Bundesgesundheitsbl
DOI 10.1007/s00103-015-2182-x

Wolfgang Ahrens[1] · **K.-H. Jöckel**[2]

[1] Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS, Bremen, Bremen, Deutschland

[2] Institut für Medizinische Informatik, Biometrie und Epidemiologie, Universitätsklinikum Essen, Essen, Deutschland

Der Nutzen großer Kohortenstudien für die Gesundheitsforschung am Beispiel der Nationalen Kohorte

Gegenwärtig vollzieht sich ein Paradigmenwechsel in der Gesundheitsforschung, bei dem neben der Optimierung der Diagnostik und Therapie von Krankheiten oder der Verbesserung der Krankenversorgung die Vorbeugung immer größere Beachtung findet. Aufgrund des demografischen Wandels wird die Anzahl der von chronischen Erkrankungen Betroffenen in unserer Bevölkerung in den kommenden Jahrzehnten deutlich zunehmen [1]. Damit verbunden ist auch ein entsprechender Anstieg der Kosten im Gesundheitswesen, die allein bei stationären Patienten mit Herzversagen in Deutschland bis 2025 um 50 % auf 1,8 Mrd. € zunehmen werden [2]. Dies hat eine umfangreiche Debatte zur Priorisierung, also zur Rationierung medizinischer Versorgungsleistungen, im Gesundheitswesen ausgelöst [3]. Parallel dazu nimmt die Evidenz zu, dass eine verbesserte Krankheitsprävention nicht nur kosteneffizient ist, sondern unter dem Strich sogar Kosten einspart [4, 5]. Ausgehend von einer Kosten-Effizienzanalyse für den Bereich kardiovaskulärer Erkrankungen gab die American Heart Association die Empfehlung, dass nicht nur wegen der erwarteten Kostenexplosion in diesem Bereich, sondern auch im Interesse einer gesünderen und produktiveren Gesellschaft die primäre Prävention einen höheren Stellenwert bekommen muss [6]. Epidemiologische Studien werden damit immer mehr zu einem unverzichtbaren Bestandteil der Gesundheitsforschung, da sie sowohl wirksame Ansatzpunkte für eine bevölkerungsbasierte Prävention identifizieren als auch die wissenschaftliche Evidenz für die Effektivität von entsprechenden Interventionsmaßnahmen liefern können. Im vorliegenden Beitrag wird die Rolle der Epidemiologie innerhalb der Gesundheitsforschung am Beispiel der bisher größten deutschen Gesundheitsstudie, der Nationalen Kohorte, dargestellt, und es wird erläutert, worin sich diese Studie von dem, was im allgemeinen unter Big Data verstanden wird, unterscheidet.

Beobachtungsstudien in der Epidemiologie

Epidemiologische Beobachtungsstudien erlauben es nicht nur, die Verteilung und zeitliche Entwicklung von Erkrankungen in der Bevölkerung (Prävalenz) zu beschreiben. Sie ermöglichen auch, die Häufigkeit von Neuerkrankungen (Inzidenz) und ihren Verlauf zu ermitteln sowie ihre Ursachen auf Bevölkerungsebene zu erforschen. Insbesondere die Identifizierung und Quantifizierung von Gesundheitsgefahren für den Menschen ist der beobachtenden Epidemiologie vorbehalten, da sich schon allein aus ethischen Erwägungen die experimentelle Erforschung von Risikofaktoren am Menschen verbietet. Das wichtigste Ziel epidemiologischer Beobachtungsstudien ist die Identifizierung von Ansatzpunkten zur Vermeidung oder Früherkennung von Erkrankungen, also die primäre oder sekundäre Prävention. Sie liefern aber auch wichtige Hinweise, die ihren Weg in die Behandlung finden können. Die drei wichtigsten epidemiologischen Studientypen werden im Folgenden jeweils kurz beschrieben (für eine detaillierte Einführung siehe [7]).

