and validated for use in German studies³¹¹.

Adverse life events, especially during early life, have an important impact on adult biological stress reactivity³¹² and have been found to be associated with risk of type 2 diabetes³¹³ and the metabolic syndrome³¹⁴⁻³¹⁶ in most, but not in all studies³¹⁷. In the National Cohort, a history of adverse life events will be obtained.

By prospectively assessing different domains of psychosocial factors in conjunction with a wide range of behavioral and biological measurements and disease outcomes we can elucidate potential causal pathways to adverse health effects.

Assessment of psychosocial factors:

Examinations & questionnaires (at baseline and during reassessment)

Level 1: Questionnaires

A.2.4.5 Socioeconomic status

“Socio-economic status” (SES) expresses the relative social and economic position of a person or of an entire household as compared to the other members of society. Measures of SES are income, education, and occupation, or combined indices, but also material possessions (such as house and car ownership) or self-rated subjective status that can be used to reflect the social status of a person³¹⁸. Social epidemiologic research has shown consistently that lower SES is associated with higher mortality and morbidity for the majority of chronic diseases³¹⁹⁻³²⁴. This inverse social gradient can be found to a different extent in most industrialized countries and is well documented for the US and European countries^{307, 325-330}. However, results concerning the inverse social gradient differ considerably between the type of disease or health outcome analyzed and between the measurements of SES that

were used^{322, 331, 332}. While findings are more consistent for CVD and diabetes, results are divergent for different types of cancer^{329, 333}. Furthermore, only little information is available on the association of SES with neurologic and psychiatric diseases, although some results point to an adverse effect on major depression³³⁴, Parkinson's disease³³⁵, and cognitive impairment²⁷⁷. Notably, considerable differences in health inequality have been observed by gender^{324, 331}.

Explanations for inequalities in health according to SES include material, cultural-behavioral, psychosocial, life course, and biomedical factors and mechanisms³³⁶⁻³³⁸. However, the results of studies attempting to estimate the relative contribution of these groups of risk factors to explain the social gradient of disease are conflicting³³⁸⁻³⁴². Various attempts have been made to explain how SES might affect health via individual health behavior³²⁷ and biomedical factors^{343, 344}. Mechanisms that have been suggested involve inflammatory factors as well as metabolic factors and obesity^{343, 344}. However, further research is needed to better understand **causal pathways** from SES to different adverse health outcomes, including psychosocial, behavioral, and biomedical risk factors as potential intermediate factors. Most importantly, longitudinal studies are lacking that address prospective associations of SES and psychosocial factors with subclinical disease, preclinical biomarkers, and functional parameters of health that would be suitable to disentangle the interplay of factors on the pathway from SES to disease. The fact that we will be able to assess prospectively the association of SES with subclinical disease stages avoids the social downward drift bias (as a consequence of disease).

SES across the life course: There is increasing evidence that there is a cumulative effect of SES across the life course on subclinical and clinical cardiovascular and metabolic diseases³⁴⁵⁻³⁴⁷ and mortality³⁴⁸, and on inflammatory factors which may play an important role in the causal chain^{344, 349} from lower SES to an increased risk of chronic diseases. Furthermore, the effect of SES on disease risk may change over the course of life. Studying variation in the association of different SES indicators with specific diseases will give us a better understanding of the etiological mechanisms relating specific diseases with specific exposures³⁵⁰. Moreover, it is important to evaluate any confounding by SES by including different indicators of SES across the life course³⁵⁰.

Regional variation of disease across Germany and role of neighborhood SES: There is increasing evidence that neighborhood SES (i. e., the socioeconomic environment of the residential area as measured by aggregate area-level measures such as unemployment rates, housing value, population density, and car ownership in the residential area) can affect health and contribute to social inequalities in health even after accounting for individual SES³⁵¹. A lower level of neighborhood SES has been found to be associated with an increased risk of mortality^{352, 353}, stroke³⁵⁴, hypertension³⁵⁵, CHD³²⁶, subclinical CVD^{347, 356, 357}, the metabolic syndrome³⁴⁵, diabetes³⁵⁸, propensity for depression^{359, 360}, and other adverse health outcomes.

While several potential mechanisms are being discussed with respect to contextual factors that shape an individual's health, such as neighborhood resources for physical activity and healthy foods^{355, 358, 361-363}, further research is needed to elucidate the pathways from area-based SES to disease^{351, 364}.

There is a remarkable **geographical gradient** across Germany for all-cause and cardiovascular mortality⁸, diabetes³²⁴, hypertension³⁶⁵, and obesity^{366, 367}. At the same time, the geographical gradient of area-based SES such as unemployment rates has been well documented (www.destatis.de). Initial analyses point to a relevant effect of neighborhood SES on cardiovascular risk factors in Germany and in Eastern Europe¹¹. However, data are lack-

ing to further analyze how neighborhood SES and regional differences in socioeconomic deprivation might help to explain the geographical gradient in disease and mortality beyond individual SES.

Assessment of SES:

Examinations & questionnaires (at baseline and during reassessment)

Level 1: Questionnaire

A.2

A.2.4.6 Sleep-related characteristics

There is evidence that sleep characteristics including sleep duration at night, sleep environment, sleep disturbances at night, daytime sleepiness, afternoon naps (siesta), perceived sleep satisfaction, sleep apnea syndrome, and sleep quality influence the risk of atherosclerotic diseases, especially coronary artery disease^{368, 369}, stroke³⁷⁰, and other diseases. Furthermore, chronic sleep complaints in epidemiologic studies have been associated with an increase in overall mortality and morbidity³⁷¹. In addition, several diseases and exposures influence sleep characteristics (for example, depression and medical drugs). Sleep characteristics are also considered as intermediates between certain exposures (e.g., shift work) and diseases (e.g., cancer). Chronodisruption is a relevant disturbance of the circadian organization of physiology and endocrine regulation and is hypothesized to be linked to the development of various diseases. It may be caused by shift work or other “zeitgebers” outside of the occupational setting such as physical activity at night, nightly exposure to light, sleep disturbance, or reduced sleep duration³⁷². There is increasing interest in determining whether chronodisruption represents a risk factor for hormone-dependent cancers, e.g., breast or prostate cancer³⁷³. Melatonin is a robust marker for the circadian phase. Physiologically, melatonin levels are low during the day and peak at night. Biologically relevant chronodisruption, which is associated with a circadian phase shift, can be established by measuring the melatonin metabolite 6-sulfatoxymelatonin in morning urine³⁷⁴.

Assessment of sleep-related characteristics:

Examinations and questionnaires (at baseline and during reassessment)

Level 1 Questionnaire

Level 2 Portable device to assess sleep apnea and other sleep characteristics

A.2.4.7 Chronic infections, immune factors, and microflora

Immunization, infection status and risk of chronic diseases

Immunizations, vaccine-preventable diseases: In view of the increasing life expectancy and the fact that immunity to “childhood diseases” is increasingly being acquired through vaccines instead of natural infection, pathogen-specific immunity is changing, and the public health relevance of vaccine-preventable diseases in the adult population is bound to increase. Based on data from the statutory health insurances, the Robert Koch Institute estimates 240,000 annual cases of herpes zoster in Germany (Robert Koch Institute, unpublished). The prevalence of carcinogenic human papillomavirus infections among women older than 20 years in Germany is estimated to be 4.3%³⁷⁵. In Germany, 10,000 to 40,000 individuals die annually from influenza-associated complications in Germany³⁷⁶. The notification rates of invasive pneumococcal infections are the highest in individuals ≥60 years compared to other age groups³⁷⁶. Investigations of the Robert Koch Institute using blood donations have shown that the prevalence for hepatitis B virus was 133 per 100,000 blood donations³⁷⁷. At

the same time, substantial immunization gaps still exist among German adults^{378, 379}. The impact of both childhood and adult vaccination strategies on the epidemiology of vaccine-preventable diseases has become an emerging topic in mathematical vaccination models. For instance, attempts to describe latency and reactivation of shingles after varicella virus infection may have paramount implications for varicella vaccination strategies, but validation through prospective study designs is urgently needed³⁸⁰. Similar questions concerning replacement, reactivation, and recombination of other pathogens such as *Streptococcus pneumoniae*, *Neisseria meningitidis*, and human papillomavirus, and the respective impact of these effects, can only be addressed adequately in prospective settings. Seromarkers for immunity exist for most vaccine-preventable diseases and will be complemented by a review of immunization cards, self-report, and secondary data analysis such as data available through the Associations of Statutory Health Insurance Physicians³⁸¹.

Effects of infections and immunizations on cancer and other chronic diseases: A variety of pathogens have been associated with increased risk of malignancies, such as *Helicobacter pylori* and gastric cancer and mucosal-associated lymphoid tissue lymphoma, human papilloma virus and cervical cancer, and hepatitis viruses and liver cancer and, recently, lymphoma³⁸². The attributable risk of forms of cancer stemming from these treatable and partly vaccine-preventable infections has been estimated by the American Cancer Society to be about 18% worldwide. The 5-year prevalence of cervical cancer in Germany is 23,800, with 1,600 deaths annually according to the Germany cancer registry¹³⁷. The fraction of these cancers truly attributable to infections, however, can only be validly estimated by using a prospective study design. The National Cohort provides the suitable age distribution and study design to address these questions, as a large variety of infections can be assessed through seroconversion markers, while direct pathogen detection may also be possible in level 3 studies.

Persistent infections, particularly by members of the herpesvirus family, have been associated with increased overall mortality and with the pathogenesis of various chronic diseases that often feature some degree of chronic inflammation^{383, 384}. Human cytomegalovirus has been linked to heart disease and other cerebrovascular diseases, including stroke³⁸⁵, autoimmune diseases, certain cancers (e.g., glioblastoma multiforme), and to immune senescence. With a baseline seroprevalence of 30% and a yearly seroconversion rate of 0.55%³⁸⁶, this virus is well suited for prospective investigations into effects of a specific pathogen on the incidence of chronic diseases. There is epidemiologic evidence that infectious agents contribute to the pathogenesis of neurodegenerative and neuroinflammatory diseases, including Alzheimer's disease and multiple sclerosis³⁸⁷. However, a single pathogen has not been identified. By prospectively collecting information on chronic disease endpoints and a variety of biomaterials from each participant, the etiology of these and other chronic diseases can be investigated in the context of (1) complex infectious exposure both cross-sectionally and longitudinally (sequential infections) and (2) currently unknown or unsuspected infectious exposure.

Assessment of immunization and infection status:

Examinations and questionnaires (at baseline and during reassessment)

- Level 1: Questionnaire, immunization cards
- Level 3: Seromarkers

Immune senescence, immune dysfunction and chronic disease risk

Immune senescence refers to the increased susceptibility to infections, poor vaccine responses, and altered chronic inflammation that is observed in about 10–30% of individu-

als over age 65, depending on the definition^{388, 389}. The observations that 95% of influenza deaths occur in 45+ year olds, although they account for <12% of all influenza infections³⁹⁰, and that 30–57% of ≥60 year olds do not mount protective titers after influenza immunization^{391, 392} imply that immune senescence will contribute increasingly to the rise in health care expenses in the aging population in Germany. In terms of risk factors, chronic infections (particularly with cytomegalovirus and other herpes viruses^{393, 394}) are currently believed to make the highest contribution, although other factors such as nutrition³⁹⁵, comorbidity and chronic stress³⁹⁶ are also being considered. *Decline in immune function unrelated to immune senescence*, too, may result in increased susceptibility to infection and increased propensity for inflammation. This issue will also be addressed since it leads to higher health care utilization and loss of working capacity and therefore presents burdens on public health and the economy. *Autoimmunity associated with non-organ-specific autoantibodies* will also be studied. These antibodies can be readily detected and constitute an early step in humoral autoimmunity. Indeed, they are not only associated with specific autoimmune diseases but also with chronic diseases such as CVD and neoplastic diseases. Declining immune function (due to immune senescence or not) will be assessed by measuring frequency and severity (assessing physician attendance and hospitalization) of various infectious diseases, including invasive bacterial infections, and lower-than-expected immunization titers. Cases at risk of immune senescence will – in addition – be identified by characteristic abnormalities in leukocyte surface markers (e.g., inverted CD4/CD8 ratio) in cryopreserved peripheral blood mononuclear cells (MNC), and by using novel biomarkers as they become available.

Assessment of immune senescence and dysfunction:

Collection of relevant biomaterials (at baseline and during reassessment):

Level 3 Analyses of blood serum and (intact) peripheral blood lymphocytes

Microbial colonization (oral, nasal, and colonic microflora) and chronic disease risk:

It is becoming increasingly clear that microbial colonization affects the health of the human host (e.g., ref. ³⁹⁷). Firstly, the intranasal and gut microflora may represent small but important community reservoirs of bacteria resistant to broad-spectrum antibiotics. Methicillin-resistant *Staphylococcus aureus*, extended-spectrum beta-lactamase-producing bacteria or Vancomycin-resistant Enterococci are each detected in 1–2% of the general population³⁹⁸, with a rising trend. Risk factors for the acquisition and persistence are prior antibiotic use, hospitalizations, exposure to colonized animals, and genetic host factors. Colonized individuals of the National Cohort will be followed up in targeted studies. Approximately 20% of adults are intranasal carriers of methicillin-sensitive *Staphylococcus aureus*, an organism that is associated with expression of virulence factors and superantigens, and invasive infections such as abscess formation, endocarditis and osteomyelitis, and the development of atopic dermatitis. Another emerging issue is carriage of *Streptococcus* spp. which, in light of the respective vaccine, may experience carriage replacement and switching variants in favor of non-vaccine-preventable invasive pneumococcal disease³⁹⁹. The recent advances in genomics research have opened previously unexpected avenues for studying associations between entire colonizing microbial populations (“microbiome”) and chronic diseases. For instance, the composition of the gut microflora was recently linked to a higher risk of obesity in the general population⁴⁰⁰. We will therefore investigate the effects of the composition of microbiomes in nasal swabs, saliva, or stool on the occurrence of acute and chronic diseases. These biomaterials will be biobanked in frozen state so that culture-independent, nucleic acid-based analyses can be subsequently conducted.

Assessment of microbial colonization:**Collection of relevant biomaterials** (at baseline and during reassessment):

Level 3 Analyses of saliva, nasal swabs, and stool samples
 (on nested case-control basis)

Zoonoses and chronic disease risk

Approximately 20% of the National Cohort participants are expected to own one or more household pet⁴⁰¹. Pets may pose health risks, for instance, by (1) causing injuries⁴⁰², (2) inducing allergies⁴⁰³, and (3) being a source of infectious agents. In addition to parasitic, fungal, and viral infections, pets transmit a number of bacterial pathogens⁴⁰⁴. Furthermore, several studies have shown a national burden of zoonotic infections caused by nutritional exposure and by direct or indirect contact to farm animals even in the general population⁴⁰⁵. Our goal is to study the impact of pathogens that are shared by humans and animals on human health. Two pathogens could act as proxies of special interest. *Bartonella* spp. are highly prevalent in cats and dogs in Germany (seroprevalence ~15%)^{406, 407} and can cause severe infections in humans, especially in immunocompromised individuals^{408, 409}. Pets and farm animals may represent community reservoirs of antimicrobial-resistant bacteria^{410, 411}. The Level 1 questionnaire will contain screening questions regarding past and present exposure to household pets and farm animals. Individuals reporting a significant exposure in their history will then be selected for targeted substudies at Level 3. For instance, we will study exposure to *Bartonella* spp. and *Staphylococcus aureus* in pets and their owners in parallel. The ability of the owners to obtain nasal swabs and stool samples from their cats or dogs and their willingness to have a veterinarian draw blood from their pets will be tested in 2011 in a feasibility study headed by the University of Veterinary Medicine in Hannover.

Assessment of zoonotic infections:**Examinations and questionnaires** (at baseline and during reassessment)

Level 1: Questionnaire (screening questions)
 Level 3: Collection of biomaterials in participants and animals
 for microbial analysis

A.2.4.8 Occupational and environmental exposures**Occupation**

Most people spend a major part of their lifetime at work, often exposing them to hazardous chemical and physical risks and to negative psychosocial stress. Exposures at the workplace are generally higher than in the environmental setting; they have been measured, regulated, and monitored for approximately 40 years now and have been addressed in particular in specialized, occupational cohort studies. In the National Cohort, complete occupational histories will be obtained for study participants by data retrieval from the National Institute for Employment Research (Federal Employment Agency). These data will be linked to job exposure matrices (JEM) to estimate possible past exposures to occupational risk factors for cancer, cardiovascular, musculoskeletal, and neurodegenerative diseases. Several kinds of exposures can be investigated this way, including physical and biomechanical work factors, exposure to fine (PM₄ <4 μm) and nanosized particles (<100 nm), specific pesticide groups and solvents, electromagnetic fields, and metals.

Work organizational constraints also affect mental health (fatigue, anxiety, depression, burn-out-syndrome, etc.), and may increase the risks of CVD^{306, 412} (e.g., hypertension) and

other diseases. There is also robust evidence that psychosocial stress at work is a strong predictor of mental disorders, including depression^{276, 413} (see also **Sect. A.2.4.4**). Exposure to an adverse psychosocial work environment (e.g., high demand, low control, low support, low reward, injustice, and job insecurity) increases the risk of CVD, depression, cognitive decline, and dementia. Status passages during working life and changes in social prestige have been reported to exert an influence on all-cause mortality and CVD, independently of the family biography and other lifestyle factors⁴¹⁴.

Night work can desynchronize biological clocks⁴¹⁵ and disrupt circadian processes, such as tissue regeneration and DNA repair, and thereby may increase the risk of cancer³⁷³. Moreover, shift work increases the incidence of CVD and diabetes mellitus⁴¹⁶.

Assessment of occupational exposures:

Examinations and questionnaires (at baseline and during reassessment)

- Level 1: Questionnaire (filter questions for later administration of specific questionnaires, e.g., on shift work)
- Level 2: Specific questionnaires (psychosocial factors at work)
- Level 3: Specific questionnaires (following filter questions, e.g., on shift work, stress at work, hazardous exposures)

Secondary data: Occupational history

Environmental exposures

Air pollution and noise:

Air pollution has been consistently associated with all-cause mortality and with CVD mortality⁴¹⁷. Studies assessing traffic-related pollution have highlighted that other components of fine particulate matter – ultrafine particles, organic carbohydrates, or other oxidative stress-inducing components such as transition metals – are associated with severe health effects. In addition to traffic-related air pollutants, environmental noise is increasingly becoming a problem as well. A multicenter European study on aircraft noise showed an excess risk of hypertension related to long-term noise exposure⁴¹⁸. The overall evidence on the association between noise and ischemic heart disease is still mixed^{419, 420} and relatively scarce. However, there is still strong suspicion that traffic-related exposures to air pollutants, noise, and stress increase the risk of CVD and other diseases. Furthermore, comprehensive assessments of these kinds of exposures are sparse and other chronic diseases related to oxidative stress, such as diabetes⁴²¹ or neurodegenerative diseases⁴²², are just emerging as being potentially affected by traffic exposures.

In the National Cohort, geocoding data can be connected with study participants' addresses of residence (small-area neighborhood) and with habitual workplaces to determine average local levels of outdoor air pollution and noise. In addition, several questions relating to indoor exposure will be addressed.

Assessment of exposure to air pollution and noise:

Examinations and questionnaires (at baseline and during reassessment)

- Level 1: Questionnaire (incl. home and work addresses)

Secondary data: Geocoding data on air pollution, noise, and microclimate

Ionizing radiation is known to increase the risk for cancer and potentially for certain noncancer endpoints. Medical exposure is the major source of exposure to ionizing radiation in the

population and represents an important confounder or effect modifier for other risk factors (genetic, environmental, etc.). Cumulative medical exposure to radiation will be estimated from individuals' health insurance data on the use of diagnostic and therapeutic devices that employ ionizing radiation.

The effective dose from external environmental radiation exposure is estimated to lie in the range of 0.5 to 1.2 mSv per year in Germany (German Radiation Protection Commission 2008). No data exist for individual external environmental radiation exposure; however, representative data on the individual distribution of radon concentrations in homes have been collected recently⁴²³.

Assessment of exposure to ionizing radiation:

Examinations and questionnaires (at baseline and during reassessment)

Level 3: Questionnaire for medical exposure
Bracelet dosimeter for assessing background exposure

Secondary data: Health insurance data (estimation of medical exposure to ionizing radiation)
Geocoding data on background radiation

A.3 Study design

A.3.1 Population sampling and recruitment

The National Cohort will represent a population-based cohort of men and women aged 20 to 69 years, who will be recruited by invitation of a random sample of inhabitants of defined geographical areas in Germany. In total, approximately 200,000 cohort participants will be recruited.

A.3.1.1 Study regions and study population

Study regions

In June 2009, a call was launched for statements of interest in participating in assembling the National Cohort. This call was broadly distributed to national scientific epidemiologic organizations, medical faculties, governing bodies of universities, and interested scientists. Prerequisites for participating in recruitment included practical experience in conducting population-based, prospective cohort studies, experience in using standardized assess-

Figure 3.1: Recruitment clusters and study centers of the National Cohort



ment instruments, demonstration of successful scientific activities in chronic disease research, and a statement of support by the hosting institution. In addition, the applicants were asked to prepare a proposal for the suggested recruitment program in their region.

The statements of interest were reviewed by an international expert panel. Interested institutions within a given region formed regional clusters. As an outcome of the review process, it was decided that eight regional clusters – Bavaria, Baden-Wuerttemberg/Saarland, North Rhine-Westphalia, Saxony/Saxony-Anhalt, Berlin/Brandenburg, Lower Saxony/Hamburg/Bremen, Schleswig-Holstein, and Mecklenburg- Western Pomerania – with a total of 18 individual study centers distributed across Germany will perform the recruitment and follow-up of study participants of the National Cohort (**Figure 3.1**).

A.3

Table 3.1: *Intended sample size of the National Cohort, by cluster and study center*

Cluster	Study center	Intended sample size
Bavaria	Augsburg	20,000
	Regensburg	10,000
Baden-Wuerttemberg/Saarland	Heidelberg/Mannheim	10,000
	Freiburg	10,000
	Saarbruecken	10,000
North Rhine-Westphalia	Essen	10,000
	Muenster	10,000
	Duesseldorf	10,000
Saxony/Saxony-Anhalt	Halle	10,000
	Leipzig	10,000
Berlin/Brandenburg	Berlin-North	10,000
	Berlin-Center	10,000
	Berlin-South/Brandenburg	10,000
Lower Saxony/Hamburg/Bremen	Brunswick/Hanover	10,000
	Hamburg	10,000
	Bremen	10,000
Schleswig-Holstein	Kiel	10,000
Mecklenburg-Western Pomerania	Neubrandenburg	20,000
	Total	200,000

The numbers of study participants that each of the 18 study centers proposed to recruit needed to be adjusted to a total number of 200,000. The final decision regarding the number of subjects to be recruited by each center was guided by the strategic plan of arriving at an appropriate balance with respect to regional distribution (South/Central/North and East/West Germany), representativeness of the overall population of Germany, rural versus urban areas, and variation in regional indicators of SES (unemployment rate, poverty risk).

The agreed contribution of each cluster and study center in terms of sample size is given in **Table 3.1**. The figures show a distribution of 60,000 participants each in the southern (clusters Bavaria and Baden-Wuerttemberg/Saarland) and northern (clusters Lower Saxony/Hamburg/Bremen, Schleswig-Holstein, and Mecklenburg-Western Pomerania) areas

of Germany, plus 50,000 subjects in the central area (clusters North Rhine-Westphalia and Saxony/Saxony-Anhalt) and 30,000 subjects in Berlin-Brandenburg (cluster Berlin/Brandenburg). A comparison of the cohort size in eastern (former GDR; including Berlin) and western regions shows proportions of 35% and 65%, respectively, which differs from the proportions in the underlying population (20% East versus 80% West).

The distribution of cohort participants across broad groups defined by population density (shown in **Table 3.2**) is as follows: about 35% of the participants live in areas with less than 400 inhabitants/km² (rural areas); 30% of the cohort live in areas with 400 to 2,000 inhabitants/km²; and about 35% of the cohort live in large cities. In particular, the clusters Bavaria, Saxony/Saxony-Anhalt, and Mecklenburg-Western Pomerania represent areas with a low population density. For reasons of cost-effectiveness, standardization of examination procedures, and quality control of the data collection, each center will recruit a minimum of 10,000 participants.

Study population

The study population was selected to represent specific population and regional characteristics. Numbers and key characteristics of the selected source populations for the random selection of eligible participants are described in detail in **Table 3.2**. The data in **Table 3.2** show that the source population is very heterogeneous with respect to several indicators. The percentage of non-German citizens in the age range of 20–69 years lies between 1.5% and 25.9%; population density data as a simple indicator of rural versus urban areas indicates regions between 41 inhabitants per km² (Mecklenburg-Western Pomerania, Müritzt) and close to 4,000 inhabitants per km² (metropolitan area of Berlin); regions with high unemployment rates (up to 18.3%; Mecklenburg-Western Pomerania, Demmin) contrast with those with low rates (as low as 3.6%; Bavaria). These differences across the source population fulfill the requirements of establishing a cohort in Germany for studying health effects in various subgroups with sufficient statistical power while maintaining a population-based design.

The cohort will include 100,000 women and 100,000 men aged 20–69 years. Subjects aged 20–29 and 30–39 years will each contribute 10% persons to the cohort, while the three decades spanning the ages of 40–49, 50–59, and 60–69 years are slightly overrepresented, with each decade contributing 26.7% to the cohort. The greater proportion of middle-aged and older subjects is justified by their greater risk of developing chronic disease in the near future, thereby generating a larger number of disease endpoints. The rationale for including persons aged 20–39 years is the strategic goal of studying intermediate phenotypes, early functional changes, deteriorations in health status, and subclinical diseases. In order to detect and follow early functional changes, both younger and older subjects need to be included.

The age and sex distribution of participants for the overall National Cohort will be the same in each study center and in the intensified and basic examination subgroups (**Table 3.3**). Twenty percent of subjects in each age and sex stratum in each study center will go through the intensified baseline assessment program.

For the calibration substudy, 15% of the Level 2 participants will be reinvited within 12 months after recruitment (on average 6 months). Here, too, the age and sex distribution of the cohort will be followed.

Baseline recruitment will take 5 years, followed by a 5-year period of follow-up examination of the entire cohort. Essentially, this results in examination of 2,000 subjects per year in a study center with n=10,000 participants. Assuming 200 working days per year, 10 participants will be recruited each day, and 2 of those 10 subjects will go through the intensified assessment program.

Table 3.2: Description of the source population (at the administrative level of city or county), by study center

Study center, community	Source population ^{§*} (N)	Females ^{§*} (%)	Non-German citizens ^{§*} (%)	Subjects below age 20 [§] (% of total population)	Subjects with age 70 and above [§] (% of total population)	Population density [§] (total inhabitants per km ²)	Population movement over the past 10 years (in % / year)	Unemployment rate [§] (%)
Germany	54,799,164	49.7	10.1	19.0	14.1	229.6	+0.14	8.2
Bavaria								
Augsburg city	177,851	50.1	19.4	17.9	14.5	1793.0	+0.55	8.3
County of Augsburg	157,544	50.0	7.7	21.9	12.6	224.3	+0.28	3.8
County of Aichach-Friedberg	84,228	49.7	6.5	22.2	11.8	163.6	+0.46	3.6
Regensburg city	93,400	50.6	13.0	16.2	13.9	1656.6	+0.80	7.2
County of Regensburg	121,688	49.5	4.8	21.6	11.9	131.3	+0.51	3.6
Baden-Wuerttemberg/Saarland								
Mannheim city	215,144	49.0	25.9	16.4	13.4	2147.8	+0.19	4.2
Freiburg city	156,918	51.4	15.5	17.0	11.5	1435.2	+0.82	6.9
County of Emmendingen	103,333	50.2	7.4	21.4	13.1	231.9	+0.53	3.9
Saarbruecken city	221,509	50.4	11.1	18.5	15.4	817	-0.04	10.4
Saar-Palatinate county	97,480	50.1	6.8	19.0	16.4	360	-0.04	6.5
North Rhine-Westphalia								
Duesseldorf city	405,859	51.4	20.4	16.8	13.7	2689.7	+0.37	9.6
Essen city	383,586	50.5	13.6	17.9	16.0	2756.7	-0.01	12.3
Muelheim a. d. Ruhr city	109,537	51.0	11.3	17.9	17.0	1843.4	+0.10	8.4
Bochum city	257,462	54.4	12.8	16.8	15.2	2603.1	+0.01	10.3
Muenster city	189,812	52.5	7.1	18.0	12.7	904.1	+0.16	6.5
County of Muenster	1,696,035	50.0	8.6	21.3	13.6	377.1	+0.12	10.8
Saxony/Saxony-Anhalt								
Halle city	162,339	50.8	4.0	15.0	15.3	1725.8	-1.07	13.9
Saalekreis	141,078	48.7	1.6	15.0	14.9	140.4	-0.16	12.8
Leipzig city	363,208	49.5	7.3	14.4	15.2	1732.1	0.57	15.0
County of Leipzig	186,837	48.9	1.5	15.5	15.7	361.5	-0.14	12.8
Berlin-Brandenburg								
Berlin city	2,223,584	49.5	16.7	16.4	18.8 (>65)	3860	+0.12	14.1
Adjacent suburban municipalities in Brandenburg	679,417	48.7	2.3	16.9	16.6	265	+0.16	5.4

Study center, community	Source population ^{§*} (N)	Females ^{§*} (%)	Non-German citizens ^{§*} (%)	Subjects below age 20 [§] (% of total population)	Subjects with age 70 and above [§] (% of total population)	Population density [§] (total inhabitants per km ²)	Population movement over the past 10 years (in % / year)	Unemployment rate (%)
Lower Saxony/Hamburg/Bremen								
Hamburg city	1,235,336	49.8	16.9	17.3	13.0	2346.2	+0.52	8.6
Bremen city	371,356	50.1	15.3	17.3	14.8	1683.0	+0.34	11.1
Delmenhorst city	49,286	49.9	10.1	20.1	14.0	1195.0	-0.27	10.7
Lilienthal	11,900	50.1	4.9	21.0	13.5	254.0	+0.59	5.6
South-West Bremen:								
Northern county of Vechta	71,800	48.8	8.6	24.7	15.2	178.8	+0.54	4.3
Parts of county of Oldenburg (Municipalities Ganderkesse, Doettingen, Wildeshausen)	36,400	49.5	5.2	22.2	12.3	170.0	+0.74	9.4
Northeast county of Cloppenburg	17,100	48.4	5.7	26.1	11.2	93.1	+0.38	6.6
Municipality Diepholz	10,750	49.1	5.2	22.2	12.3	158.4	+0.45	5.3
Municipality Hude	10,260	49.5	5.2	22.2	12.3	127.0	+0.74	9.4
Brunswick/Hanover:								
Brunswick city	168,813	48.8	11.2	16.8	14.7	1,287	+0.24	9.3
County of Peine	86,985	49.6	7.0	21.3	14.2	247.9	+0.40	7.7
Eastern region of Hanover	182,700	50.4	11.7	18.5	14.3	288.5	+0.23	9.3
Hanover city	358,804	50.5	16.4	16.8	13.7	2543.4	+0.21	10.7
Schleswig-Holstein								
Kiel city	170,916	48.7	9.4	16.8	12.6	2002.4	+0.42	10.9
County of Rendsburg-Eckernförde	175,381	59.9	3.6	21.5	13.9	124.1	+0.20	6.1
Mecklenburg-Western Pomerania								
Neubrandenburg city	52,899	49.8	2.3	14.9	14.5	866.5	-1.28	15.1
County of Demmin	64,010	47.8	1.7	15.9	16.0	50.5	-1.22	18.3
County of Uecker-Randow	58,366	47.5	3.6	15.1	16.2	53.8	-1.14	17.2
County of Mecklenburg-Strelitz	59,025	54.0	1.6	15.6	15.1	42.4	-0.79	15.1
County of Mueritz	47,038	54.7	1.7	15.9	15.0	41.0	-0.42	12.7

* figures are related to subjects in the age range of 20-69 years;

§: data on dec 31 2008,

§: data for the year 2009

(main data source: <https://www.regionalstatistik.de/genesis/online/>, status September 2010)

Table 3.3: Distribution of cohort participants by sex and 10-year age groups for the entire cohort and for a given study center with n=10,000 participants (overall and per year) (Note: participants of MRI are described in Sect. A.4.4)

	Age groups (years)					
	Total	20-<30	30-<40	40-<50	50-<60	60-<70
Entire cohort						
Total (n)	200,000 100%	20,000 10%	20,000 10%	53,333 26.7%	53,333 26.7%	53,333 26.7%
Females (n)	100,000 50%	10,000 5%	10,000 5%	26,667 13.3%	26,667 13.3%	26,667 13.3%
Males (n)	100,000 50%	10,000 5%	10,000 5%	26,667 13.3%	26,667 13.3%	26,667 13.3%
Level 2 (n)	40,000 20%	4,000 2%	4,000 2%	10,667 5.3%	10,667 5.3%	10,667 5.3%
Calibration sub-study (n)	6,000	600	600	1,600	1,600	1,600
Per recruitment center with n=10,000						
Total (n)	10,000 100%	1,000 10%	1,000 10%	2,667 26.7%	2,667 26.7%	2,667 26.7%
Females (n)	5,000 50%	500 5%	500 5%	1,333 13.3%	1,333 13.3%	1,333 13.3%
Males (n)	5,000 50%	500 5%	500 5%	1,333 13.3%	1,333 13.3%	1,333 13.3%
Level 2 (n)	2,000 20%	200 2%	200 2%	533 5.3%	533 5.3%	533 5.3%
Calibration sub-study (n)	300	30	30	80	80	80
Per year (5 year recruitment phase) and recruitment center with n=10,000						
Total (n)	2,000	200	200	533	533	533
Females (n)	1,000	100	100	267	267	267
Males (n)	1,000	100	100	267	267	267
Level 2 (n)	400	40	40	140	140	140
Calibration sub-study (n)	60	6	6	16	16	16

%, given as percentage of the total cohort (except for the calibration sub-study);

20% of subjects will undergo the intensified baseline program (level 2);

15 % of the level-2 participants will take part in the calibration sub-study

A.3.1.2 Population sampling

Sampling procedure

A random sample of the underlying source population, stratified by sex and 10-year age groups, will be drawn from population registries.

In Germany, all inhabitants are obliged to register with population registries. Upon request, the local population registries will select a sex- and age-stratified random sample of inhabitants and will transmit the following data for each subject:

- ▶ First name, surname
- ▶ Gender
- ▶ Full postal address
- ▶ Date of birth
- ▶ Nationality
- ▶ Date of move to current address

The personal data of the population sample as identified by the population registries will be transferred to and stored in a database located at the local study center in accordance with data protection rules and concepts (see **Sect. A.4**). The data will be used to initiate the recruitment process.

Random samples from the population registries will be collected twice (at the beginning of the study and again 2.5 years later) during the 5-year recruitment period.

Inclusion and exclusion criteria

The selection of eligible subjects will adhere to the following general inclusion and exclusion criteria:

Subjects are eligible for inclusion in the random sample from the population registry if:

- ▶ They are 20–69 years of age at a fixed time point (middle of each 2.5-year recruitment periods for which a random sample will be drawn).
- ▶ Their primary residence is within the selected study area.

Subjects will be excluded from baseline recruitment if:

- ▶ They are not capable of providing informed consent, including not being capable of understanding the study information.
- ▶ They are not able to respond to the questions and are unable to participate in most examinations.

Exclusion criteria will be checked when establishing contact with the invited participant and before signing the informed consent form. Detailed instructions will be provided to the staff in charge of contacting the subjects by telephone and of obtaining the informed consent.

Selection of subjects for the intensified baseline assessment program

A 20% random subsample of each stratum of the sex and 10-year age groups will be invited to participate in the intensified assessment program. Subjects included in the sample provided by the population registries will be randomly assigned either to the intensified subgroup or to the basic assessment subgroup by the respective study center. Randomization will be performed before initiating the contacting process since the informational material offered to potential study subjects is distinct from that provided to participants of the core baseline program. Self-selection of participation into the core or the intensified program by

the participants themselves will not be possible. No oversampling of a specific population subgroup is planned.

Selection of subjects for the calibration substudy

At recruitment, all Level 2 participants will be asked whether they are willing to take part in a repeated examination within 12 months (on average 6 months). According to the given sex and age distribution of the cohort, subjects agreeing to participate in the reassessment will be consecutively invited for examinations until the required numbers in each stratum and year (see **Table 3.3**) are reached.

A.3.1.3 Participation rates

A.3

Based on previous experience from population-based studies in Germany, it is estimated that at least 400,000 primary invitations must be mailed in order to recruit 200,000 subjects into the cohort, ultimately resulting in a participation proportion of about 50%, following a multistep invitation and recruitment process (see below). We aim for a high response proportion in order to obtain a reasonably representative sample of the underlying population.

Measures to achieve high response rates

The high participation proportion rate of 50% will be achieved by employing a comprehensive contacting and reminder program (see below). We will ensure that population subgroups with an expected low participation rate (e.g., subjects with a low SES) are well represented in the cohort by instituting a number of measures. For example, a public relations representative of the National Cohort responsible for general public relations activities will provide information on the cohort as a whole. In addition, each study region will authorize a communication expert who will organize public relations for the recruitment area. The aim is to implement a number of local public relations activities (adapted to local circumstances) that raise and maintain awareness of the National Cohort. These may involve:

- ▶ Information and integration of local institutions and stakeholders such as local health authorities, general practitioners, pharmacies, mayors, etc.
- ▶ Declaration of support by official representatives of health organizations (e.g., Ministry of Health, Ministry of Education and Research) and the responsible data privacy commissioners by letters of support to be mailed with the invitation letters
- ▶ Professionally designed information and advertising material (flyer, postcards, posters, etc.)
- ▶ A service telephone and a webpage to provide information on the National Cohort
- ▶ Media campaigns including coverage of issues related to the National Cohort in print media, radio, and television:
 - Features and reports in national and local press, radio and television
 - Press releases featuring interviews with members of the Epidemiologic Planning Committee, local principal investigators, local health ministers, and celebrities
 - The possibility that national and local reporters and film teams visit the study centers, create short movie clips, and interview study center staff as well as potential participants
 - Local and national press reports at certain dates (e.g., at the beginning of the study, halftime, end of baseline recruitment)
- ▶ Articles in the national scientific press (e.g., *Ärzteblatt*)

Furthermore, the promotion campaign will be supported by national and regional organizations and representatives of medical research and health care systems, such as health insurances and scientific associations, e.g., the German Epidemiologic Association (DGEpi).

Important aspects for enhancing individual participation rates include:

- ▶ Each participant will receive selected results of his/her medical examinations conducted in the study center as well as selected results of the clinical laboratory analysis of blood and urine samples.
- ▶ Reimbursement of travel costs will be offered.
- ▶ Sponsoring incentives by businesses and other institutions.
- ▶ Tombolas with prizes for participants.

According to experience from population-based cohort studies in Germany, a high participation rate during recruitment does not interfere with a very high participation rate during follow-up of the cohort, both for active follow-up by means of questionnaires as well as for repeated examinations during follow-up (see **Sects. A.3.6** and **A.3.7**).

A.3.1.4 Recruitment of study participants

After the addresses have been drawn from local population registries, the local study centers will invite eligible study subjects and perform subsequent administrative tasks. Depending on the local infrastructure, contacting and invitation of subjects may be organized either through individual organizational units located within the institution responsible for regional recruitment or in the actual local study center. The term "study center" as used here refers to the center where the assessment, i.e., interview and examination of the study subjects, will take place. Administrative tasks include: organizing the invitation, mailing invitation letters and study materials, contacting potential participants by telephone, scheduling (and administering) appointments for examinations, organizing feedback of examination results (medical reports) to study participants, and documenting contacts and recording correspondence with participants. Another important task to be performed locally is quality assurance of recruitment procedures (invitation, phone contacting, and documenting reasons for non-response or nonparticipation) and the actual data collection. The latter will involve checking for completeness of study documentation at the individual level, ensuring data storage and data transfer processes and basic plausibility checks of collected data.

Recruitment strategy: written invitation, phone contacts, and home visits

The invitation mailing will be organized by each study center and will follow a general strategy. Study subjects will be recruited by inviting consecutively drawn subsamples of the original population sample at random.

The recruitment strategy includes at least two written invitations and active contact attempts by study personnel via the telephone or home visit if the eligible person does not respond to the written invitation. The mailed invitation contains a letter of invitation, a description of the aims and examination procedures of the study, a copy of the approval letter from the responsible Data Privacy Commissioner, and, if possible, letters of support from regional authorities and project promoters / advocates. Recruitment strategies will vary according to whether a subject has a home telephone number that can be identified from the official telephone directories or whether a telephone number is not available.

After a first letter of invitation, persons with known telephone numbers will receive a subsequent phone call by study personnel in the course of the following days. At the same time, subjects are provided the phone number of the local study center for scheduling appointments or for obtaining additional information. If the person does not respond actively to the invitation letter, study personnel make at least 10 attempts to establish telephone contact at different times of the day and on different days of the week. If no contact can be made within 4 weeks, a reminder invitation is mailed to the subject, offering the option of replying

with a prepared prepaid reply card indicating convenient examination dates. If the second invitation does not lead to a contact with the subject or has not verified their response state (e. g., moved away with unknown address, too ill to participate in the study, deceased, or unwilling to participate in the study), a home visit may be conducted by trained study personnel. Several attempts will be made to locate the study subject or to retrieve information from relatives or neighbors regarding the subject's correct address of residence or telephone number, or their willingness or ability to participate. For subjects without a registered telephone number two invitation letters including prepaid reply cards will be sent before considering a home visit. All direct contact attempts are made by trained and certified study personnel and are documented in detail.

In order to accommodate the needs and time restraints of an active working population, active phone calls aimed at contacting potential study subjects will cover all times of the day and evening as well as weekends.

A.3

Important organizational tasks to be performed by the study personnel responsible for contacting the invited subjects and organizing the recruitment are:

- ▶ To schedule an appointment for the examination
- ▶ To confirm or change a prebooked appointment
- ▶ To contact people who have not responded to the reminder letter
- ▶ To record telephone/cell phone number and e-mail details in case the appointment must be canceled or changed
- ▶ To send email or SMS text reminders before the appointment
- ▶ To contact people who missed their appointment
- ▶ To supply nonresponder questionnaires or to conduct short nonresponder interviews with subjects unwilling to participate in the study
- ▶ To provide information about the study, procedures involved in the recruitment, organizational aspects, and to clarify concerns
- ▶ To cancel the invitation and ensure the invitee is not contacted further (in case of withdrawal)

Previsit reminder message

When potential participants agree to participate, they will be asked to provide an e-mail address and/or cell phone number. With the participant's agreement, these contact details may be used to send an appointment reminder by e-mail or SMS text shortly before their scheduled appointment.

Written material provided to the study participants prior to the appointment

The first invitation mailing will include a participant information leaflet providing basic information about the National Cohort, including the aims of the study, funders, and organization. Furthermore, the information leaflet will briefly inform potential participants about the examinations and interview. A reminder letter will contain the same information material.

A letter confirming the date of examination will be sent to subjects who agreed to participate in the study and contains a map of the study center, a brief, self-administered questionnaire, and specific instructions on how to prepare for the study (e. g., fasting state if needed), and what to bring to the study center (e.g., packages of medication, reading glasses, or completed questionnaire).

It will also provide the address of the study website with further information and the telephone number for further questions. Prior to the assessment, participants will receive more

detailed study information (see **Sect. A.7.1.1**). Moreover, participants will receive an informed consent form which has to be signed prior to the assessment.

Procedure after missed appointments

It is expected that about 10% of potential participants with scheduled appointments will miss an appointment, some of whom will not cancel the scheduled appointment beforehand. Subjects who missed their appointment will be recontacted by phone or will be sent a further letter to invite them to schedule a new appointment.

The method of recruitment described here has been selected as the most efficient strategy to achieve the desired response rate.

A.3.1.5 Contact results, response status, and nonresponder information

Results of contact attempts: response status

As a result of attempts to establish contact with the invited subjects, the subject's response can be classified according to the following response categories: subject participated, subject was excluded prior to examination, subject refused participation, and subject could not be contacted. Reasons for exclusion can be that the participant died prior to examination, moved away from the study region prior to invitation, or was unable to participate in the examinations due to severe illness. Classifying a subject as excluded requires explicit information on the reason for exclusion. Subjects who actively express their refusal to participate or who do not participate in spite of successfully having been contacted, or who cannot be contacted at all are classified as refusals for the purpose of calculating the response proportions.

Verification of address and vital status

If a subject cannot be located at the address provided by the population registry, and no other information about their current whereabouts can be obtained, the population registry will be requested to check and provide the current address and vital status of the subject. Likewise, if an invitation letter is returned by the post as "moved away to unknown address" or "deceased," the population registry will be asked to verify vital status and address of the subject and to provide, if applicable, their current address or date of death.

Nonresponder information

Subjects choosing not to participate in the study will be asked to give the reason for their unwillingness or inability to participate. This information is necessary to classify them as refusals or exclusions from the cohort and in order to better tailor the recruitment strategy, improve communication of study aims, and help better argue against potentially modifiable reasons for nonresponse.

Moreover, subjects who cannot participate or are unwilling to participate in the examinations will be asked to provide basic sociodemographic information and information on selected health-related behaviors and diseases in order to assess potential selection bias due to nonresponse. Nonresponder questionnaires may be answered by telephone or written questionnaires.

A.3.1.6 Equipment and organization of local study centers

In order to reduce barriers for participation and to increase participation rates, study centers should be located in an acceptable proximity to the homes of the invited subjects. For certain

rural areas, this may require the inclusion of selected communities within the greater rural area. The study centers will either be newly created in academic or clinical research facilities or will be based in pre-existing study centers. Each study center will be identically equipped for examinations that will be performed on Level 1 or Level 2 subjects, i.e., in the full cohort or in the cohort of 40,000 subjects included in the intensified examination program. Additionally, all centers need to be identically equipped with the necessary information technology (IT) facilities, including touch screens to capture self-administered questionnaire data and PCs for conducting computer-assisted interviews. Network connections to servers on which the data will be stored and internet access will be required.

Room and space requirements

A.3

Each study center will provide dedicated space for carrying out medical examinations (e. g., anthropometric measurements, ultrasound of carotid arteries, ophthalmological examinations, exercise tests, and drawing of venous blood samples) as well as space for conducting interviews, and space equipped with touch screens and furnished with tables for completing self-administered questionnaires. Laboratory facilities for preanalytical processing and interim storage of blood, urine, and other biological samples will also be needed. This includes the provision of deep-freeze storage facilities. Moreover, all centers need to be equipped with appropriate toilettes for urine sampling, including rest rooms for handicapped persons. Easy accessibility to the study center (e.g., availability of an elevator) must be ensured.

In order to enable the examination of at least 10 subjects per day with an examination and interview program of 2.5 or 4 hours as described in **Sects. A.3.2** and **A.3.3** (Baseline assessments at Levels 1 and 2), a number of rooms equipped for the following examination modules, interview, and other purposes must be provided:

- ▶ Reception/front desk area, including a waiting area for study subjects
- ▶ Two rooms for personal interviews
- ▶ One larger area with cubicles for touch screen-based self-administered questionnaires and tests (incl. selected tests of cognitive function, speech-in-noise test)
- ▶ One kitchen and breakfast area for subjects who participate in the OGTT
- ▶ One laboratory, including centrifuges, refrigerators, and deep freezer storage space
- ▶ Rooms for general examinations (including anthropometric measures, blood pressure measurements, resting ECG, pulse-wave analysis, hand grip, spirometry, AGE reader, venipuncture)
- ▶ Rooms for specialized measurements, such as
 - Ultrasound (e.g., carotid ultrasound, 3D-Echocardiography)
 - Cardiorespiratory function testing (step test)
 - Sensory measurements (ophthalmological measurements, sniffing sticks)
 - Neuropsychiatric function testing (face-to-face), posturometry / balance tests, examination of the musculoskeletal system
 - Dental examinations
 - DXA scan
- ▶ Several rooms for study personnel, e. g.,
 - One separate room for the study physician who performs the final discussion of examination results with the study participants
 - One room for medical documentalists, including archive space
 - One room for technical assistants who prepare and complete study documentation
- ▶ Toilets (suited for male, female, and handicapped persons)

The respective rooms should comply with mandatory standards for drawing and processing of blood and other specimens. Thus, they must be floored with suitable material permitting adequate disinfection and cleaning, and adequate washing facilities must be provided.

For some of the examinations, room temperature must be maintained at a certain prespecified level (e.g., at least 22° C for the rooms where blood pressure measurements and ECGs are recorded). Thus, provision of air conditioning in some of the rooms is mandatory (including the storage space for –80°C freezers).

Personnel and equipment needed for each study center

All study centers need to be fully equipped with all devices and instruments required to perform the data collection, including medical examinations, interviews, laboratory processing, and storage of biosamples. Provision of the necessary IT environment, data storage capacities, and web access as outlined in **Sect. A.4** (Integrated data management) is mandatory for all study centers, and archiving cabinets for study documentation of the participants must be in place. In addition, four of the study centers will provide dedicated MRI facilities in order to perform the specialized imaging protocol (see **Sect. A.3.4**, Assessments through medical imaging).

The personnel requirement for each center will depend on its capacity, the intended sample size, the expected number of assessments to be carried out, and its opening hours. Personnel skill-mix will include study nurses, health care technicians, data manager, receptionists, and physicians.

Logistic organization of participants' visit, interview, and physical examinations

The staff at the study centers will conduct the interviews and examinations of the study participant. The responsibility for organizing a smooth examination process of 10 or more subjects per center per day lies with the head and the personnel of the individual study centers. The appointment scheduling will be organized by each coordinating/study center following the standardized procedure as described above.

After welcoming and registering the participant at the study center, it will be ensured that the participant has understood the study information and all examination procedures; completion of the informed consent form will be checked.

The core program including examinations, interview, and collection of biosamples (Level 1) will last about 2.5 h on average, while the intensified program (Level 2) will require about 4 h. Since participants in the intensified program will need to be fasting for the OGTT, their appointments will start between 7 a.m. and 10.30 a.m. Subjects participating in the core program only may not be fasting and can start their appointments later in the day.

Workflow will be standardized according to a detailed protocol that must be adhered to at each study center. This will increase adherence to the envisaged examination times, help standardize data collection across centers, and neutralize the effect that certain aspects related to the examination order may have on study participants. The optimal workflow will be tested during the pilot study and will subsequently be implemented in the main study. Lengthy waiting times between the examination procedures and frequent changing of rooms will be avoided. Therefore, as many examinations as possible will be conducted in the same room. For example, the measurement of height, weight, waist, and hip circumferences, blood pressure, and the recording of the resting ECG can be conducted in one room so that participants do not need to take off their shoes and clothes several times.

A.3.2 Baseline interview and questionnaires

Basic information about sociodemographic variables, medical history, medication use, and lifestyle will be assessed by computer-assisted, face-to-face interviews. Furthermore, self-administered questionnaires will be applied via touch screen during the visit to the study center, as a web-based questionnaire, or as a traditional paper-and-pencil method. As for the physical examinations, the baseline interview and questionnaires will focus on further information on major pre-existing diseases, subclinical intermediates, and functions pertaining to CVD, diabetes, cancer, neurologic and psychiatric diseases, respiratory diseases, and infections, as well as on the assessment of potential risk factors for these diseases. Through the combination of extensive information on physical examinations, biomaterials, and risk factors, a wide range of innovative research questions can be addressed.

A.3

This section is subdivided into core modules (**A.3.2.1**) covering more general topics which will be applied to all 200,000 study participants and specific modules (**A.3.2.2**) covering topics that were identified and proposed by the thematic expert working groups. These working groups provided information on possible short and valid instruments covering different organs and systems, disease groups, and exposures. They also reviewed specific questionnaire modules. For some of the topics, specific modules will only be administered to the intensively phenotyped subgroup (Level 2, n=40,000).

The envisaged face-to-face total interview time of the modules presented below will be about 1 h for all 200,000 participants, supplemented by self-administered questionnaires by touch screen or on paper (partly to be filled in at home), for which the estimated time of completing the questionnaires will be less than 1 h. Personnel time only needs to be calculated for the interview. Self-administered questionnaires do not require continuous presence of study personnel, although support may be offered by trained receptionists as needed.

The web-based questionnaires which can be filled out at the participant's convenience from any location with internet access (e.g., at home) may take up to 1 h for completion. Some of the instruments will be completed around the time of the examination visit, but some of the web-based instruments are planned for later completion by the participant. Web access for completion of the respective instruments will also be offered to the participants at the study centers.

A.3.2.1 Core interview and questionnaires

An overview of the modules of the baseline core interview and questionnaires is given in **Table 3.4**, and some additional information on the modules is provided in the following text.

Sociodemographic and socioeconomic factors

As further outlined in **Sect. A.2.4.5**, there is an inverse association of SES with morbidity and mortality for a wide range of chronic diseases and with their associated risk factors^{319, 325, 327, 332, 339, 424, 425}. It has also been shown that SES at different stages of the course of life has an independent and cumulative effect on health outcomes and mortality^{344, 345, 349, 426-429}. Moreover, general sociodemographic factors are potential confounders and need to be collected routinely for epidemiologic studies.

Instruments:

General sociodemographic information (e.g., age, gender, place of birth, ethnicity/migration background, marital status, partnership, and offspring) and information on SES will be obtained using standardized instruments that have been used successfully in existing German

Table 3.4: Modules of the core interview and questionnaires

Modules	Assessment instruments			
	Inter-view	Touch Screen	SAQ (paper)	Web-based
Socio-demographic and socio-economic factors Age and gender, place of birth, ethnicity, migration background, living situation, marital status, partnership, children, education (participant, partner, parents), occupation (participant, parents), and income	X			
Medical history of the participant CVD, diabetes and other metabolic diseases, cancer, neurologic and psychiatric diseases, respiratory diseases, infections; musculoskeletal diseases, allergies, skin diseases, gastro-intestinal diseases, gallbladder and hepatic disease, kidney and urinary system diseases, eye- and ear-related diseases, operations, accidents, hospitalization during past 12 months	X			
Family history (medical history of parents and siblings) List of diseases (diabetes, myocardial infarction, stroke, asthma, severe depression, Parkinson's disease, dementia/Alzheimer disease, cancer); Age at diagnosis for the parents and number of affected brothers/sisters		X		
Medication use Medication use during past 7 days, including ATC codes, frequency, dosage and duration of use, prescription information per drug used (IDOM)	X			
Screening Cancer screening and health check up	X			
Women's questionnaire Menstruation, contraception, fertility, pregnancy, menopause, hormone replacement therapy, women-specific diseases (breast and gynecological disease) and operations	X			
Men's questionnaire Men-specific diseases (prostatic hyperplasia)		X		
Smoking Active smoking (past, current, occasional, never, average amount of smoking); Passive smoking	X			

Modules	Assessment instruments			
	Inter-view	Touch Screen	SAQ (paper)	Web-based
Alcohol				
Alcohol consumption (frequency and amount of drinking)	X			
Binge drinking pattern (short version of HAPIEE questionnaire)	X			
Alcohol abuse and dependence (BASIC questionnaire)		X		
Alcohol addiction: CAGE questionnaire		X		
Drugs				
Drug use (list of nine drugs, ever use, use during the last 12 months)		X		
Health-related quality of life				
Self-rated health, Health related quality of life (WHOQOL BREF)		X		
Health-related quality of life: QALY (EQ-5D)			X	
Other personal characteristics				
Adoption, siblings		X		
Early life factors			X	
Phenotype: natural hair color, skin complexion, eye color		X		
Weight history			X	

cohorts according to recommendations published by the Federal statistical office and by the German Epidemiologic Association⁴³⁰⁻⁴³². To assess the SES, information on own education, occupation, and income as well as parental SES at the time of the participant's childhood and the partner's SES will be obtained for operationalization of different concepts of SES, including childhood SES and SES across the course of life.

Medical history of the participant

In order to be able to correctly identify incident cases of the diseases in the focus of the National Cohort, it is important to recognize prevalent cases of these diseases at the time of recruitment. Moreover, it is important to obtain information on comorbidity and treatment since several diseases or their respective treatments may affect each other (e.g., allergies and cancer^{433, 434}, renal disease and diabetes, and corticotherapy and osteoporosis).

Instruments:

Previous medical conditions and general health status will be acquired via standardized computer-assisted interview using questions that have been used successfully in previous epidemiologic studies. Standard questions for a wide range of prevalent diseases (with the focus on diabetes, cancer, cardiovascular, respiratory, neurologic, psychiatric, and infectious diseases) include physician's diagnosis of disease, date of diagnosis, occurrence during the last 12 months, and treatment during the last 12 months. For some diseases in the focus of the National Cohort (e.g., cancer, CVD, diabetes mellitus, stroke, and neurologic and psychiatric diseases and functions), additional specific questionnaire modules will be applied. Moreover, several questions cover additional diseases (such as eye- and ear-related and gastrointestinal diseases), operations, and accidents.

Family history

Individuals with a positive family history for many types of diseases [diabetes, CVD, and cancer (breast, ovarian, colorectal, prostate)] have an increased risk of developing the disease themselves^{339, 435-437}. This may be due to shared genes and genetic susceptibilities, but also to a shared environment and common behaviors among family members.

Instruments:

Family history of disease will be collected for first-degree relatives (parents and siblings) for those most important diseases in the focus of the National Cohort because information on family history can be reliably obtained (for details, see **Table 3.4**) In addition, age at death and cause of death for parents will be recorded.

Medication

Information on medication currently taken by the participant is needed for several purposes: it may be used for the definition of specific diagnoses (e.g., antidiabetic medication as one criterion for prevalent diabetes mellitus), but also because certain drugs may affect physiological mechanisms, biomarkers, factors, and diseases that may themselves be studied as main exposures or outcomes (e.g., cardioactive medication and heart rate variability, blood pressure, statins, and cancer⁴³⁸⁻⁴⁴²). Additionally, detailed drug information can provide the basis for pharmacoepidemiologic research questions.

Instruments:

All medications taken by the participant during the past 7 days will be recorded using the computer-assisted Instrument for Database assisted Online recording for Medication (IDOM) software (developed by GSF Munich)⁴⁴³ that is based on the current statutory health insurance medication database for Germany⁴⁴⁴ and with which each drug can be exactly classified according to the anatomical therapeutic chemical classification system (ATC code). For each drug, information on prescription versus over-the-counter, frequency, amount, and duration of intake will be recorded, and defined daily doses (DDD) can be derived.

Screening and health check-ups

Health-related behavior also includes the use of preventive health services. In order to identify target groups for tailored secondary prevention strategies, it is important to obtain information on the use of preventive health services. Moreover, the use of preventive health services may have to be adjusted for as a confounder to address potential biases due to increased detection of disease in screened populations.

Instruments:

Information on use of health services for cancer screening and general health check-ups (e.g., for cardiovascular, renal, or metabolic diseases) will be obtained with a few questions.

Specific questionnaire for women and men

Onset of menarche and menopause, reproductive history, and use of exogenous hormones may have a significant impact on women's health, mainly due to hormonal effects.

For instance, early menarche has been demonstrated to be associated with increased adult BMI and CVD risk factors⁴⁴⁵⁻⁴⁴⁷, menopause is associated with depressive disorders and sleep disturbances^{448, 449}, and long-term health effects associated with the hormonal changes of menopause include an increased risk of osteoporosis and ovarian and breast cancer. The intake of oral contraceptives has been found to modify the risk of breast, cervical, or

ovarian carcinoma⁴⁵⁰ and to be associated with an increased risk of thromboembolism and cardiovascular events^{451, 452}. Hormone replacement therapy has been shown to be associated with risk of breast and ovarian cancer⁴⁵⁰ and CVD⁴⁵³. Therefore, we will obtain detailed information on menarche, reproductive history, menopause, and hormone use.

Male lower urinary tract symptoms, benign prostatic hyperplasia, enlargement of the prostate, and bladder outlet obstruction are common among aging men and will increase in socioeconomic and medical importance at a time of increased life expectancy and aging^{454, 455}. For elderly men, benign prostatic hyperplasia has considerable influence on quality of life.

Instruments:

In the women's questionnaire, information will be gathered about menstruation (date of onset, duration of the cycles), pregnancies (number, outcomes), breast feeding (duration for each child), menopause (date of onset, symptoms), and intake of exogenous hormones (for birth control and hormone replacement therapy) and about breast and other gynecological diseases and operations.

For men, we will obtain information on benign prostatic hyperplasia. Further specific questions for men that are still under discussion are related to, for example hormonal intake, fertility, and erectile function.

Smoking, alcohol, and drugs

The leading lifestyle exposures constituting risk factors for a wide range of diseases are smoking²⁶⁸ and alcohol consumption²⁶⁷ and need to be taken into account as confounders in many analyses. Thus, we will aim to quantify life-long tobacco and alcohol consumption and information on drinking patterns. Drug use is also a phenomenon of increasing concern as there is an increasing prevalence of use in the younger population⁴⁵⁶ and because some drugs have been shown to induce or "exacerbate" psychiatric disorders and injuries. For instance, the association between cannabis consumption and schizophrenia is under discussion^{457, 458}, and there may be an association with mental health. Therefore, we will obtain information on use of selected illegal drugs.

Instruments:

In order to be able to quantify smoking history, information on life-long smoking habits – active as well as passive – will be obtained, complying with published recommendations for questionnaires on smoking habits⁴⁵⁹. The questionnaire for tobacco use that follows published recommendations for questionnaires on smoking⁴⁵⁹ covers regular and occasional smoking, former smoking, smoking duration, periods of quitting, and average number smoked for filter, nonfilter and hand-rolled cigarettes, cigars, cigarillos, and pipes. Information on passive smoking will be collected using questions concerning home, working, and leisure environment.

Questionnaire modules on alcohol consumption will enable us to quantify alcohol consumption and to assess drinking patterns and have been adapted from previous epidemiologic studies (e.g., KORA, EPIC, and HAPIEE study). Alcohol consumption will be assessed by using questions about usual frequency, quantity, and type of alcohol consumed during the past 12 months such that alcohol consumption can be calculated in g/day. Consumption during weekdays and weekends will be documented separately. The Alcohol abuse and dependence (BASIC) questionnaire⁴⁶⁰ is a 6-item, short tool that can be used to screen these conditions and has previously been used in population-based studies in Germany. It can also be used to classify risky alcohol consumption. Binge drinking will be assessed using the modified HAPIEE questionnaire⁴⁶¹.

For the National Cohort, we adapted questions about drug consumption that have been used in other studies (Bundeszentrale für gesundheitliche Aufklärung⁴⁶², National Health and Nutrition Examination Survey (NHANES)⁴⁶³, LifeGene⁴⁶⁴, Constances⁴⁶⁵). Drugs that will be covered in the questionnaires include cannabis, ecstasy, LSD, amphetamine, cocaine, crack, heroin, sniffing agents, psychoactive plants, and doping substances, and lifetime and 12-month prevalence of use of the respective drugs will be queried.

Health-related quality of life, subjective health

Self-rated (subjective) health and health-related quality of life (HRQoL) are important factors related to the risk for chronic diseases (physical and mental) as well as to the degree of health care utilization. The concepts of subjective health and HRQoL include essential information about the participant's perception of their physical and mental health status⁴⁶⁶.

It has been shown that self-rated health is a valid indicator for overall health status and mortality. As shown in a review of almost 30 studies, self-rated health predicts mortality, independently of other relevant factors⁴⁶⁷. Due to the predictive value of subjective health and HRQoL, these concepts are of particular public health relevance. Evaluating the impact of several diseases on subjective health status and HRQoL, on the one hand, and self-rated health and HRQoL as exposure variables for diseases, on the other, constitute interesting topics for future research.

Instruments:

Validated and internationally established instruments will be used to assess health-related quality of life (WHOQOL-Bref⁴⁶⁸, EQ-5D^{469, 470}). The WHOQOL-Bref assesses the individual's perceptions in the context of their culture and value systems and their personal goals, standards, and concerns. It measures the following broad domains: physical health, psychological health, social relationships, and environment. The EQ-5D instrument covers the five dimensions mobility, self-care, usual activities, pain/discomfort, and anxiety/depression and can be used to calculate aggregated outcomes as quality-adjusted life years (QALYs), e.g., by using the algorithm developed by Greiner et al.⁴⁷¹.

Other personal characteristics

Additional information on personal characteristics and early life factors that may be associated with subsequent disease or that are important to correctly interpret associations with other factors will be obtained with short questionnaire modules. The modules will include questions related to, for example, adoption versus bodily descent from parents, siblings, having been breast fed, birth weight, weight history, and phenotypic information on complexion, hair, and eye color.

A.3.2.2 Specific interview and questionnaire modules

An overview of specific interview and questionnaire modules that were mainly proposed by the thematic working groups is given in **Table 3.5**.

For the more intensively investigated subgroup (n=40,000, Level 2), several questionnaire modules will be more comprehensive than those used for Level 1 participants, partly guided by additional physical examinations included in the intensified program that may require more in-depth questionnaire information.

Table 3.5: Specific interview and questionnaire modules

Modules	Assessment instruments			
	Inter-view	Touch Screen	SAQ (paper)	Web-based
Neurologic and psychiatric factors *				
Psychiatric diagnoses screening questions (parts of MINI Internat. Neuropsychiatric Interview)	X			
Depressive symptoms (CES-D)	X			
Anxiety sensitivity index questionnaire (ASI-3)		X		
Restless legs (RLS) questionnaire		X		
Headache questionnaire		X		
Sleep disturbance (PSQI)		X		
Psychosocial factors				
Personality (BFI-10)		X		
Chronic stress screening scale (TICS)		X		
Childhood trauma questionnaire (short version)		X		
Effort reward imbalance at work (ERI)		X		
<i>Occupational demands, worries, leadership, work-family conflict (COPSOQ) (level 2 only)</i>				X
<i>Occupational insecurity (BIBB/IAB) (level 2 only)</i>				X
Social networks & social support (Berkman & Syme 2004, with Berkman adapted version for Germany)		X		
Infections and immune function *				
Infections (respiratory, gastro-intestinal, etc.)		X		
Vaccination				
Contact with animals				
Musculoskeletal diseases *				
<i>Pain mannequin (body regions to be marked + pain intensity and possible impairment in daily activities due to pain) (only in subgroup with level 2 examination and/or MRI of joints)</i>		X		
Oral health *				
Short questionnaire on periodontitis and utilization of dental services		X		
Dental health-related QoL (OHIP-5)		X		
Physical activity *				
Work, leisure time, sports		X		
<i>24-h physical activity recall (web-based) (level 2 only)</i>				X

Modules	Assessment Instruments			
	Inter-view	Touch Screen	SAQ (paper)	Web-based
Diet *				
Simplified 24-h diet recalls (web-based; also available as paper version) (1x in study center, 3x at home)				X
Short food list (web-based; also available as paper version)				X
<i>Exact 24-h recalls by telephone (EPIC-SOFT, 3x) (level 2 only)</i>	X			
Environmental factors				
Job and home address (for linkage with geocoded environmental data), Medical radiation exposure		X		
Occupational questions *				
[Work-related stress; listed under “psychosocial factors”] Screening questions for specific conditions, e.g. shift work			X	
Health care utilization and compliance				
Health insurance (type, insurance member’s number) Social security number	X		X	
Health care utilization: general practitioner visits, which of 16 medical subspecialties (Fachärzte) visited how often during past 3 months, ambulatory and inpatient treatment past 12 months		X		
Adherence to medication (Morisky Score)		X		
Other modules for specific subgroups				
Questions only for cancer survivors: various modules			X	

*Further details will be clarified in feasibility studies (see Sect. **A.3.3.11**)

Neurologic and psychiatric factors

The lifetime prevalence of psychiatric disorders such as major depression, anxiety disorders, or addiction is over 45% and rising⁴⁷². The disorders usually start early in life, i.e., in adolescence or early adulthood, and very often are chronically progressive and recurrent. The individual and economic burdens caused by psychiatric disorders are very high: four psychiatric diseases are found among the 10 leading causes of disability worldwide, major depression being the number one for burden of disease in high- and middle-income countries even before CVD⁵⁵. About 10–15% of patients commit suicide during the course of these diseases, and even without clinically relevant symptoms mental disorders may enhance overall mortality⁴⁷³. Psychiatric disorders show a high rate of comorbidity with somatic diseases.

The proposed instruments include scales for assessing depressive symptoms, anxiety, personality, and chronic stress.

Depression and anxiety

The **Mini International Neuro-Psychiatric Interview (M.I.N.I.)** is a brief structured interview for the major Axis I psychiatric disorders in DSM-IV and ICD-10⁴⁷⁴. Validation and reliability studies have compared the M.I.N.I. to the SCID-P for DSM-III-R and the CIDI (a structured interview developed by the World Health Organization for lay interviewers for ICD-10). The results of these studies show that the M.I.N.I. has acceptably high validation and reliability scores, but can be administered in a much shorter period of time than the above referenced instruments. The validated German M.I.N.I. version for the assessment of depression and anxiety will be used in the interview part of the National Cohort. Only modules A, B, E, and P (major depression, dysthymia, panic attacks, and generalized anxiety) of MINI will be applied. The modules consist of 1–3 screening questions for each condition, which are extended by more detailed questions if answer patterns are positive. In addition, scales for the self-report of depressive and anxiety symptoms will be used (see below) to add a severity rating to the disease classification.

The **Center for Epidemiologic Studies Depression Scale (CES-D)** is a self-report, 20-item scale designed to measure depressive symptoms experienced during the previous week. The CES-D, with possible summary scores ranging from 0 to 60, has been shown to be a valid and reliable instrument in the general population for assessing clinically relevant depression⁴⁷⁵. To identify respondents with a relevant level of depressive symptoms, a cut-off score of 16 is commonly used. The scale has a stable four-factor structure, interpreted as the following dimensions: (1) depressive affect, (2) positive affect, (3) somatic complaints, and (4) interpersonal difficulties. A validated German-language version of CES-D is available.

Anxiety Sensitivity Index (ASI-3): Anxiety sensitivity is a psychological construct that is relevant for the etiology of anxiety disorders¹⁰⁴. Anxiety sensitivity refers to the fear of symptoms of anxiety that is based on the concern that these symptoms of anxiety might be harmful^{476, 477}. Individuals with marked anxiety sensitivity are concerned about possible somatic, social, or cognitive consequences of anxiety symptoms. A fast heartbeat, for example, is interpreted by these persons as sign of an impending MI, sweating as a precursor of public disgrace. According to the “Expectancy Model,” anxiety sensitivity can intensify fear, anxiety, and panic^{476, 477}. Empirical findings are to a great deal consistent with the assumptions of the expectancy model. In studies, anxiety disorder patients showed a higher degree of anxiety sensitivity when compared to healthy controls^{478, 479}. Recent findings suggest that anxiety sensitivity not only plays a role in the development and maintenance of anxiety disorders^{480, 481}, but that it might also be a general risk factor for diverse psychiatric disorders, such as alcohol abuse and others^{482, 483}.

The German version of the ASI-3 contains 18 items that are assessed on a 5-point Likert scale. It consists of the three subscales physical concerns, cognitive concerns, and social concerns that are each formed by six items¹⁰⁴. Its structure is thus identical to that of the original English version⁴⁸⁴. Apart from the analysis on subscale level, it is also possible to calculate a total score of all items.

RLS and headache type

For the assessment of RLS and headache type, validated German versions of symptom questionnaires will be used. They can be used for disease classification and, in the case of the headache module, to further distinguish major headache types.

The RLS module is short 5-item questionnaire that was validated against physician classification⁴⁸⁵ and used in identical form in three other German population-based studies be-

fore. It includes the minimal criteria published by the International Restless Legs Syndrome Study Group in 1995¹¹⁴ and their slight adaptations from 2003⁴⁸⁶.

The standardized headache question module assesses the new classification (i.e., 2nd edition) of the International Headache Society⁴⁸⁷. It particularly differentiates between the newly introduced term “probable” migraine (IHS code 1.6), “chronic” migraine (IHS code 1.5), and “probable” TTH (IHS code 2.4). It includes 21 questions on headache characteristics and symptom frequency. The headache module has previously been used in the Dortmund, KORA, and SHIP studies in Germany⁴⁸⁸.

Sleep disturbance

The Pittsburgh Sleep Quality Index (PSQI)⁴⁸⁹ is the most widely used questionnaire to assess sleep quality and sleep-related problems if a sleep laboratory examination is not possible, as it is the case in most epidemiologic studies. The PSQI assesses sleep-related problems retrospectively over a period of 4 weeks and consists of 18 items which are summarized in 7 components. The PSQI allows answers to be self-reported, e.g., on a touch screen or using a paper-and-pencil questionnaire. The German version⁴⁹⁰ of the PSQI will be the basis for the assessment; however, results from the recently initiated feasibility study on the assessment of sleep characteristics will produce adaptations and additions to the PSQI to better serve the multiple purposes of assessing sleep-related problems in the National Cohort. In particular, parts of the D-MEQ questionnaire to measure the chronotype⁴⁹¹, the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome⁴⁹², and the sleep-related (especially siesta) questionnaire of the Heinz Nixdorf Recall Study⁴⁹³ could serve as important extensions or partial substitutions of the PSQI.

Psychosocial factors

A wide variety of adverse psychosocial factors, such as personality, chronic stress, social isolation, lack of social support, and adverse life events, have been shown to be associated with an increased risk for several chronic diseases (see **Sect. A.2.4.4**).

Personality: Big Five Inventory (BFI-10): Currently, the so-called “Big Five” constitutes the most popular approach to studying human personality. The development of the five-factor model started in the 1940s and was based on factor analyses of large numbers of self- and peer reports on personality-relevant adjectives and questionnaire items. It is important to note that the five factors are conceived of as dimensions with two endpoints each (not types), i.e., individuals vary continuously on them.

The dimension extroversion incorporates assertiveness, an open expression of impulses, dominance, and a sense of sociability. Extroversion may also reflect a quality of happiness²⁹⁰. Persons who are highly extroverted can be thus thought of as sociable, active, talkative, optimistic, and hilarious. Introverted individuals, on the other hand, are more likely to behave rather retentively in social interactions and prefer to remain independent. Central aspects of the dimension agreeableness are warmth, friendliness, emotional supportiveness, and nurturance. Agreeable individuals tend to display cooperative behaviors and trust in others, while individuals low in agreeableness experience more conflicts and choose to resolve conflicts more often by displays of power²⁹⁰. Conscientiousness comprises planning, persistence, and purposeful striving towards goals. Conscientiousness has, for example, been linked to educational achievements²⁹⁰. The construct Neuroticism refers to individual differences with regard to the experience of negative emotions. Individuals high in neuroticism often experience anxiety, insecurity, excitement, or tension. Openness refers

to the extent of interest or engagement in new experiences. Highly open persons present as eager for knowledge, imaginative, and willing to try new behaviors.

By means of the recently developed and comprehensively validated BFI-10⁴⁹⁴, which goes back to the longer version BF-44⁴⁹⁵, human personality in terms of the five-factor model can be measured in an extremely economic way while maintaining reliability and validity of the measurements at the same time. In the BFI-10, each of the five factors is covered by only two items.

The **Trier Inventory for the Assessment of Chronic Stress (TICS)**⁴⁹⁶ captures different facets of chronic stress experience in selected life domains. It is based on an interaction-related stress concept, i.e., stress evolves when there is an insufficient fit between the demands a person has to master in everyday life and the resources this person has at his/her disposal. Depending on the type of demand, different sources of stress arise. The different kinds of chronic stress that have been described include job overload, social overload, pressure to succeed, work discontent, excessive demands at work, lack of social recognition, social tensions, social isolation, and chronic worrying. Typically, several of these kinds of stress occur at the same time and are accompanied by states of negative affectivity. Generally, chronic stress is characterized by the following features: it has a subtle, unspecified begin, can be of long or short duration, is characterized by frequently recurring strains, and can be of either high or low intensity⁴⁹⁶.

These kinds of chronic stress as captured by the TICS scales have been linked to sleep quality as well as to indicators of physical and mental performance. In particular, the observed associations with mental performance were fairly high^{497, 498}. Several psychobiological studies support systematic relationships between TICS scales and the cortisol response to awakening^{496, 499}. In addition, different professional groups seem to differ in how they experience chronic stress⁴⁹⁸.

The original version of TICS⁴⁹⁶ comprises 57 items, which are to be assessed on 5-point Likert scales covering all nine kinds of stress named above. Twelve items of the TICS constitute the Screening Subscale of Chronic Stress (TICS-SSCS), which provides an un-specific global measure of chronic stress. TICS screening items refer to aspects such as being afraid that something unpleasant may happen, not being able to suppress worrying thoughts, or not being able to cope with all duties. Higher values indicate greater stress⁵⁰⁰.

Traumatic childhood experiences: Research over the past decade has suggested that traumatic experiences during childhood are a key factor in predicting negative health outcomes, especially for psychiatric diseases, but also in several somatic disorders and substance abuse⁵⁰¹. Every year 4–16% of all children in high income countries are physically abused, 10% experience psychological abuse or neglect, and 5–10% are exposed to penetrative sexual abuse. Childhood maltreatment has been implicated as an important risk factor for major depression and somatization disorders but also obesity, CHD, and various autoimmune diseases. In addition aversive environmental factors in early life can induce long-lasting pathophysiological alterations such as an increased load of inflammatory factors and alterations in the neuroendocrine system which are subject to gene–environment interactions⁵⁰² and DNA methylation⁵⁰³. The assessment of childhood maltreatment offers therefore the opportunity to better define biological models of disease onset within the framework of the National Cohort and will improve our understanding of the pathophysiology of common disorders assessed in this study.

Instruments:

Only few questionnaires exist for retrospective assessment of traumatic experiences during childhood and adolescence. The "Childhood Trauma Questionnaire" (CTQ) is one of the most widely used scales. It includes 28 items covering several domains of early abuse and neglect on a 5-point Likert scale. The German version of the CTQ has been recently translated and tested and a short version developed^{504, 505}. This short version consists of 5 questions and has been implemented in the population-based SHIP study. The questions can be answered in self-report questionnaires and the instrument provides an economic assessment of traumatic childhood experiences.

Psychosocial work stress: The ERI model was developed in Germany^{298, 506} and shown to be predictive of adverse health outcomes. It is well established in international epidemiologic cohort studies^{301, 507}. Its theoretical assumptions focus on social reciprocity and highlight distress arising from inadequately low rewards in spite of high external efforts. Furthermore, the ERI model accounts for internal efforts (overcommitment) as an individual risk factor for getting or staying in work situations characterized by ERI or as a health-related risk factor on its own. The ERI questionnaire has been validated for assessing the dimensions efforts, rewards, and overcommitment while rewards were measured within the subdimensions esteem, job security, and job promotion. In the National Cohort the newly developed short version^{508, 509} will be used which provides an economic measure for inclusion in large-scale epidemiologic investigations. The short version assesses the same dimensions as the original version but with only 16 items instead of 22. In statistical analyses a sum-score for each dimension and a ratio can be calculated while a ratio >1 reflects a high ERI.

Further instruments that are available to assess job strain and psychosocial work environments and that have successfully been used in German populations include the *Copenhagen Psychosocial Questionnaire (COPSOQ)*⁵¹⁰⁻⁵¹² and the *BIBB/IBA questionnaire*⁵¹³⁻⁵¹⁵. The latter assesses the psychosocial effects that may be caused by reorganizing the workplace.

Social networks and social support will be assessed using a standardized instrument developed by Berkman and Syme^{309, 310, 516, 517} that has been modified and validated for use in German studies (e.g., the Heinz Nixdorf Recall study³¹¹). Network measures include availability of a confidant, partnership, close ties, social participation, and a summary index of social integration. Support measures include instrumental and emotional support. Questions first assess the availability of someone to help in daily tasks and the presence of one or more persons to approach when problems arise. In a second step the questionnaire inquires as to who actually provided support and whether that support is appropriate.

Infectious diseases

The infectious diseases to be studied in the National Cohort include respiratory infections, gastrointestinal infections, and periodontal diseases. As outlined in **Sect. A.2.2.6**, those infectious diseases present a high burden of disease in the population, but the determinants of susceptibility, course, and severity of these diseases are not fully understood. Therefore, information on these types of infectious diseases will be collected at baseline and during follow-up of the National Cohort. In this way, the burden of disease in Germany can be precisely defined, a respiratory infection risk score can be developed, new vaccine strategies can be established, and we can improve our understanding of the associations of periodontal disease with other chronic diseases such as CVD. Furthermore, information on vaccination gaps and their determinants in the population will be collected to identify barriers

ers to vaccination for vaccine-preventable diseases. It will also be possible to analyze the association of vaccinations with development of future chronic diseases.

Instruments:

The specific questionnaire module on infectious diseases will cover in detail respiratory disease, gastrointestinal infections, childhood infectious diseases, past and recent vaccinations, and contact with pets. The participant will be asked to bring his/her vaccination booklet. The questionnaire will be tested in a feasibility study.

Musculoskeletal assessment

Musculoskeletal conditions are frequently found in the general population^{222, 518}. The related pain assessment and questionnaire items target osteoarthritis, back pain, chronic widespread pain, and rheumatoid arthritis. This assessment will complement findings from MRI, a clinical examination, and biomaterials to provide a unique depth of population-based information on phenotypes and genotypes related to musculoskeletal diseases.

Instruments:

An electronic pain mannequin (touch screen) will serve to record pain prevalence at different anatomic sites of the body during the past 7 days. This time interval was chosen to improve the study of associations with clinical findings. Pain mannequins are widely used in clinical practice and research^{519, 520}. The mannequin will be based on an anterior and posterior drawing of the body as displayed in the German Pain Questionnaire⁵²¹. The final distinction of body parts on the mannequin will be guided by previous research⁵²⁰ and an ongoing methods assessment based on the population-based SHIP⁵²². When affirming pain in selected body regions, subsequent questions will address pain intensity, related disability on a 0–10 numerical rating scale, as implemented in the Graded chronic pain questionnaire, and the duration of the pain problems to identify chronic conditions^{523, 524}. Target regions for these additional questions, e.g., joints or back, may be guided by the MRI results. Five selected items of the Rheumatoid Arthritis Disease Activity Index (RADAI) will serve to assess symptoms specific for rheumatoid arthritis, e.g., morning stiffness in the joints⁵²⁵.

Oral health and periodontitis

The presence and treatment need of caries and periodontitis are strongly associated with SES^{526, 527}. In addition, quality of life is strongly related to the dental and prosthetic status⁵²⁸, and can be assessed by means of specific instruments. The short validated Oral Health Impact Profile questionnaire (OHIP-5) covers questions about the basic oral health state. In the US population-based cross-sectional study NHANES 2003-2004, the OHIP-14 was used to quantify the impact of oral disease on quality of life. Results of this study were compared with OHIP scores of an Australian cohort⁵²⁹. The OHIP instrument has been translated into several languages all over the world and is also available in a well-tested German version^{530, 531}.

Instruments:

The validated OHIP-5 questionnaire about the basic oral health situation will be administered through touch screen. Four additional questions will cover utilization of dental services and on diagnosis or treatment of periodontitis.

Physical activity

Physical activity is a complex behavior that can be subdivided into type of activity, frequency, duration, intensity, and setting. The true effect of physical activity on health outcomes is likely to be underestimated due to substantial measurement error in the existing assessments of physical activity, and the exact contribution of the different components of physical activity to chronic disease is unclear⁵³². Most previously conducted studies have only used questionnaires to assess physical activity, most of them aiming to estimate energy expenditure. However, there is wide agreement that the validity of existing physical activity questionnaires is limited for estimating energy expenditure and quantifying physical activity^{238, 240, 533}. Therefore, the main instrument for measuring physical activity will rely on accelerometry, which is an objective measure, while supplementary information on activity patterns will be obtained by questionnaire. The combination of monitoring physical activity via web-based physical activity recalls and simultaneous measurement of physical activity by an objective device (triaxial accelerometer) is needed to improve the precision of physical activity data.

Instruments:

The physical activity questionnaire for the National Cohort will address those aspects of physical activity that are not covered by objective measurement devices such as accelerometers. Information on physical activity will be obtained for the past 12 months. Thus, seasonal variation in physical activity can be analyzed, which would not be possible with the short-term accelerometric physical activity assessment. The aim of the questionnaire is to assess the activity behavior and structural and temporal patterns (whereas energy expenditure will be estimated by an objective device). Therefore, based on experience from existing, validated questionnaires, a new-generation questionnaire has been developed and is currently being tested in a substudy in EPIC-Germany.

In addition, a computer-based physical activity recall (CPAR-24) will be used to obtain more detailed information on all physical activities undertaken during the previous 24 hours. This information is necessary to derive algorithms for interpreting the accelerometer data.

Diet

Diet has been hypothesized to play an important role in the development of major chronic conditions, such as CVD, diabetes, cancer, and osteoporosis. Such hypotheses are based on an abundance of plausible biological mechanisms relating diet and nutrition to disease occurrence. To date, however, nutritional epidemiologic studies have not generated thoroughly consistent evidence regarding the role of many dietary factors in disease etiology. Thus, several hypotheses on associations between dietary factors and disease occurrence remain to be clarified²³⁹.

One possible explanation for the inconsistency in previous nutritional epidemiologic studies of chronic diseases is the methodological challenge in validly assessing habitual diet in free-living individuals. There is particular concern that dietary assessment instruments commonly used in large-scale studies (food frequency questionnaire) are prone to substantial measurement error, which might have resulted in considerable underestimation of true associations⁵³⁴. Therefore, improved instruments of data collection will be used in the National Cohort.

Instruments:

Recent advances have rendered the use of multiple, internet-based 24-h dietary recalls possible, implemented in combination with a food list (for identification on nonconsumers of

certain foods) and the measurement of biomarkers of intake. This combination will allow for a more valid dietary assessment without compromising the study participants by extensive dietary records. Thus, the chosen approach is judged both feasible and cost-effective. For participants without internet access, a paper or a touch screen version will be available. For calibration issues and portion size definition, exact 24-h diet recalls will be collected in a subgroup of the cohort, using a validated computer-assisted telephone interview (EPIC-SOFT).

Environmental exposures

The evaluation of environmental risk factors such as air pollution with particulate matter or noise pollution is important because they are widespread in the population and may have a large public health impact in spite of small RRs since highly prevalent exposures may lead to large attributable risks.

Outdoor exposure to air pollution: Associations between particulate matter and mortality have been reported⁴¹⁷, which resulted in tightening of the particulate matter standard in the European Union in 2005. In spite of the improvement in standards, particulate matter and gaseous pollution from traffic-related sources remains a matter of continuous concern⁴²², and the impact of traffic-related exposures is an important part of the assessment of environmental exposures. Noise exposure and exposure to high or low temperature are further outdoor factors to be analyzed as potential risk factors for chronic diseases.

Instruments:

Recording of the home and work address (together with the participants consent to retrieve the residential history from official registries), and the usual commuting pattern (mode of transport used and duration) will be collected by questionnaire in the entire cohort. These will form the basis for linkage with geographically coded exposure information for air pollution, noise, and temperature. Spatially and temporally resolved maps for PM10, PM2.5, SO2, NO2, O3, ultrafine particles, and elemental and organic carbon will be developed independently to assign exposure status to the geocoded residential and workplace addresses. Maps spatially resolved for noise exposure (considering differences between day and night time and in cooperating commuting times and work place exposures) as well as maps spatially resolved for temperature exposure (considering differences between seasons) will be developed and can be applied for the benefit of the National Cohort.

It is important to note that based on geographic information system (GIS) tools, other emerging exposures can be added during the course of the study. Thus, a flexible tool is being established. We propose to establish the GIS databases at different physical locations and separately from the health data. Only the derived exposure estimates but not the geocodes for addresses would be merged to the health data. Thus, confidentiality issues are taken into consideration while the novel and powerful GIS tool can be exploited as part of the National Cohort.

Ionizing radiation: Man-made exposure to ionizing radiation is dominated by exposure to medical radiation, predominantly for diagnostic purposes⁵³⁵. Medical radiation exposure is emerging as an important public health issue, but there is a lack of information on the distribution of exposure in the population, particularly in relation to age, sex, and diagnoses. Moreover, it is important to analyze the contribution of medical radiation exposure to chronic disease occurrence, e.g., to assess an association with cancer and well-defined noncancer endpoints, and to prospectively monitor the distribution of medical exposure in Germany based on individually collected data.

Instruments:

Information on medical radiation exposure and related organ doses will be collected with a newly developed questionnaire. The questionnaire is being developed and validated in a pilot study funded by the BfS in 2011/2012. The purpose of the pilot study is to assess the feasibility to validly and reliably reconstruct medical exposures.

Occupational exposure/history

Adverse health effects of common occupational exposures include the occurrence of chronic diseases such as cancer and cardiovascular, musculoskeletal, and neurodegenerative diseases, and it is therefore important to monitor both occupational exposures and disease occurrence in the general population and in particular in the aging workforce. Occupational exposures comprise a wide range of work-related circumstances. *Shift work* has increased in the Western industrialized world in the last few decades and night work and shift work may be associated with circadian disruption, which in turn may be associated with chronic disease⁵³⁶⁻⁵³⁸. *Noise exposure* is one of the major occupational hazards in many workplaces and has several health effects, including hearing loss and psychological effects⁵³⁹. Exposure to *particulate matter* (PM) is a health problem at many workplaces that is regulated for the inhalable fraction of particles and fine particulate matter reaching the alveolar region of the lung. The role of adverse psychosocial factors at work has been described above.

Instruments:

A detailed occupational history will be retrieved for each study participant from secondary data sources, namely, the IAB Database of the Institute for Employment Research, Federal Employment Agency. The procedure for linking data is currently being tested in a feasibility study. Information on occupations held and industries in which the occupation was performed will be classified according to the International Standard Classification of Occupations (ISCO) and to the International Standard Industrial Classification (ISIC). During the core interview, a few occupational screening questions for defined exposures of interest will be included; this will allow subsequent Level 3 studies (with separate funding) in which supplemental questionnaires will be applied via telephone interviews, mailed questionnaires, or web-based questionnaires at a later time. The job and industry codes as well as the screening questions will supply valuable exposure information in addition to exposure levels in biological samples. Furthermore, work-related psychosocial stress will be assessed using different validated questionnaires (see **Sect. A.2.4.4**, Psychosocial factors).

Health care utilization and compliance

Not all aspects of illness and health, health care, and health care delivery can be assessed validly by means of secondary data sources (health insurance data). In order to determine the efficiency of health care delivery and usage, the perspective of the health care recipient is crucial.

It can be expected that claims data will not be available for all study participants. Therefore, basic variables on health care utilization are assessed in the core protocol. Moreover, consultation times are of increasing interest⁵⁴⁰⁻⁵⁴² since time which has to be spent seeking health care may affect both morbidity and adherence to medical care.

Instruments:

The following items will be included in the core protocol subsequent to the assessment of current medication use (IDOM, **Sect. A.3.2.1**, Medication).

A short instrument, assessing outpatient and hospital care and some supplementary services and loss of productivity as well as a single question regarding consultation time has been adopted from the KORA surveys⁵⁴³. By assessing patient time we can estimate indirect, nonmedical costs.

Medication adherence

Approximately 50% of all patients do not take their medication as prescribed⁵⁴⁴. The degree of nonadherence to medication in patients with risk factors and/or existing diseases has been shown to be negatively associated with disease progression and disease outcomes in various patient samples. Hence, we expect that medication (non-) adherence is associated with relevant outcomes (morbidity, mortality, and quality of life) of diseases investigated in the National Cohort. Objective measures of medication adherence such as pill count or electronic monitoring are not feasible in the National Cohort. However, studies have shown that measures assessing general medication adherence correlate at least moderately with objective measures⁵⁴⁵ and are predictive of disease outcomes⁵⁴⁶. Within the National Cohort, the impact of self-reported medication adherence on morbidity and mortality for various diseases and in risk populations can be systematically studied in both cross-sectional and prospective analyses. Patients with poor adherence can be characterized on the basis of demographic, medical, and social characteristics. We expect better outcomes regarding hospitalizations and death in patients with higher scores of adherence.

Instruments:

Adherence to medication will be assessed using the German version of the 4-item Morisky questionnaire, which is most commonly used in the international context⁵⁴⁷, complemented by a single-item question about the percent medication adherence during the last month. This has been found to be predictive of cardiovascular events in patients with stable CHD⁵⁴⁶.

Other modules for specific subgroups

Other modules that have been suggested for specific subgroups include several questionnaires for cancer survivors that will be applied only in this subgroup subsequent to the basic assessment visit during the follow-up period. Proposed questionnaires that are still under discussion include the tumor-specific EORTC QLQ-C30 quality of life questionnaire, the Fatigue Assessment Questionnaire (FAQ), the Benefit Finding Scale (BFS), the Posttraumatic Growth Inventory (PGI), the Fear of progression (short form PA-F-KF), and the Lubben Social Network Scale (LSNS-6).

A.3.3 Baseline physical measurements

A.3.3.1 Overview

Decisions for including and excluding physical measurements in the National Cohort were made (i) on the basis of the prioritized diseases and exposures/risk factors and (ii) after careful consideration of specific criteria as listed below (**Table 3.6**). Standardized and validated examination modules that are used in other large epidemiologic studies were also taken into consideration.

Selection criteria used to decide which physical measurements are to be implemented included the scientific relevance of the methods, their innovative potential, the validity and reliability of the proposed methods, ethical considerations, and the required financial re-

sources. Aspects of quality assurance (e.g., staff training, ease of use, standardization of measurements), time required for performing the proposed measurements, possibility for computerized capture of data, and expected acceptance of the procedures by the study participants were also considered. Mainly due to limitations in time and budget, some measurements will only be conducted in the intensively examined subcohort. The core physical examination program for the entire cohort will take about 1 ½ h on average (including the collection of biosamples); the program for the intensified subgroup will take about 1 ½ hours more. For selected instruments, practicability and other aspects are being tested in feasibility studies.

Table 3.6: *Criteria for selecting of instruments for physical examinations*

General criteria

- ▶ Scientific relevance
- ▶ Measurement is associated with primary outcome
- ▶ Measurement is reproducible, valid, and suitable for repeated measurements in cohort
- ▶ Acceptability for subject regarding type of examination, duration, effort and potential inconvenience
- ▶ SOPs suitable for use across multiple sites; acceptable training of staff
- ▶ Capture data directly into computer where possible
- ▶ Costs acceptable with high participant throughput and large total number of subjects
- ▶ Limited time required, rapid measurement
- ▶ Widest possible coverage of body systems
- ▶ Potential for feedback of meaningful results to participants

Specific criteria for scientific relevance

- ▶ Additional information/value (as compared to other cohort studies)
- ▶ Innovation potential of the examination methods (will new information be produced? better quantification of exposure, function or disease outcomes)
- ▶ Preventive potential of the assessed exposure
- ▶ Relevance of the health outcome in terms of public health and economic burden for the society
- ▶ Relevance of exposure information for multiple health outcomes
- ▶ High frequency and/or increasing time trends of the exposure or health outcomes in the population
- ▶ Exposure or other factor needed as important confounder in statistical analyses
- ▶ Exposure assessment needed for aetiologically relevant scientific question
- ▶ Complementary measurements: availability of short and long versions or related modules to assess specific exposures or functions (modular aspect for examination of whole and intensified sub-cohort)

Baseline measurements for all subjects (Level 1) will include measurement of blood pressure, recording of a 10-s ECG, 3D-Echocardiography, combined device, cognitive functioning tests, spirometry, tooth count, anthropometric measures (body weight, body height, waist, and hip circumference), DXA scan, and physical activity tests (**Table 3.7**). For the 7-day recording of physical activity, the accelerometer device will be explained to the par-

Participants and fixed. Participants will be instructed to return the device to the study center by mail at the end of the 7-day assessment period.

In the intensively investigated group (Level 2) a more in-depth phenotyping will be conducted. In addition to the Level 1 examination program, study participants belonging to this group will get further examinations related to the cardiovascular system, glucose metabolism (diabetes), respiratory diseases, the musculoskeletal system, oral health, sense organ functions, and physical activity and fitness (**Table 3.7**).

Table 3.7: Baseline physical examinations at level 1 (n=200,000) and level 2 (n=40,000)

Instruments	Level 1 (duration)	Level 2 (duration)
Cardiovascular examinations:		
- Blood pressure and heart rate	7 min	
- Electrocardiography (10-sec. ECG; 12 leads)	8 min	
- 3-D-Echocardiography	6 min	
- Combined device (e.g., Vascular Explorer)*: Arterial stiffness, Ankle Brachial Index, etc.	10 min	
- Carotic sonography, intima-media thickness		10 min
- Long-term ECG / apnea device*		5 min
Diabetes-related measurements:		
- Oral glucose tolerance test (OGTT)		5 min
- AGE reader (skin AF)		4 min
Cognitive function:		
- Test battery*	10 min	10 min
Lung function:		
- Spirometry	8 min	
- Nitric oxide (FeNO) in exhaled air		6 min
Musculo-skeletal system:		
- Medical examination*		10 min
Oral health:		
- Tooth count	1 min	
- Oral examination*		10 min
Sense organs:		
- Ophthalmological measurements		10 min
- Hearing test (touch screen, headphone) †		4 min
- Olfactory test (sniffin'-sticks 12)		5 min

Instruments	Level 1 (duration)	Level 2 (duration)
Anthropometry:		
- Body weight and body height (standing)	3 min	
- Waist and hip circumference	2 min	
- DXA scan*	10 min	
Physical activity and physical fitness:		
- 7-d accelerometry (instructions for use)*	5 min	
- Step test*		10 min
- Hand grip strength (3 times)	3 min	
- One leg stand / posturometry*		5 min
Collection of bio-samples	15 min	
Total examination time	88 min	94 min

* Subject to feasibility study; see Sect. **A.3.3.11**; † self-administered test

A summary description of the scientific background of the selected instruments along with a short description of the procedure is provided in the following sections.

For the group of cancer and infectious diseases, no specific physical examinations are proposed. However, several of the proposed examinations are of specific interest for these disease groups, too. For malignant neoplasms, important measures include body composition, physical activity, and skin AF, and for infectious diseases physical activity and dental examination.

A.3.3.2 Cardiovascular examinations

Blood pressure and heart rate

Elevated blood pressure is a cause of stroke and CHD⁵⁴⁸ as well as of vascular and overall mortality⁵⁴⁹. Blood pressure (both hypertension and hypotension) also appears to play an important role with regard to cognitive functions and dementia⁵⁵⁰⁻⁵⁵⁴ and is strongly associated with type 2 diabetes mellitus and obesity⁵⁵⁵⁻⁵⁵⁸.

Accelerated heart rate is considered as an independent risk factor and prognostic marker for morbidity and mortality. There is evidence for a clinically relevant association between elevated heart rate and adverse CVD risk factors, cardiovascular events, and mortality among adults⁵⁵⁹⁻⁵⁶². This association is generally stronger in men than in women⁵⁵⁹⁻⁵⁶¹. It has been suggested that the central nervous system plays a key role via a sympathetic overactivation for both the increase in heart rate and for the development of hypertension^{560, 562}.

Procedure of measurement:

Systolic and diastolic blood pressure will be measured using a validated oscillometric device that has been used in other epidemiologic studies (e.g., OMRON HEM 705 CP II). Measurement will be taken according to a standardized protocol as the results will be influenced, for example, by measuring technique of the investigator, posture of the participant, and position of the device. Three blood pressure measurements will be taken in the sitting position after

a resting time of at least 5 min, with a waiting time of 3 and 1 min, respectively, before the second and third measurement. Heart rate will be manually measured after the first blood pressure measurement using a stop watch. The measurement procedure is simple and takes about 7 min.

Electrocardiography

A 12-lead resting ECG (10 s) reflects the electrical activity of the heart and is therefore one of the most frequently used diagnostic tools for cardiologic assessment. It is also one of the most efficient tools for diagnosing conduction disturbances and arrhythmias (such as atrial fibrillation), cardiac ischemia, acute coronary syndromes, and MI and left ventricular hypertrophy. It will be used to confirm prevalent and to detect incident cardiac events and disturbances.

Procedure of measurement:

ECGs are recorded under highly standardized conditions (e.g., using the DAL square)⁵⁶³ which have been applied in numerous studies (e.g., NHANES III, KORA, CARLA, Heinz Nixdorf Recall). A 12-lead ECG is a noninvasive, routine, standard procedure. Ten electrodes (six on the chest and four on the limbs) will be placed on the subject according to a standard procedure⁵⁶³. The subject will then be asked to lie down still during the recording. All ECGs will be digitally acquired.

3D-Echocardiography for the assessment of left ventricular function and diagnosis of heart failure

As outlined in **Sect. A.2**, the prevalence and incidence of heart failure are increasing and pose a considerable burden on aging societies. Impaired left ventricular systolic function and isolated diastolic dysfunction (heart failure with preserved ejection fraction) are both associated with morbidity, hospitalization, and mortality^{43, 45, 46, 564}. In addition to documenting symptoms of heart failure, cardiac imaging represents an indispensable tool for diagnosing heart failure. Echocardiography is the imaging technique of choice for comprehensive assessment of diastolic function⁴², and recently published guidelines of the European Society of Cardiology for the diagnosis of acute and chronic heart failure clearly recommend the use of echocardiography as imaging technique^{43, 48}.

Procedure of measurement:

In the National Cohort, we will assess systolic and diastolic function and determine other cardiac parameters using a 3D-Echocardiography device with a matrix probe to acquire multibeam or single-beat, full-volume data sets wide enough to encompass all four chambers and further data sets with color Doppler of the mitral valve (optionally including the aortic valve). These will be used to derive left ventricular volumes (endsystolic and end-diastolic), stroke volumes, and ejection fraction, left ventricular mass, and left atrial endsystolic volumes and also right ventricular and atrial volumes and ejection fraction. Mitral valve morphology and flow patterns can be determined as well. Moreover, tissue Doppler echocardiography will be used to determine parameters of diastolic function [e.g., early diastolic tissue Doppler velocity (e'), early mitral inflow velocity (E), atrial inflow velocity (A), E/e' , E/A , and deceleration time (DT)] based on tissue Doppler signal acquisition at the lateral mitral annulus and pulsed-wave Doppler signal acquisition of mitral inflow. Echocardiographic data will be acquired according to published recommendations^{565, 566}. Total examination time will comprise 3–5 min, optionally with additional offline reading time of 3 min to quantify left atrial and ventricular volume, e' , E , A , and DT .

The device will also be equipped with an ultrasound probe suitable for performing the measurement of intima–media thickness of the carotid artery.

Measurement of arterial stiffness and ankle–brachial index

Arterial stiffness is defined as a reduction in arterial distensibility. The physiological mechanisms underlying arterial stiffening occurring with age and the accumulation of cardiovascular risk factors include breaks in elastin fibers, accumulation of collagen, fibrosis, inflammation, medial smooth muscle necrosis, calcifications, and diffusion of macromolecules within the arterial wall⁵⁶⁷. These phenomena have been shown to develop simultaneously in coronary arteries and the aorta⁵⁶⁸. The association of specific measures of arterial stiffness (such as the augmentation index and central blood pressure) with target-organ damage and cardiovascular outcome is greater than that of brachial blood pressure⁵⁶⁹, i.e., the increase in cardiovascular risk can be estimated better from central than from brachial blood pressure measurements⁵⁷⁰. Recent guidelines for the management of hypertension call for assessment of arterial stiffness to guide therapeutic decisions⁵⁷¹.

Moreover, alterations in vascular stiffness may cause diastolic and systolic heart failure, for example, as a result of left ventricular hypertrophy, and subendocardial ischemia or coronary artery disease. However, the potential link between measures of vascular stiffness and ventricular dysfunction (“arterioventriculo coupling”) has not been addressed in large population-based cohort studies. Furthermore, there is insufficient knowledge on determinants of arterial stiffness.

The gold standard for measuring arterial stiffness is pulse wave velocity⁵⁷², as measured by the Sphygmocor device⁵⁷³, which has also been used most frequently. However, newer devices which are easier to use and less observer-dependent have recently become available (Vascular Explorer, Vicorder)⁵⁷⁴. In practice, pulse wave velocity is calculated as the traveling time or distance of the pulse wave between two measuring sites⁵⁷⁵. The assessment of the wave reflection in the arterial system, the pulse wave analysis, gives information on arterial stiffness, central blood pressure, and subendocardial flow reserve.

The *ankle–brachial index* is a noninvasive, rapid, and quantitative measure of PAD, derived from brachial and ankle systolic blood pressure values. The ankle–brachial index is inversely associated with CVD events, mortality, and the progression to critical leg ischemia, rest pain, and ulceration^{576–579}. It is one of the standard tools for diagnosing lower extremity PAD and is used for PAD screening and therapeutic efficacy monitoring. Lower extremity PAD can be symptomatic (intermittent claudication) but is often asymptomatic. The use of an ankle–brachial index < 0.9 to diagnose PAD has a sensitivity of 0.95 and specificity of 0.99 and is highly reproducible⁵⁸⁰. However, further information is needed to improve our understanding of early developments and determinants of progressive reduction in the ankle–brachial index, and to better assess the predictive ability of reduced ankle–brachial index for future cardiovascular events.

Procedure of measurement:

Ankle–brachial index, pulse wave analysis, and velocity will be measured at a controlled room temperature (22°C) after the subject has rested for at least 5–10 min in a supine position.

For these measurements, several devices are available, among them the SphygmoCor device (the current gold standard for measuring pulse wave analysis and velocity)⁵⁷², the Vascular Explorer, and the Vicorder (the latter two being new multifunctional devices for measuring ankle–brachial index and pulse wave velocity). To determine the ankle–brachial

index, conventional methods using oscillometric blood pressure measurement devices and hand-held Doppler devices to detect systolic blood pressure at the brachial artery, the dorsal artery of foot, and posterior tibial artery are available as well.

For the National Cohort, one of the multifunctional devices will be used. The potential of the Vascular Explorer or the Vicorder device to replace the more observer-dependent and time-consuming use of the Sphygmocor device in the National Cohort will be evaluated in a feasibility study.

Carotid sonography: intima–media thickness

Atherosclerotic disease begins in early life, progresses over decades⁵⁸¹, and may result in rupture or erosion of existing plaques, leading to MI or stroke⁵⁸². The degree of sub-clinical atherosclerosis as defined by CIMT or presence of carotid plaques is a predictor of incident CVD events in the absence of overt CVD⁵⁸³⁻⁵⁸⁸. Vascular imaging such as CIMT measurements with B-mode ultrasound is a noninvasive, sensitive, and reproducible technique for identifying and quantifying atherosclerotic burden and CVD risk and can be used as a screening tool that improves risk prediction beyond application of major risk factors alone⁵⁸⁹⁻⁵⁹¹.

By including carotid ultrasound measures as a surrogate of subclinical atherosclerosis in the examination protocol of the National Cohort we can compare our results with those of other large-scale epidemiologic studies.

Procedure of measurement:

High-resolution B-mode ultrasound systems with linear ultrasound transducers (frequencies ≥ 7 MHz) represent the standard equipment for measuring CIMT and plaques. For the National Cohort, a combined ultrasound system will be used for echocardiography and carotid artery imaging. Examinations will be performed by trained and certified investigators and will be done at both carotid arteries. External landmarks (e.g., the Meijer arc or similar devices) will be used to standardize transducer angle. CIMT is measured preferably on the far wall from multiple sequences of longitudinal scans of the common carotid artery. The reading will be done offline using automated reading software. Carotid plaques will be evaluated at the near and far walls of common carotid artery, bulb, and internal carotid artery segments. Scanning and measurement procedures as well as definitions of CIMT and carotid plaques will be in accordance with current guidelines^{592, 593}.

Long-term ECG recording and sleep-related characteristics

Atrial fibrillation (including silent, undiagnosed atrial fibrillation) is the most common cardiac arrhythmia. It is associated with a doubling of the death rate and with substantial increases in severe cardiovascular complications such as stroke, acute coronary syndrome, and heart failure.

Many clinically relevant ECG changes such as atrial fibrillation are transient, and the search for such changes can be lengthy and cumbersome, e.g., evaluating syncope or palpitations. Conventional Holter ECG recordings often miss transient ECG changes (e.g., intermittent atrial fibrillation, paroxysmal tachycardias, or intermittent AV block). In the National Cohort, we aim to record at least a 24-h holter ECG, if possible taking a 7-day recording. Simple, easy-to-carry-and-apply automatic and/or patient-activated ECG recording systems make it possible to record an ECG over an extended period of time (e.g., 7 days) in epidemiologic studies.

Several new devices simultaneously permit the recording of other physiological parameters, such as respiratory characteristics or triaxial accelerometric data.

The selection of the specific ECG device will be based on the results of the feasibility study that will be completed prior to beginning the pilot phase for the National Cohort. Different combinations of devices or examinations will be evaluated in feasibility studies concentrating on (1) ECG detection, (2) sleeping patterns /apnea, and (3) recording of physical activity via triaxial accelerometry.

Procedure of measurement:

We aim to use innovative mobile ECG monitoring devices to assess silent atrial fibrillation and other arrhythmias. The most suitable of several available devices will be selected according to the results of an ongoing feasibility study. Ideally, one device would be able to simultaneously assess cardiac rhythm, respiratory characteristics, and physical activity with sufficient quality for use in the National Cohort. The recording device will be worn for 7 days.

The Corventis Nuvant™ adhesive ECG recorder is an automated, carrier-independent, one-way recorder that is carried parasternally. Patients are monitored by this unobtrusive, leadless, and water-resistant device using automatic triggers that capture asymptomatic arrhythmias, but they can also be manually activated. At the same time, triaxial accelerometric data and respiratory rate can be recorded.

The Omron HeartScan™ HCG-801-E ECG recorder is a cordless, proband-operated, single-lead ECG recorder that records periods of 30 s at predefined time intervals and during symptomatic episodes when activated by the proband. Stored ECG episodes will be analyzed offline.

The Shimmer device with addition of the ECG board daughter card sensing system (Shimmer Research) is a triaxial acceleration sensor (accelerometer) combined with a continuous-recording, 3-lead ECG device.

Further multifunctional devices with which ECGs, 3-axial accelerometric data, and sleep-related respiratory characteristics (e.g., hypopnea, apnea) can be simultaneously recorded will be evaluated during feasibility studies.

The long-term ECG devices will be centrally ordered and technically prepared and then distributed among the participating centers. For purposes of data collection and quality control, a competence center will be established to store the sensor raw data obtained by each device over the assessment period.

The study participants will be instructed at the study center on the use of the respective device, and the devices will be returned by mail after the end of the recording time.

A.3.3.3 Diabetes-related measurements

Oral glucose tolerance test

Detecting early changes in glucose metabolism and discovering undiagnosed cases of diabetes is a primary prerequisite for embarking on innovative diabetes research. About 40–50% of all undiagnosed diabetes cases are detected by the 2-h OGTT alone, which would have been missed based on fasting glucose levels^{57, 594}. Although fasting indices have been frequently used as surrogate estimates of insulin sensitivity⁵⁹⁵, they have not been useful in estimating insulin secretion⁵⁹⁶. We propose applying the insulinogenic index, calculated as the ratio of the 30-min increment in insulin to glucose concentration $(\text{Ins } 30 - \text{Ins } 0) / (\text{Glu } 30 - \text{Glu } 0)$ ⁵⁹⁷, which is widely used to estimate insulin secretion and to predict risk of diabetes^{596, 598}.

Although the intraindividual variation in response to glucose administration is substantial⁵⁹⁹, its value in epidemiologic studies has been demonstrated in several large-scale studies. Major factors that may affect the test results can be largely excluded by following a strictly standardized procedure. The standardized test conditions require a careful arrangement of additional tests to be performed by the study participant. Due to insufficient evidence of possible influences on other parameters, a feasibility study will be set up to explore this further and to define measures that allow unbiased analyses of all aspects of the cohort study.

An oral glucose load of 75 g (in 300 ml water) is administered to fasted subjects, and blood sampling will be performed before and 30 min and 120 min after glucose load.

AGE reader

Recently, skin AF has emerged as a noninvasive and reproducible tool to estimate the AGE level of skin tissue^{600, 601}. Skin AF, as measured by AGE reader is associated with plasma levels of advanced glycation endproducts and has been validated against skin levels of several specific AGEs and a classic assay for AGEs – collagen-linked fluorescence – in different populations (diabetics, subjects with renal failure, and controls)⁶⁰⁰⁻⁶⁰².

Skin AF has a higher sensitivity than fasting glucose and HbA1c to detect impaired glucose tolerance and diabetes and may be useful as a screening tool⁶⁰³. Lutgers et al. reported an overall Altman error of 5% for repeated skin AF measurements in the same persons (diabetic patients and controls) taken over a single day, and an error of 5.9% for the intraindividual seasonal variance of skin AF among control participants and diabetic patients⁶⁰⁴.

Skin AF measurements by the AGE reader (DiagnOptics) are noninvasive, reliable, and reproducible (good repeatability, low intraindividual variation, low measurement error at repeated measurements) and are not affected by current glucose levels and thus not dependent on fasting state during measurement⁶⁰³.

Procedure of measurement:

The tissue accumulation of AGEs will be determined with triple measurements by means of fluorescent techniques, using the DiagnOptics AGE Reader. The total procedure takes 4 min and is noninvasive, easy to perform, and requires only a minimum of training.

A.3.3.4 Cognitive functioning tests

It is generally acknowledged that cognitive assessment must reflect several different cognitive domains (e.g., memory, attention, reasoning, speed of processing, and executive functions) which may be specifically affected in certain diseases (e.g., Alzheimer's disease, stroke, Parkinson's disease) and aging.

Cognitive ability and its decline in older age seems to be related to prior cognitive ability in early or mid-life and to educational level, as subjects with higher education or social class seem to be able to compensate longer for the age-related brain changes^{605, 606}. Decline in cognitive function and dementia may also be associated with general vascular risk factors, such as physical inactivity^{607, 608}, smoking^{609, 610}, or diabetes⁶¹¹.

Procedure of measurement:

Cognitive function will be assessed by using different tests, performed during a face-to-face interview or on a touch screen. Different domains of cognitive function will be covered with the proposed tests: general cognitive ability (e.g., number series task), working (e.g., digit span backward) and episodic memory (e.g., Hopkins-Verbal-Learning Test revised), verbal

fluency (e.g., animal names), processing speed (e.g., letter digit substitution test), and executive functioning.

Both established and more recently developed measures of cognitive functions which may be affected in diseases targeted by the National Cohort study will be examined during a feasibility study with the aim to derive a battery of feasible, age- and disease-sensitive measures of attention, executive function, and episodic memory. All tested instruments will be validated and simple to implement, even if applied by using a PC.

Based on the results of the feasibility study, a set of instruments will be selected for the National Cohort with a maximum total duration time of 15 min (10 min for testing and 5 min for touch screen).

A.3.3.5 Respiratory disease-related measures

Lung function: spirometry

Spirometry is a well-established technique for assessing lung function and can easily be performed in terms of technical requirements and with minimal inconvenience to the study participant. Spirometry assesses lung function by measuring inhaled and exhaled volumes of air as a function of time. Parameters usually assessed are FEV1 and forced vital capacity (FVC)^{202, 612-615}.

Spirometry provides an integrative signal covering the mechanically relevant aspects of lung structure by fully noninvasive means. The results demonstrate chronic repair and remodeling processes in the lung and therefore correspond with disease development and progression but also with loss of function due to organ aging. Spirometry is used to assess conditions such as asthma and COPD and to distinguish between obstructive and restrictive lung disease. In addition, spirometric data are considered to be a powerful predictor of longevity and aging.

Procedure of measurement:

To guarantee valid results, cooperation between the participant and study nurse and patient effort are necessary. After adequate explanation of the maneuver, the participant will be asked to exhale several times as fast and hard as possible until three measurements without artifacts have been obtained⁶¹⁶. It takes about 8 min to obtain a set of valid, reproducible data.

Airway inflammation: NO in exhaled air (FeNO)

The analysis of exhaled NO provides information on presence and severity of allergic airway inflammation in airways disease such as allergic asthma and allergic rhinitis. Exhaled NO (FeNO) is the only currently available measure of specific airway inflammation that is rapid and easily accessible (noninvasive) and at the same time rather informative. Determination of exhaled NO at an expiratory flow rate of 50 mL/s is clinically established and accepted^{617, 618}.

Measurement of exhaled NO is simple, quick to perform, and shows good reproducibility⁶¹⁹⁻⁶²¹, and thus has been applied in several epidemiologic studies^{617, 622-624}. Available SOPs provide for highly standardized measurements⁶¹⁸ to be performed in a multicenter study setting.

Procedure of measurement:

The NIOX[®] Flex (Aerocrine) device will be used to measure NO in exhaled air. After having placed the filter in the mouth, subjects will be asked to inhale to total lung capacity NO-free

air and then to exhale slowly, keeping a constant flow. The measurement will be repeated two to three times; overall, the tests will last about 6 min.

A.3.3.6 Examination of the musculoskeletal system

The medical examination of the musculoskeletal system will measure mobility of several joints, the presence of joint swelling or tenderness, and sensitivity to pain at baseline and at the follow-up visit. Thus, we will be able to monitor changes in function and sensitivity to pain over time and to screen for individuals at risk for specific diseases endpoints in a prospective manner. Detection of incident musculoskeletal disease based on the following examinations will include osteoarthritis, rheumatoid arthritis, and musculoskeletal pain disorders. The data will also be helpful in interpreting physical activity data.

A.3

Procedure of measurement:

Spinal flexibility will be measured with lateral spinal flexion (recorded in cm).

The straight leg raising test (degrees) will be used to analyze the risk of radicular back pain. This is complemented by an assessment of radiating back pain. The measurement of pain sensitivity at the sacroiliac joints, using algometers, will complement the detection of back pain of inflammatory origin.

Hip internal rotation and knee flexion/extension (expressed in degrees with respect to the neutral position) will be used to measure mobility in these joints and knee crepitus (recorded as “absent” or “present”) as a screening finding suggestive of osteoarthritis⁶²⁵. Sensitivity to pain on the patellar and Achilles’ tendon entheses will be measured with the Algometer™.

The presence of tender and swollen joints will be assessed by palpating and moving several joints, comprising the shoulder, elbow, wrist, metacarpophalangeal, proximal interphalangeal, and knee joints. The “joint count 28” focuses on the joints that are commonly affected by rheumatoid arthritis (RA) and it is a major component of the extensively validated Disease Activity Score (DAS) 28 for RA⁶²⁶. However, we foresee a need to adapt this to the requirements of the National Cohort. Thus, the standardizability, duration, and acceptance of this examination will be studied in a feasibility study.

A.3.3.7 Oral health

Tooth count

Tooth loss is associated with an increased risk for a number of chronic diseases. The number and distribution of remaining teeth and how missing teeth are restored modulate the effect of caries and periodontitis on subsequent tooth loss. Bearing in mind the consequences of dental destruction and tooth loss, there is a remaining need for dental restorations and prosthodontic devices in patients with reduced dentition. The dental health service research will be intensified with the aim to improve the quality of dental restorations and well-being of patients and to reduce costs.

Procedure of measurement:

The number of lost teeth will be documented by trained personnel for each of the 200,000 subjects.

Oral examination (clinical examination)

The clinical appearance/phenotype of oral disease will be described as best as possible and include current and validated diagnostic tools for subjects in the intensively investigated

group (dental examination, dental questionnaire, and microbiological data). Well-defined phenotypes need to be understood as a major prerequisite for future analyses of genome, metabolome, transcriptome, and microbiome of the intraoral flora (as determined in saliva samples). Host genome, oral microbiota, serum markers, and subclinical markers could then be analyzed to study the transition from health to disease more closely. This may help to identify the time when interventions are most promising and cost effective.

Periodontitis is a widespread infectious disease that leads to chronic inflammation and tooth loss⁶²⁷. While the presence of any gingival or periodontal inflammation has to be taken into consideration as a surrogate marker for other systemic diseases, tooth loss can be considered as a disease endpoint⁶²⁸. The severity of the disease can be assessed and followed up on the basis of the local or generalized distribution patterns and of the mean attachment loss.

Procedure of measurement:

For clinical examination of the oral cavity, only measurements that can be easily recorded and do not require additional devices were chosen. Basic equipment (e.g., a halogen lamp to illuminate the oral cavity, a Community Periodontal Index, CPI, probe, a dental mirror, and cotton rolls for drying teeth) as well as a dentist's chair are needed to carry out the proposed dental examinations. A standardized protocol adapted from the SHIP study⁶²⁹ will be available, and standardized training of the nondental professionals will be provided centrally.

To clinically assess periodontitis, the number of lost teeth as well as attachment loss, measured with the CPI probe on each of the designated teeth⁶³⁰, must be determined.

Cavitations due to caries will be determined according to the WHO standard as tooth-related D-component of the DMF-Index⁶³⁰. Furthermore, implants, dental restorations (fillings, crowns, and bridges) and the presence of prostheses will be recorded⁶³⁰.

The short clinical examination will also identify acute and/or chronic symptoms of temporomandibular disorders. This examination will encompass mouth opening (active and passive, measured in cm using disposable Therabite™ scales), and pain sensitivity on the temporomandibular joint and the masseter and temporalis muscles (Algometer™⁶³¹).

In a feasibility study, the optimal examination procedure will be tested and selected for application in the National Cohort.

A.3.3.8 Sense organ functions and diseases

Ophthalmological measurements

Eye-related measurements will focus on retinopathy; however, visual acuity and refraction will also be determined.

Retinal vascular characteristics are risk indicators for major systemic causes of morbidity, including diabetes, hypertension, stroke, ischemic heart disease, and all-cause mortality, even after adjustment for known risks factors^{178, 184}. Retinal changes may reflect changes in other vascular beds (e.g., in the brain or kidney). Unlike other organs, the vascular circulation of the retina can be directly visualized by taking and analyzing a photograph of the retina, which is a noninvasive procedure. Retinal photography has already been employed in population-based studies, for example, in the Atherosclerosis Risk in Communities study (ARIC), the Cardiovascular Health Study (CHS), and the UK Biobank study.

Using retinal photography, the intra- and interobserver agreement in classifying retinopathy can be considered as excellent for microaneurysms (K value of 0.81 to 0.91) and retinal

hemorrhages (0.85 to 0.99), and as moderate to excellent for other retinal signs (retinal vessel diameters, retinal vein occlusion, and arteriolar emboli), except for arteriovenous (AV) nicking and focal arteriolar narrowing¹⁸².

Procedure of measurement:

All measurements will be done employing a nonmydriatic procedure. Examinations for visual acuity and refraction and retinal photography will be conducted within 10 minutes, using a Nidek AR-360^a camera and a Nidek AFC – 230 (Visucam, Zeiss, Jena, Germany), respectively. The examination room must have dimmed light conditions.

Hearing impairment

A.3

Noise-induced hearing loss is a major cause of deafness and hearing impairment. Non-modifiable risk factors include increasing age, genetics, male gender, and race. Modifiable risk factors are voluntary exposure to loud noise, failure to use hearing protection, smoking, lack of exercise, poor diet, and the presence of diabetes and CVD⁶³². Aging and occupational and leisure time exposure to noises represent the most frequent risk factors. Around 10–15% (5 million) of German employees are exposed to noise >85 dB at work, and noise-induced hearing loss is the most frequent occupational disease. In Germany, the prevalence of a clinically significant hearing loss is estimated to be 37% for people aged 61–70 years and 60% for those aged 71–80 years^{539, 633}. Hearing impairment particularly affects social life and contacts to other persons, on both a private and professional level, and can lead to social isolation. It is also associated with reduced quality of life, depression, and cognitive disorders/impairment⁶³³. Hearing loss develops slowly, often going unnoticed by the subject. Among individuals with diabetes, there appears to be an increased risk of hearing loss and premature loss of hearing. Hyperglycemia may cause blood vessels in the inner ear to narrow, which disrupts the normal transmission of sound. Similarly, subjects with CVD are more likely to suffer from hearing loss⁶³².

During the last few years, different screening tests have been developed that are suitable for use in epidemiologic studies. For example, a telephone-administered audiometric screening test was introduced in a Dutch study⁶³⁴ and – in a slightly modified version – in the UK Biobank study (www.ukbiobank.ac.uk).

Procedure of measurement:

Full audiometry is not applied; instead, a fast and simple screening test (that can also be applied by telephone during follow-up) will be administered. This specific speech-recognition-in-noise test assesses hearing disability and was also used in the UK Biobank study. At baseline of the National Cohort, this test will be applied by means of touch screen and headphones. During the test, the participants must recognize different combinations of numbers presented in a noisy background with varying intensity. In this way speech reception threshold can be determined in speech-shaped noise, which represents the signal-to-noise ratio where a person correctly recognizes 50% of the spoken words. Mean measurement error of this test is within 1 dB⁶³⁴. The test takes about 4 min and will be performed as a self-administered test.

Olfactory function (smell)

The prevalence of olfactory dysfunction in the general population is a matter of debate. Many authors reported frequencies of 1–5 % of anosmia within the groups studied⁶³⁵. According to the National Institute on Deafness and Other Communication Disorders (NIDCD)

200,000 Americans visit their physicians each year with smell disorders. Similarly, a recent survey from Switzerland, Austria, and Germany indicates that this figure is at approximately 70,000 in the three German-speaking countries^{635, 636}. Causes of smell dysfunction are manifold, the most important ones being aging, sinusal disease, head trauma, or infections of the upper respiratory tract. Loss of olfactory function is also related to neurodegenerative disease, as it has been shown to be an early sign of Parkinson's disease^{197, 637} and Alzheimer's disease^{198, 638}.

Procedure of measurement:

The test of olfactory function will be performed using the "Sniffin' Sticks" (Burghart, Medizintechnik Wedel; Germany). This test is an odor-dispensing system based on felt-tip pens; it has previously been validated against large smell test batteries and used in a population-based study in Germany⁶³⁹. Twelve common odors are administered. For odor presentation the pen's cap is removed, and the pen is placed in front of both nostrils for approximately 3 s. Using a multiple forced-choice task, odors must be identified from a list of four descriptors for each odor. Following recommended cut-off scores, hyposmic subjects will be defined as subjects identifying less than 10 but more than 6 out of 12 odors correctly. "Functionally anosmic" subjects identify 8 or less odors out of 12 correctly.

A.3.3.9 Anthropometry

Body height, body weight, and obesity

Height is of scientific interest as a predictor of morbidity and mortality for CVD and respiratory diseases, metabolic diseases, cancer, and other diseases⁶⁴⁰⁻⁶⁴³. Gaining an understanding of the proposed associations and identifying further factors (e.g., environmental factors) that may explain these associations represent a challenge for future research. Furthermore, height is an important confounder or explanatory variable as it is independently associated with blood pressure and blood lipids^{644, 645}.

Differences in weight can mostly be accounted for by height and body fatness. Body weight in relation to body height is a simple and widely used indicator of body fatness.

BMI is a simple measure of relative body weight and is often used in health-related analyses of anthropometric cohort data. Although BMI does not directly measure the amount of body fat, it was found to be an appropriate indicator for obesity as it is highly correlated with body fat⁶⁴⁶.

Procedure of measurement:

Specifically trained and certified personnel will measure anthropometric parameters according to a standardized protocol. Body height will be measured with an accuracy of 0.1 cm using an electronic device (e.g., Seca²⁴²). The device has to be correctly set up and the participant should stand unsupported and upright with the back towards the pole. The measurement will take 1–2 min and requires only little staff training.

Body weight will be measured with an accuracy of 0.1 kg using, for example, the SECA⁶³⁵ device. The participant should wear light clothes only (no shoes) and will stand in the center of the device, without leaning or holding onto anything. The measurement will take 1–2 min and requires only little staff training.

Waist and hip circumference

Measures taking fat distribution and location of body fat into account are important in relation to health outcomes. A genetic basis for waist measurements that is independent of the

genetics of BMI has recently been described⁶⁴⁷. Waist circumference is a common measure of abdominal fat mass and can be obtained quickly. The ratio of waist circumference to hip circumference has been used as an indicator for abdominal fat accumulation and fat distribution among adults. The waist-to-hip ratio is a particularly informative predictor of vascular risk, e.g., it is an additional predictor of incident MI⁶⁴⁸. The waist-to-height ratio is a further simple index of body fat distribution that is only weakly associated with age but is highly correlated with visceral fat mass among adults^{649, 650}.

Procedure of measurement:

Measurements of waist and hip circumference will be taken with a flexible tape (e.g., Seca 201) according to a standardized protocol. Waist circumference will be measured at the smallest position between the lower rib and the upper margin of the iliac crest. Hip circumference will be measured at the widest part over the hips or buttocks region. The measurements will take about 2 min. Measurements will not be taken in pregnant women (>12th week of pregnancy).

A.3

Dual-energy X-ray absorptiometry

DXA is a noninvasive, easy to use, and relatively fast examination for assessing the three body compartments bone mass, fat mass, and lean body mass. It is the gold standard procedure to estimate bone mass and to predict bone fracture risk^{651, 652}.

Comparing DXA to other techniques for assessing body composition (e.g., underwater weighing) has shown that DXA correlates highly with other measures of body composition, suggesting that DXA is an accurate method⁶⁵³⁻⁶⁵⁵. Historically, underwater weighing has been considered the reference standard for the assessment of body composition. Due to its ease of use and the fact that it provides a three-compartment model, DXA has the potential of becoming a new reference standard⁶⁵⁶.

Procedure of measurement:

In a feasibility study it will be tested whether DXA measurement can also be applied in observational studies with healthy subjects in Germany. Strict legal and administrative requirements must be fulfilled. Furthermore, the specific device for DXA scans will be selected within the feasibility study. The dose of radiation is very low (<20 μ Sv), much lower than the radiation dose for a chest radiograph and comparable to the natural background radiation over several days. A whole-body scan that can quantify body composition (also for specific compartments, e.g., abdominal region) takes approximately 3.5–5 min; standard measurements of bone mineral density (lumbar spine and femur) will take less than 1 min.

A.3.3.10 Physical activity and physical fitness

Physical activity: Accelerometry

Triaxial accelerometers will allow the measurement of physical activity with good accuracy and a precise estimation of the amount of activity, the related energy expenditure⁵³² and the classification according to duration and intensity. With the recent technical and analytical refinements, these devices are partly able to identify different types of activity. Moreover, those devices can reliably identify time spent in sedentary behaviours. Despite frequent claims regarding the harmful health effects of sedentariness, investigators rarely have directly measured sedentary behaviours, which have usually been defined as the absence of higher levels of PA^{657, 658}.

By using triaxial accelerometers physical activity can be measured and classified with good accuracy and the amount of activity and the related energy expenditure⁵³² precisely estimated according to duration and intensity. With the recently introduced technical and analytical refinements, these devices are partly able to identify different types of activity. Moreover, the devices can reliably identify time spent engaging in sedentary behaviors. Despite frequent claims regarding the harmful health effects of sedentary lifestyle, investigators rarely have directly measured sedentary behaviors, which have usually been defined as the absence of higher levels of physical activity^{657, 658}.

Sedentary and light activities are better captured by triaxial accelerometers than by one-axial sensors or self-report⁶⁵⁹. In addition, several analytical techniques are available to convert the activity measurements to valid estimates of energy expenditure^{660, 661}. Past approaches to integrate measured acceleration over a fixed period of time into dimensionless activity counts have shown their inherent limitations, i.e., the effect of the interval lengths on the classification of intensity or the reduced capability to recognize activity patterns. Currently, innovative research regarding improved algorithms for estimating energy expenditure and recognizing activity patterns is in progress^{662, 663}.

Procedure of measurement:

Modern accelerometric devices are small and lightweight and can be attached to the body using elastic belts with minimal inconvenience for the participant. The hip is the preferred place of attachment, as it constitutes the center of the body mass. Devices will be worn for 7 days to cover variation during the week between work days and weekends. In order to reduce problems with compliance and nonwearing periods, small size and water resistance would allow the participants to wear the device for 7 days continuously without the need to take it off. Although two assessment periods of 1 week (spaced 6 months apart) would be optimal for quantifying physical activity on an individual level, measurement in the cohort will be restricted to one 7-day period per participant. The device will be explained and handed over to the participant during the visit at the study center and will be returned by mail after 7 days of recording.

For selecting the best-suited device; a feasibility study will be conducted. A combined device with additional features would be desirable.

Step test

Cardiorespiratory fitness as defined by maximal oxygen uptake (VO₂max) is usually assessed during maximal exercise testing. In the step test, VO₂max can be extrapolated from the heart rate response to the preset work load with good precision⁶⁶⁴. Implementing a submaximal exercise test increases feasibility, avoids having to exclude participants due to the extensive exclusion criteria of a maximal exercise test, reduces the burden for the participants, and also saves costs and time and the need for sophisticated instruments, and qualified personnel.

Procedure of measurement:

A step test is simple, requires little equipment (a heart rate monitor and a stepper) and can be conducted at every study center with almost no space requirements. A standardized step test simultaneously monitors heart rate and the increasing workload is defined by a recorded time signature.

Hand grip strength

Hand grip strength, particularly in late life, is important for many daily activities. It has been shown that muscle strength determined by hand grip strength in midlife predicts functional limitations and disability in later life⁶⁶⁵. Grip strength measurement has been shown to be associated with all-cause and cardiovascular mortality and disease risk in previous studies⁶⁶⁶⁻⁶⁷⁰. Low or impaired hand grip strength was associated with lower bone mass⁶⁷¹, with an increased risk of complications or prolonged length of stay after hospitalization or surgery⁶⁶⁷ and, among women, with an increased risk of developing incident vertebral fracture⁶⁷¹.

Although muscle strength ideally should be assessed in different muscle groups in order to obtain a measure of overall muscle strength of the entire body^{667, 672}, hand grip strength is the most common and the most feasible test.

Validity and relative and absolute reliability of the hand grip strength measurement have been examined in several populations. When standardized protocols were used, good reliability and validity were found^{673, 674}. This is specifically true for the Jamar dynamometer, a frequently used hand grip device⁶⁷⁵⁻⁶⁷⁷.

Procedure of measurement:

Hand grip strength will be measured using a JAMAR hand grip dynamometer. The participant holds the portable device in a standardized way and pushes against the device with maximal force. The measurement of hand grip strength depends on maximal effort by the participant but is easy to handle. Three repeated measurements will be conducted, which takes about 3 min overall.

One-leg-stand test / posturometry

Standing on one leg with eyes open can be used to assess balance and coordination competence of participants. The one-leg-stand test can easily and rapidly be administered to participants after short instruction by the study personnel. Good interrater reliability (intra-class correlation coefficients between 0.76 and 0.92) and reproducibility (coefficients of variation between 4.7% and 5.0%) have been shown by using standardized protocols for the balance test, respectively⁶⁷⁸.

Procedure of measurement:

For the one-leg-stand test, the participant positions the nonsupporting foot against the inside knee of the supporting leg. After becoming familiar with the position, the balance time of standing on one leg is measured in seconds with a stopwatch. The best of three trials is recorded unless the time limit of 60 s is reached in the first trial. Then the test is repeated for the other leg. Alternatively, posturometry can be used to assess the subjects' balance ability. The one-leg-stand test versus posturometry is subject to a feasibility study.

A.3.3.11 Feasibility studies for optimization of the final study program

As part of the planning phase, different feasibility studies are currently underway in order to test devices for their suitability in large epidemiologic settings and to develop SOPs. In some studies, questionnaires and web tools planned for self-administration are also being tested and optimized. A list of the currently running projects is given in **Table 3.8**. Most of the projects are funded by BMBF. Further details on these projects are given in the electronic annex of this application.

Table 3.8: Feasibility studies currently underway for optimizing the study program of the National Cohort

Project title	Main objectives
A. Feasibility projects with direct relevance for the programme of the National Cohort	
Long-term ECG monitoring [§]	Comparison of two different systems of mobile long-term ECG recording against a reference method ("gold standard"); Evaluation of the suitability and acceptance by study participants.
Diagnostic methods for measurement of subclinical CVD [§]	Evaluation of different non-invasive diagnostic methods for the examination of the peripheral and central arterial system with respect to feasibility and acceptance; Assessment of reliability and validity; Development of SOPs.
Mono- and multifunctional devices to measure physical activity [§]	Test of different commercially available mono- and multifunctional devices that measure body movements for at least 7 days, for detailed, objective and reliable assessments of physical activity and energy expenditure.
Test battery for determination of physical fitness [§]	Development and evaluation of a test battery for the combined determination of different aspects of physical fitness for implementation in the National Cohort; Analysis of acceptance by study subjects, assessment time, and of test limitations.
Multipurpose sleep assessment module [§]	Test of the feasibility of selected measurement tools that enable the objective assessment of sleep apnoea and further sleep characteristics (multi-functional and mono-functional devices; sleep questionnaire; measurement of melatonin).
Web-based 24-h diet recall [§]	Adaption of a web-based British 24-h diet recall to the German situation; Determination of included food items through analysis of representative national food intake data; Test of acceptance of the web-based 24-h diet recalls by study participants.
Biomaterials [§]	Several sub-studies to minimize pre-analytical artefacts and thus achieving optimal quality and maximal standardization of specimen collection, primary processing, transport and storage of the samples; development of SOPs for application in the National Cohort; Evaluation of the acceptance by study participants.
Centralized image acquisition, evaluation and data storage management system [§]	Evaluation of the feasibility of different data management systems to store large data volumes centrally by assuring optimal data safety.
Oral Health [§]	Examination of the feasibility of proposed investigations (tooth count, periodontal examination, questionnaire) and the required time for the assessment; Evaluation of the training methods for study nurses.
Neurocognitive and neuropsychiatric phenotypes [§]	Examination of established and more recently developed measures of cognitive functions for use in the National Cohort; Evaluation of neuropsychiatric and neuropsychological instruments to assess early and later changes in the course of mental disease development.
Epidemiologic measurement instruments for infectious and immunological diseases	Exploration of the feasibility of proposed assessment instruments for infectious diseases and immunoepidemiology (e.g. assessment of immune risk phenotypes with a questionnaire, active surveillance tools for acute infections).
Optimal integration of the OGTT into the study program	Evaluation of possible influences of the OGTT on other parameters such as blood pressure, ECG, or cognitive function in order to optimally embed the OGTT in the examination program of the National Cohort.

Project title	Main objectives
Implementation of DXA measurement in population-based studies	Evaluation of the feasibility and acceptance of DXA measurement as a precise, noninvasive and quick method for the assessment of bone density and body composition.
Eye-related measurements	Evaluation of suggested examinations (retinal photography, visus and refraction, etc.) concerning their application in a population-based cohort study; Assessment of examination time and optimal training of study nurses.
Comparison of MRT devices for use in the National Cohort	A newly available 3.0 Tesla MRT device will be compared with a standard 1.5 Tesla MRT device with respect to image quality and handling.
Adaption of clinical musculoskeletal examinations to the needs of a multicentric epidemiologic study	The suggested clinical musculoskeletal examinations will be tested for its standardizability, duration, and acceptance by the study participants. Development of an optimal protocol for use in the National Cohort.
B. Additional projects with no direct implication for the study programme of the National Cohort as applied for	
Migrant recruitment §	Development and testing of recruitment strategies for population-based studies assessing health status among immigrants; Focus will be given to the two biggest immigrant groups in Germany, the Turks and the (German) immigrants from the former Soviet Union.
Visceral and abdominal subcutaneous fat assessment using ultrasonography §	Examination of the feasibility and reproducibility of using ultrasonography as a non-invasive method to estimate visceral adipose tissue (VAT) and subcutaneous adipose tissue (SCAT) in study participants.
3-D body scanner for anthropometry	Evaluation and validation of the laser-based 3-D body-scanning technology as an effective means of measuring anthropometric variables; Evaluation of the intra- and interobserver variability of measurements, of correlation and agreement of measurements with "classical" anthropometric measures, on performance of the technology for extreme body builds, on limits of the technology, and on acceptance of the technology by participants.
Focus group analysis §	Evaluation of the populations' general knowledge and understanding of cohort studies and biobanks, opinions and perceptions according to data collection, data handling, data protection, consent, and confidentiality.
Evaluation of markers for skin ageing by digital imaging	Development of a protocol for the assessment of skin aging with digital imaging in an epidemiologic context; Application of the protocol by investigating the association between markers of skin aging and early markers of diabetes.
Low-dose cardiac computed tomography	Determination of the feasibility of cardiac computed tomography (CT) and CT angiography (CTA); Determination of the technical, personnel, and logistical pre-requisites to perform coronary CT and CTA in the context of an epidemiological study such as the National Cohort.

Project title	Main objectives
Significance of fat distribution and mitochondrial dysfunction for aging processes (magnetic resonance spectroscopy, MRS)	Development of optimized SOPs for phenotyping of subgroups in multicenter studies (ectopic lipids, energy metabolism) using MRI and MRS (1H, 31P MRS); Development of standardized processes to analyze raw data and set up databases for rapid access to data of selected groups of patients of different study centers.
Assessment of zoonotic infections for chronic disease risk	The ability of the pet owners to obtain nasal swabs and stool samples from their cats or dogs, and their willingness to have a veterinarian draw blood from their pets will be tested in a feasibility study.
Access to detailed data on occupational history	Evaluation of the procedure of data linkage for a detailed occupational history retrieved from secondary data sources (IAB Database of the Institute for Employment Research, Federal Employment Agency).
Use of bracelet dosimeters for assessment of natural radiation exposure	Evaluation of the use of bracelet dosimeters to measure natural background radiation in an epidemiological study; Testing of feasibility and acceptance by the study subjects.
Estimation of cumulative ionizing radiation through medical exposures	Evaluation of the feasibility of the use of data from health insurance companies (in combination with questionnaire data from the study participant) to get information on the type and frequency of applied diagnostic and therapeutic devices using ionizing radiation for the assessment of cumulative medical exposure.

[§] funded by BMBF grant

A.3.4 Assessments through medical imaging

A.3.4.1 Rationale

Biomedical imaging has become paramount for the early detection and diagnosis of various disease states and is also increasingly being used in nonclinical settings for research purposes to provide comprehensive phenotyping of the human body in its various physiological and (pre-)pathological states. This rapid development has been fostered by continuous technological advances which have made it possible to perform imaging examinations non-invasively at very high spatial and temporal resolution, providing morphologic and functional information of the highest detail. Moreover, labor-intensive, complex imaging procedures such as MRI protocols have been established for acquiring highly standardized images in heterogeneous (patient) populations, thus forming the methodological prerequisites for population-based imaging.

In the National Cohort, a comprehensive whole-body MRI protocol will be implemented in a subcohort of 40,000 study participants, so as to complement the biochemical and genetic phenotyping by detailing various morphologic and functional biomarkers for cross-sectional and longitudinal analysis. This section delineates the scientific background and objectives for this highly innovative study component and presents the study logistics.

A.3.4.2 Background

In contrast to many other imaging modalities, MRI permits comprehensive and simultaneous imaging of multiple organ systems using highly standardized acquisition protocols. Today, MRI is the technique of choice for highly sensitive and specific detection of many different

diseases in various organs and is routinely applied in low-prevalence populations. However, to date, whole-body MRI has been rarely implemented in larger prospective cohort studies, given the long acquisition and postprocessing times and insufficient resolution in the past. Except for the SHIP study^{679, 680}, whole-body MRI has not been implemented in any major cohort nationally and internationally. Using a 3-Tesla scanner, the National Cohort will be in a unique position to determine the predictive role of advanced, high-resolution, morphologic phenotyping.

In the past, the assessment of multiple organs was restricted by a long examination time since the patient had to be repositioned for single organ examinations and coils had to be changed several times. Faster approaches for imaging of the entire body could only be implemented at the expense of spatial and temporal resolution. Indeed, the capability of performing multiregion and whole-body MR scans in clinical routine is a relatively new development⁶⁸¹⁻⁶⁸³.

A.3

New technical improvements such as the introduction of parallel acquisition techniques (PAT)⁶⁸⁴, continuous table movement techniques, multichannel receiver coils, and new sequences^{685, 686} make it possible to examine different organ systems in a whole-body approach within a reasonable scanning time of approximately 45 min⁶⁸⁷⁻⁶⁹¹. Continuous table movement (CTM), for example, has recently been introduced as a promising new concept to accelerate MRI acquisition^{689, 690, 692-696}. By using this technique, real-time whole-body MRI is a seamless process during data acquisition, thereby reducing acquisition time by a factor of 2–3. Recently, the technique has been employed in MR angiography (MRA)^{691, 697-699} and oncological imaging⁷⁰⁰⁻⁷⁰⁵.

MRA shows a higher accuracy, less operator dependence, a larger field-of-view, three-dimensionality, and better contrast resolution than ultrasonography. So far, contrast-enhanced MRA using extracellular contrast agents represents standard practice. However, a large variety of recent studies have proven that non-contrast-enhanced MRA may provide an equivalent alternative imaging approach for patients who are at risk of developing nephrogenic systemic fibrosis and other contrast-related complications and might therefore be of high interest for imaging a large cohort of subjects⁷⁰⁶⁻⁷⁰⁸.

A.3.4.3 Objectives

The proposed MRI imaging substudy will apply a standardized comprehensive MRI protocol to a random set of asymptomatic individuals both at baseline and follow-up. This will generate a comprehensive morphologic and functional biorepository, which will serve as a source for numerous analyses to understand the role of MRI-derived morphologic changes and abnormalities in the natural history of various conditions as well as their potential value as risk factors for the prediction and development of subclinical and overt disease states. The specific objectives detailed below were formulated in collaboration with each thematic working group, indicating their focus and major interest, but may be considered as examples among a rich set of other novel, predominantly hypothesis-generating scientific foci.

The overall objective is supported by parallel and complementary objectives related to five body organ systems. These objectives are addressed by the specific steps within the imaging protocol (see below).

Brain / nervous system

Subclinical cerebrovascular disorders

MR imaging is the only available method to comprehensively examine brain morphology noninvasively and without radiation exposure. Particularly, the prevalence and incidence as well as cross-sectional and longitudinal associations and predictive value of subclini-

cal cerebrovascular disorders, including gliosis, microbleeds, and angiopathy and carotid and intracranial plaque burden, for the development of cerebrovascular and cardiovascular events can be evaluated.

Gliosis is a proliferation of astrocytes in damaged areas of the central nervous system. Gliosis and neuronal loss in brain regions are seen in various neurodegenerative disorders, such as Alzheimer's disease, Korsakoff's syndrome, and Parkinson's disease and following acute episodes of multiple sclerosis. Although gliosis belongs to the neurodegenerative disease category and some other diseases in this category are caused by vascular disorders, there is uncertainty as to the possible relation between gliosis and atherosclerotic diseases.

Microbleeds and amyloid angiopathy are accurately detected by brain MRI and share common risk factors. Particularly, hypertension is related to an increased risk of both⁷⁰⁹. The role of further atherosclerotic risk factors including smoking, diabetes, and obesity is less clear. We hypothesize that the risk factor profiles for incident carotid artery plaques, intracranial vascular plaques, and microbleeds are different. In addition, we assume that carotid atherosclerosis is not associated with amyloid angiopathy. Rather, genetic factors and gene-to-environment interactions predispose individuals to specific subclinical cerebrovascular disorders.

Neurodegeneration

MRI in a population setting offers various research options for assessing prevalence and incidence of neurodegenerative changes of the brain, which include measuring brain volume (total volume, gray and white matter volume), alterations in microstructural integrity (through diffusion tensor imaging), vascular brain lesions (white and gray matter lesions, lacunar infarcts, microbleeds), subclinical atherosclerosis (lacunar infarcts, white matter lesions, carotid plaques), and functional connectivity networks and their role in cognitive impairment, emotional functioning, and dementia. On the one hand, information on neurodegenerative changes at baseline will serve as exposure to assess the future risk of incident MI, stroke, mild cognitive impairment, and dementia. On the other hand, by performing follow-up MRI examinations incident neurodegenerative disorders can be defined for incidence and risk factor analyses. Finally, the changes in neurodegenerative disorders will be analyzed in relation to incident fatal and nonfatal cardiovascular events that have occurred after follow-up MRI examinations.

Cardiovascular system

Cardiac function

MRI is the most accurate noninvasive and nonradiation-based method to determine left ventricular mass. Using this technique left ventricular diastolic and systolic function can be accurately assessed globally and regionally. In addition, morphology and function of the right ventricle can be evaluated. While there is clear evidence that impaired systolic function of the left ventricle is one of the strongest predictors of cardiac mortality, only older studies indicate that right ventricular dilation and dysfunction may add predictive value to this association⁷¹⁰. Follow-up examinations of cardiac MRI will form the basis for identifying risk factors for incident cardiac hypertrophy and dysfunction and for associating the change in these cardiac parameters with incident cardiac morbidity and mortality.

Vascular morphology and stiffness

Recent technologies have substantially improved the use of non-contrast-enhanced MRI for detecting and quantifying impaired vascular structure and function. Thus, in the National Cohort MRI will also be used to determine the prevalence, cross-sectional and longitudinal

associations, and the predictive value of vascular disease, i.e., stenosis, stiffness, for the development of vascular disease and diabetes.

Musculoskeletal system

Osteoarthritis

No large-scale population-based studies have implemented MRI to ascertain the incidence of osteoarthritis; nor is any MRI-based information on prevalence and incidence of osteoarthritis-related MRI findings in the general population currently available in Germany. While the time constraints in the MRI substudy of the National Cohort will not allow intensive and comprehensive assessments of single joints, the substudy will detect early degenerative changes of the spine and the hip joints. This information will be used to address the question of the extent to which these early degenerative changes are associated with the development of hip arthropathy and back pain. We will also investigate the hypothesis that early degenerative changes in joints and spine are strong predictors of loss in quality of life and health care utilization. Furthermore, we will study putative interactions between genetic and environmental factors, including occupational factors, smoking, and physical activity with respect to the progression of degenerative changes of joints and spine.

Inflammatory arthropathy and sacroiliitis

Given the additional time effort for detailed MRI-based examinations of the hands, it will not be possible to investigate signs of rheumatoid arthritis. However, in the framework of the MRI substudy of the National Cohort we will determine the prevalence, cross-sectional and longitudinal associations, and predictive value of detected early changes that suggest spondyloarthropathy in order to assess subsequent resource utilization of sacroiliitis for the development of back pain and development of chronic arthropathy. We assume that yet unknown genetic factors determine the onset of spondyloarthropathy and sacroiliitis. Furthermore, we will study whether subclinical sacroiliitis is associated with prevalent and incident back pain and the development of chronic arthropathy and with substantial resource utilization.

Volumetry and structure of large organs

Volumetric and structure analyses of organs will be preferably conducted using fully automatic reading systems. In part, this work will be done using existing software⁷¹¹, but for other readings, dedicated algorithms must be developed. The development of these systems, however, lies beyond the scope of this study and will be done during and following data collection.

The two major advantages of whole-body MRI over ultrasound are the comprehensive imaging of many organs within one examination and the higher accuracy. Within this project, we will identify prevalent organ shape and size in cross-sectional relation to risk factors and diseases and, in the longitudinal approach, the determinants of change in organ shape and size as well as their predictive value for incident diseases, placing particular emphasis on the risk of organ failure developing in the future. Examples for research questions to be addressed are the associations between the volume of the spleen and incident inflammatory diseases, the volume of the pancreas and incident diabetes mellitus, and the volume of the liver and incident cardiovascular disease, type 2 diabetes, and mortality.

Volumetric analyses will be supplemented by studies of structural changes of organs to identify prevalence and incidence, cross-sectional and longitudinal associations, and pre-

dictive value of structural alterations and tissue composition for the development of metabolic and inflammatory disorders. Thus, we hypothesize, for example, that early changes in lung and airway structure are strong predictors for inflammatory disease and that early changes in kidney structure are strong predictors of cardiovascular and metabolic diseases.

Metabolism / energy balance

MRI can explicitly distinguish between body fat compartments, including fat in subcutaneous, visceral, intrahepatic, intrapancreatic, and intraskeletal localizations. Particularly with respect to the risk of future cardiovascular and metabolic diseases – including insulin resistance, beta cell dysfunction, and type 2 diabetes – the dedicated predictive value of volume of each of these compartments can be assessed. Thus, we will determine which specific fat compartment has the strongest predictive value for incident cardiometabolic diseases and hypothesize that the volume of visceral fat and the amount of liver fat are stronger predictors for incident type 2 diabetes and CVD than are other fat compartments or somatometric measurements. In addition to the specific role of each of the fat compartments, fat-to-muscle ratio may add further predictive value to the risk of cardiometabolic diseases. By using this database, investigations can be conducted on the interactions of genetic factors with nutrition and physical activity and predict the fat-to-muscle ratio and changes over time.

A.3.4.4 Study population, imaging methods, and planned study logistics

Overview

The comprehensive native MRI scans will be performed at dedicated imaging sites in close proximity to the study centers. These imaging facilities will be uniformly equipped with identical 3-Tesla MRI technology and adequately trained and certified staff will ensure high standardization and internal validity of the acquired MRI data sets. Upon completion of the examination, data will be transferred to a centralized reading core at the competence center on imaging for a "first-line" reading to determine the presence of relevant (major) incidental findings and for quality control. It is planned that serious incidental findings will be reported immediately to the imaging site, which will initiate appropriate algorithms to manage conditions that have been detected; the latter procedures, however, are still subject to feasibility studies.

Data will also be transferred to an integration center for long-term archiving and so that they are available for subsequent "second-line" reading and scientific analysis. Using similar technical equipment, a follow-up examination is scheduled after 5 years.

Study population

Anticipated subresponse: Currently, no final decision has been reached concerning how to design the invitation procedure at the MRI sites. The two variants under discussion are: 1) To treat the MRI examination as a substudy. Here, all participants to be examined at the site are offered to participate in the MRI examination on a voluntary basis. According to the SHIP experience, a subresponse (participation rate) of up to 60% can be expected. This would have the consequence that only National Cohort study centers or regions with >15,000 participants will be qualified as a MRI site; 2) only those subjects are included at MRI sites who are not only willing to participate in Level 1 and 2 examinations, but also in the MRI examination. At present, the first variant is favored. However, currently, too few study centers exist to examine >10,000 subjects.

Informed Consent: Informed consent for the MRI examination will be obtained at the study center by a physician. A flyer containing all relevant information in written form will be available to the potential participants. In addition, the flyer will contain 2–3 figures showing the MRI scanner and the imaging set-up. A video screen will be installed to show a brief movie detailing the scanning procedure and the coil positioning as well as highlighting the duration of the examination.

Set-up of imaging sites

Four imaging sites will be established to perform the whole-body MRI examinations in order to safeguard highest image quality, internal validity, and cost-effectiveness. Scientific evidence strongly suggests that standardization of MRI is not feasible over different MRI devices⁷¹²⁻⁷¹⁶. Thus, to ensure that acquisitions are highly standardized and comparable, new MRI devices from the same vendor and identical equipment are provided to selected imaging sites. Association with further MRI sites will be allowed only in the case that these sites provide evidence of sufficient comparability of key findings with the four National Cohort sites. Trained and certified MRI technologists will be available at each site; they will receive centralized training and certification.

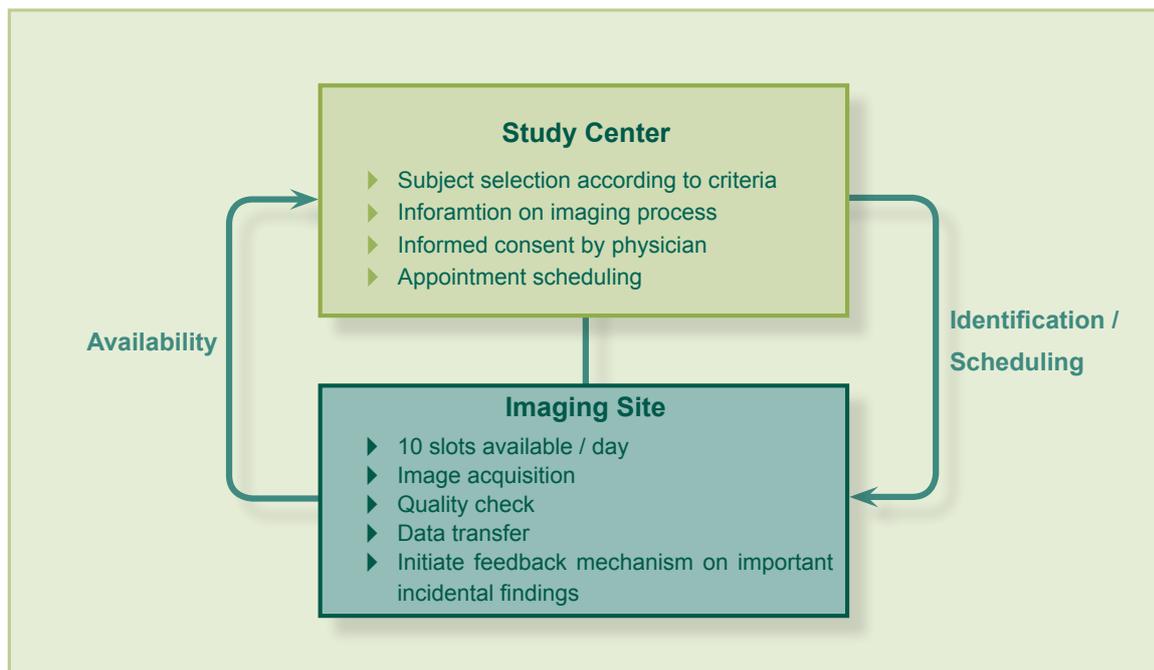
Selection process of appropriate imaging sites: Upon evaluation of the current proposal, the process for selecting appropriate imaging sites will be initiated. This will include an official and publicly available request to apply as a potential MRI imaging site for the National Cohort. The proposed selection criteria are:

- ▶ Proximity to study center.
- ▶ Maximum of one MRI device per cluster; the final distribution should take regional aspects into account (North–South, West–East, rural–urban) to assure generalizability.
- ▶ Strong expertise of local radiologists in performing large-scale clinical whole-body MRI or epidemiologic studies with high-quality standards.
- ▶ To assure optimal resource utilization, a minimum number of 5,000–10,000 subjects per site will be examined for optimal facility utilization. Consequently, not only subjects of Level 2 examinations, but also those with Level 1 examinations will be offered to participate in the MRI substudy (see MRI follow-up examination).

Interaction between study center and imaging site: Interaction between the study center and the imaging site will be of critical importance. Given the close relationship between the study center and all study participants, the study center staff will have responsibility for all information pertaining to the MRI examination and for obtaining consent and scheduling the MRI appointments. This will assure highest participation rates as relevant information can be conveyed personally.

The imaging site staff is required to inform participants about the long- and short-term availability of imaging slots using a dedicated scheduling system. In addition, the imaging site staff will need to welcome each participant, provide additional information if necessary, and detail the specific procedures. Upon completion of the scan, the imaging site should perform a quality assurance algorithm (as detailed in **Sect. A.5.7**) and transfer the data. Should an important and urgent incidental finding be detected at the first-line reading in the centralized reading core, this will be reported within 30 min to the imaging site to initiate the feedback mechanism as detailed in **Sect. A.7.2.5**. For low-priority incidental findings, this information will be transferred to the study center, which will contact the subject via a general practitioner.

Figure 3.2: Interaction between study center and imaging site.



A.3

Technical scanner specifications

Given the aforementioned objectives of the National Cohort, the current proposal takes into account the following requirements:

1. Highest flexibility of the MRI system as regards physiognomy of the volunteers and especially to body size and weight; otherwise, a potential bias by exclusion of certain cohorts in the imaging part may be introduced. For example, improper selection of a system with limited capability to scan obese volunteers will potentially bias the evaluation of cardiometabolic disorders.
2. The system should support technologies for assuring a high predictability of examination durations and also support new imaging techniques that would reduce the total rooming time required for scanning a volunteer. The number of volunteers scanned will have a direct correlation to the achieved evidence level of the National Cohort and consequently also to recruitment for the subgroups.
3. To reduce the influence of potential variation in technologists' skill levels, intelligent guidance and (semi-automation) of the scanning process is mandatory. This is expected to improve the robustness of the examination (and therefore predictability of scan time) and assure a standard of image quality throughout the cohort and imaging centers. In addition, an intelligent reduction in variables that can be influenced by the technologist during the MRI examination is mandatory to avoid excluding scans due to violation of sequence parameters of the individual MRI examination / sequences.
4. For multicenter studies, the distribution of sequence protocols and the assurance of their integrity during scanning is associated with a high level of quality checks and is often an error-prone process. Scanner technologies should therefore assist in the distribution of protocols and their application at the individual MRI site.
5. MRI screening generates large data volumes with the highest level of complexity of all imaging modalities used in clinical routine. In combination with the high volume of MRI studies, data storage, data handling, data interpretation, and an efficient reading process are crucial. In addition, reproducibility of image findings and standardization of reporting has to be considered.

- Since being introduced into clinical routine, MRI technology and application have changed dramatically; these very short development cycles in imaging technology must be accounted for in selecting appropriate up-to-date sequences and scanner technology. Considering the life-cycle of the scanners integrated in the National Cohort and to avoid inconsistencies in image quality, the latest generation of MRI scanners will be used.

Therefore, although higher in cost, a 3-Tesla system, which offers higher spatial resolution with higher diagnostic accuracy and faster image acquisition than 1.5-Tesla, is planned to be used at all imaging sites. In addition, scanners that offer semi-automated scanning processes are highly desirable to obtain a high level of standardization.

A.3

Proposed MRI imaging protocol

Table 3.9: Proposed basic whole-body MRI protocols

Sequence	Duration (min)	Phenotypic measures
Brain / neurodegenerative		
T2w FLAIR cor	2:30	Chronic vascular changes, gliosis / sclerosis
DTI ax	8:00	Microstructure, tractography
SWI ax	2:30	Microbleeds, vascular malformations, amyloid angiopathy
3D MPRAGE T1-w	5:00	Morphology, atrophy, volumetry
SUBTOTAL TIME	18:00	
Cardiovascular		
Noncontrast-enhanced WB MRA (TimCT)	3:00	Carotid, renal, aortic, mesenteric vascular disease, PAOD
Fat saturated T1-w, PD-, T2-w	5:00	Carotid plaque imaging
short-axis-, 2-, 3-, 4- CV breathhold multislice trueFISP, T-SENSE	5:00	Cardiac function: regional and global LV and RV function, ESV, EDV, myocardial mass, EF, atrial volumes and function, DDF
TrueFISP	0:30	Aortic distensibility
Valve flow quantification	2:00	Assessment of aortic and mitral valve
3D time-of-flight angiography (3D TOF MRA)	2:30	Arterial vasculature brain
SUBTOTAL TIME	18:00	
Musculoskeletal		
3D-HR-PD hip	6:00	Hip: Degenerative joint disease + inguinal lymph nodes measurement
WB STIR ax (TimCT)	3:00	MSK: inflammatory joint disease (incl. SI joint), joint effusions, spondylarthropathy
T2 TSE sag	3:00	Spine, disks, vertebral fx
SUBTOTAL TIME	12:00	

Sequence	Duration (min)	Phenotypic measures
Volumetry / others		
HASTE ax, sag, cor lung, alternative 3D VIBE / alternative TrueFisp	3:00	Lung: structure, emphysema, COPD, lung nodules
3D VIBE Abdomen	3:00	Liver and spleen volumetry, pancreas
DWI Abdomen	3:00	Abdomen, esp. kidneys (morphology and function)
DIXON liver	1:30	Liver fat assessment
SUBTOTAL TIME	10:30	
Metabolic		
WB T1-w ax (TimCT)	3:00	Subcutaneous and visceral/intra-abdominal fat quantification
SUBTOTAL TIME	3:00	
+ Localizers	3:00	
TOTAL TIME	64:30	
Protocol optimization with vendor	approx. -10%	
	60 min	

[§]Please refer to the list of abbreviations for definitions

Table 3.10: Proposed optional whole-body MRI protocol[§]

Brain		
fMRI (resting state)	6:00	Neurodegeneration, i.e., dementia, depression
Cardiovascular		
T2*, T2 and T1	10:00	Quantitative cardiac biomarkers
TrueFISP PSIR – post contrast (DCE)	2:00	Myocardial scar, fibrosis, and inflammation
Musculoskeletal		
3D-HR-PD knee (separate coil with repositioning)	6:00	Knee: degenerative joint disease
PD cor	4:00	Shoulder: rotator cuff lesions
Volumetry / others		
Prostate: HR T2-w	4:00	BPH, PC
H1 spectroscopy liver	2:00	Quantification of liver fat / structure

[§]Please refer to the list of abbreviations for definitions

Imaging workflow and participant scheduling

Given a 12-h per day MRI capacity and a 60-min examination program (40 min for data acquisition and 20 min for pre- and postprocessing), a realistic mean number of 10 subjects per day will be examined. Over a period of 5 years with 200 working days a year, 2,000 participants will be scanned per year. Thus, 10,000 participants could be examined over the 5-year recruitment period per MRI site, resulting in a total of 40,000 subjects that could be scanned at the four MRI sites.

An alternative approach is currently being discussed: Given the budget constraints, we need to consider that high numbers of cohort participants are inversely correlated with

the possible depth of examination. The arbitrary requirement to examine one subject in a 40-min examination time imposes a limitation on the amount of information that can be acquired. Longer MRI examination times of 60–70 min, for example, which are well tolerated by study subjects according to the experience at the SHIP site, will make it possible to additionally cover the research interests of most groups. Such a program would not only result in good comparability of the National Cohort data with other major studies in the field, but also provide extra and innovative options that have not been addressed in any other study. We expect a final decision on this issue in early 2011.

A further alternative is under evaluation: Instead of implementing four stationary MRI centers, vendors are asked to calculate costs for mobile MRI centers, either as a truck-based or a container-based solution. The advantage of mobile centers would be that MRI could be performed at more than four study centers. However, costs would have to fit into the financial budget.

A.3

MRI follow-up examination

Protocol: After performing baseline MRI examinations over a period of 5 years, follow-up MRI examinations will be conducted in all subjects willing to participate. The follow-up scans will similarly be performed over a period of 5 years. Identical technical equipment, technologist staff, and a similar imaging protocol will be applied to ensure highest internal validity. Thus, no adjustments will be made to the protocol detailed in the paragraph Proposed MRI Imaging Protocol.

Also, there will be no update of software or hardware.

Follow-up rate: Based on experience of the SHIP study and the Framingham Heart Study, we anticipate that approximately 50% of subjects scanned at baseline will receive a repeat MRI examination.

Workflow: The study center will maintain all contact with the participants. Subjects will be informed that a repeat MRI exam is scheduled after a follow-up period of 5 years (± 2 months) when consent is obtained for the baseline scan. All subjects who were scanned at baseline will receive a written invitation for the repeat exam which specifically highlights the importance and value of the follow-up examination. They are invited to individually schedule their examination. A follow-up phone call will be made to ensure that the participant received the written invitation.

An approximate follow-up response of 50% would imply that, in an average study center with 20,000 participants, an imaging site would have an overcapacity with respect to the number of study subjects who accept to be examined. In addition, if only 20% of the more intensively characterized participants are offered MRI examinations, this would result in a total of only 3,000 participants (50% of 6,000 participants). Thus, an additional proportion of approximately 30–50% of participants in the Level 1 examination program can be offered to take part in the MRI subproject.

Data transfer and archive

To ensure data safety and availability of all images for the centralized first-line reading in the centralized reading core, all acquired data will be stored two-fold: (1) All data will be sent to the reading core at the competence center on imaging to perform the first-line reading; and, (2) as detailed in **Sect. A.4.2**, all images will be long-term archived at (at least) one of the two national Integration Centers for the National Cohort. The preferred way of organizing the data transfer is to send the images from the imaging site to the integration centers,

and from there to the competence center on imaging. However, as a considerable amount of time may be required to transmit this amount of data, further analysis in the course of the respective feasibility study may find that it is appropriate to deviate from this standard procedure (described in **Sect. A.4**) in the case of imaging.

Image evaluation; centralized reading core

The present proposal does not include funding for detailed imaging evaluation with respect to the scientific hypotheses. However, in order to address the issue of incidental findings, a so-called "first-line reading" of the images will be performed centrally. Subsequently a "second-line", hypothesis-driven reading will be performed. Thus, similar to the approach proposed by the Imaging Group of the UK Biobank, MRI data will be regarded as infrastructure for future scientific projects that are funded by subsequent grant applications.

The "first-line reading" will only include those findings which will be potentially reported to the participants. They include acute diseases such as stroke or acute infectious diseases, which urgently need to be further diagnosed, and abnormalities suggesting tumor diseases.

Given the standardization requirements and financial constraints, we propose centralized reading with several participating radiologists. The reading is performed during or directly after the MRI examination, and the subjects receive the results directly after the examinations from the radiologists via phone calls or video conferences. This interaction will include recommendations for further diagnostic work-up.

Use of contrast agents

Gadolinium-based contrast agents, which are associated with a very low risk, have been routinely used in clinical MRI mainly to characterize suspected tumors, to detect inflammation, and for myocardial enhancement and vessel visualization. However, there are four disadvantages of using contrast agents in an epidemiologic study: Firstly, contrast dye requires intravenous administration, which is invasive (although minimally) and relatively time consuming. Secondly, there is a risk, albeit low, of adverse events and, considering the large size of the study population, it is likely that some of the participants will suffer from such adverse events. Thirdly, logistic concerns would require that valid information on allergies and renal dysfunction be available. Fourthly, there are financial constraints (additional cost of contrast agent administration per patient is approx. 75 Euro).

Therefore, contrast agents will not be administered, and only non-contrast-enhanced whole body MRI protocols will be applied.

A.3.4.5 International scientific advisory and data safety monitoring board

The proposed MRI substudy was developed and will be further specified in close collaboration with an international expert scientific advisory board with advisers from the United States, the Netherlands, and Great Britain who have extensive experience in population-based imaging.

We anticipate that other experts and experienced investigators will be added to the advisory board, which will be continued over the study period and meet on a regular basis (i.e., once a year). We also anticipate that some members of the advisory board will be elected into the Data Safety and Monitoring Board (DSMB), which will also contain statisticians and ethicists. The DSMB will meet on a regular basis (i.e., every 4 months). The major task of the DSMB will be to ensure the safety of the substudy by reviewing potential adverse events.

A.3.5 Collection, preanalytical processing, storage, and retrieval of biomaterials

A.3.5.1 Overview and general principles

The collection of biomaterials for future phenotyping at the molecular level represents a basic component central to the entire National Cohort. For such molecular phenotyping, a comprehensive array of biomaterials in optimal quality is indispensable.

Optimal preservation of quality relies on minimizing preanalytical artifacts that may be incurred during specimen collection, primary processing, transport and/or storage of the samples, including:

- ▶ Artifacts due to cell lysis, which promotes release of intracellular components with concentrations that are several magnitudes higher in the intracellular than in the extracellular compartment. This is exemplified by release of potassium, lactate dehydrogenase, or catecholamines from red blood cells in hemolysis, or of proteolytic enzymes from leukocytes, which not only alters their serum or plasma concentration but may also degrade target analytes such as insulin.
- ▶ Artifacts due to cell metabolism, exemplified by the decrease in glucose concentration upon prolonged storage of blood, or the continuing in vitro production by cells of the amino acid homocysteine (a marker of cardiovascular risk) by blood cells in vitro
- ▶ Artifacts due to the enzymatic degradation of molecular species upon prolonged exposure to 4°C or higher.
- ▶ Molecular artifacts due to repeated freezing and thawing.

Given the huge number of potential analytes and taking into account that analytes of interest and techniques may both change over the study period to an extent that cannot be foreseen today, it is mandatory to avoid all artifacts. This requires:

- ▶ The prompt and complete separation, ideally within 1 h of collection, of all particulate components of full blood to obviate the cell-derived artifacts detailed above.
- ▶ No delay in preparing aliquots and freezing to obviate enzymatic degradation during prolonged transportation at 4°C or higher.
- ▶ Volumes small enough (190 µl) to guarantee single use only as opposed to repeated thaw–freeze cycles necessarily implicated in the storage of larger volumes.

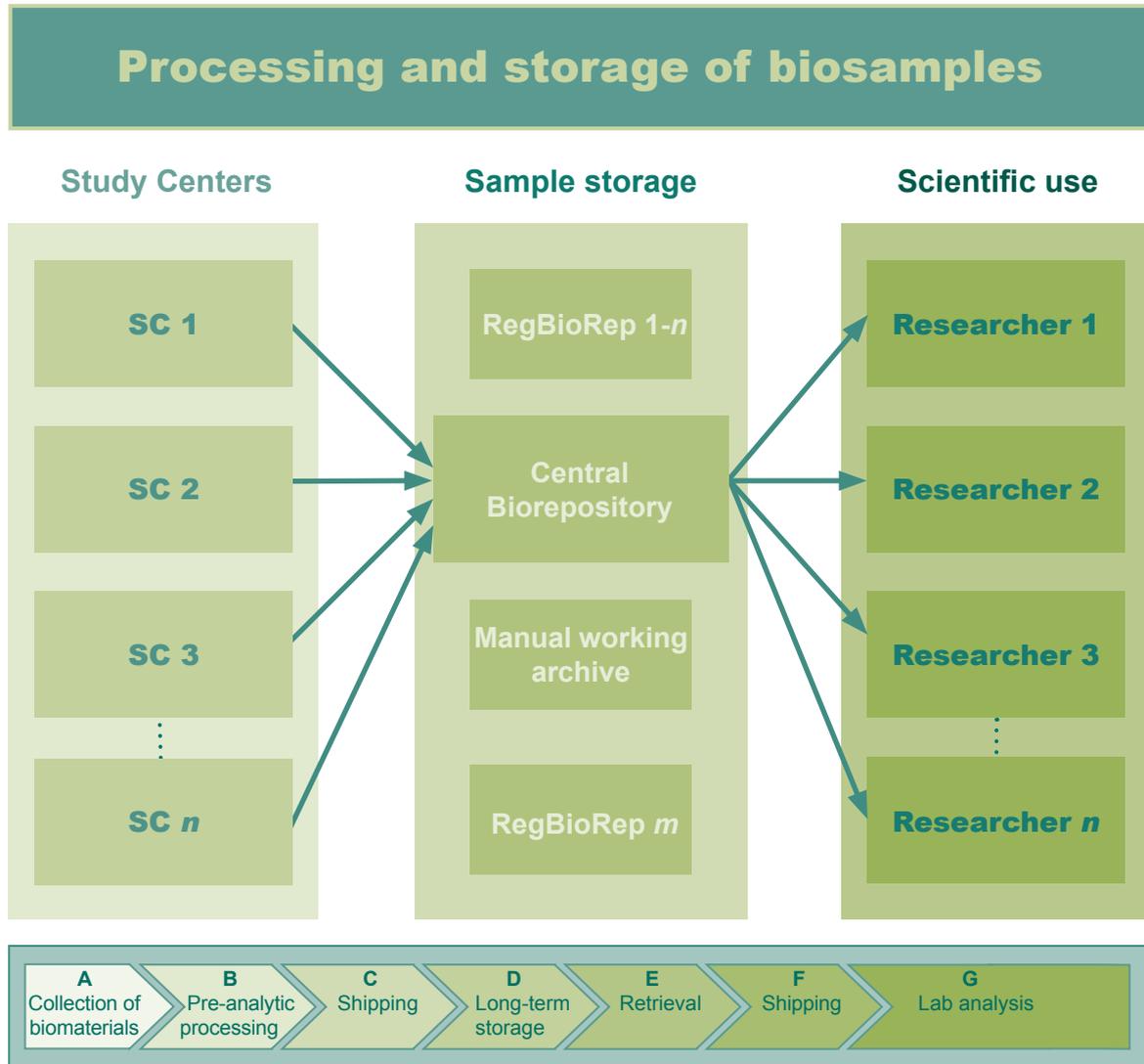
High quality and wealth of the National Cohort Biobank will be assured by the following major principles:

- ▶ Local, and not centralized processing of blood, urine, and other biomaterials, which is complete enough to reach the level of ready-prepared small aliquots that can be transported to the central storage unit on dry ice in a deep-frozen state (except for viable blood cells). This would obviate the enzymatic disintegration incurred upon prolonged exposure to 4°C or higher.
- ▶ Automation of all steps in preparation, storage, and retrieval of stored materials, promoting strict adherence to SOPs, maximizing reproducibility, and obviating artifacts that inevitably occur during manual processing in the long run due to individual failure. Thus, each study center will be equipped with a liquid handling platform.
- ▶ Storage of biomaterials from all participants throughout Germany in one central automated biorepository and decentralized back-up storage.
- ▶ The strategy also involves storing many small aliquot volumes to avoid thaw–freeze cycles and thereby to increase sample quality.
- ▶ The quality assurance and quality control of the collection and processing of bio-samples is described in **Sect. A.5.6**. Detailed SOPs and specific training materials for

sample drawing, collection, processing, transportation, and storage will be compiled by the competence unit on biomaterials.

- ▶ The concept of collection, preanalytic processing, storage, and retrieval is also shown in **Figure 3.3**.

Figure 3.3: Concept for the collection, processing, storage, and retrieval of biosamples in the National Cohort



A.3.5.2 Types of biomaterials considered

Blood obviously is considered the most important biomaterial. For biomaterial other than venous blood, samples will meet several selection criteria:

- ▶ The samples will provide major information in addition to that from blood samples in terms of physiologic coverage or coverage of organ systems (added value).
- ▶ No major complications are expected with the collection of the samples (feasibility).
- ▶ There is no indication for a major (negative) impact on participation of the invited study subjects.
- ▶ Samples can be processed and stored at appropriate costs.

Different types of biosamples were considered. The reasons for including them are listed in **Table 3.11**. **Table 3.12** gives an overview of the selected biosamples that shall be collected from all participants of the National Cohort.

Blood

In addition to the classic biological/medical data obtained by analysis of blood samples, blood fractions (DNA, RNA, proteins, and metabolites) can also be analyzed by means of high-throughput techniques available in the areas of genetics, transcriptomics, proteomics, and metabolomics; thus, the analysis of blood samples will provide a wealth of new information that will strengthen the molecular epidemiologic part of the National Cohort. Different additives and preservatives are available to separate and obtain the different blood fractions (plasma, serum, and cells). Apart from inhibitors of RNAses used to stabilize RNA, the question of using stabilizing additives first needs to be viewed in the light of the processing strategy pursued. Thus, inhibitors of glycolysis can be omitted if cells are instantaneously removed from serum or plasma following blood specimen collection since the enzymes that metabolize glucose are located in the cellular fraction. Similarly, avoiding the release of proteolytic enzymes from leukocytes by avoiding mechanical cell lysis and cell lysis following apoptosis upon prolonged transport needs to be considered first before considering the use of protease inhibitors, such as aprotinin to inhibit the enzyme systems of leukocytes. Due to the preanalytic processing in the local study centers, there is no need for most of these stabilizing agents. However, blood collection tubes containing RNase inhibitors will be used; for plasma collection, ethylenediamine-tetra-acetic acid (EDTA)-coated tubes were chosen. In addition, serum will be prepared from clotted whole blood.

Table 3.11: *Biological samples: rationale for inclusion*

Sample type	Rationale	Collection
Blood	<ul style="list-style-type: none"> > Different fractions available (plasma, serum, white cells, red cells, and peripheral blood mononuclear cells) > Different types of components present (cells, proteins, DNA, RNA, hormones, nutrients, etc.) > Suitable for a wide range of analytic procedures, including -omics technologies 	<ul style="list-style-type: none"> > Easy, low risk, well-established, standardized procedures available > Low costs of collection, except for RNA collection tubes and BD 'CPT' tubes for PBMC separation
Urine	<ul style="list-style-type: none"> > Different types of components present (cells, proteins, renal excretion products, etc.) > Suitable for a wide range of analytic procedures, including -omics technologies > Provides additional (supplemental or new) information to blood samples 	<ul style="list-style-type: none"> > Easy, well-established, standardized procedures available > Low costs
Saliva	<ul style="list-style-type: none"> > Provides specific information about oral microbiota > Different types of components present (cells, proteins, DNA, hormones, etc.) > Suitable for several analytic procedures, including -omics technologies 	<ul style="list-style-type: none"> > Easy, well-established, standardized procedures available > Low costs
Stool	<ul style="list-style-type: none"> > Provides specific information about gut microbiota > Suitable for several analytic procedures, including -omics technologies 	<ul style="list-style-type: none"> > Issue of standardization of sample collection > Specific logistics for sample collection necessary
Swabs	<ul style="list-style-type: none"> > Provides specific information about nasopharyngeal microbiota 	<ul style="list-style-type: none"> > Standardized collection according to SOPs

Urine

Urine provides useful information on exposure (e.g., diet, exposure to hazardous agents), human metabolism, and kidney function. Many different analytic techniques can be applied to investigate urinary samples. Biomarkers in urine have been studied for purposes of diagnosis and monitoring of kidney or urogenital diseases (chronic kidney disease, diabetic or obstructive nephropathy, renal transplantation, kidney, and bladder or prostate cancer), but also for several systemic conditions (colon cancer, CVD, stem cell transplantation, etc.)⁷¹⁷⁻⁷²⁰.

Collection of different types of urine has been considered: spot urine, 24-h urine, and overnight urine. As it is more complicated to organize overnight and 24-h urine collection and provide little additional information, sampling of spot urine directly at the study center has been chosen as means of collection. Due to the immediate processing (centrifugation and freezing) of the samples, no additives are necessary for urine collection.

Saliva

The possibility to characterize the oral cavity microflora is the ultimate reason for collecting saliva. The effects of the oral microbiota on oral health and other chronic diseases can only be studied by investigating such data⁷²¹. Saliva also has the advantage that hormones and some other metabolites that may be compromised in blood samples as a result of the blood collection procedure can be measured. It is well documented that stress hormones and similarly sensitive biomarkers can be studied in saliva samples. Notably, the proteome of saliva is well characterized. The relationship between biomarkers in saliva and the development of oral cancer, oral diseases (periodontitis, caries, etc.), or systemic health (e.g., celiac disease, Sjögren's syndrome, and nonoral cancer) is of particular interest⁷²²⁻⁷²⁸.

Stool

Intestinal bacteria play an important role in human health, from maturation of the immune system to the onset and course of chronic diseases. By using fecal samples, the composition (and activity) of microbiota in the large bowel can be analyzed.