Epidemiologische Studiendesigns

Querschnittstudien

Querschnittstudien, auch Prävalenzstudien genannt, stellen ein einfaches, schnell durchführbares Studiendesign dar, bei dem typischerweise eine zufällig ausgewählte, klar definierte Bevölkerungsgruppe hinsichtlich der Häufigkeit von Krankheiten, Krankheitssymptomen, bekannten Einflussfaktoren oder dem Gesundheitsverhalten untersucht wird. In den meisten Fällen werden diese Informationen mittels standardisierter Fragebögen von den Personen der untersuchten Zielgruppe selbst berichtet. Auf diese Weise werden Daten zum augenblicklichen Gesundheitszustand einer Bevölkerung im Sinne einer Momentaufnahme, z. B. für die Gesundheitsberichterstattung oder für die Abschätzung des Versorgungsbedarfs, ermittelt (**Abb. 1**). Krankheitsverläufe oder Krankheitsursachen lassen sich mit einer solchen Momentaufnahme aufgrund des Fehlens einer zeitlichen Abfolge jedoch kaum erforschen.

Fall-Kontrollstudien

Fall-Kontrollstudien versuchen diesen Nachteil dadurch zu beheben, dass sie Personen, die bereits an einer bestimmten Erkrankung leiden (die Fälle), mit Kontrollpersonen vergleichen, die (noch) nicht erkrankt sind. Dieser Vergleich richtet sich auf Faktoren (Expositionen),

Published online: 16 June 2015

Abb. 1 ◀ Querschnittsstudie: In einer Momentaufnahme werden die Prävalenzen von Erkrankungen und von krankheitsbezogenen Faktoren (Expositionen) in einer Bevölkerungsstichprobe (Auswahlbevölkerung) ermittelt

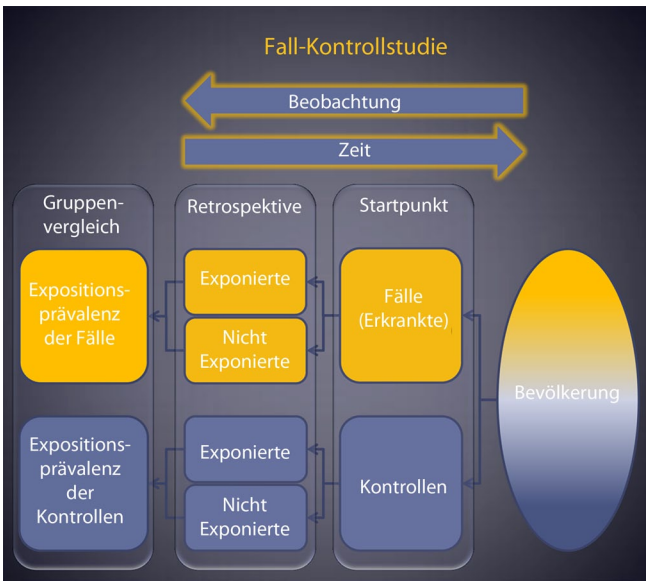

Abb. 2 ◀ Fall-Kontrollstudie: Die Prävalenzen zurückliegender krankheitsbezogener Faktoren (Expositionen) werden zwischen Erkrankten (Fällen) und Vergleichspersonen aus der gleichen Ursprungsbevölkerung (Kontrollen) verglichen

die unter Verdacht stehen, das Erkrankungsrisiko zu erhöhen, und die bereits vor Eintritt der Erkrankung aufgetreten sind. Zur Ermittlung dieser Faktoren ist die Epidemiologie typischerweise auf Angaben der untersuchten Personen angewiesen, die durch Interviews oder selbst auszufüllende Fragebögen erhoben werden. Dies hat den Vorteil, dass viele mögliche Krankheitsursachen gleichzeitig erforscht werden können. Ist eine unter Verdacht stehende Exposition in der Vergangenheit unter den Fällen häufiger aufgetreten als unter den Kontrollpersonen, so legt dies den Schluss nahe, dass diese Exposition das Erkrankungsrisiko erhöht hat (**Abb. 2**). Da die erhobenen Angaben jedoch rückwirkend, also retrospektiv ermittelt werden müssen, besteht die Gefahr, dass sie durch Erinnerungslücken unzuverlässig sind oder sogar dadurch verzerrt werden, dass die Fälle ein anderes Antwortverhalten zeigen als die Kontrollpersonen. Dies ist besonders dann von Nachteil, wenn viele Jahre zwischen Krankheitsursache und nachfolgender Erkrankung vergehen, wie das z. B. bei Krebserkrankungen die Regel ist.