Gut bacteria are known to be involved in several diseases of the gut (e.g., irritable bowel syndrome; Crohn's disease, etc.); other chronic diseases are also likely to be modulated by the gut microflora⁷²⁹⁻⁷³¹. For example, obesity is characterized by changes in the occurrence of two main gut bacterial phyla, Firmicutes and Bacteroidetes. These changes can be reversed by nutritional intervention and are associated with changes in gut bacterial functions. Furthermore, intestinal bacteria have recently been shown to be involved in type 2 diabetes using antibiotics^{732, 733}. Specific aims of future studies in this area are likely to include measurement of the composition of the gut microbiota (phylogenetic fingerprinting and microarray analyses) and – if a suitable collection method can be applied – analyses of both metabolomic and physicochemical characteristics of fecal microbial cells/cell clusters⁷³⁴⁻⁷³⁶. Although there are indications that part of the gut microbiome can be profiled in urine⁷³⁷, the potential significance of the gut microbiota for future chronic disease research is considered to be very high and argues for the collection of fecal samples⁷³⁸.

Nasal swabs

The anterior nares (the space distal to the turbinates) represent an easily accessible space for sampling nasal epithelial cells, secretions, and microorganisms. It has been known for some time that the anterior nares represents a unique microbiologic niche at the interface between the respiratory tract, the skin, and the ambient air, which may be colonized by potentially pathogenic microorganisms, for instance Staphylococci (including MRSA, see also **Sect. A.2.4.7**) and Streptococci. Recent nucleic-based descriptive studies have shown that

the anterior nares harbor complex microbial populations with distinct interactions among species⁷³⁹. A nasal swab is a noninvasive, safe, and inexpensive method of sampling this unique microbial niche. We plan to primarily use culture-independent genomic tools to describe the nasal “microbiomes” and, analogously to the approach outlined for stool (see previous section), to test whether there are associations between the composition of these microbiomes (also considering expression of specific virulence factors, antimicrobial resistance factors, or superantigens) and the occurrence of acute and chronic diseases.

Other types of biosamples

Other sample types were also considered. The collection of induced sputum is time consuming, difficult to standardize, and does not provide important additional information in a substantial number of participants. Hair and nails were not expected to provide sufficient additional information to justify the costs of collection and storage. Subcutaneous fat and muscle biopsies were also considered; as biopsies are invasive methods and may have a negative impact on the participation rate, we decided not to perform them.

A.3.5.3 Concept for collection and preanalytic processing of biological samples

As a core program, 30 aliquots of serum, 48 aliquots of EDTA plasma, 6 aliquots of erythrocytes, white blood cells of 30 ml blood for later DNA extraction, and 12 aliquots of urine will be stored (**Table 3.12**). All aliquots will have a volume of 0.19 ml. In addition, RNA-stabilized blood, live peripheral blood MNC, as well as saliva, stool, and nasal swabs will be collected and stored. For reasons of quality (see above) the samples will be processed in the local study centers. An automated robotic system that will be installed at all recruitment sites will be used to prepare aliquots.

Table 3.12: Biomaterials to be collected in the National Cohort

Primary material	Volume	Processed material	Aliquots and storage at -180°C (unless stated otherwise)
Blood + clot activator	2x10 ml	Serum	30x0.19 ml
Blood + EDTA	2.0 ml	EDTA blood → hematology	Local lab., within <6 h
		EDTA plasma	48x0.19 ml
	3x10 ml	EDTA-packed erythrocytes	6x0.19 ml
		Buffy coat + 90% of red cell layer for later DNA extraction	1x9.0 ml (in manual – 80°C freezers)
Blood + RNase inhibitors	2.5 ml	RNA (Tempus/PAXgene)	1x2.5 ml (in manual – 80°C freezers)
Blood in BD CPT	10 ml	Ficoll-isolated PBMC + DMSO ¹	4x2 Mio PBMC (in liquid nitrogen vapor phase)
Total blood	65 ml		92 aliquots stored

Primary material	Volume	Processed material	Aliquots and storage at -180°C (unless stated otherwise)
Urine (supernatant)	2.4 ml	Small aliquot volumes following centrifugation	12x0.19 ml
Urine (supernatant)	50 ml	Large aliquot volumes following centrifugation	5x10 ml aliquots (in manual -80°C freezers)
Saliva, spitted	2 ml	No processing	2 aliquots (in manual -80°C freezers)
Stool		(To be determined) minimum processing, to be elaborated during feasibility study	2 aliquots (in manual -80°C freezers)
Nasal swabs		No processing	Swab tip (in manual -80°C freezers)

¹In feasibility studies, both absolute cell recovery and representation of the various mononuclear cell fractions using Becton Dickinson 'CPT' tubes will be assessed in comparison with conventional Ficoll-Hypaque gradient centrifugation of acid citrate dextrose (ACD) blood.

All steps of biosample collection, preanalytic handling, storage, and shipping to the central biobank will follow specific SOPs. In addition, the study personnel will be trained to ensure highest and identical quality across all study centers (see also **Sect. A.5** on quality assurance and quality control).

Blood

Blood will be obtained through venipuncture and the procedure carried out by qualified medical personnel. Even though venipuncture is invasive, it is generally well accepted and any associated risks can be considered as minimal.

- ▶ Overall, 65 ml blood will be obtained from each participant.
- ▶ Serum/plasma will be separated from blood cells as quickly and completely as possible to avoid the occurrence of any artifacts. Serum and plasma will be cooled down and aliquots prepared that will be stored in the freezer (-80°C) until transport to the central biobank or storage in the local biobank.
- ▶ For hematological analyses, 2 ml EDTA blood will be sent to a certified (clinical) laboratory.
- ▶ The buffy coat plus a substantial part of the red blood cell layer (from EDTA tubes) will be used for isolating peripheral mononuclear cells (PBMC) (after lysis of red blood cells). PBMC will be resuspended, aliquots prepared, and stored for later DNA extraction.
- ▶ Six aliquots of 0.2 ml packed erythrocytes from EDTA-coated tubes will be stored.
- ▶ For each subject, one tube containing RNase inhibitors will be filled, handled according to the manufacturer's instructions, and stored for later RNA extraction. RNA tubes from PaxGene or Tempus will be used.
- ▶ In addition, viable cells will be prepared from a subgroup of participants by means of Becton Dickinson 'CPT' tubes. For further laboratory work-up, the CPT tubes will be sent to the laboratory of the central biobank.

A detailed workflow for preparation and storage of locally processed blood and urine samples has been established. However, some details of this concept are subject to a feasibility

study. For example, PAXgene and Tempus tubes will be compared for quantitative RNA recovery and RNA quality. Both absolute cell recovery and representation of the various mononuclear cell fractions using Becton Dickinson 'CPT' tubes will be assessed in comparison to conventional Ficoll-Hypaque gradient centrifugation of ACD blood.

Urine

- ▶ Spot urine will be collected in 100-ml collection tubes (without additives). After centrifugation, the supernatant will be cooled in ice water, and aliquots will be prepared by using the liquid handling platform.
- ▶ The pellet and five 10-ml aliquots will be stored separately; a final decision regarding this, however, is pending.

Saliva

- ▶ Saliva will be collected using SaliCap collection tubes (IBL International) with the aim of obtaining a final saliva volume of 2 ml. Two aliquots of saliva will be stored.

Stool

- ▶ Study participants will be provided with the stool collection tube (and supplemental material) beforehand, along with a description of how and when the stool sample will be selected and how the tube will be handled before transport to the study center (at visit). Each participant will be asked for one stool sample, at best from stool produced in the morning of the day of the study center visit. Two aliquots of stool will be stored.

SOPs for collection and (minimum) processing of stool samples compatible with the constraints of field work in epidemiologic studies need to be definitively elaborated, and testing of sample quality necessary for high-throughput sequencing of microbial DNA and RNA will be conducted in the frame of a feasibility study.

Nasal swabs

- ▶ A trained study nurse will take a nasal swab using flocked nylon fiber sterile applicators (Copan Italia, Brescia, Italy). The tip of the swab will be frozen at the study center and shipped frozen to the central biobank, where it will be stored at -80°C . The bio-banked swabs will be analyzed with culture-independent methods only.

Elaboration of SOPs for standardized collection, comparison between different types of swabs regarding their ease of collection, and their use in microbial/virological studies will be tested in the feasibility study.

A.3.5.4 Concept for transport, storage, and retrieval of biomaterials

Transfer of biosamples to central and local storage centers

Transportation will be required for the aliquots of serum, plasma, urine, blood, and the saliva, nasal swabs, and stool that are prepared ready-for-use in the local study centers. They will be transferred on dry ice to either local or central storage facilities. The biomaterials to be processed in the central laboratory must be transferred at ambient temperature (viable cells) or at 4°C . Professional transportation companies will be hired at defined intervals to transport the materials to the central and local biobanks.

Overview of storage of biomaterials

The collected biomaterials will be stored in three different ways:

- ▶ **Automated central biorepository at the Helmholtz Zentrum Muenchen (HMGU) in Munich/Neuherberg (-80°C)**

- ▶ **Manual working archive at the HMGU** in Munich/Neuherberg (vapor phase of liquid nitrogen, -180°C)
- ▶ **Manual local back-up archives** at a defined number of centers (vapor phase of liquid nitrogen, -180°C)

Of the collected plasma and urine samples, two thirds of the aliquots will be sent to the central biobank at the HMGU. For cost reasons only one sixth of the aliquots will be imported primarily into the automated biorepository (5x serum, 8x plasma, 1x erythrocytes, and 2x urine). The automated working archive will be fed from the manual working archive reservoir as soon as materials have proved to be "very important materials" (VIM). Thus, over time the automated working archive will be enriched with VIM.

The other third of the plasma and urine aliquots will be stored in local back-up archives.

Blood for later DNA extraction, RNA-stabilized blood, saliva samples, stool samples, and swabs will be stored centrally in manual -80°C freezers. Similarly, also large-volume urine aliquots and urine sediments will be stored in -80°C freezers.

Construction of a new building for central storage at HMGU

Centralized storage at HMGU of biomaterials obtained from all cohort participants throughout Germany requires a new building that accommodates

- ▶ The automated biorepository.
- ▶ The liquid nitrogen tanks for the additional blood and urine samples as well as the viable cell storage.
- ▶ The manual -80°C freezers for the additional materials (blood for later DNA extraction, RNA stabilized blood, stool, saliva, and nasal swabs).
- ▶ The laboratory required for the central processing of samples and for quality control. One laboratory task is the separation and cryopreservation of viable cells from BD CPT tubes sent by the study centers; Furthermore, on-demand extraction of DNA and RNA (with separate funding) will be offered.
- ▶ The space required for the central biosample data management.
- ▶ The offices of personnel running the store, the central laboratory, and the management of laboratory data.

The planning for the building has started and it will be finished in 2012. The automated archive (repository) will be installed in the middle of 2013. Interim storage will be performed in the vapor phase of liquid nitrogen tanks in an existing vacant building of HMGU.

Automated biorepository at HMGU (-80°C)

Storage and retrieval from the automated biorepository represents an indispensable prerequisite not only for storing and tracking large numbers of aliquots, but also, and even more so, for retrieving single aliquots. Manual access to single aliquots (as opposed to complete racks) would not only require an immense time expenditure but also create difficulties in controlling exposure to higher temperature of both the selected and adjacent aliquots and, hence, in the long run would result in the progressive deterioration of the store's contents and value.

The HMGU Biobanking Planning Group, along with an international "Advisory Committee" experienced in biobanking (from UK and Sweden), have already visited all major suppliers of -80°C automated biorepositories [in alphabetical order: Fraunhofer (Saarland), Hamilton (Switzerland), Liconic (Liechtenstein), REMP-Nexus (Switzerland), TAP (UK), and the two

large-size stores already in operation (REMP store at Pfizer, Groton, Connecticut, USA; TAP store at UK Biobank in Manchester, UK). A European call for tenders will be launched for final selection.

For storage in the automated repository (and beyond), 0.2-ml Micronic tubes equipped with a 2D barcode and stable in the vapor phase of liquid nitrogen will be used.

Manual centrally operating archive at HMGU (–180°C)

Half (three-sixths) of the sample aliquots will be stored in a manually accessible working archive in the vapor phase of liquid nitrogen tanks (per participant: 15x serum, 24x plasma, 3x erythrocytes, and 6x urine), located adjacent to the automated repository.

A.3

The selection of one out of two liquid nitrogen storage tank systems will be made on the basis of a call for tenders in Europe.

Once a participant has proven to be of special interest for scientific research, i.e., due to disease occurrence, two-sixths of his/her aliquots will additionally be transferred to the automated biorepository, thus, in the long term, enriching the biorepository with so-called VIM.

Cryopreserved viable cells will be stored exclusively in the vapor phase in liquid nitrogen tanks.

Manual local back-up archives (–180°C)

One third of the blood and urine samples will be stored in local centers, preferably in one center per cluster participating in the National Cohort. The aliquots will be stored in the vapor phase of liquid nitrogen tanks (per participant: 10x serum, 16x plasma, 2x erythrocytes, and 4x urine). It forms a back-up storage system for the central biobank.

Number of stored aliquots

The total number of aliquots stored centrally will amount to about 20 million over the whole course of the cohort study, including samples obtained at recruitment, at short-term reassessment (for calibration purposes), and at reassessment of all participants after 5 years. Another ~8 million aliquots will be stored in local back-up archives.

Laboratory information management system

All steps from blood extraction until storage and later retrieval of aliquots will be controlled by a laboratory information management system (LIMS) registering each single aliquot by a unique defined barcode and monitoring and recording each individual step in the laboratory processing and in the storage when linked to the store's own inventory and tracking system. Currently, the four companies LabVantage, LisaLIMS, StarLIMS, and Lablynx satisfy these requirements. A call for tenders in Europe will be launched to select the LIMS provider/system.

The LIMS system will be integrated into the overall data management system of the National Cohort.

Retrieval of biomaterials

After finishing the recruitment phase, researchers can apply for data and biomaterials of the National Cohort in the context of a scientific project (see also **Figure 3.3**). After a positive vote of the Use and Access Committee, the samples will be selected and processed as agreed. Researchers must finance the transport, the extraction of DNA and RNA, and molecular analyses of the samples from their own budget.

Suggested site for the central biobank (HMGU, Munich)

Helmholtz Center Munich (HMGU) has a long-lasting experience in biobanking. Currently, approx. 1.6 million plasma and serum aliquots from 60,000 subjects are stored, as well as 15,000 DNA samples, 6,000 RNA samples, and 4,000 cell lines. During the last 3 years, about biosamples for 150 studies have been retrieved for collaborative projects.

HMGU is coordinating the “Biobank Alliance” as part of the “m4 Excellence Cluster” funded by BMBF. H.-E. Wichmann is German coordinator of the European Biobank Research Initiative BBMRI. In Munich the BBMRI Catalogue of 280 biobanks registered in BBMRI is hosted.

In addition, HMGU has broad experience in extraction of DNA, genotyping, transcriptomics, metabolomics, and (beginning) epigenomics.

A.3.5.5 Development of a tumor biobank for histologic and molecular tumor subclassification

Concept for tumor tissue biobank

As discussed in **Sect. A.2**, tumors from many organ sites exhibit different molecular subtypes, and it is increasingly being recognized that subtypes of tumors develop along different pathways and show different epidemiologic risk factors. Thus, for future research on the risk factors and etiology of cancer, it will be of utmost importance to have detailed knowledge about the morphologic and molecular characteristics of the tumors from all cases with the most frequent forms of incident cancer within the National Cohort. For morphologic and molecular subclassification of tumor types within the National Cohort, we propose that a quality-assured central biorepository of tumor samples be installed at the Pathology Institute of the University of Heidelberg in collaboration with the National Center for Tumor Diseases (NCT) in Heidelberg. After the paraffin-embedded tumor blocks have been collected to the central repository, relevant derivatives of these will be prepared, such as sections and tissue microarrays for immunohistochemistry, as well as nuclear acid extracts for analyzing epigenetic markers and genetic mutation profiles.

The collection and subsequent analyses of tumor tissues of incident cancer cases, in conjunction with the extensive risk factor information and further measurements that can be made in blood and other biomaterials, will make it possible to perform statistical analyses addressing cancer etiology that are of unsurpassed depth. In addition, the systematic collection of tumor materials within the National Cohort will provide a systematic description of molecular tumor characteristics for an epidemiologically defined population across Germany.

Requirements for tumor collection

Tissue: For diagnosis (biopsy) and during surgery (resection) tumor tissues generally are obtained and sent to the respective pathological institutes, where the samples are then

routinely prepared for diagnostic evaluation. Prepared in this way, tissue samples remain stable for a long time (in principle for periods up to 100 years) and can be used further for histological and molecular analyses. These formalin-fixed, paraffin-embedded tumor tissues represent the basis for preparing histological slides.

Generally, tumor materials are not used up by the diagnostic process and remain in storage for several years at the respective pathology institutes, until they are discarded because of space limitations and financial restrictions. Thus, most diagnostic archives are relatively short-term and, in addition, they do not generally have common standards and rules for access to materials. For future research based on the tumor tissues within the National Cohort it is, therefore, essential to establish a central tumor tissue bank that provides quality-assured storage, documentation, assessment, and release of these materials for research purposes.

A.3

The general collection of fresh-frozen tumor samples by the National Cohort is neither feasible nor is it necessary for the majority of research questions. Owing to various recent technical innovations, the vast majority of relevant molecular analyses can be performed with formalin-fixed, paraffin-embedded tumor tissues. Largely unrestricted analyses will be possible in the areas of DNA analysis (genomic imbalances, translocations, mutations), epigenetics (methylation assays), transcript (RNA expression) analysis, and analysis of non-coding RNAs, and histology (including special stains and immunohistology).

From formalin-fixed, paraffin-embedded tumor tissues, standardized derivatives such as nucleic acids, protein extracts, and multitissue arrays can be produced for further analyses in the context of well-planned, nested projects, making efficient use of the limited tumor materials available. For reasons of efficiency, quality control, and subsequent archiving and distribution, these derivatives will be produced in the context of the central tumor tissue bank. Furthermore, the use of tissue materials for various laboratory analyses generally requires that the material be re-examined by pathology experts, and this, too, should be performed at an institution that can provide the relevant histopathological expertise.

Suggested site for the tumor biobank (NCT, Heidelberg)

The existing infrastructures of the tissue biobank of NCT in Heidelberg will provide optimal support for the tissue biobank of the National Cohort for the following reasons:

- ▶ Ethical and legal situation: The tissue biobank of NCT in Heidelberg has received broad ethical approval for prospective tissue collection and retrospective use of previously collected samples for research purposes.
- ▶ Project management and experience: With more than 620 completed projects, the NCT tissue biobank has the highest project experience in centralized tumor banking in central Europe. Active project management in standardized project tracking and specific experience in supporting large population-based epidemiologic studies are available.
- ▶ Technical standards: The tissue biobank of NCT in Heidelberg offers a broad technology platform (all relevant histological methods, multitissue array production, nucleic acid technology, microdissection, virtual microscopy, and image analysis). Especially tissue microarray technology joined with virtual microscopy is of great use for managing decentralized tissue-based projects and was implemented successfully in the past.
- ▶ Quality management: The tissue biobank at NCT has implemented a comprehensive quality management system and is the only accredited tissue biobank in central Europe. All processes are guided by SOPs and are monitored by internal and external audits. The tissue biobank is supervised twice annually by its advisory board, annually by NCT, and every 3 years by an international review board.

- ▶ Sustainability: The tissue biobank has a regular budget that is provided by NCT and finances all its basic functions and infrastructure.
- ▶ Networking: The tissue biobank at NCT is founder and coordinator of the German Working Group Tissue Banks of the Cancer Centers, a forum which is funded by the German Cancer Aid.

A.3.6 Reassessment of intermediate (preclinical) phenotypes, measurements of function, and risk factors

For the National Cohort, all cohort participants will be reinvited 5 years after initial enrollment for a second examination. This second assessment will comprise a large part of the questionnaires used at the baseline visit and the collection of blood, urine, saliva, nasal swabs, and fecal samples and a reassessment of physical and medical examinations. In general, the physical and medical examinations in the reassessment program of the National Cohort will follow study protocols identical to those used at recruitment.

The re-examination of study participants after 5 years will serve two related objectives:

1. To assess longitudinal changes in functions and intermediate phenotypes
2. To assess changes in risk factors for disease (including risk factors that may be determined at a later date in collected biomaterials)

In addition to the 5-year reinvitation of the full study cohort, representative subsets of cohort participants will be asked after each of the two main visits to take part in short-term reliability / "calibration" substudies as well, within a time interval of maximally 1 year. These additional short-term studies will serve a third objective, namely, estimation of within-subject variations in risk factors that, in the context of their anticipated longer-term effects on chronic disease risks, can be considered as "random" errors.

These three objectives and some aspects of practical implementation of the 5-year reassessments are discussed further in **Sects. A.3.6.1 to A.3.6.3**, below.

A.3.6.1 Assessments of longitudinal changes in intermediate (preclinical) phenotypes and measurements of function

The first objective – to assess changes in function and intermediate phenotypes – represents an important means of assessing longitudinal functional decline or development of clinical disease. This, in turn, will then also allow us to conduct studies on the possible determinants of these longitudinal changes. Major examples of such longitudinal changes are decline in cognitive function, considered in conjunction with possibly related changes in brain structure on MRI scans, the new occurrence of cardiac arrhythmias as read from long-term ECG, or changes in musculoskeletal examinations (also possibly in conjunction with MRI) indicative of developing inflammatory joint disease. A full list of functional and (sub-) clinical assessments that will be reassessed after 5 years is given in **Table 3.13**. Neurologic and psychiatric changes will be evaluated by using specific questionnaires that produce scores.

In practice, it is expected that at least 75% of all participants in the first (baseline) visit will also agree to participate in the second visit, 5 years later. This estimate is based on the experience of existing cohort studies in Germany, including KORA, SHIP, and the Heinz Nixdorf RECALL studies, where participation rates on repeat visits have been systematically above 75%^{522, 740, 741}. Furthermore, in the two German EPIC cohorts, in Heidelberg and Potsdam, an ongoing re-examination of a random sample of study participants after an average of 12 years of follow-up shows a success rate of about 70%.

Subjects who move out of a region will be offered a second study visit as close as possible to their new place of residence. Experience with ongoing studies such as the EPIC cohorts show that, at least among subjects above age 40, only few study participants are likely to move far away from their original place of residence. Thus, we expect the vast majority of participants to be available for re-examination at their initial study center, at least in these older age groups.

Table 3.13: Reassessment of intermediate (preclinical) phenotypes and measurements of function in the National Cohort

	Measurement		
	At level 1	At level 2	Specific subgroups
Cardiovascular diseases			
Atherosclerosis	Arterial stiffness, ankle-brachial index, etc.	Ultrasonography of carotid artery, IMT	MRI, 3D TOF Angiography
Cardiac dysfunction	Echocardiography 10-sec ECG	Long-term ECG	MRI
Elevated blood pressure	Blood pressure		
Diabetes			
Impaired fasting glucose Impaired glucose tolerance		Fasting glucose OGTT	
AGE accumulation in the skin Retinopathy		Skin AF Retinal photography	
Hepatic steatosis, visceral fat			MRI
Cancer			
Precursor stages of hematologic malignancies	(blood-based analyses)		
Neurologic and psychiatric diseases			
Mild cognitive impairment	Cognitive tests		
Brain structures			MRI
Olfactory function	Smell test		
Questionnaires	CES-D, ASI-3, Headache, RLS questionnaires		
Respiratory diseases			
Lung function	Spirometry		MRI (lung volume)
Airway inflammation		Exhaled NO	
Infectious diseases			
Respiratory infections	Questionnaire		
Gastroenteric infections	Questionnaire		Active surveillance; stool samples by mail
Periodontal disease	Questionnaire, Tooth count	Periodontal examination	
Musculoskeletal diseases			
Osteoarthritis		Musculoskeletal examination	MRI
Rheumatoid arthritis (RA)		Musculoskeletal examination Biomarkers	MRI
Musculoskeletal pain disorders		Pain mannequin	
Osteoporosis	Dual-energy X-ray absorptiometry (DXA)		

A.3

A.3.6.2 Assessment of changes in risk factors for disease (including risk factors that may be determined at a later date in collected biomaterials).

The second objective – to assess possible changes in risk factors for disease – will make it possible to take such possible systematic changes into account, and will thus provide more accurate estimations of RRs for subsequent disease. A list of risk factors that will be reassessed and obtained by questionnaire after 5 years is given in **Table 3.14**.

Table 3.14: Reassessment and questionnaires of changes in risk factors for diseases in the National Cohort

	Measurement		Questionnaire
	At level 1	At level 2	
Physical activity	7-day accelerometry		X
Physical fitness	Hand grip strength	Step test Posturometry	
Body composition	Weight and height Waist and hip circumference DXA		
Diet	Collection of biomaterials		X
Smoking and alcohol consumption			X
Psychosocial factors			X
SES			X
Sleep			X
Chronic infections, immune factors:			
Immunization/infection status	Collection of biomaterials		X
Immune senescence and dysfunction	Peripheral blood lymphocytes		
Microbial colonization	Collection of biomaterials		
Zoonotic infections	Collection of biomaterials		X
Occupational and environmental exposures			X

In only few large-scale prospective studies have data been collected to account for possible systematic changes in an individual's risk factors over the longer term and for the effects that such changes may have on subsequent risks of chronic diseases. Failure to account for medium-term variability and systematic changes in risk factors can produce considerably biased relative and attributable risk estimates for disease, and thus an inaccurate quantitative evaluation of the importance of specific risk factors. For example, in prospective studies that have spanned time periods of 20 years or more and which used repeat measurements at 5- to 10-year intervals, the association of blood pressure measurements with risk of CHD has been shown to decrease progressively as the time lag from exposure measurement until diagnosis increases^{742, 743}. In this type of situation, repeat measurements taken after a longer time interval will have predictive potential for disease risk independently of (i.e., even when adjusting for) the exposure measurements (and true exposure levels) at the earlier points, and variations in exposure measurements cannot be simply considered as random error that is nondifferential with respect to prospective disease outcomes^{743, 744}. In **Sect. A.6**, basic risk estimation models that may be used to integrate two risk factor measurements taken over a medium-term time interval are discussed further, accounting for possible systematic changes in risk that may be related to medium-term changes in risk factors⁷⁴².

A.3.6.3 Reliability / calibration substudies

As stated in the introduction of this section, a third objective for the National Cohort is to estimate short-term, within-subject “random” variations in risk factors in calibration substudies, too. The availability of such data will make it possible to correct for attenuation biases in RR estimates or other measures of association with disease that are caused by random misclassification of subjects by risk factor levels⁷⁴⁵.

To meet this third objective, replicate measurements over shorter time periods will be collected in validation/calibration subsets of cohort participants. With the availability of data on short-term variability, in addition to the systematic medium-term (5-year) reassessments of all cohort participants, we can combine the use of regression calibration approaches and longitudinal data analysis methods for repeat risk factor assessments over longer time periods, so as to obtain quantitatively more accurate estimates of RRs of disease associated with long-term exposures or risk factors⁷⁴² (see also **Sect. A.6**).

For the calibration substudies, only subjects who participated in the intensive (Level 2) study program will be invited, as otherwise the number of subjects for whom data are available would be too small for the intensified research program. For the first 5-year study period, we will reinvoke a total of 6,000 study subjects across all centers within 1 year after their first study visit. Within the second 5-year period, 4,000 subjects across all centers will be reinvited within 1 year after their 5-year re-examination. Motivations for the sample sizes of these calibration substudies in the first and second phases of the overall data collection are given in **Annex, Sect. C.2**.

Subjects for the calibration substudies will be sampled within strata of sex and 10-year age groups, proportionally to numbers of cohort participants within these strata. The invitation for the calibration substudies will follow within less than 12 months (on average 6 months) after their previous (first or second) study visit. For the participants of the calibration study, the complete examination program will be applied, with exception of the OGTT. Regarding interview and questionnaire modules, only those parts will be repeated that assess relevant risk factors that may show short-term, “random” variability.

A.3.7 Procedures for prospective follow-up of vital status and disease occurrences

A.3.7.1 Introduction

The value of the National Cohort depends not only on its ability to obtain comprehensive baseline data and biomaterials but also on detailed follow-up information about the vital status and disease occurrences in the participants. The aim of the follow-up is to obtain information on a regular basis about incident diseases of interest, changes in exposures and risk factors, incident symptoms, vital status, and cause of death during follow-up for each participant included at the baseline examination. A further aim is to maintain contact with the subjects over the whole study period so that at any point in time the participant’s address is known and he/she can be contacted, if necessary.

By signing the informed consent form prior to the baseline assessment, participants will give their permission to be contacted again during the follow-up period at regular intervals in order to provide information about changes in lifestyle or other risk factors and incident diseases, to confirm that information can be obtained from medical files and registries, and to confirm that both their physician and health insurance provider can be contacted for further information on diseases and medical management, along with the necessary contact information (names of physicians, their speciality, health insurance of the participant, etc.).

A.3.7.2 Recontacting and tracking of study participants

Personal identifying information such as name, date of birth, and address will be obtained prior to the invitation process at baseline, will be checked at the study center during recruitment, and recorded in a separate file. If not yet available, further information such as the participants' cellular telephone numbers and e-mail addresses will be collected. This information may help to ensure that participants are not lost during follow-up.

During the follow-up period, participants will be informed about interesting developments and results of the National Cohort by regular newsletters. These newsletters can also advise participants to communicate any changes in contact details actively to the study center.

Every 2–3 years after the first visit in the study center, participants will be contacted by mail or e-mail, including a follow-up questionnaire. For subjects who do not respond to the initial follow-up questionnaire, a reminder letter will be sent. If a contact cannot be established by these means, participants will be, where applicable, contacted by phone and a computer-assisted telephone interview will be offered at the same time or a later time scheduled. In addition, if the subjects do not (for various reasons) want to answer the questionnaire, a short version, assessing the health status related to the main diseases of interest, will be asked, to enable the comparison of a few core characteristics of responders and nonresponders.

For all study subjects who do not respond to the follow-up procedure invitations (undelivered mail/e-mail, no contact or response) an inquiry at the residents' registration office will be initiated that clarifies the vital status of the subject and whether subjects have moved away (all residents in Germany are required to provide their the new address once they move away).

In case of death, as the registries record the date and place of death, the date of death (population registries) and cause of death (public health office) will be obtained.

For subjects who have moved, the new address will be obtained from the original registration office.

In the study database, the tracking status for each study subject will therefore be regularly updated. The most likely status is "full participant" (has sent back a completed questionnaire). The other final status categories are "dead", "final drop out" (has left the study) or "permanently unavailable" (e.g., no permanent residence, has moved to another country).

A.3.7.3 Follow-up procedures for ascertainment of vital status and incident diseases

Prospective ascertainment of incident diseases will primarily be based on *active follow-up procedures*. These include the assessment of novel onset disease by means of questionnaires, which will be verified, for defined diseases, by contacting the participant's pertinent physician(s).

Questionnaires will be sent to the participants every 2–3 years (see above) to request information about any major, physician-diagnosed diseases that have occurred since the baseline visit, including the treating physician and the date of diagnosis (see **Table 3.15**).

For certain diseases, we will also use additional information obtained via questionnaire from the participant to increase the sensitivity of disease ascertainment (**Table 3.15**). For example, to increase sensitivity to detect incident cases of type 2 diabetes, we will also consider information on new onset of antidiabetic treatment.

In addition, as many subjects are not able to give details about newly diagnosed diseases (e.g., invasive breast cancer versus in situ breast cancer; angina pectoris versus MI; stenting without prior MI versus stenting because of acute MI, etc.), the questionnaires will query diagnosis and treatment circumstances that can indicate incident outcomes of interest as closely as possible. For example, data from subjects who were admitted to hospitals for a cardiologic diagnostic check-up or treatment or for a cancer check-up will be validated by hospital chart review, etc.

In a second step, selected self-reported incident diseases (and dates of diagnosis) will be systematically **verified by contacting the participant's pertinent physician(s)** who made the diagnosis (i.e., general practitioners, specialists, or hospitals/clinics, respectively). Only verified cases will finally be considered for analyses, so as to ensure high specificity.

A.3

For certain diseases, we will also obtain more specific medical information from the physician or use organized access to patients' clinical records and/or pathology records for direct information retrieval, e.g., for further classification into specific subtypes of disease. For example, information from patients' clinical records will be used to classify cases of stroke into specific subtypes (ischemic versus hemorrhagic). Likewise, information from clinical records and pathology records will be used for the specific coding of histologic subtypes of cancer.

In addition to active follow-up, we will make use of **linkage with external data sources** for detecting and verifying incident diseases. This is particularly relevant in the field of cancer, where **epidemiologic cancer registries** (with varying degrees of completeness of coverage) have now been established in every state in Germany. These registries will also be used to ascertain incident disease. Supporting epidemiologic research in the field of cancer etiology is among the primary aims of cancer registries in Germany. For many cases of cancer and for many cancer types, the use of cancer registry data will reduce the workload for further active verification of clinical and/or retrieval of pathology records, and thus will both facilitate the follow-up process and allow for major cost savings. Several successful examples exist in Germany to support the aforementioned strategy to ascertain incident disease. Similar strategies have been used in the Heinz Nixdorf Recall Study, CARLA, SHIP, EPIC, and KORA studies^{746, 747}.

Information on vital status will be collected from population registries at regular intervals for the participants who have not taken part in the follow-up or reassessment procedures. For subjects who have died, the **death certificate** that provides information on the cause of death will be obtained from the public health offices. A National Mortality Registry is presently being planned in Germany. As this national registry is expected to become progressively operational and increase its coverage – which will probably be not the case before year 3 of the study – the procedure described above will be gradually substituted by data retrieval from this registry following a validation phase with parallel ascertainment of approximately another 2 years.

The use of secondary data from the health system as described in **Sect. A.3.8** will be tested for its usability to support the follow-up process. Individual linkage to secondary data has a significant potential as an additional source of follow-up information⁷⁴⁸.

For assessment of prevalent diseases at baseline we plan to use similar strategies, in conjunction with the results from physical and medical examinations conducted as part of the baseline recruitment protocol.

For ascertainment of dementia, depression, headache, and RLS, scores based on questionnaires will be established that can be used to supplement the self-reported data on physician-diagnosed diseases. These scores are described in more detail in **Sect. A.2.2** and **A.2.3**.

Table 3.15: Assessment of incident diseases in the National Cohort

Disease	Question to participant (Physician-diagnosed disease, date of diagnosis)	Additional information from participant considered to support self-report (and increase sensitivity)	Physicians' diagnosis (including diagnosis, date of diagnosis, interventions) subtypes	Passive follow-up	Death certificate
Cardiovascular diseases*					
CHD: MI	X	PTCA, Stent, CABG Angina pectoris	X	MI registries (where available)	X
Heart Failure	X	NYHA stage	X, systolic or diastolic heart failure		X
Atrial fibrillation	X (Arrhythmia)	Medication (anticoagulants)	X		X
Diabetes mellitus	X	Medication (insulin, oral antidiabetics)	Type 1 or 2		
Cancer					
Breast, prostate, lung, colorectum endometrium, ovarian, others	X		Grade and stage, molecular subtypes, prognostic markers	Cancer registries	X
Neurological and psychiatric diseases*					
Cerebrovascular disease: stroke	X		Ischemic, haemorrhagic	Stroke registries (where available)	X
Dementia	(X)	next-of-kin, admission to nursing homes; nonresponder questionnaire	X		X
Parkinson's Disease	(X)	next-of-kin, admission to nursing homes; nonresponder questionnaire	X		
Depression	X		X		
Headache Migraine, tension- type headache	X	Symptoms	X, subtype		
RLS	X		X		

Disease	Question to participant (Physician-diagnosed disease, date of diagnosis)	Additional information from participant considered to support self-report (and increase sensitivity)	Physicians' diagnosis (including diagnosis, date of diagnosis, interventions) subtypes	Passive follow-up	Death certificate
Respiratory diseases					
COPD	X (COPD/Chronic bronchitis)	Questionnaire on specific symptoms, medication (bronchodilators, inhalative glucocorticoids)	X, small airway disease, emphysema, medication		X
Asthma	X	Questionnaire on specific symptoms, medication (bronchodilators, inhalative glucocorticoids)	X, allergic, nonallergic		
Infectious diseases					
Respiratory infections Gastrointestinal infections Vaccine preventable diseases, Immune senescence	(Questionnaire) Active surveillance**		X Immune status, lab results, differential diagnoses		(X)

* Optional use of information from assessment modules where appropriate, e.g. ECG information for the diagnosis of MI

** See also **Sect. A.2.2.6**

A.3.7.4 Prospective assessment of changes in exposures/risk factors

The repeated contact to the study participants by mail to ascertain incident disease will also be used to update information about exposures/risk factors. This includes information on changes in lifestyle such as smoking behavior, physical activity, etc. In the National Cohort we will evaluate whether some of these factors can be assessed by internet-based instruments that cover the habits during the previous day (e.g., physical activity, diet) or events during the past weeks or months (acute infectious diseases such as respiratory infections and gastroenteritis).

A.3.7.5 Endpoint committees

Standard criteria to verify endpoints of interest will be used. Data from subjects with endpoint information that appears inconsistent, implausible, or arbitrary will be reviewed by a specialized endpoint criteria committee that includes clinical experts and various specialists such as pathologists, radiologists, and others. As the National Cohort study includes several medically diverse endpoints, several committees will be set up. These committees are:

- ▶ CHD endpoint committee
- ▶ Cancer endpoint committee
- ▶ Stroke endpoint committee

As the cohort is very large, the workload of the endpoint committees will become too large if every event is reviewed by the committees. Based on experience from the Heinz Nixdorf Recall Study⁷⁴⁶, events can be distinguished in:

- a) Clearly documented, unambiguous events
- b) Documented events that cannot be clearly classified by the local study centers

Only events of category b) will be reviewed by the endpoint committees. In addition, a random sample of events from category a) will be submitted to the endpoint committees in order to control the quality of categorization. The committees meet 1–2 times per year for reviewing ambiguous endpoints.

For all other endpoints dedicated expert panels will be established that assess the validity of the diagnoses based on a random sample.

A.3.8 Use of secondary data sources

A.3.8.1 Rationale

In addition to the primary data obtained by means of the various methods described in this application, we will also make use of secondary data as complementary information concerning exposure and disease status of the participants (i.e., for *etiological questions*). In addition, such data may provide valuable information about health care utilization of the study subjects (i.e., testing of hypotheses in the fields of *health service research*, health-related quality of life research, and *health economy*). Secondary data will be primarily obtained from statutory health insurances and statutory pension insurances (i.e., *social security data*) and from environmental surveillance systems (i.e., *environmental data*).

A.3.8.2 Social security data

Social security data may provide valuable information concerning *health care services*. The utilization, quality, and perception of health services may be important predictors for morbidity. Amount, manner, and specific parameters of the use of medical services for patients with chronic diseases or with multimorbidity can be seen as a risk factor if utilization of medical services (health care overuse, misuse, and underuse) or if access or availability, of care is suboptimal. In addition, social security data on *occupational history* can contribute in multivariate models analyzing risk factors for certain diseases. Depending upon the specific research question, secondary data can thus reflect either a determinant or an endpoint of interest.

About 85% of the population in Germany is insured by the statutory health insurance system. Apart from civil servants and self-employed individuals, most employees are also subject to the statutory social insurance contributions to unemployment, to long-term care insurance, and to the statutory pension insurance scheme. Personal data related to the use of health services (outpatient and hospital) and the occupational history (job codes) are routinely collected and stored over long time periods on a legal basis for a majority of the population in Germany.

Data routinely collected as part of the statutory health insurance system

The following data that are routinely collected as part of the statutory health insurance system and are of primary interest with regard to the National Cohort: (1) Outpatient medical care; (2) inpatient medical care; and (3) prescribed medication.

Outpatient contacts to general practitioners and medical specialists are regularly documented on a quarterly basis and include information on date of each contact, diagnoses, and services according to the EBM code (“Einheitlicher Bewertungsmaßstab”). More than 90 % of the population has at least one documented contact per year. On average more than ten different ICD codes (acute or chronic diseases) are documented per patient and per year. Another parameter of interest for the National Cohort may be the utilization of screening data (e.g., cancer screening).

A.3

In-patient stays are documented with information on time, diagnoses (admission, discharge, and claims diagnoses) as well as information about the treatment, coded via the “Operationen- und Prozeduren-Schlüssel (OPS)”, the German-language modification of the International Classification of Procedures in Medicine. With this tool events leading to hospitalization and many clinical interventions can be identified, e.g., hysterectomy, hip replacements, coronary bypass grafts, or PTCA.

Medication qualifying for reimbursement is recorded by date dispensed, number of packages dispensed, and the specific ID for each dispensed product (‘Pharmazentralnummer’). Therefore, product-based information such as the Anatomical Therapeutic Chemical Classification System (ATC), the pharmacological strength, defined daily doses (DDD), contraindications, and interactions can be unequivocally assigned.

The aforementioned data must be sent by providers in strictly standardized formats and are held by the various statutory health insurances together with the health insurance ID of the insured person.

In addition to the aforementioned data, total **health-related spending** of the insurance and **work inability data** are documented at a person-level by health insurances and could possibly be used for the National Cohort. As we expect that the organization of the health care system will evolve in the course of the cohort, adaptations regarding changes in legal regulations and access to additional data will be considered.

Data routinely collected as part of the German statutory pension insurance scheme and administrative data of the Federal Employment Agency / Institute for Employment Research

Membership in the statutory pension insurance system is compulsory in Germany for more than 95 % of the German workforce. While the main focus of the pension insurance is on old-age retirement, the system also covers work disability pensions. These pensions apply in the event of early retirement due to permanent severe reduction of an employee’s ability to work. The German statutory pension insurance provides access to claims data at the individual level. A major portion of rehabilitation measures are also covered by pension insurance rather than health insurance. Thus, these data are relevant for assessing efforts to keep people (with chronic disease) in work, given the high importance of an active working status for a person’s health and well-being.

Most of the service claims in the German social insurance system are related to labor force participation and income of insurance members. Data are provided by employers at least annually and transferred by the health insurance to the pension and unemployment insurances. Data include **sociodemographic information** of each employee as well their **occupational code** and the **economic sector of employment**, both classified according

to standard classifications of the Federal Office of Statistics. However, for reasons of data confidentiality, only certain institutions are given access to these data. The Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung) is part of the Federal Employment Agency and holds what is known as an "Integrated Employment Biography". Here, individuals who either have been employed subject to social security, have received benefits according to the German Social Code Book II or III, participated in programs of active labor market policies, or have officially been registered as job-seekers at the German Federal Employment Agency are covered for the periods 1975 to 2008 and 2000 to 2008 in formerly West and East Germany, respectively. Therefore, the data provide a life-course follow-up of work and unemployment/job-seeking histories. A former version of that file has been used, for example, to create occupational biographies and estimate risks for early retirement⁷⁴⁹.

All access to secondary data will be based on individual informed consent. Only pseudonymized data will be used in the analyses.

A.3.8.3 Environmental data

Several sources for secondary data on environmental hazards are already available. More and spatially more refined data will be generated in the future. We envision combining publicly available databases with research-based databases and developing integrated model predictions. For the National Cohort, the following data will be of specific interest:

- ▶ **Noise maps** are mandatory in cities with inhabitants of more than 200,000 individuals under a European directive (http://europa.eu/legislation_summaries/environment/noise_pollution/l21180_en.htm). Complete coverage and refinements of exposure from all sources of traffic are envisioned for the study areas included in the National Cohort.
- ▶ Spatially resolved **weather maps** (http://www.wetterpool.de/wetterkarte_temperatur.php) will enable us to characterize increased chronic heat stress during summers or exposure to weather extremes throughout the seasons.
- ▶ Spatially resolved maps on natural **radiation** (environmental radioactivity based on the integrated measurement and information system (IMIS) <http://www.bfs.de/ion/imis> and, indoor and outdoor radon

http://www.bfs.de/de/ion/radon/radon_boden/radonkarte.html) are currently available from the BfS. Other parameters (e.g., distance from high-voltage power lines) will be made available in the future.

A.3.8.4 Strengths and possible limitations of secondary data use for the National Cohort

How secondary data can enhance the scientific value of the National Cohort can be summarized as follows:

1. **Reduce misclassification and complete data:** Individual population-based data on health and utilization of health care, employment, and unemployment are collected continuously by statutory health insurances and statutory pension insurances according to legal standards and requirements. Similarly, data from environmental surveillance systems are collected continuously on a routine basis in Germany. As data acquisition is close in time and independent of individual constraints, this data source is largely independent of self-reports and thus may be less susceptible to errors in reporting. Use of secondary data may therefore help reduce differential or nondifferen-

tial misclassification, which may decrease systematic and random error in data analysis. In addition, secondary data use may help complete data collection in instances where primary data from the participant may not be (or no longer are) available, e.g., for deceased study subjects.

2. **Support follow-up procedures:** Secondary data may support the active and passive follow-up procedures described in **Sect. A.3.7**. Since these data are largely independent of self-report of the participants, their use has considerable potential to improve completeness and validity of the traditional follow-up procedures. They may also help in studying "softer" outcomes that are difficult to assess via active or passive follow-up procedures.
3. **Provide alternative means of acquiring occupational data:** Access to data of the Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung) represents an alternative means of acquiring occupational data on an individual basis and thus will reduce the burden for study participants and interviewers.
4. **Completion of information on individual medications:** All reimbursed medications are documented with prescription date, ATC code, and prescription dosage by statutory health insurances. Using these data, medication histories can be assessed over extended time periods, which can augment and improve the information provided by the participant.
5. **Information on individual health care utilization:** Both operations and procedures during a hospital stay and most interventions performed by the treating physician are documented by statutory health insurances. These are valuable data that are difficult to collect via a personal interview. The information includes utilization of screening examinations, surgical procedures, etc.
6. **Evaluate health economics:** Since statutory health insurances in Germany cover a broad range of medical services, most of the direct health-related costs are documented (as are some of the indirect costs). By using such data we can estimate diagnosis-associated costs and carry out health economics analyses in Germany.
7. **Information on the physical environment:** Data collected at the federal, state, and city level provides a multifaceted characterization of the physical environment in Germany and is widely used for city and regional planning. The approach proposed here provides the opportunity to comprehensively assess multiexposure scenarios at the place of residence or work with relatively high precision at comparably low cost and provides scientific evidence on the effectiveness of ongoing abatement measures.

One major limitation to the use of secondary data is that data may not be available for all cohort participants, depending on the specific insurance companies that participants adhere to and on whether participants will allow extraction and linkage of their insurance data to the National Cohort data base. Furthermore, the content and format of available data, e.g. on use of medications and health care, can vary considerably across different insurances and it may not always be possible to obtain uniform information for all cohort members. It is anticipated that health insurance data may be obtained for about 60-70 percent of cohort participants, but feasibility studies will need to be conducted to confirm this. Finally, it is likely that the availability of health insurance data, in particular, will be associated with socio-economic indicators, because the data from private insurances, which generally can be adhered to only by higher-income groups, are not as easily accessible as those from statutory insurances. Thus, it cannot be ruled out that overall health status or disease risks differ between subjects whose data can be retrieved, and the rest of the cohort. These, and possible further complexities will have to be properly addressed and accounted for in statistical analyses, so as to examine and avoid possible biases that could be induced by the selective availability of data.

A.3.8.5 Experiences with data linkage of secondary data in epidemiologic studies in Germany

Use of secondary data in cohort studies represents a rather new field in Germany, which, however, can be built on some existing experience. In the KORA region (Augsburg) in two health economics studies on allergic skin and respiratory diseases and on acute care of MI, claims data were linked to primary data from interviews and questionnaires. In these two studies (including 1,100 participants), 64% and 78% of insured subjects, respectively, agreed to pass on their individual health insurance data.

The health economic complementary evaluation of the Heinz Nixdorf Recall (Essen) study conducted a cost-effectiveness analysis of coronary calcium deposit screening. Individual data from the statutory health insurance were collected based on general individual consent of subjects covering all study modules. Informed consent was given by 90% of the participants. A similarly high response rate is reported by the study “Investigating work related determinants of psychological and physical health in an ageing work force” (lidA-Study, started in 2009) funded by the German Federal Ministry of Education and Research, which is examining the long-term effect of occupation on the aging work force. About 70% of the participants in the lidA pretest agreed to linking primary data with their health insurance data and their work history data.

A.3.8.6 Data quality and quality assurance of secondary data

Experience shows that it is mandatory to conduct a standardized assessment of the quality of secondary data sources. The Bremen Institute for Prevention Research and Social Medicine (BIPS) has linked and updated data from four statutory health insurances for a number of consecutive years. Routine plausibility checks aid in detecting incomplete data delivery by the statutory health insurances or delivery of nonconcordant data over different years caused, for example, by different definitions for data extraction in the statutory health insurances over time.

In addition to routine plausibility checks, it is important to further examine the validity of information obtained from secondary data sources. For example, data obtained from statutory health insurances can be validated by using different approaches: internal validation, comparison with external statistics to discover incomplete recording or “overcoding” of diagnoses, linkage with data from selected hospital information systems to validate diagnoses based on the results of diagnostic tests and procedures in hospital, and linkage with disease registries to validate diagnoses in statutory health insurance data against diagnoses in these registries. Trust in the validity of statutory health insurance data can further be built when known associations can be replicated in these data. For example, BIPS invites statutory health insurance members with a particular diagnosis to undergo a standardized assessment to either confirm or refute this diagnosis. The Bremen Institute for Prevention Research and Social Medicine has also participated in benchmarking the incidence of defined diagnoses by comparing several European databases. This has helped to identify errors in coding data at the national level^{748, 750}.

A.3.9 The National Cohort in the context of other cohorts of adults in Germany and internationally

Ongoing cohort studies in Germany

Currently in Germany a variety of prospective cohorts are being maintained by different institutions:

- ▶ The KORA cohort in Augsburg since 1984/85 (PI H.E. Wichmann, HMGU Munich) with 18,000 participants. <http://www.helmholtz-muenchen.de/kora>
- ▶ The EPIC cohort in Heidelberg since 1994 (PI R. Kaaks, DKFZ) with 25,500 participants http://www.dkfz.de/de/epidemiologie-krebserkrankungen/arbeitsgr/ernaerpi/ee_p01_epichd.html
- ▶ The “SHIP” Study in Greifswald since 1997 (PI H. Völzke, University of Greifswald) with 4308 participants (SHIP-0) and recently another 5000 (SHIP-TREND) participants; http://www.medizin.uni-greifswald.de/cm/fv/english/ship_en.html
- ▶ The EPIC cohort in Potsdam since 1998 (PI H. Boeing, DIfE), with 27,600 participants <http://www.dife.de/de/index.php?request=/de/forschung/projekte/epic.php>
- ▶ The “HEINZ NIXDORF RECALL” Study in Essen since 2000 (PIs R. Erbel and K.-H. Jöckel University of Duisburg -Essen) with 4,814 participants; <http://www.recall-studie.uni-essen.de/>
- ▶ The “CARLA” study in Halle since 2002 (PIs H. Greiser and J. Haerting, University of Halle) with 1,779 participants <http://www.medizin.uni-halle.de/imebi/index.php?cid=143>
- ▶ The DGS (Dortmunder Gesundheitsstudie)in Dortmund since 2003 (PI K. Berger, University of Münster) with 1,312 participants
- ▶ The popgen Study in Schleswig Holstein since 2003 (PIs M. Krawczak and S. Schreiber, University of Kiel) with 10,000 population controls, <http://www.popgen.de/>
- ▶ The Gutenberg Heart Study since 2007 (PI S. Blankenberg, University of Mainz) with 17,000 projected participants <http://www.herzstiftung-mainzer-herz.de/herzstiftung/forschung/gutenberg-herz-studie.html>

The existing German cohorts have varying thematic research emphases. The European EPIC study, which includes the two German subcohorts in Heidelberg and Potsdam, focuses on the relationships of nutrition, lifestyle, and metabolism with risks of cancer and other chronic diseases. The KORA, SHIP, Heinz Nixdorf Recall, and CARLA studies are smaller than the EPIC study, but include much more comprehensive subclinical and clinical phenotyping of cohort participants. The National Cohort will not only extend and complement these existing studies, but will address many additional research questions and form the basis for high-quality epidemiologic training in Germany over the next two to three decades.

Recently established European cohorts studies

As described above, the National Cohort will be large enough to accommodate powerful stand-alone studies on a variety of major chronic disease endpoints, including diabetes, CVD, specific forms of cancer, aging-related degenerative diseases, and long-term health outcomes of chronic infection. For rarer disease outcomes or for types of analyses (e.g., of gene–environment interactions) that may require very large numbers of observed disease outcomes, our strategy will be to conduct joint analyses for specific exposures and endpoints based on data from the National Cohort and those from other large-scale cohorts in Europe. These latter cohorts include, in particular, the UK Biobank study (500,000 study subjects, who have already been recruited; <http://www.ukbiobank.ac.uk/>), the Swedish “LifeGene” study (in planning phase; 300,000 participants projected, <https://www.lifegene.se/>), and the French “CONSTANCES” study (currently in the pilot phase; 200,000 cohort participants projected, <http://www.constances.fr/>). As some physical examinations and target outcomes in the National Cohort are similar to those included in the one or the other European cohort studies (see **Table 3.16**), the prerequisites for joint analyses are given.

Table 3.16 demonstrates the comprehensive program foreseen for the National Cohort in

comparison to other European large-scale studies. For many fields of investigation, the National Cohort will include a more advanced phenotyping, e.g., carotid ultrasound, long-term ECG in the area of cardiovascular examinations, or application of accelerometers in the entire cohort to precisely assess physical activity. In none of these studies are dental examinations and examinations of the musculoskeletal system included, and no study has applied whole-body MRI, although brain and cardiac MRI and MRI-based quantification of visceral fat volume are planned in the reassessment of a subgroup of the UK Biobank cohort.

Concerning biosamples, tubes containing RNase inhibitors, and thus suitable for RNA isolation, will be used for the entire cohort (**Table 3.17**). Furthermore, saliva, feces, and nasal swabs will be collected, biomaterials which are not collected in the other studies (except for saliva in the enhancement 2 program of the UK Biobank). In addition, the preparation of viable cells is foreseen for the intensively investigated group of 40,000 subjects.

A number of cohort studies are being planned in other German-speaking regions in Europe, including South-Tirol (Northern Italy), and regions in Austria, Luxemburg, and Switzerland. In these latter studies it is planned or being discussed to adhere to a study protocol that to a very large extent will be standardized with that of the National Cohort.

	The National Cohort	UK Biobank	LifeGene Sweden	CONSTANCES France
Anthropometry (incl. bone density)	Body height and body weight	X (basic)	X	X
	Waist and hip circumference	X (level 1)	X	X
	BIA	X (basic)	X	
	DXA [#]	(heel bone ultrasound, basic)		
Physical activity	Accelerometer (7 days)	X (enhancement 2)		
	Fitness test	X (enhancement 2)		
	Hand grip strength	X (basic)		X (senior)
	One-leg-stand test			X (senior)
	Ophthalmological measurements	X (level 2)		X (visual acuity)
Sense organ functions	Audiometric test	X (enhancement 2) (touch screen questionnaire)	X (audiometry)	X (audiometry)
	Olfactory tests [#]	X (level 2)		
	Whole-body MRI [#]	X (in 40.000)		
MRI	Cardiac MRI [#]	X (in 40.000)		
	Brain MRI [#]	X (in 40.000)		

[#] The National Cohort as the only cohort study including these measurements at baseline

Table 3.17 : Comparison of biosamples collected in recent European Cohort Studies with the ones planned for the National Cohort study

	The National Cohort			UK Biobank			LifeGene Sweden			CONSTANCES France		
	Vol	Number	Aliquots	Vol	Number	Aliquots	Vol	Number	Aliquots	Vol	Number	Aliquots
Blood												
EDTA												
Whole blood												
Plasma	10 ml	3	48x0.19 ml	9 ml	2	8x1.4 ml	4 ml	1	8x0.3 ml	10 ml	1	4x0.5 ml
Buffy coat						4 x1.4 ml						8x0.5 ml
Buffy coat +90% of red cell layer for DNA			1x9 ml									
Red blood cells						2x1.4 ml						
Packed erythrocytes#			6x0.19 ml									
Plasma							9 ml	2	16x0.3 ml	10 ml	1	8x0.5 ml
EDTA	2 ml	1	-	4 ml	1	-	4 ml	1	-			
Citrate 3.8%							4 ml	1	4x0.3 ml			
Li-Hep with gel (PST)				8 ml	1	4x1.4 ml	4 ml	1	8x0.3 ml	10 ml	1	8x0.5 ml
Clot activator (SST)	10 ml	2	30x0.19 ml	8 ml	1	4x1.4 ml				10 ml	1	4x0.5 ml
RNAse inhibitors#	2.5 ml	1	1x2.5 ml	2-3ml*		6x1 ml						
Acid citrate dextrose				6 ml	1	2x1.4 ml						
Blood in BD CPT	10 ml	4x2 Mio. MNC										
Trace metal tube							7 ml	1	8x0.3 ml			
Total blood	65 ml		84x0.19 ml	44 ml		24x1.4 ml	49 ml		44x0.3 ml	40 ml		32x0.5 ml
Urine												
Spot Urine	2.4 ml		12x0.19ml	9 ml	1	6x1.4 ml	9 ml	1	8x0.3 ml	10 ml	1	8x0.5ml
Urine supernatant	50 ml		5x10ml									
Saliva #	2 ml		2	2-4ml*		2x1 ml						
Feces #	X		2									
Nasal swabs #	X		X									

The National Cohort as the only cohort study collecting these biomaterials for all participants

* Enhancement 2

A.4 Integrated data management

A.4.1 Aims and requirements

The general aim of the common data management infrastructure is to support the National Cohort in all phases and parts by using professional IT. This includes the following:

- ▶ IT support of recruitment and management of participants and schedule management at the local study centers
- ▶ Efficient, standardized data collection both with manual input and using diagnostic devices; support of collecting, processing, storing, and managing biological samples
- ▶ Secure storage and archival of data according to data privacy protection rules and such that availability, integrity, and confidentiality of all data are guaranteed at any time
- ▶ Quality control of decentrally and centrally collected data (standardization, reliability, validity, completeness, and plausibility) by means of automatic quality control procedures as well as support of manual quality checks
- ▶ Integration of data collected at the study centers into a common data set, ensuring a high quality standard and on-demand availability and which supports efficient analysis
- ▶ Providing and controlling access to study data for use in internal and external research projects, and reintegrating data resulting from internal and external research projects in the central data base of the National Cohort.

Two different types of data must be distinguished: One is the personal identifying data of the study subjects; the other is the actual study data, i.e., all medical information collected from the subjects after they have given their consent to participate in the study. Personal identifying data are needed only for contacting study subjects during recruitment and follow-up, possibly for submitting incidental findings according to the regulations described in the ethics concept, and for requesting secondary data from external sources. Study data are needed for scientific analysis only.

The data for the National Cohort with its large number of recruiting study centers and other involved institutions can only be efficiently managed by standardizing data collection and processing. The implementation of a centralized data management infrastructure represents an important resource for all institutions that take part in this large, long-term cohort study. The central organization supports standardization of instruments and processes and optimization of data quality, guarantees high availability of all data, and enables study-wide common handling of data privacy protection regulations. The effort for data management of individual institutions is at the same time minimized through the use of common IT solutions.

Requirements for integrated data management comprise both general requirements for cohort studies and large-scale distributed systems and specific requirements resulting from the large number of institutions involved in the National Cohort:

- ▶ **Data quality and standardization:** Data quality is essential in order for the results obtained from scientific analysis of study data to be valid. Data quality management can be considered with regard to content (see **Sect. A.5**) and from a technical perspective. Technical components of data quality management, which must be provided by the data management infrastructure, are:
 - Consistent reference to the data collection instruments and their respective parts, structure, presentation, and resulting values throughout all data structures and applications
 - Automatic checks for completeness, plausibility, and consistency wherever possible

- Transparent documentation of the history of each data item in the system (documentation of modified values, date, reason, and executing person for any modification of any record) prospectively throughout the entire follow-up
- Version management (keeping all versions of instruments, data models, and software and denoting used versions with the stored data to enable complete reproducibility of any specific set-up and respective data set at any later point in time)
- Minimizing sources of error by integrating data from technical devices electronically
- Safe and efficient integration and archiving of all data

Wherever possible, quality management measures should be implemented centrally in order to maximize their usefulness.

- ▶ **Data privacy protection:** Data privacy protection must be considered conceptually during all phases of the project, as described in detail in **Sect. A.7** of this document. Data privacy protection must be supported by organizational and technical measures according to the present state of technology and must fulfill all protection requirements at any time. Confidentiality, integrity, availability, and authenticity of person-related data must be ensured, as must transparency and auditability in processing person-related data.
- ▶ **IT safety and security:** With respect to IT safety (i.e., protection against accidents) and security (i.e., protection against threats), data management concepts and procedures must guarantee availability, integrity, and confidentiality of all processed data and availability and integrity of the data management infrastructure. In detail, this comprises the following aspects:
 - Examination of study subjects must be possible at any time without serious disturbances due to data management issues. For maximum safety, electronic data capture procedures should be supplemented by a paper-based back-up procedure.
 - Electronically captured data must be immediately stored safely and permanently. The organization of data management must guarantee fast, complete, consistent, and up-to-date restoration of data in case of data loss as a result of human or technical failure or application malfunction, by using a data back-up and recovery concept.
 - Unauthorized access to any data must be reliably prevented. IT security provisions must conform to the relevant standards of the Federal Office for Information Security.
- ▶ **Specific requirements:** Data will be acquired in a large number of study centers distributed over Germany; however, central integration of all locally captured data is required for efficient scientific analysis. Some instruments of data acquisition, e.g., MRI, ECG, or accelerometry devices, create data that require special knowledge, specific hardware, and/or specific software to be useful for scientific analysis. Such “complex” data must be interpreted using the professional expertise and resources available at the participating institutions. Study data collected in the study centers will be complemented by secondary data from other sources (e.g., public registries, health insurance companies, and environmental / geographic exposure data). Requests for such information and integrating data from central, external sources should be handled by a central unit within the National Cohort.

A.4.2 Organizational model

Based on the requirements described above, an organizational model for data management has been developed, which also forms the basis for defining processes and dataflow. It involves three types of main organizational units:

- ▶ **Study centers: recruitment, study subject management, and data collection**
At the study centers, study subjects are recruited and managed, examined, and interviewed. In the process, study data and biosamples are obtained. Secondary data from local sources (physicians, local registries) are collected by the study centers as well. Quality management tasks involving local records are executed by study centers personnel.
- ▶ **Integration centers: integration and processing at the national level**
Two integration centers collect and integrate all study data from the study centers and competence units. They execute automatic checks for consistency and completeness and generally carry out all the technical procedures for quality assurance and quality control that are not covered by a specialized competence unit. They provide access to standardized study data to all the study centers and offer a technical platform for data analysis by internal and external scientists. The two integration centers offer identical services and both store all data, so they can take over each other's tasks if needed, thus guaranteeing redundancy of all main processes and high service availability.
- ▶ **Competence units: topic-specific integration, processing, quality assurance, and standardization**
Competence units for particular, thematically defined areas carry out centralized tasks of enriching, standardizing, assuring quality, and controlling quality of data. These tasks include:
 - Processing and interpreting “complex” kinds of data (such as MRI or ECG data)
 - Standardizing, assuring quality, and performing comparative quality control over all study centers
 - Collecting data at the national level, e.g., secondary data from health insurances
 - Performing central tasks in endpoint-specific follow-up

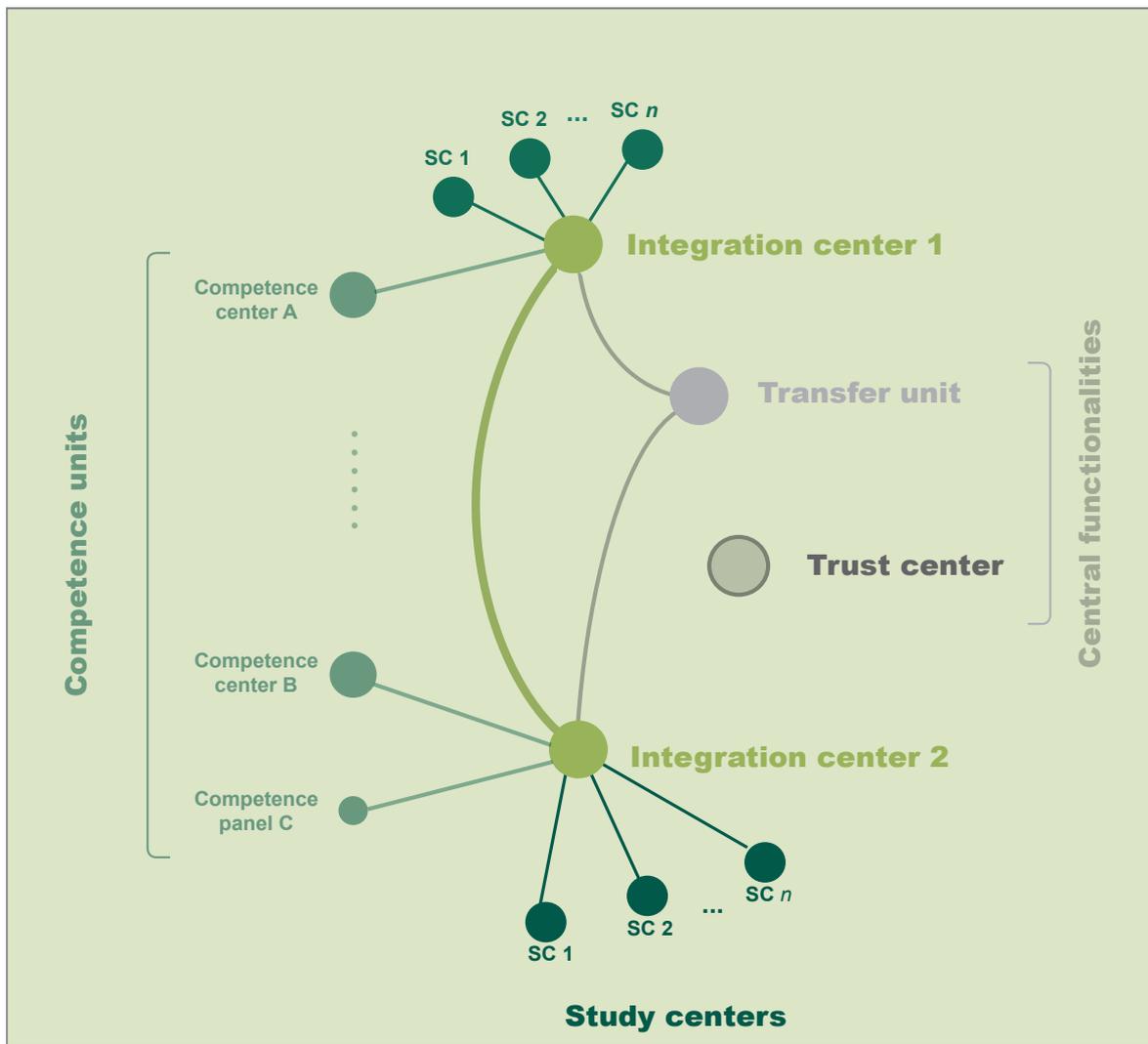
Competence units operate either as competence centers or as competence panels, depending on the technical needs of the area being covered; see **Sect. A.4.2.3** for details.

Figure 4.1 gives an overview of the resulting organizational structure, and **Figure 4.2** visualizes the relationships between the different types of organizational units by showing the most important data flows between them.

Apart from the data management tasks assigned above, two more functions will be implemented at the central level:

- ▶ A trust center is needed for tasks where personal data are to be processed or study subjects need to be reidentified at the central level.
- ▶ The administrative tasks related to clearing and transfer of data from the national cohort for analysis through internal and external scientists will be taken on by a central transfer and documentation unit.

Figure 4.1: Organizational structure for data management of the National Cohort



A.4

Figure 4.2: Data flow among main organizational units

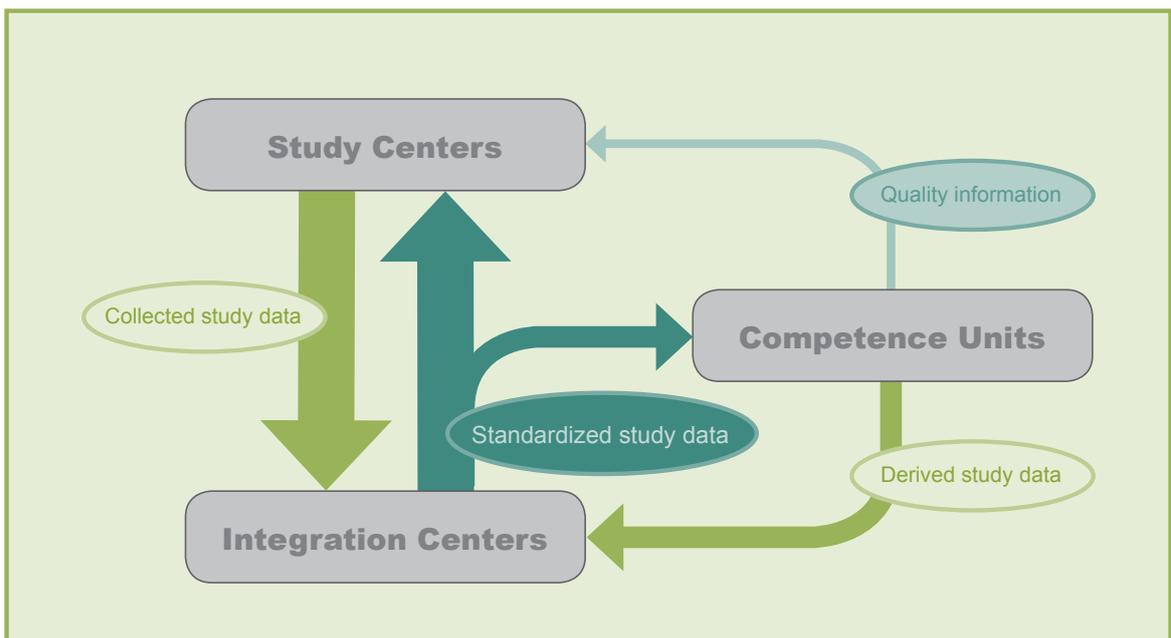
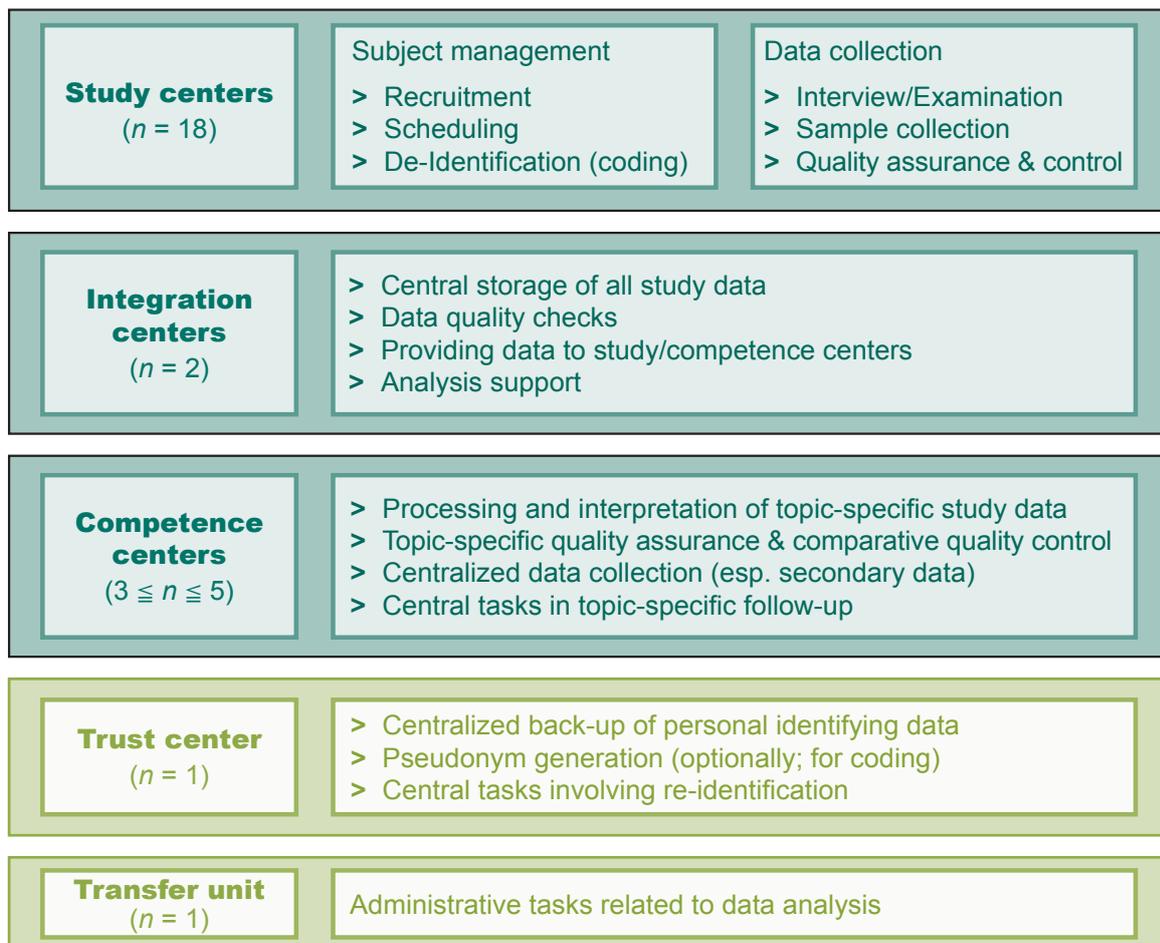


Figure 4.3: Organizational units and their functions

A.4

A.4.2.1 Study centers

The tasks of study centers are divided into two main areas: study subject management and data collection. This division is also reflected in **Figure 4.4**, which schematically depicts the basic structure of a study center from a technical point of view: The study subject management part (including recruitment, schedule management, registration, and (re-)contacting) is shown on the left-hand side, the data collection part in the middle. In the two parts, separate kinds of data are processed: Personal identifying information for the study subjects is handled in the subject management part and stored by the unique identification number ID-P (see **Sect. A.7**). In the data collection part, all collected study data regarding a particular subject are identified by the pseudonym ID-S. The two parts are linked by the relation between ID-P and ID-S, which is created during recruitment, and is handled with special care concerning confidentiality.

Study centers are built as similarly to each other as possible, including uniform hardware, software, and data models that will be developed centrally.

Personal identifying information for the study subjects is stored in a local database at each study center, separately from any medical study data. Access to this database is restricted to a few authorized persons within the study center. The information is also sent to the overall subject list (“master patient index”) at the trust center for a central check for double entries, pseudonym generation (for coding), and as back-up for data safety reasons. However,

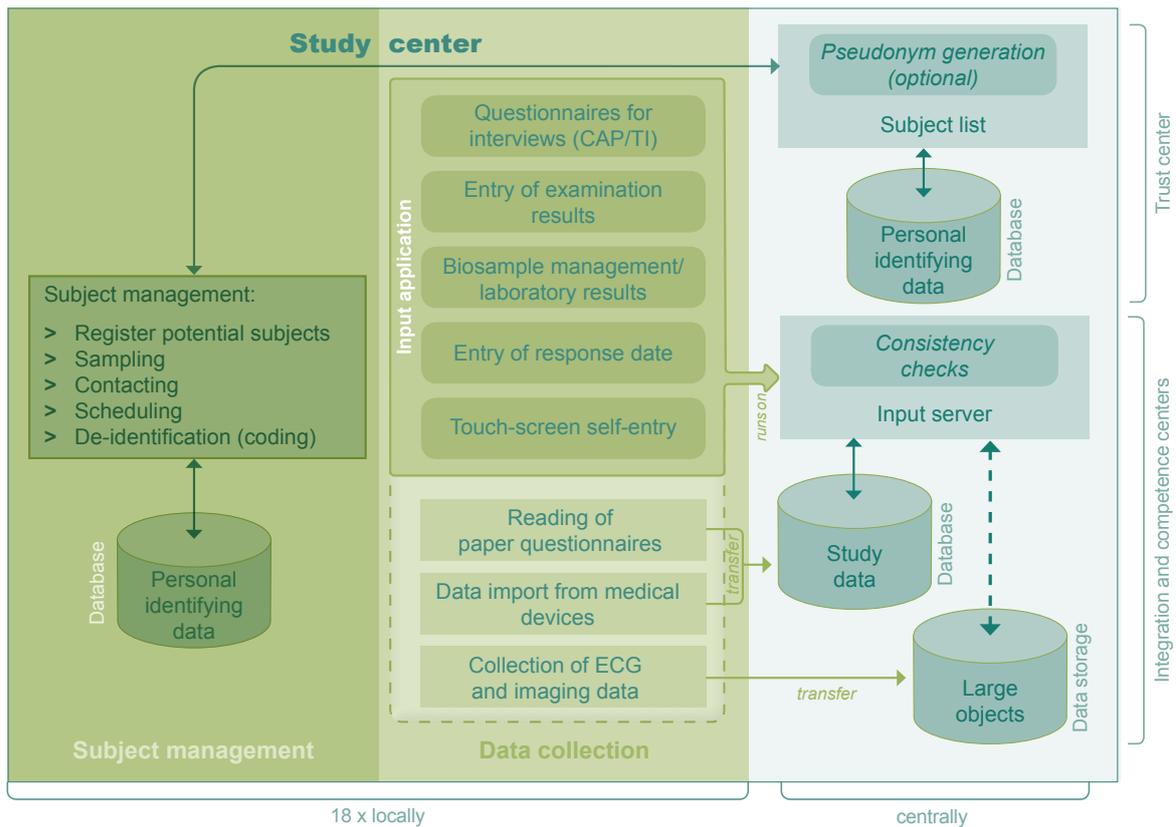


Figure 4.4: Technical structure of study centers

the responsibility for subject management and personal identifying information remains with the study centers, as they represent the main contact for the study participants.

For data collection, electronic forms will be used for all types of manually entered data, including questionnaires (both assisted and self-administered), interviews, examination progress documentation and results, biosample management, etc. All entered data are immediately and automatically checked for plausibility, consistency, and completeness, and are instantly available throughout the study center. Electronic forms are preferably provided by software running on a central server, which stores all collected data immediately in a central database.

Data stored on medical devices are exported to the central study database automatically insofar as possible.

Data collection for the National Cohort includes integration of secondary data from local sources, such as medical data to be requested from the general practitioners or treating physicians of the study subjects or subject-related status data requested from the local population registries. The study centers request these data according to the informed consent given from the local data source. Resulting data are transmitted to the integration center in de-identified form, in a manner similar to transmission of other study data.

A.4.2.2 Integration centers

Two integration centers are to be established for the National Cohort, one in Heidelberg and one in Greifswald. Their main tasks are

- ▶ To collect, integrate, and store all study data from study centers and competence units
- ▶ To operate highly available central tools for electronic data capture

- ▶ To run automated procedures of quality control on the collected data
- ▶ To make all collected data immediately available to the study centers that collected them
- ▶ To provide competence units with access to the respective topic-specific data
- ▶ To support scientific analysis of study data in collaboration with the transfer unit

These tasks are described in more detail in **Sect. A.4.2.5** below.

For software and data model development, a number of planning, preparatory, and supporting tasks can be carried out by experts from all participating institutions. However, this work needs to be coordinated centrally; the integration centers are responsible for this task.

The task includes developing an integrated data model for standardized storage of all study data and its change history; the model will be based on a formal specification of all data collection instruments, from which all other data structures and forms (both electronic and paper forms) are generated. This specification (the common “data dictionary”) contains information about the meaning, structure, and presentation of the collected data. Furthermore, all software that is needed, e.g., for subject management and data collection, will be developed or adapted from existing software; a hierarchical system for controlling access to the different areas of data collection and processing needs to be developed; and study center personnel must be trained in using the software.

Finally, integration centers execute some technical tasks in the area of remote and local system administration, including technical support for study centers or operating the central access control system.

A.4.2.3 Competence units

Competence units have the task of processing, interpretation, and quality management of complex or topic-specific study data in the aforementioned (**Sect. A.4.2**) four areas:

- ▶ Integration and processing of “complex” kinds of data, i.e., data that need to be further processed or interpreted by specialists in order to become useful for statistical analysis, at the national level. Through processing and interpretation, additional “simple” study data are derived from the complex data.
- ▶ Standardization, quality assurance, and comparative quality control concerning particular assessment instruments for the respective topics. This includes defining SOPs and conditions for result validity as well as comparative analysis of data from different study centers, e.g., for recognizing and sustainably correcting systematic divergences which may be caused by differences in operating procedures, coding criteria, or the calibration or functioning of medical devices, etc.
- ▶ Data collection at the national or at least regional level, e.g., from external sources such as registers or health insurances, or directly from the study subjects via internet.
- ▶ Central tasks in endpoint-specific follow-up; particularly, endpoint committees (see **Sect. A.3.7.5**) can be considered competence units.

Competence units will be established in two different forms, depending on the technical needs of the areas being covered: If the allocated tasks include electronic data processing using a specialized software or hardware infrastructure (which is the case for MRI, for example), a dedicated competence center is set up and provided with that infrastructure. Otherwise, without such needs, a competence panel of qualified experts is set up; such competence panels use the IT infrastructure provided by the integration centers.

The selection of competence centers implies a trade-off between the requirements and advantages of specialized processing and quality management and the additional effort and

cost of setting up and maintaining a greater number of specialized units each with their own infrastructure. Currently planned competence centers are:

- ▶ Imaging: Handling of all MRI, DXA, ultrasound, and echocardiography data. This includes providing teleradiology services to different places where experts on specific parts of the images are located; running separate competence centers for this would not be cost efficient.
- ▶ Physical activity and ECG: Both of these instruments require application of algorithms on the collected data to derive characteristic values. For this reason, they could be handled together in one competence center with a common infrastructure and supported by the same IT personnel. While the algorithms are relatively well-established for ECG, methods for deriving measures of physical activity from accelerometry data are in a more experimental state and will probably be improved continuously and re-applied multiple times while the study is running.
- ▶ Secondary data: The selection of external sources and data elements, the collection of data from those sources into temporary storage, and the mapping of such data into the common data model of the study database requires not only a considerable amount of manual work but also specialized hardware and software. Possible procedures for the use of health insurance data and social insurance data are described in Sect. A.7.2.3 of this document.
- ▶ Biobanking: A biobank is similar to a competence center as it uses special infrastructure and processing, even though not (only) for data but for biosamples.

In addition to these proposed competence centers, a number of competence panels will be set up for the quality management of data collection instruments not covered by competence centers and also in the field of disease-specific follow-up.

A.4.2.4 Trust center

The trust center generally takes on all tasks in which personal data of the study subjects need to be handled at the national level:

- ▶ Generation of unique pseudonyms for (potential) study subjects: Although it is possible to generate pseudonyms at the study centers and make them unique, for example, by adding a study center ID, central generation of pseudonyms is considered as the preferred alternative because it offers the additional possibility to recognize and avoid repeated entry of a single person via different study centers with different pseudonyms (so-called synonyms); such synonyms become more likely when sampling is done in multiple tranches. For the best effect, potential participants will be checked for double entry immediately after sampling and before contacting them.
- ▶ Central back-up storage of study participant data: Personal identifying data for the study subjects are kept at the trust center as a back-up for cases of data loss in the study centers. Especially in later phases of the study, these data may be useful for reidentifying participants (e.g., in case of incidental findings).
- ▶ Data linkage with external sources at the national level: External sources of secondary data that operate at a national level must also be queried centrally (rather than by each of the study centers individually). Such queries must contain personal identifying information about the study subjects, which must not be stored centrally together with medical study data. Therefore, queries for data from external sources are prepared by the competence center for secondary data and supplemented afterwards with personal identifying information about the respective subjects by the trust center (which, however, never sees the replies to those queries). See **Sect. A.7.2.3** for details.

A.4.2.5 Data transfer and project documentation unit

A central transfer and project documentation unit (in short: transfer unit) will handle all administrative tasks related to access to data intended for scientific analysis by internal and external scientists:

- ▶ Development of a formal procedure for the processes related to study data analysis, including regulations and contracts regarding data access, project monitoring, data export, as well as return and integration of resulting new data.
- ▶ Management of the administrative part of the aforementioned formal procedure (while the technical part will be managed by the integration centers), both for “internal projects” (study data analysis by scientists who have contributed to generating the data) and for “associated projects” (data analysis by cooperating scientists or together with institutions outside of the National Cohort).
- ▶ Definition of subsets of data to be provided (in coordination with the applicant) and additional coding of IDs. Possible reidentification risks inherent in particular subsets of study data will be addressed in this process.
- ▶ Retrieval and integration of new results emerging from scientific projects (e.g., laboratory results).

The transfer unit will fulfill these tasks using the IT infrastructure provided by the integration centers.

A.4

A.4.3 Technical aspects of data collection, processing, storage, and analysis

A.4.3.1 Software infrastructure

Software design, development, and maintenance will take advantage of widely used open source tools for integrated development and software version management (e.g., Eclipse IDE, Subversion). Adaptation of software components provided by contributing institutions and use of open source frameworks will be preferred to completely new development, if this fulfills all requirements.

As a relational database management system, MySQL will be used during the feasibility studies. Should this prove inadequate, other options will be considered for the main study, including Oracle database, Microsoft SQL server, or (other) open source systems.

A.4.3.2 Electronic data capture

The preferred means of collecting all kinds of manually captured data (including, for example, self-administered questionnaires, interview results, and results of examinations without any devices that capture results electronically) is by using electronic forms provided through an electronic data capture forms application (“EDCF application”). This application integrates automatic checks of plausibility and consistency of all entered data and provides for storing and safeguarding the data as quickly as possible. In addition, paper forms are available as a back-up if technical problems with electronic data capture arise.

This EDCF application will be developed and maintained centrally and made available on workstations at the study centers by automated processes. The preferred means of implementation is as a web application running on central servers at the integration centers. Workstations in study centers need to be equipped with a standard web browser only, then, and no additional software needs to be distributed. However, with that solution, network availability is critical. If an analysis (which is currently still in progress) shows that there is a considerable need for enabling offline data capture (i.e., with no network connection

between workstation and integration center), a special client-server implementation will be developed, where the client part running on the study center workstations is distributed automatically and can run without a network connection.

All data captured through the EDCF application will generally be transferred immediately to an integration center. This is implicit if it is implemented as a web application using central servers at the integration centers. If it is implemented as a special client-server solution, all data are also transferred immediately when a network connection is available. When it is not, data are temporarily stored locally and synchronized with the server as soon as possible.

In any case, the EDCF application will be generated automatically from a formal description of the data collection instruments (the “data dictionary”). The application is modularized along examination topics and methods; it can be extended through additional modules, and individual modules can be modified without affecting other areas. Access to the application is differentiated for modules and user roles, i.e., access to a certain application module is restricted to users in participating institutions using this module who are acting in a role that requires use of this module. Insofar as possible, the application will reflect the chronological course of the examinations.

A.4

Immediate data transfer in no way affects the responsibility of the study centers for their data and the need to conduct quality checks on their data. All study data concerning a specific study subject are fully available to the study center that recruited that participant. The integration centers operate servers that allow the study centers to view and export all data regarding their own participants for local use at any time.

A.4.3.3 Data from medical devices

For collecting data from medical devices, automated procedures for electronic data transfer will be implemented. Generally, all data will be transferred to an integration center for central storage and archiving; if the data are also needed at a competence unit, the integration centers will make them available to that competence unit.

In most cases, medical devices will store captured raw data locally first. Where these temporary storage capabilities are not sufficient to guarantee the safety of the captured data even during temporary network unavailability, additional software or hardware will be added for that purpose. This may also be necessary for implementing processes for exporting the stored data from the device and transferring it to an integration center.

The technical means and procedures for storing, exporting, and transferring the data to the integration centers will use clinical interface standards (e.g., DICOM, HL7) where possible and will be developed centrally (coordinated by the integration centers) for all the devices that are uniformly used at all study centers. Details for the processes and the necessary software and hardware will be developed after final selection of such devices; in some cases, this selection depends on the results of the ongoing feasibility studies.

Where extra steps are necessary for obtaining “usable” data from raw data, these will preferably be implemented as automated processes running at integration or competence centers.

A.4.3.4 Handling of paper forms

In addition to electronic data capture, paper forms will also be used for two different purposes: Firstly, paper questionnaires will be distributed to study subjects, filled-in at home,

and sent back or brought in by subjects visiting study centers. Secondly, paper versions of all electronic forms will be available as a back-up for electronic data capture.

Paper forms will be prepared and automatically generated centrally (based on the “data dictionary”), distributed electronically to the study centers, and printed there with individual pseudonyms (bar codes, for example) added. (Alternatively, depending on cost evaluation, forms could be printed centrally, distributed to the study centers by mail, and supplemented with individual pseudonyms there before shipping them to the subjects). The preferred way of handling filled-in paper forms is to scan them in the study centers (including data capture) and transfer the data to the integration centers electronically, while paper archiving is done at the study centers (if necessary). Should the cost prove too high, alternative solutions will be considered. Options include (a) manual data capture in the study centers, (b) scanning of paper forms at the study centers and centralized automatic data capture from the images, and (c) paper forms being sent to the integration centers for central processing.

A.4.3.5 Processing of study data by competence units

Competence units generally work on standardized study data provided to them by the integration centers. In order to enhance the standardization of captured data and dataflows, direct transfer of data from study centers to competence units will be avoided insofar as possible. All topic-specific study data required for the work of a competence unit are provided in the form of specific views via central servers in the integration centers. More “complex” data objects (i.e., images, binary data objects) may also be exported automatically to the competence centers (i.e., copied to their own infrastructure). Since permanent storage and archiving is centralized at the integration centers, no data need to be stored permanently at the competence centers.

The results of any processing, interpretation, and review by competence units are captured in similar ways as in study centers, i.e., by electronic data capture forms centrally provided by the integration centers or by direct data transfer from technical devices. Both competence centers and study centers have full access to these results for their respective subjects at any time.

A.4.3.6 Integration and storage of study data

Through the use of electronic data capture via servers in the integration centers insofar as possible, all data collected through electronic forms will immediately (or at least as soon as possible) be integrated and stored centrally. Data from medical devices will also be transferred to the integration centers as soon as possible, using device-specific procedures, which will be automated to the greatest possible extent.

The stored data include a change history that documents the original value of each study data record and the date, reason, and executing person of every change that has been made afterwards. A version management system keeps all versions of data models and software and denotes used versions with the stored data to ensure that any specific set-up and respective data can be completely reproduced at a later point in time.

Data that are no longer actively needed are archived in ways that guarantee their long-term availability. All data are backed up at regular intervals, and a procedure is in place for restoring data, if needed, with adequate effort in adequate time. This also includes a procedure for point-in-time recovery of data, which is able to restore study data in the state they were in immediately before a failure event.

All study data will be mirrored between the integration center sites in order to achieve a high degree of availability. Together with all services also being provided in identical ways, each integration center can stand in for the other in case of any failure.

Because of the complexity of these tasks and the effort required to fulfill them with the necessary degree of reliability, the option of a centralized service being provided by the integration centers saves considerable effort in comparison to any other solution that relies on local data storage at the study centers.

A.4.3.7 Analysis and use

The integration centers implement the technical part of making study data available for scientific analysis, while the administrative part is managed by the transfer unit.

Analyses can either be executed on a computation platform provided by the integration centers (by granting access to project-specific views of subsets of study data, such that study data never leave the platform if not necessary), or specific subsets can be prepared, additionally coded with new pseudonyms, exported, and transferred physically to the project partner analyzing the data.