Kohortenstudien

Kohortenstudien, auch Längsschnittstudien oder Prospektivstudien genannt, haben den Vorteil, dass sie die Abfolge zwischen einer Krankheitsursache (Exposition) und dem Neuauftreten einer Erkrankung eindeutig beschreiben und die Ergebnisse weniger durch Erinnerungslücken verzerrt werden können, insbesondere dann nicht, wenn sowohl die Expositionserfassung als auch die Diagnose der Erkrankung auf objektiven (Mess-)Daten basieren. In ihrer einfachsten Form beginnen Kohortenstudien mit einer Gruppe von exponierten Personen, der so genannten Kohorte, z. B. der Belegschaft eines Betriebes, die einem bestimmten Schadstoff ausgesetzt ist, in der dann nachfolgend (prospektiv) das Neu-Auftreten verschiedener Erkrankungen ermittelt wird. Eine solche Kohorte lässt sich als expositionsbasierte Kohorte charakterisieren (**Abb. 3**). Die Exposition wird dabei vor dem Auftreten und damit unabhängig von einer späteren Erkrankung erfasst. Die Inzidenz verschiedener Erkrankungen kann dann ihrer Inzidenz in einer Vergleichsbevölkerung gegenübergestellt werden, die dieser Exposition nicht oder in erheblich geringerem Maße ausgesetzt war. Eine erhöhte Krankheitsinzidenz in der exponierten Bevölkerung lässt – sofern biologisch plausibel und nicht durch die Beteiligung eines Störfaktors erklärbar – auf einen ursächlichen Zusammenhang mit der Exposition schließen.

Expositionsbasierte Kohortenstudien können zwar verschiedene Krankheitsendpunkte gleichzeitig untersuchen, sind aber bezüglich der untersuchten Exposition festgelegt. Die Konzentration auf eine Exposition hat darüber hinaus den Nachteil, dass das Zusammenwirken verschiedener Krankheitsursachen nicht erforschbar ist. Es kann sogar zu einer fälschlichen Zuschreibung einer Ursache-Effekt-Beziehung kommen, wenn ein relevanter Störfaktor (Confounder) nicht erfasst werden konnte. Ein klassisches Beispiel hierfür ist z. B. das Zigarettenrauchen, das in vielen Industriebereichen unter Exponierten häufiger vorkommt als im Durchschnitt der Bevölkerung. In diesen Branchen treten Krankheiten, die durch Rauchen verursacht werden, gehäuft auf, sodass das Rauchen in der statistischen Datenanalyse berücksichtigt werden muss, um berufliche Ursachen dieser Erkrankungen herausfiltern zu können.

Diese Schwachstelle tritt in Vielzweck-Kohortenstudien in geringerem Ausmaß auf oder kann sogar ganz vermieden werden. Bei diesem Ansatz werden gleich zu Beginn – ähnlich wie in einer Querschnittstudie – viele verschiedene Lebensstil- und Umweltfaktoren für jeden einzelnen Studienteilnehmer ermittelt, um sie mit dem späteren Krankheitsauftreten in Beziehung setzen zu können (**Abb. 4**). Das Spektrum der zu erfassenden Faktoren muss in solchen Studien sehr breit ge-

fasst sein. Neben bereits bekannten Risikofaktoren, die als z. B. Confounder in Betracht kommen können oder die in ihrer Wirkung mit anderen Krankheitsursachen in einer Wechselbeziehung stehen können, muss eine Vielzahl möglicher Krankheitsursachen ermittelt werden, die jeweils für eine oder verschiedene Erkrankungen relevant sein können. Je kleiner eine Kohorte ist, umso geringer ist die Anzahl an Krankheitsfaktoren und ihrer Wechselwirkungen, die mit ausreichend großer Häufigkeit in der Kohorte vorkommt, um wissenschaftlich untersuchbar zu sein.

Der besondere Vorteil des prospektiven Designs von Kohortenstudien muss damit erkauft werden, dass die Studien für Erkrankungen, die eine lange Entstehungsgeschichte haben und deren Ursachen weit zurück in der Vergangenheit einer erkrankten Person gesucht werden müssen, langfristig mit einer Beobachtungsdauer von mehreren Jahrzehnten angelegt sein müssen. Dies trifft für die Mehrzahl der chronischen Erkrankungen zu, die in den modernen Industriegesellschaften das Krankheits- und Sterbegeschehen dominieren. Je kleiner eine Kohorte ist, umso länger muss sie beobachtet werden, um eine ausreichende Anzahl dieser interessierenden Krankheitsereignisse beobachten zu können, und umso geringer ist die Anzahl verschiedener Erkrankungen, die häufig genug beobachtet werden können, um zu wissenschaftlich belastbaren Schlussfolgerungen zu gelangen. Ein spezielles methodisches Problem kann sich durch die Einbeziehung prävalenter Expositionen ergeben, wie am Beispiel von Medikamentengebrauch beschrieben wurde [8]. Wenn z. B. die mit einer Behandlung verbundenen Risiken zeitveränderlich sind, wie z. B. nach einem chirurgischen Eingriff, so kann eine Verzerrung durch selektives Überleben (Survivor Bias) auftreten. Außerdem können sich gemeinsam bzw. als Folge der Behandlung die Kovariaten ändern, was bei Nichtberücksichtigung zu Confounding führen kann. Um die Gefahr für derartige Verzerrungen zu vermindern, ist die Studiengruppe ggf. auf Personen zu beschränken, die zu Beginn der Exposition bereits unter Beobachtung standen (New User Design) [8].

Zusammenfassung · Abstract

Bundesgesundheitsbl DOI 10.1007/s00103-015-2182-x

W. Ahrens · K.-H. Jöckel

Der Nutzen großer Kohortenstudien für die Gesundheitsforschung am Beispiel der Nationalen Kohorte

Zusammenfassung

Groß angelegte epidemiologische Vielzweck-Kohortenstudien mit langer Laufzeit erlauben aufgrund ihres prospektiven Charakters die Untersuchung von komplexen Krankheitsursachen, die Analyse von Krankheitspfaden und die Identifizierung von neuen präklinischen Krankheitszeichen. Die Nationale Kohorte (NAKO) ist eine bevölkerungsbezogene, hoch standardisierte und sehr detailliert phänotypisierte Kohorte. Sie soll die Basis für neue Strategien zur Ursachenforschung, Früherkennung und Prävention multifaktorieller Erkrankungen schaffen. Die NAKO ist die bisher größte bevölkerungsbezogene Kohortenstudie Deutschlands. Im Jahr 2014 hat die Untersuchung von insgesamt 200.000 Frauen und Männern im Alter von 20–69 Jahren in 18 Studienzentren begonnen. Die Studie ermöglicht die Erforschung der Ursachen chronischer Krankheiten im Zusammenhang mit Lebensgewohnheiten, genetischen, sozioökonomischen, psychosozialen und umweltbedingten Faktoren. Damit schafft die NAKO die Basis zur Entwicklung von Maßnahmen zur Vorbeugung und Früherkennung dieser Erkrankungen. Im Fokus stehen Erkrankungen des Herz-Kreislauf-Systems und der Atemwege, Krebs, Diabetes, neurodegenerative/-psychiatrische und muskuloskelettale Erkrankungen sowie Infektionskrankheiten. Aufgrund der schieren Größe der Studie stellt sich die Frage, in wieweit es sich hier um ein typisches Big-Data-Projekt handelt. Es wird abgeleitet, dass dies nicht der Fall ist.

Schlüsselwörter

Big Data · Epidemiologie · Früherkennung · Krankheitsursachen · Prävention

The benefit of large-scale cohort studies for health research: the example of the German National Cohort

Abstract

The prospective nature of large-scale epidemiological multi-purpose cohort studies with long observation periods facilitates the search for complex causes of diseases, the analysis of the natural history of diseases and the identification of novel pre-clinical markers of disease. The German National Cohort (GNC) is a population-based, highly standardised and in-depth phenotyped cohort. It shall create the basis for new strategies for risk assessment and identification, early diagnosis and prevention of multifactorial diseases. The GNC is the largest population-based cohort study in Germany to date. In the year 2014 the examination of 200,000 women and men aged 20–69 years started in 18 study centers. The study facilitates the investigation of the etiology of chronic diseases in relation to lifestyle, genetic, socioeconomic, psychosocial and environmental factors. By this the GNC creates the basis for the development of methods for early diagnosis and prevention of these diseases. Cardiovascular and respiratory diseases, cancer, diabetes, neurodegenerative/-psychiatric diseases, musculoskeletal and infectious diseases are in focus of this study. Due to its mere size, the study could be characterized as a Big Data project. We deduce that this is not the case.