A.4

Background literature to the topics covered in this chapter is in references 751-780.

A.5 Methods for quality assurance and quality control

A.5.1 Organizational structure for quality assurance and quality control

To ensure that the study conduct as a whole, including data collection and data handling, is of high quality over time and across study centers, a quality management system will be established, as described in Part B. During the set-up phase of the study, we will decide whether this system will be certified according to ISO 9001, as, for example, in the Heinz-Nixdorf-Recall-Study.

This section concentrates on standardized procedures for quality assurance and quality control, which have been developed and are being established irrespective of the presently open certification issue. The following figure depicts the organizational structure of quality assurance and quality control in the National Cohort (**Figure 5.1**).

Figure 5.1: Organization and communication of external and internal quality management



As shown in the figure above, we distinguish between internal quality management and external quality management.

A.5.1.1 Internal quality management

The organization of internal quality management is congruent with the general organizational model introduced in Sect. A.4 and supplements it with an additional organizational unit, the central quality office. The roles of the organizational units in respect to quality management are detailed in the following.

Study centers: The study centers will implement quality standards at the local level according to study-wide SOPs. Study centers will continuously document their study procedures in quality protocols. These documents will include feedback from study personnel along with suggestions for improvements. A local quality manager will be responsible for all quality management-related processes inside each study center. He/she reports directly to the head of the study center and represents the center in the working group for quality management at the central level.

Competence units: For each theme-specific data assessment module (e.g., data from ECG, MRI), competence units are responsible for issues related to quality assurance and quality control. Each competence unit is organizationally and scientifically responsible for that particular theme. The competence units will formulate quality requirements, ensure that quality requirements are fulfilled, take measures when quality issues arise, and communi-

cate with the integration centers. For each theme-specific data assessment module, the specific tasks of each competence unit include:

- ▶ Defining quality requirements of each assessment module
- ▶ Defining requirements concerning the study center layout and the standardization of infrastructure and facilities across the individual study centers (e.g., room size and room temperature)
- ▶ Establishing and updating SOPs to ensure standard procedures across all study centers and considering gold standard assessment criteria (e.g., for MRI), if needed
- ▶ Developing and testing training materials for study personnel
- ▶ Defining the qualification profile for study personnel, including personnel for training, certification, and recertification
- ▶ Determining the equipment required for examinations (e.g., medical devices and computers) to ensure that identical assessment tools are used in all study centers
- ▶ Defining the interface between devices and the online data storage system
- ▶ Developing calibration and maintenance procedures for the devices
- ▶ Defining exclusion criteria of study participants for critical examinations (e.g., no radiographic imaging during pregnancy)
- ▶ Establishing plausibility checks to detect conspicuous data (e.g., missing information, incorrect or implausible values, and inconsistent items)
- ▶ Establishing procedures for data cleaning
- ▶ Developing quality protocols to provide standardized quality documentation tools for the study centers
- ▶ Monitoring theme-specific data to ensure ongoing supervision of all procedures
- ▶ Analyzing the records of monitoring tools (e.g., quality protocols) from the study centers, preparing quality reports, and forwarding them to the central quality office and to the study centers

A.5

Integration centers: The integration centers are responsible for central storage of all study data, data quality checks, and data transfer to competence centers. They ensure automated quality management procedures (including automated checks for plausibility as well as consistency with existing data and automated generation of quality protocols for quality control). They also implement the technical aspects of making the study data available for analysis, and they coordinate the selection, adaptation, or development of all software.

Central quality office: The central quality office supervises all aspects of data quality. The office ensures that all tasks are carried out according to the SOPs. A representative from the central quality office will participate in every meeting of the working group for quality management. It also liaises with the external quality management.

A.5.1.2 External quality management

External quality management is implemented in the form of a panel of experts (e.g., laboratory experts or imaging experts) that evaluate and systematically review the assessment procedures and prepare quality reports detailing the results of the review. These experts will furthermore be responsible for the annual site visits of the study centers as described below.

A.5.2 Principles for quality assurance and quality control in the National Cohort

All data collection procedures will be standardized, rendering all the data collected across the individual study centers comparable. This will reduce intra- and intercenter variability.

An essential component of quality assurance and quality control is training, retraining, and certification of study staff. A training description with information on contents of the training course, definition of the number of test runs, and runs conducted under supervision will be described in the SOPs, which will also include training materials. Face-to-face training will be held centrally (or regionally, with special attention on having the same training in all regions) and is supplemented with local training sessions held by instructors selected by the competence units. Newly hired staff will receive the same training procedures as the core personnel. Annual retraining and recertification are necessary to avoid drift over time. Study staff will be certified upon completion of the training course; the certification procedure will directly follow the training sessions. Certification criteria will be defined by the competence units and approved by the working group on quality management. The certification results will be documented for future reference. Training will be coordinated by the central quality office; organizational issues will partly be delegated to regional or local units.

Before training and certification start, the respective procedures will be piloted using mock study subjects and, if approved, with actual study participants.

Site visits represent an important part of the quality monitoring process. All centers will be visited by external quality management personnel once per year. The aim of these site visits is to assess whether study procedures are performed according to the SOPs (interviews, medical examinations, calibration, and use of devices). Information on how procedures are performed at the study centers will be collected using quality protocols and checklists which are derived from the SOPs. The protocols of the site visits comprise part of the quality reports. Monitoring procedures will be aimed at (1) checking the quality and completeness of the data on a regular basis, (2) identifying intra- and inter-(examiner, study nurse, interviewer, technician, instrument, and study center) variability, (3) identifying temporal changes in study parameters or examinations over time, and (4) identifying data outliers.

Monitoring procedures at the level of the study center. Specific checklists and quality protocols will document the quality assessment at each study center. The preferred means of applying checklists and quality protocols is in electronic form, integrated with the centralized EDCF application; paper may be used where this is not feasible. Checklists will be used to ensure that each study subject participates in all medical examinations and interviews. Additionally, particulars that arise during the assessment visit are documented (e.g., reason why a study subject did not receive a certain examination). Quality protocols include data analyses and data presentations regarding specific quality aspects of the study. For this purpose, gold standards or alloyed gold standards will be identified by the competence units to monitor the accuracy of the examiners or instruments and to analyze intra- and intervariability in study center staff, assessment instruments, and study centers. For each theme-specific assessment module, specific quality control requirements will be assessed and adequate standards will be identified by the competence units to verify the accuracy of the examiners or instruments and to analyze intra- and intervariability in study center staff, assessment instruments, and study centers.

Missing, incorrect, interchanged, or nonclassifiable values will be documented and interpreted by the local quality manager and discarded if necessary. Conspicuous values will be discussed with experts and documented for future reference. The plausibility of data outliers will be monitored regularly both by the local quality manager and by the respective competence units. Calibration and technical maintenance of the technical equipment will be documented, and the records will be analyzed by the competence units responsible for the respective kind of equipment. In order to assess intra- and intervariability, each assessment item will be tested against a predefined “gold standard.” The variation will be assessed, documented, and interpreted as part of the quality protocol, which will subsequently be circulated to study staff and will result in corrections, instruction updates, and retraining

procedures, if required; the local quality manager is responsible for accomplishing such measures .

Monitoring procedures at the level of the central quality office: The central quality office will coordinate the monitoring procedures. Therefore, the monitoring results of all study centers will be analyzed for entire study population sample and for stratified subsamples (e.g., by study centers, gender, season, age, and examination). The central quality office will prepare quality reports in cooperation with the competence units, which are evaluated by external quality management. Subsequently, the central quality office will refer the comments and instructions given by the external quality management to the competence units, who may propose new guidelines. Quality reports will be generated upon completion of the feasibility and pilot studies. For the main study, overall quality reports will be prepared bi-annually. In addition, interim quality reports will be prepared at shorter time intervals (weekly or monthly), depending on the specific assessment in question.

A.5.3 Standardization of sampling methods and assurance of high participation

The representativeness and number of study participants will be assessed in each study center by comparing the composition of the population sample with the general population in the recruitment area according to age, sex, and other demographic variables. Instructions for subject recruitment will be checked for completeness and plausibility with respect to the population to be recruited. All contact procedures described in the study protocol with respect to the contact letters and sequence and timing of contacts will be monitored. For this monitoring a uniform electronic appointment system will be used. An initial letter, at least one postal reminder, and subsequent phone calls are foreseen for recruiting subjects into the cohort as described in **Sect. A.3.1.4**.

A short nonresponder questionnaire soliciting basic demographic data will be employed to detect possible selection biases at an early stage. As part of quality control, the study centers will thus assess the reasons for nonresponse or drop-out; the number, content, and contact time of successful and unsuccessful contact attempts; and documentation and correct classification of the reasons for nonparticipation and verification of quality-neutral drop-outs. The collected information is examined both locally and by the central quality office.

As soon as deviations from anticipated response rates are detected, specific counteractive measures will be taken (e.g., target population-specific measures). It is also important to ensure that subjects willing to participate are included in the cohort and can be convinced to participate in future follow-up. Means to increase response rates will be applied and documented to measure positive (or negative) effects on response rates. Measures of cohort retention to be applied include:

- ▶ Incentives
- ▶ Sending information flyers and brochures
- ▶ Sending general or preliminary study results
- ▶ Updating addresses at postal offices or residential registries or by phone call
- ▶ Providing subjects the opportunity for indicating a change of address (free postal card, email account, or password-protected website)

A.5.4 Quality assurance and quality control of questionnaires

A.5.4.1 Quality assurance of questionnaires and interviews

A variety of data collection instruments will be employed, including paper-and-pencil self-administered questionnaires, web-based self-administered questionnaires, touch screen questionnaires, computer-based face-to-face interviews, and computer-assisted telephone interviews. A comprehensive plan of operations for using each data collection instrument will be established by the respective competence unit. Furthermore, a back-up assessment method will be available should the primary assessment method prove to be impossible. The web-based self-administered questionnaire, touch screen questionnaire, computer-based face-to-face interview, and computer-assisted telephone interview will be designed to include plausibility checks and skip patterns to avoid implausible data, discrepancies in answers regarding one questionnaire, or inconsistencies among data collected through different modules of the questionnaire.

All interviews are intended to be recorded (e.g., using an MP3 format) at each study center. Experience from other studies in Germany has shown that acceptance of interview recording is generally high. Recording interviews provides the opportunity to monitor the standardization of study personnel assistance and also makes it possible to correct incompatible or implausible data. Special aspects of the self-administered questionnaires include a standardized manner of distributing the questionnaires (before or after the clinical examination, via postal service, or web-based), standardized data entry from a paper format to a digital format (scanner-readable format versus data entry by two different persons), and provision of a prepaid and preaddressed envelope for each participant (if needed).

A.5.4.2 Quality control of questionnaires and interviews

Quality control for questionnaires and interviews will be applied on a regular basis throughout all phases of the study. For each data assessment instrument applied in the National Cohort, distinct procedures will be employed to monitor study personnel and study equipment. Plausibility checks, which will be automated insofar as possible and be integrated into the data management infrastructure, will help control data quality on a continuous basis. The quality of the interviews will be monitored through direct observation during site visits, supplemented by the recording of all interviews. By recording all interviews (e.g., using an MP3 format) any interview can be selected for quality control afterwards. The responsible experts from competence units or from external quality management will play back random samples to verify adherence to the SOPs, the good quality, and the validity of the collected data and provide feedback to the interviewers. Such systematic evaluation of interviewer performance will further ensure high quality of interviews.

A.5.5 Quality assurance and quality control of medical examinations

A.5.5.1 Quality assurance of medical examinations

To minimize the variability created by using different physical examination instruments, all centers will be equipped with the same devices. The medical examinations will be tested during the feasibility and pilot studies at each study center over the study period; progress and new technologies can arise, rendering the currently used devices obsolete. If devices are replaced, validation studies will be conducted that are coordinated by the respective competence units. Authorization for changes or modification to study devices must also be given by the competence units before approval by the epidemiologic steering committee.

For certain assessments, replicate measurements will be carried out in order to identify data errors. All measured values will be recorded, so that implausible values can be discarded afterwards and more accurate data obtained.

A.5.5.2 Quality control of medical examinations

Checklists will document potential problems that arise during the medical examinations and will provide pertinent information to the competence units. Assessment devices will be calibrated and inspected regularly using the same algorithms at each study center to ensure reliable and comparable data collection. The procedures will be developed by competence units, compiled and distributed by the central quality office, and carried out by study center personnel under supervision of the local quality manager.

A.5.6 Quality assurance and quality control of the collection and processing of biological specimens

Aspects that potentially impact upon the quality of biological specimens include the technical skill of study examiners and laboratory technicians, the accuracy in collecting and processing biological specimens, accuracy in liquid handling, transport, and storage conditions, and the time elapsed between sample collection and sample storage.

A.5

A.5.6.1 Quality assurance of biological specimens

To ensure comparability between the study centers and minimize variability regarding pre-analytic artifacts, all study centers will apply the same liquid handling procedures. The use of a high-throughput liquid handling platform in each study center will ensure precision and accuracy in the workflow.

Detailed SOPs and specific training materials for sample drawing, collection, processing, transportation, and storage will be compiled by the competence unit on biomaterials. Different techniques for sample drawing, collection, processing, transportation, and storage will be tested in feasibility and pilot studies designed to address adequate testing of all procedures and techniques and to identify and solve potential problems in the workflow.

Criteria and intervals for accuracy checks will also be defined by the competence unit on biomaterials. All results of accuracy tests will be documented in a data quality protocol.

The collection of blood, urine, and saliva will be performed by study nurses, who will participate in a specific training and certification course on the standardized collection of biological specimens. Laboratory technicians responsible for sample processing and storage will also be trained according to specific SOPs. Examiners and technicians will be certified upon completion of the training course. A video presentation on the collection and handling of biological specimens is currently being considered as an additional training option. The training will be coordinated by the central quality office, based on recommendations from the competence unit on biomaterials. Regional structures will be used for practical organizational issues where possible.

Varying time intervals between sample drawing and processing will be tested in a feasibility study to analyze the impact of elapsed time on the quality and viability of the various biomaterials. Criteria related to sample transport (e.g., transport on dry ice or in ice water) will include details regarding the time of transportation, the sample temperature, and temperature deviations during transportation.

The influence of temperature, storage time, preservatives and additives to the collection tubes, and repeated rounds of freezing and thawing on the sample quality will also be tested during the feasibility studies using appropriate and sensitive markers. To guarantee temperature stability over time, the temperature of all freezers will be documented daily. The number of stored aliquots and all deviations from the standardized protocol will be documented in a central database. Furthermore, a contingency plan will ensure sample integrity over the intended lifetime of the study, including storage solutions designed to assure that conditions are maintained (e.g., mechanical breakdown of the refrigeration plant, the robotics system, or the electrical supply).

The procedure for sample aliquot retrieval will be established according to predefined criteria and follow specific SOPs. Furthermore, sample movements will be tracked and logged in a central database using the bar codes of each container.

Deviations from any of the standardized protocols will be documented in the central database and therefore be readily available for examination by any unit involved in quality management.

Novel technologies that arise during the recruitment or follow-up phases of the study will be tested in parallel with existing methods. Only if old and new technologies yield comparable data will we consider replacing existing methods.

A subset of clinically relevant laboratory data will be provided to study participants within several weeks of their visit at the study center. Clinical parameters will be analyzed in certified laboratories associated with the study centers. Measurements between the clinical laboratories will be verified to determine interlaboratory variability.

If necessary, the epidemiologic planning committee and competence units will define whether certain groups of participants will need to be excluded from the study. To identify appropriate exclusion criteria, specific questions will be included in the core questionnaire.

A.5.6.2 Quality control of biological specimens

To monitor adherence to the SOPs, experts selected by the competence unit on biomaterials will visit the study centers at regular intervals. The competence unit is also accredited to authorize deviations from SOPs, if necessary. All study centers will be visited by external quality management once each year to ensure adherence to the SOPs and to certify the examiners and technicians on site. Furthermore, the examiners and technicians will be retrained and recertified once per year to ensure adherence to the SOPs over time. If quality disparities are identified in biological samples collected by a certain individual, reasons which could explain such variation will be ascertained and used to retrain that particular technician. Newly hired personnel will receive the same training procedure as the core staff.

Replicate measurements are used in a number of settings. During the feasibility and pilot studies, they are used to test the impact of time and storage conditions on sensitive biomarkers and to test different verification procedures within the study centers or clinical laboratories. During the main study, replicate measurements are used to identify data errors. Replicate measurements will be appropriately documented and analyzed. Local accuracy checks will be used to detect missing, incorrect, interchanged, or nonclassifiable parameters. Interchanged or nonclassifiable biological specimens or parameter values will be discarded. Conspicuous values will be checked promptly and, if necessary, will be discussed with the clinical laboratory responsible for that particular value.

The liquid handling equipment will be calibrated and inspected regularly to ensure accuracy and precision of the robotics procedures. The precision in preparing aliquots will be checked

by a technical support team from the vendor of the equipment twice a year in each study center to minimize errors. The resulting technical report will be checked by the competence unit on biomaterials on a regular basis. Quality reports will be prepared to determine intra- and interreliability between examiners/technicians as well as among study centers.

A.5.7 Measures to assure high and consistent image quality in the MRI substudy

Quality assurance of the imaging procedures represents a challenge, given the multicenter design of the National Cohort. Standardization procedures established for imaging examinations in the Study of Health in Pomerania (SHIP) will serve as a basic template for the National Cohort.

For all examinations, SOPs are available for obligatory perusal by the examiners. The examination procedure is tested in pilot studies. All examiners are trained over several months and certified following the training period. Potential observer bias or trends are continuously monitored. These efforts to ensure high quality standards result in low method and observer variability and are particularly important given the long duration of the data acquisition phase.

Technologist training: The MRI technologists on-site will already have adequate experience in performing MRI examinations. In general, a minimum of 3 years of MRI experience is required. All technologists will undergo centralized training, consisting of lectures and practical lessons covering scanning and data transfer. Furthermore, they will be trained to perform a first-line quality check (see below) after each study participant has been examined. All technologists will be recertified on an annual basis. The same application engineer from the vendor will also provide on-site training for the technologists of all imaging sites. Automated operation and user guidance (DOT engine) are provided by the vendor to assure highly standardized acquisitions.

Phantom measurements: It is anticipated that regular measurements on standard and organ-specific MRI phantoms will be performed (phantoms will be specified after final decisions have been made concerning the examination program). This ensures external validity and also addresses the challenge of a long-lasting field phase.

Imaging devices: To minimize device-related bias, new high-end imaging devices with identical hardware and software equipment from the same company are required for the National Cohort, for which no updates (e.g., software updates) will be applied as long as data are collected at baseline and during follow-up. Comprehensive imaging examinations including MRI will be conducted only at centers where radiological expertise in large-scale clinical and epidemiologic studies with high-quality standards is available.

Pilot studies: Before the start of the main study, pilot studies will be conducted to test the feasibility of methods and procedures. During each pilot study, volunteers will be examined. Information gained in these pilot studies will be used to optimize the procedures at each study center, to detect real or potential leaks or blocks in the information flow, and to further ensure the quality of the study.

First-line quality check: The first-line quality check will be performed at the imaging site by the technologists for each examination. In general, this will include completion of a checklist (preferably in electronic form, with paper as back-up) to ensure that all required sequences were acquired, no major artifacts are present, and that adequate measures were taken to store the acquired images. The criteria for the first-line check will be specified in advance

by the competence center on imaging and implemented and trained during the centralized technologist training courses. If insufficient image quality is detected, additional MRI sequences will be acquired to maintain a complete protocol. If there is any technical failure, the technical service of the vendor will be notified and a re-examination will be offered to the subjects.

Second-line quality check: A second-line quality check will be performed at the centralized reading core at the competence center on imaging prior to the first-line reading. This check will include similar items as for the first-line quality check but will additionally involve subjective assessment by a radiologist. If impaired image quality is detected, the respective imaging center will be contacted and any causes will be investigated. If a higher frequency of cases with impaired image quality is detected at one imaging site, retraining and site visits will be employed.

MRI reader certification: All imaging examiners will undergo training in the respective clinical departments before they enter the study. The potential examiner will first be certified after he/she has been trained for at least 3 months. During this period, each trainee is required to examine at least 2 x 25 images for continuously distributed findings and at least 2 x 100 findings for dichotomous variables. Intra- and interobserver variabilities for selected major characteristics are determined. Statistical analyses are conducted using Bland and Altman plots for continuously distributed variables. The mean bias will not exceed 5%, and 2 SD of the bias will not exceed 25%. Statistical analyses for dichotomized variables are performed by κ statistics, whereby $\kappa > 0.8$ is expected. Trainees who do not fulfill the quality criteria after 3 months of training are subjected to further mandatory training. The additional training includes an individualized calibration of the trainee based on the results of his/her first certification. Should there be no improvement, the trainee is replaced by another person. A web-based certification procedure is planned.

Further efforts focus on the external validity of the data. Imaging modalities will be intensively discussed with national and international experts in advance of the field phase of the National Cohort. To support comparability of the National Cohort with other population-based studies, we will invite external observers to take part in the web-based certification procedures on a regular basis.

A.5.8 Quality assurance and quality control during follow-up

In parallel to the processes described above, the study centers will be busy with continuously maintaining contact to the other members of the cohort, deriving information from these sites, from their own doctors, and from institutions and other data sources as described in Sect. A.3.8. In addition to the quality management aspects presented in that section, the more general task will be to implement a system that makes it possible to compare the results of the different center-specific monitoring processes at the central level. A set of quality indicators, e.g., percentage of participants who are successfully tracked, percentage of responders giving information on contacts to the health system, completeness of vital status or death certificate information, will be created and compared across centers. By using this benchmark process we can identify possible problems, but also chances for improvement.

Concerning endpoint information, a step-wise validation procedure is envisaged:

1. For cancer, stroke, and CHD all events of category (a) and a random sample of category (b), as described in **Sect. A.3.7.5** will be seen by a central endpoint committee.
2. For all other endpoints, a random sample of the documented events will be assessed by dedicated expert panels.

All diagnoses will be coded according to the ICD version being effective at the start of the study. Revisions and/or diagnostic progress will be considered as add-on diagnoses, if deemed appropriate. Information that leads to a specific ICD code will be stored electronically such that later revisions can be used with only minimal additional effort.

The feedback from the expert assessments to the study centers will enhance the diagnostic quality and comparability of the diagnoses across centers and over time.

The results of the abovementioned benchmarking process, quality indicators comparing study center-based diagnoses with the results from the experts' assessments, and the number of events stratified by study center will be disseminated on an annual basis.

A.5.9 Quality assurance and quality control of data management

High standards for data acquisition, data processing, and data storage will be enforced by implementing quality assurance and quality control procedures in data management.

A.5.9.1 Quality assurance of data management

Quality assurance during the system analysis phase

A system specification document will specify the following points:

- ▶ Overall project aims:
 - “Must” criteria: performance parameters that the future data management system will comply with (see **Sect. A.4.2**);
 - “Can” criteria: it is not absolutely necessary to fulfill these criteria, but will be strived for if sufficient resources are available.
 - “Borderline” criteria: parameters and functions beyond the reach of the anticipated software system (e.g., quality control of the data at the level of the competence unit for data management can be individually performed with other statistical programs and therefore is not part of this solution).
- ▶ Product application:
 - Application domain: the system to be created will be used only for the National Cohort. It is possible that a similar system could be used for other future studies, but this is beyond the scope of the current project.
 - Target groups: the main target group for this system consists of study personnel from individual study centers.
 - Working conditions: the system will run as a web application on servers in the integration centers, with permanent surveillance of IT systems, data items, and data flow by central data administrators.
- ▶ Product functions: exact and detailed descriptions of each function (see **Sect. A.4.2**)
- ▶ Product data: the data that need to be stored over a longer period of time, seen from a user's point of view.
- ▶ Product performance: specifications concerning processing time and data accuracy, simultaneous updates on all servers.
- ▶ User interface: the software will be designed for both keyboard and touch screen use. In addition, a concept for a role-based access model will be defined.
- ▶ Nonfunctional requirements: external rules and laws that will be respected by the software, security requirements, and platform dependencies. The system will work with pseudoanonymous data. Access to personalized data will be restricted to a limited number of persons; the components that manage personalized data will be specially protected.

- ▶ Technical product environment:
 - Software: the software will run in a client-server environment within a browser. All current browsers, such as Internet Explorer, Mozilla Firefox, or Opera will be supported. The use of existing frameworks is preferred to developing individual solutions.
 - Hardware: the list of devices for both server and workstations which will be used throughout the study. For optimal performance, all study centers will use identical hardware.
 - Product interfaces: interfaces between product components or to peripheral devices will be described.
- ▶ Glossary: All technical terms in this document will be explained in a glossary for non-professional readers.

Quality assurance during the design of the IT architecture and modularization according to the system specification

National Cohort IT specialists will translate the system specification document into IT language and create IT architecture modules. These modules will be updated until they are found to be in full agreement with the system specification. Where this is sensible and practical, standards such as the “Unified Modeling Language” will be used to produce formal models that can be automatically translated into code (see next section).

Quality assurance during the implementation phase

A software developer will implement the designed modules into code, taking into consideration the following quality assurance aspects:

Choosing an appropriate programming method: The quality of the resulting code will be assured by using innovative programming methods (so-called “agile software development”). The software teams will be self-organized and adapt rapidly to changing project circumstances.

Using standard interfaces to peripheral devices: The IT system will be connected with various peripheral devices that assess and deliver data into the data management system. The system will support digital connections to peripheral devices and provide interfaces to automatically read the measurements so that errors due to human input are excluded. All study centers will use identical peripheral devices.

Accounting for specific aspects of the biobanking module: The biobanking module will maintain (if possible in an automatic mode) a current record of the samples stored. For this module, it is important to create preprepared scenarios (e.g., in the form of UML sequences or activity diagrams) so that the software can reflect such situations as they occur. Examples of such situations could include the case of when a subject requests deletion of a stored sample or when a sample is sent from one study center to another.

Testing the newly developed software: Once the implementation phase has been completed, the system will be tested with respect to its functionality and documentation. Initial feasibility tests will be carried out before the software is finalized. The results of those tests will then be integrated in the final software version. Testing of the communication pipeline will be established at the level of the IT specialist, at the level of the study personnel during the feasibility studies to enable a first round of error elimination, and at the level of the study personnel during an actual recruiting situation during the pilot study. An external software certification or testing procedure will be considered for the IT system.

A.5.9.2 Quality control of data management

Once the National Cohort IT system has been installed in its final productive environment, the quality of the data items and the data flow will be continuously monitored and controlled. As the IT system is installed in different geographic locations, we present four individual perspectives to approach the main quality control issues: a general point of view, a study center point of view, an integration center point of view, and a competence unit point of view.

Quality control issues from a general point of view

- ▶ Standardization: each organizational unit works with the same version of the software and delivers data in the same way.
- ▶ Historiography: the software will allow data changes at any point. These changes will be documented. It should be possible to track every change in a history table and restore each value that each stored parameter took at different times. Recovery procedures to obtain previous versions of the data will be tested before the software is used.
- ▶ Data safety: new security breaches are discovered for older software technologies over time; therefore, it is necessary to regularly monitor the relevant sources of information about security leaks, and to adapt the software in a timely manner in order to eliminate potential vulnerabilities.
- ▶ Specification of the data flow and responsibilities to solve problems: users with dedicated roles (e.g., competence units) monitor the data flow during recruitment to detect possible interobserver variability within or across study centers.
- ▶ Flagging problems: each part of the system will be allowed to flag problems to other parts. Other parts that work with the same data will be warned about these problems. If necessary, flagged data will be banned from further processing until the flag has been resolved.
- ▶ Data versioning: all data packages that are sent from one entity to another will be versioned, including the name of the entity that sent the data, the exact date when the data were sent, and an incremental versioning number.

A.5

Quality control issues at the level of the study center

- ▶ Plausibility checks during data entry: the data being entered will be checked for completeness and plausibility. Violations of such conditions will be made evident to the examiner; where appropriate, it will be impossible to submit such data.
- ▶ Check for completeness: at the end of the input phase, the system will check all data and issue warnings if incomplete records are encountered. Whenever a value is missing, an appropriate explanation will be available.
- ▶ Data flow control: Data from study centers will be transferred to the integration centers immediately, or at least as soon as possible. Independently of the data transfer, the study centers will be able to mark data records as complete and qualitatively sound and ready for further processing or analysis.

Quality control issues at the level of the integration center

- ▶ Cross-center differences: these will be controlled on a regular basis by the data management team to ensure that changes over time (e.g., turnover of study personnel) do not impact the data.
 - ▶ Unit testing and software updates: the quality of the IT system will be maintained through software updates, including during the recruitment process. Such software updates ensure that changes in real-life procedures to correct last-minute problems are also reflected in the IT. For example, changes made to one component also affect
-

other parts of the system. For this reason, unit-testing procedures will be integrated into the software development process. These use predefined dummy inputs for all data for which results are already known. In the case of a software update, the unit tests ensure that the software still runs according to the original specifications.

- ▶ Concurrent software updates: we anticipate that most parts of the software will run on central servers. Any additional software that needs to be installed at the study centers will be updated at all centers simultaneously to avoid the situation that several centers are still working with older versions of the software. As such a scenario cannot be avoided in every instance; the new version of the software will be able to upgrade older versions of stored data and provide information regarding data that have changed. The stored data need to be tagged as to the software version that created the data.
- ▶ Data flow control: the integration center releases the data to the competence unit for data management.
- ▶ Error handling: as the integration center is the organizational unit that releases the software, it will be automatically informed of runtime errors that may occur while working with the program. The error protocol will include general information about the computer at which the error took place and the data that generated that particular error.

Quality issues at the level of the competence unit:

- ▶ Standard format: the data from the integration center will be converted into a standard format (e.g., XML) for analysis at the competence unit.

A.6 Planned statistical analyses and statistical power considerations

A.6.1 General issues and structure of the section

As stated in **Sect. A.1**, the major specific aims for the National Cohort include identifying lifestyle, nutritional, occupational, environmental, (pre)clinical, and (epi)genetic risk factors for the development of chronic diseases and aging-related declines in health and determinants of the progression of early-stage (preclinical) morbidity to overt clinical-stage disease and/or related losses of autonomy. A related aim is to develop risk prediction models and algorithms that may help identify individuals at increased risk of developing major chronic diseases. These latter models may include questionnaire assessments of risk factors, as well as genetic and other biological markers, and clinical examination data. For each of these objectives, quantitative risk associations – in terms of relative and attributable risk – will be estimated for a range of chronic disease outcomes. Finally, it is also anticipated that the cohort will be used to evaluate candidate blood or urine-based markers or clinical examinations (e.g., imaging) for early detection of disease.

The design of the National Cohort as a large-scale, population-based prospective cohort study guides the general strategy for planned statistical analyses. This has a number of direct implications:

- ▶ Whenever indicated and feasible, and contingent upon the exposure or risk factor information collected, statistical analyses will be based on data from the full cohort. If a study question of interest is only relevant for a certain subpopulation, relevant subcohorts will be considered.
- ▶ This study does not have one single hypothesis or even a limited number of hypotheses that can be taken as a single basis for confirmatory statistical testing and for sample size considerations and power calculations. Rather, research questions as outlined in **Sects. A.1** and **A.2** have been defined from diverse areas in epidemiology.

A wide range of scientific questions and related statistical analyses follow from these two general statements, which are summarized in **Sect. A.6.2**.

To structure the statistical models used for the National Cohort, we need to account for the different types of outcome that will be observed during study follow-up. Thus, we have outlined a possible series of general types of analysis and a list of selected approaches, without any claim of completeness, in **Sect. A.6.2.1**.

Within this general model (by type of outcome), the risk factors or intermediate factors presenting along the pathway from exposure to a health event will be related using a wide range of specialized models. Therefore, the directional and undirectional association and correlation structures between the different exposure variables must be incorporated in statistical analyses and the general models must be adapted accordingly. Models for exposures with a clustered structure such as environmental exposures, which require (variance) component models, for repeated measurements, and for missing values (e.g., due to losses by follow-up) must also be adapted. On the basis of the general models, some of these **model adaptations** are therefore described in **Sect. A.6.2.3**.

Calculations for the expected numbers of incident cases of several relevant diseases are outlined in **Sect. A.6.3**. These calculations are based on the source population presented in **Sect. A.3.1**. The numbers, combined with the aforementioned statistical methods, serve as examples for the general power calculations of the cohort of 200,000.

Furthermore, these figures are linked to those statistical methods for assessing statistical power. In Sect. A.6.4, therefore, we have provided some detailed examples of statistical power calculations with respect to the aforementioned structure.

A.6.2 Statistical analyses for the National Cohort

A.6.2.1 General methods for analyzing different outcomes

For the full cohort as well as for subcohorts, time-to-event methods constitute the major approach for identifying and assessing risk factors for mortality and incidence or progression of diseases. Thus, when studying a single event type, e.g., overall mortality, Cox proportional hazards regression models will typically be used to simultaneously investigate the effects of risk factors of interest while adjusting for potential confounders.

For measurements that are repeated after the 5 years of follow-up, regression models for longitudinal data will be employed. A typical example of such a model is the generalized estimating equations (GEE) approach, where the effects of various factors will be investigated while accounting for associations between measurements in the same individuals. Special attention will be paid to potentially informative drop-out or failure to obtain further measurements due to the occurrence of a specific event. These issues will be addressed by joint models for longitudinal and time-to-event data⁷⁸¹.

For disease-specific mortality or incidence of a specific disease as endpoint, we also need to account for competing risks. This will be done by using a Fine and Gray regression model⁷⁸², which summarizes the effects of risk factors on the occurrence of the event of interest while adjusting for potential confounders. Time-dependent risk factors will be considered in these approaches, too.

A.6

Table 6.1: Statistical analysis methods for the National Cohort: general examples by type of outcome variable.

Endpoint	Example	Analysis	Remarks
Binary and time-to-event	Cancer	Cox proportional hazards model	Major approach for identifying and assessing risk factors for mortality and incidence or progression of diseases
		Multistate models	Regression models for the transition hazards
		Time-varying covariate effects	Aalen's Additive Hazards Model
Multiple binary endpoints, time-to-event data	Multiple neurological outcome	Competing risk model	Fine and Gray model
Binary	Chronic infection	Conditional logistic regression	Major approach for nested case-control studies
Continuous	Blood pressure	(Generalized) linear regression	
Semi-quantitative / ordinal	Depression	Ordinal regression	Proportional odds model

For more complex outcome structures, multistate models will be employed. Adequately taking the correct timing of events into account is important in all these modeling approaches. If the time scale involves time since enrollment into the study, right censoring will be considered; if age or time since a specific event is the chosen time scale, left truncation will also be accounted for⁷⁸³.

Within these approaches different methods of confounder adjustment (general fixed and random mixed models) will be used, for example, for different association patterns of the risk factors and for risk factor–outcome relationships in cases of clustered risk factors. This is essential for clustered exposures such as environmental risk factors or even nutritional exposures, which may be clustered for substantial parts of the entire cohort.

Some general examples of these methods are outlined in **Table 6.1**.

A.6.2.2 Approaches to multicenter analysis

To account for the multicenter study design, multilevel statistical models can be used that integrate information on exposure-disease relationships on two complementary levels of observation, namely: (a) within-center relationships, which reflect the relationships at the individual level in each of the centers; and (b) a between-center relationship, which captures the association that may exist between exposure and disease risk at the aggregate level. Such multilevel models can take confounding variables fully into account, can be extended to simultaneously correct for bias in estimated exposure–disease relationships that may result from measurement errors, using data from calibration substudies, and are also well-suited for studying or identifying specific area effects on health⁷⁸⁴⁻⁷⁸⁹.

A.6.2.3 Special statistical methods

This section describes some of the statistical analysis methods that go beyond the standard statistical methods commonly used in cohort studies which were described in the previous **Sect. A.6.2.1**. These special methods will play a major role in the analysis in the National Cohort for two main reasons:

- ▶ The specific design of the study (e.g., repeated measurements for a subsample of all cohort members within less than 1 year) requires special statistical methods.
- ▶ This study will have a large public health impact. Concepts such as attributable risk or disability adjusted life-years, therefore, deserve wider attention.

A.6.2.3.1 Replicate risk factor assessments, and measurement errors

General remarks on measurement error for exposure assessment

In previous sections, the effects of random measurement errors and misclassification by exposure level were not discussed. In practice, however, measurements of risk factors or exposures generally contain a considerable degree of error. If no corrections are made for these effects, estimates of RR corresponding to measured exposure levels generally will not reflect RR levels with respect to the true underlying exposure variable⁷⁹⁰⁻⁷⁹⁴.

In unifactor analyses, assuming there are no confounding effects related to other covariates, random measurement errors in a continuous and quantitatively measured exposure variable will underestimate RRs. This statistical phenomenon is often referred to as “attenuation bias” or “regression dilution” bias⁷⁹⁵. Likewise, when a quantitative exposure variable measured with random error is dichotomized or otherwise categorized, random misclassification by exposure categories will occur, and in unifactor, nonconfounded situations RRs by exposure category will also tend to be underestimated. In multifactorial analyses, by contrast, in which confounding relationships between the multiple risk factors can be found, RRs with respect to a defined exposure measurement may be either over- or underestimated when each of the risk factors is measured with error⁷⁹⁶. In most epidemiologic studies so far, including most existing prospective cohorts, this intertwined problem has presented

a major obstacle to obtaining quantitatively accurate estimates of relative disease risks for specific exposure factors.

Calibration substudies

An extensive body of literature has been published on the design and use of “validation” or “calibration” substudies to correct for regression dilution biases in RR estimates; see, for example, Rosner et al. (1990, 1992)^{797, 798}. An important observation in this context is that “gold standard” reference measurements that are totally error-free are not required for regression calibration adjustments to be accurate, as long as a minimum set of model assumptions can be made. One such key assumption is that, conditionally on the true exposure level, replicate measurements are not correlated over time (assumption of independent random errors). This definition of random error also implies that, conditionally on true exposure level, the exposure measurements should have no association with risk of developing disease or any other “outcome” variable. Assuming independent random errors, Rosner’s multifactorial regression calibration approach produces a matrix of calibration (correction) factors that, under the above model assumptions, mathematically link observed (logistic) regression coefficients (vector) to the unbiased (“true”) regression coefficients (vector β^*), i.e.

$$\underline{\beta} \approx (\Sigma_b + \Sigma_w)^{-1} \Sigma_b \underline{\beta}^* \quad , \quad .$$

In this equation, Σ_w and Σ_b are, respectively, the within- and between-subject matrices of (co)variance for the exposure measurements.

In practice, the assumption of uncorrelated random errors may, or may not, be plausible, depending on the type of measurement considered. For example, individuals may differ strongly but relatively randomly in their tendency to systematically underestimate dietary intake levels by food frequency questionnaires, and this may result in strongly correlated errors between replicate measurements. Similar error correlations may also occur for other exposure assessments primarily based on self-reports. By contrast, the assumption of independent random errors often may be considered plausible for measurements based on laboratory assessments in blood or urine samples, or based on clinical examinations or other “objective” instruments (e.g, accelerometry), at least as long as the true exposure is defined for a comparatively short time window (e.g., <1 year). Within a comparatively short time interval, it is unlikely that important systematic changes in true exposure level occur that could be related to parallel and related changes in disease risk.

Adjustments for regression dilution biases, e.g., using the regression calibration approach, will generally entail a statistical cost in terms of a widening of the confidence interval of RR estimates^{745, 797, 798}. To limit such losses in overall precision of corrected, *versus* uncorrected, RR estimates, the adjustment factors themselves should be estimated with sufficiently high accuracy, which implies a minimal sample size for the validation/calibration substudy. Annex C.2.4 provides estimates of sample size requirements for validation/calibration studies. Based on these calculations we propose to collect repeat measurements for a representative subsample of 10,000 cohort participants, proportionally spread over all participating study centers. The overall number of 10,000 subjects may be sufficient to calibrate blood- or urine-based measurements for multiple, parallel, nested case-control studies on different disease endpoints, depending both on the future sample size of these studies and on values observed in practice for RRs and for $\rho_{m1,m2}$. To allow measurement error calibration for exposure and risk factor assessments obtained in both visits – at baseline and after 5 years – the calibration sample will be split into 6,000 subjects for the first 5-year period and 4,000 subjects for the second 5-year period (i.e., weighted by expected numbers of study participants in the first study visit [N=200,000] and second visit [up to N=150,000]).

Longer-term replicate risk factor assessments and measurements errors

For many specific exposure types, short-term variations (e.g., a 1-year interval) may reasonably be considered “random” variations around average exposure level at that time point (or within a comparatively short time window around that time point) that are nondifferential with respect to disease outcome. Over longer time intervals, however, this assumption may progressively lose validity and risk of disease may show stronger associations with risk factors assessed at a more recent time before diagnosis. For the latter type of situation, Frost and White (2005)⁷⁴² also demonstrated that a regression calibration approach to correct for (presumed “random”) errors in baseline measurements, using replicate measurements taken over long time intervals, may overcorrect and lead to biased RR estimates.

Frost and White (2005)⁷⁴² proposed a basic longitudinal (“life course”) modeling approach that provides a useful, basic framework for the statistical modeling of disease risks in relation to long-term repeat measurements in prospective cohort studies. Following that approach, a simple statistical model for the (first) two risk factor/exposure measurements planned within the National Cohort is:

$$\log(\text{risk}) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} \quad ,$$

where, for individual i , x_{i2} is the centered error-free (true) risk factor level in the latest index period of risk factor assessment, and x_{i1} is the true risk factor level in first (initial recruitment) period. In this basic model, coefficient β_2 can be interpreted as the “history-adjusted” (log) RR (or log-odds ratio) associated with an error-free risk factor (which may be expressed for a 1-unit increase), adjusted for all past error-free levels. The history-adjusted current association describes the short-term, potentially modifiable risk, measuring the difference in risk between two subjects who have different risk factor levels now but had identical levels in the past. With just two exposure periods, the model can be rewritten as

$$\log(\text{odds}) = \alpha + (\beta_1 + \beta_2)x_{i1} + \beta_2(x_{i2} - x_{i1}) \quad ,$$

which indicates that the coefficient in the above models actually estimates the change in (log)disease risk associated with the change in over time ($x_{i2}-x_{i1}$) true exposure level.

A second way of rewriting the above, basic model is:

$$\log(\text{risk}) = \alpha + \beta_L (w_1 x_{i1} + w_2 x_{i2}) \quad ,$$

where $\beta_L = \sum_{i2} \beta_i$ and $w_i = \beta_i / \beta_L$, that is, exposures assessed in visits (time periods) 1 and 2 are weighted by the strength of their log-linear relationship with risk (i.e., weighted by their respective β -coefficients). In this second model, β_L reflects the long-term association for subjects whose risk factor level has differed by 1 unit throughout the full cohort study period, and for modifiable exposures this can be interpreted as measuring the long-term potentially modifiable risk, with respect to a long-term exposure history. In the special case where $\beta_1 = \beta_2$, the latter model will provide estimates for a single regression coefficient, β_L ($= \beta_1 = \beta_2$), relating (log) disease risk to the (equally weighted) average of the exposures, x_{i1} and x_{i2} , in the two index periods. Averaging the exposures from two index periods will give a corresponding increase in statistical power and estimated accuracy (confidence interval) of RR estimates, as intraindividual (time-related) variations in exposure will be reduced by taking an integrated measure of the exposures at the two time points. However, if $\beta_1 > \beta_2$ (or *vice versa*, $\beta_1 < \beta_2$), then unequal weights will be given to the exposures in the different index periods and, depending on the degree to which the relative weighting predominantly favors the exposure at just one of the time points, x_1 or x_2 , the potential gains in statistical power by averaging exposures over time (testing the null hypothesis $\beta_L = 0$ for long-term association) will be smaller.

In practice the true exposure or risk factor levels x_{iN} will not be known exactly, but will be measured with error. Therefore, the estimated β -coefficients and relative weights of exposure $w_i = \beta_i / \beta_L$ should be corrected for regression dilution effects, e.g., following the regression calibration methods proposed by Rosner et al. Ideally, the latter ideally should be conducted independently in each of the index (exposure/risk factor assessment) periods. Therefore, of the 10,000-subject subsample for short-term calibration studies, 6,000 will be allocated to the first assessment/recruitment period and 4,000 to the second (5-year) repeat assessment period to reflect our assumption that up to about 75% of cohort participants may accept to take part in the second visit.

A.6.2.3.2 Estimation of further population based parameters using the National Cohort

It is of major interest to estimate the effect of a risk factor on the population at large in order to assess the general burden of disease for different risk factors and to increase the general public health impact of the cohort. For this purpose, attributable risks, life expectancy, and the concept of disability-adjusted life years (DALY) will be used.

Attributable risks

The population attributable risk (PAR) or, more generally, the generalized impact fraction (GIF) is an appropriate concept in this context. PAR is defined as the proportion of cases for a particular disease that can be attributed to the factor of interest,

$$\underline{PAR} = \frac{P(D) - P(D | \bar{E})}{P(D)} ,$$

where $P(D)$ denotes the general disease risk and $P(D|\bar{E})$ the disease risk for the nonexposed part of the entire population. GIF gives that same quantity if the distribution of the risk factor is changed to an alternative distribution, i.e.,

$$\underline{GIF} = \frac{P(D) - P(D^*)}{P(D)} ,$$

where $P(D^*)$ denotes the probability of the disease under the alternative distribution⁷⁹⁹.

PAR and GIF can be estimated since all quantities can be directly estimated from cohort studies. Concepts will be applied to partition the PARs and the GIFs, respectively, such that the overall PAR and the PARs can be meaningfully interpreted for each factor separately.

Life expectancy

Life expectancy is commonly estimated using age-specific mortality in the German population. The National Cohort provides a detailed estimate of the joint risk factor distribution, which can be used to calculate different scenarios for future development.

Disability-adjusted life years

DALY is a measure of overall disease burden, expressed as the number of years lost due to ill-health, disability, or early death. DALYs for a disease or injury are calculated as the sum of the years of life lost due to premature mortality (YLL) in the population and the years lost due to disability (YLD) for incident cases of the disease or injury. YLL are calculated from the number of deaths at each age multiplied by a global standard life expectancy of the age at which death occurs. YLD for a particular cause in a particular time period are estimated as:

YLD = # of incident cases in that period × average duration of the disease × disability weight.

The disability weight reflects the severity of the disease on a scale from 0 (perfect health) to 1 (death)⁸⁰⁰.

The DALY concept has been used relatively little in epidemiologic research in Germany. Terschüren et al. (2009)⁸⁰¹ estimated the future burden of disease using the population forecast with respect to age and sex distribution only. In the National Cohort, long-term observations can be used to assess the effect of changing risk factor patterns on population health.

A.6.2.4 Risk prediction models

In risk prediction modeling, a first step will be to develop a discriminative RR model that includes the relevant predictor variables. For an analysis within the full cohort, "time to disease onset" can be analyzed by means of a Cox-proportional-hazards regression. In a nested case-control setting with a (bivariate) disease outcome, a risk prediction model can be derived by means of a logistic regression model. Within such models new markers can be combined with established risk information to assess their significance and their relevance with respect to risk prediction. The general goal of risk prediction models will be to integrate the information from a potentially large series of risk factors; thus, model selection procedures such as stepwise-model selection or boosting need to be applied⁸⁰². The choice of a model selection method will depend on the specific situation regarding, for example, data quality and existence of prior, additional evidence. A relevant feature in a model-building process is "overfitting" as the effect of a statistical model to grasp aspects that are unique to the study sample. By incorporating cross-validation or bootstrapping methods into the model selection process we can keep this effect under control⁸⁰³.

An important criterion for developing a risk prediction model is its "discriminative quality." The discriminative quality of a risk prediction model can be measured in terms of the "area under the receiver operator curve" (AUROC), which summarizes the sensitivity of a risk prediction model in relation to one-minus-specificity across all possible cut-off points. For comparison of different risk models (e.g., a conventional model and an extended model including new markers), the difference in the models' AUROC values can be tested⁸⁰⁴. In the presence of established risk limits the "net reclassification gain" (NRI) can be calculated as an alternative measure to compare classification resulting from different risk models. As an extension to this idea of capture reclassification, the "integrated discrimination index" (IDI) summarizes the general improvement in discrimination. These alternative statistics provide valuable additional insight into the discriminative quality of risk models⁸⁰⁵.

Finally, a further important step in constructing risk prediction models involves translating RRs into predicted absolute risk levels. Absolute risk thresholds may be interpreted more directly in terms of potential risks and benefits if the model is used to identify subjects for interventions or treatments. With a cohort design absolute risk levels can be estimated within the study⁸⁰⁶. External data from registries can be included as an additional source. The two sources can be used to validate or optimize calibration of a discriminative model⁸⁰⁷. In addition to case numbers, the cohort can provide valuable information on the distribution of some, if not all, risk factors. To which extent the cohort is representative of the general population may have to be critically evaluated at this point.

A.6.2.5 Statistical analyses for studies embedded into the National Cohort

In situations in which data from the full cohort are not available due to financial, organizational, or other constraints, "hybrid" designs will be used. These can be properly defined

case-cohort studies and nested case-control studies. For each specific application, the particular choice of a specific design will be carefully considered and a rationale for its use will be given. Specific attention will be paid to establishing an adequate sampling scheme in order to minimize avoidable biases. For the statistical analyses, logistic regression models, if indicated, will be used in addition to the modeling approaches outlined above. Since analyses based on the nested case-control design will, in general, be the most limited in terms of statistical power, power calculations presented further in this section (**Sect. A.6.4**) will be preferentially based on this design.

Diagnostic studies will be performed as phase III studies according to the classification given by Pepe et al (2001)⁸⁰⁸. Only biomarkers or scores validated in phase II studies as defined therein will be investigated. If several markers or scores are combined, predefined methods for defining diagnostic rules such as logistic regression or tree-based methods will be used in order to determine sensitivity and specificity by resampling methods such as cross-validation.

A.6.3 Basic population profile and expected numbers of incident chronic disease cases in the National Cohort

The National Cohort will include a total of 100,000 women and 100,000 men (Level 1, with a randomly sampled subcohort of 20,000 women and 20,000 men in whom intensified phenotyping and medical examinations will be conducted (Level 2). The planned extension of the cohort with 25,000 migrants and related sample size issues are not considered here.

Both the overall cohort and the intensified subcohort will follow an age-stratified population sampling scheme that aims recruitment of 10% of cohort participants within each of the two age groups of 20–29 and 30–39 years, respectively, and about 27% in each of the three 10-year age groups between age 40 and age 69 (**Table 6.2**). The rationale for this targeted age distribution, with higher population sampling fractions for the age groups above 40 years, is that the majority of chronic diseases – heart disease, stroke, (type 2) diabetes, cancer, and dementias – show the highest incidence after this age. However, incidence for diabetes, in particular, has been increasing rapidly in the last several decades for younger adults as well, and the targeted sampling of 20% of the cohort below age 40 will make it possible to study risk factors, etiology, and possible modes for early diagnosis (e.g., OGTT in combination with blood-based, metabolic markers) of diabetes during early adulthood.

Table 6.2: Recruitment scheme: underrepresentation of younger age groups

Age group	Women		Men		Total	
	Overall	Intensified	Overall	Intensified	Overall	Intensified
20-29	10,000	2,000	10,000	2,000	20,000	4,000
30-39	10,000	2,000	10,000	2,000	20,000	4,000
40-49	26,660	5,332	26,660	5,332	53,320	10,664
50-59	26,670	5,334	26,670	5,334	53,340	10,668
60-69	26,670	5,334	26,670	5,334	53,340	10,668
TOTAL	100,000	20,000	100,000	20,000	200,000	40,000

The study sizes and targeted age distributions for the overall cohort (N=200,000) and a 20% subcohort with intensified phenotyping (N=40,000) are motivated mainly by statistical power calculations and criteria for the statistical detection of minimally relevant quantitative associations of risk for frequent chronic diseases and premature mortality with their most common causes (see **Sect. A.6.4**).

For the more frequent forms of cancer, estimates of expected incident events per disease type were obtained by applying age- and sex-specific incidence values from German cancer registries to the projected age distribution of the cohort (**Table 6.3**), adjusting for progressive attrition of the cohort as a result of overall mortality. Likewise, for MI and diabetes estimates of cumulative disease incidence were obtained, using age-specific incidence as ascertained in the MONICA/KORA registry for MIs in Augsburg as well as in the population-based MONICA/KORA cohort study. For stroke, age-specific incidence from the population-based registries of the Erlangen Stroke Project and the Ludwigshafen Stroke Study were used. As for the German population no incidence data for rheumatoid arthritis and COPD were available, data on age-specific incidence in the UK were used (**Table 6.4**).

In these various estimations of expected cumulative disease incidence, simplifying assumptions were made, in that no account was made of attrition of the cohort that will result from subjects' future requests for withdrawal from the study or losses to follow-up due to other causes. These simplifications would be expected to cause some degree of overestimation of cumulative cancer incidence. However, based on experience from ongoing studies, we can anticipate low active withdrawal rates of study participants and very high case ascertainment rates through record linkage with cancer registries or active follow-up for the other chronic diseases. Thus, the above factors are unlikely to cause more than a 5–10% overestimation of over the first 10- to 15-year follow-up period.

Table 6.3: Expected counts of incident cancer cases after 5, 10, 15, or 20 years of average follow-up, for the overall cohort (N=200,000) or for the intensified subcohort (N=40,000).*

Disease	Expected cumulative incidence at study level 1 (200,000 subjects)				Expected cumulative incidence at study level 2 (40,000 subjects)			
	Average follow-up duration (years) (+ expected corresponding calendar date)							
	5 yrs (2022)	10 yrs (2027)	15 yrs (2032)	20 yrs (2037)	5 yrs (2022)	10 yrs (2027)	15 yrs (2032)	20 yrs (2037)
Any cancer	5,100	13,000	21,000	29,000	1,000	2,600	4,000	6,000
Breast	780	1,800	2,900	4,000	160	370	590	800
Prostate	720	1,900	3,200	4,600	140	380	640	910
Colon, Rectum	670	1,800	3,100	4,500	130	360	620	890
Lung	560	1,400	2,400	3,400	110	290	480	680
Bladder	260	710	1,200	1,800	50	140	250	360
Kidney	190	500	850	1,200	40	100	170	240
non-Hodgkin L.	140	340	580	820	30	70	120	160
Pancreas	120	330	580	830	20	70	120	170
Corpus Uteri	120	320	540	770	20	60	110	200
Brain+CNS	90	200	330	450	20	40	70	90
Ovary	110	260	440	610	20	50	90	120

Calculations based on age-specific incidence from German Cancer Registry⁸⁰⁹

Table 6.4: Expected counts of noncancer incident cases after 5, 10, 15, or 20 years of average follow-up, for the total cohort (N=200,000) or the intensified subcohort (N=40,000).

Disease	Data source	Expected cumulative incidence at study level 1 (200,000 subjects)				Expected cumulative incidence at study level 2 (40,000 subjects)			
		Average follow-up duration (years) (+ expected corresponding calendar date)							
		5 yrs (2022)	10 yrs (2027)	15 yrs (2032)	20 yrs (2037)	5 yrs (2022)	10 yrs (2027)	15 yrs (2032)	20 yrs (2037)
Myocardial infarction	KORA Augsburg Registry	1,700	4,400	7,300	10,000	350	870	1,500	2,100
	MONICA/KORA study	2,400	5,800	9,400	13,000	480	1,200	1,900	2,500
Stroke	Erlangen and Ludwigshafen incidence rates	1,600	4,300	7,500	11,000	310	860	1,500	2,200
Diabetes	MONICA/KORA study	5,800	13,000	21,000	28,000	1,200	2,700	4,200	5,600
	Study of Health in Pomerania	7,700	19,000	31,000	41,000	1,600	3,800	6,100	8,100
Rheum. arthritis	Norfolk Arthritis Register	250	590	940	1,300	50	120	190	250
COPD	UK General Practice Database	2,300	5,800	9,700	13,000	460	1,160	1,900	2,700
Heart failure	Framingham Heart study and Rotterdam study	1,600	4,600	8,200	12,000	320	920	1,700	2,400
Mortality	German Federal Statistical Office	4,600	14,000	26,000	47,000	910	2,700	5,300	9,300

A.6

Figure 6.1: Expected recruitment scheme and accrual of follow-up time assuming uniform speed of subjects recruited over a 5-year period for first and second visit and assuming 75% of subjects recruited at first visit participate in the second visit.

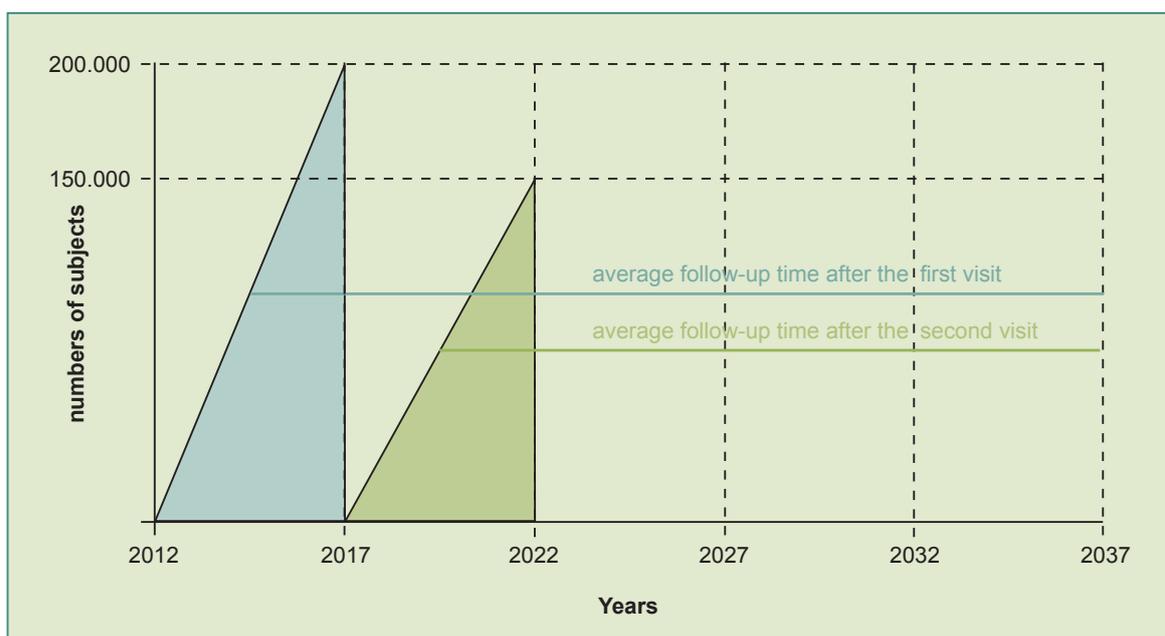


Table 6.5: Expected counts of incident cases of disease after second visit among participants presenting for the second study visit*, after 10, 15, or 20 years of follow-up**, for the full cohort (N=200,000) or for the intensified subcohort (N=40,000).

Disease	Expected cumulative incidence at study level 1 (200,000 subjects)			Expected cumulative incidence at study level 2 (40,000 subjects)		
	Average follow-up duration (years) (+ expected corresponding calendar date)					
	10 years (2027)	15 years (2032)	20 years (2037)	10 years (2027)	15 years (2032)	20 years (2037)
Any cancer	7,200	14,000	20,000	1,900	3,600	5,200
Breast	1,000	1,800	2,600	190	360	520
Prostate	1,100	2,100	3,100	210	410	620
Colon, rectum	1,000	2,000	3,000	270	530	810
Lung	790	1,500	2,300	210	410	600
Bladder	400	800	1,200	110	210	320
Kidney	270	530	800	70	140	210
Non-Hodgkin L.	190	370	540	50	100	140
Pancreas	190	370	560	50	100	150
Corpus uteri	170	340	510	30	70	100
Brain+CNS	110	200	290	30	50	80
Ovary	140	270	400	30	50	80
MI (registry)	2,400	4,700	6,900	640	1,200	1,800
MI (study)	3,100	5,900	8,500	830	1,600	2,200
Stroke	2,500	5,000	7,500	660	1,300	2,000
Diabetes (MONICA/KORA)	7,200	13,000	19,000	1,900	3,500	4,900
Diabetes (SHIP)	11,000	20,000	29,000	2,800	5,200	7,400
Rheumatoid arthritis	310	570	810	80	150	220
COPD	3,200	6,200	9,100	850	1,600	2,400
Heart failure	2,700	5,500	8,500	720	1,500	2,200

*Assuming that 75% of subjects recruited at first visit present for the second visit; ** counting follow-up time from the beginning of the cohort study

Another aspect that is difficult to account for in calculations of cumulative disease incidence is self-selection of cohort participants. Other prospective studies have often shown “healthy participant” effects, as reflected by lower incidence of chronic diseases than expected on the basis of general population registry data. Possible impact of self-selection on cumulative incidence is addressed in **Annex C.2.1**, where we compare cumulative chronic disease incidence estimated from national registry data or from current observations in the combined German EPIC cohorts in Heidelberg and Potsdam. These comparisons show that, to a variable degree, healthy participant effects may indeed result in cumulative disease incidences that can be either up to 30% lower, but for some diseases (e.g., breast cancer, prostate cancer) can also be up to 30% higher than predicted from national incidence rates.

A further important consideration in practice is that the duration of follow-up periods mentioned in **Tables 6.3** and **6.4**, set at 5, 10, 15, or 20 years for calculating expected cumulative disease incidences, will not correspond to an equal duration of project time since start of recruitment. Study plans foresee a 5-year period to complete a first “baseline” recruitment of cohort members. Assuming a uniform speed of subject recruitment into the cohort over a 5-year period, an average of only 2.5 years of follow-up time will accrue during this

period (see **Figure 6.1**). In addition, we anticipate that about 2.5 additional years of time will be needed to reach complete ascertainment of chronic disease cases, either through active follow-up or through record linkage with cancer registries or other health information systems. We thus anticipate that, for example, a 10-year average follow-up duration will be reached at about 15 years after initiation of recruitment – that is, with an overall delay of about 5 years. **Table 6.4** shows the calendar dates on which case ascertainment can be expected to be complete for (median) follow-up times of 5, 10, 15, and 20 years, if recruitment of the cohort starts in 2012.

An important part of the study protocol involves reinviting the full study cohort for a second visit after 5 years, at which time a second blood and urine sample will be collected and repeat measurement will be taken for a series of clinical and other measurements. The repeat measurements will be used to monitor changes in risk factors or (pre-)clinical conditions, and to examine possible relationships of these changes with subsequent risks of chronic diseases (see also **Sect. 6.4.2**, below). **Table 6.5** shows the expected numbers of incident cases of chronic disease diagnosed after the second measurement among the 75% of cohort participants that we anticipate will agree to participate in the second round of investigation.

A.6.4 Statistical power considerations of the National Cohort

A.6.4.1 General remarks on power calculations in cohort studies

In Cox proportional hazards analyses of full cohort data and also in the statistical analyses of hybrid (nested case-cohort or case-control) study designs, the statistical power for tests of association of disease risks with specific risk factors or determinants and precision of RR estimates crucially depend on the numbers of incident cases of disease. Furthermore, in the hybrid study designs, a further determinant of the statistical power will be the number of nondiseased study participants with whom exposure or risk factor information from the cases can be compared.

In **Sects. A.6.4.2** and **A.6.4.3** (below), and in **Annex C.2.2**, we present graphs and tables that show statistical power or minimally detectable effects for nested case-control studies with control-to-case ratios of 1:1, 2:1, or 4:1.

In nested case-control studies, and for statistical analyses focusing on main effects for a given exposure or risk factor, the effective sample size of a study will asymptotically double when the ratio of control subjects to case subjects is increased from 1:1 to infinity⁸¹⁰. The greatest single-step gain in effective study size will be obtained, however, when the control-to-case ratio is increased from 1:1 to 2:1. Beyond a ratio of 4:1, the incremental increases in effective study size and statistical power for main effects analyses generally become too small to motivate further investments in exposure measurements for additional controls. **Calculations for nested case-control studies with a 4:1 control-to-case ratio thus also provide only slightly conservative estimates for the statistical power in full-cohort settings, where numbers of incident cases of disease generally are smaller than numbers of nondiseased individuals.**

Estimates are made with different levels of statistical significance, corresponding to situations in which only a single test is performed ($\alpha = 0.05$), a few tests are performed simultaneously ($\alpha = 0.01$), or a large number of tests are performed simultaneously ($\alpha = 10^{-4}$). The latter situation corresponds, for example, to multiple tests performed when hypothesis-free searches are made for proteomic or metabolomic markers of disease risk.

All power calculations in **Sects. A.6.4.2** and **A.6.4.3** are based on the following assumptions:

- ▶ Possible effects of exposure measurement errors are ignored. (**Sect. A.6.2.3.1**, however, addresses the issue of random errors, particularly for continuous exposure measurements, and presents a strategy for reproducibility/calibration substudies embedded in the National Cohort, to improve quantitative estimation of RRs, adjusting for regression dilution effects due to random measurement errors.)
- ▶ The possibility of missing data was ignored. While imputation methods will be used in the analysis when necessary, the effects on power can approximately be translated with the percentage of missing values for a given covariable⁸¹¹.
- ▶ No confounding is assumed. However, the variance inflation factor (VIF), which is discussed below, directly translates into the factor for the sample size.

In addition to sample size requirements for detecting RRs, **Sect. A.6.4.5** presents estimates of sample size requirements for nested case-control studies aiming to estimate sensitivity of a continuous diagnostic marker at a prefixed level of specificity.

A.6.4.2 Minimally detectable odds ratios in main effects models

Exposures or risk factors are frequently measured or coded as binary, multiple categorical, or continuous variables. In nested case-control studies, typical examples of binary exposure variables based on laboratory measurements include classifying individuals into carriers or noncarriers of a given infectious agent, or carrier status for a given (polymorphic) genetic risk allele. In cohort studies, a typical example of binary risk factor classifications is a comparison between use versus nonuse of exogenous hormones or medical drugs. Examples of continuously measured risk factors include many biomarkers (e.g., markers of metabolism) measured in blood or urine samples or quantitative scores based a larger number of questionnaire responses (e.g., nutrient intakes estimated from a food frequency questionnaire). In a first step, continuous exposure variables may also be broken down into quantile categories and RRs (or odds ratios) estimated for each quantile, taking one of the quantile categories as reference. A statistical instrument for the latter situation is a categorical model, counting events in cells of a multidimensional contingency table and assuming an order between the categories of the new quantile variable. This latter approach can thus be regarded as a general and robust method that should be followed by dose–response analyses⁸¹².

For continuous exposure or risk factor measurements, it is often appropriate to test the hypothesis that increasing levels of the exposure are associated with increases in risk on a continuous (log-)linear scale. The statistical power of such tests is based on a *standardized difference (A)*. Definition and interpretation of standardized difference are given in **Annex C.2.2**.

For continuous exposure/risk factor variables that follow skewed (e.g., approximately log-normal) probability distributions or more complex distributional types, including a spike at zero situation (e.g., cigarette smoking or alcohol consumption), more complex modeling strategies, such as fractional polynomials, may be needed⁸¹³. Rigorous sample size or statistical power calculations for these more complex situations are not being considered here.

In multifactorial statistical models that include a number of covariates as adjustment or confounding factors, the estimates of sample size requirements will increase, compared to the sample size needed for nonadjusted analyses. A general approximation to this increase in sample size requirements is given by the VIF⁸¹⁴:

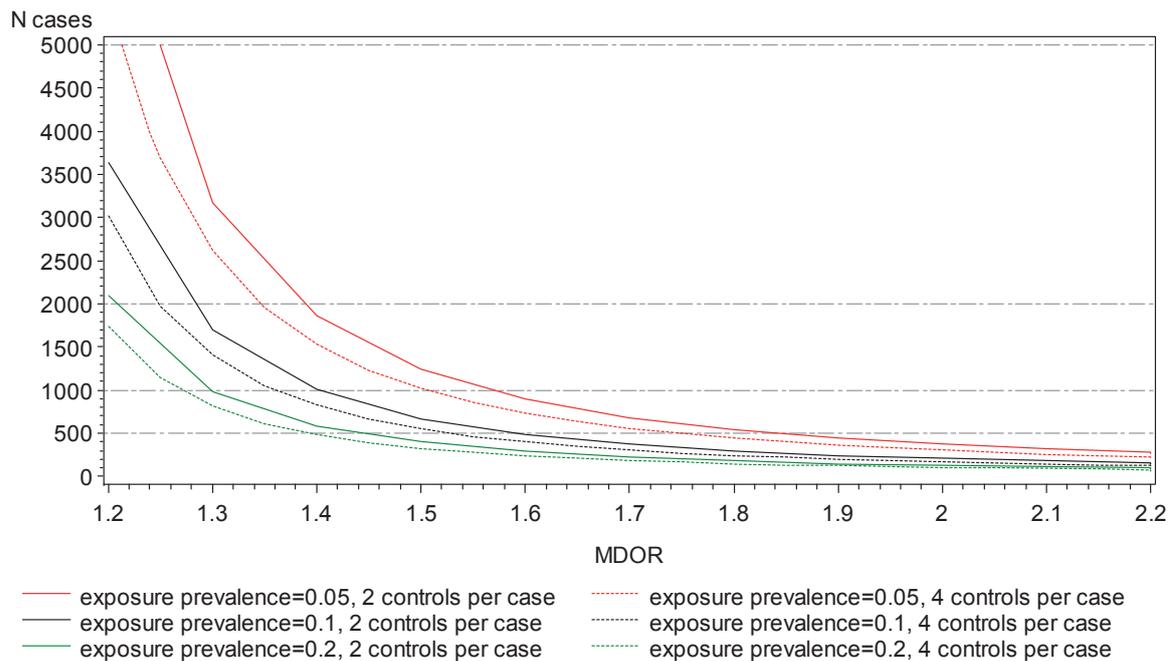
$$VIF = \left(\frac{1}{1 - \rho_{CE}^2} \right) ,$$

where ρ_{CE}^2 is the squared multiple correlation of the exposure variable (E) with the other covariates (C) included in the statistical model. The VIF helps examine the effects of covariate adjustments on statistical power and minimally detectable RRs. Furthermore, the required study size for detecting a minimal odds ratio [MDOR] with minimal power and for main effects analyses increases proportionally to the VIF. **Tables S6.5** in **Annex C.2.2** provide estimates of sample size requirements for situations both with and without confounding adjustments, assuming different values for the VIF.

Figures 6.2 to 6.4 show the study size needed for nested case-control studies to be able to detect a MDOR or standardized exposure differences of various magnitudes with 0.80 power and variable significance levels for exposure/risk factor variables that follow a binary distribution, distributed into quartile levels, or continuously (and approximately normally) distributed. These estimates were calculated with standard software, in particular the programs “QUANTO”⁸¹⁵ and “POWER” from the US National Cancer Institute⁸¹⁶. Additional detailed supplementary tables (**Tables S6.2** and **S6.3**) showing the MDOR that can be detected at various study sizes are also presented in **Annex C.2.2**. In addition, **Annex C.2.2 (supplementary Figure S6.1 and supplementary Table S6.6)** also provides estimates of the MDOR (power of 0.80, at significance levels 0.05 or 10^{-4}) associated with carrier status for a given genetic (dominant) allele, as a function of number of disease cases (with control:case ratios of 2:1 or 4:1) and allele frequency.

A.6

Figure 6.2: Study size (number of disease cases, with 2 or 4 controls per case) required for detection of a minimum odds ratio [MDOR] associated with a binary exposure variable and with statistical power 0.80 and significance level 0.05



alpha=0.05, overall disease prevalence 1%

Figure 6.3: Study size (number of disease cases, with 2 or 4 controls each) required for detection of a minimum odds ratio [MDOR] varying from 1.2 to 2.5, between top to bottom quartiles of a continuous exposure variable (statistical Power 0.80, at significant levels of 0.05 or 10^{-4})

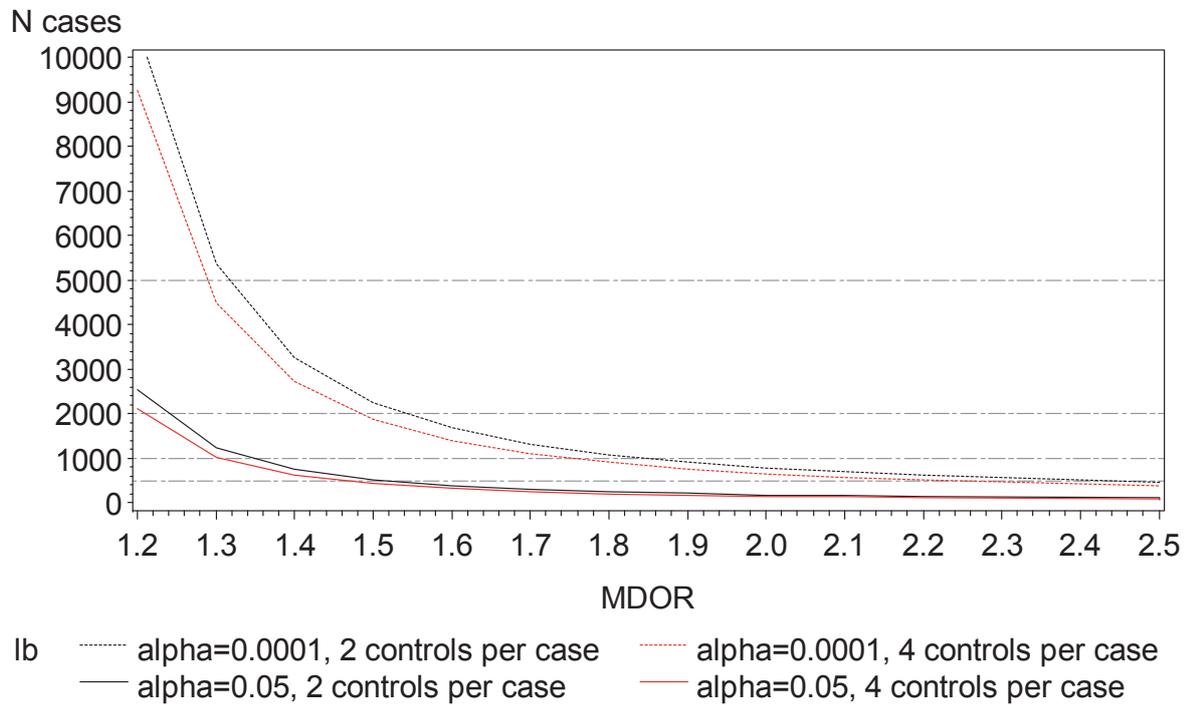
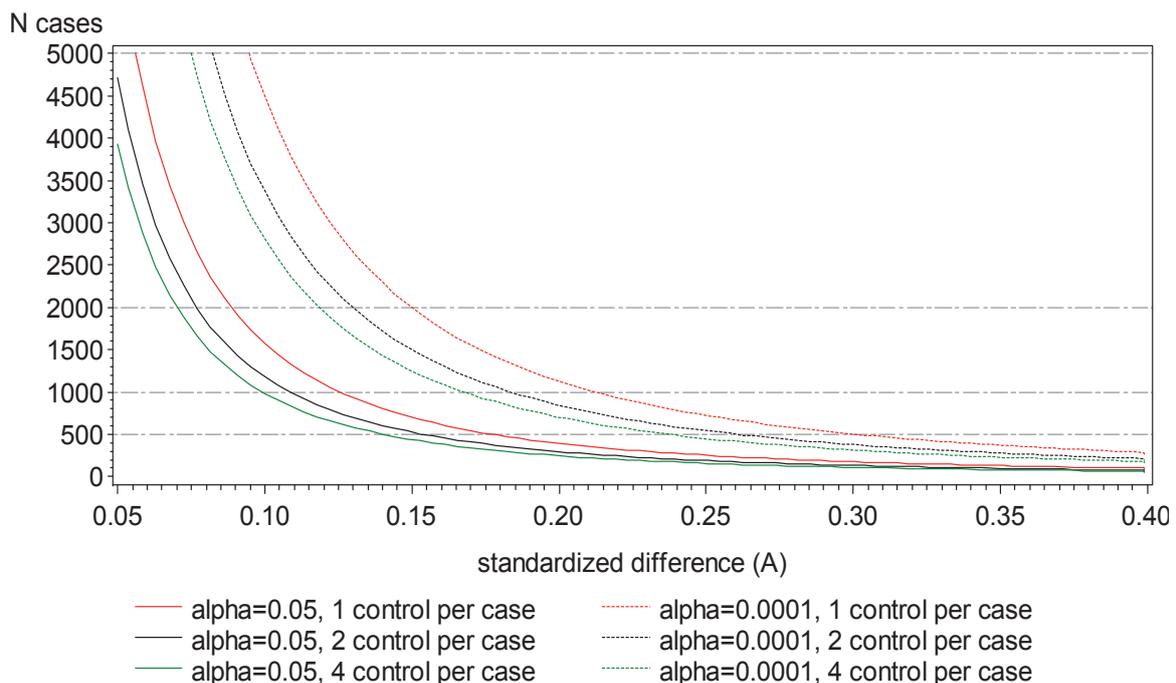


Figure 6.4: Study size (number of disease cases, with 1, 2, or 4 controls each) required to detect a minimal standardized difference A, in a continuous exposure variable* (statistical power 0.80, at significance levels of 0.05 or 10^{-4})



* Normal distributed covariable, equal variance in cases and controls

Figures 6.2 and 6.3 (and **supplementary Tables S6.2 and S6.3**) indicate that studies which include at least 300–500 cases of a specific disease, with control-to-case ratios varying between 2:1 and 4:1, will have a statistical power of 0.80 (at a significance level of 0.05) to detect an odds ratio of about 1.4–1.6 for a binary exposure with 20% population prevalence, or an odds ratio of about 1.5–1.7 for top vs. bottom quartile categories of an exposure or risk factor in the controls. For exposures or risk factors measured on a continuous scale, and which are approximately normally distributed, we will be able to detect a standardized exposure difference (A) of about 0.14–0.20, again at a significance level of 0.05 (**Figure 6.4, supplementary Table S6.5**) with a study of this size. The number of 300–500 incident cases of disease corresponds, for example, to the rarer forms of cancer listed in **Table 6.3** or to incident cases of type 2 diabetes among adults <40 years of age in the subcohort with intensified phenotyping.

With 1,000–2,500 cases, as obtained for more frequent diseases at both study levels after 10 years, a statistical power (at significance level 0.05) of 0.80 or higher is given to detect standardized differences of 0.06–0.10 – a mean exposure difference corresponding to an expected odds ratio of around 1.15 to 1.29 between the extreme quartiles. Alternatively, when strong confounding factors are included, with $\rho^2_{CE} = 0.50$ (i.e., VIF = 2.0), mean standardized differences of about 0.10–0.15 can be detected, corresponding to an expected odds ratio of around 1.29–1.46 between the extreme quartiles. In the full cohort (N=200,000), 1,000–2,500 incident cases are expected to be observed for the more frequent cancer types, and in the intensified subcohort (N=40,000) similar case numbers are also expected for diabetes and common CVD endpoints.

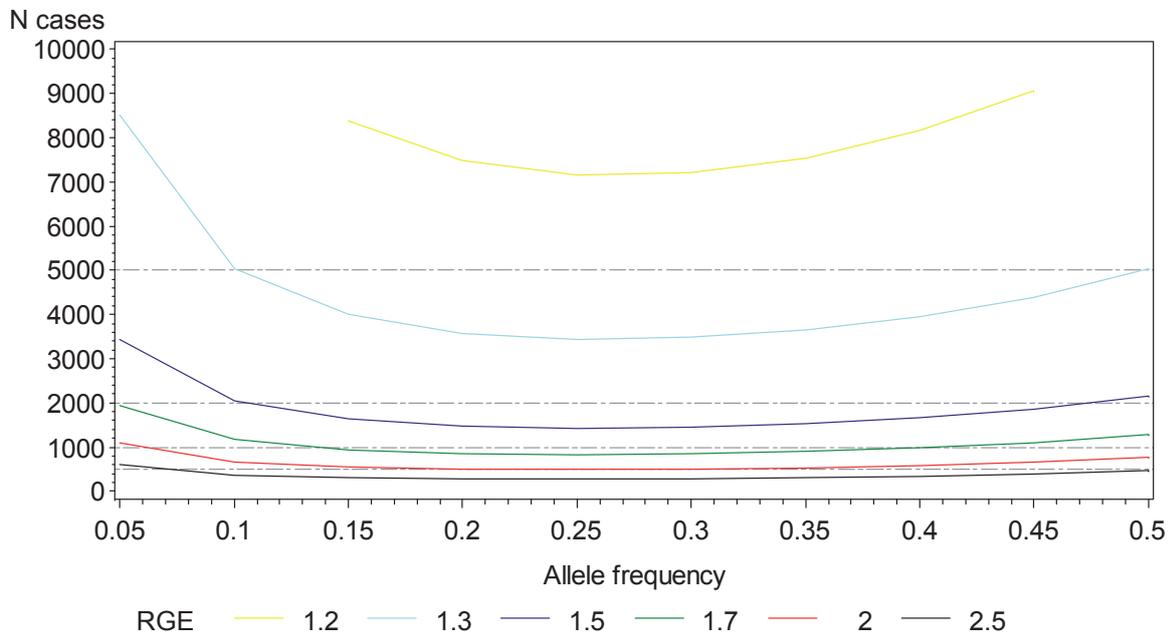
At still higher levels of incidence, with 5,000 events and more, the MDOR for a binary exposure factor with a 20% population prevalence decreases to 1.13 and lower. Binary exposures corresponding to the latter type of prevalence and odds ratio could be, for example, carrier status for a genetic risk allele. By including strong confounding factors (i.e., $\rho^2_{CE} = 0.50$, VIF = 2.0), a MDOR of around 1.18–1.2 can be detected for a binary exposure factor with a 20% population prevalence. This highest level of incident events will include diabetes, frequent forms of CVD, and overall mortality, in particular, within the full study cohort.

A.6.4.3 Interaction effects between genetic and nongenetic (“environmental”) risk factors

Figure 6.5 shows the MDOR for gene–environment interaction (power 0.80, at significance level of 0.05), in a nested case-control study with N cases of disease, and control:case ratios of 4:1, at fixed (dominant) main effect odds ratios of 1.2 for the genetic variant, and a main effects odds ratio of 2.0 for the environmental factor. Here, the odds ratio for interaction is defined as a departure from the null model which, assuming a multiplicative model for combined effects, predicts an odds ratio of 2.4 (1.2 x 2.0) for subjects exposed to both the genetic and nongenetic factors. Further estimates, under varying model assumptions, are given in **Annex C.2.3, Figure S6.2, and supplementary Table S6.7**. As indicated by **Figure 6.5** and the supplementary tables, 1,500–5,000 cases of disease will generally be required to detect an odds ratio for interaction of 1.5 or stronger, depending on allele frequency and on prevalence of the nongenetic risk factors. In the National Cohort, this corresponds to expected numbers of cases for the most frequent diseases, including diabetes, MI, stroke, and cancers of the breast and prostate.

Similar calculations, but for a more rigorous significance level of 10⁻⁴, are presented in **Sect. C.2.3**.

Figure 6.5: Study size (number of incident cases of disease) required to detect minimal odds ratios [MDOR] of various magnitudes, for gene–environment interaction with power 0.80, at significance level of 0.05; nested case-control study with 4 controls per case*, assuming main effects odds ratios of 1.2 for the genetic and 2.0 for the nongenetic risk factor, plus a prevalence of the nongenetic risk factor of 0.20



*assuming main effects odds ratios of 1.2 for genetic and 2.0 for non-genetic risk factor
alpha=0.05, prevalence of non-genetic risk factor 20%

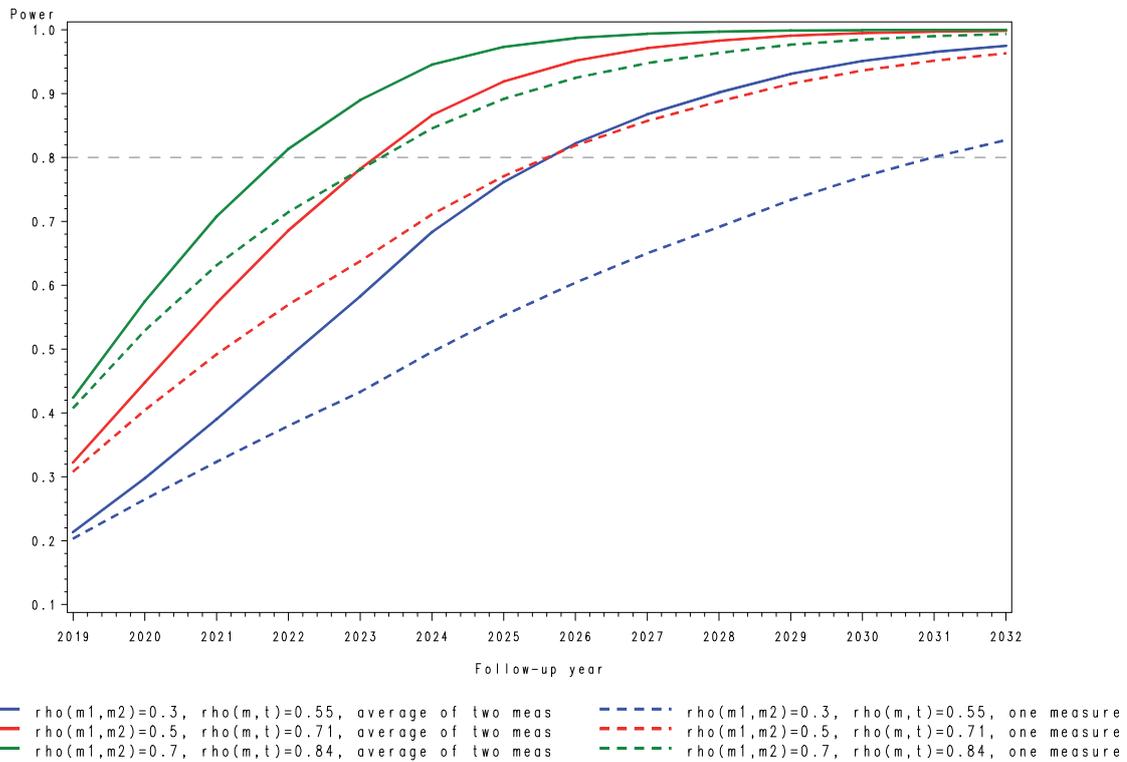
A.6.4.4 Effects of random measurement errors and use of repeat measurements

The degree of correlation between exposures in the two time periods, $\rho_{m1,m2}$, is an important determinant of the power of possible tests of “history-adjusted” exposure associations with disease risk. For example, when correlation $\rho_{m1,m2}$ is as high as 0.80, a test whether risk is related to exposure in the second index period, adjusted for exposure at the earlier period (i.e., testing the null hypothesis that $\beta_2 = 0$, conditionally on exposure level x_1) will have a high VIF (VIF = 2.8), and a corresponding reduction in statistical power. Likewise, a comparatively high correlation $\rho_{m1,m2}$ will also limit possible power gains that could be achieved by basing the test of exposure–disease association on a weighted average of the exposures (i.e., testing the null hypothesis that $\beta_L = 0$). For exposures (e.g., blood- or urine-based markers) that have a correlation of $\rho_{m1,m2} = 0.30$ – 0.70 between replicate measurements taken over time, a theoretical (maximal) average gain in effective sample size of about 40–60% can be calculated, under the assumption that random errors of the replicate measurements are not correlated and that disease risk is equally associated with the exposure measurements at either time point (see **Annex C.2.4**).

Figure 6.6 illustrates both relative and absolute gains in statistical power that can be obtained by (equally) averaging exposures from two time periods. This specific example focuses on kidney cancer, showing how statistical power will increase with increasing numbers of incident cases until the year 2032. The assumed strength of association is that of

Figure 6.6: Gain in statistical power when, for 75% of the cohort, replicate measurements taken 5 years apart are combined into an average exposure assessment, as compared to a situation with only one baseline measurement available.

Example of a nested case-control study with cases based on expected number of cases for kidney cancer, one control per case, with a true standardized exposure difference between cases and controls corresponding to an expected odds ratio of 1.8 between the extreme quartiles, and assuming 2 years' delay in case ascertainment.



Example of kidney cancer, OR[Q4-Q1]=1.8

a standardized continuous exposure difference corresponding to an expected odds ratio of 1.8 between the extreme quartiles for a given exposure, in each of the two time windows corresponding to the first and second visits of study subjects in the National Cohort. **Figure 7.6** also demonstrates the greater relative increase in statistical power in the lower range of correlation ($\rho_{m1,m2} = 0.30$) as compared to the higher range ($\rho_{m1,m2} = 0.50$). The figure also shows that, only with a correlation between replicate measurements of at least $\rho_{m1,m2} \geq 0.30$ can sufficient absolute power be reached to detect a standardized difference of the above magnitude if the nested case-control study includes 1,300 cases or fewer. In **Sect. C.2.4, supplementary Table S6.9** presents similar calculations for a wider range of correlations and sample sizes for nested case-control studies (of various possible disease outcomes). These additional tables show that, when the correlation $\rho_{m1,m2}$ between replicate measurements is lower than 0.3 (i.e., $\rho_{m,x} = 0.55$), then any further decrease in this correlation coefficient causes absolute statistical power values to quickly drop off to levels at which meaningful analyses of exposure–disease relationships will no longer be possible.

A.6.4.5 Statistical power for studies of sensitivity and specificity of diagnostic markers for early detection

The National Cohort's biorepository of blood (serum, plasma, and intact white blood cells) and urine samples will provide a valuable resource for validating blood-based, candidate biomarkers for the early detection of specific forms of chronic disease. Due to the prospective study design, it is possible to examine whether a given biomarker (or combination of markers) can predict disease occurrence ahead of its "natural" date at diagnosis (estimation of "lead time"). Furthermore, the prospective study design helps avoid biases in the estimation of sensitivity and specificity of diagnostic markers that can occur when such estimations are based on comparisons between cases with advanced (clinically detectable) disease and control subjects (for example, identification of nonspecific markers that reflect a general inflammatory response to a tumor). In addition to blood- or urine-based biomarkers, imaging methods (e.g., measures of brain structures) for early diagnosis and prediction of clinical disease occurrence (e.g., specific forms of dementia) can also be prospectively evaluated in the National Cohort.

In most situations, we anticipate that markers for early detection will be measured on a continuous scale. Concerning sample size calculation, to guarantee an appropriate amount of specificity, the minimally acceptable false-positive rate [FPR_0] must first be fixed. Then we test whether the marker provides a minimally acceptable level (TPR_0) of true positive detection rates (sensitivity). Sample size requirements can then be calculated corresponding to tests with sufficient statistical power to detect disease with TPR_0 , assuming alternative true positive rates TPR_1 ⁸¹⁷.

For relatively rare forms of disease, very high rates of specificity are needed (compared to the number of true-positive cases) to avoid a massive excess of invasive diagnostic work-up in individuals who are in reality free of the disease. Depending on the prevalence of disease, this generally implies specificity levels of 99% or higher. However, the specificity criterion can be relaxed if the marker is to be used for a first screening, to be followed by a second, still relatively noninvasive screening tool with which specificity can be further increased. For example, in the case of ovarian cancer, blood-based markers could be used as a first screening, followed by transvaginal echography as a second-level screening tool⁸¹⁸. In this type of situation, a novel screening marker may be useful, for example, when it can detect disease up to 2 years in advance of its natural diagnosis, with a specificity of 98% and TPR_0 of 20%.

Table 6.7 shows the numbers of cases with disease required for the statistical evaluation of a candidate diagnostic test that has a specificity of 0.95, 0.98, or 0.995 and achieving 80% statistical power in a one-sided statistical test with a significance level of 0.05, testing the null hypothesis that sensitivity is less than TPR_0 , against the alternative hypothesis that sensitivity is at least TPR_0 . The table indicates, for example, that for a diagnostic continuous marker test for which a specificity threshold is set at 0.98, a statistical test that the diagnostic tool has at least 20% sensitivity, when TPR_1 is at least 0.40, will require a nested case-control study of at least 57 cases of disease and 570 control subjects for comparison. Although a targeted level of specificity (FPR_0) must be fixed, the calculation method used for **Table 6.7** also includes an estimation step for specificity, using a prespecified tolerance (constant ϵ) for the degree of accuracy of this estimate. In our calculations, the tolerance ϵ was fixed at 0.005. As reflected by the data in Table 6.7, sample size requirements generally decrease with improving sensitivity – that is, the higher the FPR is expected to be, the higher the number of cases required to establish a true-positive rate (TPR) at TPR_0 . Furthermore, sample size requirements increase when the underlying TPR_1 is closer to the estimated TPR_0 . With regard to the FPR, which should be at a level of maximally FPR_0 (plus the tolerance constant ϵ) particularly the number of control subjects must be large enough

for correct estimation of FPR. Thus, as the underlying TPR_1 is further away from the targeted TPR_0 , a smaller number of cases would be needed, but at some point the numbers of cases and controls are kept constant by the demand of securing accurate joint estimation of FPR_0 (shaded areas in **Table 6.7**).

In practice, the variance of a diagnostic marker may differ between subjects developing disease (cases) and control subjects who remain disease-free and often will show wider spread among the cases. As the FPR is estimated from control subjects, a smaller marker variance among controls, compared to that among cases, will generally improve estimation of FPR_0 . Thus, comparing the situation in which biomarker variances for cases and controls are equal to the one in which variance of biomarker among control patients is half of the biomarker variance of cases, smaller numbers of cases are needed in the latter situation.

To be able to study the validity of candidate markers within a prospective cohort, the cohort should have accumulated certain numbers of prospectively identified (“incident”) cases of disease, detected within a reasonably short period of the blood sample collection. In general, a diagnostic marker, e.g., for cancer, may be able to detect a disease only within a limited time span (often no more than 1–2 years) before its natural diagnosis. **Table 6.6** indicates numbers of cancer cases occurring within 1, 2, or 3 years after visit to a study center and collection of biological specimens, assuming either a single recruitment visit for the full cohort or also a second visit.

A.6

Table 6.6: Expected counts of incident cancer cases within 1, 2, or 3 years after blood donation

	Number of cases after baseline blood donation from full cohort			Baseline blood sample from full cohort, and second blood sample from 75% of participants		
	within 1 year	within 2 years	within 3 years	within 1 year	within 2 years	within 3 years
Prostate	182	365	547	319	639	958
Breast	161	322	483	282	563	845
Colon	88	177	265	155	309	464
Lung	59	117	176	103	205	308
Kidney	34	68	101	59	118	177
Bladder	24	49	73	43	85	128
Pancreas	19	38	57	33	66	99
Brain	18	36	54	32	63	95
Non-Hodgkin	17	35	52	30	61	91
Corpus Uteri	13	27	40	24	47	71
Ovary	12	25	37	22	43	65

*3-year incident cases calculated as 0.2-fold 10-year incidence from EPIC-HD-sample; 2-year incident cases as 2/3 of cases in 3 years and 1-year incidence cases as 1/3 of cases in 3 years.

Table 6.7: Number of cases and controls required for nested case-control studies aiming to test for minimal sensitivity, at pre-fixed specificity, of a continuous diagnostic biomarker or test*

TPR ₀		var(case biomarker)/var(control biomarker) = 1.0												var(case biomarker)/var(control biomarker) = 2.0											
		Specificity (1-FPR)						Specificity (1-FPR)						Specificity (1-FPR)						Specificity (1-FPR)					
		0.995 (1-0.005)		0.98 (1-0.02)		0.95 (1-0.05)		0.995 (1-0.005)		0.98 (1-0.02)		0.95 (1-0.05)		0.995 (1-0.005)		0.98 (1-0.02)		0.95 (1-0.05)							
TPR ₁	2	4	10	2	4	10	2	4	10	2	4	10	2	4	10	2	4	10	2	4	10				
0.20	3435	1950	1058	1508	986	673	1022	743	576	1207	836	613	725	595	516	673	534	492	673	534	492				
0.30	1020	575	308	443	287	193	673	337	164	353	241	175	278	169	146	673	337	139	673	337	139				
0.35	512	287	153	278	142	95	673	337	135	175	119	85	278	139	71	673	337	135	673	337	135				
0.40	312	175	92	278	139	57	673	337	135	106	72	51	278	139	56	673	337	135	673	337	135				
0.45	209	117	62	278	139	56	673	337	135	71	48	34	278	139	56	673	337	135	673	337	135				
0.50	148	83	44	278	139	56	673	337	135	71	36	24	278	139	56	673	337	135	673	337	135				
0.55	107	60	32	278	139	56	673	337	135	71	36	18	278	139	56	673	337	135	673	337	135				
0.60	78	44	23	278	139	56	673	337	135	71	36	15	278	139	56	673	337	135	673	337	135				
0.65	71	36	17	278	139	56	673	337	135	71	36	15	278	139	56	673	337	135	673	337	135				
0.70	71	36	15	278	139	56	673	337	135	71	36	15	278	139	56	673	337	135	673	337	135				
0.75	71	36	15	278	139	56	673	337	135	71	36	15	278	139	56	673	337	135	673	337	135				

A.6

TPR ₀		var(case biomarker)/var(control biomarker) = 1.0												var(case biomarker)/var(control biomarker) = 2.0													
		Specificity (1-FPR)						Specificity (1-FPR)						Specificity (1-FPR)						Specificity (1-FPR)							
		0.995 (1-0.005)			0.98 (1-0.02)			0.95 (1-0.05)			0.995 (1-0.005)			0.98 (1-0.02)			0.95 (1-0.05)			0.995 (1-0.005)			0.98 (1-0.02)			0.95 (1-0.05)	
control to case ratio		control to case ratio		control to case ratio		control to case ratio		control to case ratio		control to case ratio		control to case ratio		control to case ratio		control to case ratio		control to case ratio		control to case ratio		control to case ratio		control to case ratio		control to case ratio	
2	4	10	2	4	10	2	4	10	2	4	10	2	4	10	2	4	10	2	4	10	2	4	10	2	4	10	
0.50	5222	2917	1534	2232	1422	936	1478	1045	786	1765	1189	843	1017	815	694	1765	1189	843	1017	815	694	1765	1189	843	1017	815	694
0.60	1247	698	368	535	342	226	673	337	190	423	286	204	278	197	168	423	286	204	278	197	168	423	286	204	278	197	168
0.65	512	287	153	278	142	95	673	337	135	175	119	85	278	139	71	175	119	85	278	139	71	175	119	85	278	139	71
0.70	255	144	77	278	139	56	673	337	135	89	61	44	278	139	56	89	61	44	278	139	56	89	61	44	278	139	56
0.75	138	78	43	278	139	56	673	337	135	71	36	25	278	139	56	71	36	25	278	139	56	71	36	25	278	139	56
0.80	76	44	24	278	139	56	673	337	135	71	36	15	278	139	56	71	36	15	278	139	56	71	36	15	278	139	56
0.85	71	36	15	278	139	56	673	337	135	71	36	15	278	139	56	71	36	15	278	139	56	71	36	15	278	139	56
0.90	71	36	15	278	139	56	673	337	135	71	36	15	278	139	56	71	36	15	278	139	56	71	36	15	278	139	56
0.80	1915	1116	636	878	597	428	673	466	376	716	516	396	456	386	344	716	516	396	456	386	344	716	516	396	456	386	344
0.90	283	169	101	278	139	72	673	337	135	113	84	67	278	139	60	113	84	67	278	139	60	113	84	67	278	139	60
0.95	71	36	21	278	139	56	673	337	135	71	36	15	278	139	56	71	36	15	278	139	56	71	36	15	278	139	56

* Highlighted are the number of cases kept constant to at the level that is determined by the demand of securing FPR₀.

A.7 Ethical aspects

The questions of ethics and data protection in the National Cohort have been discussed with the working group of data protection officers (“Arbeitskreis Wissenschaft der Datenschutzbeauftragten des Bundes und der Länder”) to develop a concept which is acceptable for all federal states. The following section describes the current status. Consensus has been reached for most points; however, some questions are still open and some changes may be necessary in the final version.

The study protocol for feasibility studies within the framework of the National Cohort has a similar structure but is simpler. It was first submitted to one ethics committee, which approved it. This approved protocol has been sent to the other ethics committees involved and we expect to also receive their approval in January 2011.

A.7.1 Invitation letters and informed consent

A.7.1.1 Information for potential participants

Potential participants will receive, by postal mail, an invitation letter as well as general information about the National Cohort. The invitation strategy is described in detail in **Sect. A.3.1.4**.

The invitation letter will be an official letter signed by the principal investigators and the coordinators of the clusters. It will contain a brief description of the purpose of the study and involved institutions and funders and outline the intended research program, including the time frame. A general information flyer will be sent along with the invitation and will contain general information about the project and the aim of the study. Furthermore, the invitation will contain a prepaid postal reply for the participants to make an appointment at the study center and advice on how to contact the National Cohort team and how to take part in the study. Letters of support from the Federal Ministry of Research and Education, the Federal Ministry of Health and local stakeholders are also attached.

Furthermore, participants will receive a detailed study information leaflet. This leaflet contains detailed information about the project, the aim of the study, funders, organization, conditions for participation, research program (personal interview, questionnaire, physical examination, and collection of blood and other biomaterials), risks due to participation (e.g., due to collection of blood samples, etc.), voluntariness, right to withdraw, recontact, currently intended and possible further analyses (including genetic analyses), storage of samples, communication of personal results, and confidentiality and data protection. We will confirm that the highest level of data protection will be ensured in accordance with all relevant legislations. Potential participants for MRI will receive detailed specific information describing the procedure of feedback in case of incidental findings.

Both documents, the invitation letter and the information leaflet, will be designed with the greatest possible care since they will be of particular importance for providing the participants with information and for developing positive attitudes and motivation to participate. For further information or concerns, a telephone service and a website are provided. Further questions will be answered personally at the study center.

A.7.1.2 Informed consent

Since it will be impossible to foresee all future research projects, consent will be sought for the broadest possible range of future medical research proposals, including the use of bio-

logical/genetic markers measured in blood samples and other biomaterials, with full respect of current laws and regulations for human data protection. It will be endeavored to ensure that participants understand what they are consenting to. This will be done by training the National Cohort personnel well in communicating the information and assessing the participants' understanding and to give further advice and clarification if necessary.

Participants will be requested to give written consent:

- ▶ That they have read and understood the information leaflet and that they had the possibility to ask further questions and that the appropriate answers were understood.
- ▶ To participate in the entire research program (personal interview, questionnaire, medical examination, and collection of blood and other biomaterials).
- ▶ To be recontacted (via mail, telephone, or internet) and reinterviewed (and thus re-identified) at regular intervals during the follow-up period to provide information about changes in lifestyle or other risk factors and incident diseases and to be invited to repeated examinations.
- ▶ That personal identifying data will be recorded strictly separated from study data and biosamples and will only be used for the purpose of recontacting. Home and work address will be used to generate scientific data concerning environmental exposures such as outdoor air quality, noise, etc. These data will only be analyzed in coded form.
- ▶ That their data and biosamples will be stored in the long-term for future research, including tests of new methods or extension of the initial research proposals due to new techniques or increased knowledge/new insights that had not been defined or were unknown at the time of recruitment. A new request of consent will not be necessary in such a case.
- ▶ That their data and biosamples can be used to analyze genetic risk factors for chronic diseases.
- ▶ That for future scientific analyses there may be collaborations with universities and research institutes (national and international). After ethical approval, coded data for medical research could also be transferred to foundations and industries, for projects that may or may not, be based on cost sharing.
- ▶ That they will communicate the name and address of their general practitioner or treating physician/clinic and that they allow the National Cohort e.V. to check their health status from secondary data (e.g., from medical files and registries). Participants will release their present and future physicians from maintaining confidentiality, including the permission to get the information on cause of death and information on diseases from disease and death registries.
- ▶ That information needed for the study can be transferred and scientifically used in coded form from social insurance carriers and health insurance (Institute for Employment Research, statutory health insurance). For this, participants will communicate the name of their health insurance, their health insurance number, and their social security number.
- ▶ That participation will not serve as a substitute for an individual health check, but that there will be communication of the most important findings (at baseline and later on) with the advice to contact a physician in case of any concern about their health status. Participants can choose what kind of information they will receive. They will have the right to refuse to receive any information about their health status.
- ▶ That the National Cohort e. V. will be the legal owner of the data and samples and that participants transfer property rights on the data and samples to the National Cohort e. V.
- ▶ That participation is voluntary and that they have the right to withdraw at any time during the study with different options of withdrawal. Otherwise, the consent form will retain validity throughout the lifetime of the National Cohort.

The consent form for participants who will take part in the MRI will additionally include a paragraph on handling of incidental findings, including an option to refuse to be informed about such findings. There will be categories from “I want to be informed in any case” to “I do not want to be informed at all”. The procedure is described in more detail in the section “Feedback of incidental findings”.

Participants will have two alternatives to give written consent. They can give full consent to the entire consent form or they can give differentiated consent for specific parts of the study program. In the differentiated consent, single examinations can be skipped.

A.7.1.3 Voluntariness and withdrawal

It will be emphasized that participation is voluntary and that no disadvantages or adverse consequences will arise due to refusal. At any time, the study participants will have the right to withdraw without having to give any explanation and without any disadvantage. Withdrawal must be requested in written form (or by telephone with subsequent written confirmation). There will be several types of withdrawal with different consequences for data and biosample use and access:

- ▶ “No further contact”: The National Cohort team would no longer contact the participant directly, but would still have their permission to use provided data and biosamples as well as to obtain further information from secondary data sources (e.g., official registries, local health departments).
- ▶ “No further access”: The National Cohort team would no longer contact the participant or obtain further information from secondary data sources (e.g., official registries, local health departments), but would still have the permission to use the data and biosamples provided previously.
- ▶ “No further use”: The National Cohort team would no longer contact the participant or obtain further information. Furthermore, any data and biosamples provided previously will no longer be available to researchers. All samples and data will be no longer used for further analyses. However, it will not be possible to remove their data from analyses that have already been done (and published). For technical reasons, biosamples will not be destroyed but blocked. Data will be anonymized by deleting all personal identifying data. The discussion of this type of withdrawal has not yet been completed.

A.7.1.4 Risks

Participation in the National Cohort will not cause any considerable harm. The questionnaire includes some sensitive questions that study participants may feel uncomfortable about answering. However, participants will have the possibility to skip such questions if they are unable or unwilling to answer them.

Blood will be collected by experienced and trained staff. However, there will be a small risk of bruising (hematoma) at the puncture site and a minimal risk for a local or systemic infection and the injury of a nerve.

The lung function examination might cause dizziness. Trained personnel will be available all the time.

The step test will exclusively be conducted under medical control. Prior to the test, the study physician will check contraindications and will decide whether the test will be carried out or not. If there should be any problem during the step test (e.g. chest pain, arrhythmia, discomfort, dizziness etc.), the test will be terminated.

The more extensive dental examination could cause slight bleeding of the gum.

During the MRI, participants are enclosed and the MRI device will produce some noise, which may be considered uncomfortable/unpleasant. If the participants should become scared or uncomfortable, the imaging procedure will be terminated. MRI will not be conducted in pregnant women or in participants with claustrophobia, with embedded metallic objects, e.g., from surgery or accidents, or with pacemakers.

During the DXA measurement, participants are not at all enclosed. The dose of radiation is very low (less than 20µSv). In comparison, the average dose of natural radiation is 2100 µSv per year. DXA will not be conducted in pregnant women.

For the OGTT, participants must be in a fasting state, which could cause hypoglycemia. In the study center, participants must drink a solution of sugar. During or after the test, a small risk for short-term hypoglycemia or gastrointestinal problems is present.

No other direct risks are associated with the foreseen research program (**Table 7.1**).

Table 7.1: Physical examinations in the National Cohort and their risks and contraindications

Examination	Risks	Contraindications
<i>1. Level examinations</i>		
Weight and height	-	-
Waist (and hip) circumference	-	-
DXA scan	Low dose radiation (< 20µSv)	Pregnancy (ask for known or possible pregnancy)
Blood pressure	-	-
Arterial stiffness, ankle brachial index (combined device)	-	-
10-sec ECG	-	-
33 echocardiography	-	-
Spirometry	Dizziness	Heart attack Collapsed lung Infection of the upper airways
Accelerometry	-	-
Hand grip strength	-	Swollen, inflamed, or painful hands Operation on the hand within the last 6 months
Cognitive function tests	-	-
Tooth count	-	-
Collection of blood	Blue marks (hematoma) Local or systemic infection Injury of a nerve	Known haemophilia

Examination	Risks	Contraindication
<i>2. Level examinations</i>		
OGTT	Hypoglycaemia Gastrointestinal symptoms	Fever Gastrointestinal infections/diseases Pregnancy Allergy to currant juice Known diabetes mellitus
Carotid sonography, intima media thickness	-	-
24h-ECG/apnae device	-	-
AGE reader/skin AF	-	-
Step test	Chest pain Dizziness (Muscle) fatigue Physical discomfort	Angina pectoris Aortic stenosis Dysrhythmias Pacemaker
One-leg-stand test	-	-
Analysis of NO in exhaled air	-	-
Audiometry	-	-
Eye-related measurements	-	-
Dental examinations	Slight gum bleeding	Known Haemophilia Known heart valve problems
Smelling test	-	-
Examination of the musculoskeletal system	-	-
<i>3. Level examinations</i>		
MRI	Incidental findings	Claustrophobia Pregnancy Embedded metallic objects from surgery or accidents Pacemaker

Indirect risks may present due to findings during the examination visit and the biosample analyses that deserve further medical examination. In the case of an incidental finding, participants will be informed and advised to contact a physician, if they consented to be informed. The strategy for how participants will be informed about incidental findings is described in detail below, in the section “Feedback of incidental findings”.

Finally, a minimal risk is associated with the use of personal information, but greatest care will be taken to maintain data protection and confidentiality as described below.

A.7.1.5 Expense allowance

Participants will receive a small expense allowance, including travel costs. This aims to strengthen the idea of donating important information and material for research on the prevention and treatment of serious disease. Samples and data should be seen as public goods.

A.7.1.6 Participant insurance

Due to the minimal risk of the study, general participation insurance will not be required. For any damage, the public liability insurance of the study centers will be liable. For the partici-

pants, travel accident insurance may be required. A special insurance will cover participants who receive a DXA or MRI examination.

A.7.2 Procedures for ethical clearance of National Cohort proposals

A.7.2.1 Regulatory framework

The National Cohort will be conducted in accordance with all relevant legislation, including the Federal Data Protection Act, the Data Protection Acts of the German States (which are comparable to the Federal Data Protection Act), the German Constitution, the registration laws of the Federal States, and the Cancer Registries Law. Overall, according to §1 of the German Constitution, human dignity will be protected. The right of self-determination, and the right to life and physical integrity are regulated in §2, German Constitution. Since, for example, the collection of blood samples may affect physical integrity, participants must give their written consent to the study.

In the following, the term “personal data” includes names, addresses, and other contact data as well as all medical study data that are collected during the study. In order to better protect the privacy of study participants, personal data will be split up into the two parts “personal identifying data” (name, contact data) and “study data” (only medical data).

For recruitment, the responsible registry office will provide the study centers with randomly selected registration data. According to the respective registry law of the Federal States¹, personal data can be transferred from the registry office to an official office (university, research institute) if these institutions need these data to fulfill their functions. According to §14, Federal Data Protection Act, the mentioned institutions are allowed to store, alter, and use personal data if this is necessary to fulfill their functions.

Personal data must be deleted once no longer needed to fulfill the specific functions (§20, Federal Data Protection Act). Within the scope of the National Cohort, study data will be kept for the long-term, because it is anticipated to be of scientific value for a long time. Personal identifying data will be deleted after the end of the study, after withdrawal and when no further direct contact to the participants or reidentification of individuals is needed. During the follow-up period, direct contact to the participants will be necessary, e.g., to update information on health status and relevant exposures. Due to the study design, it is planned to use personal identifying data for at least 20–30 years.

According to §3a, Federal Data Protection Act, personal data must be anonymized or coded (pseudonymized). Personal identifying data will be stored strictly separated from study data with additional restrictions for access. Only few staff members will have access to personal identifying data. All National Cohort staff members with direct contact to participants or personal identifying data will be committed to follow data secrecy according to §5, Data Protection Act and will be made aware of §203 Criminal Code in the case of breach of duty. Within the scope of the National Cohort, exclusively coded scientific study data will be transferred from the study centers to the integration centers for centralized recording.

The access to and use of secondary data will be conducted in accordance with §287, Social Security Code V, and §75, Social Security Code X.

According to §1, Cancer Registries Law, cancer registries must provide coded data for scientific research. According to §8, Cancer Registries Law, in which measures concerning

¹ § 24 Schleswig-Holstein; § 25 Berlin; Art. 28 Bavaria; § 28 Brandenburg; § 29 Baden-Wuerttemberg, Lower Saxony, Saxony, Saxony-Anhalt; § 30 Bremen; § 31 Hamburg, Mecklenburg-Western Pomerania, Rhineland-Palatinate, North Rhine-Westphalia, Saarland

health protection and important research questions are regulated, data can be transferred from the cancer registries. The transmission of personal data requires written consent. The transferred data will only be used for the requested and ethically approved purpose.

A.7.2.2 Data protection

The National Cohort will undertake the commitment to protect the confidentiality of collected data and samples. The entire research program as well as intended further studies and analyses will be approved by the responsible data protection commissioner and the ethics committee. The study will be conducted under consideration of the advice from the German Ethics Council⁸¹⁹, the “Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.” (TMF)⁸²⁰ and the guidelines signed and approved by the German scientific societies of epidemiology for Good Epidemiologic Practice (GEP)⁸²¹. The study will act in accordance with the Federal Data Protection Act and all other relevant legislation to protect the participants (see above). There will be a maximum of information security to block unauthorized access using state-of-the-art technology, according to the recommendations of the Federal Office for Information Security (BSI). Personnel in the several study centers will have direct contact to the participants and will check their personal identifying data. Their contract will include the commitment to ensure data protection and professional discretion and they will need to sign a confidentiality agreement. After recruitment, personal identifying data will be stored on a separate computer in the study centers and the trust center with no direct link to study data (i.e., data from interviews, questionnaires, and physical examinations) and biosamples. Access will be password-protected and only authorized staff will have access to these data. At baseline, participants’ home and work address will be used to generate data concerning environmental exposures such as outdoor air quality, noise etc. Geocoding will include current and future addresses by means of spatial codes (geocodes). A scientific institute will be in charge of producing exposure variables derived from geocodes and the respective spatial exposure information. Only coded scientific data will be transferred to the database. Reidentification of data will be necessary, for example, for follow-up activities or to contact participants concerning serious diagnostic findings. For follow-up activities, it will be necessary to link baseline data with follow-up data (study data and biosamples) as well as with registry data (e.g., cancer, mortality).

A.7.2.3 Data handling

Definitions

Coding	Substituting a code for personal identifying information in such a way that linkage is only possible through a key
Pseudonymization	Coding
Anonymization	The irreversible removal of personal identifiers from data or samples, such that no specific individual can be identified
ID-P	Personal identification number for recording personal identifying data
ID-S	Subject identification number for recording study data (->coded data)
ID-A	Analysis number for data transferred to scientific users
ID-HI	Identification number for health insurance data
ID-SI	Identification number for social insurance data

Data handling (coding, transfer, linkage, and storage) will be performed with regard to very high standards of security, bound by contract. Data handling will be conducted at different levels of the organizational units, which are described in detail in **Sect. A.4**.

The first level will comprise the local study centers, where all the data will be obtained. Data quality control, plausibility tests, and first check for completeness will be performed at the study center. It remains to be decided whether the pseudonym for coding (pseudonymization) will be generated in the study center or in the trust center.

The second level will be formed by two integration centers. All study data from the different study centers, the laboratory analyses, and information from registries, health departments, and other external sources will be recorded in a centralized database at the integration centers. As the two integration centers will not have any direct or indirect contact with the participant, personal identifying data will not be sent to the central database.

Further units are the different competence centers, a data transfer unit, and a trust center, of which only the trust center will receive the personal identifying information and no study data.

Coding of data / pseudonymization

For the identification of participants, each participant receives a unique personal identification number (ID-P). Personal identifying data (name, address, telephone number, and date of birth) will be recorded with the ID-P (see box **Definitions**).

In order to separate personal identifying data from study data, each participant will additionally be assigned a second code/pseudonym (subject ID, ID-S) (at the study center or the trust center). The ID-S will be used for the linkage and storage of all obtained study data and samples and for communication between the different units involved in data management. At the study centers, all documents, questionnaires (electronic or paper), and containers for collecting biosamples will be provided with a barcode label containing the ID-S. To allow re-identification (only if necessary), the ID-S will be recorded together with the respective ID-P in a separate file at the study centers and the trust center (**Table 7.2**).

A.7

The ID-P will either be transferred together with the respective personal identifying data from the study centers to the trust center, or it will be generated at the trust center and transferred back to the study centers. Only few authorized staff members will have access to the ID-P and personal identifying data. ID-P and personal identifying data will not be recorded in the scientific database or anywhere in the integration centers, competence centers, or data transfer unit.

The ID-S will be transferred together with the study data from the study centers to the responsible integration center and then recorded in the central database. Data which must be processed and analyzed at a specific competence center will be transferred with the ID-S from the integration center to the respective competence center. The competence center sends the data resulting from the analysis and the ID-S back to the integration center.

Precoded biosamples will be aliquoted into 2-D barcode tubes. During this process, the ID-S will be assigned to the respective 2-D barcode.

Data and sample transfer to scientific users will be organized by the transfer unit. In the process, the ID-S will be replaced by a random analyzing number (ID-A). This number characterizes explicitly one participant within a specific data set. For each data set produced for scientific analysis, a new ID-A will be assigned. Study data will be made available to scientific users only with an ID-A.

Use of secondary data

The use of secondary data requires particular data protection provisions. Therefore, we intend to establish a competence center for secondary data. The procedures for the collection and use of secondary data from different sources have not been finalized yet and may differ depending on the source of secondary data. Both central and local data requests are possible. Procedures being discussed for using health insurance data and social insurance data are described in the following.

a) Health insurance data

The procedure for using health insurance data may be as follows:

- ▶ Participants will be asked to give consent to communicate the name and address of the health insurance and their health insurance number. In the participants' information leaflet, they will be informed about the use of their data.
- ▶ After they give their consent, the name and address of the health insurance and their health insurance number will be recorded together with the personal identifying data at the study center and the trust center.
- ▶ The trust center or the competence center for secondary data generates a new identification number for health insurance data (ID-HI) for a given ID-S. The trust center sends the ID-HI to the respective health insurance together with the participants' health insurance number and the request.
- ▶ The health insurance sends the requested data with the ID-HI to the competence center for secondary data. The competence center for secondary data processes the data, replaces the ID-HI, and sends the standardized data together with the ID-S to the integration center. By following this procedure, the trust center would not receive personal medical or social data.

b) Social insurance data

The procedure for the use of social insurance data may be as follows:

- ▶ Participants will be asked to consent to providing their social insurance number to the study center (on a separate sheet). In the information leaflet they will be informed that this will only be done in cooperation with the Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung, IAB), the research institute of the Federal Employment Agency. Furthermore, participants will be informed about the use of their data and about the content of the social insurance data. They will also give consent that study data can be linked to social insurance data.
- ▶ After they give their consent, a new identification number for social insurance data (ID-SI) will be generated (at the study center or the competence center for secondary data). This ID-SI will be transferred together with the social insurance number to the IAB. Additionally, the ID-SI will be sent to the competence center (together with the ID-S). The social insurance number will only be recorded temporarily in the study center.
- ▶ The IAB records the social insurance number together with the ID-SI for the long-term. Upon request, the IAB selects the requested data (e.g., occupational history on a yearly basis). The IAB will generate a coded data set for scientific analysis and will transfer this data set with the ID-SI to the competence center.
- ▶ The competence center replaces the ID-SI with the ID-S and sends the standardized data together with the ID-S to the integration center.
- ▶ The IAB will contact the German statutory pension fund (Deutsche Rentenversicherung) to clarify the modalities of German statutory pension fund participation.

Table 7.2: Availability of identification numbers

	ID-P	Personal Identifying Data	Link between ID-P and ID-S	ID-S and study data	ID for secondary data – central data request	ID for secondary data – local data request	ID-A
Trust center	✓	✓	✓	X	✓	X	X
Study center: Subject management	L	L	L	X	X	L	X
Study center: Examination unit	TL	TL	TL	L	X	X	X
Integration center	X	X	X	✓	X	X	✓
Competence center	X	X	X	S	✓	X	X
Source of secondary data	X	✓	X	X	✓	✓	X
Transfer unit	X	X	X	✓	X	X	✓

✓ = available; X = not available; L = only for local participants; TL = temporary, only for local participants; S = only specific data

A.7.2.4 Data access

Access to scientific data

General aspects

The National Cohort e.V. would like to encourage research using resources collected as part of the National Cohort. Data collected as part of the National Cohort are for scientific use only and should benefit high-quality research. To this end, data access must be fair, transparent, and open. As there may be a high demand for data access on various projects, careful management of the resources and coordination of such research projects are important, especially as some resources are finite (e.g., biological samples).

The National Cohort e.V. will be the legal owner of all data and biological samples provided by the participants and will retain overall control of all access to the data and samples. An access policy will be developed, in which the application process, the criteria for prioritization, the handling of conflicts of interest, and the composition of the “Use and Access Committee” will be described. Different request procedures may coexist, depending on the data required (questionnaire data vs. biological samples).

Approval

Third parties cannot contact participants directly; they can only apply for access to data and material via the National Cohort.

It will take years until (longitudinal) scientific analysis can be carried out. The National Cohort team will call for proposals, once enough data are available. The first call will be made public by posting on the webpage and direct distribution to universities, research institutions, and national and international epidemiologic associations. Any researcher interested in analyzing the data is then welcome to send a proposal to the National Cohort. Later on, researchers can apply for access, independent of calls.

The National Cohort participants' consent includes a question concerning the fact that national and international universities and research institutions, foundations, and industries will be able to apply for access to data. Participants can decide during the consent process whether they agree to this condition. In the information leaflet, participants will be informed about the possibility of access by additional institutions.

Researchers must be affiliated with an institution that is competent in conducting health-related research. Inexperienced researchers may also be asked to seek the supervision of a more experienced researcher, affiliated with one of the accepted institutions. For international projects, collaboration with a German group of scientists as well as full acceptance of the German ethics rules may be required. No access to data and samples will be given to insurance companies, employers, the police or other law enforcement agencies and only limited access to commercial organizations.

Review process

Proposals for data use will be reviewed by the "Use and Access Committee" to ensure that they are in accordance with the participants' consent and all ethical aspects of the National Cohort, that they have relevant ethical approval, and that they fit the aims of the National Cohort.

If necessary, the "Use and Access Committee" reserves the right to ask external experts for advice on the received proposals. The "Use and Access Committee" will thereafter prioritize the projects that have been proposed, will give recommendations for those which should be supported, and advise the Epidemiologic Planning Committee.

The National Cohort e.V. will have the final decision-making authority over access to and use of data and biosamples and will ensure that the data and samples will be used scientifically and in the public interest. It will promote the greatest potential benefit for public health.

Concerning biological samples, an intelligent handling over time of the resources and of the conducted analyses is necessary to avoid depletion of the stocks in the course of the first few years. To facilitate the management of the resources, in time and in space, the analyses may be conducted in a few central laboratories, which will then be in charge of delivering the results to the teams which have asked for them.

Contract

All users (universities, research institutions, foundations, and industry) must follow the same ethical standards. Access to data and/or samples will only be conceded under licenses for scientifically and ethically approved research. Rights to use data and samples will be issued for specific uses under strict conditions, including compliance with the consent given.

The researcher will need to sign an access agreement contract with the National Cohort e.V. The contract will cover the aim of the research, the proposed analyses, the timeline, and the publication of the results. The conditions for data utilization will also be detailed. The contract will be valid for a defined period of time.

Should biosamples need to be handed over, a specific contract related to sample transfer, use, and analysis will be signed between the National Cohort and the laboratory.

Users will be provided with exclusively coded data and/or samples. Users may not undertake to identify and/or contact participants. Since National Cohort e. V. is the legal owner of all data and samples, it holds the right to take legal action against unauthorized use or abuse of the data and samples.

Fees

Fees will be charged for the right to use data and samples in order to cover the costs generated by data processing, preparation, and analyses. Higher fees may be charged to organizations that might expect to derive financial benefit from use of data.

Data sharing

All research users will be requested to publish results (positive and negative) from all analyses made of National Cohort data and samples in peer-reviewed scientific journals whenever possible.

Researchers must agree to provide the results of their analyses to the National Cohort group. This includes analyses performed on the data which could, after entry in the database, be available and useful for other approved researchers. Methodology, analyses, results, and any other information of interest would be published in a public domain (e.g., on the webpage). This would then benefit many people and would also keep the participants informed about the results of the study in which they were involved.

The National Cohort website will contain updated information about the use of data and samples and about which researchers and institutions have been granted or denied access. Participants can thus see which researchers have access to their data. Participants who may become uncomfortable with the idea of access may consider their right to withdraw from the study.

Results from a feasibility study, including public opinions concerning access to data and samples, will be considered in the final data access concept.

A.7

Access to personal information

The access to personal identifying data is described in **Sects. A.7.2.1** and **A.7.2.2**.

Folders prepared for the assessment visit, including the consent form, registration sheet, documentation sheet, etc. should be stored daily in lockable cabinets near the registration desk. After the assessment visit all organizational data from the documentation sheet must be entered into the organizational database. The folders finalized thus far must be kept in lockable cabinets in a lockable room. All sheets containing exclusively scientific data must be transferred from the study centers to the coordinating study center once a week/every 2 weeks. However, whenever possible, data will be obtained directly electronically. Once the study is finished, all documents in the folders that are no longer needed (bar code sheets, routing slips, etc.) will be destroyed. Only the registration sheet, the consent form, and the documentation sheet will be sorted according to the organization number (org-number) and stored for the long-term in a specific archive. Only few persons will have access to this archive. These documents can also be archived in electronic form.

Control of data protection

The responsible data protection commissioner for each study center will ensure that data collection and handling comply with the data protection concept that has been developed and in accordance with the Federal Data Protection Act and all other legislation concerning data protection and confidentiality.

A.7.2.5 Provision of health information to participants

At the end of the visit in the study center or soon after, if desired, participants will receive the results from the medical examinations conducted at the study center and analyzed baseline diagnostic results. By reporting standard ranges and giving respective explanations concerning their individual results, participants will be provided with information that gives meaning to the examinations. If there is any concern about their health status, participants will be advised to contact their general practitioner or another physician. Moreover, participants will be informed and advised to contact a physician in the case of severe findings which need to be further clarified. No medical diagnosis will be provided. A detailed feedback at the time of the baseline examination also goes beyond the scope of the study. On the consent form, participants will advise as to what kind of information they want and how they want to be informed. New biomarkers implicating a very high risk of a preventable and serious disease may be found. If such findings become available and are corroborated in agreement with relevant medical expertise, participants may be contacted and these findings will be communicated in accordance with an ethics commission. To respect the “right of nescience,” this will only be done if the participant agreed to be contacted for such information.

We aim to ensure that subjects understand that study participation does not provide a personal health check but rather will provide important and valuable information concerning the etiology of chronic diseases and for future prevention and treatment. Thus, participation will help enhance other people’s health in the future.

Feedback of incidental findings

Concerning the use of MRI, one important issue warrants specific attention in connection with the target of asymptomatic voluntary participants, namely incidental findings.

Incidental findings represent an important and very critical issue within the imaging sub-study of the National Cohort and dedicated guidelines will be developed in collaboration with the pertinent ethics committees. Overall, the guidelines will balance the moral obligation of reporting findings, anticipated benefits for the participants, and the scientific validity without intervention in terms of an unbiased assessment. Based on the experience in the SHIP study⁸²², only those incidental findings that are of potential relevance for the participants are provided to them.

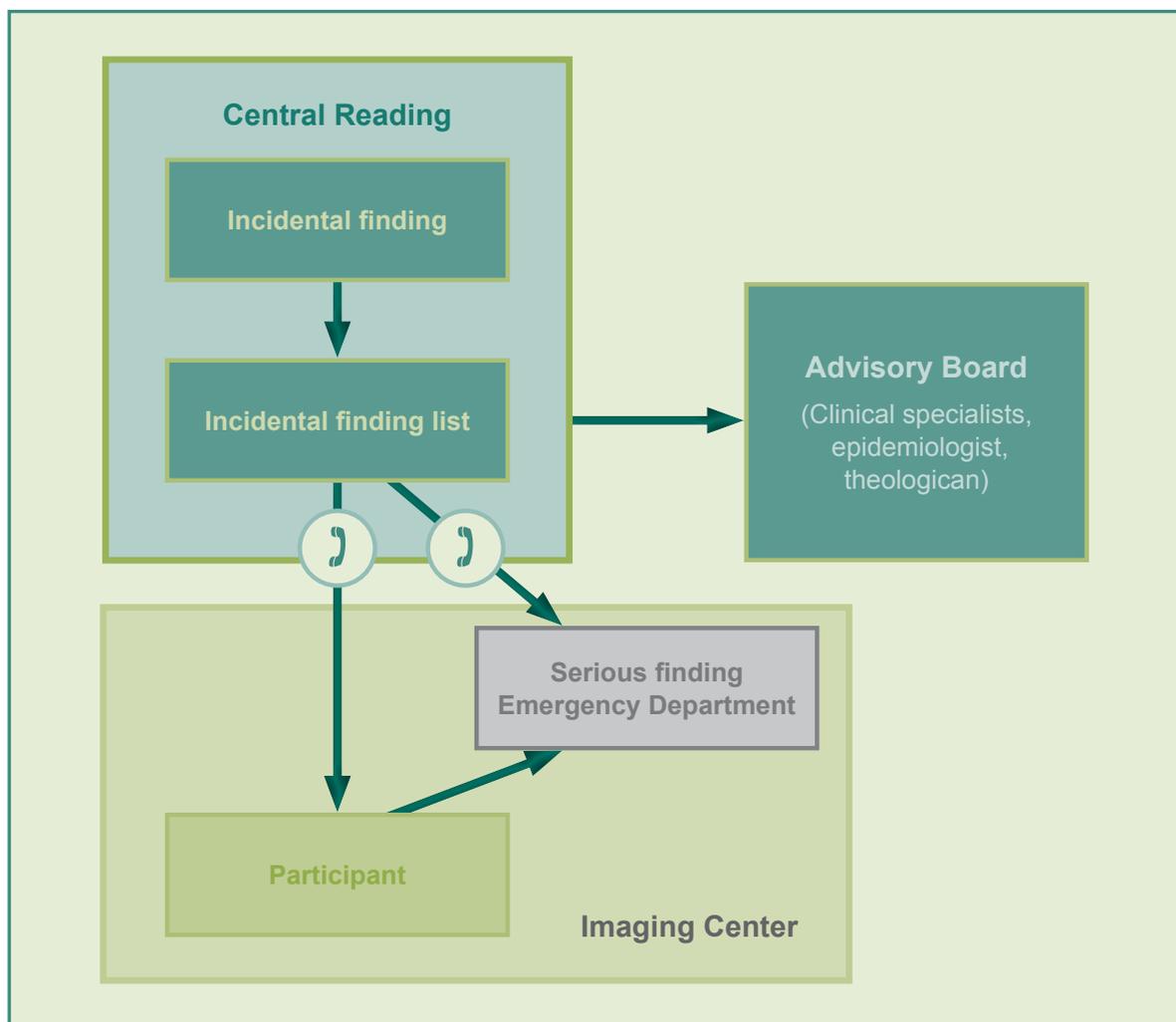
We have already prepared a draft list of serious incidental findings to be provided to the participants on head, ENT/neck, thorax, heart, abdomen, urogenital system, spine, vascular system, and musculoskeletal system. This draft will be finalized together with specialists from the different fields from Germany and the UK.

For further incidental findings, which are not predefined in the list, but are of potential relevance, we will establish an advisory board consisting of experts of all major, imaging-related clinical specialties, as well as epidemiologists and ethicists who will meet on a regular basis to discuss and develop guidelines related to distinct imaging findings on real cases (**Figure 7.1**). Over the course of the preparatory phase, guidelines will be in place to classify imag-

ing findings according to relevance and acuteness of findings and how the finding will be reported to the participant and to the primary care physician. Specifically, there will be a detailed list of findings that need to be reported without delay. Furthermore, recommendations for follow-up examinations will be developed that will be made available to the primary care physicians.

All serious incidental findings detected at the first-line reading will be communicated directly to the study participants, and measures will be taken to ensure appropriate care. The immediate phone/video call which will include recommendations for further work-up will prevent participants from unnecessary worries. Whether incidental findings detected at the second-line scientific reading should be communicated to the recruitment site, which may contact the study participant and/or the primary care physician, if necessary, still needs to be discussed.

Figure 7.1: Workflow of the centralized reading core and advisory board with respect to incidental findings.



A.7

A.7.2.6 Communication of study results

Since the results from the National Cohort will be a resource of scientific and public interest, they will be made available using different means of communication. The results may influence public health strategies such as new prevention or screening strategies.

To reach the general public, information and study results will be communicated via the National Cohort website, articles in newspapers, press reports, radio, and television. Regular communication will be important so as to inform participants of general findings and to encourage continued participation. Therefore, it is planned that each participant receives regular newsletters with up-to-date information about the study. Scientific results will be published in national and international, peer-reviewed scientific journals. The state of the project and results will also be presented at national and international conferences.

Participants will receive their own individual study results as described above (**Sect. A.7.2.5**).

A.7.2.7 Value for the public and strategy to involve the public

The National Cohort will provide unique data for current and future biomedical research. The study will build up a resource for research in all medical disciplines, enabling new and groundbreaking research on the relationships between heredity, environment, and lifestyle and their effect on the risk of chronic diseases. The results of the National Cohort will provide new information about the causes of diseases that will lead to new means of prevention, refined diagnostic methods, and therapeutic opportunities. The overall findings and implications of results that derive from the National Cohort will be made available to participants and the wider community so that they can influence public health strategies. Thus, data obtained with the National Cohort will serve as a public good. The National Cohort seeks to augment the value of the resource by stimulating best possible research projects in order to ensure that the greatest potential benefit for public health may be realized from this work.

Prior to and during recruitment, in particular in the recruitment areas, well-organized public relations activities will be conducted, including advertisement in newspapers, radio stations, and public transports. Moreover, information about the National Cohort will be communicated via internet (National Cohort website, Wikipedia, and Facebook). For all potential participants there will be a free telephone service.

To reach people with a migration background, specific public relations activities will be set up. For instance, we will aim to promote the National Cohort by advertising in migrant-specific non-German-language media (e.g., radio stations) and informing the commissioners for integration, local organizations, and specific information centers for immigrants about the purpose of the study.

A.7.3 Pre-evaluation of the proposed study from an ethical perspective

The questions of how the public perceives the National Cohort project, which issues the public identify as key for the society, and how the fact of the growing (transnational) cooperation of such projects and biobanks is perceived are topics of considerable importance for the project and might influence its development and acceptance. During the National Cohort pilot period, a focus group project will deal with the issues of general knowledge and understanding of cohort studies and biobanks, opinions and perceptions according to data collection, data handling, data protection, consent, confidentiality, etc.

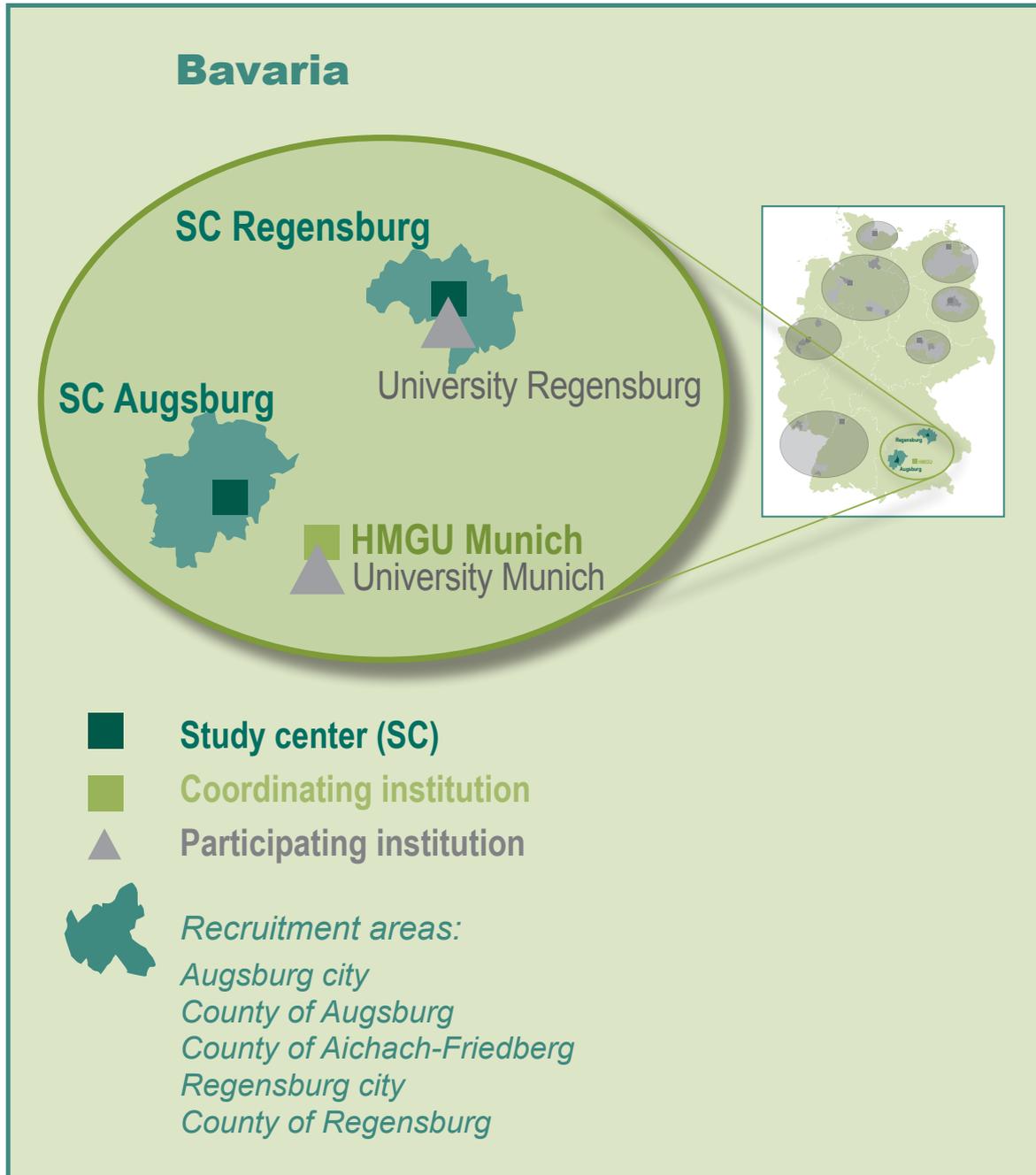
The results of the focus group project can support several decision-making processes with regard to different key issues of the study. Critical issues, in particular ethical issues, identified by the focus group participants will be discussed and procedures for dealing with these issues established.

C.1 Description of the study centers

C.1.1 Cluster Bavaria

The Cluster Bavaria comprises two study centers in the areas of Augsburg and Regensburg.

Figure 1: Geographical overview of recruitment areas and study centers for the cluster Bavaria



C.1 Description of the study centers

C.1.1 Cluster Bavaria

The Cluster Bavaria comprises two study centers in the areas of Augsburg and Regensburg.

C.1.1.1 Cluster coordinator

Table 1: Cluster coordinator/co-coordinator

Coordinator	
Name	Prof. Dr. Dr. H.-Erich Wichmann
Institution Town	Helmholtz Center Munich (HMGU) 85764 Neuherberg
Institute/department/division	Institute of Epidemiology I
Address, telephone, e-mail	Ingolstaedter Landstrasse 1, 85764 Neuherberg +49 89-3187-4066 Wichmann@helmholtz-muenchen.de
Co-coordinator	
Name	PD Dr. Jakob Linseisen
Institution Town	Helmholtz Centre Munich (HMGU) 85764 Neuherberg
Institute/department/division	Institute of Epidemiology I
Address, telephone, e-mail	Ingolstaedter Landstrasse 1, 85764 Neuherberg + 49 89-3187-3202 j.linseisen@helmholtz-muenchen.de

C.1.1.2 Study centers

Study center Augsburg

Table 2: Overview of participating institutions and investigators/scientists

Institution	Helmholtz Center Munich (HMGU)
Town	85764 Neuherberg
Principal investigator	Prof. Dr. Dr. H.-Erich Wichmann
Institute/department/division	Institute of Epidemiology I
e-mail	wichmann@helmholtz-muenchen.de
Co-principal investigator	PD Dr. Jakob Linseisen
Institute/department/division	Institute of Epidemiology I
e-mail	j.linseisen@helmholtz-muenchen.de
Institution	Helmholtz Center Munich (HMGU)
Town	85764 Neuherberg/
Associated scientist	PD Dr. Christa Meisinger
Institute/department/division	Institute of Epidemiology II, KORA study center Augsburg, Myocardial infarction register
e-mail	christa.meisinger@helmholtz-muenchen.de
Associated scientist	PD Dr. Thomas Illig
Institute/department/division	Molecular Epidemiology
e-mail	illig@helmholtz-muenchen.de
Associated scientist	Prof. Dr. Annette Peters
Institute/department/division	Institute of Epidemiology II
e-mail	peters@helmholtz-muenchen.de
Institution	Ludwigs-Maximilians-University
Town	80539 Munich
Associated scientist	Prof. Dr. Maximilian Reiser, Dean of the Medical Faculty
Institute/department/division	Institute of Radiology
e-mail	Maximilian.Reiser@med.uni-muenchen.de

Description of the institution responsible for establishing and operating the study center

The **Helmholtz Center Munich** is the German Research Center for Environmental Health (HMGU). It is a research institution of the Federal Government and the State of Bavaria within the Helmholtz Association of German Research Centers. Research at HMGU concentrates on diseases related to the environment such as lung diseases and illnesses of the immune system, and contributes to elucidating the mechanisms underlying neurodegenerative diseases, cancer, diabetes, and other chronic diseases. It is divided into 28 research in-

stitutes and independent departments, which are interlinked and cooperate on various topics and in various research programs. The center runs diverse technology platforms which function as central service units. To ensure rapid and efficient transfer of findings from basic research into medical applications, scientists of Helmholtz Center Munich work closely in translational centers and clinical cooperation groups together with medical partners in the universities and hospitals in Munich.

Research at the Institute of Epidemiology at HMGU focuses on the effects of environmental factors and genetic predisposition – as well as their interaction – on human health by means of epidemiologic methods. The main research areas in terms of disease endpoints of the institute are diabetes and the metabolic syndrome, cardiovascular diseases (CVD), diseases of the respiratory system, and allergies. Exposures of interest include indoor and outdoor air pollution, ionizing radiation, diet, physical activity, and psychosocial factors. The integration of molecular epidemiologic methods has become an indispensable part of the epidemiologic research activities of the institute. The scientific output of the institute is reflected in the publication record (<http://www.helmholtz-muenchen.de/epi/publikationen/uebersicht/index.html>) which also demonstrates that numerous national and international collaborations are a major part of the successful scientific work. In particular, the area of genome-wide association studies (GWAS) has proven to be a very efficient source of scientific collaborations and publications, thanks to the abundant phenotypic data available in MONICA/KORA.

The Institute of Epidemiology at the HMGU manages several population-based epidemiologic studies covering a broad age range, from birth to old age. The largest and most prominent study is the population-based MONICA/KORA cohort with about 18,000 subjects in the age range of 25–74 years at recruitment. Participants were enrolled during four recruitment periods in 1984/85, 1989/90, 1994/95, and 1999/2001 and were subsequently followed up for disease outcomes. Further cohort studies include two birth cohorts, the GINI study (n~6,000; recruitment 1995-98), and the LISA study (n~3,100; recruitment 1997–99), both with regular follow-up since birth. All studies include physical examinations and, with the exception of earlier study components, have incorporated the collection of biological specimens (blood, urine, faeces, and saliva).

Laboratory expertise is provided by the Molecular Epidemiology Group at HMGU, headed by T. Illig. T. Illig's lab has strong expertise in both genotyping and metabolomics. Of specific importance is the expertise in standardized collection and storage of biomaterials obtained in epidemiologic studies. Members of the HMGU group have developed a concept for long-term management of the biospecimens to be collected from the National Cohort participants.

With respect to the baseline examinations, HMGU will cooperate with researchers from outside, especially from the Medical Faculties of the Universities in Munich (LMU, TUM); for magnet resonance imaging a close cooperation with M. Reiser (Institute of Radiology, LMU) has been established.

Description of the study center Augsburg

The region of Augsburg – consisting of the city of Augsburg and the two neighboring counties of Augsburg and Aichach-Friedberg – has a source population of 420,000 inhabitants in the age range of 20–69 years, of which 20,000 subjects shall be recruited into the National Cohort.

The study center for the recruitment and follow-up of these study participants will be located in the city of Augsburg, about 70 km west of Munich. The study center is close to the railway station and thus easy to access for the study participants (by train, bus, or car).

For many years, the HMGU has been running a study center in Augsburg in which population-based studies, especially the MONICA/KORA studies, were conducted; since 2001, it has been under the supervision and management of C. Meisinger. For the recruitment and follow-up of the National Cohort, another floor (about 500 m²) in the same building will be rented. As previous studies have been carried out there, a wide experience and a functional infrastructure can be found, which will be used for the purposes of the National Cohort.

Successful cooperation with medical facilities, in particular with the Augsburg Hospital (Klinikum Augsburg) will be continued to the benefit of the National Cohort. Augsburg is the only study region in Germany with the possibility of linkage to a population-based myocardial infarction (MI) registry. The MI registry, also headed by C. Meisinger, is located at the Klinikum Augsburg and has now been operational for 25 years.

Study center Regensburg

Table 3: Overview of participating institutions and investigators/scientists

Institution	Regensburg University (RU)
Town	93053 Regensburg
Principal investigator	Prof. Dr. Dr. Michael Leitzmann
Institute/department/division	Department of Epidemiology and Preventive Medicine
e-mail	Michael.Leitzmann@klinik.uni-regensburg.de
Co-principal investigator	Prof. Dr. Iris M. Heid
Institute/department/division	Department of Epidemiology and Preventive Medicine
e-mail	Iris.Heid@klinik.uni-regensburg.de
Institution	Regensburg University (RU)
Town	93053 Regensburg
Associated scientist	Dr. Beate Fischer
Institute/department/division	Department of Epidemiology and Preventive Medicine
e-mail	Beate.Fischer@klinik.uni-regensburg.de
Institution	Medical Faculty of Regensburg University
Town	93053 Regensburg
Associated scientist	Prof. Dr. Bernhard Weber
Institute/department/division	Institute of Human Genetics at the University of Regensburg
e-mail	Bernhard.Weber@klinik.uni-regensburg.de
Institution	University Hospital Regensburg
Town	93053 Regensburg
Associated scientist	Prof. Dr. Günter Riegger
Institute/department/division	Regensburg University Medical Center
e-mail	Guenther.Riegger@klinik.uni-regensburg.de

Description of the institution responsible for establishing and operating of the study center

Regensburg University was founded in 1962 and was designed as a comprehensive university offering a diverse spectrum of educational and research disciplines. An average of 20,000 students are currently enrolled at Regensburg University. The **Department of Epidemiology and Preventive Medicine** at Regensburg University is embedded in the Regensburg University Medical Center, a tertiary care facility that provides inpatient and outpatient diagnostics and therapy for all serious diseases. It accommodates 20 clinical departments and five biomedical institutes and provides the majority of health care in North-Eastern Bavaria, with 29,000 inpatients and 111,000 outpatients being treated here every year. The Faculty of Medicine at Regensburg University Medical Center has made a firm commitment to support a comprehensive research program on disease prevention. As such, there are strong research collaborations between the epidemiology department and clinical partners at the Regensburg University Medical Center, including collaborations in the areas of cardiovascular and metabolic diseases, cancer, neurodegenerative diseases, and biostatistics. Clinical research infrastructure is available at the Regensburg University Medical Center for all pertinent aspects of medical and epidemiologic data collection and analysis, including experienced study personnel (study nurses, data managers, IT specialists, and quality assurance managers), laboratory facilities for blood sampling and biospecimen storage, and in-house facilities for imaging examinations (MRI).

The primary areas of research of the Department of Epidemiology and Preventive Medicine at Regensburg University include the interrelationship between body size, energy intake, physical activity, and chronic disease with a focus on cancer. Research encompasses methodological research on the assessment of physical activity and its individual components. Although physical activity measurement techniques have evolved considerably over the past few years, the main obstacle in quantifying physical activity is the complexity of precisely measuring its individual components. In addition, the department is pursuing research to identify biological mechanisms linking the adverse effects of obesity, energy oversupply, and physical inactivity to disease risk. Possible biological mechanisms include insulin resistance and lipid metabolism, insulin-like growth factors, steroid hormones, immune function, and inflammatory factors. The department also has a strong focus on molecular and genetic epidemiology and it is an active participant of several large consortia, such as GIANT (Genetic Investigation of ANthropometric parameters), one of the largest international genetic epidemiology consortia including over 40 studies involving n~120,000 subjects and n~100 researchers. In addition to research, the department provides teaches epidemiology to medical students and students of molecular medicine through lectures, seminars, and mentorships for masters, and doctoral theses.

Description of the study center Regensburg

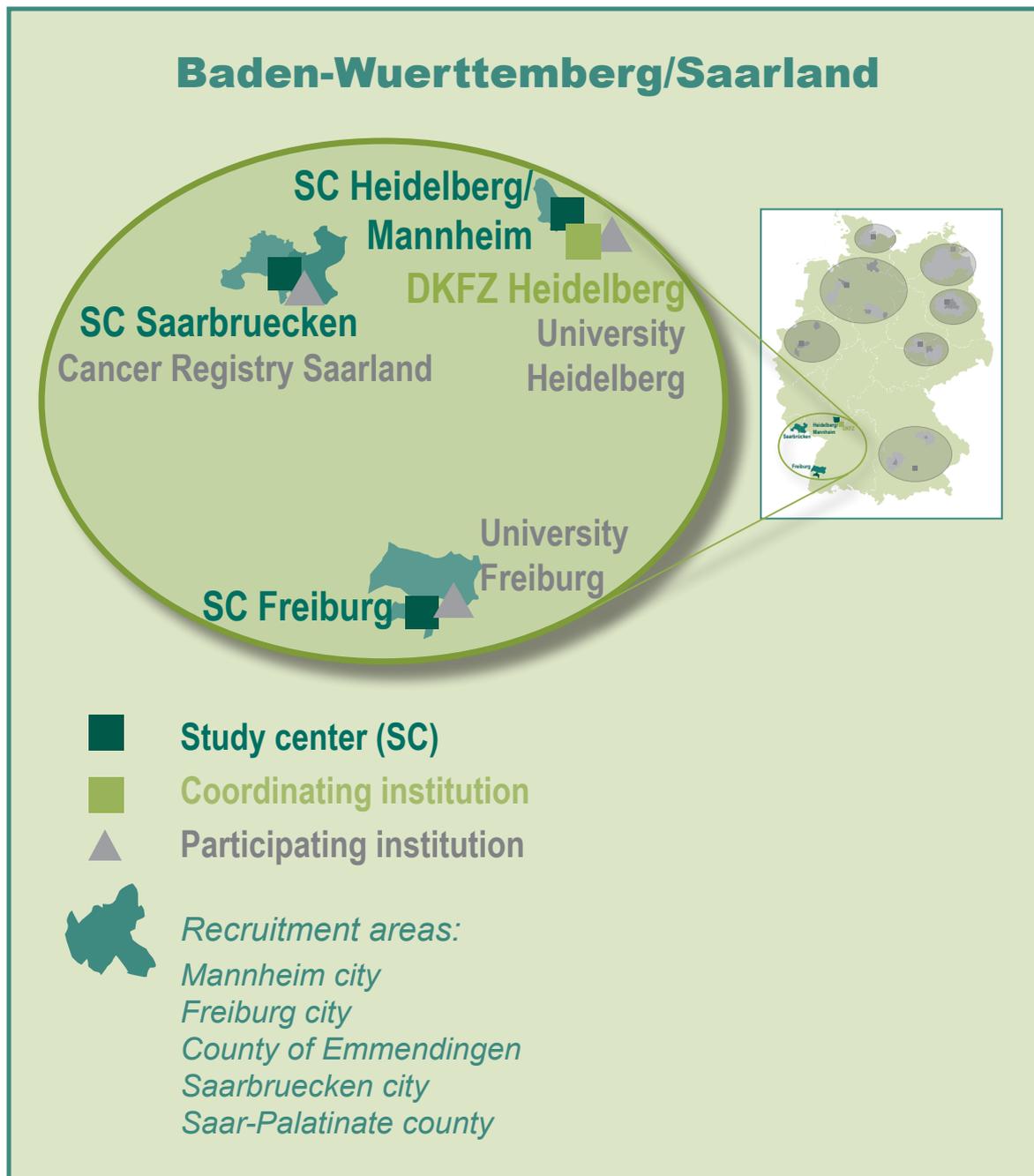
The sampling frame for the National Cohort includes a source population of 215,000 inhabitants in Regensburg city and neighboring communities. Under the leadership of M. Leitzmann 10,000 study participants in the age range of 20 - 69 years will be recruited at the study center in Regensburg. The study center includes ample space (approximately 500 m²) for interviews, clinical investigations, blood draws, and interim storage of biosamples. The study center is located within the Regensburg University Medical Center, which is easily accessible to study participants via public transportation or by car. The direct location of the study center at Regensburg University Medical Center and its existing collaborations with numerous clinicians and associated scientists will allow for an optimal conduct and oversight of the study operations and will facilitate identifying and validating a wide range

of phenotypes relevant to the National Cohort. In addition, the close affiliation with the Regensburg Cancer Registry will be instrumental efficiently and comprehensively tracing study participants, particularly for subjects who develop cancer during follow-up.

C.1.2 Cluster Baden-Wuerttemberg/Saarland

The Cluster Baden-Wuerttemberg/Saarland comprises three study centers in the area of South-West Germany, located in Mannheim, Freiburg, and Saarbruecken.

Figure 2: Geographical overview of recruitment areas and study centers for the cluster Baden-Wuerttemberg/Saarland



C.1.2.1 Cluster coordinator

Table 4: Cluster coordinator/co-coordinator

Coordinator	
Name	Prof. Dr. Rudolf Kaaks
Institution	German Cancer Research Center (DKFZ)
Town	69120 Heidelberg
Institute/department/division	Division of Cancer Epidemiology
Address, telephone, e-mail	Im Neuenheimer Feld 280, 69120 Heidelberg +49 6221 42 2219 r.kaaks@dkfz.de
Co-coordinator	
Name	Dr. Karin Halina Greiser, MPH
Institution	German Cancer Research Center (DKFZ)
Town	69120 Heidelberg
Institute/department/division	Division of Cancer Epidemiology
Address, telephone, e-mail	Im Neuenheimer Feld 280, 69120 Heidelberg +49 6221 42 2219 h.greiser@dkfz.de

C.1.2.2 Study centers

Study center Heidelberg/Mannheim

Table 5: Overview of participating institutions and investigators/scientists

Institution	German Cancer Research Center (DKFZ)
Town	69120 Heidelberg
Principal investigator	Prof. Dr. Rudolf Kaaks
Institute/department/division	Division of Cancer Epidemiology
e-mail	r.kaaks@dkfz.de
Co-principal investigator	Dr. Karin Halina Greiser, MPH
Institute/department/division	Division of Cancer Epidemiology
e-mail	h.greiser@dkfz.de
Institution	University of Heidelberg
Town	69120 Heidelberg
Co-principal investigator	Prof. Dr. Heiko Becher
Institute/department/division	Medical Faculty, Institute of Public Health
e-mail	heiko.becher@urz.uni-heidelberg.de
Institution	University Heidelberg Medical Center
Town	69120 Heidelberg
Associated scientist	Prof. Dr. Hans-Ulrich Kauczor
Institute/department/division	Dept. of Radiology, Diagnostic and Interventional Radiology
e-mail	hans-ulrich.kauczor@med.uni-heidelberg.de

Description of the institutions responsible for establishing and operating the study center

The **German Cancer Research Center (DKFZ)**, the largest biomedical research center in Germany, was established in Heidelberg in 1964. DKFZ's long-term goal is to unravel the causes and mechanisms of cancer and apply this knowledge to develop novel tools for diagnosis, early detection, treatment, and prevention. The **Medical Faculty of the University of Heidelberg** – the oldest university in Germany – is one of the largest and most prestigious medical centers in Europe. The university hospital and the cooperating facilities in the region offer inpatient and outpatient diagnostics and therapy for all severe diseases. All major departments of the Medical Faculty of Heidelberg are located on the Biomedical Campus Heidelberg in close proximity to DKFZ.

The **Division of Cancer Epidemiology** at DKFZ studies the causes of cancer in population groups with the aim of identifying and, if possible, avoiding risk factors so as to prevent cancer. A wide range of epidemiologic studies have been successfully performed in the division. Among them is the EPIC Heidelberg study, for which the DKFZ was responsible for recruiting 25,545 subjects aged 35–64 years at the baseline examination in 1994 to 1998. Direct access to the population registries has been established for follow-up. Experienced staff trained in data collection (interviews, examination, and medical information) and recruitment

procedures (documentaries) is available. A well-functioning laboratory unit exists for the management of blood sampling and large biorepositories. Facilities for specialized examinations (MRI) are available in-house. New rounds of follow-up repeat exposure assessment started in EPIC, including MRI (whole body), movement detectors, and other examinations. DKFZ also houses the cancer registry of Baden-Wuerttemberg which will facilitate follow-up of incident cancer events. In addition to the EPIC study, several epidemiologic studies that are led by division members are ongoing, such as the multicenter case-control study MARIE or the large-scale cohort study LUSI.

The Institute of Public Health is part of the Medical School/Medical faculty at Heidelberg University. The aim of the institute is to contribute to improving health through research, teaching, and direct services (patient care, consulting) in developing countries and at home. The staff at the unit of Epidemiology and Biostatistics is well trained in conducting large epidemiologic studies. The Institute hosts the PhD program epidemiology (Graduiertenkolleg 793, see <http://grk.dermis.net>) that currently enrolls about 40 PhD students and that will be linked to the planned cohort with several PhD projects. In this cohort study, the collaboration with clinical partners at the Medical Faculty of the Heidelberg University will be particularly strong, encompassing cardiovascular, pharmacological, psychiatric, dental, biostatistical, and other specific expertise.

The DKFZ will run the study center Heidelberg/Mannheim, coordinated by H. Greiser and R. Kaaks. It is planned to establish a migrant cohort which will be strongly linked to the National Cohort, with 5,000 migrants in the study center Mannheim/Heidelberg. H. Becher will be PI of this migrant study component, and in view of this will also act as a Co-PI for the Heidelberg/Mannheim study center overall.

For medical examinations, study center Heidelberg/Mannheim will collaborate with the Heidelberg Medical Center. H.-U. Kauczor, head of the department of radiology, will be collaborating scientist for research involving magnetic resonance imaging.

Description of the study center Heidelberg/Mannheim

In Mannheim 10,000 individuals will be recruited and will include inhabitants of the city of Mannheim. The study center will be established close to the railway station in Mannheim to ensure a good access for the study participants of the region. The central location of the study center in Mannheim is advantageous, because this facilitates the recruitment of the Turkish migrants as an important subpopulation of the cohort as they represent 5.7% of the residents in Mannheim. If additional funding for a migrant cohort can be acquired, this will become relevant. Under this condition, it is foreseen to recruit 5,000 migrants from this area.

The study center will comprise rooms for investigation, interviews and preanalytical preparation and interim storage of biosamples (in total covering about 500 m²). The close cooperation with the cancer registry Baden-Wuerttemberg, which is located at DKFZ, ensures that incident cancer cases can be efficiently traced. The close vicinity to the study coordinating center at DKFZ, the Institute of Public Health, and the Medical Faculty at Heidelberg facilitates the close cooperation with clinicians and associated scientists from the faculty of Medicine.

Study center Freiburg

Table 6: Overview of participating institutions and investigators/scientists

Institution Town	University Medical Center Freiburg 79110 Freiburg
Principal investigator Institute/department/division e-mail	Prof. Dr. Dr. Karin Michels Comprehensive Cancer Center Freiburg (CCCF), Division of Cancer Epidemiology tumorepidemiologie@uniklinik-freiburg.de
Co-principal investigator Institute/department/division e-mail	Dr. Claudia Schmoor University Medical Center Freiburg's Clinical Trials Unit (ZKS – recently re-named as “Studienzentrum Freiburg”) claudia.schmoor@uniklinik-freiburg.de
Associated scientist Institute/department/division e-mail	PD DR. Alexandra Nieters, MPH Center of Chronic Immunodeficiency (CCI), Molec. Epidem. alexandra.nieters@uniklinik-freiburg
Associated scientist Institute/department/division e-mail	PD Dr. Martin Schumacher Institute of Biomedical Biometry and Medical Informatics ms@imbi.uni-freiburg.de
Associated scientist Institute/department/division e-mail	Dr. Anna Köttgen, MPH Dept. of Internal Medicine anna.koettgen@uniklinik-freiburg.de

Description of the institutions responsible for establishing and operating the study center

Founded in 1457, the **University of Freiburg** is one of the oldest German universities and is now one of the nation's leading research and teaching institutions, with a main focus on medicine. The **University Medical Center Freiburg** is a tertiary care hospital and one of the largest medical centers in Europe. All medical specialties and subspecialties are represented. As one of the general principles, future-oriented research ensures optimal diagnosis, treatment, and rehabilitation of our patients and more recently, additional emphasis is placed on prevention research and application.

The **Comprehensive Cancer Center Freiburg (CCCF)** has been involved in cutting-edge innovative treatment strategies for a multitude of cancer types. The unique position of the CCCF Freiburg was once again confirmed in November 2010 by including CCCF as a partner in the German Consortium for Translational Cancer Research. This association of elite centers aims to bring the latest research results into patient care faster.

The research focus of the **Division of Cancer Epidemiology** at CCCF at the University of Freiburg is to identify risk factors for cancer and to explore mechanistic pathways, including epigenetics. Under the leadership of K. Michels, the Division of Cancer Epidemiology has been part of CCCF since July 2008. The division builds on previous research which has demonstrated that cancer develops largely as a result of modifiable lifestyle and environmental factors, which interact with genetic predisposition and the epigenetic make-up. Therefore, the division mainly focuses on identifying and investigating protective and risk factors for cancer development and cancer-related diseases (e.g., diabetes) in populations and on epigenetic epidemiology. The Division of Cancer Epidemiology is currently pursuing several epigenetic projects, including the identification of prognostic indicators for endometrial cancer and epigenetic changes associated with diabetes, a risk factor for cancer.

In addition to research, the Division of Cancer Epidemiology participates in teaching epidemiology to students of medicine and molecular medicine through lectures and assistance with their doctoral thesis.

The **University Medical Center Freiburg's Clinical Trials Unit** (ZKS – recently renamed “Studienzentrum Freiburg“), is a central facility of the University Medical Center Freiburg and the Medical Faculty of the Albert-Ludwigs-University Freiburg. The Clinical Trials Unit as a core facility of the Comprehensive Cancer Center Freiburg provides comprehensive professional services for planning, conducting, and analyzing of clinical trials.

The Clinical Trials Unit has extensive expertise in conducting clinical trials, including clinical effectiveness research, and epidemiologic studies. Since its foundation in 1998 – initially funded by the BMBF – the Clinical Trials Unit Freiburg has been involved in more than 400 trials, large epidemiologic studies, and oncological long-term studies involving several thousand patients. The existing infrastructure of the Clinical Trials Unit and its qualified and experienced personnel (50 staff members, among them study nurses, data managers, project managers, study physicians, IT specialists, clinical monitors, and a quality assurance manager) can be used for recruitment and examination of study participants of the National Cohort.

The project management for the National Cohort and its organizational structure (recruitment, implementation of the examinations and interviews, preanalytical preparation and interim storage of biosamples, and long-term follow-up of study participants) will be a collaborative effort of K. Michels (PI) from the Division of Epidemiology and the Clinical Trials Unit under C. Schmoor's (Co-PI) collaborative leadership.

Description of the study center Freiburg

The city of Freiburg has a population of 220,000 inhabitants and a total of 1.2 million in the immediate vicinity of Freiburg. The sampling frame for the 10,000 study participants to be recruited will be the city of Freiburg and the county of Emmendingen, which includes both urban and rural areas with a source population of 260,000. The study center will be located in the premises of the Division of Cancer Epidemiology and the University Medical Center Freiburg's Clinical Trials Unit within the premises of the University Medical Center Freiburg. Due to the epidemiologic expertise of the Division of Epidemiology and the experience of the Clinical Trial Unit under C. Schmoor's collaborative leadership in conducting large clinical trials and epidemiologic studies, their collaboration provides synergies to facilitate implementation of the National Cohort. Through the joint location of study center and research divisions, conduct and oversight of the study can be optimized and the proximity to the laboratory and hospital buildings eases collaboration. Existing collaboration with several clinicians and associated scientists from the University Medical Center Freiburg and access to the local Clinical Cancer Registry, which is under the scientific direction of K. Michels, further integrate relevant elements of the National Cohort. If necessary, the proximity to the University Medical Center will make it possible to use the infrastructure available at the Medical Center for specialized examinations such as MRI.

The study center is located in the city center of Freiburg, which will ensure easy accessibility for the study participants by public transportation or by car, which will foster high participation rates.

Study center Saarbruecken

Table 7: Overview of participating institutions and investigators/scientists

Institution	German Cancer Research Center (DKFZ)
Town	69120 Heidelberg
Principal investigator	Prof. Dr. Hermann Brenner
Institute/department/division	Division of Clinical Epidemiology and Aging Research
e-mail	h.brenner@dkfz.de
<hr/>	
Institution	Saarland Cancer Registry
Town	66024 Saarbruecken
Co-principal investigator	Dipl. med. inform. Christa Stegmaier, R'din
Institute/department/division	
e-mail	krebsregister@gbe-ekr.saarland.de

Description of the institutions responsible for establishing and operating the study center

The main areas of research at the **Division of Clinical Epidemiology and Aging Research** at DKFZ include clinical cancer epidemiology, epidemiology of chronic age-related diseases, and epidemiologic methods. In the field of clinical cancer epidemiology, the division conducts large-scale epidemiologic studies on new avenues of more effective cancer prevention and early detection, and on issues of quality of medical care, prognosis, and quality of life of cancer patients. In the field of aging research, multiple large-scale, population-based and patient cohort studies are run in cooperation with regional public health and clinical partners. The division is actively involved in multiple national and international research consortia, such as the EU-funded Consortium on Health and Aging: Network of Cohorts in Europe and the United States (CHANCES), or the Emerging Risk Factors Collaboration. The division has extensive experience with external study centers, including a long-standing study center located in Saarbruecken next to the Saarland Cancer Registry, where long-term follow-up of statewide cohorts in Saarland is conducted.

The well-established statewide **Saarland Cancer Registry** and its excellent working relations with clinical and pathology partners, in particular those of the Medical Faculty of the University of Saarland, provide an excellent basis for specialized baseline and follow-up examinations and for follow-up of participants with respect to cancer incidence and validation of diagnoses.

Description of the study center Saarbruecken

The study center Saarbruecken will be run as an external study center of DKFZ's Division of Clinical Epidemiology and Cancer Research headed by H. Brenner. It will be affiliated with the Saarland Cancer Registry at the Ministry for Health and Consumer Protection and will be located in Saarbruecken, the state capital of Saarland. At this site, 10,000 inhabitants of the city of Saarbruecken and surrounding counties will be recruited. The location of the site in Saarbruecken guarantees good accessibility for the inhabitants of the city of Saarbruecken and adjacent counties.

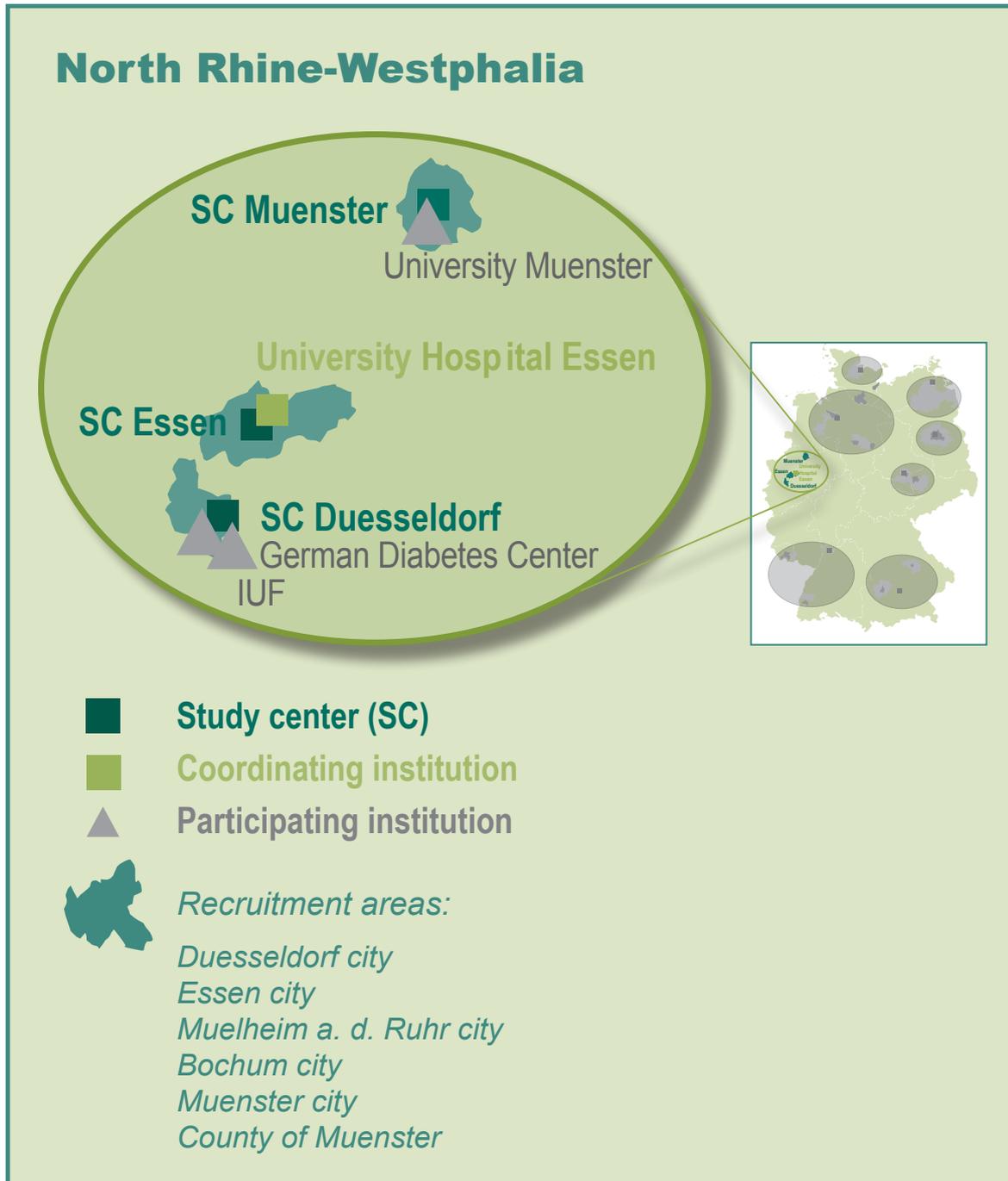
The close affiliation to the renowned Saarland Cancer Registry will facilitate the recruitment and will be instrumental for efficient follow-up of study participants, in particular for

identifying and tracing incident cancer cases. The study center will comprise rooms for investigation, interviews, and preanalytical preparation and interim storage of biosamples. There will be regular site visits of DKFZ scientists and staff at Saarbruecken (2 h by car). Implementation and running of the study center will benefit from the long-standing collaboration between DKFZ and the Saarland Cancer Registry and the excellent experience with external DKFZ study centers at Saarbruecken in preceding and ongoing projects. The vicinity of the Saarland Medical Faculty at Homburg/Saar (located between Heidelberg and Saarbruecken, 20 min from Saarbruecken by car) facilitates close cooperation with clinicians and associated scientists.

C.1.3 Cluster North Rhine-Westphalia

The Cluster North Rhine-Westphalia comprises three study centers in the area of the Rhine-Ruhr area and Westphalia, located in Essen, Duesseldorf, and Muenster.

Figure 3: Geographical overview of recruitment areas and study centers for the cluster North-Rhine-Westphalia



C.1.3.1 Cluster coordinator

Table 8: Cluster coordinator/co-coordinator

Coordinator	
Name	Prof. Dr. Karl-Heinz Jöckel
Institution Town	University of Duisburg-Essen, University Hospital Essen 45122 Essen
Institute/department/division	Institute for Medical Informatics, Biometry and Epidemiology (IMIBE)
Address, telephone, e-mail	Hufelandstr. 55, 45122 Essen, +49 201 723 4514 k-h.joeckel@uk-essen.de
Co-coordinator	
Name	PD Dr. Susanne Moebus MPH
Institution Town	University of Duisburg-Essen, University Hospital Essen Essen
Institute/department/division	Institute for Medical Informatics, Biometry and Epidemiology (IMIBE)
Address, telephone, e-mail	Hufelandstr. 55, 45122 Essen +49 201 723 4514 susanne.moebus@uk-essen.de

C.1.3.2 Study centers

Study center Essen

Table 9: Overview of participating institutions and investigators/scientists

Institution Town	University of Duisburg-Essen, University Hospital Essen 45122 Essen
Principal investigator Institute/department/division e-mail	Prof. Dr. Karl-Heinz Jöckel Institute for Medical Informatics, Biometry und Epidemiology (IMIBE) k-h.joeckel@uk-essen.de
Co-principal investigator Institute/department/division e-mail	PD Dr. Susanne Moebus MPH Institute for Medical Informatics, Biometry und Epidemiology (IMIBE) susanne.moebus@uk-essen.de
Associated scientist Institute/department/division e-mail	Dr. Nico Dragano Institute for Medical Informatics, Biometry und Epidemiology (IMIBE) nico.dragano@uk-essen.de

Description of the institution responsible for establishing and operating the study center

The **University of Duisburg-Essen** (UDE), created in 2003 by the merger of the universities of Duisburg and Essen, is the youngest university in North Rhine-Westphalia and the eighth largest university in Germany. Both campuses are easy to reach and offer some 34,000 students a broad academic spectrum with an international orientation – ranging from the humanities and social sciences to economics and the engineering and natural sciences and medicine. Students from 130 countries are currently enrolled at the UDE.

The **Medical Faculty of the University of Duisburg-Essen** emerged from the municipal clinics in 1963. With 1,800 students, of which about ten per cent are foreign, about 300 students successfully complete their medical studies annually. Fifty per cent of these graduates have gained a medical doctorate at the faculty. Research focuses on the main areas of medical research, namely oncology, immunology, transplantation, cardiovascular system and medical genetics and on the biological disciplines of genetics, bioinformatics, and developmental, cellular, molecular and structural biology. Work also includes medical technology solutions and new imaging procedures.

The **Institute of Medical Informatics, Biometry and Epidemiology** (IMIBE) is part of the Medical Faculty of the University of Duisburg-Essen and located on the campus of the University Hospital Essen. The IMIBE (K.H. Jöckel and S. Moebus) has considerable experience in performing large-scale cohort studies. A wide range of both population and patient based epidemiologic and clinical studies has been successfully performed by the Institute. One important example is the ongoing population based Heinz Nixdorf Recall cohort-study, which started in the year 2000 and is conducted in close cooperation with the Clinic for Cardiology. Experienced staff trained in data collection (interviews, examination, and medical information), and recruitment procedures (documentaries) is available, including extended field experience with established quality management procedures.

The Director of the IMIBE, hosting also the clinical tumor registry, is head of the Center for Clinical Trials Essen (ZKSE), with both institutions being affiliated with the Comprehensive Cancer Center (CCC) in Essen. The CCC offers treatment for a wide range of cancers using a multidisciplinary approach to patient care. The assignment is to reduce cancer incidence, morbidity, and mortality through treatment, training and research. The CCC emphasizes interdisciplinary basic, clinical, and translational cancer research and a research program comprising epidemiology, early detection and prevention.

The IMIBE has long-standing experience in evaluating research concepts and institutions, moderating the field work and research process, and consultancy for diverse institutions. The IMIBE will run the study center Essen, coordinated by S. Moebus and N. Dragano.

Description of the study center Essen

The study center of Essen is located in the center of the metropolitan Ruhr Area, along with London and Paris one of the largest conurbations in Europe. The Ruhr area forms the largest network of cities in Germany and is home to the largest proportion of first-, second- and third-generation migrants in Germany. Within the region, considerable differences exist regarding environmental burden and social factors at both the individual level and the community level.

At the study center in Essen, 10,000 study participants from the cities of Muelheim, Essen and Bochum, comprising 1.1 million inhabitants, will be recruited. The well-established epidemiologic assessment center *Robert-Koch-Erhebungszentrum* will serve as the study center for the National Cohort. The center is located across from the University Hospital with good connections to public transport.

The study center has rooms for clinical investigations, computer-assisted interviews, and preanalytical preparation and interim storage of biosamples (in total covering about 210 m²). The close cooperation with the Comprehensive Cancer Center in Essen, which is located at the University Hospital, ensures the efficient tracing of incident cancer cases. The close vicinity to the Medical Faculty in Essen facilitates close cooperation with clinicians and associated scientists from the faculty of Medicine.

The *Robert-Koch-Erhebungszentrum* is regularly certified and re-certified according to quality management procedures of DIN ISO 9001:2001/2008.

Figure 1: An ongoing examination at the Robert-Koch-Erhebungszentrum in Essen



Study center Muenster

Table 10: Overview of participating institutions and investigators/scientists

Institution	University of Muenster
Town	48149 Muenster
Principal investigator	Prof. Dr. Klaus Berger, MPH, MSc
Institute/department/division	Institute of Epidemiology and Social Medicine
e-mail	bergerk@uni-muenster.de
Co-principal investigator	Prof. Dr. Hans-Werner Hense
Institute/department/division	Institute of Epidemiology and Social Medicine
e-mail	hense@uni-muenster.de

Description of the institution responsible for establishing and operating the study center

The Institute of Epidemiology and Social Medicine is part of the medical faculty of the University of Muenster and located on the campus of the university hospital. It has been a WHO Collaborating Center for epidemiology and prevention of cardiovascular and other chronic diseases since 2000. A wide range of both population based and patient register based epidemiologic studies has been successfully performed by the institute. Experienced staff trained in data collection (interviews, examination, and medical information), and recruitment procedures (documentaries) is available. A close collaboration (in person) has been established with the Cancer Registry of North Rhine-Westfalia. Several ongoing epidemiologic studies are being led by members of the institute, in the fields of neurologic and psychiatric epidemiology, aging, cardiovascular epidemiology, and cancer. The institute will run the study center Muenster, coordinated by K. Berger and H.W. Hense.

Description of the study center Muenster

The Muenster area (county) includes a population of 2.6 million, of which 259,000 inhabitants live in the city of Muenster. The sampling frame will be the city and surrounding communities, which includes both urban and rural areas. At this site 10,000 participants will be recruited. The study center will be located in the city of Muenster with good accessibility for the study participants of the region. It will comprise rooms for investigation, interviews, and preanalytical preparation and interim storage of biosamples.

Study center Duesseldorf

Table 11: Overview of participating institutions and investigators/scientists

Institution	German Diabetes Center (DDZ)
Town	40225 Duesseldorf
Principal investigator	Prof. Dr. Guido Giani
Institute/department/division	Institute of Biometrics and Epidemiology
e-mail	giani@ddz.uni-duesseldorf.de
Institution	IUF - Leibniz Research Institute for Environmental Medicine
Town	40225 Duesseldorf
Co-principal investigator	Prof. Dr. Ursula Krämer
Institute/department/division	Division of Epidemiology
e-mail	kraemeru@uni-duesseldorf.de
Institution	German Diabetes Center (DDZ)
Town	40225 Duesseldorf
Associated scientist	Prof. Dr. med. Michael Roden
Institute/department/division	Institute of Clinical Diabetology
e-mail	michael.roden@ddz.uni-duesseldorf.de

Description of the institutions responsible for establishing and operating the study center

The **German Diabetes Center** (DDZ) incorporates three scientific institutes, the Institute for Clinical Diabetology, the Institute for Clinical Biochemistry and Pathobiochemistry, and the Institute for Biometrics and Epidemiology. The DDZ – a member of the Leibniz Association – is the only institution in Germany where interdisciplinary research areas that span from the causes and treatment of diabetes to epidemiology and prevention of the disease and its complications.

The **Institute for Clinical Diabetology** at the DDZ has a special research expertise in the area of combining multitracer dilution techniques and multinuclear magnetic resonance spectroscopy (MRS) to non-invasively study metabolic pathways related to insulin resistance and glucose metabolism. Together with the other institutes it is conducting a large prospective cohort study of newly diagnosed type 2 diabetic patients that includes extensive phenotyping of metabolic traits and early markers of diabetic complications. The main purpose of the Institute of Biometrics and Epidemiology is to generate current population-based epidemiologic data on diabetes and its complications. These research goals include health economics such as cost-effectiveness analyses of diabetes prevention. To achieve its objectives the Institute builds its own patient collectives (e.g., a nationwide type 1 diabetes registry) and uses platforms from external partners to implement diabetes-related research questions (KORA, SHIP). The Institute also has experience in establishing and analyzing large epidemiologic databases.

The **IUF – Leibniz Research Institute for Environmental Medicine** – major expertise is available in molecular preventive medical research in the area of environmental health. Molecular mechanisms of premature aging and degenerative diseases and environmentally induced impairment of immune reactions, especially allergies, are in the center of IUF's work. At the moment, research activities mainly focus on two environmental noxae: particles and nonionizing radiation. The internal scientific structure of IUF results from the cooperation of at present six working groups “cell biology“, “molecular immunology“, “molecular toxicology“, “particle research“, “epidemiology“, and “molecular aging research“. The working group “epidemiology” focuses on investigating the pathogenesis and development of respiratory diseases and allergies in children and premature aging processes and associated diseases (skin aging, COPD). The group has long-standing experience in conducting and analyzing epidemiologic studies, beginning in the 1970s.