Keywords

Big Data · Causes of disease · Early detection of disease · Epidemiology · Prevention

In der Regel werden neben den Befragungs- und Untersuchungsdaten auch biologische Materialien wie Blut, Urin oder Speichel gesammelt und für spätere Analysen tiefgefroren aufbewahrt. Diese Materialien können im Rahmen sogenannter eingebetteter Fall-Kontrollstudien effizient genutzt werden, indem die teilweise sehr teuren und Material verbrauchenden Laboranalysen später gezielt nur bei Erkrankten und einer zufällig aus der Kohorte ausgewählten Kontrollgruppe durchgeführt werden.

Hinsichtlich der Anzahl untersuchbarer Krankheitsursachen, der Wechselwirkungen zwischen ihnen, der Breite des erforschbaren Krankheitsspektrums, der erforderlichen Beobachtungsdauer und der Möglichkeit, Fall-Kontrollstudien einzu-

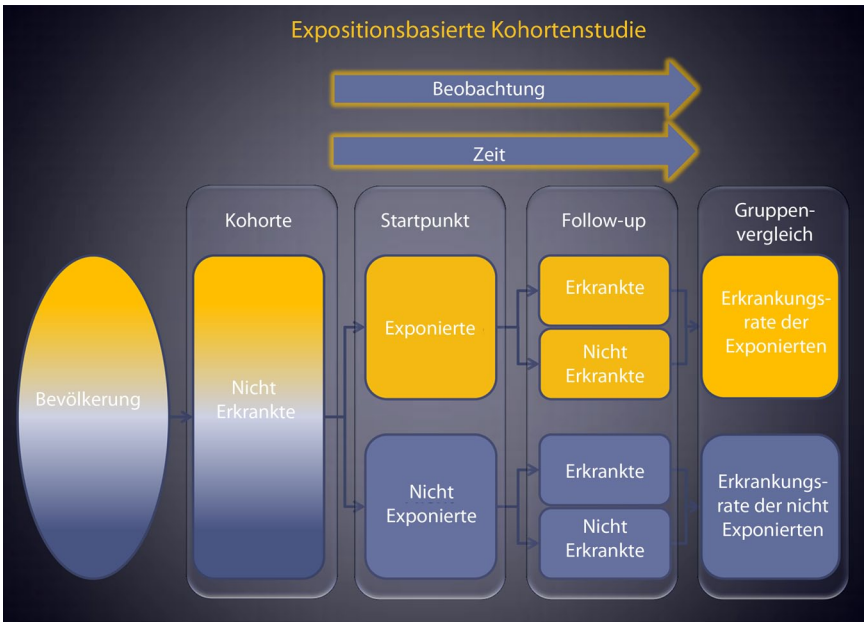

Abb. 3 ▲ Expositionsbasierte Kohortenstudie: Beginnend mit einer vermuteten Krankheitsursache (Exposition) wird im Rahmen einer (mehrjährigen) Nachbeobachtung (Follow-up) das Neuauftreten von Erkrankungen (Krankheitsrate) zwischen Exponierten und nicht Exponierten verglichen

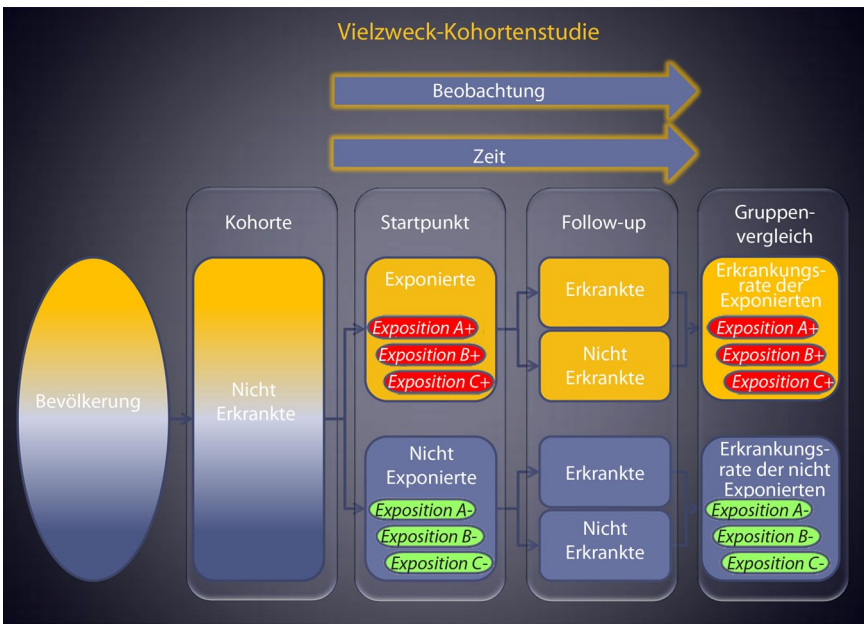

Abb. 4 ▲ Vielzweck-Kohortenstudie: Expositionen A, B und C werden im Querschnitt erfasst. Im Follow-up wird die Neuerkrankungsrate zwischen Exponierten (+) und nicht Exponierten (–) für verschiedene Kombinationen von A, B und C verglichen

betten, ist eine groß angelegte Kohortenstudie also von Vorteil.