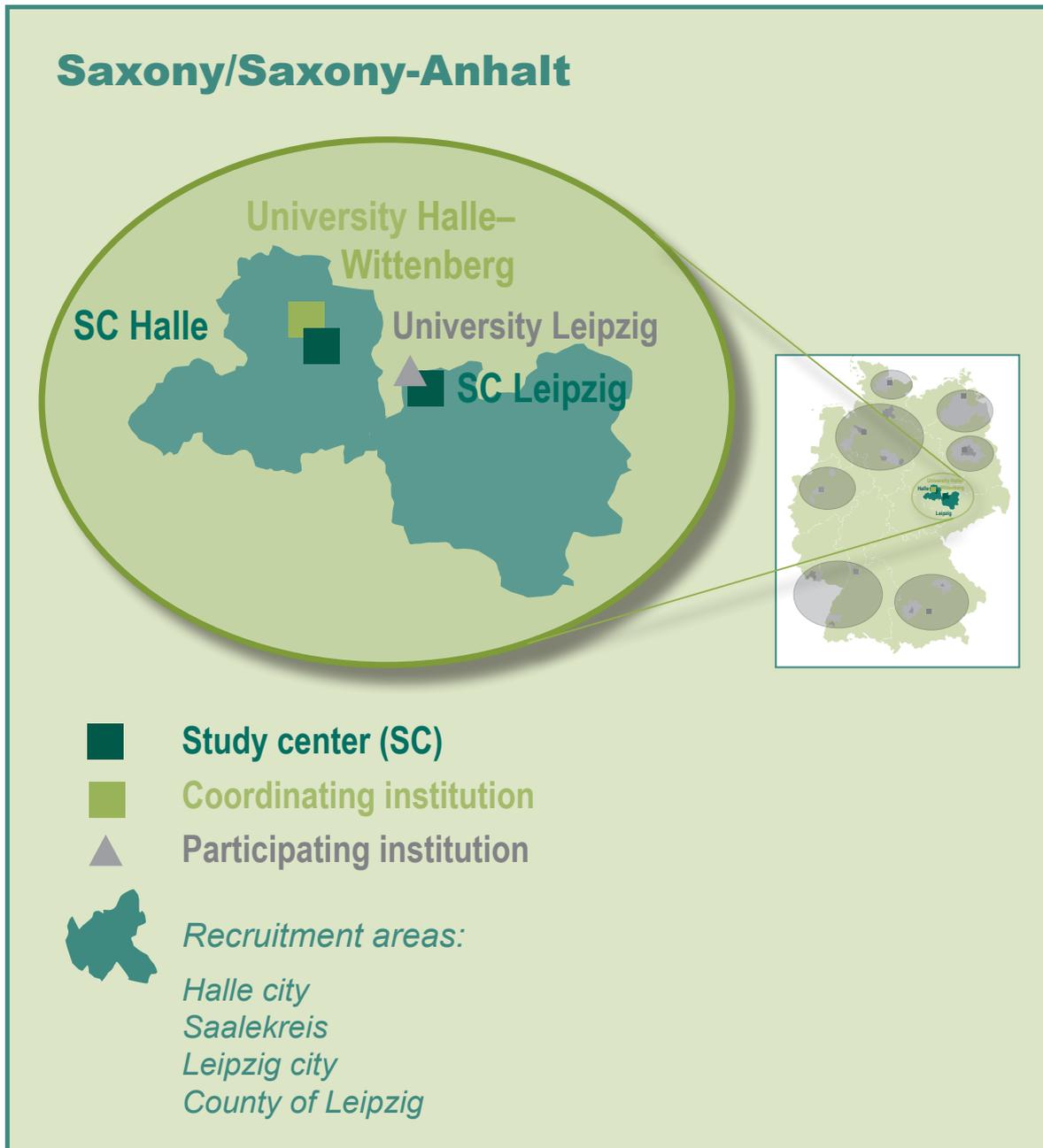
Description of the study center Duesseldorf

The recruitment of 10,000 study participants will take place in Duesseldorf, including inhabitants of the city of Duesseldorf (584,000 inhabitants in 2010). A total of 2,000 individuals will receive an extensive examination (e.g., including MRI and digital imaging of extrinsic skin aging). The study center will be located at the DDZ and will be run together with the IUF, which is located nearby. The DDZ can be easily reached by public transportation and is close to the main central railway station of Duesseldorf (10–15 min). The DDZ study center will comprise rooms for interviews, examinations, preparation, and storage of biosamples (in total covering more than 500 m²). The various existing facilities at the DDZ and IUF (MRI/MRS, technical laboratory, clinical study center, data management, quality control, training and supervision of investigators) will be embedded in the infrastructure.

C.1.4 Cluster Saxony/Saxony-Anhalt

The Cluster Saxony/Saxony-Anhalt comprises two study centers in the area of Central Germany (former GDR), located in Halle and Leipzig. The region Halle–Leipzig includes about 1.3 million inhabitants, comprising a source population (20–69 years) of about 930,000 inhabitants.

Figure 4: Geographical overview of recruitment areas and study centers for the cluster Saxony/Saxony-Anhalt



C.1.4.1 Cluster coordinator

Table 12: Cluster coordinator

Coordinator	
Name	Prof. Dr. Andreas Stang, MPH
Institution Town	Medical Faculty of the Martin-Luther-University of Halle-Wittenberg, 06112 Halle (Saale)
Institute/department/division	Institute of Clinical Epidemiology
Address, telephone, e-mail	Magdeburger Str. 8, 06112 Halle (Saale) + 49 345-557-3596 e-Mail: andreas.stang@medizin.uni-halle.de

C.1.4.2 Study centers

Study center Halle

Table 13: Overview of participating institutions and investigators/scientists

Institution Town	Medical Faculty of the Martin-Luther-University of Halle-Wittenberg, 06112 Halle (Saale)
Principal investigator Institute/department/division e-mail	Prof. Dr. Andreas Stang, MPH Institute of Clinical Epidemiology (IKE) e-mail andreas.stang@medizin.uni-halle.de
Co-principal investigator Institute/department/division e-mail	Prof. Dr. Johannes Haerting Institute of Medical Epidemiology, Biometry and Informatics (IMEBI) e-mail: johannes.haerting@medizin.uni-halle.de

Description of the institutions responsible for establishing and operating the study center

The **Institute of Clinical Epidemiology** (director: A. Stang) was originally founded as a section of the Institute of Medical Epidemiology, Biometry and Informatics at the Medical Faculty of the Martin-Luther-University of Halle-Wittenberg in 2004 with the appointment of A. Stang as the head of the section. In 2009, the Medical Faculty founded the Institute of Clinical Epidemiology, which is headed by A. Stang. The scientific work of the Institute is concentrating on clinical epidemiology, focusing mainly on evaluating diagnostic tests, screening and therapeutic interventions, and prognostic factors.

The **Institute of Medical Epidemiology, Biometry and Informatics** (IMEBI; director: J. Haerting) is part of the Medical Faculty of Halle University and of the Halle (Saale) Medical Center. The institute has extensive expertise in biostatistical and epidemiologic methods and in planning and conducting epidemiologic studies and clinical trials. J. Haerting is principal investigator of the CARLA study and head of the DSMC of the SHIP study Greifswald.

The Medical Faculty and University Hospital are located in Halle (Saale), in the center of Germany. With a history spanning over 300 years, it is one of the oldest medical schools in Germany. Germany's first university hospital was founded here. All medical specialties and subspecialties are represented. The Medical Faculty Halle is currently focusing on the main subjects oncology/tumor biology and clinical epidemiologic care and rehabilitation research. The Medical Faculty is part of the Martin Luther University Halle-Wittenberg, which is the largest university in the state of Saxony-Anhalt. At present, more than 18,000 students are enrolled at the university, which encompasses over 190 courses of study in 18 faculties and departments.

Description of the study center Halle

The city of Halle and the neighboring administrative district (Saalekreis) includes a population of about 440,000 inhabitants. The source population comprises about 300,000 inhabitants aged 20–69 years. In Halle 10,000 individuals will be recruited and will mainly comprise inhabitants of the city of Halle and a subsample of the Saalekreis. The study center is located at the inner-city campus of the University Hospitals of the University of Halle in vicinity of IKE and IMEBI and is easily accessible by the local public transport system. The study center will comprise rooms for investigation, interviews and preanalytical preparation and interim storage of biosamples (in total covering about 500 m²). Furthermore, very close to the study center, infrastructure is available for specialized examinations such as MRI. Currently, the study center is being renovated, incurring costs of about 650,000 € and funded by the state of Saxony-Anhalt. This renovation is going to be finalized by June 2011.

Study center Leipzig

Table 14: Overview of participating institutions and investigators/scientists

Institution	Medical Faculty, University of Leipzig and University Hospital Leipzig
Town	04103 Leipzig
Principal investigator	Prof. Dr. Markus Löffler
Institute/department/division	Institute for Medical Informatics, Statistics, and Epidemiology
e-mail	markus.loeffler@imise.uni-leipzig.de
Co-principal investigator	Prof. Dr. Joachim Thiery
Institute/department/division	Institute of Laboratory Medicine, Clinical Chemistry, and Molecular Diagnostics
e-mail	joachim.thiery@medizin.uni-leipzig.de

Description of the institutions responsible for establishing and operating the study center

The University of Leipzig is the second oldest university in Germany. In its 600-year history a broad spectrum of scientific disciplines was established, with special emphasis on the humanities, natural sciences, and medicine. The university is dedicated to the pursuit of excellence in research and teaching. The Medical Faculty of the University of Leipzig is one of the oldest in Germany, tightly interwoven with the university hospitals, it offers comprehensive diagnostics and therapy. With a central medical campus and several off-campus institutions,

such as the renowned heart center, more than half a million inpatients and outpatients are served per year.

The **Institute of Medical Informatics, Statistics, and Epidemiology (IMISE)** comprises strong expertise in clinical trials and biometry, medical informatics, and systems biology and bioinformatics. IMISE is involved in research networks such as “Verbundprojekt ‘Familärer Darmkrebs’” and “Deutsche Studiengruppe für hochmaligne Non-Hodgkin Lymphome” (DSNHL). IMISE offshoots include the Center for Clinical Studies (ZKS), conducting more than 100 investigator-initiated or industry-led clinical trials, and the Interdisciplinary Center for Bioinformatics (IZBI). A main goal of IMISE is to contribute to improving health care and medical science through high-quality clinical trials and research into pathomechanisms of disease. In addition to the scientific expertise required in this context, a strong IT infrastructure has been established.

The **Institute of Laboratory Medicine, Clinical Chemistry, and Molecular Diagnostics (ILM)** with its accredited laboratories provides a wide range of laboratory diagnostic services for the university hospitals and beyond. ILM performs newborn screening for the states of Saxony and Thuringia, is involved in numerous clinical trials as the central laboratory, and has a strong research agenda, for instance, for CVD.

IMISE and ILM have been leading partners in establishing the “Leipzig Research Center for Diseases of Civilization” (LIFE). In LIFE, population cohorts of children and adults are recruited along with several disease cohorts, totaling more than 20,000 individuals. Detailed phenotypic characterization of participants comprises questionnaires and technical assessments of sociodemographics, disease history, anthropometry, and clinical and laboratory measurements. LIFE brings together scientists from many institutes and clinics and has strong links to the IFB Adiposity Diseases and the Max-Planck-Institute for Human Cognitive and Brain Sciences in Leipzig.

IMISE in conjunction with ILM will run the study center Leipzig. Additional epidemiologic expertise is available from the study center Halle. We plan to recruit 10,000 participants for the National Cohort. In addition to level 1 assessments, assessments for levels 2 and 3 are envisioned.

Description of the study center Leipzig

In Leipzig 10,000 individuals will be recruited and will include inhabitants of the City of Leipzig with its population of 500,000. A subcohort comprising inhabitants of surrounding rural areas may be established. The study center has been established on the medical campus of the University of Leipzig, about 15 min walking distance to the center of the city. It is directly accessible by several tram lines, and a bus line and by car, linking the center to all neighborhoods of Leipzig as well as to the main railway station.

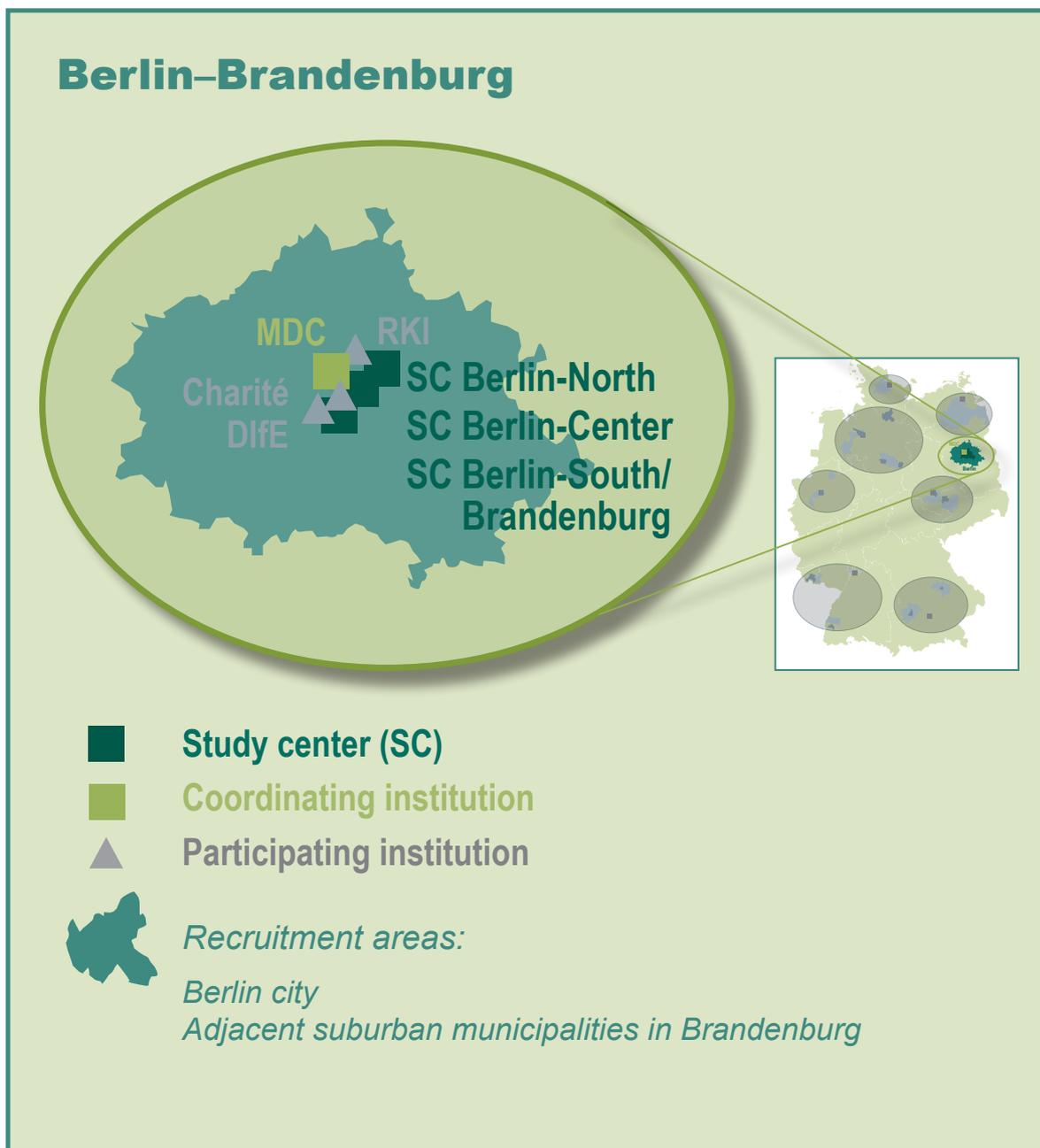
The study center comprises rooms for interviews, questionnaires, anthropometry, ultrasound, echocardiography, ECG, physical assessments, and preanalytical preparation and interim storage of biosamples. A direct pneumatic delivery link to the main laboratory exists. Currently, IT infrastructure is being established for large cohort studies including data entry and storage, data transfer from instruments and instant laboratory measurements, and tracking of biomaterials.

The location of the study center on the medical campus of the University of Leipzig facilitates close cooperation with clinicians and scientists and provides options for additional assessments for subcohorts in specialized facilities.

C.1.5 Cluster Berlin-Brandenburg

The Cluster Berlin-Brandenburg consists of the Max Delbrueck Center for Molecular Medicine (MDC), the Charité – University Medical Center Berlin, the German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), and the Robert Koch Institute (RKI). The Cluster comprises three study centers that are located at or near the sites of the partner institutions (Berlin-North, located in the northern part of Berlin; Berlin-Center, located in the central part of Berlin; and Berlin-South, located in the southern part of Berlin). They are spread across the study region, the metropolitan area of Berlin-Brandenburg, thereby minimizing costs and time for study participants to travel to the centers and increasing their willingness to participate. Within a radius of 60 km, these study centers cover a total source population of 4.4 million inhabitants (3.4 million in Berlin and 1.0 million in the adjacent districts of Brandenburg).

Figure 5: Geographical overview of recruitment areas and study centers for the cluster Berlin-Brandenburg



C.1.5.1 Cluster coordinator

Table 15: Cluster coordinator

Coordinator	
Name	Prof. Dr. Tobias Pischon, MPH
Institution	Max Delbrueck Center for Molecular Medicine (MDC) Berlin-Buch
Town	13125 Berlin
Institute/department/division	Molecular Epidemiology Group
Address, telephone, e-mail	Robert Roessle Straße 10, 13125 Berlin + 49 30 9406 4563 tobias.pischon@mdc-berlin.de

C.1.5.2 Study centers

Study center Berlin-North

Table 16: Overview of participating institutions and investigators/scientists

Institution	Max Delbrueck Center for Molecular Medicine (MDC) Berlin-Buch
Town	13125 Berlin
Principal investigator	Prof. Dr. Tobias Pischon, MPH
Institute/department/division	Molecular Epidemiology Group
e-mail	tobias.pischon@mdc-berlin.de

Description of the institution responsible for establishing and operating the study center

The **Max Delbrueck Center for Molecular Medicine (MDC) Berlin-Buch** combines basic research in molecular biology and genetics with clinical research in the fields of cardiovascular and metabolic diseases, cancer, and function and dysfunction of the nervous system. The MDC and the Charité – University Medical Center Berlin have jointly established the Experimental and Clinical Research Center (ECRC), which is dedicated to translational medical research, in particular human subjects-oriented mechanistic research. Based on this research, MDC scientists and clinicians jointly aim at developing new methods for the diagnosis, prevention, and treatment of cancer, cardiovascular, and neurological diseases.

With the appointment of a new professorship and the development of a new research group, MDC has recently expanded its research portfolio to the field of Molecular Epidemiology, with the aim of studying at the molecular level the relationship of genetic, metabolic, and environmental factors with risk and outcome of cardiovascular and metabolic diseases in human populations.

Together with the Charité – University Medical Center Berlin, the German Institute of Human Nutrition (DIfE), and the Robert-Koch-Institute, MDC has built up the epidemiologic research network for the metropolitan area Berlin-Brandenburg (EPI2B), which combines

the existing expertise and thereby provides a platform for epidemiologic research to support the National Cohort.

Being part of the Helmholtz association, MDC is among the Helmholtz centers that have taken the initiative for setting up a prospective cohort for chronic disease epidemiology in Germany (Helmholtz cohort). An international epidemiology expert panel evaluated this proposal positively in 2008, which provided specific start-up funds for the planning of the cohort at MDC along with the other Helmholtz institutions involved. As part of the planning phase of the Helmholtz cohort, MDC has set up the working group on cardiovascular diseases which, based on the expertise of its more than 40 members, developed between 2009 and 2010 an expert report with detailed suggestions for scientific approaches and protocols for studying cardiovascular disease in the National Cohort.

MDC will coordinate the Cluster Berlin-Brandenburg and establish and operate the study center Berlin-North. In addition, MDC will provide the infrastructure for storage and maintenance of the biological samples within the cluster.

Description of the study center Berlin-North

The study center will be located in Berlin-Buch in rooms of the Experimental and Clinical Research Center (ECRC). Together with the study centers Berlin-Center and Berlin-South this study center will cover the metropolitan area Berlin-Brandenburg, with a total source population of 4.4 million inhabitants (3.4 million in Berlin and 1.0 million in the adjacent districts of Brandenburg). These three study centers, each recruiting 10,000 study participants, will help to minimize costs and time for the study participants to travel to the centers and increase their willingness to participate.

Study center Berlin-Center

Table 17: Overview of participating institutions and investigators/scientists

Institution	Charité - University Medical Center Berlin
Town	10117 Berlin
Principal investigator	Prof. Dr. Stefan N. Willich, MPH, MBA
Institute/department/division	Institute for Social Medicine, Epidemiology, and Health Economics
e-mail	stefan.willich@charite.de
Co-principal investigator	Prof. Dr. Peter Heuschmann
Institute/department/division	Center for Stroke Research Berlin (CSB)
e-mail	peter.heuschmann@charite.de
Associated scientist	PD Dr. Thomas Keil, MSc
Institute/department/division	Institute for Social Medicine, Epidemiology, and Health Economics
e-mail	thomas.keil@charite.de

Description of the institution responsible for establishing and operating the study center

The **Charité – University Medical Center Berlin** is one of the largest university clinics in Europe. Scientists and physicians engage in state-of-the-art research, patient care, and education at this institution. More than half of the German Nobel laureates in medicine and physiology come from the Charité, among them Emil von Behring, Robert Koch and Paul Ehrlich. The Charité also has an international reputation for excellence in training. It extends over four campuses with more than 100 clinics and institutes bundled under 17 CharitéCenters.

The Institute for Social Medicine, Epidemiology, and Health Economics is the largest institute at the CharitéCenter 1 for Health and Human Sciences. The institute initiated and coordinated a number of large prospective outpatient-population cohort studies, including the ORBITAL Study (Open Label Primary Care Study of Rosuvastatin-based Compliance Initiatives to Achievement of LDL Goal) on secondary prevention of CVD with over 8,000 patients that were recruited and followed nationwide and followed up for 5 years, and the Acute Stroke Care Study (n= 1094, Competence Net Stroke), which aimed to improve the care of patients with acute stroke in the Berlin inner city districts. Moreover, the Institute has been involved in several large prospective birth cohort studies on chronic respiratory and allergic disorders, such as coordination of the 20-year follow-up of the Multi-center Allergy Study (MAS, n=1,300), the first German birth cohort study on the development of atopic diseases from childhood to adulthood, and the EuroPrevall birth cohort study (nine countries, n=12,500) investigating lifestyle, environmental, and genetic factors of food allergy (EU FP6). Furthermore, the institute has been coordinating a collaboration of more than 20 ongoing European birth cohorts as part of the EU Network of Excellence GA2LEN (Global Allergy and Asthma European Network) and is currently organizing a harmonized follow-up assessment among the older European birth cohorts that reached adolescence (MeDALL, EU FP7). In 2007 the Institute established the Charité Ambulanz für Prävention und Integrative Medizin (CHAMP). CHAMP is designed as a research health care center in the fields of prevention and integrative medicine that aims to bridge the gap between science and practice and between traditional and alternative medicine.

The Center for Stroke Research Berlin (CSB) was funded by the Federal Ministry of Education and Research as an integrated research and treatment center. The CSB investigators have extensive experience in determining subclinical and clinical vascular disease phenotypes, designing and coordinating prospective hospital- and population-based stroke registries, performing large patient cohort studies, and combining large data-sets (e.g., German Stroke Register Study Group, prospective CSB cohort, EU FP7 funded European Implementation Score Collaboration). The CSB performs investigator-initiated stroke trials based on their discoveries (e.g., PANTHERIS) and is also coordinating a number of nationally and internationally recognized scientific stroke programs (e.g., EU FP7 funded ARISE network performing biomarker studies). Standards have been developed and validated for the determination of blood pressure, hematological, biochemical, and immunological studies, as well as the whole range of subclinical and clinical vascular disease phenotypes including access to advanced clinical diagnostic facilities (e.g., brain and vascular imaging). As mentioned above, Charité and MDC have jointly established the Experimental and Clinical Research Center (ECRC), as a platform for translational medical research, and, in addition, together with MDC, DIfE, and RKI the epidemiologic research network for the metropolitan area Berlin-Brandenburg (EPI2B) to support the National Cohort.

As part of the planning phase, scientists were actively engaged in several of the working groups that provided detailed suggestions for scientific approaches and protocols for the most important chronic diseases to be addressed within the National Cohort. The Charité will establish and operate the study center Berlin-Center.

Description of the study center Berlin-Center

The study center will be located in Berlin-Center in rooms of the Charité, Campus Center. Together with the study centers Berlin-North and Berlin-South this study center will cover the metropolitan area Berlin-Brandenburg (see above). These three study centers will help to minimize costs and time for the study participants to travel to the centers and increase their willingness to participate.

Study center Berlin-South/Brandenburg

Table 18: Overview of participating institutions and investigators/scientists

Institution	German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE)
Town	Potsdam-Rehbruecke, 14558 Nuthetal
Principal investigator	Prof. Dr. Heiner Boeing, MSPH
Institute/department/division	Department of Epidemiology
e-mail	boeing@dife.de
Associated scientist	Dr. Cornelia Weikert, MPH
e-mail	Department of Epidemiology weikert@dife.de

Description of the institution responsible for establishing and operating the study center

The **German Institute of Human Nutrition at Potsdam-Rehbruecke (DIfE)** is conducting studies in the field of nutrition and health with the aim of understanding the molecular basis of nutrition-dependent diseases and to develop new strategies for prevention, treatment, and nutritional recommendations. It is a member of the Leibniz Association, an alliance of scientific institutions in Germany. The Department of Epidemiology, headed by H. Boeing, established from 1994 to 1998 the ongoing EPIC-Potsdam Cohort which has been followed prospectively since recruitment. The cohort consists of initially 27,548 population-based participants from the Potsdam area (south of Berlin) and is part of the EU-wide European Prospective Investigation into Cancer and Nutrition (EPIC) with a total of 519,000 subjects. The EPIC-Potsdam study is the largest prospective study so far in Germany on life style and risk of chronic diseases. The local EPIC-Potsdam study has a record of 12 years of follow-up of the cohort for diseases such as type 2 diabetes, MI, stroke, hypertension, heart failure, and cancer. The follow-up procedure includes biannual questionnaires sent to the study participants, record linkage with the local hospitals and cancer registries, and validation of potential diagnoses by medical records and treating physicians. Within the framework of the EPIC-Potsdam study, blood samples have been collected from more than 26,000 participants and are stored in a cryobank that is located at DIfE and maintained by the principal investigator. The publications of investigators from this cohort have mainly addressed the effects of diet, anthropometry, metabolic and genetic factors, and physical activity on end-points such as mortality, cancer, type 2 diabetes, and CVD.

As mentioned above, DIfE has jointly established with MDC, Charité, and RKI the epidemiologic research network for the metropolitan area Berlin-Brandenburg (EPI2B) to support the National Cohort.

As part of the planning phase of the Helmholtz cohort, DIfE has (together with the German Cancer Research Center, DKFZ) set up the working group physical activity and (together with the Helmholtz-Center Munich) the working group on diet. The two working groups could make use of the expertise of their about 30 members and developed an expert report with detailed suggestions for scientific approaches and protocols for Diet and Physical Activity to be addressed in the National Cohort. The DIfE will establish and operate the study center Berlin-South.

Description of the study center Berlin-South/Brandenburg

The study center will be located in Berlin-South at the Campus Benjamin Franklin. In this campus, DIfE already supports a metabolic unit established due to a joint professorship with the Charité. The study center in Berlin-South as well as the study centers Berlin-North and Berlin-Center cover the EPI2B metropolitan area Berlin-Brandenburg. It offers nearby examination facilities with a minimum of travel time for those participants living in the southern part of the study area.

Study support center

Table 19: Overview of participating institutions and investigators/scientists

Institution	Robert Koch Institut (RKI) Berlin
Town	13302 Berlin
Principal investigator	Dr. Bärbel-Maria Kurth
Institute/department/division	Department of Epidemiology and Health Reporting
e-mail	KurthB@rki.de
Associated scientist	PD Dr. Martin Schlaud
e-mail	SchlaudM@rki.de

Description of the institution

The **Robert Koch Institute (RKI)** is the central federal institution responsible for disease control and prevention and is therefore the central federal reference institution for both applied and response-oriented research and for the public health sector. The Department of Epidemiology and Health Reporting of the RKI is responsible for health reporting to the German federal government and thus for initiating and conducting nationwide health surveys, many of which are embedded in the European Health Monitoring Network. RKI has been conducting large population-based studies for more than 25 years, starting with cross-sectional prevalence studies in the 1980s, and later prospective cohort studies in adults (1990s) and children/adolescents (2000s). Since 2008, a national health monitoring system has been established at the RKI. In this system, representative health interview and examination surveys are conducted according to a schedule of fixed intervals. The aim of the health surveys is to collect up-to-date information on the health and health behavior of the German population, to determine trends over time and to identify risk factors. Furthermore, the additionally integrated panel component makes it possible to reveal temporary sequences, developmental trajectories, and potential causal relations. Currently, one cohort study in adults (n = 8,000) and one cohort study in children/adolescents (n = 17,500) are ongoing. The full planning, development, and testing of survey instruments, recruitment, address management, field work,

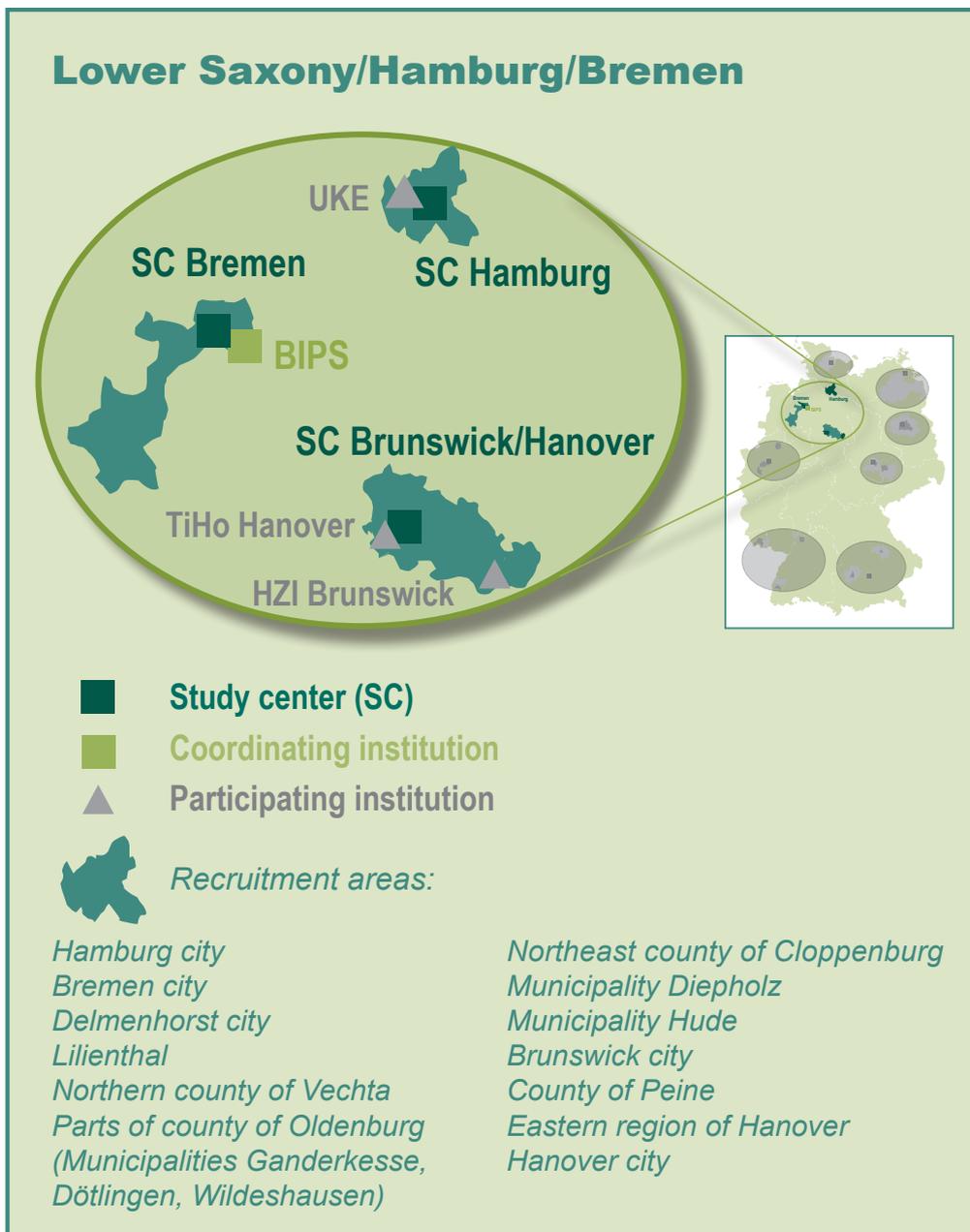
laboratory testing, quality and data management, feed-back of individual results, statistical analysis, critical evaluation, and publication of results is being done by the RKI.

Together with MDC, DIfE, and Charité, RKI has jointly established the epidemiologic research network for the metropolitan area Berlin-Brandenburg (EPI2B) to support the National Cohort. Within the context of the National Cohort and based on its extensive experience, the RKI will support the recruitment of study participants.

C.1.6 Cluster Lower Saxony/Hamburg/Bremen

The Cluster Lower Saxony/Hamburg/Bremen comprises three study centers in the area of Lower Saxony (Brunswick/Hanover) and the city states of Bremen and Hamburg.

Figure 6: Geographical overview of recruitment areas and study centers for the cluster Lower Saxony/Hamburg/Bremen



C.1.6.1 Cluster coordinator

Table 20: Cluster coordinator

Coordinator	
Name	Prof. Dr. Wolfgang Ahrens
Institution Town	Bremen Institute of Prevention Research and Social Medicine (BIPS) 28359 Bremen
Institute/department/division	Department of Epidemiologic Methods and Etiologic Research
Address, telephone, e-mail	Achterstr. 30, D-28359 Bremen +49 421 218-56822 ahrens@bips.uni-bremen.de

C.1.6.2 Study centers

Study center Brunswick/Hanover

Table 21: Overview of participating institutions and investigators/scientists

Institution Town	Helmholtz Center for Infection Research (HZI) 38124 Brunswick
Principal investigator Institute/department/division e-mail	PD Dr. Dr. Frank Pessler (provisional) Department of Infection Genetics Frank.Pessler@helmholtz-hzi.de
Institution Town	University of Veterinary Medicine Hanover, Foundation 30559 Hanover
Co-principal investigator* Institute/department/division e-mail	Prof. Dr. Lothar Kreienbrock Department of Biometry, Epidemiology and Information Processing Lothar.Kreienbrock@tiho-hannover.de
Institution Town	Helmholtz Center for Infection Research (HZI) 38124 Brunswick
Associated scientist Institute/department/division e-mail	Dr. Manas Akmatov Department of Infection Genetics Manas.Akmatov@helmholtz-hzi.de

*The role includes active participation in field work and research activities

Institution	University of Veterinary Medicine Hanover, Foundation
Town	30559 Hanover
Associated scientist	Katja Hille
Institute/department/division	Department of Biometry, Epidemiology and Information Processing
e-mail	Katja.Hille@tiho-hannover.de
Associated scientist	Dr. Roswitha Merle
Institute/department/division	Department of Biometry, Epidemiology and Information Processing
e-mail	Roswitha.Merle@tiho-hannover.de

Description of the institutions responsible for establishing and operating the study center

The **Helmholtz Center for Infection Research (HZI)** in Brunswick is a multi-department research institute focusing on diverse aspects of infectious disease research, including genetics, immunology, and vaccinology. More than 700 staff members, including medical scientists, chemists, biologists, and biotechnologists, work in over 40 research groups.

HZI has long-standing expertise in fundamental aspects of microbial pathogen research, including: characterization of bacterial and viral pathogens, immunophenotyping (e.g., flow cytometry, and immunohistochemistry), large-scale sequencing, proteomics and microarrays, and automated high-throughput screening of biologically active compounds. Within the year 2011, a senior infection epidemiologist with extensive experience in population-based infectious disease research will assume the leadership of the newly founded HZI Department of Epidemiology and will become PI of the study center Brunswick/Hanover. The provisional PI for the study center, F. Pessler, will function as Co-PI after that time. Currently, the epidemiology research group at HZI is working full time for the National Cohort. It consists of a physician-scientist project leader, one full-time post doctoral fellow, one doctoral student, and two laboratory technicians. This group collaborates closely with laboratory based-scientists at HZI in planning aspects of the National Cohort relating to infection and immunity. It also coordinates the cohort's working groups "infection/immunity" (together with the Robert Koch Institute) and "musculoskeletal system/autoimmunity", and participates in the working group "biorepository".

The epidemiologists at HZI have participated in nationwide prospective cohort studies on clinical manifestations and outcomes of chronic inflammatory disorders (e.g., the German Center for Rheumatology Research JIA Core Documentation, the German Pediatric Enbrel Registry, and the Dresden JIA Cohort), and have been conducting prospective studies on the occurrence of acute respiratory infections in humans. Moreover, population-based longitudinal, cross-sectional, and case-control studies have been and are being performed by associated scientists at HZI. They are therefore highly experienced in the use of standardized instruments for clinical and epidemiologic research.

The **University of Veterinary Medicine Hanover (TiHo)**, founded in 1778, is the only self-contained veterinary university in Germany. Six clinics, 20 departments, and two research stations in rural Lower Saxony belong to the University.

The Department of Biometry, Epidemiology and Information Processing (IBEI-TiHo) has experience in designing, conducting and analyzing population-based studies in human and an-

imal populations. IBEI-TiHo hosts the WHO-Collaborating Center for Research and Training in Veterinary Public Health and therefore has expertise in the research field of veterinary medicine in a public health context. The principal investigator at IBEI-TiHo, L. Kreienbrock, has international experience in planning, conducting, and analyzing epidemiologic studies. Next to the head of the institute the TiHo-staff comprises two veterinary specialists in epidemiology ('Fachtierarzt'), two biometricians (post doc), two statisticians, nine doctoral students, and four medical record managers. Currently, IBEI-TiHo is a member of the research network "FBI-Zoo" – Foodborne human infections of zoonoses and coordinator of the network "RESET ESBL and (fluoro)quinolone resistance in Enterobacteriaceae", both funded by the German Ministry for Research.

IBEI-TiHo will provide its epidemiologic expertise in the field of recruitment, collection and storage of biological samples in infection epidemiology. Particular experience has been gained in co-operations such as FBI-Zoo or multicenter studies of sporadic bacterial infections and also in population-based studies in cancer epidemiology or in studies in Veterinary Public Health dealing with farmers as source population. Moreover, experience in setting up and supporting large and complex databases and also in the statistical analyses of epidemiologic studies exists. This will be incorporated into the study, especially within the framework of zoonoses research.

Description of the study center Brunswick/Hanover

The proposed area of recruitment comprises the cities of Hanover and Brunswick, the county of Peine as well as some counties east of Hanover which can be reached in less than 1 h by train or car. Approx. 800,000 inhabitants aged 20–69 years live in this area and represent the source population from which 10,000 subjects are intended to be enrolled.

The study center will be located in a newly constructed, state-of-the-art complex dedicated to clinical and epidemiologic studies, the Hanover Center for Translational Medicine (HCTM). HCTM is jointly operated by the Fraunhofer Society, the Hanover Medical School (MHH) and HZI and will begin operations in January 2013. The outpatient research facilities of the adjacently located Fraunhofer Institute will be used until then. The HCTM will be located in Hanover on the campus of MHH and will contain all facilities for interventional and noninterventional clinical and epidemiologic studies. This four-level complex will include 650 m² for outpatient examination and interview rooms, 550 m² for imaging facilities (including MRI, sonography, and DEXA), laboratories licensed for pathogen safety level S2, a flow cytometry unit, and a 500 m² biorepository. A vacuum tube system will make it possible to send microbial specimens directly to the microbiological laboratories of MHH. Office space will be available on the level above the examination rooms, thus placing research staff in close proximity to the study facilities. Ample parking space will be available for participants arriving by car. HCTM will be located within walking distance from the bus (5 min) and tram (10 min) stops. A taxi shuttle from public transportation to HCTM will be offered to all participants to ensure high participation rates, particularly of older participants.

Study center Hamburg

Table 22: Overview of participating institutions and investigators/scientists

Institution	University Medical Center Hamburg-Eppendorf
Town	20246 Hamburg
Principal investigator	Prof. Dr. Dieter Flesch-Janys
Institute/department/division	Hubertus-Wald-Tumor-Center – University Cancer Center Hamburg Department of Cancer Epidemiology and Clinical Cancer Registry and Department of Medical Biometry and Epidemiology
e-mail	flesch@uke.de
Co-principal investigator	Dr. Nadia Obi
Institute/department/division	Department of Cancer Epidemiology and Clinical Cancer Registry
e-mail	n.obi@uke.de

Description of the institutions responsible for establishing and operating the study center

The **University Cancer Center Hamburg (UCCH)** is the central organization of comprehensive cancer care and research at the University Medical Center Hamburg-Eppendorf (UKE). UCCH is distinguished as one of only ten research and care centers nationwide to be designated as a Center of Excellence in Oncology by the German Cancer Aid (Deutsche Krebshilfe), the leading private sponsor of cancer research in Germany. Our mission is to provide patients with the best comprehensive care available today, and to accelerate the development of enhanced therapeutic, diagnostic and preventive means to fight cancer more effectively tomorrow. The focus of the Department of Cancer Epidemiology and Clinical Cancer Registry is identifying preventable causes of cancer and the establishing and evaluating interventions and feasibility studies for cancer prevention. In addition it runs the Clinical Cancer Registry for the UKE. The department has longstanding experience in large epidemiologic studies. It homes the MARIE study for identifying preventable breast cancer risk factors, the MARIEplus study for identifying of modifiable prognostic factors as well as a cohort study in postmenopausal women for identifying determinants and changes in early detection behaviour.

The department cooperates closely with the Hamburg Cancer Registry, the oldest cancer registry in Germany and with the Department of Medical Biometry and Epidemiology (headed by K. Wegscheider). There is long-standing collaboration with the central inhabitant registry for recruiting representative population samples and for follow-up. Experienced staff trained in data collection (interviews, examination, and medical information), and recruitment procedures (documentaries) is available. The department can draw on the laboratory unit for the management of blood sampling and large biorepositories are available at the UKE. Facilities for specialized examinations (MRI) are available in-house.

Description of the study center Hamburg

At the University Medical Center Hamburg, 10,000 individuals will be recruited. The center is located close to the city center of Hamburg. It is directly connected via two bus lines operating every 5 min and the underground is nearby. A central garage is available directly beneath the building.

The study center is a separate building on the campus of the university medical center. It comprises rooms for reception, investigations, interviews and preanalytical preparation and interim storage of biosamples (in total covering about 400 m²). The Department of Radiology for possible MRI examinations is located directly on the opposite side.

Study center Bremen

Table 23: Overview of participating institutions and investigators/scientists

Institution	Bremen Institute of Prevention Research and Social Medicine (BIPS)
Town	28359 Bremen
Principal investigator	Prof. Dr. Wolfgang Ahrens
Institute/department/division	Department of Epidemiologic Methods and Etiologic Research
e-mail	ahrens@bips.uni-bremen.de
Co-principal investigator	PD Dr. Thomas Behrens
Institute/department/division	Department of Epidemiologic Methods and Etiologic Research
e-mail	behrens@bips.uni-bremen.de
Associated scientist	Stefanie Dreger
Institute/department/division	Department of Epidemiologic Methods and Etiologic Research
e-mail	dreger@bips.uni-bremen.de
Institution	Georg-August-University Goettingen, Medical School
Town	37073 Goettingen
Associated scientist	Prof. Dr. Heike Bickeböller
Institute/department/division	Department of Genetic Epidemiology, Center for Informatics, Statistics and Epidemiology
e-mail	hbickeb@gwdg.de
Associated scientist	Prof. Dr. Peter Falkai
Institute/department/division	Department of Psychiatry and Psychotherapy
e-mail	pfalkai@gwdg.de

Description of the institutions responsible for establishing and operating the study center

The **Bremen Institute for Prevention Research and Social Medicine (BIPS)** was established on January 1, 1981, and joined the University of Bremen as a central research unit in 2007. On November 28, 2009 BIPS became an associated member of the Research Community Gottfried Wilhelm Leibniz e.V. (Leibniz Community). As an epidemiologic research institute BIPS covers the entire spectrum of epidemiologic research, including the description of health and health-related factors in populations, identification of determinants of health and disease, and implementation and evaluation of health interventions and the associated transfer of results to politics and citizens. For that purpose employees from various fields and qualifications are cooperating within the institute. The staff at BIPS comprises 44 scientists, 48 nonscientists and 35 student assistants who are working in four departments. The BIPS has long-standing experience in coordinating and executing large population-based cohort studies within national and international cooperations. The ongoing IDEFICS study ("Identification and prevention of dietary- and lifestyle-induced health effects in children and infants") was set up in 11 countries across Europe and is currently being coordinated by BIPS. The study is following the health status of approx. 16,000 2- to 9-year-old children. Furthermore, BIPS has coordinated the external quality control of field work and organized and coordinated the central field work training for all 16 study centers. Several studies with mortality follow-up have been conducted at BIPS. To obtain information on causes of death from death certificates, standard data protection procedures have been developed and applied in various studies, e.g., in the BIPS Bitumen Cohort Study. Procedures also comprise identification of next-to-kin subjects. The epidemiologic cancer registry of the State of Bremen is located at BIPS. It is complemented by the Bremen death index (Bremen Mortality Index, BreMI) for the State of Bremen, which registers the cause of death and all information from the death certificate of all Bremen residents since 01.01.1998 in an electronic database. The information in the BreMI database can be personally linked with subjects in a cohort study to establish the cause of death for cohort members residing in Bremen. According to §9, Abs. 6 Nr. 2 Gesetz des Leichenwesens, the BREMI may be used for public research projects. In addition, BIPS hosts a pharmacoepidemiologic database covering 14 mill. members (or roughly 1/6 of the entire German population) enrolled in four Statutory Health Insurances (SHIs). About 85% of the total population in Germany is registered as member of one of these SHIs.

In its long standing experience in coordinating and conducting large national and international population-based studies, BIPS has acquired special expertise in biosample quality management and biosample logistics (shipping, storage and electronic documentation of samples).

BIPS has an experienced data management unit that develops tools for the implementation of studies and surveys. A set of routines and procedures for data management, questionnaires for computer-assisted interviewing, and a study management system to operate and supervise data assessment are under continuous development. These tools are also employed outside the institute in multicentric, international studies. A special software tool to control compliance with standard contact procedures and to monitor follow-up of study participants has been designed to standardize all field procedures, increase response proportion, and facilitate quality assurance of contact procedures. This electronic contact documentation records all attempted contacts with study participants, including their outcome (interview completed, agreed to interview, agreement pending, or interview refused). The data management unit has experience in setting and supporting large and complex databases.

In addition, BIPS has a long-standing experience in quality assurance of population-based cohort studies. BIPS has coordinated the evaluation of the German Cardiovascular Prevention Study (DHP) and the Southern German Nasal Cell Carcinoma study and implemented various control measures regarding survey quality. BIPS also held the exclusive responsibility for quality control of the second and third survey of the DHP. Moreover, the institute is in charge of external quality control for the 'Deutsche Erwachsenen Gesundheitssurvey' (DEGS) conducted by the Robert-Koch-Institute (RKI) in 180 examination centers across Germany (N=7,500). H. Bickeböller and P. Falkai from the University Medical School of the Georg-August-University Goettingen (Universitätsmedizin Goettingen, UMG) will contribute their scientific experience and will harmonize their Neuro Chort Goettingen (n=3,000) with the protocol of the National cohort (see Outlook). H. Bickeböller from the Department of Genetic Epidemiology has long standing experience in epidemiological studies with a genetic focus. P Falkai from the section on Psychiatric Genetics has longstanding experience in the establishment and extensive phenotype characterization of large cohorts of patients and controls for psychiatric genetic epidemiological studies.

The **Department of Epidemiologic Methods and Etiological Research** integrates core areas of epidemiologic research in occupational, environmental, and lifestyle epidemiology. The increasing need to investigate the interplay of complex exposures of the modern environment and to overcome difficulties regarding validation and quantification of these exposures requires both sophisticated statistical methods as well as considerable efforts in exposure assessment. BIPS will head the study center Bremen coordinated by W. Ahrens.

Description of the study center Bremen

In Bremen and the surrounding area, 10,000 study participants will be recruited. The study center will be located on the ground floor of the BIPS. Accessibility to the study center is very good since a tram stop is located right in front of the building. The trip from the central station to the study center takes ca. 10 min. The study center is equipped with four rooms for physical examinations, and interviews, and a laboratory for preanalytical preparation (in total ca. 200 m²), plus rooms for storing biosamples. Currently, there is a cold storage room (35 m³) where biosamples are kept at -20 °C, 22 freezers (-20 °C), and 14 ultra deep freezers (-80 °C).

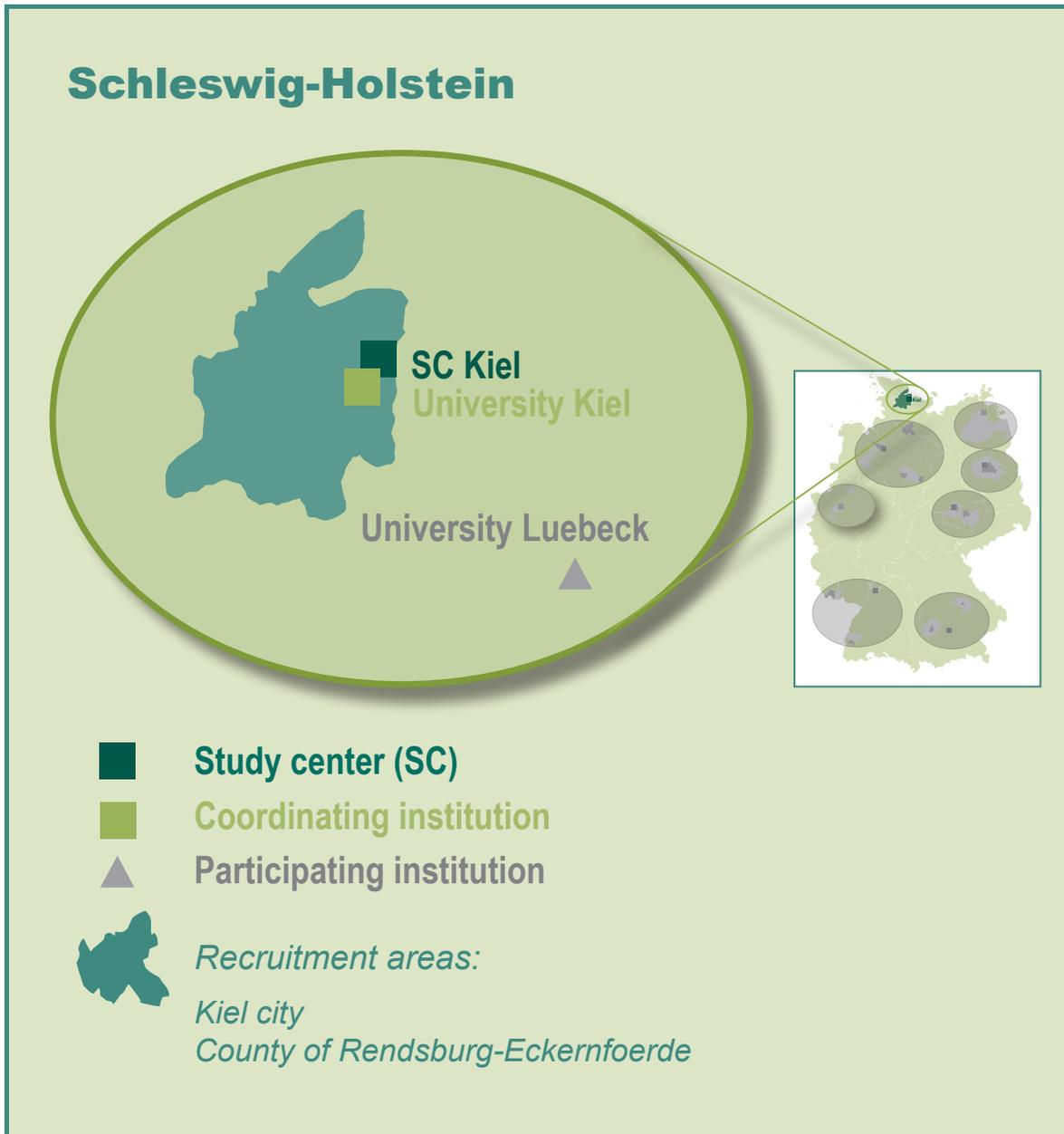
BIPS has a special unit „field work“ that conducts and supervises the recruitment of study participants. The unit has long-standing experience in adopting and evaluating standardized survey instruments and in recruiting and following up study participants. The close collaboration with the cancer registry of Bremen, which is located at BIPS, ensures an efficient tracing of incident cancer cases.

The Center for advanced imaging (CAI) at Bremen University is one of five outstanding regional imaging centers in Germany to be designated a "Center of Excellence" (<http://www.med.uni-magdeburg.de/cai/einfuehrung.html>). It has expertise in developing new neuroimaging and visualization techniques. CAI has extensive brain-imaging laboratories at its disposal. These laboratories are equipped with MRI units for the functional and spectroscopic examination of human subjects (1.5 T, 3x3 T, and 7 T), animals, and tissues (2x4.7 T, 2x8.4 T, and 14.1 T). It also has a whole-head MEG apparatus with integrated EEG (248/96 channels). In addition, the Fraunhofer MEVIS - Institute for Medical Image Computing is located in the neighbouring building of the BIPS. Fraunhofer MEVIS is one of the leading global and internationally networked research and development centers for computer assistance in image-based medicine. It follows a patient-centered and workflow-oriented approach to resolve clinically relevant issues in image-based diagnosis and therapy. Fraunhofer MEVIS focuses on epidemiologically significant diseases of the cardiovascular system, the brain, liver and lung, as well as oncological disorders (www.mevis.fraunhofer.de/en/Institute.html).

C.1.7 Cluster Schleswig-Holstein

The Cluster Schleswig-Holstein comprises one study center in the area of Schleswig-Holstein, located in Kiel.

Figure 7: Geographical overview of recruitment areas and study centers for the cluster Schleswig-Holstein



C.1.7.1 Cluster coordinator

Table 24: Cluster coordinator/co-coordinator

Coordinator	
Name	Prof. Dr. Ute Nöthlings
Institution	Christian-Albrechts-University Kiel
Town	24105 Kiel
Institute/department/division	Section of Epidemiology, Institute for Experimental Medicine
Address, telephone, e-mail	Arnold-Heller Str. 3, Haus 3 24105 Kiel +49 431 597 3677 u.noethlings@iem.uni-kiel.de
Co-coordinator	
Name	Prof. Dr. Alexander Katalinic
Institution	University of Luebeck
Town	23562 Luebeck
Institute/department/division	Institute for Clinical Epidemiology and Institute for Cancer Epidemiology
Address, telephone, e-mail	Ratzeburger Alle 160 23562 Luebeck +49 451 5005440 alexander.katalinic@uk-sh.de

C.1.7.2 Study center

Study center Kiel

Table 25: Overview of participating institutions and investigators/scientists

Institution	University Hospital Schleswig-Holstein, Campi Kiel and Luebeck
Town	24105 Kiel
Principal investigator (Prof. Dr. Ute Nöthlings
Institute/department/division	Biobank popgen, Section for Epidemiology Institute for Experimental Medicine Christian-Albrechts-University Kiel
e-mail	u.noethlings@iem.uni-kiel.de
Co-principal investigator	Prof. Dr. Alexander Katalinic
Institute/department/division	Institute for Clinical Epidemiology and Institute for Cancer Epidemiology / University of Luebeck
e-mail	alexander.katalinic@uk-sh.de

Description of the institutions responsible for establishing and operating the study center

The **Christian-Albrechts-University (CAU)** Kiel has long-standing expertise in life sciences research. At **medical faculty** the **Biobank popgen** represents an institutionalized infrastructure for recruitment, field work, and biobanking located at the University Hospital Schleswig-Holstein Campus Kiel. Popgen is run by the **Section for Epidemiology in the Institute for Experimental Medicine**. Popgen continuously recruits study participants for ongoing population-based studies. Recruitment generally entails administration of standardized questionnaires, physical examinations, collection and storage of biospecimens (whole blood, serum, RNA, DNA, urine, and stool), accompanied by the ascertainment of specific clinical and phenotypic data. Popgen maintains its large biobank of more than 75,000 study participants in close collaboration with the Institute for Clinical Molecular Biology (ICMB), which also runs the genotyping platform associated with popgen. Popgen has a sophisticated data management system essentially following the data protection concept of TMF, developed in close collaboration with the Institute for Medical Informatics and Statistics (IMIS). The research activities of IMIS are focused upon genetic epidemiology, population genetics and bioinformatics. IMIS hosts one of the German Centers of Expertise in Genetic Epidemiology, established through the German national genome research network (NGFN-2), and as such co-founded popgen in 2003. Practical experience in the field work for large prospective cohort studies was also gained by the head of the Section of Epidemiology at previous working places, particularly with the European Prospective Investigation into Cancer and Nutrition-Study (EPIC) – which comprises more than 500,000 study participants across Europe, and the Multiethnic Cohort Study, which consists of more than 215,000 participants of multiethnic origin in Hawaii and Southern California, USA. Both studies use active follow-up procedures, in particular the German part of the EPIC-Study adopted case ascertainment procedures applicable to the German health care infrastructure.

The popgen Biobank pursues an active follow-up of selected general populations and disease specific cohorts, which includes repeated participant contact. In 2010, popgen started the first follow-up of the central control population of 1,300 participants. In addition to completing of comprehensive medical and lifestyle (nutrition) questionnaires – also via the web – participants are invited to receive a physical examination at the study center on the campus of the University Hospital in Kiel, which includes blood pressure measurement, ECG, anthropometric measurement, ultrasonography of liver, gallbladder and adipose tissue, and whole-body MRI imaging. Close cooperations with clinical partners in Kiel (cardiology, 1st medical department, radiology) are established. In this study a range of biosamples, including DNA, RNA, blood, serum, plasma, erythrocytes, urine, and stool, are collected.

In the popgen 2.0 network (P2N) popgen will interlink with further biobanking activities at the University Hospital Schleswig-Holstein Campus Kiel, including the biobanks of the Comprehensive Cancer Center North, the Institutes of Pathology, Neurosurgery, Pharmacology, the University Lung Center North and the Center for family medicine. This network will assist in tracing study participants and collecting disease-specific biomaterial such as tumor tissue.

The **University of Luebeck (UzL)** also has long-standing expertise in epidemiologic research with a history of more than 20 years. To strengthen this expertise in 2009 an academic center of population medicine and health care research (aZBV) was founded as one core area of the university. About 15 institutes/working groups dealing with epidemiology in a wider context are members of the aZBV. It is run by the Department of Population Medicine (SP-PM).

The working groups of the Institute for Social Medicine (IfSM) have conducted and coordinated numerous cross-sectional and longitudinal pain- and rheumatic disorder-related surveys in Germany: the European Vertebral Osteoporosis Study EVOS, two population

surveys which challenged the concept of fibromyalgia as a distinct rheumatological disease entity, a large cohort study of initially 10,000 subjects to develop an algorithm to assess rehabilitation needs of insureds of a regional statutory pension fund for blue collar workers with severe, disabling BP (Luebeck Algorithm). A previous population-based survey in Hannover addressed more than 11,000 adult residents of the city studying the prevalence and health care for rheumatoid arthritis and various rheumatic complaints, including back pain.

Since 1997, the **Institute for Cancer Epidemiology** (IKE) is running the population-based cancer registry of Schleswig-Holstein (in cooperation with the medical association Schleswig-Holstein). Each year, approximately 20,000 incident cases occur in the area. The registry processes some 50,000 notifications (mandatory notification) and 30,000 death certificates per year. Currently more than 600,000 notifications representing more than 250,000 cancer patients are stored in the registry. These data are used for multiple purposes, including public health reporting, small area analyses, or epidemiologic research. Furthermore, several large cohort studies were or are being conducted at the institute, as the worldwide largest controlled cohort study of children conceived after employing artificial reproductive techniques (ICSI follow-up study = intra cytoplasmic sperm injection, n=3,300) or the OVIS study (health care research in cancer survivors, n=5,000). Since 2001, a cohort of women after diagnostic mammography is being built up and analyzed. Up to now more than 300,000 women have been documented, and a record linkage to the cancer registry was established.

In 2011 the University of Luebeck launched the recently established Institute for Clinical Epidemiology (IfKE) (W3 professorship in epidemiology). The new institute will work in close collaboration with the **Institute for Cancer Epidemiology** and support the academic center of population medicine and health care research.

Research activities at Medical Department II of the UzL are focused upon epidemiologic aspects of cardiac geometry and function and on the molecular genetics of MI. The investigators have more than 15 years of experience in conducting epidemiologic research in the context of the MONICA studies. In particular, ECG studies were introduced and repeatedly carried out in the Augsburg KORA study, producing multiple novel observations. With respect to MI genetics the group coordinates the German (NGFN, Atherogenomics), European (EU frame-work 6, Cardiogenics), and global initiatives (CARDIoGRAM) on genome-wide association studies for this pertinent disease.

Description of the study center Kiel

For the National cohort, the Biobank popgen will be responsible for the field work of recruiting 10,000 study participants in Schleswig-Holstein and therefore a study center will be established at the University Hospital Schleswig-Holstein, campus Kiel. Popgen will be responsible for all tasks related to recruitment and follow-up, with assistance by the partner institutes at the CAU and UzL. For example the ICMB will assist in collecting biomaterials whereas IMIS and SP-PM will supervise ethical and data safety issues. IfSM and IfKE will be responsible for exposure assessments related to health care research, IKE will be responsible for record linkages to the cancer registry Schleswig-Holstein during follow-up and the Medical Department II will be responsible for cardiovascular phenotyping following the general protocol of the National Cohort. Comprehensive non responder analyses will be done at IfKE.

In Kiel 10,000 individuals will be recruited including inhabitants of the city of Kiel and adjacent municipalities. The catchment area for the SH cluster will include the city of Kiel (population 170,000) and adjacent municipalities (population 175,000). This results in a total source population of about 345,000. Registration offices for drawing random population-based samples have been successfully accessed in the past.

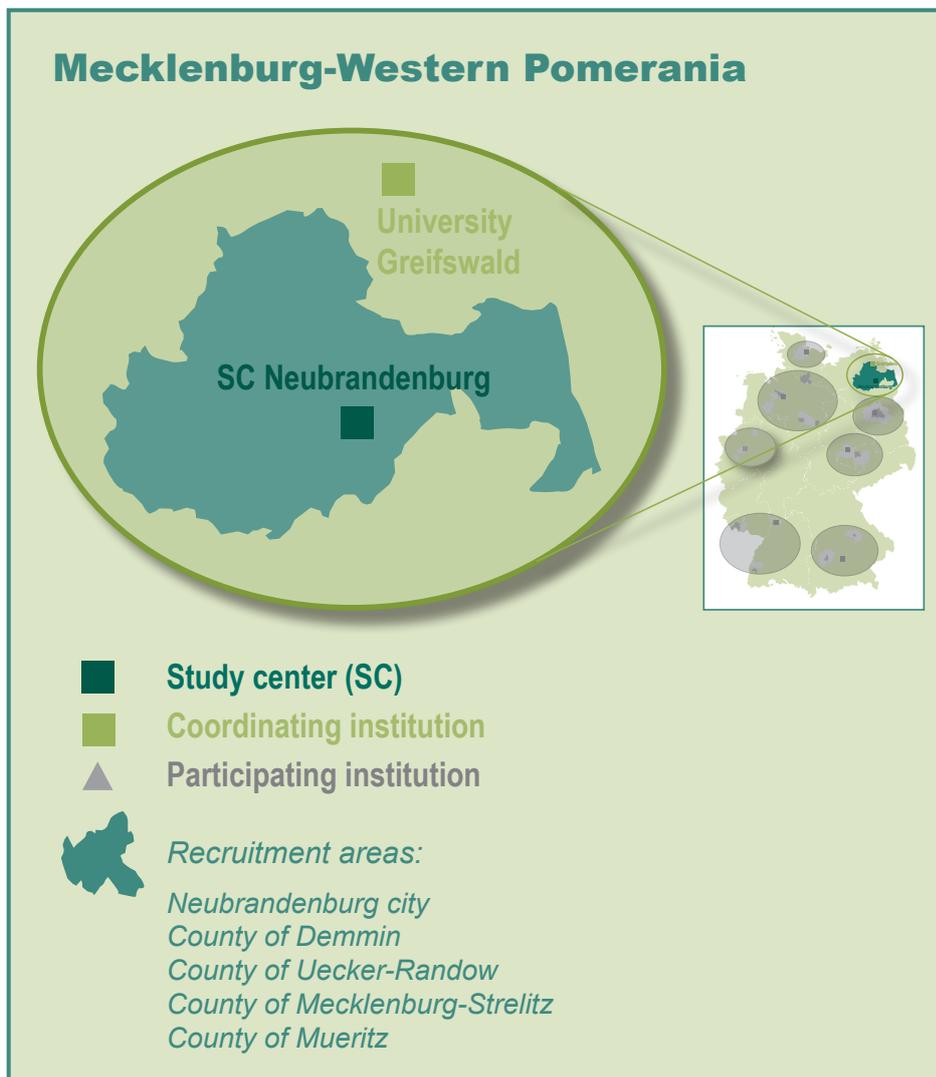
The study center will be located at the University Hospital Schleswig-Holstein Campus Kiel. The campus of the University Hospital Schleswig-Holstein in Kiel is centrally located in Kiel and easily accessible by public transport.

The study center will comprise rooms for investigations, and interviews and terminals for self-administered questionnaires. The laboratory of the Biobank popgen is located on the same campus and will be used for preanalytical preparation of biosamples. The biorepository of the Biobank popgen is located at the university campus Center for Molecular Biosciences (ZMB) in Kiel. Close cooperation with the cancer registry of Schleswig-Holstein, which is a partner in this project, ensures efficient tracing of incident cases. By using the biobanking network popgen 2.0 tumor samples can be collected and individuals traced for incident diseases. Close cooperation with clinicians is ensured by established connections to institutes of the Medical Faculties of CAU, Kiel and UzL, Luebeck.

C.1.8 Cluster Mecklenburg-Western Pomerania

The Cluster Mecklenburg-Western Pomerania comprises one study center in the area of North-East Germany, located in the city of Neubrandenburg.

Figure 8: Geographical overview of recruitment areas and study center the the cluster Mecklenburg-Western Pomerania



C.1.8.1 Cluster coordinator

Table 26: Cluster coordinator/co-coordinator

Coordinator	
Name	Prof. Dr. Henry Völzke
Institution	Ernst Moritz Arndt University
Town	17487 Greifswald
Institute/department/division	Institute for Community Medicine
Address, telephone, e-mail	Walther Rathenau Str. 48; D-17487 Greifswald +49 3834 867541 voelzke@uni-greifswald.de
Co-coordinator	
Name	Prof. Dr. Wolfgang Hoffmann
Institution	Ernst Moritz Arndt University
Town	17487 Greifswald
Institute/department/division	Institute for Community Medicine
Address, telephone, e-mail	Ellernholtzstr. 1-2; D-17489 Greifswald +49 3834 867751 wolfgang.hoffmann@uni-greifswald.de

C.1.8.2 Study center

Study center Neubrandenburg

Table 27: Overview of participating institutions and investigators/scientists

Institution	Ernst Moritz Arndt University
Town	17487 Greifswald
Principal investigator	Prof. Dr. Henry Völzke
Institute/department/division	Institute for Community Medicine
e-mail	voelzke@uni-greifswald.de
Co-principal investigator	Prof. Dr. Wolfgang Hoffmann
Institute/department/division	Institute for Community Medicine
e-mail	wolfgang.hoffmann@uni-greifswald.de
Institution	Ernst Moritz Arndt University
Town	17487 Greifswald
Associated scientist	Prof. Dr. Stephan B. Felix
Institute/department/division	Clinic for Internal Medicine B
e-mail	felix@uni-greifswald.de
Associated scientist	Prof. Dr. Reiner Biffar
Institute/department/division	Policlinic for Prosthetic Dentistry, Gerostomatology and Dental Materials
e-mail	biffar@uni-greifswald.de
Associated scientist	PD Dr. Ralf Puls
Institute/department/division	Institute for Radiology
e-mail	puls@uni-greifswald.de
Associated scientist	Dr. Carsten Oliver Schmidt
Institute/department/division	Institute for Community Medicine
e-mail	carsten.schmidt@uni-greifswald.de
Associated scientist	Dr. Sebastian Baumeister
Institute/department/division	Institute for Community Medicine
e-mail	sebastian.baumeister@uni-greifswald.de
Associated scientist	Sabine Schipf, MSc (Epi)
Institute/department/division	Institute for Community Medicine
e-mail	sabine.schipf@uni-greifswald.de

Description of the institution responsible for establishing and operating the study center

The **University of Greifswald**, founded in 1456, is one of the oldest universities in Europe. Based on recommendations of the German Science Council, Community Medicine has been developed to a main research field of the Medical Faculty of the University of Greifswald. The Community Medicine Research Network was founded in 1995 and comprises scientists and clinicians of different specialities, including epidemiologists, internists, cardiologists, neurologists, psychiatrists, social scientists, dentists, molecular biologists, pediatricians, and many more. Within this network, not only do the different specialities within the medical faculty collaborate with each other, but also with scientists from different departments. Because after German reunification in 1990, scientifically valid data from East Germany to explain the regional differences in life expectancy were lacking and, consequently, there was a need for population-based research in northeast Germany. The Study of Health in Pomerania (SHIP) thus representing the central tool for the Community Medicine Research Network. The formation of the Institute for Community Medicine in 2002 established permanent institutional structures within the Community Medicine Research Network. Since 2007, SHIP has been integrated in the Institute of Community Medicine as a third section of Clinical-Epidemiologic Research. SHIP is actively involved in multiple national research consortia, such as the Competence Network Heart Failure and the Diabetes Collaborative Research of Epidemiologic Studies (DIAB-CORE), an epidemiologic subproject of the Competence Network Diabetes, and the Transregio SFB 19 “Inflammatory cardiomyopathy”. SHIP is also part of international research consortia that perform genome-wide association studies.

Currently, the overall project consists of two studies, the second examination follow-up of the first SHIP cohort (SHIP-2) and baseline examinations of the second SHIP cohort (SHIP-TREND). Both studies have been designed to investigate the long-term progression of subclinical findings, their determinants and prognostic values (SHIP-2); to analyze the secular trend of subclinical and overt diseases and their determinants in a high-risk population (SHIP-TREND vs SHIP-0); and to assess the prevalence of subclinical findings defined by highly innovative no-invasive methods (SHIP-TREND). The aforementioned details on the comprehensive project of SHIP and the successful efforts with respect to quality assurance reflect the expertise of the existing personnel in implementing recruitment and long-term follow-up of large-scale population-based cohort studies. Given the expertise from SHIP, where data were collected at two examination centers (Stralsund and Greifswald), sufficient experience is available to organize examinations outside Greifswald. There will be regular site visits of the coordinating staff at Neubrandenburg (1 h by car).

Description of the study center Neubrandenburg

The sample frame is located south of the SHIP region, including the city of Neubrandenburg and the surrounding communities of Uecker-Randow, Demmin, Mueritz, and Mecklenburg-Strelitz. The population living in the study region comprises 281,338 (20–69 years). The central population registry in Mecklenburg-Western Pomerania will be used to draw a population-representative sample of 20,000 participants. The study center Neubrandenburg is located at the Dietrich Bonhoeffer Hospital in Neubrandenburg, an academic teaching hospital of the University Hospital of Greifswald, which provides rooms for investigations, interviews, terminals for self-administered questionnaires and preanalytical preparation and interim storage of biosamples. The location in the hospital supports an optimal conduction of the study, for utilization of the laboratory or the infrastructure for examinations including MRI. The study center is easily accessible by public transport or by car.

Mecklenburg-Western Pomerania is the least densely populated and one of the economically weakest states in Germany, with an area of more than 23,000 km² and a population of approximately 1.7 million. During the 1980s, mortality of adults living in the East worsened relative to West Germany. In addition to these West-East differences, further regional disparities became evident within East Germany with a South-to-North gradient becoming more and more apparent during the late 1980s. In addition, North-East Germany, which is less industrialized, suffers not only from a decline in birth rates but also from migration of young, healthy and well educated people to metropolitan areas. These determinants add to economic changes. Presently, communities such as Uecker-Randow and Demmin have the highest unemployment and poverty rates in Germany. These changes, although particularly rapid and pronounced, are not limited to this region. Rather, changes observed here are predictive for many other European regions, and therefore North-East Germany may be regarded as a model region for many other parts of Europe.

C.2 Methods for Statistical Power and Sample Size Calculations; Supplementary Tables and Figures

C.2.1 Possible effects of self-selection on cumulative disease incidence

In main section of the document (**Sect.A.6.3**), the expected number of chronic disease were calculated without adjustment for possible biases bias that may arise when individuals recruited into National Cohort do not fully reflect the general population. Other prospective studies have often shown “healthy participant” effects, as reflected by lower incidence of chronic diseases than expected on the basis of general population registry data. Therefore, for comparison purposes, expected cumulative cancer incidence within the National Cohort was also calculated by extrapolating from observed incidence data in the two German EPIC cohorts, in Heidelberg and Potsdam, through re-weighting of sex- and age-specific incidence these cohorts (**Table S6.1**). The two EPIC cohorts include a total of about 53,000 women and men, recruited between 1994 and 1998, and jointly constitute the largest prospective cohort on lifestyle in relation to cancer and other chronic diseases in Germany.

In the EPIC cohort, observed cumulative incidences for cancers of the breast, prostate, pancreas and brain were 10-40% higher than expected, whereas observed numbers of most other cancer types were 10-30% lower. The latter contrasts may be explained by socio-economic and behavioral differences between the EPIC participants and the general German population. The EPIC participants have higher than average levels of education and income, higher than average participation in screening for breast and prostate cancers, higher rates of colonoscopy (with removal of adenomas), and lower than average prevalence of smoking, certain occupational exposures and obesity in EPIC than in the general German population.

Table S6.1. Possible impact of self-selection on cumulative incidences of chronic diseases – approximate ratios of observed to expected numbers of cases in German EPIC cohort.

	Estimated expected number of cases in the cohort in 10 years follow-up as % of expected cases in general population
Any cancer	80%
Breast	110%
Prostate	140%
Colon, Rectum	70%
Lung	80%
Bladder	100%
Kidney	90%
non-Hodgkin L.	70%
Pancreas	110%
Corpus Uteri	90%
Brain+CNS	120%
Ovary	80%

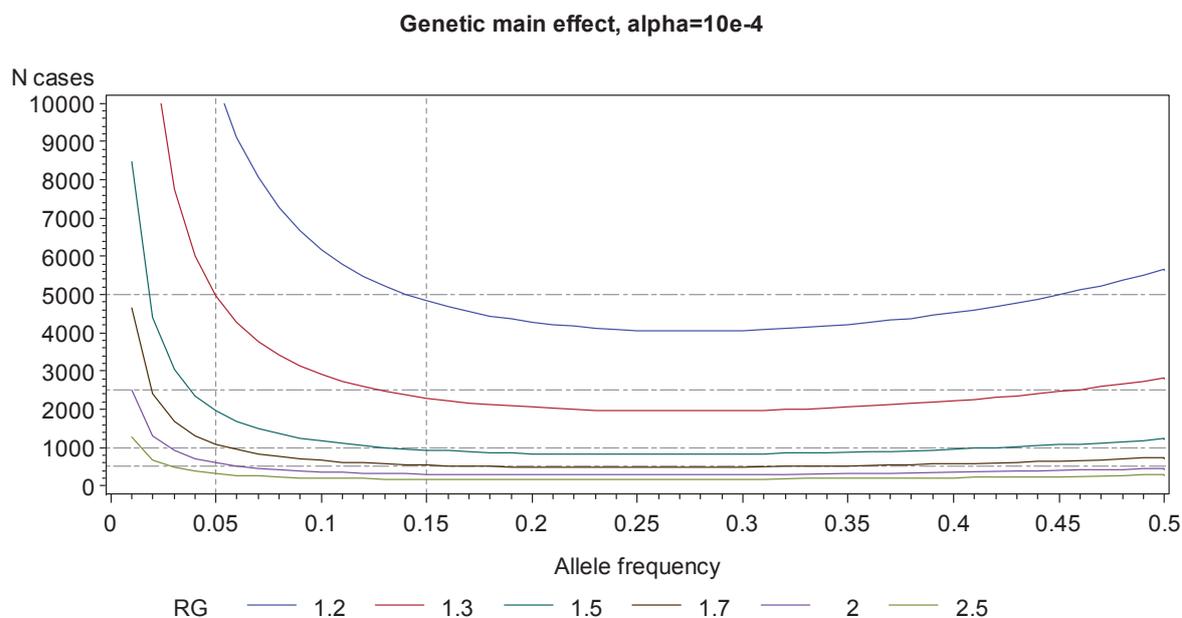
The differences in case numbers for cancer over 10 or 15 years' follow-up based on registry data or on extrapolations from different ongoing German cohort studies document a degree of uncertainty with regard to numbers of cases with disease that may be expected in practice, depending on possible self-selection of cohort participants. Apart from current health status, these self-selection effects are likely to be associated with different background risks for chronic disease in different geographic and socio-economic strata of the German population, which are difficult to fully describe in advance.

C.2.2 Minimally detectable odds ratios in main effects models

In addition to the **Figures 6.2 to 6.4** in the main protocol text (Chapter 5), **Table S6.2** provides minimum detectable odds ratios (MDOR; statistical power 0.80, significance level of either 0.05 or 0.01) for a *binary main effect*, as function of exposure prevalence, number of disease cases, and number of controls per case. Likewise, **Table S6.3** provides minimum detectable odds ratios (MDOR; statistical power 0.80, significance level of either 0.05 or 10^{-4}) for comparison of top to bottom quartiles of a continuous exposure variable, for a number of scenarios.

Figure S6.1 provides estimates for study size (number of cases with 2 control each) required to detect with statistical power 0.80 minimal odds ratios [MDOR]), at a significance level of 10^{-4} . This figure shows that a nested case-control study with 1000 cases of disease and two controls per case will allow detection of a MDOR of 1.5 for a dominant allele that has carrier frequencies of about 0.14 or higher. For smaller MDOR, considerably larger study sizes are generally needed.

Figure S6.1: Study size (number of cases with 2 control each) required to detect minimally detectable odds ratios [MDOR] varying from 1.2 to 2.5 for a genetic main effect (statistical power 0.80, significance level 10^{-4}).**



Assuming dominant model, unmatched case-control study with 2 controls per case

** Assuming Hardy-Weinberg equilibrium between high- and low-risk alleles – that is, for allele frequency of the carrier frequency cf is computed as $cf = af^2 + 2af[1 - af]$.

With regard to continuous exposure or risk factor measurements, as mentioned in the main section of the document (Section A.6.4.2), it will be often appropriate to test the hypothesis that increasing levels of the exposure are associated with increases in risk on a continuous (log-) linear scale. When the exposure variable in question is normally distributed with different mean and equal variance, the relative risk function for dose difference is log-linear. If the disease is comparatively rare, a test for the regression coefficient β , $\beta = 0$ is asymptotically equivalent to a tests for a difference in mean exposure levels between cases and controls. The statistical power of such tests are based on a standardized difference A between cases and controls, defined as,

$$A = \frac{\mu_1 - \mu_0}{\sigma} = \beta\sigma$$

where μ_0 and μ_1 are means for control subjects and disease cases, respectively, σ is the population standard deviation of the exposure variable, and β is the logistic regression coefficient relating increase in log-disease risk to a standardized unit increase in exposure. The measure A (or the product $\beta\sigma$) presents a measures of overall strength of association of the continuous risk factor with disease risk. An advantage of this measure, A , is that is independent of exposure measurement scales. To help interpret the practical significance of a standardized difference of magnitude A , for specification of an alternative hypothesis, it is useful to recall that this entity relates mathematically to an expected relative risk between quantiles (e.g., quartiles or quintiles), as indicated in the **Table S6.4** In addition, **Table S6.5** provides estimates of study size (number of disease cases, with variable control:case ratios) needed for detection of a given standardized difference A with statistical power 0.80, (significance level 0.05).

Table S6.2: Minimum detectable odds ratios (MDOR); statistical power 0.80, significance level of either 0.05 or 0.01 for a binary main effect, as function of exposure prevalence, number of disease cases, and number of controls per case.

Case to control ratio (unmatched)	Exposure prevalence	Critical p-value	Minimum detectable OR for binary main effect															
			150 cases	200 cases	300 cases	400 cases	500 cases	700 cases	1,000 cases	1,500 cases	2,000 cases	2,500 cases	3,000 cases	4,000 cases	5,000 cases	10,000 cases		
1:1	0.05	0.05	3.2	2.8	2.4	2.2	2.0	2.0	1.8	1.7	1.5	1.5	1.4	1.4	1.3	1.2	1.2	
	0.1	0.05	2.5	2.2	2.0	1.8	1.7	1.6	1.5	1.5	1.4	1.3	1.3	1.3	1.2	1.2	1.14	
	0.2	0.05	2.1	1.9	1.7	1.6	1.5	1.4	1.4	1.3	1.2	1.2	1.2	1.19	1.17	1.15	1.10	
	0.4	0.05	1.9	1.8	1.6	1.5	1.4	1.4	1.3	1.2	1.2	1.2	1.17	1.16	1.14	1.12	1.08	
	0.05	0.01	3.9	3.4	2.8	2.5	2.3	2.0	1.8	1.7	1.6	1.5	1.5	1.4	1.3	1.2	1.2	
	0.1	0.01	3.0	2.6	2.2	2.0	1.9	1.7	1.6	1.5	1.4	1.4	1.4	1.3	1.3	1.2	1.17	
	0.2	0.01	2.4	2.2	1.9	1.8	1.7	1.5	1.4	1.3	1.3	1.2	1.2	1.2	1.2	1.18	1.13	
	0.4	0.01	2.2	2.0	1.8	1.6	1.5	1.4	1.4	1.3	1.2	1.2	1.2	1.2	1.17	1.15	1.10	
	1:2	0.05	0.05	2.8	2.5	2.1	2.0	1.8	1.7	1.6	1.5	1.4	1.4	1.3	1.3	1.3	1.2	1.17
		0.1	0.05	2.2	2.0	1.8	1.7	1.6	1.5	1.4	1.3	1.3	1.2	1.2	1.2	1.19	1.17	1.12
		0.2	0.05	1.9	1.8	1.6	1.5	1.4	1.4	1.3	1.2	1.2	1.2	1.18	1.16	1.15	1.13	1.09
		0.4	0.05	1.8	1.6	1.5	1.4	1.4	1.3	1.2	1.2	1.17	1.15	1.14	1.12	1.10	1.07	
0.05		0.01	3.3	2.9	2.4	2.2	2.1	1.9	1.7	1.6	1.5	1.4	1.4	1.3	1.3	1.2	1.2	
0.1		0.01	2.6	2.3	2.0	1.8	1.7	1.6	1.5	1.4	1.3	1.3	1.3	1.3	1.2	1.2	1.15	
0.2		0.01	2.2	2.0	1.7	1.6	1.6	1.5	1.4	1.3	1.3	1.2	1.2	1.2	1.17	1.16	1.11	
0.4		0.01	2.0	1.8	1.6	1.5	1.5	1.4	1.3	1.2	1.2	1.2	1.18	1.17	1.14	1.13	1.09	
1:4		0.05	0.05	2.5	2.3	2.0	1.8	1.7	1.6	1.5	1.4	1.3	1.3	1.3	1.2	1.2	1.15	
		0.1	0.05	2.1	1.9	1.7	1.6	1.5	1.4	1.4	1.3	1.2	1.2	1.2	1.17	1.15	1.11	
		0.2	0.05	1.8	1.7	1.5	1.4	1.4	1.3	1.3	1.2	1.19	1.16	1.15	1.13	1.11	1.08	
		0.4	0.05	1.7	1.6	1.4	1.4	1.3	1.3	1.2	1.18	1.15	1.14	1.12	1.11	1.09	1.07	
	0.05	0.01	3.0	2.6	2.3	2.1	1.9	1.8	1.6	1.5	1.4	1.4	1.3	1.3	1.3	1.18		
	0.1	0.01	2.4	2.1	1.9	1.8	1.7	1.5	1.4	1.3	1.3	1.2	1.2	1.2	1.19	1.13		
	0.2	0.01	2.0	1.8	1.7	1.6	1.5	1.4	1.3	1.3	1.2	1.2	1.2	1.18	1.16	1.10		
	0.4	0.01	1.9	1.7	1.6	1.5	1.4	1.3	1.3	1.2	1.2	1.19	1.17	1.15	1.12	1.08		

* highlighted are the combinations where MDOR is 1.5 or smaller

Table S6.3: Minimum detectable odds ratio (MDOR); statistical power 0.80, significance level of either 0.05 or 10⁻⁴ for comparison of top to bottom quartiles of a continuous exposure variable by number of controls per cases.

Case to control ratio (unmatched)	Critical p-value	Minimum detectable OR for comparison of top to bottom quartiles													
		150 cases	200 cases	300 cases	400 cases	500 cases	700 cases	1,000 cases	1,500 cases	2,000 cases	2,500 cases	3,000 cases	4,000 cases	5,000 cases	10,000 cases
1:1	0.05	2.4	2.1	1.9	1.7	1.6	1.5	1.4	1.3	1.3	1.2	1.2	1.18	1.16	1.11
	10 ⁻⁴	7.2	5.3	3.8	3.1	2.8	2.3	2.0	1.8	1.6	1.6	1.5	1.4	1.4	1.2
1:2	0.05	2.1	1.9	1.7	1.6	1.5	1.4	1.3	1.3	1.2	1.2	1.18	1.16	1.14	1.10
	10 ⁻⁴	5.4	4.2	3.1	2.7	2.4	2.1	1.8	1.6	1.5	1.5	1.4	1.4	1.3	1.2
1:4	0.05	2.0	1.8	1.6	1.5	1.5	1.4	1.3	1.2	1.2	1.18	1.17	1.14	1.13	1.09
	10 ⁻⁴	4.6	3.6	2.8	2.4	2.2	2.0	1.7	1.6	1.5	1.4	1.4	1.3	1.3	1.2

* highlighted are the combinations where MDOR is 1.5 or smaller

Table S6.4: Relationship between standardized difference A and expected Odds Ratio between extreme tertiles or quartiles of the exposure distribution.

Standardized difference A	OR over quartiles	OR over tertiles
0,05	1,14	1,12
0,08	1,23	1,19
0,10	1,29	1,25
0,15	1,46	1,39
0,20	1,66	1,55
0,25	1,89	1,73
0,30	2,14	1,93
0,35	2,43	2,16
0,40	2,76	2,41

Table S6.5: Sample size needed for detection of a given standardized difference A (statistical power 0.80, significance level 0.05) by number of controls per cases.

Case to control ratio (unmatched)	Critical p-value	correlation (ρ_{CE}^2)	Variance inflation factor (VIF)	Sample size									
				0.05	0.08	0.10	0.15	0.20	0.25	0.30	0.35	0.40	
1:1	0.05	No adjustment		6280	2455	1572	700	395	253	177	130	100	
	0.05	0.3	1.10	6908	2701	1729	770	435	278	195	143	110	
	0.05	0.5	1.33	9188	3592	2300	1024	578	370	259	190	146	
	0.05	0.7	1.96	18008	7040	4508	2007	1133	725	508	373	287	
	10 ⁻⁴	No adjustment		17921	7004	4485	1997	1126	723	504	372	286	
	10 ⁻⁴	0.3	1.10	19713	7704	4934	2197	1239	795	554	409	315	
	10 ⁻⁴	0.5	1.33	26218	10247	6562	2922	1647	1058	737	544	418	
	10 ⁻⁴	0.7	1.96	51388	20084	12861	5726	3229	2073	1445	1067	820	
	1:2	0.05	No adjustment		4711	1841	1179	525	296	190	132	98	75
		0.05	0.3	1.10	5182	2025	1297	578	326	209	145	108	83
0.05		0.5	1.33	6892	2693	1725	768	433	278	193	143	110	
0.05		0.7	1.96	13509	5279	3381	1505	849	545	379	281	215	
10 ⁻⁴		No adjustment		13440	5253	3363	1497	844	542	377	278	241	
10 ⁻⁴		0.3	1.10	14784	5778	3699	1647	928	596	415	306	265	
10 ⁻⁴		0.5	1.33	19663	7685	4920	2190	1235	793	552	407	353	
10 ⁻⁴		0.7	1.96	38539	15063	9643	4293	2420	1554	1081	797	691	
1:4		0.05	No adjustment		3925	1534	982	437	246	158	110	81	62
		0.05	0.3	1.10	4318	1687	1080	481	271	174	121	89	68
	0.05	0.5	1.33	5742	2244	1437	639	360	231	161	119	91	
	0.05	0.7	1.96	11255	4399	2816	1253	705	453	315	232	178	
	10 ⁻⁴	No adjustment		11199	4376	2802	1247	702	450	314	231	178	
	10 ⁻⁴	0.3	1.10	12319	4814	3082	1372	772	495	345	254	196	
	10 ⁻⁴	0.5	1.33	16384	6402	4099	1824	1027	658	459	338	260	
	10 ⁻⁴	0.7	1.96	32113	12548	8035	3576	2013	1290	900	662	510	

Table S6.6a: Minimum detectable odds ratios (MDOR); statistical power 0.80 at significance level of either 0.05 or 10⁻⁴ for a genetic main effect, by allele frequency; nested case-control study with 2 controls per case.