Die Nationale Kohorte

Ziele und Design

Die Nationale Kohorte (NAKO) ist eine solche neue und groß angelegte Studie, die schon allein aufgrund ihrer Größe einen Erkenntnisfortschritt gegenüber früheren Kohorten kleineren Umfangs verspricht [9]. Es handelt sich um eine Vielzweckkohorte, die als Forschungsinfrastruktur aufgebaut wird, um die Ursachen, die Entstehung und den Verlauf der wichtigsten chronischen Erkrankungen, nämlich Herz-Kreislauf-Erkrankungen, Atemwegserkrankungen, Krebs, Diabetes, Erkrankungen des Muskel-/Skelettsystems, Infektionskrankheiten sowie neurodegenerativer und psychiatrischer Erkrankungen umfassend zu erforschen. Die Entstehung dieser Erkrankungen ist multifaktoriell. Daher wird eines der wichtigsten Ziele die Aufklärung des Zusammenwirkens von genetischen Faktoren, Umwelteinwirkungen, des sozialen Umfelds und des individuellen Lebensstils sein. Neben der Identifizierung von Risikofaktoren ist die Studie aber auch so angelegt, dass mit ihr funktionelle Störungen untersucht und Frühzeichen für das spätere Auftreten der genannten Erkrankungen entdeckt werden können. Zusammengenommen sollen aus den gewonnenen Erkenntnissen Strategien für eine bessere Vorbeugung und Behandlung der wichtigsten Volkskrankheiten abgeleitet sowie die Früherkennung von Krankheiten verbessert werden [10].

Die Nationale Kohorte wird 200.000 Frauen und Männer im Alter von 20–69 Jahren aus ganz Deutschland einschließen, die medizinisch untersucht und nach ihren Lebensgewohnheiten (z. B. körperliche Aktivität, Rauchen, Ernährung, Medikamenteneinnahme, Beruf) sowie weiteren Einflussgrößen (soziale Faktoren, Lebensqualität, Vorerkrankungen, Umwelt) befragt werden. Dabei werden alle Studienteilnehmer um biologische Proben gebeten. So werden Blut-, Urin-, Speichel- und Stuhlproben sowie ein Nasenschleimhautabstrich entnommen und für die spätere wissenschaftliche Untersuchung in einer zentralen Bioprobenbank – je nach Probenmaterial bei – 80 bis – 196 °C – kryokonserviert.

Die Entwicklung der verschiedenen Untersuchungs- und Befragungsmodule basiert auf einer umfangreichen Begleitforschung zu ihrer Machbarkeit im Rahmen zahlreicher Pretests [11], wobei insbesondere auch die Akzeptanz der Sammlung und Asservierung von Bioproben untersucht wurde [12]. Eine Übersicht über alle Module wurde kürzlich publiziert [10].

Fallzahlüberlegungen

Durch die Einlagerung von Bioproben, aber auch durch die Speicherung von Daten der Bildgebung, wird überdies die Möglichkeit geschaffen, zu einem späte-

Tab. 1 Im Rahmen einer eingebetteten Fall-Kontrollstudie aufdeckbare Risiken für verschiedene Matchingverhältnisse und Fallzahlen bei einer Irrtumswahrscheinlichkeit von 5 % ($\alpha = 0{,}05$) und einer statistischen Macht von 80 % ($1-\beta = 0{,}8$)