Allele frequency	Critical p Value	Minimum detectable OR for genetic effect (2 controls)														
		150 cases	200 cases	300 cases	400 cases	500 cases	700 cases	1000 cases	1500 cases	2000 cases	2500 cases	3000 cases	4000 cases	5000 cases	10 000 cases	
0.05	0.05	2.2	2.0	1.8	1.7	1.6	1.5	1.4	1.3	1.3	1.2	1.2	1.2	1.2	1.18	1.12
0.05	10 ⁻⁴	3.5	3.0	2.6	2.3	2.1	1.9	1.7	1.6	1.5	1.4	1.4	1.3	1.3	1.3	1.2
0.1	0.05	1.9	1.8	1.6	1.5	1.5	1.4	1.3	1.3	1.2	1.19	1.17	1.15	1.13	1.10	
0.1	10 ⁻⁴	2.9	2.5	2.2	2.0	1.8	1.7	1.6	1.4	1.4	1.3	1.3	1.3	1.2	1.16	
0.25	0.05	1.8	1.6	1.5	1.4	1.4	1.3	1.3	1.2	1.17	1.15	1.14	1.12	1.11	1.08	
0.25	10 ⁻⁴	2.6	2.3	2.0	1.8	1.7	1.6	1.5	1.4	1.3	1.3	1.2	1.2	1.18	1.13	
0.5	0.05	2.1	1.9	1.6	1.5	1.5	1.4	1.3	1.2	1.2	1.18	1.16	1.14	1.13	1.09	
0.5	10 ⁻⁴	3.9	3.2	2.5	2.1	2.0	1.7	1.6	1.5	1.4	1.3	1.3	1.3	1.2	1.15	

* highlighted are the combinations where MDOR is 1.5 or smaller

Table S6.6b: Minimum detectable odds ratios (MDOR); statistical power 0.80, at significance level of either 0.05 or 10⁻⁴ for a genetic main effect, by allele frequency; nested case-control study with 4 controls per case

Allele frequency	Critical p Value	Minimum detectable OR for genetic effect (4 controls)													
		150 cases	200 cases	300 cases	400 cases	500 cases	700 cases	1000 cases	1500 cases	2000 cases	2500 cases	3000 cases	4000 cases	5000 cases	10 000 cases
0.05	0.05	2.0	1.9	1.7	1.6	1.5	1.5	1.4	1.3	1.3	1.2	1.2	1.2	1.16	1.11
0.05	10 ⁻⁴	3.2	2.8	2.4	2.1	2.0	1.8	1.7	1.5	1.5	1.4	1.4	1.3	1.3	1.2
0.1	0.05	1.8	1.7	1.5	1.5	1.4	1.3	1.3	1.2	1.19	1.17	1.16	1.14	1.12	1.09
0.1	10 ⁻⁴	2.6	2.3	2.0	1.9	1.7	1.6	1.5	1.4	1.3	1.3	1.3	1.2	1.2	1.15
0.25	0.05	1.7	1.6	1.4	1.4	1.3	1.3	1.2	1.19	1.16	1.14	1.13	1.11	1.10	1.07
0.25	10 ⁻⁴	2.4	2.1	1.9	1.7	1.6	1.5	1.4	1.3	1.3	1.2	1.2	1.19	1.17	1.12
0.5	0.05	1.9	1.8	1.6	1.5	1.4	1.3	1.3	1.2	1.19	1.16	1.15	1.13	1.11	1.08
0.5	10 ⁻⁴	3.5	2.8	2.3	2.0	1.8	1.7	1.5	1.4	1.3	1.3	1.3	1.2	1.2	1.14

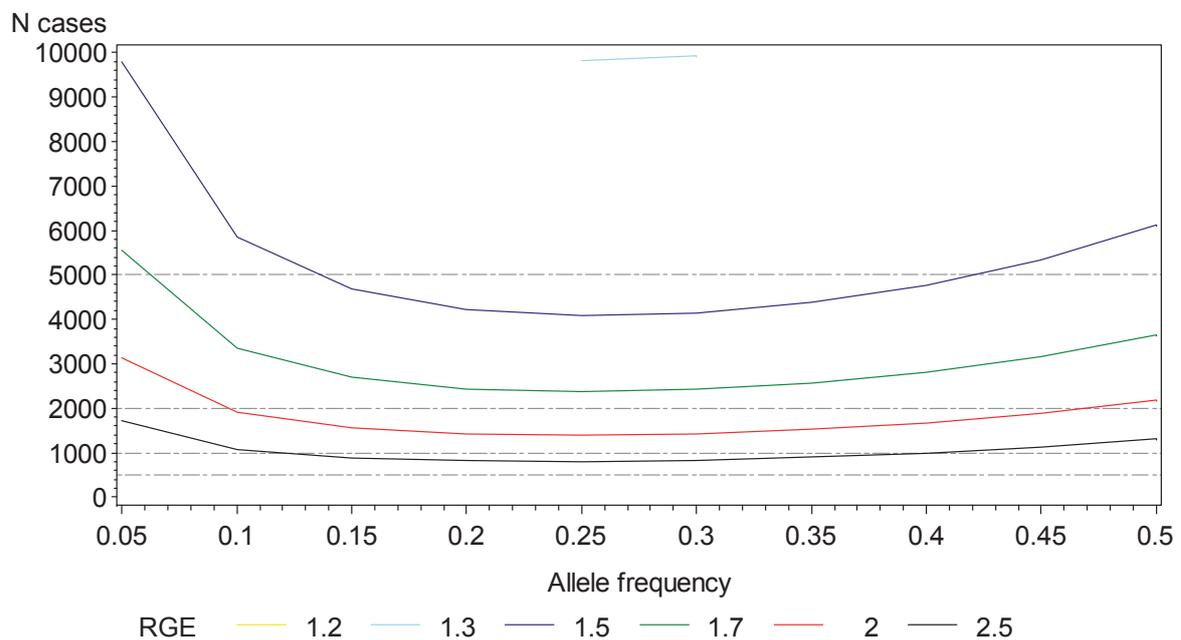
* highlighted are the combinations where MDOR is 1.5 or smaller

C.2.3 Interaction effects between genetic and non-genetic (“environmental”) risk factors.

In **Sect. A.6.4.3** of the study protocol (main text), **Figure 6.5** shows the minimal detectable odds ratio for gene-environment interaction (power 0.80, at significance level of 0.05), in a nested case-control study with N cases of disease, and control:case ratios of 4:1, at fixed (dominant) main effect odds ratios of 1.2 for the genetic variant, and a main effects odds ratio of 2.0 for the environmental factor.

Figure S6.2 below provides similar estimates, but at a significance level of 10^{-4} , which may be more appropriate when larger numbers of interactions are being tested simultaneously. This figure confirms that, for statistical interaction effects of moderate magnitude (MDOR for interaction around 1.5, a sample size of at least 4,000-5,000 cases of disease will be required for tests at this significance level. In the National Cohort, cumulative incidences of this magnitude will be reached only for the most common disease outcomes, including diabetes and myocardial infarction.

Figure S6.2: Study size (number of incident cases of disease) required to detect minimal odds ratios [MDOR] of various magnitudes, for gene-environment interaction with power 0.80, at significance level of 10^{-4} ; nested case-control study with **4 controls per case***, assuming main effects odds ratios of 1.2 for the genetic and 2.0 for the non-genetic risk factor, plus a prevalence of the non-genetic risk factor of 0.20.



*assuming main effects odds ratios of 1.2 for genetic and 2.0 for non-genetic risk factor
alpha=0.0001, prevalence of non-genetic risk factor 20%

Table S6.7a: Minimum detectable odds ratios [MDOR] for gene-environment interaction (statistical power of 0.80; dominant effects model for the genetic variant), by allele frequency, by non-genetic risk factor prevalence, and by significance level; nested case-control study with 2 controls per case.

Genotype prevalence	Non-genetic factor prevalence	Critical P-value	Minimum detectable OR for gene-environment interaction (2 controls)													
			150 cases	200 cases	300 cases	400 cases	500 cases	700 cases	1000 cases	1500 cases	2000 cases	2500 cases	3000 cases	4000 cases	5000 cases	10,000 cases
0.05	0.05	0.05	13.5	9.8	6.7	5.4	4.6	3.8	3.1	2.6	2.3	2.1	2.0	1.8	1.7	1.5
0.05	0.05	10 ⁻⁴	51.2	22.3	14.8	11.3	8.0	6.0	4.5	3.8	3.3	3.0	2.7	2.4	1.9	
0.1	0.05	0.05	8.8	6.7	4.8	4.0	3.5	2.9	2.5	2.1	2.0	1.8	1.7	1.6	1.5	1.4
0.1	0.05	10 ⁻⁴	47.2	24.8	13.2	9.4	7.5	5.6	4.3	3.4	2.9	2.6	2.4	2.2	2.0	1.7
0.25	0.05	0.05	7.4	5.6	4.1	3.4	3.0	2.6	2.2	1.9	1.8	1.7	1.6	1.5	1.4	1.3
0.25	0.05	10 ⁻⁴	43.3	22.0	11.4	8.0	6.4	4.8	3.7	2.9	2.5	2.3	2.2	2.0	1.8	1.5
0.5	0.05	0.05	16.6	10.5	6.4	4.9	4.1	3.3	2.7	2.2	1.9	1.9	1.8	1.6	1.6	1.4
0.5	0.05	10 ⁻⁴	36.1	18.9	9.5	6.7	5.5	4.1	3.2	2.6	2.2	2.1	2.0	1.8	1.7	1.7
0.05	0.1	0.05	8.0	6.2	4.6	3.8	3.3	2.8	2.4	2.1	1.9	1.8	1.7	1.6	1.5	1.4
0.05	0.1	10 ⁻⁴	34.9	20.2	11.6	8.5	6.9	5.2	4.1	3.3	2.8	2.6	2.4	2.1	2.0	1.6
0.1	0.1	0.05	5.5	4.4	3.4	2.9	2.6	2.3	2.0	1.8	1.7	1.6	1.5	1.4	1.4	1.3
0.1	0.1	10 ⁻⁴	18.2	11.9	7.5	5.8	4.9	3.9	3.1	2.6	2.3	2.1	2.0	1.8	1.7	1.5
0.25	0.1	0.05	4.7	3.8	3.0	2.6	2.3	2.0	1.8	1.6	1.5	1.5	1.4	1.4	1.3	1.2
0.25	0.1	10 ⁻⁴	16.6	10.4	6.4	4.9	4.1	3.3	2.7	2.3	2.0	1.9	1.8	1.7	1.6	1.4
0.5	0.1	0.05	8.4	5.9	4.1	3.3	2.9	2.4	2.1	1.8	1.7	1.6	1.5	1.4	1.4	1.3
0.5	0.1	10 ⁻⁴	50.6	25.3	15.3	11.1	9.2	6.9	5.6	4.8	4.1	3.6	3.2	2.9	2.7	1.5
0.05	0.2	0.05	5.8	4.6	3.5	3.0	2.7	2.4	2.1	1.8	1.7	1.6	1.5	1.5	1.4	1.3
0.05	0.2	10 ⁻⁴	20.2	12.9	8.0	6.1	5.1	4.0	3.2	2.7	2.4	2.2	2.0	1.9	1.7	1.5
0.1	0.2	0.05	4.1	3.4	2.8	2.4	2.2	2.0	1.8	1.6	1.5	1.4	1.4	1.3	1.3	1.2
0.1	0.2	10 ⁻⁴	11.8	8.1	5.5	4.3	3.7	3.1	2.6	2.2	2.0	1.8	1.7	1.6	1.5	1.4
0.25	0.2	0.05	3.6	3.0	2.5	2.2	2.0	1.8	1.6	1.5	1.4	1.4	1.3	1.3	1.2	1.17
0.25	0.2	10 ⁻⁴	11.6	7.4	4.7	3.8	3.2	2.7	2.3	2.0	1.8	1.7	1.6	1.5	1.4	1.3
0.5	0.2	0.05	6.4	4.5	3.2	2.7	2.4	2.1	1.8	1.6	1.5	1.5	1.4	1.3	1.3	1.2
0.5	0.2	10 ⁻⁴	14.0	8.5	5.1	4.1	3.4	2.9	2.4	2.1	1.9	1.9	1.8	1.7	1.6	1.4

* assuming main effects odds ratios of 1.2 for genetic and 2.0 for non-genetic risk factor

** highlighted are the combinations where MDOR is 1.5 or smaller

Table S6.7.b: Minimum detectable odds ratios [MDOR] for gene-environment interaction (statistical power of 0.80; dominant effects model for the genetic variant), by allele frequency, by non-genetic risk factor prevalence, and by significance level; nested case-control study with 4 controls per case.

Genotype prevalence	Non-genetic factor prevalence	Critical P-value	Minimum detectable OR for gene-environment interaction (4 controls)													
			150 cases	200 cases	300 cases	400 cases	500 cases	700 cases	1000 cases	1500 cases	2000 cases	2500 cases	3000 cases	4000 cases	5000 cases	10 000 cases
0.05	0.05	0.05	8.7	6.8	5.0	4.2	3.7	3.1	2.6	2.3	2.1	1.9	1.8	1.7	1.6	1.4
0.05	0.05	10 ⁻⁴	35.5	21.4	12.6	9.3	7.6	5.8	4.5	3.6	3.1	2.8	2.6	2.3	2.1	1.7
0.1	0.05	0.05	6.1	4.9	3.8	3.2	2.9	2.5	2.2	1.9	1.8	1.7	1.6	1.5	1.5	1.3
0.1	0.05	10 ⁻⁴	19.7	13.1	8.4	6.5	5.4	4.3	3.5	2.8	2.5	2.3	2.2	2.0	1.8	1.6
0.25	0.05	0.05	5.3	4.3	3.3	2.8	2.6	2.2	2.0	1.8	1.6	1.6	1.5	1.4	1.4	1.3
0.25	0.05	10 ⁻⁴	17.8	11.6	7.3	5.6	4.7	3.7	3.0	2.5	2.2	2.1	1.9	1.8	1.7	1.5
0.5	0.05	0.05	9.3	6.7	4.6	3.8	3.3	2.7	2.3	2.0	1.8	1.7	1.6	1.5	1.5	1.3
0.5	0.05	10 ⁻⁴	38.6	16.0	10.2	7.7	6.1	5.5	4.1	3.2	2.7	2.4	2.3	2.0	1.9	1.6
0.05	0.1	0.05	5.8	4.7	3.7	3.1	2.8	2.5	2.1	1.9	1.8	1.7	1.6	1.5	1.4	1.3
0.05	0.1	10 ⁻⁴	17.2	11.9	7.8	6.1	5.1	4.1	3.4	2.8	2.4	2.3	2.1	1.9	1.8	1.5
0.1	0.1	0.05	4.3	3.6	2.9	2.5	2.3	2.1	1.8	1.7	1.6	1.5	1.4	1.4	1.3	1.2
0.1	0.1	10 ⁻⁴	10.9	7.9	5.5	4.5	3.9	3.2	2.7	2.3	2.1	1.9	1.8	1.7	1.6	1.4
0.25	0.1	0.05	3.7	3.1	2.5	2.3	2.1	1.9	1.7	1.5	1.5	1.4	1.4	1.3	1.3	1.19
0.25	0.1	10 ⁻⁴	10.1	7.1	4.8	3.9	3.4	2.8	2.4	2.0	1.9	1.7	1.7	1.6	1.5	1.3
0.5	0.1	0.05	6.0	4.5	3.3	2.8	2.5	2.2	1.9	1.7	1.6	1.5	1.5	1.4	1.3	1.2
0.5	0.1	10 ⁻⁴	24.4	14.4	9.6	6.5	5.1	3.8	3.0	2.5	2.2	2.0	1.9	1.7	1.6	1.4
0.05	0.2	0.05	4.5	3.7	3.0	2.6	2.4	2.1	1.9	1.7	1.6	1.5	1.5	1.4	1.3	1.2
0.05	0.2	10 ⁻⁴	12.0	8.6	5.9	4.7	4.1	3.3	2.8	2.3	2.1	2.0	1.9	1.7	1.6	1.4
0.1	0.2	0.05	3.4	2.9	2.5	2.2	2.0	1.8	1.7	1.5	1.4	1.4	1.3	1.3	1.3	1.18
0.1	0.2	10 ⁻⁴	8.1	6.0	4.3	3.6	3.1	2.7	2.3	2.0	1.8	1.7	1.6	1.5	1.5	1.3
0.25	0.2	0.05	3.1	2.6	2.2	2.0	1.8	1.7	1.5	1.4	1.4	1.3	1.3	1.2	1.2	1.15
0.25	0.2	10 ⁻⁴	8.2	5.7	3.9	3.2	2.8	2.4	2.1	1.8	1.7	1.6	1.5	1.4	1.4	1.3
0.5	0.2	0.05	5.2	3.8	2.8	2.4	2.2	1.9	1.7	1.5	1.5	1.4	1.4	1.3	1.3	1.18
0.5	0.2	10 ⁻⁴	16.0	10.2	6.5	5.6	4.3	3.2	2.6	2.1	1.9	1.8	1.7	1.6	1.5	1.3

* assuming main effects odds ratios of 1.2 for genetic and 2.0 for non-genetic risk factor

** highlighted are the combinations where MDOR is 1.5 or smaller

C.2.4 Sample size requirements for validation/calibration sub-studies

When a sub-study with replicate measurements is used for regression calibration, e.g. of blood-based exposure measurements, the accuracy of the adjustment will depend jointly on the strength of correlation ($\rho_{m1,m2}$) between the replicate measurements, the size of the sub-study (N), the strength of association (standardized difference $A=\beta\sigma$) between exposure measurements and disease risk and the number of cases with disease used for relative risk modelling^{824, 825}. This can be expressed by the following equation, using the ratio of observed number of cases (D) over the effective number of cases (\tilde{D}) left after calibration correction as a relative efficiency criterion⁸²⁴:

$$\frac{D}{\tilde{D}} = 1 + A^2 \frac{1 - \rho_{m1,m2}^2}{\rho_{m1,m2}^2} \frac{D}{N}$$

For example, when the correlation $\rho_{m1,m2}$ between replicate measurements is at least 0.40, and for a standardized difference (A) in exposure level between cases and controls corresponding to an expected relative risk of 4.0 between the extreme quintiles of the exposure measurement (i.e., $A=0.50$), a calibration sub-study should include at least 5.25 times the number of cases with disease in a nested case-control study to limit the loss in effective number of cases to 20% (i.e., $D/\tilde{D} = 1/0.80 = 1.25$). At higher correlations between replicate measurements, or in situations where relative risk estimates (association measure A) are lower, this same criterion can be met by a smaller reproducibility/calibration sub-study. For example, when the standardized difference in exposure between cases and controls corresponds to an expected relative risk [odds ratio] of 2.0 between the extreme quintiles (i.e., $A=0.25$), the size of the sub-study should be only about 1.3 times the number of cases, which is four times fewer than in the previous example. Table S6.8 shows the sample size numbers (N) required for different combinations of $\rho_{m1,m2}$, A, D.

On the basis of the estimated sample size requirements for calibration sub-studies (**Table S6.8**), and anticipating that most relevant exposure measurements (i.e. those that in practice can be expected to allow detection of a true relative risk of at least 1.5 or higher; see also **Table S6.9**) will generally have a correlation of at least 0.30 between short-term replicate measurements (e.g. over a one-year time interval), we propose a study size for validation/calibration sub-studies of 10,000 subjects in total. Of these, 6,000 will be in the first (baseline) recruitment round, and 4,000 in the second round of data collection, five years later.

Table S6.8: Sample size (*N*) required for different combinations of association measure (*A*), number of cases (*D*) and correlation between two repeated measurements ($\rho_{m1,m2}$), when the loss in effective number of cases is limited to 20%

Association measure (<i>A</i>)	Number of cases (<i>D</i>)	Correlation between repeat measurements ($\rho_{m1,m2}$) (correlation between measured and true exposure; $\rho_{m,x}$)								
		0.10 (0.32)	0.20 (0.45)	0.30 (0.55)	0.40 (0.63)	0.50 (0.71)	0.60 (0.77)	0.70 (0.84)	0.80 (0.89)	0.90 (0.95)
0.15	300	2673	648	273	142	81	48	28	15	6
	400	3564	864	364	189	108	64	37	20	8
	500	4455	1080	455	236	135	80	47	25	11
	700	6237	1512	637	331	189	112	66	35	15
	1000	8910	2160	910	473	270	160	94	51	21
0.25	300	7425	1800	758	394	225	133	78	42	18
	400	9900	2400	1011	525	300	178	104	56	23
	500	12375	3000	1264	656	375	222	130	70	29
	700	17325	4200	1769	919	525	311	182	98	41
	1000	24750	6000	2528	1313	750	444	260	141	59
0.50	300	29700	7200	3033	1575	900	533	312	169	70
	400	39600	9600	4044	2100	1200	711	416	225	94
	500	49500	12000	5056	2625	1500	889	520	281	117
	700	69300	16800	7078	3675	2100	1244	729	394	164
	1000	99000	24000	10111	5250	3000	1778	1041	562	235

* highlighted are the situations where less than 6000 subjects are required for calibration sub-study (index period 1), or less than 4000 (index period 2).

Table S.6.9: Statistical power for detecting true minimal standardized differences A, corresponding to expected true odds ratios ranging from 1.5 to 2.0, when a continuous exposure variable is measured with random error; nested case-control study with 2 controls per case, and with either single exposure measurements, or an average of two (replicate) measurements (for 75% of subjects).

Number of cases	Correlation between repeat measurements $\rho(m_1, m_2)$		Correlation between exposure and true exposure $\rho(m, t)$		Standardized difference corresponding to odds ratio OR [Q4-Q1]											
	0.10	0.20	0.30	0.40	1.5		1.6		1.7		1.8		1.9		2.0	
	initial power	power with repeats	initial power	power with repeats	initial power	power with repeats	initial power	power with repeats	initial power	power with repeats	initial power	power with repeats	initial power	power with repeats	initial power	power with repeats
300	0.09	0.12	0.09	0.12	0.11	0.15	0.13	0.18	0.15	0.21	0.17	0.24	0.19	0.28		
	0.14	0.19	0.14	0.19	0.18	0.24	0.21	0.30	0.25	0.35	0.29	0.41	0.33	0.46		
	0.19	0.25	0.19	0.25	0.24	0.32	0.30	0.39	0.35	0.46	0.41	0.53	0.46	0.60		
	0.24	0.30	0.24	0.30	0.31	0.39	0.38	0.47	0.44	0.55	0.51	0.63	0.57	0.69		
	0.29	0.35	0.29	0.35	0.37	0.44	0.45	0.54	0.53	0.62	0.60	0.70	0.67	0.77		
	0.34	0.39	0.34	0.39	0.43	0.49	0.52	0.59	0.61	0.68	0.68	0.76	0.75	0.82		
	0.38	0.42	0.38	0.42	0.49	0.54	0.59	0.64	0.67	0.73	0.75	0.80	0.81	0.86		
400	0.11	0.15	0.11	0.15	0.13	0.19	0.16	0.23	0.18	0.27	0.21	0.31	0.24	0.35		
	0.18	0.24	0.18	0.24	0.22	0.31	0.27	0.38	0.32	0.44	0.37	0.51	0.42	0.57		
	0.24	0.32	0.24	0.32	0.31	0.41	0.38	0.50	0.44	0.58	0.51	0.65	0.57	0.72		
	0.31	0.38	0.31	0.38	0.39	0.49	0.48	0.59	0.56	0.68	0.63	0.75	0.70	0.81		
	0.37	0.44	0.37	0.44	0.47	0.56	0.56	0.66	0.65	0.75	0.73	0.82	0.79	0.87		
	0.43	0.49	0.43	0.49	0.54	0.61	0.64	0.72	0.73	0.80	0.80	0.87	0.86	0.91		
	0.48	0.53	0.48	0.53	0.60	0.66	0.71	0.76	0.79	0.84	0.86	0.90	0.91	0.94		
500	0.13	0.18	0.13	0.18	0.15	0.22	0.18	0.27	0.22	0.32	0.25	0.37	0.28	0.42		
	0.21	0.29	0.21	0.29	0.26	0.37	0.32	0.45	0.38	0.53	0.44	0.60	0.50	0.67		
	0.29	0.38	0.29	0.38	0.37	0.49	0.45	0.59	0.53	0.67	0.60	0.75	0.67	0.81		

Number of cases	Correlation between repeat measurements $\rho(m_1, m_2)$	Correlation between measured and true exposure $\rho(m, t)$	Standardized difference corresponding to odds ratio OR [Q4-Q1]											
			1.5		1.6		1.7		1.8		1.9		2.0	
			initial power	power with repeats	initial power	power with repeats	initial power	power with repeats	initial power	power with repeats	initial power	power with repeats	initial power	power with repeats
	0.40	0.63	0.37	0.46	0.47	0.58	0.56	0.68	0.65	0.77	0.73	0.84	0.79	0.89
	0.50	0.71	0.44	0.53	0.56	0.65	0.66	0.76	0.75	0.84	0.82	0.89	0.87	0.93
	0.60	0.77	0.51	0.58	0.63	0.71	0.74	0.81	0.82	0.88	0.88	0.93	0.92	0.96
	0.70	0.84	0.57	0.63	0.70	0.75	0.80	0.85	0.87	0.91	0.92	0.95	0.96	0.97
700	0.10	0.32	0.16	0.23	0.20	0.29	0.24	0.36	0.28	0.42	0.33	0.49	0.37	0.55
	0.20	0.45	0.27	0.38	0.35	0.49	0.43	0.59	0.50	0.67	0.57	0.75	0.64	0.81
	0.30	0.55	0.38	0.50	0.49	0.63	0.59	0.73	0.67	0.81	0.75	0.88	0.81	0.92
	0.40	0.63	0.48	0.60	0.60	0.72	0.71	0.82	0.79	0.89	0.86	0.94	0.91	0.96
	0.50	0.71	0.57	0.67	0.70	0.79	0.80	0.88	0.87	0.94	0.92	0.97	0.96	0.98
	0.60	0.77	0.65	0.73	0.78	0.84	0.87	0.92	0.93	0.96	0.96	0.98	0.98	0.99
	0.70	0.84	0.72	0.77	0.84	0.88	0.91	0.94	0.96	0.97	0.98	0.99	0.99	1.00
1000	0.10	0.32	0.21	0.31	0.26	0.39	0.32	0.48	0.38	0.56	0.44	0.64	0.50	0.70
	0.20	0.45	0.37	0.51	0.47	0.63	0.56	0.74	0.65	0.82	0.73	0.88	0.79	0.92
	0.30	0.55	0.51	0.65	0.63	0.78	0.74	0.87	0.82	0.93	0.88	0.96	0.92	0.98
	0.40	0.63	0.63	0.75	0.76	0.86	0.85	0.93	0.91	0.97	0.95	0.99	0.98	0.99
	0.50	0.71	0.73	0.82	0.84	0.91	0.92	0.96	0.96	0.99	0.98	1.00	0.99	1.00
	0.60	0.77	0.80	0.86	0.90	0.94	0.96	0.98	0.98	0.99	0.99	1.00	1.00	1.00
	0.70	0.84	0.86	0.90	0.94	0.96	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00

C.3 References

1. **RKI**, Health in Germany/Gesundheit in Deutschland. *Robert Koch Institut*, 2007(ISBN 3-89606-173-9).
2. **Van Baal, P., et al.**, Lifetime medical costs of obesity: prevention no cure for increasing health expenditure. *PLoS Med*, 2008. 5(2): p. e29.
3. **Rappange, D., et al.**, Lifestyle intervention: from cost savings to value for money. *Journal of Public Health*, 2010. 32(3): p. 440.
4. **Razum, O., et al.**, Social Epidemiology after the German Reunification: East vs. West or Poor vs. Rich? *International Journal of Public Health*, 2008. 53(1): p. 13-22.
5. **Holford, T. and C. Stack**, Study design for epidemiologic studies with measurement error. *Statistical Methods in Medical Research*, 1995. 4(4): p. 339.
6. **White, E., J. Hunt, and D. Casso**, Exposure measurement in cohort studies: the challenges of prospective data collection. *Epidemiologic reviews*, 1998. 20(1): p. 43.
7. **Anderson, G.L., et al.**, Assessing Lead Time of Selected Ovarian Cancer Biomarkers: A Nested Case–Control Study. *Journal of the National Cancer Institute*, 2010. 102(1): p. 26-38.
8. **Muller-Nordhorn, J., et al.**, Regional variation and time trends in mortality from ischaemic heart disease: East and West Germany 10 years after reunification. *British Medical Journal*, 2004. 58(6): p. 481.
9. **Klapper, A., D. Bardehle, and O. Razum**, Alters-und geschlechtsspezifische Mortalität im Ruhrgebiet von 1994 bis 2004 [Age-and sex-specific mortality in the Ruhr region from 1994 to 2004]. *Das Gesundheitswesen*, 2007. 69(10): p. 521–526.
10. **Dragano, N., et al.**, Subclinical coronary atherosclerosis and neighbourhood deprivation in an urban region. *European journal of epidemiology*, 2009. 24(1): p. 25-35.
11. **Dragano, N., et al.**, Neighbourhood socioeconomic status and cardiovascular risk factors: a multilevel analysis of nine cities in the Czech Republic and Germany. *BMC Public Health*, 2007. 7(1): p. 255.
12. **Ouzounian, M., et al.**, Predict, prevent and personalize: Genomic and proteomic approaches to cardiovascular medicine. *The Canadian journal of cardiology*, 2007. 23(Suppl A): p. 28A.
13. **Dent, T.H.S.**, Predicting the risk of coronary heart disease: I. The use of conventional risk markers. *Atherosclerosis*, 2010. 213(2): p. 345-351.
14. **Dent, T.H.S.**, Predicting the risk of coronary heart disease. II: The role of novel molecular biomarkers and genetics in estimating risk, and the future of risk prediction. *Atherosclerosis*, 2010. 213(2): p. 352-362.
15. **Wilson, P., et al.**, Prediction of coronary heart disease using risk factor categories. *Circulation*, 1998. 97(18): p. 1837.
16. **Conroy, R., et al.**, Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European heart journal*, 2003. 24(11): p. 987.

17. **Keil, U., et al.**, Risikoabschätzung tödlicher Herz-Kreislauf-Erkrankungen. Die neuen SCORE-Deutschland-Tabellen für die Primärprävention. *Deutsches Ärzteblatt*, 2005. 102(25): p. A1808-1812.
18. **Schulze, M., et al.**, Der Deutsche Diabetes-Risiko-Score (DRS). *ERNAHRUNGSSUMSCHAU*, 2007. 54(3): p. 122.
19. **Lindström, J. and J. Tuomilehto**, The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care*, 2003. 26(3): p. 725-31.
20. **Erbel, R., et al.**, Coronary Risk Stratification, Discrimination, and Reclassification Improvement Based on Quantification of Subclinical Coronary Atherosclerosis: The Heinz Nixdorf Recall Study. *Journal of the American College of Cardiology*, 2010. 56(17): p. 1397-1406.
21. **Gail, M.H., et al.**, Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually. *Journal of the National Cancer Institute*, 1989. 81(24): p. 1879-1886.
22. **Wacholder, S., et al.**, Performance of Common Genetic Variants in Breast-Cancer Risk Models. *New England Journal of Medicine*, 2010. 362(11): p. 986-993.
23. **Greenland, P., et al.**, 2010 ACCF/AHA Guideline for Assessment of Cardiovascular Risk in Asymptomatic Adults: Executive Summary: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines Developed in Collaboration With the American Society of Echocardiography, American Society of Nuclear Cardiology, Society of Atherosclerosis Imaging and Prevention, Society for Cardiovascular Angiography and Interventions, Society of Cardiovascular Computed Tomography, and Society for Cardiovascular Magnetic Resonance. *Journal of the American College of Cardiology*, 2010. 56(25): p. 2182-2199.
24. **Ridker, P.M., et al.**, Development and Validation of Improved Algorithms for the Assessment of Global Cardiovascular Risk in Women. *JAMA: The Journal of the American Medical Association*, 2007. 297(6): p. 611-619.
25. **Melander, O., et al.**, Novel and Conventional Biomarkers for Prediction of Incident Cardiovascular Events in the Community. *JAMA: The Journal of the American Medical Association*, 2009. 302(1): p. 49-57.
26. **van Zitteren, M., et al.**, Genome-Based Prediction of Breast Cancer Risk in the General Population: A Modeling Study Based on Meta-Analyses of Genetic Associations. *Cancer Epidemiology Biomarkers & Prevention*, 2011. 20(1): p. 9-22.
27. **Thanassoulis, G. and R.S. Vasan**, Genetic Cardiovascular Risk Prediction: Will We Get There? *Circulation*, 2010. 122(22): p. 2323-2334.
28. **Ruchat, S.-M., et al.**, Combining genetic markers and clinical risk factors improves the risk assessment of impaired glucose metabolism. *Annals of Medicine*, 2010. 42(3): p. 196-206.
29. **Blennow, K., et al.**, Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nat Rev Neurol*, 2010. 6(3): p. 131-144.
30. **Prentice, R.**, Design issues in cohort studies. *Statistical Methods in Medical Research*, 1995. 4(4): p. 273.
31. **Wickramaratne, P.**, Sample size determination in epidemiologic studies. *Statistical Methods in Medical Research*, 1995. 4(4): p. 311.

32. **Burton, P., et al.**, Size matters: just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology. *International journal of epidemiology*, 2008.
33. **World Health Organization**, The global burden of disease: 2004 update. 2008, Geneva: *WHO Press*.
34. **Statistisches Bundesamt Deutschland**. Todesursachenstatistik 2007. 2009 [cited 2009 July 29]; Available from: <http://www.destatis.de/>.
35. **Statistisches Bundesamt Deutschland**. Krankheitskosten 2006. 2009 [cited 2009 July 29]; Available from: <http://www.destatis.de/>.
36. **Leal, J., et al.**, Economic burden of cardiovascular diseases in the enlarged European Union. *Eur Heart J*, 2006. 27(13): p. 1610-9.
37. **Sanderson, J.E., et al.**, Global burden of cardiovascular disease. *Heart*, 2007. 93(10): p. 1175.
38. **Robert Koch Institut**, Koronare Herzkrankheit und akuter Myokardinfarkt, in Gesundheitsberichterstattung des Bundes (Heft 33). 2006, *Robert Koch Institut*: Berlin.
39. **Mosterd, A. and A.W. Hoes**, Clinical epidemiology of heart failure. *British Medical Journal*, 2007. 93(9): p. 1137.
40. **Kannel, W.B. and A.J. Belanger**, Epidemiology of heart failure. *American heart journal*, 1991. 121(3): p. 951-957.
41. **Lloyd-Jones, D.M., et al.**, Lifetime risk for developing congestive heart failure: the Framingham Heart Study. *Circulation*, 2002. 106(24): p. 3068.
42. **Leong, D.P., C.G. De Pasquale, and J.B. Selvanayagam**, Heart Failure With Normal Ejection Fraction: The Complementary Roles of Echocardiography and CMR Imaging. *JACC Cardiovascular Imaging*, 2010. 3(4): p. 409.
43. **Paulus, W.J., et al.**, How to diagnose diastolic heart failure: a consensus statement on the diagnosis of heart failure with normal left ventricular ejection fraction by the Heart Failure and Echocardiography Associations of the European Society of Cardiology. *European heart journal*, 2007.
44. **Owan, T.E., et al.**, Trends in prevalence and outcome of heart failure with preserved ejection fraction. *New England Journal of Medicine*, 2006. 355(3): p. 251.
45. **Bursi, F., et al.**, Systolic and diastolic heart failure in the community. *Jama*, 2006. 296(18): p. 2209.
46. **Redfield, M.M., et al.**, Burden of systolic and diastolic ventricular dysfunction in the community: appreciating the scope of the heart failure epidemic. *Jama*, 2003. 289(2): p. 194.
47. **Lee, D.S., et al.**, Relation of disease pathogenesis and risk factors to heart failure with preserved or reduced ejection fraction: insights from the framingham heart study of the national heart, lung, and blood institute. *Circulation*, 2009. 119(24): p. 3070.
48. **Dickstein, K., et al.**, ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2008. *European heart journal*, 2008. 29(19): p. 2388.
49. **Lloyd-Jones, D.M., et al.**, Lifetime risk for development of atrial fibrillation: the Framingham Heart Study. *Circulation*, 2004. 110(9): p. 1042-1046.

50. **Benjamin, E.J., et al.**, Prevention of atrial fibrillation: report from a national heart, lung, and blood institute workshop. *Circulation*, 2009. 119(4): p. 606-618.
51. **Miyasaka, Y., et al.**, Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. *Circulation*, 2006. 114(2): p. 119-125.
52. **Das, R.R., et al.**, Prevalence and correlates of silent cerebral infarcts in the Framingham offspring study. *Stroke*, 2008. 39(11): p. 2929-2935.
53. **Knecht, S., et al.**, Atrial fibrillation in stroke-free patients is associated with memory impairment and hippocampal atrophy. *Eur Heart J*, 2008.
54. **Grote, V.A., S. Becker, and R. Kaaks**, Diabetes mellitus type 2-an independent risk factor for cancer? *Exp Clin Endocrinol Diabetes*, 2010. 118(1): p. 4-8.
55. **WHO**, The Global Burden of Disease: 2004 Update. 2008: *World Health Organization*.
56. **Franse, L.V., et al.**, Type 2 Diabetes in Older Well-Functioning People: Who Is Undiagnosed? *Diabetes care*, 2001. 24(12): p. 2065.
57. **Rathmann, W., et al.**, High prevalence of undiagnosed diabetes mellitus in Southern Germany: target populations for efficient screening. The KORA survey 2000. *Diabetologia*, 2003. 46(2): p. 182-189.
58. **Passa, P.**, Diabetes trends in Europe. *Diabetes/Metabolism Research and Reviews*, 2002. 18(S3): p. S3-S8.
59. **Chiasson, J.L., et al.**, Acarbose for the prevention of Type 2 diabetes, hypertension and cardiovascular disease in subjects with impaired glucose tolerance: facts and interpretations concerning the critical analysis of the STOP-NIDDM Trial data. *Diabetologia*, 2004. 47(6): p. 969-975.
60. **Kosaka, K., M. Noda, and T. Kuzuya**, Prevention of type 2 diabetes by lifestyle intervention: a Japanese trial in IGT males. *Diabetes research and clinical practice*, 2005. 67(2): p. 152-162.
61. **Li, G., et al.**, The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. *The Lancet*, 2008. 371(9626): p. 1783-1789.
62. **Lindström, J., et al.**, Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. *The Lancet*, 2006. 368(9548): p. 1673-1679.
63. **Pan, X.R., et al.**, Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. *Diabetes care*, 1997. 20(4): p. 537.
64. **Tuomilehto, J., et al.**, Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *New England Journal of Medicine*, 2001. 344(18): p. 1343.
65. **Borcuk, A.C. and C.A. Powell**. Expression profiling and lung cancer development. 2007: *Am Thoracic Soc*.
66. **Doll, A., et al.**, Novel molecular profiles of endometrial cancer--new light through old windows. *The Journal of steroid biochemistry and molecular biology*, 2008. 108(3-5): p. 221-229.

67. **Hecht, J.L. and G.L. Mutter**, Molecular and pathologic aspects of endometrial carcinogenesis. *Journal of Clinical Oncology*, 2006. 24(29): p. 4783.
68. **Jass, J.R.**, Serrated adenoma of the colorectum and the DNA-methylator phenotype. *Nature Clinical Practice Oncology*, 2005. 2(8): p. 398-405.
69. **Jass, J.R.**, Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology*, 2007. 50(1): p. 113-130.
70. **Rakha, E.A., et al.**, Expression profiling technology: its contribution to our understanding of breast cancer. *Histopathology*, 2008. 52(1): p. 67-81.
71. **Ricciardelli, C. and M.K. Oehler**, Diverse molecular pathways in ovarian cancer and their clinical significance. *Maturitas*, 2009. 62(3): p. 270-275.
72. **Sato, M., et al.**, A translational view of the molecular pathogenesis of lung cancer. *Journal of Thoracic Oncology*, 2007. 2(4): p. 327.
73. **Scott, M. and W.G. McCluggage**, Current concepts in ovarian epithelial tumorigenesis: correlation between morphological and molecular data. *Histology and histopathology*, 2006. 21(1): p. 81.
74. **Snover, D.C., et al.**, Serrated polyps of the large intestine. *American journal of clinical pathology*, 2005. 124(3): p. 380.
75. **Toyota, M., et al.**, CpG island methylator phenotype in colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 1999. 96(15): p. 8681.
76. **Yehiely, F., et al.**, Deconstructing the molecular portrait of basal-like breast cancer. *Trends in Molecular Medicine*, 2006. 12(11): p. 537-544.
77. **Felix, A.S., et al.**, Factors associated with Type I and Type II endometrial cancer. *Cancer Causes and Control*, 2010: p. 1-6.
78. **Anderson, W.F., et al.**, Comparison of age-specific incidence rate patterns for different histopathologic types of breast carcinoma. *Cancer Epidemiology Biomarkers & Prevention*, 2004. 13(7): p. 1128.
79. **Colditz, G.A., et al.**, Risk factors for breast cancer according to estrogen and progesterone receptor status. *Journal of the National Cancer Institute*, 2004. 96(3): p. 218.
80. **de Vogel, S., et al.**, Dietary folate and APC mutations in sporadic colorectal cancer. *Journal of Nutrition*, 2006. 136(12): p. 3015.
81. **Janssen-Heijnen, M.L.G. and J.W.W. Coebergh**, The changing epidemiology of lung cancer in Europe. *Lung Cancer*, 2003. 41(3): p. 245-258.
82. **Weijenberg, M.P., et al.**, Dietary fat and risk of colon and rectal cancer with aberrant MLH1 expression, APC or KRAS genes. *Cancer Causes and Control*, 2007. 18(8): p. 865-879.
83. **Zhang, S.M., et al.**, Folate intake and risk of breast cancer characterized by hormone receptor status. *Cancer Epidemiology Biomarkers & Prevention*, 2005. 14(8): p. 2004.
84. **Kaaks, R. and P. Stattin**, Obesity, Endogenous Hormone Metabolism, and Prostate Cancer Risk: A Conundrum of "Highs" and "Lows". *Cancer Prevention Research*, 2010. 3(3): p. 259.

References

85. **Calle, E.E. and R. Kaaks**, Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nature Reviews Cancer*, 2004. 4(8): p. 579-591.
 86. **Chen, Y.C. and D.J. Hunter**, Molecular epidemiology of cancer. *CA: A Cancer Journal for Clinicians*, 2005. 55(1): p. 45-54.
 87. **Colditz, G.A., T.A. Sellers, and E. Trapido**, Epidemiology—identifying the causes and preventability of cancer? *Nature Reviews Cancer*, 2005. 6(1): p. 75-83.
 88. **Peto, J.**, Cancer epidemiology in the last century and the next decade. *Nature*, 2001. 411(6835): p. 390-395.
 89. **Amarenco, P., et al.**, Classification of stroke subtypes. *Cerebrovasc Dis*, 2009. 27(5): p. 493-501.
 90. **Qiu, C., D. De Ronchi, and L. Fratiglioni**, The epidemiology of the dementias: an update. *Current Opinion in Psychiatry*, 2007. 20(4): p. 380-385
10.1097/YCO.0b013e32816ebc7b.
 91. **Torre, J.C.d.I.**, Alzheimer's disease: How does it start? *Journal of Alzheimer's Disease*, 2002. 4(6): p. 497-512.
 92. **Wells, K., J. Golding, and M. Burnam**, Affective, substance use, and anxiety disorders in persons with arthritis, diabetes, heart disease, high blood pressure, or chronic lung conditions. *General Hospital Psychiatry*, 1989.
 93. **Rugulies, R.**, Depression as a predictor for coronary heart disease:: a review and meta-analysis1. *American Journal of Preventive Medicine*, 2002. 23(1): p. 51-61.
 94. **Campayo, A., C.H. Gómez-Biel, and A. Lobo**, Diabetes and Depression. *Current Psychiatry Reports*, 2010: p. 1-5.
 95. **De Groot, M., et al.**, Association of depression and diabetes complications: a meta-analysis. *Psychosomatic Medicine*, 2001. 63(4): p. 619.
 96. **Lloyd, C.E., et al.**, Cross cultural comparisons of anxiety and depression in adults with type 1 diabetes. *Diabetes/Metabolism Research and Reviews*, 2003. 19(5): p. 401-407.
 97. **Penninx, B., et al.**, Inflammatory markers and depressed mood in older persons: results from the Health, Aging and Body Composition study. *Biological Psychiatry*, 2003. 54(5): p. 566-572.
 98. **Maes, M.**, A review on the acute phase response in major depression. *Reviews in the Neurosciences*. 4(4): p. 407.
 99. **Maes, M.**, Evidence for an immune response in major depression: a review and hypothesis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 1995. 19(1): p. 11-38.
 100. **Maes, M., et al.**, Psychomotor retardation, anorexia, weight loss, sleep disturbances, and loss of energy: psychopathological correlates of hyperhaptoglobinemia during major depression. *Psychiatry research*, 1993. 47(3): p. 229-241.
 101. **Maes, M., et al.**, Alterations in iron metabolism and the erythron in major depression: Further evidence for a chronic inflammatory process. *Journal of affective disorders*, 1996. 40(1-2): p. 23-33.
-

102. **Fairbanks, V., et al.**, Modern nutrition in health and disease. *Lea and Febiger, Philadelphia, PA*, 1988: p. 193-226.
103. **Brock, J.**, Iron in infection, immunity, inflammation and neoplasia. *Iron metabolism in health and disease*, 1994: p. 353–372.
104. **Kemper, C.J., M. Ziegler, and S. Taylor**, Überprüfung der psychometrischen Qualität der deutschen Version des Angstsensitivitätsindex-3. *Diagnostica*, 2009. 55(4): p. 223-233.
105. **Stovner, L., et al.**, The global burden of headache: a documentation of headache prevalence and disability worldwide. *Cephalalgia*, 2007. 27(3): p. 193-210.
106. **Lambert, J., et al.**, Impact of migraine symptoms on health care use and work loss in Canada in patients randomly assigned in a phase III clinical trial. *Can J Clin Pharmacol*, 2002. 9(3): p. 158-64.
107. **Lipton, R.B., W.F. Stewart, and D. Simon**, Medical consultation for migraine: results from the American Migraine Study. *Headache*, 1998. 38(2): p. 87-96.
108. **Holzhammer, J. and C. Wöber**, Non-alimentary trigger factors of migraine and tension-type headache. *Schmerz (Berlin, Germany)*, 2006. 20(3): p. 226.
109. **Holzhammer, J. and C. Wöber**, [Alimentary trigger factors that provoke migraine and tension-type headache]. *Schmerz (Berlin, Germany)*, 2006. 20(2): p. 151.
110. **Panconesi, A.**, Alcohol and migraine: trigger factor, consumption, mechanisms. A review. *J Headache Pain*, 2008. 9(1): p. 19-27.
111. **Spierings, E.L., A.H. Ranke, and P.C. Honkoop**, Precipitating and aggravating factors of migraine versus tension-type headache. *Headache*, 2001. 41(6): p. 554-8.
112. **Aamodt, A.H., et al.**, Headache prevalence related to smoking and alcohol use. The Head-HUNT Study. *European Journal of Neurology*, 2006. 13(11): p. 1233-1238.
113. **Bigal, M.E., J.N. Liberman, and R.B. Lipton**, Obesity and migraine: a population study. *Neurology*, 2006. 66(4): p. 545-50.
114. **Walters, A., et al.**, Toward a better definition of the restless legs syndrome. *Movement Disorders*, 1995. 10(5): p. 634-642.
115. **Garcia-Borreguero, D., et al.**, Epidemiology of restless legs syndrome: the current status. *Sleep Medicine Reviews*, 2006. 10(3): p. 153-167.
116. **Von Campenhausen, S., et al.**, Prevalence and incidence of Parkinson's disease in Europe. *European Neuropsychopharmacology*, 2005. 15(4): p. 473-490.
117. **Kleinhenz, J., et al.**, The prevalence of Parkinson disease in West Germany--are general practice data a suitable survey instrument? *Das Öffentliche Gesundheitswesen*, 1990. 52(4): p. 181.
118. **Schenkman, M., et al.**, Longitudinal evaluation of economic and physical impact of Parkinson's disease. *Parkinsonism & Related Disorders*, 2001. 8(1): p. 41-50.
119. **Pritzkeleit, R., F. Beske, and A. Katalinic**, [Disease numbers in pneumology - a projection to 2060]. *Pneumologie*, 2010. 64(9): p. 535-40.
120. **Sullivan, S.D., S.D. Ramsey, and T.A. Lee**, The Economic Burden of COPD*. *Chest*, 2000. 117(2 suppl): p. 5S.

121. **Walter, R., D.J. Gottlieb, and G.T. O'Connor**, Environmental and genetic risk factors and gene-environment interactions in the pathogenesis of chronic obstructive lung disease. *Environmental health perspectives*, 2000. 108(Suppl 4): p. 733.
122. **Anderson, D. and W. MacNee**, Targeted treatment in COPD: a multi-system approach for a multi-system disease. *International Journal of Chronic Obstructive Pulmonary Disease*, 2009. 4: p. 321.
123. **Tuder, R.M., et al.** State of the art. Cellular and molecular mechanisms of alveolar destruction in emphysema: an evolutionary perspective. 2006: *Am Thoracic Soc*.
124. **Barnes, P.J.** Future treatments for chronic obstructive pulmonary disease and its comorbidities. 2008: *Am Thoracic Soc*.
125. **Königshoffa, M., N. Kneidingerb, and O. Eickelberga**, TGF- signalling in COPD: deciphering genetic and cellular susceptibilities for future therapeutic regimens. *Swiss medical weekly: official journal of the Swiss Society of Infectious Diseases, the Swiss Society of Internal Medicine, the Swiss Society of Pneumology*, 2009. 139(39-40): p. 554.
126. **Masoli, M., et al.**, The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy*, 2004. 59(5): p. 469-478.
127. **Eder, W., M.J. Ege, and E. von Mutius**, The asthma epidemic. *New England Journal of Medicine*, 2006. 355(21): p. 2226.
128. **Sleiman, P. and H. Hakonarson**, Recent advances in the genetics and genomics of asthma and related traits. *Current Opinion in Pediatrics*, 2010. 22(3): p. 307.
129. **Mutius, E.v., et al.**, Prevalence of asthma and allergic disorders among children in united Germany: a descriptive comparison. *Br. Med. J.*, 1992. 305 p. 1395-1399.
130. **Nowak, D.**, Information on the new 4111 occupational disease: "chronic obstructive bronchitis or emphysema in miners of bituminous coal mines with evidence of the exposure to a cumulative fine dust dose around 100 ([mg/m³] x years). *Pneumologie*, 1996. 50(9): p. 652.
131. **Monto, A.S.**, Epidemiology of viral respiratory infections. *The American journal of medicine*, 2002. 112(6): p. 4-12.
132. **Armstrong, G.L. and R.W. Pinner**, Outpatient visits for infectious diseases in the United States, 1980 through 1996. *Archives of Internal Medicine*, 1999. 159(21): p. 2531.
133. **Nicholson, K.G., et al.**, Risk factors for lower respiratory complications of rhinovirus infections in elderly people living in the community: prospective cohort study. *British Medical Journal*, 1996. 313(7065): p. 1119.
134. **Sato, M., et al.**, Antibody response to influenza vaccination in nursing home residents and healthcare workers during four successive seasons in Niigata, Japan. *Infection control and hospital epidemiology*, 2005. 26(11): p. 859-866.
135. **Mossong, J., et al.**, Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*, 2008. 5(3): p. e74.
136. **Barnett, J.B., D.H. Hamer, and S.N. Meydani**, Low zinc status: a new risk factor for pneumonia in the elderly? *Nutrition reviews*. 68(1): p. 30-37.
137. **RKI**, Infektionsepidemiologisches Jahrbuch meldepflichtiger Krankheiten für 2009. 2010, *Robert Koch Institut*: Berlin.

138. **Wheeler, J.G., et al.**, Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. *Bmj*, 1999. 318(7190): p. 1046.
139. **De Wit, M.A.S., et al.**, Sensor, a population-based cohort study on gastroenteritis in the Netherlands: incidence and etiology. *American journal of Epidemiology*, 2001. 154(7): p. 666.
140. **Simonsen, J., et al.**, Usefulness of seroconversion rates for comparing infection pressures between countries. *Epidemiology and Infection*, 2006: p. 1-8.
141. **Shirato-Horikoshi, H., et al.**, Binding activity of norovirus and sapovirus to histo-blood group antigens. *Archives of virology*, 2007. 152(3): p. 457-461.
142. **O'Brien Sarah, R., et al.**, Methods for determining disease burden and calibrating national surveillance data in the United Kingdom: the second study of infectious intestinal disease in the community (IID2 study). *BMC medical research methodology*, 2010. 10.
143. **Völzke, H., et al.**, Cohort Profile: The Study of Health in Pomerania. *International journal of epidemiology*.
144. **Holtfreter, B., et al.**, Prevalence of periodontal disease and treatment demands based on a German dental survey (DMS IV). *Journal of clinical periodontology*. 37(3): p. 211-219.
145. **Bollen, A.M., et al.**, The effects of orthodontic therapy on periodontal health: a systematic review of controlled evidence. *The Journal of the American Dental Association*, 2008. 139(4): p. 413.
146. **Chávarry, N.G., et al.**, The relationship between diabetes mellitus and destructive periodontal disease: a meta-analysis. *Oral health & preventive dentistry*, 2009. 7(2): p. 107.
147. **Demmer, R.T. and M. Desvarieux**, Periodontal infections and cardiovascular disease: the heart of the matter. *The Journal of the American Dental Association*, 2006. 137(2): p. 14S.
148. **Demmer, R.T., D.R. Jacobs Jr, and M. Desvarieux**, Periodontal disease and incident type 2 diabetes: results from the First National Health and Nutrition Examination Survey and its epidemiologic follow-up study. *Diabetes care*, 2008. 31(7): p. 1373.
149. **Desvarieux, M., et al.**, Gender differences in the relationship between periodontal disease, tooth loss, and atherosclerosis. *Stroke*, 2004. 35(9): p. 2029.
150. **Detert, J., et al.**, The association between rheumatoid arthritis and periodontal disease. *Arthritis Research & Therapy*, 2010. 12(5): p. 218.
151. **Grau, A.J., et al.**, Periodontal disease as a risk factor for ischemic stroke. *Stroke*, 2004: p. 10000110.
152. **Pischon, N., et al.**, Obesity, inflammation, and periodontal disease. *Journal of dental research*, 2007. 86(5): p. 400.
153. **Cameron, A.J., J.E. Shaw, and P.Z. Zimmet**, The metabolic syndrome: prevalence in worldwide populations. *Endocrinology and Metabolism Clinics*, 2004. 33(2).
154. **McGill, H.C., Jr.**, The pathogenesis of atherosclerosis. *Clin Chem*, 1988. 34(8B): p. B33-9.

155. **Ross, R.**, Atherosclerosis--an inflammatory disease. *N Engl J Med*, 1999. 340(2): p. 115-26.
156. **Wang, T.J., et al.**, Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: the Framingham Heart Study. *Circulation*, 2003. 107(23): p. 2920.
157. **Biaggioni, I.**, Sympathetic control of the circulation in hypertension: lessons from autonomic disorders. *Curr Opin Nephrol Hypertens*, 2003. 12(2): p. 175-180.
158. **Burns, J., et al.**, Relationship Between Central Sympathetic Drive and Magnetic Resonance Imaging-Determined Left Ventricular Mass in Essential Hypertension. *Circulation*, 2007. 115(15): p. 1999-2005.
159. **Grassi, G.**, Assessment of Sympathetic Cardiovascular Drive in Human Hypertension: Achievements and Perspectives. *Hypertension*, 2009. 54(4): p. 690-697.
160. **Carnethon, M.R., et al.**, Prospective investigation of autonomic nervous system function and the development of type 2 diabetes: the Atherosclerosis Risk In Communities study, 1987-1998. *Circulation*, 2003. 107(17): p. 2190-2195.
161. **Straznicky, N.E., et al.**, Mediators of sympathetic activation in metabolic syndrome obesity. *Curr Hypertens Rep*, 2008. 10(6): p. 440-447.
162. **Masuo, K., et al.**, Sympathetic nerve hyperactivity precedes hyperinsulinemia and blood pressure elevation in a young, nonobese Japanese population. *Am J Hypertens*, 1997. 10(1): p. 77-83.
163. **Pagani, M., et al.**, Low and high frequency components of blood pressure variability. *Ann N Y Acad Sci*, 1996. 783:10-23.: p. 10-23.
164. **Pomeranz, B., et al.**, Assessment of autonomic function in humans by heart rate spectral analysis. *Am J Physiol*, 1985. 248(1 Pt 2): p. H151-H153.
165. **Stumvoll, M., B.J. Goldstein, and T.W. van Haeften**, Type 2 diabetes: principles of pathogenesis and therapy. *The Lancet*, 2005. 365(9467): p. 1333-1346.
166. **Hanley, A.J.G., et al.**, Prediction of type 2 diabetes using simple measures of insulin resistance. *Diabetes*, 2003. 52(2): p. 463.
167. **Stumvoll, M., A. Fritsche, and H. Häring**, The OGTT as test for beta cell function? *European Journal of Clinical Investigation*, 2001. 31(5): p. 380-381.
168. **Goldin, A., et al.**, Advanced glycation end products: sparking the development of diabetic vascular injury. *Circulation*, 2006. 114(6): p. 597.
169. **Bierhaus, A. and P. Nawroth**, Multiple levels of regulation determine the role of the receptor for AGE (RAGE) as common soil in inflammation, immune responses and diabetes mellitus and its complications. *Diabetologia*, 2009. 52(11): p. 2251-2263.
170. **Mulder, D.J., et al.**, Skin autofluorescence, a novel marker for glycemic and oxidative stress-derived advanced glycation endproducts: an overview of current clinical studies, evidence, and limitations. *Diabetes Technology & Therapeutics*, 2006. 8(5): p. 523-535.
171. **Genuth, S., et al.**, Glycation and carboxymethyllysine levels in skin collagen predict the risk of future 10-year progression of diabetic retinopathy and nephropathy in the diabetes control and complications trial and epidemiology of diabetes interventions and complications participants with type 1 diabetes. *Diabetes*, 2005. 54(11): p. 3103.

172. **Gerrits, E.G., et al.**, Skin Autofluorescence. *Diabetes care*, 2008. 31(3): p. 517.
173. **Lutgers, H.L., et al.**, Skin autofluorescence as a noninvasive marker of vascular damage in patients with type 2 diabetes. *Diabetes care*, 2006. 29(12): p. 2654.
174. **Monnier, V.M., et al.**, Skin collagen glycation, glycooxidation, and crosslinking are lower in subjects with long-term intensive versus conventional therapy of type 1 diabetes: relevance of glycated collagen products versus HbA1c as markers of diabetic complications. DCCT Skin Collagen Ancillary Study Group. Diabetes Control and Complications Trial. *Diabetes*, 1999. 48(4): p. 870.
175. **Lutgers, H., et al.**, Skin autofluorescence provides additional information to the UK Prospective Diabetes Study (UKPDS) risk score for the estimation of cardiovascular prognosis in type 2 diabetes mellitus. *Diabetologia*, 2009. 52(5): p. 789-797.
176. **Ghidoni, R., et al.**, Decreased plasma levels of soluble receptor for advanced glycation end products in mild cognitive impairment. *Journal of Neural Transmission*, 2008. 115(7): p. 1047-1050.
177. **Li, K., et al.**, Association between the RAGE G82S polymorphism and Alzheimer's disease. *Journal of Neural Transmission*, 2010. 117(1): p. 97-104.
178. **Cheung, N., P. Mitchell, and T.Y. Wong**, Diabetic retinopathy. *The Lancet*, 2010. 376(9735): p. 124-136.
179. **Peng, X.Y., et al.**, Retinopathy in Persons without Diabetes:: The Handan Eye Study. *Ophthalmology*, 2010. 117(3): p. 531-537.
180. **Wong, T.Y., et al.**, Relation between fasting glucose and retinopathy for diagnosis of diabetes: three population-based cross-sectional studies. *The Lancet*, 2008. 371(9614): p. 736-743.
181. **Wong, T.Y., et al.**, Three-year incidence and cumulative prevalence of retinopathy: the atherosclerosis risk in communities study. *American journal of ophthalmology*, 2007. 143(6): p. 970-976.
182. **Baker, M.L., et al.**, Retinal signs and stroke: revisiting the link between the eye and brain. *Stroke*, 2008. 39(4): p. 1371.
183. **Wong, T.Y., et al.**, A prospective cohort study of retinal arteriolar narrowing and mortality. *American journal of Epidemiology*, 2004. 159(9): p. 819.
184. **Wong, T.Y. and R. McIntosh**, Hypertensive retinopathy signs as risk indicators of cardiovascular morbidity and mortality. *British medical bulletin*, 2005. 73(1): p. 57.
185. **Nieto, W.G., et al.**, Increased frequency (12%) of circulating chronic lymphocytic leukemia-like B-cell clones in healthy subjects using a highly sensitive multicolor flow cytometry approach. *Blood*, 2009. 114(1): p. 33.
186. **Landgren, O., et al.**, Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. *Blood*, 2009. 113(22): p. 5412.
187. **Rabkin, C.S., et al.**, t (14; 18) Translocations and risk of follicular lymphoma. *JNCI Monographs*, 2008. 2008(39): p. 48.
188. **Chiorazzi, N. and M. Ferrarini**, B CELL CHRONIC LYMPHOCYTIC LEUKEMIA: Lessons Learned from Studies of the B Cell Antigen Receptor. *Annual Review of Immunology*, 2003. 21(1): p. 841-894.

189. **Ghiotto, F., et al.**, Remarkably similar antigen receptors among a subset of patients with chronic lymphocytic leukemia. *Journal of Clinical Investigation*, 2004. 113(7): p. 1008-1016.
190. **Hamblin, T.**, Is chronic lymphocytic leukemia a response to infectious agents? *Leukemia research*, 2006.
191. **Luck, T., et al.**, Mild cognitive impairment: incidence and risk factors: results of the leipzig longitudinal study of the aged. *J Am Geriatr Soc*, 2010. 58(10): p. 1903-10.
192. **Blacker, D., et al.**, Neuropsychological measures in normal individuals that predict subsequent cognitive decline. *Arch Neurol*, 2007. 64(6): p. 862-71.
193. **Landis, B.N. and T. Hummel**, New evidence for high occurrence of olfactory dysfunctions within the population. *The American journal of medicine*, 2006. 119(1): p. 91.
194. **Brämerson, A., et al.**, Prevalence of Olfactory Dysfunction: The Skövde Population Based Study. *The Laryngoscope*, 2004. 114(4): p. 733-737.
195. **Hoffman, H.J., E.K. Ishii, and R.H. Macturk**, Age Related Changes in the Prevalence of Smell/Taste Problems among the United States Adult Population: Results of the 1994 Disability Supplement to the National Health Interview Survey (NHIS). *Annals of the New York Academy of Sciences*, 1998. 855(1): p. 716-722.
196. **Bacon, A.W., et al.**, Very Early Changes in Olfactory Functioning Due to Alzheimer's Disease and the Role of Apolipoprotein E in Olfaction. *Annals of the New York Academy of Sciences*, 1998. 855(OLFACTION AND TASTE XII: AN INTERNATIONAL SYMPOSIUM): p. 723-731.
197. **Haehner, A., et al.**, Olfactory loss may be a first sign of idiopathic Parkinson's disease. *Movement Disorders*, 2007. 22(6): p. 839-842.
198. **Murphy, C.**, Loss of olfactory function in dementing disease. *Physiology & behavior*, 1999. 66(2): p. 177-182.
199. **Ponsen, M.M., et al.**, Idiopathic hyposmia as a preclinical sign of Parkinson's disease. *Annals of Neurology*, 2004. 56(2): p. 173-181.
200. **Zielinski, J., et al.**, Increasing COPD awareness. *Eur Respir J*, 2006. 27(4): p. 833-52.
201. **Mannino DM, H.F., Pavlin BI, Ferdinands JM**, Risk factors for prevalence of and mortality related to restriction on spirometry: findings from the First National Health and Nutrition Examination Survey and follow-up. *The International Journal of Tuberculosis and Lung Disease*, 2005. 9: p. 613-621.
202. **Scarlata, S., et al.**, Restrictive pulmonary dysfunction at spirometry and mortality in the elderly. *Respiratory Medicine*, 2008. 102(9): p. 1349-1354.
203. **Kohansal, R., et al.**, The Natural History of Chronic Airflow Obstruction Revisited: An Analysis of the Framingham Offspring Cohort. *Am. J. Respir. Crit. Care Med.*, 2009. 180(1): p. 3-10.
204. **Jacobsen, P.K., et al.**, LUNG FUNCTION AS A PREDICTOR OF SURVIVAL IN VERY ELDERLY PEOPLE: THE DANISH 1905 COHORT STUDY. *Journal of the American Geriatrics Society*, 2008. 56(11): p. 2150-2152.
205. **Stavem, K., et al.**, Lung function, smoking and mortality in a 26-year follow-up of healthy middle-aged males. *Eur Respir J*, 2005. 25(4): p. 618-25.
206. **Sin, D.D., L. Wu, and S.F.P. Man**, The Relationship Between Reduced Lung Function and Cardiovascular Mortality*. *Chest*, 2005. 127(6): p. 1952-1959.

207. **ATS**, ATS Workshop Proceedings: Exhaled nitric oxide and nitric oxide oxidative metabolism in exhaled breath condensate: Executive summary. *Am J Respir Crit Care Med*, 2006. 173(7): p. 811-3.
208. **Barnes, P.J., et al.**, Exhaled nitric oxide in pulmonary diseases: a comprehensive review. *Chest*, 2010. 138(3): p. 682-92.
209. **Grob, N.M. and A.D. Raed**, Exhaled nitric oxide in asthma: progress since the introduction of standardized methodology. *Journal of Breath Research*, 2008. 2(3): p. 037002.
210. **Merx, H., K.E. Dreinhöfer, and K.P. Günther**, Sozialmedizinische Bedeutung der Arthrose in Deutschland Socioeconomic Relevance of Osteoarthritis in Germany. *Z Orthop Ihre Grenzgeb*, 2007. 145: p. 421-429.
211. **Thiem, U., et al.**, Prevalence of musculoskeletal complaints and self-reported joint osteoarthritis in the population of Herne: a telephone survey. *Zeitschrift für Rheumatologie*, 2008. 67(5): p. 432.
212. **Schneider, S., et al.**, Prevalence and correlates of osteoarthritis in Germany Representative data from the First National Health Survey. . *Der Orthopade*, 2005. 34(8): p. 782-90.
213. **Valdes, A.M. and T.D. Spector**, The genetic epidemiology of osteoarthritis. *Current opinion in rheumatology*, 2010. 22(2): p. 139.
214. **Englund, M., et al.**, Prevalence and incidence of rheumatoid arthritis in southern Sweden 2008 and their relation to prescribed biologics. *Rheumatology*, 2010.
215. **Symmons, D., et al.**, The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. *Rheumatology*, 2002. 41(7): p. 793.
216. **Woolf, A.D. and B. Pfleger**, Burden of major musculoskeletal conditions. *Bulletin of the World Health Organization*, 2003. 81: p. 646-656.
217. **Zink, A., D. Huscher, and M. Schneider**, How closely does rheumatology treatment follow the guidelines?: Ambition and reality. *Zeitschrift für Rheumatologie*, 2010.
218. **Glaeske, G., C. Schick Tanz, and K. Janhsen**, GEK-Arzneimittel-Report 2008. *Asgard-Verlag, St. Augustin*, 2008.
219. **Kramer, H.R. and J.T. Giles**, Cardiovascular disease risk in rheumatoid arthritis: Progress, debate, and opportunity. *Arthritis Care & Research*, 2010.
220. **Scott, D., F. Wolfe, and T.W. Huizinga**, Rheumatoid arthritis. *Lancet*, 2010. 376(9746): p. 1094-108.
221. **Solomon, D.H., et al.**, Patterns of cardiovascular risk in rheumatoid arthritis. *Annals of the rheumatic diseases*, 2006. 65(12): p. 1608.
222. **Nöthen, M. and K. Böhm**, Krankheitskosten in Deutschland: Welchen Preis hat die Gesundheit im Alter?, in Beiträge zur Gesundheitsberichterstattung des Bundes: Gesundheit und Krankheit im Alter. 2009, *Robert-Koch-Institut*: Berlin.
223. **Picavet, H.S.J. and J. Schouten**, Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC3-study. *Pain*, 2003. 102(1-2): p. 167-178.
224. **Breivik, H., et al.**, Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *European Journal of Pain*, 2006. 10(4): p. 287-333.

225. **Schmidt, C.O., et al.**, Back pain in the German adult population: prevalence, severity, and sociodemographic correlates in a multiregional survey. *Spine*, 2007. 32(18): p. 2005.
226. **Häuser, W., et al.**, Prevalence and predictors of pain in several body regions. Results of a representative German population survey. *Schmerz (Berlin, Germany)*, 2009. 23(5): p. 461.
227. **Wenig, C.M., et al.**, Costs of back pain in Germany. *European Journal of Pain*, 2009. 13(3): p. 280-286.
228. **Schmidt, C.O. and S.E. Baumeister**, Simple patterns behind complex spatial pain reporting? Assessing a classification of multisite pain reporting in the general population. *Pain*, 2007. 133(1-3): p. 174-182.
229. **Kanis, J.A., et al.**, European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporos Int*, 2008. 19(4): p. 399-428.
230. **Eastell, R., C. Newman, and D.C. Crossman**, Cardiovascular Disease and Bone. *Archives of Biochemistry and Biophysics*, 2010.
231. **Harvey, N., E. Dennison, and C. Cooper**, Osteoporosis: impact on health and economics. *Nature Reviews Rheumatology*, 2010. 6(2): p. 99-105.
232. **Khazai, N.B., G.R. Beck Jr, and G.E. Umpierrez**, Diabetes and fractures: an overshadowed association. *Current Opinion in Endocrinology, Diabetes and Obesity*, 2009. 16(6): p. 435.
233. **DVO**, DVO-Leitlinie 2009 zur Prophylaxe, Diagnostik und Therapie der Osteoporose bei Erwachsenen. 2009, *Dachverband Osteologie e. V.*: Essen.
234. **Duncan, E.L. and M.A. Brown**, Genetic Determinants of Bone Density and Fracture Risk--State of the Art and Future Directions. *J Clin Endocrinol Metab*, 2010. 95(6): p. 2576-2587.
235. **Richards, J.B., et al.**, Changes to osteoporosis prevalence according to method of risk assessment. *Journal of Bone and Mineral Research*, 2007. 22(2): p. 228-234.
236. **Blair, S.N. and J.N. Morris**, Healthy hearts--and the universal benefits of being physically active: physical activity and health. *Annals of epidemiology*, 2009. 19(4): p. 253.
237. **International Agency for Research on Cancer**, Weight Control and Physical Activity. 2002, Lyon: *World Health Organization*.
238. **Neilson, H.K., et al.**, Estimating activity energy expenditure: how valid are physical activity questionnaires? *American Journal of Clinical Nutrition*, 2008. 87(2): p. 279.
239. **Schatzkin, A., et al.**, Observational epidemiologic studies of nutrition and cancer: the next generation (with better observation). *Cancer Epidemiology Biomarkers & Prevention*, 2009. 18(4): p. 1026.
240. **Westerterp, K.R.**, Assessment of physical activity: a critical appraisal. *European journal of applied physiology*, 2009. 105(6): p. 823-828.
241. **Jain, S.H., et al.**, Cross-Sectional Associations Between Abdominal and Thoracic Adipose Tissue Compartments and Adiponectin and Resistin in the Framingham Heart Study. *Diabetes care*, 2009. 32(5): p. 903.

242. **Bonora, E., et al.**, Is it possible to derive a reliable estimate of human visceral and subcutaneous abdominal adipose tissue from simple anthropometric measurements? 1. *Metabolism*, 1995. 44(12): p. 1617-1625.
243. **Ludescher, B., et al.**, Correlation of fat distribution in whole body MRI with generally used anthropometric data. *Investigative radiology*, 2009. 44(11): p. 712.
244. **Stenholm, S., et al.**, Sarcopenic obesity: definition, cause and consequences. *Current Opinion in Clinical Nutrition & Metabolic Care*, 2008. 11(6): p. 693.
245. **Zamboni, M., et al.**, Sarcopenic obesity: a new category of obesity in the elderly. *Nutrition, Metabolism and Cardiovascular Diseases*, 2008. 18(5): p. 388-395.
246. **Blair, S., et al.**, Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. *Jama*, 1996. 276(3): p. 205.
247. **Kodama, S., et al.**, Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. *Jama*, 2009. 301(19): p. 2024.
248. **Ruiz, J.R., et al.**, Muscular strength and adiposity as predictors of adulthood cancer mortality in men. *Cancer Epidemiology Biomarkers & Prevention*, 2009. 18(5): p. 1468.
249. **Ruiz, J.R., et al.**, Association between muscular strength and mortality in men: prospective cohort study. *British Medical Journal*, 2008. 337(jul01 2): p. a439.
250. **Friedenreich, C.M. and M.R. Orenstein**, Physical activity and cancer prevention: etiologic evidence and biological mechanisms. *Journal of Nutrition*, 2002. 132(11): p. 3456S.
251. **Pischon, T.**, Use of obesity biomarkers in cardiovascular epidemiology. *Disease Markers*, 2009. 26(5): p. 247-263.
252. **Rundle, A.**, Molecular epidemiology of physical activity and cancer. *Cancer Epidemiology Biomarkers & Prevention*, 2005. 14(1): p. 227.
253. **Farrell, S., et al.**, Cardiorespiratory Fitness, Different Measures of Adiposity, and Cancer Mortality in Men. *Obesity*, 2007. 15(12): p. 3140-3149.
254. **De Caterina, R., et al.**, Nutritional mechanisms that influence cardiovascular disease. *American Journal of Clinical Nutrition*, 2006. 83(2): p. 421S.
255. **Thompson, P.D., et al.**, Exercise and Acute Cardiovascular Events. Placing the Risks Into Perspective. A Scientific Statement From the American Heart Association Council on Nutrition, Physical Activity, and Metabolism and the Council on Clinical Cardiology. In Collaboration With the American College of Sports Medicine. *Circulation*, 2007.
256. **Hu, F.B.**, Plant-based foods and prevention of cardiovascular disease: an overview. *American Journal of Clinical Nutrition*, 2003. 78(3): p. 544S.
257. **Kromhout, D., et al.**, Prevention of coronary heart disease by diet and lifestyle: evidence from prospective cross-cultural, cohort, and intervention studies. *Circulation*, 2002. 105(7): p. 893.
258. **Zarraga, I.G.E. and E.R. Schwarz**, Impact of dietary patterns and interventions on cardiovascular health. *Circulation*, 2006. 114(9): p. 961.
259. **Jenab, M., et al.**, Plasma and dietary vitamin C levels and risk of gastric cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST). *Carcinogenesis*, 2006. 27(11): p. 2250.

References

260. **Eussen, S., et al.**, Plasma folate, related genetic variants, and colorectal cancer risk in EPIC. *Cancer Epidemiology Biomarkers & Prevention*, 2010. 19(5): p. 1328.
 261. **Johansson, M., et al.**, Serum B vitamin levels and risk of lung cancer. *Jama*, 2010. 303(23): p. 2377.
 262. **Key, T.J., et al.**, Plasma carotenoids, retinol, and tocopherols and the risk of prostate cancer in the European Prospective Investigation into Cancer and Nutrition study. *American Journal of Clinical Nutrition*, 2007. 86(3): p. 672.
 263. **Bingham, S.A. and N. Day**, Commentary: Fat and breast cancer: time to re-evaluate both methods and results? *International journal of epidemiology*, 2006. 35(4): p. 1022.
 264. **Bingham, S.A., et al.**, Are imprecise methods obscuring a relation between fat and breast cancer? *The Lancet*, 2003. 362(9379): p. 212-214.
 265. **Freedman, L.S., et al.**, A comparison of two dietary instruments for evaluating the fat-breast cancer relationship. *International journal of epidemiology*, 2006. 35(4): p. 1011.
 266. **Schatzkin, A., et al.**, A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) study. *International journal of epidemiology*, 2003. 32(6): p. 1054.
 267. **International Agency for Research on Cancer**, IARC Alcohol Drinking, in IARC Monographs. 1988: Lyon, France
 268. **International Agency for Research on Cancer**, Tobacco smoke and involuntary smoking, in IARC Monographs. 2003, IARC: Lyon, France.
 269. **Kuper, H., M. Marmot, and H. Hemingway**. Systematic review of prospective cohort studies of psychosocial factors in the etiology and prognosis of coronary heart disease. 2002.
 270. **Strike, P.C. and A. Steptoe**, Psychosocial factors in the development of coronary artery disease. *Progress in cardiovascular diseases*, 2004. 46(4): p. 337-347.
 271. **Albus, C., et al.**, Psychosocial factors in coronary heart disease--scientific evidence and recommendations for clinical practice. *Gesundheitswesen (Bundesverband der Ärzte des Öffentlichen Gesundheitsdienstes (Germany))*, 2005. 67(1): p. 1.
 272. **Agardh, E.E., et al.**, Work stress and low sense of coherence is associated with type 2 diabetes in middle-aged Swedish women. *Diabetes care*, 2003. 26(3): p. 719.
 273. **Norberg, M., et al.**, Work stress and low emotional support is associated with increased risk of future type 2 diabetes in women. *Diabetes research and clinical practice*, 2007. 76(3): p. 368-377.
 274. **Heraclides, A., et al.**, Psychosocial Stress at Work Doubles the Risk of Type 2 Diabetes in Middle-Aged Women. *Diabetes care*, 2009. 32(12): p. 2230.
 275. **Brunner, E.J., T. Chandola, and M.G. Marmot**, Prospective effect of job strain on general and central obesity in the Whitehall II Study. *American journal of Epidemiology*, 2007. 165(7): p. 828.
 276. **Bonde, J.P.E.**, Psychosocial factors at work and risk of depression: a systematic review of the epidemiological evidence. *Occupational and environmental medicine*, 2008. 65(7): p. 438.
-

277. **Koster, A., et al.**, Socioeconomic differences in incident depression in older adults: the role of psychosocial factors, physical health status, and behavioral factors. *Journal of psychosomatic research*, 2006. 61(5): p. 619-627.
278. **Gute, G. and E.M. Eshbaugh**, Personality as a predictor of hooking up among college students. *Journal of Community Health Nursing*, 2008. 25(1): p. 26-43.
279. **Hampson, S.E., et al.**, Forty years on: Teachers' assessments of children's personality traits predict self-reported health behaviors and outcomes at midlife. *Health psychology: official journal of the Division of Health Psychology, American Psychological Association*, 2006. 25(1): p. 57.
280. **Hong, R.Y. and S.V. Paunonen**, Personality traits and health-risk behaviours in university students. *European Journal of Personality*, 2009. 23(8): p. 675-696.
281. **Raynor, D.A. and H. Levine**, Associations between the five-factor model of personality and health behaviors among college students. *Journal of American College Health*, 2009. 58(1): p. 73-82.
282. **Schmitt, D.P.**, The Big Five related to risky sexual behaviour across 10 world regions: Differential personality associations of sexual promiscuity and relationship infidelity. *European Journal of Personality*, 2004. 18(4): p. 301-319.
283. **Gallo, L.C., et al.**, Job characteristics, occupational status, and ambulatory cardiovascular activity in women. *Annals of Behavioral Medicine*, 2004. 28(1): p. 62-73.
284. **Collins, S.M., R.A. Karasek, and K. Costas**, Job strain and autonomic indices of cardiovascular disease risk. *American journal of industrial medicine*, 2005. 48(3): p. 182-193.
285. **Riese, H., et al.**, Job strain in relation to ambulatory blood pressure, heart rate, and heart rate variability among female nurses. *Scandinavian journal of work, environment & health*, 2004. 30(6): p. 477-485.
286. **Hintsanen, M., et al.**, Effort—reward imbalance, heart rate, and heart rate variability: the cardiovascular risk in young finns study. *International Journal of Behavioral Medicine*, 2007. 14(4): p. 202-212.
287. **Garszen, B.**, Psycho-oncology and cancer: linking psychosocial factors with cancer development. *Annals of Oncology*, 2002. 13(Supplement 4): p. 171.
288. **Schwarz, S., H. Messerschmidt, and M. DÄ ren**, Psychosocial risk factors for cancer development. *Medizinische Klinik (Munich, Germany: 1983)*, 2007. 102(12): p. 967.
289. **Schraub, S., H. Sancho-Garnier, and M. Velten**, Should psychological events be considered cancer risk factors? *Revue d'Épidémiologie et de Santé Publique*, 2009. 57(2): p. e7-e16.
290. **Carver, C.S. and M.F. Scheier**, Perspectives on personality. 4 ed. 2000, Boston: Allyn & Bacon.
291. **Soldz, S. and G.E. Vaillant**, The Big Five Personality Traits and the Life Course: A 45-Year Longitudinal Study* 1,* 2. *Journal of Research in Personality*, 1999. 33(2): p. 208-232.
292. **Jang, K.L., et al.**, Heritability of Facet-Level Traits in a Cross-Cultural Twin Sample: Support for a Hierarchical Model of Personality* 1. *Journal of Personality and Social Psychology*, 1998. 74(6): p. 1556-1565.
293. **Loehlin, J.C., et al.**, Heritabilities of common and measure-specific components of the Big Five personality factors. *Journal of Research in Personality*, 1998. 32: p. 431-453.

294. **McCrae, R.R. and P.T. Costa Jr**, Personality trait structure as a human universal. *American Psychologist*, 1997. 52(5): p. 509-516.
295. **Buss, D.M.**, Social adaptation and five major factors of personality. *The five factor model of personality: Theoretical perspectives*, 1996: p. 180-207.
296. **Chapman, B.P., et al.**, Can the Influence of Childhood Socioeconomic Status on Men's and Women's Adult Body Mass Be Explained by Adult Socioeconomic Status or Personality? Findings From a National Sample. *Health Psychology*, 2009. 28(4): p. 419-427.
297. **Lemos-Giráldez, S. and A.M. Fidalgo-Aliste**, Personality dispositions and health-related habits and attitudes: A cross-sectional study. *European Journal of Personality*, 1997. 11(3): p. 197-209.
298. **Siegrist, J., et al.**, The measurement of effort-reward imbalance at work: European comparisons. *Social Science & Medicine*, 2004. 58(8): p. 1483-1499.
299. **Karasek, R. and T. Theorell**, Healthy work. 1990: *Basic Books New York*.
300. **Salavecz, G., et al.**, Work stress and health in Western European and post-communist countries: an East–West comparison study. *British Medical Journal*, 2010. 64(01): p. 57.
301. **Chandola, T., J. Siegrist, and M. Marmot**, Do changes in effort-reward imbalance at work contribute to an explanation of the social gradient in angina? *Occupational and environmental medicine*, 2005. 62(4): p. 223.
302. **Siegrist, J. and N. Dragano**, Psychosoziale Belastungen und Erkrankungsrisiken im Erwerbsleben. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, 2008. 51(3): p. 305-312.
303. **Kivimäki, M., et al.**, Work stress and risk of cardiovascular mortality: prospective cohort study of industrial employees. *Bmj*, 2002. 325(7369): p. 857.
304. **Kuper, H., et al.**, Psychosocial determinants of coronary heart disease in middle-aged women: a prospective study in Sweden. *American journal of Epidemiology*, 2006. 164(4): p. 349.
305. **Rosengren, A., et al.**, Association of psychosocial risk factors with risk of acute myocardial infarction in 11 119 cases and 13 648 controls from 52 countries (the INTERHEART study): case-control study. *The Lancet*, 2004. 364(9438): p. 953-962.
306. **Eller, N.H., et al.**, Work-related psychosocial factors and the development of ischemic heart disease: a systematic review. *Cardiology in review*, 2009. 17(2): p. 83.
307. **Marmot, M.G., et al.**, Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *The Lancet*, 1997. 350(9073): p. 235-239.
308. **Cohen, S.**, Social Relationships and Health. *American Psychologist*, 2004. 59(8): p. 676-684.
309. **Berkman, L.**, Social Networks and Social Support, in Encyclopedia of Public Health, L. Breslow, Editor. 2002, *Macmillan References*: Gale. p. 1128-1130.
310. **Berkman, L.F. and S.L. Syme**, Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *American journal of Epidemiology*, 1979. 109(2): p. 186.

311. **Weyers, S., et al.**, Low socio-economic position is associated with poor social networks and social support: results from the Heinz Nixdorf Recall Study. *International Journal for Equity in Health*, 2008. 7(1): p. 13.
312. **Meaney, M.J.**, Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience*, 2001. 24: p. 1161-1192.
313. **Mooy, J.M., et al.**, Major stressful life events in relation to prevalence of undetected type 2 diabetes: the Hoorn Study. *Diabetes care*, 2000. 23(2): p. 197.
314. **Vogelzangs, N., et al.**, Psychosocial risk factors and the metabolic syndrome in elderly persons: findings from the Health, Aging and Body Composition study. *The Journals of Gerontology: Series A*, 2007. 62(5): p. 563.
315. **Räikkönen, K., K.A. Matthews, and L.H. Kuller**, Depressive symptoms and stressful life events predict metabolic syndrome among middle-aged women. *Diabetes care*, 2007. 30(4): p. 872.
316. **Pyykkönen, A.J., et al.**, Stressful Life Events and the Metabolic Syndrome. *Diabetes care*, 2010. 33(2): p. 378.
317. **Kumari, M., J. Head, and M. Marmot**, Prospective study of social and other risk factors for incidence of type 2 diabetes in the Whitehall II study. *Archives of Internal Medicine*, 2004. 164(17): p. 1873.
318. **Singh-Manoux, A., M.G. Marmot, and N.E. Adler**, Does subjective social status predict health and change in health status better than objective status? *Psychosomatic Medicine*, 2005. 67(6): p. 855.
319. **Kaplan, G.A. and J.E. Keil**, Socioeconomic factors and cardiovascular disease: a review of the literature. *Circulation*, 1993. 88(4): p. 1973.
320. **Rosvall, M., et al.**, Contribution of main causes of death to social inequalities in mortality in the whole population of Scania, Sweden. *BMC Public Health*, 2006. 6(1): p. 79.
321. **Whitehead, M. and G. Dahlgren**, Concepts and principles for tackling social inequities in health: Levelling up Part 1. *Copenhagen: WHO*, 2007.
322. **Herd, P., B. Goesling, and J.S. House**, Socioeconomic position and health: the differential effects of education versus income on the onset versus progression of health problems. *Journal of Health and Social Behavior*, 2007. 48(3): p. 223.
323. **Brown, A.F., et al.**, Socioeconomic position and health among persons with diabetes mellitus: a conceptual framework and review of the literature. *Epidemiologic reviews*, 2004. 26(1): p. 63.
324. **Tamayo, T., C. Herder, and W. Rathmann**, Impact of early psychosocial factors (childhood socioeconomic factors and adversities) on future risk of type 2 diabetes, metabolic disturbances and obesity: a systematic review. *BMC Public Health*, 2010. 10(1): p. 525.
325. **Diez-Roux, A.V., et al.**, Social Inequalities and Atherosclerosis The Atherosclerosis Risk in Communities Study. *American journal of Epidemiology*, 1995. 141(10): p. 960.
326. **Diez-Roux, A.V., et al.**, Neighborhood environments and coronary heart disease: a multilevel analysis. *American journal of Epidemiology*, 1997. 146(1): p. 48.

327. **Lynch, J.W., G.A. Kaplan, and J.T. Salonen**, Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse. *Social Science & Medicine*, 1997. 44(6): p. 809-819.
328. **Kunst, A.E., et al.**, Trends in socioeconomic inequalities in self-assessed health in 10 European countries. *International journal of epidemiology*, 2005. 34(2): p. 295.
329. **Huisman, M., et al.**, Educational inequalities in cause-specific mortality in middle-aged and older men and women in eight western European populations. *The Lancet*, 2005. 365(9458): p. 493-500.
330. **Mackenbach, J.P., et al.**, Health inequalities: Europe in profile. 2006: *Produced by COI for the Dept. of Health*.
331. **Dalstra, J.A.A., et al.**, Socioeconomic differences in the prevalence of common chronic diseases: an overview of eight European countries. *International journal of epidemiology*, 2005. 34(2): p. 316.
332. **Nocon, M., T. Keil, and S.N. Willich**, Education, income, occupational status and health risk behaviour. *Journal of Public Health*, 2007. 15(5): p. 401-405.
333. **Menvielle, G., et al.**, Educational differences in cancer mortality among women and men: a gender pattern that differs across Europe. *British journal of cancer*, 2008. 98(5): p. 1012-1019.
334. **Wang, J.L., N. Schmitz, and C.S. Dewa**, Socioeconomic status and the risk of major depression: the Canadian National Population Health Survey. *Journal of Epidemiology and Community Health*, 2009. 64(5): p. 447.
335. **de Lau, L.M.L. and M.M.B. Breteler**, Epidemiology of Parkinson's disease. *The Lancet Neurology*, 2006. 5(6): p. 525-535.
336. **Wilkinson, R.G. and M.G. Marmot**, Social determinants of health: the solid facts. 2003: *World Health Organization*.
337. **Bartley, M.**, Health inequality: an introduction to theories, concepts, and methods. 2004: *Polity*.
338. **Skalicka, V., et al.**, Material, psychosocial, behavioural and biomedical factors in the explanation of relative socio-economic inequalities in mortality: evidence from the HUNT study. *International journal of epidemiology*, 2009.
339. **Lynch, H.T. and J.F. Lynch**, Breast cancer genetics: family history, heterogeneity, molecular genetic diagnosis, and genetic counseling. *Current Problems in Cancer*, 1996. 20(6): p. 329-365.
340. **Lynch, J., et al.**, Socioeconomic status and carotid atherosclerosis. *Circulation*, 1995. 92(7): p. 1786.
341. **van Oort, F.V.A., F.J. van Lenthe, and J.P. Mackenbach**, Material, psychosocial, and behavioural factors in the explanation of educational inequalities in mortality in the Netherlands. *Journal of Epidemiology and Community Health*, 2005. 59(3): p. 214.
342. **Rask, K., E. O'Malley, and B. Druss**, Impact of socioeconomic, behavioral and clinical risk factors on mortality. *Journal of Public Health*, 2009. 31(2): p. 231.
343. **Rathmann, W., et al.**, Is inflammation a causal chain between low socioeconomic status and type 2 diabetes? Results from the KORA Survey 2000. *European journal of epidemiology*, 2006. 21(1): p. 55-60.

344. **Loucks, E.B., et al.**, Life course socioeconomic position is associated with inflammatory markers: The Framingham Offspring Study. *Social Science & Medicine*, 2010.
345. **Chichlowska, K.L., et al.**, Life Course Socioeconomic Conditions and Metabolic Syndrome in Adults: The Atherosclerosis Risk in Communities (ARIC) Study. *Annals of epidemiology*, 2009. 19(12): p. 875-883.
346. **Lemelin, E.T., et al.**, Life-course socioeconomic positions and subclinical atherosclerosis in the Multi-Ethnic Study of Atherosclerosis. *Social Science & Medicine*, 2009. 68(3): p. 444-451.
347. **Kim, D., et al.**, Do Neighborhood Socioeconomic Deprivation and Low Social Cohesion Predict Coronary Calcification?: The CARDIA Study. *American journal of Epidemiology*, 2010.
348. **Galobardes, B., J. Lynch, and G. Smith**, Is the association between childhood socioeconomic circumstances and cause-specific mortality established? Update of a systematic review. *British Medical Journal*, 2008. 62(5): p. 387.
349. **Pollitt, R.A., et al.**, Cumulative life course and adult socioeconomic status and markers of inflammation in adulthood. *British Medical Journal*, 2008. 62(6): p. 484.
350. **Galobardes, B., J. Lynch, and G. Davey Smith**, Measuring socioeconomic position in health research. *British medical bulletin*, 2007.
351. **Diez Roux, A.V. and C. Mair**, Neighborhoods and health. *Annals of the New York Academy of Sciences*, 2010. 1186(1): p. 125-145.
352. **Borrell, L.N., et al.**, Neighbourhood characteristics and mortality in the Atherosclerosis Risk in Communities Study. *International journal of epidemiology*, 2004. 33(2): p. 398.
353. **van Lenthe, F.J., et al.**, Neighbourhood unemployment and all cause mortality: a comparison of six countries. *Journal of Epidemiology and Community Health*, 2005. 59(3): p. 231.
354. **Lisabeth, L.D., et al.**, Neighborhood environment and risk of ischemic stroke. *American journal of Epidemiology*, 2007. 165(3): p. 279.
355. **Mujahid, M.S., et al.**, Neighborhood characteristics and hypertension. *Epidemiology*, 2008. 19(4): p. 590.
356. **Nordstrom, C.K., et al.**, The association of personal and neighborhood socioeconomic indicators with subclinical cardiovascular disease in an elderly cohort. The cardiovascular health study* 1. *Social Science & Medicine*, 2004. 59(10): p. 2139-2147.
357. **Murray, E.T., et al.**, Trajectories of Neighborhood Poverty and Associations With Subclinical Atherosclerosis and Associated Risk Factors: The Multi-Ethnic Study of Atherosclerosis. *American journal of Epidemiology*, 2010. 171(10): p. 1099.
358. **Auchincloss, A.H., et al.**, Neighborhood resources for physical activity and healthy foods and incidence of type 2 diabetes mellitus: the Multi-Ethnic Study of Atherosclerosis. *Archives of Internal Medicine*, 2009. 169(18): p. 1698.
359. **Henderson, C., et al.**, Neighbourhood characteristics, individual level socioeconomic factors, and depressive symptoms in young adults: the CARDIA study. *Journal of Epidemiology and Community Health*, 2005. 59(4): p. 322.

References

360. **Mair, C., A.V. Roux, and S. Galea**, Are neighbourhood characteristics associated with depressive symptoms? A review of evidence. *Journal of Epidemiology and Community Health*, 2008. 62(11): p. 940.
361. **Auchincloss, A.H., et al.**, Neighborhood resources for physical activity and healthy foods and their association with insulin resistance. *Epidemiology*, 2008. 19(1): p. 146.
362. **Franco, M., et al.**, Neighborhood characteristics and availability of healthy foods in Baltimore. 2008.
363. **Moore, L.V., et al.**, Fast-Food Consumption, Diet Quality, and Neighborhood Exposure to Fast Food. *American journal of Epidemiology*, 2009. 170(1): p. 29.
364. **Chaix, B., et al.**, Disentangling contextual effects on cause-specific mortality in a longitudinal 23-year follow-up study: impact of population density or socioeconomic environment? *International journal of epidemiology*, 2006. 35(3): p. 633.
365. **Meisinger, C., et al.**, Regional disparities of hypertension prevalence and management within Germany. *Journal of hypertension*, 2006. 24(2): p. 293.
366. **Moebus, S., et al.**, Regional Differences in the Prevalence of the Metabolic Syndrome in Primary Care Practices in Germany. *Deutsches Arzteblatt International*, 2008. 105(12): p. 207.
367. **Stang, J. and C. Bayerl**, Position of the American Dietetic Association: child and adolescent nutrition assistance programs. *Journal of the American Dietetic Association*, 2010. 110(5): p. 791.
368. **Garvey, J., C. Taylor, and W. McNicholas**, Cardiovascular disease in obstructive sleep apnoea syndrome: the role of intermittent hypoxia and inflammation. *European Respiratory Journal*, 2009. 33(5): p. 1195.
369. **Malhotra, A. and J. Loscalzo**, Sleep and cardiovascular disease: an overview. *Progress in cardiovascular diseases*, 2009. 51(4): p. 279.
370. **Cardona Portela, P., et al.**, Sleep-disordered breathing and acute stroke. *Cerebrovascular Diseases*, 2009. 27(1): p. 104-110.
371. **Mullington, J., et al.**, Cardiovascular, inflammatory, and metabolic consequences of sleep deprivation. *Progress in cardiovascular diseases*, 2009. 51(4): p. 294-302.
372. **Erren, T. and R. Reiter**, Defining chronodisruption. *Journal of pineal research*, 2009. 46(3): p. 245-247.
373. **Straif, K., et al.**, Carcinogenicity of shift-work, painting, and fire-fighting. *The lancet oncology*, 2007. 8(12): p. 1065.
374. **Schernhammer, E., et al.**, Urinary 6-sulfatoxymelatonin levels and risk of breast cancer in postmenopausal women. *Journal of the National Cancer Institute*, 2008. 100(12): p. 898.
375. **Klug, S., et al.**, Prevalence of human papillomavirus types in women screened by cytology in Germany. *Journal of medical virology*, 2007. 79(5): p. 616-625.
376. **der Ständigen Impfkommision, E.**, am Robert Koch-Institut/Stand Juli 2007. *Epidemiologisches Bulletin*, 2007. 50: p. 267-286.
377. **Offergeld, R., et al.**, Infektionsepidemiologische Daten von Blutspendern in Deutschland 2007. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, 2010: p. 1-9.

378. **Sauerbrei, A., et al.**, Antikörper gegen impfpräventable Erkrankungen bei Schwangeren und deren Neugeborenen. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, 2004. 47(1): p. 10-15.
379. **Voigt, K., et al.**, Aktueller Impfstatus in unterschiedlichen medizinischen Berufsgruppen in Brandenburg, Sachsen und Sachsen-Anhalt= Current Vaccination Status of Health-Care Personnel in Brandenburg, Saxony and Saxony-Anhalt. *Das Gesundheitswesen*, 2008. 70(7): p. 408-414.
380. **Forde, J. and B. Meeker**, A model of varicella-zoster reactivation. *Mathematical Biosciences and Engineering (MBE)*, 2010. 7(4): p. 765-777.
381. **Reuss, A., et al.**, Sind Daten der Kassenärztlichen Vereinigungen für die Surveillance von Impfungen und impfpräventablen Erkrankungen geeignet. *Gesundheitswesen*, 2010. 72(6): p. 340-6.
382. **Marcucci, F. and A. Mele**, Hepatitis viruses and non-Hodgkin lymphoma: epidemiology, mechanisms of tumorigenesis and therapeutic opportunities. *Blood*, 2010.
383. **Roberts, E.T., et al.**, Cytomegalovirus antibody levels, inflammation, and mortality among elderly Latinos over 9 years of follow-up. *American journal of Epidemiology*, 2010. 172(4): p. 363.
384. **Wang, G.C., et al.**, Cytomegalovirus infection and the risk of mortality and frailty in older women: a prospective observational cohort study. *American journal of Epidemiology*, 2010. 171(10): p. 1144.
385. **Blankenberg, S., et al.**, Cytomegalovirus infection with interleukin-6 response predicts cardiac mortality in patients with coronary artery disease. *Circulation*, 2001. 103(24): p. 2915.
386. **Hecker, M., et al.**, Continuous cytomegalovirus seroconversion in a large group of healthy blood donors. *Vox sanguinis*, 2004. 86(1): p. 41-44.
387. **Honjo, K., R. van Reekum, and N. Verhoeff**, Alzheimer's disease and infection: Do infectious agents contribute to progression of Alzheimer's disease? *Alzheimer's and Dementia*, 2009. 5(4): p. 348-360.
388. **Gross, P.A., et al.**, The efficacy of influenza vaccine in elderly persons. *Annals of Internal Medicine*, 1995. 123(7): p. 518.
389. **Wikby, A., et al.**, Changes in CD8 and CD4 lymphocyte subsets, T cell proliferation responses and non-survival in the very old: the Swedish longitudinal OCTO-immune study. *Mechanisms of ageing and development*, 1998. 102(2-3): p. 187-198.
390. **Couch, R.B., et al.**, Influenza: its control in persons and populations. *The Journal of Infectious Diseases*, 1986. 153(3): p. 431-440.
391. **Govaert, T., et al.**, Immune response to influenza vaccination of elderly people. A randomized double-blind placebo-controlled trial. *Vaccine*, 1994. 12(13): p. 1185-1189.
392. **Castle, S.**, Clinical relevance of age-related immune dysfunction. *Clinical Infectious Diseases*, 2000. 31(2): p. 578-585.
393. **Pawelec, G., et al.**, Cytomegalovirus and human immunosenescence. *Reviews in Medical Virology*, 2009. 19(1): p. 47-56.
394. **Pawelec, G., et al.**, Human Immunosenescence. *Annals of the New York Academy of Sciences*, 2006. 1067(Understanding and Modulating Aging): p. 56-65.

395. **Lesourd, B.**, Nutrition: a major factor influencing immunity in the elderly. *The journal of nutrition, health & aging*, 2004. 8(1): p. 28-37.
396. **Dhabhar, F.**, A hassle a day may keep the pathogens away: The fight-or-flight stress response and the augmentation of immune function. *Integrative and Comparative Biology*, 2009. 49(3): p. 215.
397. **Shanahan, F.**, The colonic microflora and probiotic therapy in health and disease. *Curr Opin Gastroenterol.*, 2010. Epub ahead of print.
398. **Salgado, C. and B. Farr**, Outcomes associated with vancomycin-resistant enterococci: a meta-analysis. *Infection control and hospital epidemiology*, 2003. 24(9): p. 690-698.
399. **Hanage, W., et al.**, Evidence that pneumococcal serotype replacement in Massachusetts following conjugate vaccination is now complete. *Epidemics*, 2010. 2(2): p. 80-84.
400. **Tsai, F. and W.J. Coyle**, The microbiome and obesity: Is obesity linked to our gut flora? *Current gastroenterology reports*, 2009. 11(4): p. 307-313.
401. **IVH** (2009) Der deutsche Heimtiermarkt. Struktur & Umsatzdaten.
402. **Gilchrist, J., K. Gotsch, and G. Ryan**, Non-fatal dog bite-related injuries treated in hospital emergency departments—United States, 2001. *Morbidity and Mortality Weekly Report*, 2003. 52(26): p. 605-610.
403. **Kerkhof, M., et al.**, Effects of pets on asthma development up to 8 years of age: the PIAMA study. *Allergy*, 2009. 64(8): p. 1202-1208.
404. **Rabinowitz, P.M., Z. Gordon, and L. Odofin**, Pet-related infections. *American family physician*, 2007. 76(9): p. 1314.
405. **Werber, D., et al.**, Shiga Toxin-producing *Escherichia coli* Infection in Germany--Different Risk Factors for Different Age Groups. *American journal of Epidemiology*, 2007. 165(4): p. 425.
406. **Haimerl, M., et al.**, Seroprevalence of *Bartonella henselae* in cats in Germany. *Journal of medical microbiology*, 1999. 48(9): p. 849.
407. **Just, F.T., et al.**, Molecular evidence for *Bartonella* spp. in cat and dog fleas from Germany and France. *Zoonoses and Public Health*, 2008. 55(8 10): p. 514-520.
408. **Lamas, C.C., et al.**, *Bartonella* spp. infection in HIV positive individuals, their pets and ectoparasites in Rio de Janeiro, Brazil: Serological and molecular study. *Acta Tropica*, 2010. 115(1-2): p. 137-141.
409. **Pape, M., et al.**, Occurrence of *Bartonella henselae* and *Bartonella quintana* among Human Immunodeficiency Virus Infected Patients. *Annals of the New York Academy of Sciences*, 2005. 1063(1): p. 299-301.
410. **Morris, D.O., et al.**, The prevalence of carriage of methicillin-resistant staphylococci by veterinary dermatology practice staff and their respective pets. *Veterinary Dermatology*, 2010.
411. **Van Duijkeren, E., et al.**, Transmission of a Pantone-Valentine leucocidin-positive, methicillin-resistant *Staphylococcus aureus* strain between humans and a dog. *Journal of clinical microbiology*, 2005. 43(12): p. 6209.
412. **Kivimäki, M., et al.**, Work stress in the etiology of coronary heart disease: a meta-analysis. *Scandinavian journal of work, environment & health*, 2006. 32(6): p. 431-442.

413. **Netterstrøm, B., et al.**, The relation between work-related psychosocial factors and the development of depression. *Epidemiologic reviews*, 2008. 30(1): p. 118.
414. **Wu, Z. and R. Hart**, The effects of marital and nonmarital union transition on health. *Journal of Marriage and Family*, 2002. 64(2): p. 420-432.
415. **Siepkka, S.M., et al.**, Genetics and neurobiology of circadian clocks in mammals. *Clocks and rhythms*, 2007: p. 251.
416. **Rüger, M. and F. Scheer**, Effects of circadian disruption on the cardiometabolic system. *Reviews in Endocrine & Metabolic Disorders*, 2009. 10(4): p. 245-260.
417. **World Health Organization**, Air Quality Guidelines Global Update 2005. 2006, WHO Regional Office for Europe: Copenhagen.
418. **Jarup, L., et al.**, Hypertension and exposure to noise near airports: the HYENA study. *Environmental health perspectives*, 2008. 116(3): p. 329.
419. **Babisch, W., et al.**, Traffic noise and risk of myocardial infarction. *Epidemiology*, 2005. 16(1): p. 33.
420. **Van Kempen, E., et al.**, The association between noise exposure and blood pressure and ischemic heart disease: a meta-analysis. *Environmental health perspectives*, 2002. 110(3): p. 307.
421. **Krämer, U., et al.**, Traffic-related air pollution and incident type 2 diabetes: results from the SALIA Cohort Study. *Environmental health perspectives*, 2010. 118(9): p. 1273.
422. **Peters, A., et al.**, Translocation and potential neurological effects of fine and ultrafine particles a critical update. *Part Fibre Toxicol*, 2006. 3(13): p. 1743-8977.
423. **Menzler, S., et al.**, Population attributable fraction for lung cancer due to residential radon in Switzerland and Germany. *Health physics*, 2008. 95(2): p. 179.
424. **Nagel, G., et al.**, The impact of education on risk factors and the occurrence of multimorbidity in the EPIC-Heidelberg cohort. *BMC Public Health*, 2008. 8(1): p. 384.
425. **Rathmann, W., et al.**, Sex differences in the associations of socioeconomic status with undiagnosed diabetes mellitus and impaired glucose tolerance in the elderly population: the KORA Survey 2000. *The European Journal of Public Health*, 2005. 15(6): p. 627.
426. **Carson, A.P., et al.**, Cumulative socioeconomic status across the life course and subclinical atherosclerosis. *Annals of epidemiology*, 2007. 17(4): p. 296-303.
427. **Davey Smith, G., et al.**, Education and occupational social class: which is the more important indicator of mortality risk? *British Medical Journal*, 1998. 52(3): p. 153.
428. **Lawlor, D.A., S. Ebrahim, and G. Davey Smith**, Adverse socioeconomic position across the lifecourse increases coronary heart disease risk cumulatively: findings from the British women's heart and health study. *Journal of epidemiology and community health*, 2005. 59(9): p. 785.
429. **Lawlor, D.A., et al.**, Association of childhood socioeconomic position with cause-specific mortality in a prospective record linkage study of 1,839,384 individuals. *American journal of Epidemiology*, 2006. 164(9): p. 907.
430. **Robert Koch-Institut**, Selbstauffüll-Fragebögen im Rahmen von DEGS. *Epidemiologisches Bulletin*, 2009. 10: p. 89.

431. **Statistisches Bundesamt**, Methoden-Verfahren-Entwicklungen. Demographische Standards. Eine gemeinsame Empfehlung des Arbeitskreises Deutscher Markt und Sozialforschungsinstitute e. V. (ADM), der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V. (ASI) und des Statistischen Bundesamtes. 2004, *Statistisches Bundesamt*: Wiesbaden
432. **Ahrens, W., B.M. Bellach, and K.H. Jöckel**, Messung soziodemographischer Merkmale in der Epidemiologie. RKI-Schrift 1/98. 1998, München *MMV Medizin Verlag*.
433. **Linos, E., et al.**, Atopy and risk of brain tumors: a meta-analysis. *Journal of the National Cancer Institute*, 2007. 99(20): p. 1544.
434. **Wang, H. and T.L. Diepgen**, Atopic dermatitis and cancer risk. *British Journal of Dermatology*, 2006. 154(2): p. 205-210.
435. **Lloyd-Jones, D.M., et al.**, Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *Jama*, 2004. 291(18): p. 2204.
436. **Rathmann, W., et al.**, Incidence of Type 2 diabetes in the elderly German population and the effect of clinical and lifestyle risk factors: KORA S4/F4 cohort study. *Diabetic Medicine*, 2009. 26(12): p. 1212-1219.
437. **Wilson, P.W.F., et al.**, Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Archives of Internal Medicine*, 2007. 167(10): p. 1068.
438. **Coogan, P.F., J. Smith, and L. Rosenberg**, Statin use and risk of colorectal cancer. *Journal of the National Cancer Institute*, 2007. 99(1): p. 32.
439. **Friis, S., et al.**, Cancer risk among statin users: A population based cohort study. *International Journal of Cancer*, 2005. 114(4): p. 643-647.
440. **Hoffmeister, M., J. Chang Claude, and H. Brenner**, Individual and joint use of statins and low dose aspirin and risk of colorectal cancer: A population based case-control study. *International Journal of Cancer*, 2007. 121(6): p. 1325-1330.
441. **Poynter, J.N., et al.**, Statins and the risk of colorectal cancer. *New England Journal of Medicine*, 2005. 352(21): p. 2184.
442. **Shadman, M., et al.**, Non-steroidal anti-inflammatory drugs and statins in relation to colorectal cancer risk. *World Journal of Gastroenterology: WJG*, 2009. 15(19): p. 2336.
443. **Döring, A., et al.**, KORA-Survey 2000. Manual of Operation. Untersucher-Handbuch. 2000, Neuherberg: *GSF*
444. **Fricke, U., J. Günther, and A. Zawinell**, Anatomisch-therapeutisch-chemische Klassifikation mit Tagesdosen für den deutschen Arzneimittelmarkt. *Methodik der ATC-Klassifikation und DDD-festlegung. ATC-Index mit DDD-Angaben. GKV-Arzneimittelindex*, 2005. 4.
445. **Freedman, D.S., et al.**, The relation of menarcheal age to obesity in childhood and adulthood: the Bogalusa heart study. *BMC pediatrics*, 2003. 3(1): p. 3.
446. **Kivimaki, M., et al.**, Association of age at menarche with cardiovascular risk factors, vascular structure, and function in adulthood: the Cardiovascular Risk in Young Finns study. *American Journal of Clinical Nutrition*, 2008. 87(6): p. 1876.

447. **Lakshman, R., et al.**, Early age at menarche associated with cardiovascular disease and mortality. *Journal of Clinical Endocrinology & Metabolism*, 2009. 94(12): p. 4953.
448. **Ensrud, K.E., et al.**, Frequency and severity of hot flashes and sleep disturbance in postmenopausal women with hot flashes. *Menopause*, 2009. 16(2): p. 286.
449. **Joffe, H., et al.**, Vasomotor symptoms are associated with depression in perimenopausal women seeking primary care. *Menopause*, 2002. 9(6): p. 392.
450. **International Agency for Research on Cancer**, Combined Estrogen-Progestogen contraceptives and combined Estrogen-progestogen menopausal therapy, in IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. 2007, *IARC*: Lyon.
451. **Kemmeren, J.M., A. Algra, and D.E. Grobbee**, Third generation oral contraceptives and risk of venous thrombosis: meta-analysis. *Bmj*, 2001. 323(7305): p. 131.
452. **World Health Organization Collaborative Study of Cardiovascular Disease and Steroid Hormone Contraception**, Effect of different progestagens in low oestrogen oral contraceptives on venous thromboembolic disease. *Lancet*, 1995. 346: p. 1582-1588.
453. **Writing Group for the Women's Health Initiative Investigators**, Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *Jama*, 2002. 288(3): p. 321.
454. **Roehrborn, C.G.**, Male lower urinary tract symptoms (LUTS) and benign prostatic hyperplasia (BPH). *Med Clin North Am*, 2011. 95(1): p. 87-100.
455. **Verhamme, K.M.C., et al.**, Incidence and prevalence of lower urinary tract symptoms suggestive of benign prostatic hyperplasia in primary care--the Triumph project. *European urology*, 2002. 42(4): p. 323-328.
456. **Bundesagentur für gesundheitliche Aufklärung**, Die Drogenaffinität Jugendlicher in der Bundesrepublik Deutschland 2008. *Alkohol-, Tabak-und Cannabiskonsum. Erste Ergebnisse zu aktuellen Entwicklungen und Trends. Köln: BZgA*, 2008.
457. **Hall, W., L. Degenhardt, and M. Teesson**, Cannabis use and psychotic disorders: an update. *Drug and Alcohol Review*, 2004. 23(4): p. 433-443.
458. **Kawohl, W. and W. Rössler**, Cannabis and Schizophrenia: new findings in an old debate. *Neuropsychiatrie: Klinik, Diagnostik, Therapie und Rehabilitation: Organ der Gesellschaft Österreichischer Nervenärzte und Psychiater*, 2008. 22(4): p. 223.
459. **Latza, U., W. Hoffmann, and C. Terschuren**, Rauchen als möglicher Confounder in epidemiologischen Studien: Standardisierung der Erhebung, Quantifizierung und Analyse.[Smoking as potential confounder in German epidemiological studies: standardization of assessment, quantification, and analysis]. *Gesundheitswesen*, 2005. 67: p. 795-802.
460. **Bischof, G., et al.**, Development and evaluation of a screening instrument for alcohol-use disorders and at-risk drinking: the brief alcohol screening instrument for medical care (BASIC). *Journal of studies on alcohol and drugs*, 2007. 68(4): p. 607.
461. **Peasey, A., et al.**, Determinants of cardiovascular disease and other non-communicable diseases in Central and Eastern Europe: rationale and design of the HAPIEE study. *BMC Public Health*, 2006. 6(1): p. 255.

462. **Bundeszentrale für gesundheitliche Aufklärung (BZgA)**, Die Drogenaffinität Jugendlicher in der Bundesrepublik Deutschland 2008. Verbreitung des Konsums illegaler Drogen bei Jugendlichen und jungen Erwachsenen. Eine Wiederholungsbefragung der Bundeszentrale für gesundheitliche Aufklärung. 2010: Köln
463. **NHANES**. 2009 December 6, 2010]; Available from: http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm.
464. **LifeGene**. 2010 December 6, 2010]; Available from: <https://www.lifegene.se/For-scientists/Questionnaire-Overview/>
465. **Constances**. 2010 December 6, 2010]; Available from: http://www.constances.fr/index.php?option=com_rokdownloads&view=folder&Itemid=126&id=237%3AAnne-xes-du-protocole-de-la-phase-pilote&limitstart=10
466. **Radoschewski, M.**, HEALTH RELATED QUALITY OF LIFE-CONCEPTS AND MEASURES-THE STATE OF THE ART. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, 2009. 43(3): p. 165-189.
467. **Idler, E.L. and Y. Benyamini**, Self-rated health and mortality: a review of twenty-seven community studies. *Journal of health and social behavior*, 1997. 38(1): p. 21-37.
468. **Angermeyer, M.C., R. Kilian, and H. Matschinger**, WHOQOL-100 und WHOQOL-BREF: Handbuch für die deutschsprachige Version der WHO-Instrumente zur Erfassung von Lebensqualität. 2000: *Hogrefe & Huber*.
469. **Brooks, R.**, EuroQol: the current state of play* 1. *Health policy*, 1996. 37(1): p. 53-72.
470. **Szende, A., M. Oppe, and N. Devlin**, EQ-5D value sets: inventory, comparative review and user guide series: EuroQol Group Monographs. Vol. 2. 2007: *Springer Verlag*.
471. **Greiner, W., et al.**, Validating the EQ-5D with time trade off for the German population. *The European Journal of Health Economics*, 2005. 6(2): p. 124-130.
472. **Kessler, R.C., et al.**, Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of general psychiatry*, 2005. 62(6): p. 593.
473. **Everson-Rose, S.A., J.S. House, and R.P. Mero**, Depressive symptoms and mortality risk in a national sample: confounding effects of health status. *Psychosomatic Medicine*, 2004. 66(6): p. 823.
474. **Sheehan, D., et al.**, The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, 1998. 59: p. 22-33.
475. **Miller, D., et al.**, Clinically Relevant Levels of Depressive Symptoms in Community Dwelling Middle Aged African Americans. *Journal of the American Geriatrics Society*, 2004. 52(5): p. 741-748.
476. **Reiss, S.**, Expectancy model of fear, anxiety, and panic. *Clinical Psychology Review*, 1991. 11(2): p. 141-153.
477. **Reiss, S. and R.J. McNally**, The expectancy model of fear, in Theoretical issues in behaviour therapy, S. Reiss and R.R. Bootzin, Editors. 1985, *Academic*: New York. p. 107-121.

478. **Deacon, B.J., et al.**, The Anxiety Sensitivity Index-Revised: psychometric properties and factor structure in two nonclinical samples. *Behaviour Research and Therapy*, 2003. 41(12): p. 1427-1449.
479. **Peterson, R.A. and K. Plehn**, Measuring anxiety sensitivity, in *Anxiety sensitivity - Theory, research, and treatment of the fear of anxiety*, S. Taylor, Editor. 1999, *Erlbaum*: Mahwah. p. 61-82.
480. **Plehn, K. and R.A. Peterson**, Anxiety sensitivity as a predictor of the development of panic symptoms, panic attacks, and panic disorder: A prospective study. *Journal of anxiety disorders*, 2002. 16(4): p. 455-474.
481. **Schmidt, N.B.**, Prospective evaluation of anxiety sensitivity, in *Anxiety sensitivity: theory, research, & treatment of the fear of anxiety*, S. Taylor, Editor. 1999, *Lawrence Erlbaum Associates*: Mahwah NJ. p. 217-238.
482. **Schmidt, N.B., J.D. Buckner, and M.E. Keough**, Anxiety sensitivity as a prospective predictor of alcohol use disorders. *Behavior modification*, 2007. 31(2): p. 202.
483. **Schmidt, N.B., M.J. Zvolensky, and J.K. Maner**, Anxiety sensitivity: Prospective prediction of panic attacks and Axis I pathology. *Journal of Psychiatric Research*, 2006. 40(8): p. 691-699.
484. **Taylor, S., et al.**, Robust dimensions of anxiety sensitivity: Development and initial validation of the Anxiety Sensitivity Index-3. *Psychological Assessment*, 2007. 19(2): p. 176-188.
485. **Berger, K., et al.**, Iron metabolism and the risk of Restless Legs Syndrome in an elderly general population – The MEMO-Study. *Journal of Neurology*, 2002. 249(9): p. 1195-1199.
486. **Allen, R., et al.**, Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology:: A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. *Sleep Medicine*, 2003. 4(2): p. 101-119.
487. Headache Classification Subcommittee of the International Headache Society. The International Classification of Headache Disorders, 2nd Edition. *Cephalalgia*, 2007. 24 (Supp. 1): p. 23-160.
488. **Pfaffenrath, V., et al.**, Regional variations in the prevalence of migraine and tension type headache applying the new IHS criteria: the German DMKG Headache Study. *Cephalalgia*, 2009. 29(1): p. 48-57.
489. **Buysse, D., et al.**, The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry research*, 1989. 28(2): p. 193-213.
490. **Riemann, D. and J. Backhaus**, *Behandlung von Schlafstörungen*. 1996: *Psychologie Verlags Union Weinheim*.
491. **Griefahn, B., et al.**, Zur Validität der deutschen Übersetzung des Morningness Eveningness Questionnaires von Horne und Östberg. *Somnologie*, 2001. 5(2): p. 71-80.
492. **Netzer, N., et al.**, Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. *Annals of Internal Medicine*, 1999. 131(7): p. 485.
493. **Stang, A., et al.**, Daily siesta, cardiovascular risk factors, and measures of subclinical atherosclerosis: results of the Heinz Nixdorf Recall Study. *Sleep*, 2007. 30(9): p. 1111.

494. **Rammstedt, B. and O. John**, Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 2007. 41(1): p. 203-212.
495. **John, O., E. Donahue, and R. Kentle**, The Big Five Inventory—Versions 4a and 54. *Berkeley: University of California, Berkeley, Institute of Personality and Social Research*, 1991.
496. **Schulz, P., W. Schlotz, and P. Becker**, Trierer Inventar zum chronischen Stress. 2004: Hogrefe.
497. **Becker, P., P. Schulz, and W. Schlotz**, Persönlichkeit, chronischer Stress und körperliche Gesundheit. *Zeitschrift für Gesundheitspsychologie*, 2004. 12(1): p. 11-23.
498. **Schulz, P. and W. Schlotz**, Das Trierer Inventar zur Erfassung von chronischem Stress—Version 2 (TICS2). *Universität Trier, Fachbereich*, 2002.
499. **Geiss, A., et al.**, Predicting the failure of disc surgery by a hypofunctional HPA axis: evidence from a prospective study on patients undergoing disc surgery. *Pain*, 2005. 114(1-2): p. 104-17.
500. **Klaghofer, R., et al.**, Development of life satisfaction in young physicians: results of the prospective SwissMedCareer Study. *International Archives of Occupational and Environmental Health*: p. 1-8.
501. **Gilbert, R., et al.**, Burden and consequences of child maltreatment in high-income countries. *The Lancet*, 2009. 373(9657): p. 68-81.
502. **Binder, E., et al.**, Association of FKBP5 polymorphisms and childhood abuse with risk of posttraumatic stress disorder symptoms in adults. *Jama*, 2008. 299(11): p. 1291.
503. **McGowan, P., et al.**, Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nature Neuroscience*, 2009. 12(3): p. 342-348.
504. **Wingenfeld, K., et al.**, The German Version of the Childhood Trauma Questionnaire (CTQ): Preliminary Psychometric Properties. *Psychotherapie, Psychosomatik, medizinische Psychologie*, 2010.
505. **Wingenfeld, K., et al.**, The Reliable, Valid and Economic Assessment of Early Traumatization: First Psychometric Characteristics of the German Version of the Adverse Childhood Experiences Questionnaire (ACE). *Psychotherapie, Psychosomatik, medizinische Psychologie*, 2010.
506. **Siegrist, J.**, Adverse health effects of high-effort/low-reward conditions. *Journal of occupational health psychology*, 1996. 1(1): p. 27-41.
507. **Siegrist, J. and N. Dragano**, Psychosocial stress and disease risks in occupational life. Results of international studies on the demand-control and the effort-reward imbalance models. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 2008. 51(3): p. 305.
508. **Siegrist, J., et al.**, A short generic measure of work stress in the era of globalization: effort–reward imbalance. *International Archives of Occupational and Environmental Health*, 2009. 82(8): p. 1005-1013.
509. **Leineweber, C., et al.**, How valid is a short measure of effort–reward imbalance at work? A replication study from Sweden. *Occupational and environmental medicine*, 2010. 67(8): p. 526.

510. **Kristensen, T.S., et al.**, The Copenhagen Psychosocial Questionnaire - a tool for the assessment and improvement of the psychosocial work environment. *Scand J Work Environ Health*, 2005. 31: p. 438-49.
511. **Nübling, M., et al.**, Methoden zur Erfassung psychischer Belastungen. Erprobung eines Messinstrumentes (COPSOQ), in Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. 2005, *Bundesanstalt für Arbeitsschutz und Arbeitsmedizin*: Dortmund/Berlin/Dresden.
512. **Nübling, M., et al.**, Measuring psychological stress and strain at work-Evaluation of the COPSOQ Questionnaire in Germany. *GMS Psycho-Social Medicine*, 2006. 3.
513. **Dragano, N., P.E. Verde, and J. Siegrist**, Organisational downsizing and work stress: testing synergistic health effects in employed men and women. *Journal of epidemiology and community health*, 2005. 59(8): p. 694.
514. **Jansen, R.**, "Die BIBB/IAB-Erhebung und die Gesundheitsberichterstattung", in Arbeitsweltbezogene Gesundheitsberichterstattung in Deutschland. Stand und Perspektiven. Beiträge zur Gesundheitsberichterstattung des Bundes. . 2002, *Robert Koch-Institut*: Berlin. p. 86-97.
515. **Hall, A.**, Die BIBB/BAuA-Erwerbstätigenbefragung 2006 : Methodik und Frageprogramm im Vergleich zur BIBB/IAB-Erhebung 1998. 2009, *Bundesinstitut für Berufsbildung*: Bonn.
516. **Berkman, L.F., et al.**, Social Integration and Mortality: A Prospective Study of French Employees of Electricity of France–Gas of France. *American journal of Epidemiology*, 2004. 159(2): p. 167.
517. **Seeman, T.E. and L.F. Berkman**, Structural characteristics of social networks and their relationship with social support in the elderly: who provides support. *Social Science & Medicine*, 1988. 26(7): p. 737-749.
518. **Picavet, H.S.J. and J.M.W. Hazes**, Prevalence of self reported musculoskeletal diseases is high. *British Medical Journal*, 2003. 62(7): p. 644.
519. **Schott, G.D.**, The cartography of pain: The evolving contribution of pain maps. *European Journal of Pain*, 2010. 14(8): p. 784-791.
520. **van den Hoven, L.H., K.J. Gorter, and H.S. Picavet**, Measuring musculoskeletal pain by questionnaires: the manikin versus written questions. *Eur J Pain*, 2010. 14(3): p. 335-338.
521. **Pfingsten, M., et al.**, Psychometric properties of the pain drawing and the Ransford technique in patients with chronic low back pain. *Schmerz*, 2003. 17(5): p. 332.
522. **Volzke, H., et al.**, Cohort profile: the Study of Health in Pomerania. *International journal of epidemiology*, 2010.
523. **Jensen, M.P. and P. Karoly**, Self-report scales and procedures for assessing pain in adults, in Handbook of pain assessment, D.C. Turk, Melzack, R, Editor. 2005, *The Guilford Press*: New York.
524. **Von Korff, M.**, Epidemiological and survey methods: assessment of chronic pain, in Handbook of pain assessment, D.C. Turk, Melzack, R., Editor. 2005, *The Guilford Press*: New York.
525. **Leeb, B.F., et al.**, Patient-centered rheumatoid arthritis disease activity assessment by a modified RADAI. *The Journal of rheumatology*, 2008. 35(7): p. 1294.

526. **Mack, F., et al.**, Study of Health in Pomerania (SHIP): Relationship among socioeconomic and general health factors and dental status among elderly adults in Pomerania. *Quintessence international*, 2003. 34(10): p. 772-778.
527. **Mundt, T., et al.**, Socioeconomic indicators and prosthetic replacement of missing teeth in a working age population—Results of the Study of Health in Pomerania (SHIP). *Community Dentistry and Oral Epidemiology*, 2009. 37(2): p. 104-115.
528. **Mack, F., et al.**, The impact of tooth loss on general health related to quality of life among elderly Pomeranians: results from the study of health in Pomerania (SHIP-O). *The International journal of prosthodontics*, 2005. 18(5): p. 414.
529. **Sanders, A.E., et al.**, Impact of oral disease on quality of life in the US and Australian populations. *Community Dentistry and Oral Epidemiology*, 2009. 37(2): p. 171-181.
530. **John, M.T., W. Micheelis, and R. Biffar**, Normwerte mundgesundheitsbezogener Lebensqualität für Kurzversionen des Oral Health Impact Profile. *SCHWEIZER MONATSSCHRIFT FÜR ZAHNHEILKUNDE*, 2004. 114(8): p. 784-791.
531. **John, M.T., et al.**, German short forms of the oral health impact profile. *Community Dentistry and Oral Epidemiology*, 2006. 34(4): p. 277-288.
532. **Lagerros, Y.T. and P. Lagiou**, Assessment of physical activity and energy expenditure in epidemiological research of chronic diseases. *European journal of epidemiology*, 2007. 22(6): p. 353-362.
533. **Ferrari, P., C. Friedenreich, and C.E. Matthews**, The role of measurement error in estimating levels of physical activity. *American journal of epidemiology*, 2007. 166(7): p. 832.
534. **Bingham, S., et al.**, Associations between dietary methods and biomarkers, and between fruits and vegetables and risk of ischaemic heart disease, in the EPIC Norfolk Cohort Study. *International journal of epidemiology*, 2008. 37(5): p. 978.
535. **Bundesministeriums für Umwelt, N.u.R.**, Umweltradioaktivität und Strahlenbelastung 2006. Unterrichtung durch die Bundesregierung Drucksache.16/6835.16. Wahlperiode. 2007, *Bundesministeriums für Umwelt, Naturschutz und Reaktorsicherheit*: Berlin. p. 1-75.
536. **European Foundation for the Improvement of Living and Working Conditions**, Fourth European Working Conditions Survey. 2007, *Office for Official Publications of the European Communities*: Luxembourg.
537. **Harrington, J.M.**, Health effects of shift work and extended hours of work. *Occupational and environmental medicine*, 2001. 58(1): p. 68.
538. **International Agency for Research on Cancer**, Painting, Firefighting, and Shiftwork., in *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. 2010, *IARC*: Lyon.
539. **European Agency for Safety and Health at Work, E.**, Noises in figures. 2005, *Office for Official Publications of the European Communities*: Luxembourg.
540. **Jonas, D.E., et al.**, Patient time requirements for screening colonoscopy. *The American journal of gastroenterology*, 2007. 102(11): p. 2401-2410.
541. **Russell, L.B.**, Completing costs: patients' time. *Medical care*, 2009. 47(7_Supplement_1): p. S89.
542. **Russell, L.B., D.C. Suh, and M.A. Safford**, Time requirements for diabetes self-management: too much for many. *J Fam Pract*, 2005. 54(1): p. 52-56.

543. **Holle, R., et al.**, KORA--a research platform for population based health research. *Gesundheitswesen (Bundesverband der Ärzte des Öffentlichen Gesundheitsdienstes (Germany))*, 2005. 67: p. S19.
544. **Haynes, R.B., et al.**, Interventions for enhancing medication adherence. *Cochrane Database of Systematic Reviews*, 2008. 2.
545. **Shi, L., et al.**, Correlation between adherence rates measured by MEMS and self-reported questionnaires: a meta-analysis. *Health and Quality of Life Outcomes*, 2010. 8: p. 99.
546. **Gehi, A.K., et al.**, Self-reported medication adherence and cardiovascular events in patients with stable coronary heart disease: the heart and soul study. *Archives of Internal Medicine*, 2007. 167(16): p. 1798.
547. **Morisky, D.E., et al.**, Five-year blood pressure control and mortality following health education for hypertensive patients. *American Journal of Public Health*, 1983. 73(2): p. 153.
548. **Lawes, C.M.M., S.V. Hoorn, and A. Rodgers**, Global burden of blood-pressure-related disease, 2001. *The Lancet*, 2008. 371(9623): p. 1513-1518.
549. **Lewington, S., et al.**, Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet*, 2002. 360(9349): p. 1903.
550. **Duron, E. and O. Hanon**, Hypertension, cognitive decline and dementia. *Archives of Cardiovascular Diseases*, 2008. 101(3): p. 181-189.
551. **Kennelly, S.P., B.A. Lawlor, and R.A. Kenny**, Blood pressure and the risk for dementia--A double edged sword. *Ageing research reviews*, 2009. 8(2): p. 61-70.
552. **Moretti, R., et al.**, Risk factors for vascular dementia: Hypotension as a key point. *Vascular Health and Risk Management*, 2008. 4(2): p. 395.
553. **Paglieri, C., et al.**, Hypertension and cognitive function. *Clinical and Experimental Hypertension*, 2008. 30(8): p. 701-710.
554. **Qiu, C., B. Winblad, and L. Fratiglioni**, The age-dependent relation of blood pressure to cognitive function and dementia. *The Lancet Neurology*, 2005. 4(8): p. 487-499.
555. **Davy, K.P. and J.E. Hall**, Obesity and hypertension: two epidemics or one? *American Journal of Physiology- Regulatory, Integrative and Comparative Physiology*, 2004. 286(5): p. R803.
556. **Francischetti, E.A. and V.A. Genelhu**, Obesity-hypertension: an ongoing pandemic. *International journal of clinical practice*, 2007. 61(2): p. 269-280.
557. **Mathieu, P., et al.**, Visceral obesity: the link among inflammation, hypertension, and cardiovascular disease. *Hypertension*, 2009. 53(4): p. 577.
558. **Meisinger, C., A. Döring, and M. Heier**, Blood pressure and risk of type 2 diabetes mellitus in men and women from the general population: the Monitoring Trends and Determinants on Cardiovascular Diseases/Cooperative Health Research in the Region of Augsburg Cohort Study. *Journal of hypertension*, 2008. 26(9): p. 1809.
559. **Palatini, P.**, Elevated heart rate as a predictor of increased cardiovascular morbidity. *Journal of hypertension. Supplement: official journal of the International Society of Hypertension*, 1999. 17(3): p. S3.

560. **Palatini, P., A. Benetos, and S. Julius**, Impact of increased heart rate on clinical outcomes in hypertension: implications for antihypertensive drug therapy. *Drugs*, 2006. 66(2): p. 133-144.
561. **Perret-Guillaume, C., L. Joly, and A. Benetos**, Heart rate as a risk factor for cardiovascular disease. *Progress in cardiovascular diseases*, 2009. 52(1): p. 6-10.
562. **Tjogen, T.B., A. Flaa, and S.E. Kjeldsen**, High Heart Rate as Predictor of Essential Hypertension::: The Hyperkinetic State, Evidence of Prediction of Hypertension, and Hemodynamic Transition to Full Hypertension. *Progress in cardiovascular diseases*, 2009. 52(1): p. 20-25.
563. **Rautaharju, P.M., et al.**, A simple procedure for positioning precordial ECG and VCG electrodes using an electrode locator*. *Journal of Electrocardiology*, 1976. 9(1): p. 35-40.
564. **Owan, T.E., et al.**, Secular trends in renal dysfunction and outcomes in hospitalized heart failure patients. *Journal of cardiac failure*, 2006. 12(4): p. 257-262.
565. **Kühl, H.P.**, Left ventricular function, in Three-dimensional echocardiography, T. Buck, A. Frank, and M. Monaghan, Editors. 2010, *Springer*: Berlin.
566. **Nagueh, S.F., et al.**, Recommendations for the evaluation of left ventricular diastolic function by echocardiography. *European Journal of Echocardiography*, 2009. 10(2): p. 165.
567. **Laurent, S., P. Boutouyrie, and P. Lacolley**, Structural and genetic bases of arterial stiffness. *Hypertension*, 2005. 45(6): p. 1050.
568. **Schwartzkopff, B., et al.**, Structural and functional alterations of the intramyocardial coronary arterioles in patients with arterial hypertension. *Circulation*, 1993. 88(3): p. 993.
569. **Roman, M.J., et al.**, High central pulse pressure is independently associated with adverse cardiovascular outcome: the Strong Heart Study. *Journal of the American College of Cardiology*, 2009. 54(18): p. 1730.
570. **Roman, M.J., et al.**, Central pressure more strongly relates to vascular disease and outcome than does brachial pressure: the Strong Heart Study. *Hypertension*, 2007. 50(1): p. 197.
571. **Mancia, G., et al.**, ESH-ESC practice guidelines for the management of arterial hypertension: ESH-ESC Task Force on the Management of Arterial Hypertension. *J Hypertens*, 2007. 25(9): p. 1751-62.
572. **Laurent, S., et al.**, Expert consensus document on arterial stiffness: methodological issues and clinical applications. *European heart journal*, 2006. 27(21): p. 2588.
573. **Wilkinson, I.B., et al.**, Reproducibility of pulse wave velocity and augmentation index measured by pulse wave analysis. *Journal of hypertension*, 1998. 16(12): p. 2079.
574. **Hickson, S.S., et al.**, Validity and repeatability of the Vicorder apparatus: a comparison with the SphygmoCor device. *Hypertension Research*, 2009. 32(12): p. 1079-1085.
575. **Hirata, K., M. Kawakami, and M.F. O'Rourke**, Pulse wave analysis and pulse wave velocity: a review of blood pressure interpretation 100 years after Korotkov. *Circulation journal: official journal of the Japanese Circulation Society*, 2006. 70(10): p. 1231.

576. **Criqui, M.H., et al.**, Mortality over a period of 10 years in patients with peripheral arterial disease. *New England Journal of Medicine*, 1992. 326(6): p. 381-386.
577. **Leng, G.C., et al.**, Incidence, natural history and cardiovascular events in symptomatic and asymptomatic peripheral arterial disease in the general population. *International journal of epidemiology*, 1996. 25(6): p. 1172.
578. **Newman, A.B., et al.**, Ankle-arm index as a marker of atherosclerosis in the cardiovascular health study. Cardiovascular health study (CHS) collaborative research group. *Circulation*, 1993. 88(3): p. 837.
579. **Sikkink, C.J., et al.**, Decreased ankle/brachial indices in relation to morbidity and mortality in patients with peripheral arterial disease. *Vascular medicine (London, England)*, 1997. 2(3): p. 169.
580. **Hirsch, A.T., et al.**, ACC/AHA 2005 Practice Guidelines for the Management of Patients With Peripheral Arterial Disease (Lower Extremity, Renal, Mesenteric, and Abdominal Aortic): A Collaborative Report from the American Association for Vascular Surgery/Society for Vascular Surgery,* Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, Society of Interventional Radiology, and the ACC/AHA Task Force on Practice Guidelines (Writing Committee to Develop Guidelines for the Management of Patients With Peripheral Arterial Disease): Endorsed by the American Association of Cardiovascular and Pulmonary Rehabilitation; National Heart, Lung, and Blood Institute; Society for Vascular Nursing; TransAtlantic Inter-Society Consensus; and Vascular Disease Foundation. *Circulation*, 2006. 113(11): p. e463.
581. **McGill Jr, H.C., et al.**, Effects of coronary heart disease risk factors on atherosclerosis of selected regions of the aorta and right coronary artery. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 2000. 20(3): p. 836.
582. **Virmani, R., et al.**, Pathology of the vulnerable plaque. *Journal of the American College of Cardiology*, 2006. 47(8 Suppl C): p. C13.
583. **Chambless, L.E., et al.**, Association of coronary heart disease incidence with carotid arterial wall thickness and major risk factors: the Atherosclerosis Risk in Communities (ARIC) Study, 1987–1993. *American journal of Epidemiology*, 1997. 146(6): p. 483.
584. **Greenland, P., et al.**, Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals. *Jama*, 2004. 291(2): p. 210.
585. **Hunt, K.J., et al.**, Acoustic shadowing on B-mode ultrasound of the carotid artery predicts CHD. *Ultrasound in medicine & biology*, 2001. 27(3): p. 357-365.
586. **O'Leary, D.H., et al.**, Carotid-artery intima and media thickness as a risk factor for myocardial infarction and stroke in older adults. *New England Journal of Medicine*, 1999. 340(1): p. 14.
587. **Stork, S., et al.**, Carotid artery plaque burden, stiffness, and mortality risk in elderly men: a prospective, population-based cohort study. *Circulation*, 2004. 110(3): p. 344.
588. **van der Meer, I.M., et al.**, Predictive value of noninvasive measures of atherosclerosis for incident myocardial infarction: the Rotterdam Study. *Circulation*, 2004. 109(9): p. 1089.
589. **Antonopoulos, S.**, Third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III) final report. *Circulation*, 2002. 106: p. 3143-421.

590. **Greenland, P., et al.**, Prevention Conference V: Beyond secondary prevention: identifying the high-risk patient for primary prevention: noninvasive tests of atherosclerotic burden: Writing Group III. *Circulation*, 2000. 101(1): p. e16.
591. **Taylor, A.J., C.N. Merz, and J.E. Udelson**, 34th Bethesda Conference: Executive summary--can atherosclerosis imaging techniques improve the detection of patients at risk for ischemic heart disease? *Journal of the American College of Cardiology*, 2003. 41(11): p. 1860.
592. **Stein, J.H., et al.**, Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force endorsed by the Society for Vascular Medicine. *Journal of the American Society of Echocardiography*, 2008. 21(2): p. 93-111.
593. **Touboul, P.J., et al.**, Mannheim carotid intima-media thickness consensus (2004–2006). *Cerebrovascular diseases*, 2006. 23(1): p. 75-80.
594. **Cowie, C.C., et al.**, Full accounting of diabetes and pre-diabetes in the US population in 1988–1994 and 2005–2006. *Diabetes care*, 2009. 32(2): p. 287.
595. **Matthews, D.R., et al.**, Homeostasis model assessment: insulin resistance and -cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*, 1985. 28(7): p. 412-419.
596. **Hanson, R.L., et al.**, Evaluation of simple indices of insulin sensitivity and insulin secretion for use in epidemiologic studies. *American journal of Epidemiology*, 2000. 151(2): p. 190.
597. **Phillips, D.I.W., et al.**, Understanding oral glucose tolerance: comparison of glucose or insulin measurements during the oral glucose tolerance test with specific measurements of insulin resistance and insulin secretion. *Diabetic Medicine*, 1994. 11(3): p. 286-292.
598. **Haffner, S.M., et al.**, Decreased insulin secretion and increased insulin resistance are independently related to the 7-year risk of NIDDM in Mexican-Americans. *Diabetes*, 1995. 44(12): p. 1386.
599. **Mooy, J.M., et al.**, Intra-individual variation of glucose, specific insulin and proinsulin concentrations measured by two oral glucose tolerance tests in a general Caucasian population: the Hoorn Study. *Diabetologia*, 1996. 39(3): p. 298-305.
600. **Meerwaldt, R., et al.**, Simple non-invasive assessment of advanced glycation endproduct accumulation. *Diabetologia*, 2004. 47(7): p. 1324-1330.
601. **Meerwaldt, R., et al.**, Skin autofluorescence, a measure of cumulative metabolic stress and advanced glycation end products, predicts mortality in hemodialysis patients. *Journal of the American Society of Nephrology*, 2005. 16(12): p. 3687.
602. **den Hollander, N.C., et al.**, Advanced glycation end products and the absence of premature atherosclerosis in glycogen storage disease Ia. *Journal of inherited metabolic disease*, 2007. 30(6): p. 916-923.
603. **Ediger, M.N., B.P. Olson, and J.D. Maynard**, Noninvasive optical screening for diabetes. *J Diabetes Sci Technol*, 2009. 3(4): p. 776-780.
604. **Lutgers, H.L., et al.**, Skin autofluorescence provides additional information to the UK Prospective Diabetes Study (UKPDS) risk score for the estimation of cardiovascular prognosis in type 2 diabetes mellitus. *Diabetologia*, 2009. 52(5): p. 789-797.

605. **Kaplan, G.A., et al.**, Childhood socioeconomic position and cognitive function in adulthood. *International journal of epidemiology*, 2001. 30(2): p. 256.
606. **Qiu, C., et al.**, The influence of education on clinically diagnosed dementia incidence and mortality data from the Kungsholmen Project. *Archives of neurology*, 2001. 58(12): p. 2034.
607. **Chang, M., et al.**, The Effect of Midlife Physical Activity on Cognitive Function Among Older Adults: AGES—Reykjavik Study. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 2010.
608. **Larson, E.B., et al.**, Exercise is associated with reduced risk for incident dementia among persons 65 years of age and older. *Annals of Internal Medicine*, 2006. 144(2): p. 73.
609. **Nooyens, A.C.J., B.M. van Gelder, and W.M. Verschuren**, Smoking and cognitive decline among middle-aged men and women: the Doetinchem Cohort Study. *American Journal of Public Health*, 2008. 98(12): p. 2244.
610. **Peters, R., et al.**, Smoking, dementia and cognitive decline in the elderly, a systematic review. *BMC geriatrics*, 2008. 8(1): p. 36.
611. **Cukierman, T., H.C. Gerstein, and J.D. Williamson**, Cognitive decline and dementia in diabetes—systematic overview of prospective observational studies. *Diabetologia*, 2005. 48(12): p. 2460-2469.
612. **Sin, D.D., L.L. Wu, and S.F. Man**, The Relationship Between Reduced Lung Function and Cardiovascular Mortality*. *Chest*, 2005. 127(6): p. 1952.
613. **Young, R.P., R. Hopkins, and T.E. Eaton**, Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. *European Respiratory Journal*, 2007. 30(4): p. 616.
614. **Koch, B., et al.**, Lung function reference values in different German populations. *Respiratory medicine*, 2010.
615. **Schnabel, E., et al.**, Regional differences in prediction models of lung function in Germany. *Respiratory Research*, 2010. 11(1): p. 40.
616. **Miller, M.R., et al.**, ATS/ERS Task Force (2005) Standardisation of spirometry. *Eur Respir J*, 2005. 26(2): p. 319–338.
617. **Dressel, H., et al.**, Exhaled nitric oxide: Independent effects of atopy, smoking, respiratory tract infection, gender and height. *Respiratory Medicine*, 2008. 102(7): p. 962-969.
618. **American Thoracic Society Workshop**, ATS Workshop Proceedings: Exhaled Nitric Oxide and Nitric Oxide Oxidative Metabolism in Exhaled Breath Condensate: Executive Summary. *Am. J. Respir. Crit. Care Med.*, 2006. 173(7): p. 811-813.
619. **Bohadana, A., et al.**, Reproducibility of exhaled nitric oxide in smokers and non-smokers: relevance for longitudinal studies. *BMC pulmonary medicine*, 2008. 8(1): p. 4.
620. **Grob, N.M. and R.A. Dweik**, Exhaled nitric oxide in asthma: progress since the introduction of standardized methodology. *Journal of Breath Research*, 2008. 2: p. 037002.
621. **Kharitonov, S.A., et al.**, Reproducibility of exhaled nitric oxide measurements in healthy and asthmatic adults and children. *European Respiratory Journal*, 2003. 21(3): p. 433.
622. **Olin, A.C., B. Bake, and K. Torén**, Fraction of Exhaled Nitric Oxide at 50 mL/s*. *Chest*, 2007. 131(6): p. 1852.

623. **Olin, A.C., et al.**, Height, Age, and Atopy Are Associated With Fraction of Exhaled Nitric Oxide in a Large Adult General Population Sample*. *Chest*, 2006. 130(5): p. 1319.
624. **Karrasch, S., et al.**, Exhaled nitric oxide and influencing factors in a random population sample. *Respiratory Medicine*, 2010.
625. **Zhang, W., et al.**, EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis. *Annals of the rheumatic diseases*, 2010. 69(3): p. 483-489.
626. **Prevo, M.L.L., et al.**, Modified disease activity scores that include twenty-eight-joint counts development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis & Rheumatism*, 1995. 38(1): p. 44-48.
627. **König, J., B. Holtfreter, and T. Kocher**, Periodontal health in Europe: future trends based on treatment needs and the provision of periodontal services—position paper 1. *European Journal of Dental Education*, 2010. 14: p. 4-24.
628. **Hensel, E., et al.**, Study of Health in Pomerania (SHIP): a health survey in an East German region. Objectives and design of the oral health section. *Quintessence international (Berlin, Germany: 1985)*, 2003. 34(5): p. 370.
629. **John, U., et al.**, Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Sozial-und Präventivmedizin/Social and Preventive Medicine*, 2001. 46(3): p. 186-194.
630. **WHO**, Oral Health Surveys. Basic methods. 1997, *World Health Organization*: Geneva.
631. **Gerecz-Simon, E.M., et al.**, Measurement of pain threshold in patients with rheumatoid arthritis, osteoarthritis, ankylosing spondylitis, and healthy controls. *Clinical Rheumatology*, 1989. 8(4): p. 467-474.
632. **Daniel, E.**, Noise and hearing loss: a review. *Journal of School Health*, 2007. 77(5): p. 225-231.
633. **Robert-Koch-Institut**, Hörstörungen und Tinnitus, in Gesundheitsberichterstattung des Bundes, Heft 29. 2006, *Robert-Koch-Institut*: Berlin.
634. **Smits, C. and T. Houtgast**, Results from the Dutch speech-in-noise screening test by telephone. *Ear and hearing*, 2005. 26(1): p. 89.
635. **Landis, B. and T. Hummel**, New evidence for high occurrence of olfactory dysfunctions within the population. *The American journal of medicine*, 2006. 119(1): p. 91.
636. **Damm, M., et al.**, Epidemiologie und Therapie von Riechstörungen in Deutschland, Österreich und der Schweiz. *HNO*, 2004. 52(2): p. 112-120.
637. **Ponsen, M., et al.**, Idiopathic hyposmia as a preclinical sign of Parkinson's disease. *Annals of Neurology*, 2004. 56(2): p. 173-181.
638. **Bacon, A., et al.**, Very Early Changes in Olfactory Functioning Due to Alzheimer's Disease and the Role of Apolipoprotein E in Olfactiona. *Annals of the New York Academy of Sciences*, 1998. 855(OLFACTION AND TASTE XII: AN INTERNATIONAL SYMPOSIUM): p. 723-731.
639. **Vennemann, M., T. Hummel, and K. Berger**, The association between smoking and smell and taste impairment in the general population. *Journal of Neurology*, 2008. 255(8): p. 1121-1126.
640. **Gunnell, D., et al.**, Height, leg length, and cancer risk: a systematic review. *Epidemiologic reviews*, 2001. 23(2): p. 313.

641. **Samaras, T.T., H. Elrick, and L.H. Storms**, Is short height really a risk factor for coronary heart disease and stroke mortality? A review. *Med Sci Monit*, 2004. 10(4): p. 76.
642. **Silventoinen, K., et al.**, Association between height and coronary heart disease mortality: a prospective study of 35,000 twin pairs. *American journal of Epidemiology*, 2006. 163(7): p. 615.
643. **Smith, G.D., et al.**, Height and risk of death among men and women: aetiological implications of associations with cardiorespiratory disease and cancer mortality. *Journal of epidemiology and community health*, 2000. 54(2): p. 97.
644. **Henriksson, K.M., et al.**, Associations between body height, body composition and cholesterol levels in middle-aged men. the coronary risk factor study in southern Sweden (CRISS). *European journal of epidemiology*, 2001. 17(6): p. 521-526.
645. **Langenberg, C., et al.**, Influence of height, leg and trunk length on pulse pressure, systolic and diastolic blood pressure. *Journal of hypertension*, 2003. 21(3): p. 537.
646. **Lear, S.A., et al.**, The Use of BMI and Waist Circumference as Surrogates of Body Fat Differs by Ethnicity. *Obesity*, 2007. 15(11): p. 2817-2824.
647. **Heid, I.M., et al.**, Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature genetics*, 2010. 42(11): p. 949-960.
648. **Yusuf, S., et al.**, Obesity and the risk of myocardial infarction in 27 000 participants from 52 countries: a case-control study. *The Lancet*, 2005. 366(9497): p. 1640-1649.
649. **Ashwell, M., T.J. Cole, and A.K. Dixon**, Ratio of waist circumference to height is strong predictor of intra-abdominal fat. *British Medical Journal*, 1996. 313(7056): p. 559.
650. **Ashwell, M. and S.D. Hsieh**, Six reasons why the waist-to-height ratio is a rapid and effective global indicator for health risks of obesity and how its use could simplify the international public health message on obesity. 2009.
651. **Osteologie., D.**, DVO Leitlinie zur Prophylaxe, Diagnostik und Therapie der Osteoporose bei Erwachsenen - Langfassung -. 2009, *Dachverband Osteologie e.V.*
652. **Syed, Z. and A. Khan**, Bone densitometry: applications and limitations. *Journal of obstetrics and gynaecology Canada: JOGC= Journal d'obstétrique et gynécologie du Canada: JOGC*, 2002. 24(6): p. 476.
653. **Laskey, M.**, Dual-energy X-ray absorptiometry and body composition. *Nutrition*, 1996. 12(1): p. 45-51.
654. **Prior, B.M., et al.**, In vivo validation of whole body composition estimates from dual-energy X-ray absorptiometry. *Journal of Applied Physiology*, 1997. 83(2): p. 623.
655. **Salamone, L.M., et al.**, Measurement of fat mass using DEXA: a validation study in elderly adults. *Journal of Applied Physiology*, 2000. 89(1): p. 345.
656. **BlueCross BlueShield of Kansas City**, Whole body Dual X-Ray absorptiometry to determine Body Composition. 2010.
657. **Helmerhorst, H.J.F., et al.**, Objectively measured sedentary time may predict insulin resistance independent of moderate-and vigorous-intensity physical activity. *Diabetes*, 2009. 58(8): p. 1776.

References

658. **Pate, R.R., J.R. O'Neill, and F. Lobelo**, The Evolving Definition of " Sedentary". *Exercise and sport sciences reviews*, 2008. 36(4): p. 173.
 659. **Bouten, C.V., et al.**, Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Medicine & Science in Sports & Exercise*, 1994. 26(12): p. 1516.
 660. **Crouter, S.E., K.G. Clowers, and D.R. Bassett Jr**, A novel method for using accelerometer data to predict energy expenditure. *Journal of Applied Physiology*, 2006. 100(4): p. 1324.
 661. **Plasqui, G., et al.**, Measuring Free-Living Energy Expenditure and Physical Activity with Triaxial Accelerometry. *Obesity*, 2005. 13(8): p. 1363-1369.
 662. **Pober, D.M., et al.**, Development of novel techniques to classify physical activity mode using accelerometers. *Medicine & Science in Sports & Exercise*, 2006. 38(9): p. 1626.
 663. **Staudenmayer, J., et al.**, An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *Journal of Applied Physiology*, 2009. 107(4): p. 1300.
 664. **Siconolfi, S.F., et al.**, A simple, valid step test for estimating maximal oxygen uptake in epidemiologic studies. *American journal of Epidemiology*, 1985. 121(3): p. 382.
 665. **Rantanen, T., et al.**, Midlife hand grip strength as a predictor of old age disability. *Jama*, 1999. 281(6): p. 558.
 666. **Bohannon, R.W.**, Dynamometer measurements of hand-grip strength predict multiple outcomes. *Perceptual and Motor skills*, 2001. 93(2): p. 323.
 667. **Bohannon, R.W.**, Is it Legitimate to Characterize Muscle Strength Using a Limited Number of Measures? *The Journal of Strength & Conditioning Research*, 2008. 22(1): p. 166.
 668. **Cesari, M., et al.**, Physical function and self-rated health status as predictors of mortality: results from longitudinal analysis in the iLSIRENTE study. *BMC geriatrics*, 2008. 8(1): p. 34.
 669. **Gale, C.R., et al.**, Grip strength, body composition, and mortality. *International journal of epidemiology*, 2007. 36(1): p. 228.
 670. **Metter, E.J., et al.**, Skeletal muscle strength as a predictor of all-cause mortality in healthy men. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 2002. 57(10): p. B359.
 671. **Dixon, W.G., et al.**, Low grip strength is associated with bone mineral density and vertebral fracture in women. *Rheumatology*, 2005. 44(5): p. 642.
 672. **Bohannon, R.W.**, Dynamometer measurements of grip and knee extension strength: are they indicative of overall limb and trunk muscle strength? *Percept Mot Skills*, 2009. 108(2): p. 339-342.
 673. **Knols, R.H., et al.**, Hand-held dynamometry in patients with haematological malignancies: Measurement error in the clinical assessment of knee extension strength. *BMC musculoskeletal disorders*, 2009. 10(1): p. 31.
 674. **Li, R.C., et al.**, The development, validity, and reliability of a manual muscle testing device with integrated limb position sensors. *Archives of physical medicine and rehabilitation*, 2006. 87(3): p. 411-417.
-

675. **Bellace, J.V., et al.**, Validity of the Dexter Evaluation System's Jamar dynamometer attachment for assessment of hand grip strength in a normal population. *Journal of hand therapy: official journal of the American Society of Hand Therapists*, 2000. 13(1): p. 46.
676. **Lindstrom-Hazel, D., A. Kratt, and L. Bix**, Interrater reliability of students using hand and pinch dynamometers. *The American Journal of Occupational Therapy*, 2009. 63(2): p. 193.
677. **Mathiowetz, V.**, Comparison of Rolyan and Jamar dynamometers for measuring grip strength. *Occupational Therapy International*, 2002. 9(3): p. 201-209.
678. **Suni, J.H., et al.**, Health-related fitness test battery for adults: aspects of reliability. *Archives of physical medicine and rehabilitation*, 1996. 77(4): p. 399-405.
679. **Hegenscheid, K., et al.**, Whole-body magnetic resonance imaging of healthy volunteers: pilot study results from the population-based SHIP study. *Rofo*, 2009. 181(8): p. 748-59.
680. **Volzke, H., et al.**, Cohort Profile: The Study of Health in Pomerania. *Int J Epidemiol*, 2010.
681. **Antoch, G., et al.**, Whole-body dual-modality PET/CT and whole-body MRI for tumor staging in oncology. *Jama*, 2003. 290(24): p. 3199-206.
682. **Schlemmer, H.P., et al.**, Fast whole-body assessment of metastatic disease using a novel magnetic resonance imaging system: initial experiences. *Invest Radiol*, 2005. 40(2): p. 64-71.
683. **Weckbach, S., et al.**, Systemic cardiovascular complications in patients with long-standing diabetes mellitus: comprehensive assessment with whole-body magnetic resonance imaging/magnetic resonance angiography. *Invest Radiol*, 2009. 44(4): p. 242-50.
684. **Sodickson, D.K., et al.**, Recent advances in image reconstruction, coil sensitivity calibration, and coil array design for SMASH and generalized parallel MRI. *MAGMA*, 2002. 13(3): p. 158-63.
685. **Kallmes, D.F., F.K. Hui, and J.P. Mugler, 3rd**, Suppression of cerebrospinal fluid and blood flow artifacts in FLAIR MR imaging with a single-slab three-dimensional pulse sequence: initial experience. *Radiology*, 2001. 221(1): p. 251-5.
686. **Lichy, M.P., et al.**, Magnetic resonance imaging of the body trunk using a single-slab, 3-dimensional, T2-weighted turbo-spin-echo sequence with high sampling efficiency (SPACE) for high spatial resolution imaging: initial clinical experiences. *Invest Radiol*, 2005. 40(12): p. 754-60.
687. **Goyen, M., et al.**, Whole-body three-dimensional MR angiography with a rolling table platform: initial clinical experience. *Radiology*, 2002. 224(1): p. 270-7.
688. **Quick, H.H., et al.**, High spatial resolution whole-body MR angiography featuring parallel imaging: initial experience. *Rofo*, 2004. 176(2): p. 163-9.
689. **Zenge, M.O., M.E. Ladd, and H.H. Quick**, Novel reconstruction method for three-dimensional axial continuously moving table whole-body magnetic resonance imaging featuring autocalibrated parallel imaging GRAPPA. *Magn Reson Med*, 2009. 61(4): p. 867-873.
690. **Zenge, M.O., et al.**, Whole-body magnetic resonance imaging featuring moving table continuous data acquisition with high-precision position feedback. *Magn Reson Med*, 2005. 54(3): p. 707-11.

691. **Zenge, M.O., et al.**, High-resolution continuously acquired peripheral MR angiography featuring partial parallel imaging GRAPPA. *Magn Reson Med*, 2006. 56(4): p. 859-65.
692. **Aldefeld, B., P. Bornert, and J. Keupp**, Continuously moving table 3D MRI with lateral frequency-encoding direction. *Magn Reson Med*, 2006. 55(5): p. 1210-6.
693. **Bornert, P. and B. Aldefeld**, Principles of whole-body continuously-moving-table MRI. *J Magn Reson Imaging*, 2008. 28(1): p. 1-12.
694. **Fautz, H.P. and S.A. Kannengiesser**, Sliding multislice (SMS): a new technique for minimum FOV usage in axial continuously moving-table acquisitions. *Magn Reson Med*, 2006. 55(2): p. 363-70.
695. **Kruger, D.G., et al.**, Dual-velocity continuously moving table acquisition for contrast-enhanced peripheral magnetic resonance angiography. *Magn Reson Med*, 2005. 53(1): p. 110-7.
696. **Kruger, D.G., et al.**, Recovery of phase inconsistencies in continuously moving table extended field of view magnetic resonance imaging acquisitions. *Magn Reson Med*, 2005. 54(3): p. 712-7.
697. **Kramer, H., et al.**, Peripheral magnetic resonance angiography (MRA) with continuous table movement at 3.0 T: initial experience compared with step-by-step MRA. *Invest Radiol*, 2008. 43(9): p. 627-34.
698. **Kruger, D.G., et al.**, Continuously moving table data acquisition method for long FOV contrast-enhanced MRA and whole-body MRI. *Magn Reson Med*, 2002. 47(2): p. 224-31.
699. **Vogt, F.M., et al.**, Peripheral vascular disease: comparison of continuous MR angiography and conventional MR angiography--pilot study. *Radiology*, 2007. 243(1): p. 229-38.
700. **Baumann, T., et al.**, Continuously moving table MRI with sliding multislice for rectal cancer staging: Image quality and lesion detection. *Eur J Radiol*, 2009.
701. **Baumann, T., et al.**, Detection of pulmonary nodules with move-during-scan magnetic resonance imaging using a free-breathing turbo inversion recovery magnitude sequence. *Invest Radiol*, 2008. 43(6): p. 359-67.
702. **Bley, T.A., et al.**, Sliding multislice MRI for abdominal staging of rectal gastrointestinal stromal tumours. *In Vivo*, 2007. 21(5): p. 891-4.
703. **Brauck, K., et al.**, Feasibility of whole-body MR with T2- and T1-weighted real-time steady-state free precession sequences during continuous table movement to depict metastases. *Radiology*, 2008. 246(3): p. 910-6.
704. **Schaefer, J.F. and H.P. Schlemmer**, Total-body MR-imaging in oncology. *Eur Radiol*, 2006. 16(9): p. 2000-15.
705. **Sommer, G., et al.**, Sliding multislice MRI for abdominal staging of patients with pelvic malignancies: a pilot study. *J Magn Reson Imaging*, 2008. 27(3): p. 666-72.
706. **Gebker, R., et al.**, Comparison of different MRI techniques for the assessment of thoracic aortic pathology: 3D contrast enhanced MR angiography, turbo spin echo and balanced steady state free precession. *Int J Cardiovasc Imaging*, 2007. 23(6): p. 747-56.
707. **Utsunomiya, D., et al.**, Clinical role of non-contrast magnetic resonance angiography for evaluation of renal artery stenosis. *Circ J*, 2008. 72(10): p. 1627-30.

708. **Wilson, G.J. and J.H. Maki**, Non-contrast-enhanced MR imaging of renal artery stenosis at 1.5 tesla. *Magn Reson Imaging Clin N Am*, 2009. 17(1): p. 13-27.
709. **Smith, E.E., et al.**, MRI markers of small vessel disease in lobar and deep hemispheric intracerebral hemorrhage. *Stroke*, 2010. 41(9): p. 1933-8.
710. **Polak, J.F., et al.**, Right ventricular ejection fraction: an indicator of increased mortality in patients with congestive heart failure associated with coronary artery disease. *J Am Coll Cardiol*, 1983. 2(2): p. 217-24.
711. **Gloger, O., et al.**, A fully automatic three-step liver segmentation method on LDA-based probability maps for multiple contrast MR images. *Magn Reson Imaging*, 2010. 28(6): p. 882-97.
712. **Huang, B.Y. and M. Castillo**, Neurovascular imaging at 1.5 tesla versus 3.0 tesla. *Magn Reson Imaging Clin N Am*, 2009. 17(1): p. 29-46.
713. **Huppertz, H.J., et al.**, Intra- and interscanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures. *Neuroimage*. 49(3): p. 2216-24.
714. **Kruggel, F., J. Turner, and L.T. Muftuler**, Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. *Neuroimage*. 49(3): p. 2123-33.
715. **Saam, T., et al.**, Reader and platform reproducibility for quantitative assessment of carotid atherosclerotic plaque using 1.5T Siemens, Philips, and General Electric scanners. *J Magn Reson Imaging*, 2007. 26(2): p. 344-52.
716. **Theisen, D., et al.**, Myocardial perfusion imaging with Gadobutrol: a comparison between 3 and 1.5 Tesla with an identical sequence design. *Invest Radiol*, 2007. 42(7): p. 499-506.
717. **Decramer, S., et al.**, Urine in clinical proteomics. *Molecular & Cellular Proteomics*, 2008. 7(10): p. 1850.
718. **Thongboonkerd, V.**, Urinary proteomics: towards biomarker discovery, diagnostics and prognostics. *Molecular BioSystems*, 2008. 4(8): p. 810-815.
719. **Ward, D.G., et al.**, Proteomic profiling of urine for the detection of colon cancer. *Proteome Science*, 2008. 6(1): p. 19.
720. **Zürbig, P., et al.**, A low molecular weight urinary proteome profile of human kidney aging. *Proteomics*, 2009. 9(8): p. 2108.
721. **Goodson, J.M., et al.**, Is obesity an oral bacterial disease? *Journal of dental research*, 2009. 88(6): p. 519.
722. **Carvalho, A.L., et al.**, Evaluation of promoter hypermethylation detection in body fluids as a screening/diagnosis tool for head and neck squamous cell carcinoma. *Clinical Cancer Research*, 2008. 14(1): p. 97.
723. **Hu, S., et al.**, Salivary proteomics for oral cancer biomarker discovery. *Clinical Cancer Research*, 2008. 14(19): p. 6246.
724. **Hu, S., et al.**, Salivary proteomic and genomic biomarkers for primary Sjögren's syndrome. *Arthritis & Rheumatism*, 2007. 56(11): p. 3588-3600.
725. **John, M.S., et al.**, IL-6 and IL-8: potential biomarkers for oral cavity and oropharyngeal SCCA. *Otolaryngology-Head and Neck Surgery*, 2004. 131(2): p. P112-P113.
726. **Li, Y., et al.**, Salivary transcriptome diagnostics for oral cancer detection. *Clinical Cancer Research*, 2004. 10(24): p. 8442.

727. **Ryu, O.H., et al.**, Identification of parotid salivary biomarkers in Sjogren's syndrome by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry and two-dimensional difference gel electrophoresis. *Rheumatology*, 2006. 45(9): p. 1077.
728. **Streckfus, C. and L. Bigler**, The use of soluble, salivary c-erbB-2 for the detection and post-operative follow-up of breast cancer in women: the results of a five-year translational research study. *Advances in Dental Research*, 2005. 18(1): p. 17.
729. **Baker, P.I., D.R. Love, and L.R. Ferguson**, Role of gut microbiota in Crohns disease. *Expert Review of Gastroenterology and Hepatology*, 2009. 3(5): p. 535-546.
730. **Isolauri, E., et al.**, Obesity—Extending the Hygiene Hypothesis. *Microbial host-interaction: tolerance versus allergy*, 2009: p. 75.
731. **Possemiers, S., et al.**, The intestinal environment in health and disease—recent insights on the potential of intestinal bacteria to influence human health. *Current Pharmaceutical Design*, 2009. 15(18): p. 2051-2065.
732. **Musso, G., R. Gambino, and M. Cassader**, Gut microbiota as a regulator of energy homeostasis and ectopic fat deposition: mechanisms and implications for metabolic disorders. *Current opinion in lipidology*, 2010. 21(1): p. 76.
733. **Musso, G., R. Gambino, and M. Cassader**, Obesity, Diabetes, and Gut Microbiota. *Diabetes care*, 2010. 33(10): p. 2277.
734. **Andoh, A., et al.**, Recent advances in molecular approaches to gut microbiota in inflammatory bowel disease. *Current Pharmaceutical Design*, 2009. 15(18): p. 2066-2073.
735. **Jansson, J., et al.**, Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS one*, 2009. 4(7): p. e6386.
736. **Tuohy, K.M., et al.**, Studying the human gut microbiota in the trans-omics era—focus on metagenomics and metabolomics. *Current Pharmaceutical Design*, 2009. 15(13): p. 1415-1427.
737. **Williams, H.R.T., et al.**, Characterization of inflammatory bowel disease with urinary metabolic profiling. *The American journal of gastroenterology*, 2009. 104(6): p. 1435-1444.
738. **Jia, W., et al.**, Gut microbiota: a potential new territory for drug targeting. *Nature Reviews Drug Discovery*, 2008. 7(2): p. 123-129.
739. **Wos-Oxley, M., et al.**, A poke into the diversity and associations within human anterior nares microbial communities. *The ISME Journal*, 2010. 4(7): p. 839-851.
740. **Rathmann, W., et al.**, Hemoglobin A1c and glucose criteria identify different subjects as having type 2 diabetes in middle-aged and older populations: The KORA S4/F4 Study. *Annals of Medicine*, 2010(0): p. 1-8.
741. **Tiller, D., Kluttig A, B Schumann B, Werdan K, Haerting J, Greiser K.**, Response der CARLA-Follow-up Untersuchung (2007-2010). *Gesundheitswesen*, 2010. 72.
742. **Frost, C. and I.R. White**, The effect of measurement error in risk factors that change over time in cohort studies: do simple methods overcorrect for'regression dilution'? *International journal of epidemiology*, 2005. 34(6): p. 1359.

743. **Boshuizen, H.C., et al.**, Effects of past and recent blood pressure and cholesterol level on coronary heart disease and stroke mortality, accounting for measurement error. *American journal of Epidemiology*, 2007. 165(4): p. 398.
744. **Bartlett, J., et al.**, RE: "Effects of past and recent blood pressure and cholesterol level on coronary heart disease and stroke mortality, accounting for measurement error". *American journal of Epidemiology*, 2008. 167(4): p. 502.
745. **Rosner, B., D. Spiegelman, and W.C. Willett**, Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *American journal of Epidemiology*, 1990. 132(4): p. 734.
746. **Erbel, R., et al.**, Coronary Risk Stratification, Discrimination, and Reclassification Improvement Based on Quantification of Subclinical Coronary Atherosclerosis: The Heinz Nixdorf Recall Study. *Journal of the American College of Cardiology*, 2010.
747. **Bergmann, M.M., U. Bussas, and H. Boeing**, Follow-up procedures in EPIC-Germany—data quality aspects. *Annals of Nutrition and Metabolism*, 2000. 43(4): p. 225-234.
748. **Hoffmann, W., C. Bobrowski, and K. Fendrich**, Secondary data analysis in the field of epidemiology of health care. Potential and limitations. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 2008. 51(10): p. 1193.
749. **Boedeker, W., et al.**, The impact of work on morbidity-related early retirement. *Journal of Public Health*, 2008. 16(2): p. 97-105.
750. **Andersohn, F. and E. Garbe**, Pharmacoepidemiological research with large health databases. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 2008. 51(10): p. 1135.
751. Bundesdatenschutzgesetz in der Fassung der Bekanntmachung vom 14. Januar 2003 (BGBl. I S. 66), das zuletzt durch Artikel 1 des Gesetzes vom 14. August 2009 (BGBl. I S. 2814) geändert worden ist (http://www.gesetze-im-internet.de/bundesrecht/bdsg_1990/gesamt.pdf. Last access 10.2010).
752. Gesetz zum Schutz des Bürgers bei der Verarbeitung seiner Daten (Landesdatenschutzgesetz – DSGVO M-V) vom 28. März 2002 (GVObI. M-V S. 154), Zuletzt geändert am 25. Oktober 2005 (GVObI. M-V S. 535): http://www.lfd.mv.de/dschutz/ges_ver/guv/guv_c_20.html, Last access 10.2010.
753. Meldegesetz für das Land Mecklenburg-Vorpommern (Landesmeldegesetz - LMG) in der Fassung der Bekanntmachung vom 30. Januar 2007 (GVObI. M-v 2007, S. 34): http://mv.juris.de/mv/MG_MV_2007_rahmen.htm, Last access 10.2010.
754. ITIL Lifecycle Publication Suite. 2007.
755. BSI-Standard 100-2: IT-Grundschutz-Vorgehensweise, Version 2.5. 2008, Bonn *Bundesamt für Sicherheit in der Informationstechnik (BSI)*.
756. BSI-Standard 100-2: IT-Grundschutz-Vorgehensweise, Version 2.5. 2008 [cited 2010 October]; Available from: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/ITGrundschutzstandards/standard_1002.pdf?__blob=publicationFile.

757. BSI-Standard 100-3: Risikoanalyse auf Basis von IT-Grundschutz, Version 2.0. 2008 [cited 2010 October]; Available from: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/ITGrundschutzstandards/standard_1003.pdf?__blob=publicationFile.
758. IT-Grundschutz-Kataloge. 11. Ergänzungslieferung - November 2009. 2009 [cited 2010 October]; Available from: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/Download/it-grundschutz-kataloge_2009_EL11_de.pdf?__blob=publicationFile.
759. **Angelow, A., et al**, Methods and implementation of a central biosample and data management in a three-centre clinical study. *Computer Methods and Programs in Biomedicine*, 2008. 91 p. 82-90.
760. **Benson, T.**, Principles of Health Interoperability HL7 and SNOMED. 2009: *Springer Verlag*.
761. **Cartlidge, A., et al.** (2007) An Introductory Overview of ITIL® V3, Version 1.0.
762. **Chroust, K., et al.**, Experience in data management of the clinical retrospective project in Czech and Slovak oncology centres (IKARUS Project). *Klinická onkologie : casopis České a Slovenské onkologické společnosti*, 2009. 22: p. 163-167.
763. **Eckel, B.**, Thinking in Java: The definitive introduction to object-oriented programming in the language of the world wide web. 4th ed. 2006: *Prentice Hall International*.
764. **Frankel, D.**, Model Driven Architecture: Applying MDA to Enterprise Computing. 2003: *John Wiley & Sons*.
765. **Glock, J., R. Herold, and K. Pommerening**, Personal identifiers in medical research networks: Evaluation of the personal identifier generator in the Competence Network Paediatric Oncology and Haematology. *GMS Med Inform Biom Epidemiol*, 2006
766. **Harris, P.A., et al.**, Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 2009(42): p. 377-381.
767. **Hoffmann, W., U. Latza, and C. Terschüren**, Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP) – überarbeitete Auflage nach Evaluation, mit Ergänzungen durch Ausführungsbestimmungen zur Guten Praxis Sekundärdaten Analysen (GPS). 2008.
768. **Hoffmann, W., U. Latza, and C. Terschüren**, Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP) – überarbeitete Auflage nach Evaluation, mit Ergänzungen durch Ausführungsbestimmungen zur Guten Praxis Sekundärdaten Analysen (GPS). 2008.
769. **Holdener, A.T., et al.**, Ajax. The Definitive Guide. 1st ed. 2008: *O'Reilly Media*.
770. **Kriegel, A.**, SQL Bible. 2nd ed. 2008: *Wiley*.
771. **Loney, K.**, Oracle Database 11G: The Complete Reference. Har/Pas edition 2009: *Mcgraw-Hill Professional*.
772. **Meyer, J., et al.**, Implementation of an Efficient Data Management within a Large-Scale Population-Based Medical Research Project. . *Submitted to: Computer Methods and Programs in Biomedicine*, 2010.

773. **Ostrowski, C. and B.D. Brown**, Oracle Application Server 10g Web Development: Build and Deploy Web-Based Applications in the Oracle Environment. (illustrated edition). 2005: *Mcgraw-Hill Professional*.
774. **Paek, S.H., et al.**, Data management system in a movement disorder center: technical report. *Stereotactic and functional neurosurgery*, 2010. 88 p. 216-223.
775. **Pavlović, I. and D. Miklavcic**, Web-based electronic data collection system to support electrochemotherapy clinical trial. *IEEE transactions on information technology in biomedicine*, 2007(11): p. 222-230.
776. **Pianykh, O.S.**, Digital Imaging and Communications in Medicine (DICOM): A practical introduction and survival guide. 2008: *Springer Verlag*.
777. **Rainsberger, J.B.**, Junit Recipes: Practical Methods for Programmer Testing (illustrated edition). 2004: *Manning*.
778. **Reng, C.-M., et al.**, Generische Lösungen zum Datenschutz für die Forschungsnetze in der Medizin. *Medizinische Wissenschaftliche Verlagsgesellschaft*, 2006.
779. **Schlösser-Faßbender, M., et al.**, Pseudonymisierungsdienst für wissenschaftliche Forschungsnetze – Spezifikation http://www.tmf-ev.de/Themen/Projekte/V000_01_PSD/tabid/303/EntryId/2525/Command/Core_Download/Method/attachment/Default.aspx, Last access 10.2010.
780. **Thwin, S.S., et al.**, Automated inter-rater reliability assessment and electronic data collection in a multi-center breast cancer study. *BMC medical research methodology*, 2007. 7: p. 23.
781. **Diggle, P.J., I. Sousa, and A.G. Chetwynd**, Joint modelling of repeated measurements and time-to-event outcomes: the fourth Armitage lecture. *Statistics in medicine*, 2008. 27(16): p. 2981-2998.
782. **Fine, J.P. and R.J. Gray**, A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*, 1999. 94(446): p. 496-497.
783. **Aalen, O.O., Ø. Borgan, and H.K. Gjessing**, Survival and event history analysis: a process point of view. 2008: *Springer Verlag*.
784. **Begg, M. and M. Parides**, Separation of individual level and cluster level covariate effects in regression analysis of correlated data. *Statistics in medicine*, 2003. 22(16): p. 2591-2602.
785. **Ferrari, P., et al.**, A Bayesian multilevel model for estimating the diet/disease relationship in a multicenter study with exposures measured with error: the EPIC study. *Statistics in medicine*, 2008. 27(29): p. 6037-6054.
786. **Sheppard, L., R. Prentice, and M. Rossing**, Design considerations for estimation of exposure effects on disease risk, using aggregate data studies. *Statistics in medicine*, 1996. 15(17): p. 1849-1858.
787. **Haneuse, S.**, Hierarchical models for combining ecological and case-control data. *Biometrics*, 2007. 63(1): p. 128-136.
788. **Jackson, C., S. Richardson, and N. Best**, Studying place effects on health by synthesising individual and area-level outcomes. *Social Science & Medicine*, 2008. 67(12): p. 1995-2006.
789. **Jackson, C., N. Best, and S. Richardson**, Improving ecological inference using individual level data. *Statistics in medicine*, 2006. 25(12): p. 2136-2159.

References

790. **Davis, C.E., et al.**, A single cholesterol measurement underestimates the risk of coronary heart disease: an empirical example from the Lipid Research Clinics Mortality Follow-up Study. *Jama*, 1990. 264(23): p. 3044.
 791. **Gardner, M.J. and J.A. Heady**, Some effects of within-person variability in epidemiological studies. *Journal of Chronic Diseases*, 1973. 26(12): p. 781-795.
 792. **Hughes, M.D. and S.J. Pocock**, Within-subject diastolic blood pressure variability: implications for risk assessment and screening. *Journal of clinical epidemiology*, 1992. 45(9): p. 985-998.
 793. **MacMahon, S., et al.**, Blood pressure, stroke, and coronary heart disease* 1:: Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *The Lancet*, 1990. 335(8692): p. 765-774.
 794. **Wood, A.M., et al.**, Regression dilution methods for meta-analysis: assessing long-term variability in plasma fibrinogen among 27,247 adults in 15 prospective studies. *Int J Epidemiol*, 2006. 35(6): p. 1570-8.
 795. **Klerk, N., D.R. English, and B.K. Armstrong**, A review of the effects of random measurement error on relative risk estimates in epidemiological studies. *International journal of epidemiology*, 1989. 18(3): p. 705.
 796. **Clarke, R., et al.**, Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. *American journal of Epidemiology*, 1999. 150(4): p. 341.
 797. **Rosner, B., D. Spiegelman, and W.C. Willett**, Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *American journal of Epidemiology*, 1992. 136(11): p. 1400.
 798. **Rosner, B., W.C. Willett, and D. Spiegelman**, Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in medicine*, 1989. 8(9): p. 1051-1069.
 799. **Benichou, J. and M. Palta**, Rates, risks, measures of association and impact. *Handbook of Epidemiology*, 2007: p. 91-156.
 800. **Lopez, A.D.**, Global burden of disease and risk factors. 2006: *Oxford University Press, USA*.
 801. **Terschüren, C., et al.**, Health status of 'Ruhr-City' in 2025—predicted disease burden for the metropolitan Ruhr area in North Rhine-Westphalia. *The European Journal of Public Health*, 2009. 19(5): p. 534.
 802. **Hothorn, T. and P. Bühlmann**, Model-based boosting in high dimensions. *Bioinformatics*, 2006. 22(22): p. 2828.
 803. **Harrell, F.E.**, Regression modeling strategies. 2001: *Springer New York*.
 804. **Hanley, J.A. and B.J. McNeil**, A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 1983. 148(3): p. 839.
 805. **Pencina, M.J., et al.**, Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in medicine*, 2008. 27(2): p. 157-172.
 806. **Benichou, J. and M.H. Gail**, Estimates of absolute cause-specific risk in cohort studies. *Biometrics*, 1990. 46(3): p. 813-826.
-

807. **Gail, M.H.**, Estimation and interpretation of models of absolute risk from epidemiologic data, including family-based studies. *Lifetime Data Analysis*, 2008. 14(1): p. 18-36.
808. **Pepe, M.S., et al.**, Phases of biomarker development for early detection of cancer. *Journal of the National Cancer Institute*, 2001. 93(14): p. 1054.
809. **Husmann, G., et al.**, Krebs in Deutschland 2005/2006: Häufigkeiten und Trends. Berlin: Robert Koch-Institut, 2010.
810. **Lachin, J.M.**, Sample size evaluation for a multiply matched case-control study using the score test from a conditional logistic (discrete Cox PH) regression model. *Statistics in medicine*, 2008. 27(14): p. 2509-2523.
811. **Qi, L., Y.F. Wang, and Y. He**, A comparison of multiple imputation and fully augmented weighted estimators for Cox regression with missing covariates. *Statistics in medicine*, 2010.
812. **Becher, H.**, General principles of data analysis: continuous covariables in epidemiological studies. *Handbook of Epidemiology*, 2004: p. 595-624.
813. **Royston, P. and W. Sauerbrei**, Multivariable model-building. 2008: John Wiley.
814. **Hsieh, F.Y., D.A. Bloch, and M.D. Larsen**, A simple method of sample size calculation for linear and logistic regression. *Statistics in medicine*, 1998. 17(14): p. 1623-1634.
815. **Gauderman, W.J. and J.M. Morrison**, QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies. *Internet: <http://hydra.usc.edu/gxe>*, 2006.
816. **García-Closas, M. and J.H. Lubin**, Power and sample size calculations in case-control studies of gene-environment interactions: comments on different approaches. *American journal of Epidemiology*, 1999. 149(8): p. 689.
817. **Pepe, M.S., et al.**, Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *Journal of the National Cancer Institute*, 2008. 100(20): p. 1432.
818. **Menon, U., et al.**, Results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Lancet Oncology*, 2009. 10: p. 327-340.
819. Deutscher Ethikrat: Humanbiobanken für die Forschung. Stellungnahme. *Deutscher Ethikrat, Berlin, 2010.*
820. **Reng, C., et al.**, im Auftrag des Koordinierungsrates der Telematikplattform für Medizinische Forschungsnetze: Generische Lösungen zum Datenschutz für die Forschungsnetze in der Medizin. Schriftenreihe der Telematikplattform für Medizinische Forschungsnetze. *Schriftenreihe der Telematikplattform für Medizinische Forschungsnetze Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin, 2006.*
821. Deutschen Region der Internationalen Biometrischen Gesellschaft (DR-IBS): Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP). *Arbeitsgruppe Epidemiologische Methoden der Deutschen Arbeitsgemeinschaft für Epidemiologie (DAE) In Zusammenarbeit mit der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS), Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) 2004.*

References

822. **Hegenscheid, K., et al.** Whole-Body Magnetic Resonance Imaging of Healthy Volunteers: Pilot Study Results from the Population-Based SHIP Study= Ganzkörper-Magnetresonanztomografie von gesunden Probanden: Pilot-studienergebnisse der populationsbasierten SHIP Studie. 2009: *Thieme*.
823. **Smit AJ, et al.** Advanced glycation endproducts in chronic heart failure. *Ann N Y Acad Sci* 2008; 1126: 225-30.
824. **Kaaks, R., E. Riboli, and W. Van Staveren,** Sample size requirements for calibration studies of dietary intake measurements in prospective cohort investigations. *American journal of epidemiology*, 1995. 142(5): p. 557.
825. **Plummer, M., D. Clayton, and R. Kaaks,** Calibration in multi-centre cohort studies. *International journal of epidemiology*, 1994. 23(2): p. 419.