Expositions-prävalenz (Kontrollen)	Matching-verhältnis (Fälle:Kontrollen)	Fallzahl				
		100	500	1000	2000	5000
0,1 %	1:1			10,6	6,5	3,8
	1:4		10,0	6,3	4,3	2,9
1 %	1:1	11,0	3,8	2,8	2,2	1,7
	1:4	6,5	2,8	2,2	1,9	1,5
5 %	1:1	4,1	2,0	1,7	1,5	1,3
	1:4	3,0	1,8	1,5	1,4	1,3
10 %	1:1	3,0	1,7	1,5	1,4	1,2
	1:4	2,4	1,6	1,4	1,3	1,2
20 %	1:1	2,3	1,5	1,4	1,3	1,2
	1:4	2,0	1,4	1,3	1,2	1,2

ren Zeitpunkt innovative Biomarker oder neue Bildsequenzen, die heute noch gar nicht bekannt oder untersucht sind, hinsichtlich ihres prädiktiven Potenzials im Rahmen einer eingebetteten Fall-Kontrollstudie zu analysieren. Basierend auf den erwarteten Fallzahlen für die interessierenden Erkrankungen, die je nach Krankheitsgruppe und Beobachtungsdauer zwischen 100 und mehr als 10.000 Fällen variieren, lassen sich die im Rahmen eingebetteter Fall-Kontrollstudien aufdeckbaren Risiken abschätzen. In **Tab. 1** sind kleinsten noch aufdeckbaren Erkrankungsrisiken für verschiedene Fallzahlen dargestellt. So werden nach einer Beobachtungsdauer von fünf Jahren in der Kohorte ca. 110 maligne Eierstocktumoren erwartet. Wählt man für diese Fälle die vierfache Anzahl an Kontrollpersonen aus, lässt sich bei einer niedrigen Expositionsprävalenz von 1 % nur ein sehr hohes relatives Risiko von ungefähr 6,5 aufdecken. Bei einer Fallzahl von 500, wie sie z. B. für Nierenkarzinome nach 10 Jahren Beobachtungsdauer erwartet wird, lässt sich bei einer häufigen Exposition (20 % Prävalenz) und gleicher Anzahl von Fällen und Kontrollen (1:1-Matching) bereits eine 50 %ige Risikoerhöhung entdecken. Die Untersuchung des Risikos durch sehr seltene Expositionen (0,1 %) erfordert dagegen eine Fallzahl von mehreren Tausend.

Implementierung der Studie

Die NAKO wird von einem Netzwerk deutscher Forschungseinrichtungen aus der Helmholtz-Gemeinschaft, den Universitäten, der Leibniz-Gemeinschaft und der Ressortforschung getragen [13]. Deutschlandweit beteiligen sich 25 Forschungsinstitute am Aufbau dieser groß angelegten Langzeit-Bevölkerungsstudie. In 18 Studienzentren werden bis 2018 jeweils mindestens 10.000 Studienteilnehmer medizinisch untersucht und befragt, um anschließend ihre gesundheitliche Entwicklung über viele Jahre weiter zu beobachten. Bereits fünf Jahre nach der Basisuntersuchung werden alle Teilnehmer erneut zu einer Untersuchung und zweiten Befragung in die Studienzentren eingeladen. Im Laufe der Nachbeobachtung über einen Zeitraum von geplant 25–30 Jahren werden die bei einigen Teilnehmern aufgetretenen Erkrankungen mit den viele Jahre vor Erkrankungseintritt erhobenen Daten und biologischen Proben in Verbindung gebracht. Die wissenschaftliche Aussagekraft der Studie wird durch die geplante Verknüpfung mit vorhandenen Datenquellen wie z. B. ortsbezogenen Lärm- und Luftschadstoffmessungen vergrößert. Durch dieses Vorgehen bietet die Studie zukünftig eine Plattform für eine Vielzahl von wissenschaftlichen Untersuchungen, um die sich nicht nur wissenschaftliche Einrichtungen aus dem NAKO-Konsortium, sondern auch davon unabhängige Forschergruppen bewerben können. Die Regeln für die Datennutzung wurden in einer speziell für die Studie erarbeiteten Nutzungsordnung festgelegt [13].

Bezug zu anderen Kohortenstudien

Die Studie hat zahlreiche bedeutende internationale Kohortenstudien wie z. B. die Framingham Heart Study zum Vorbild, die bereits in der Vergangenheit wichtige Erkenntnisse zu den Ursachen von chronischen Erkrankungen erbracht haben [14]. Sie berücksichtigt auch die Erfahrungen verschiedener nationaler Kohorten kleineren Umfangs. Der Umfang der Befragungen und der biologischen Untersuchungen geht jedoch deutlich über das hinaus, was bisher in vergleichbaren Studien praktiziert wurde, und erlaubt so eine außergewöhnlich detaillierte Phänotypisierung der Studienteilnehmer mit modernsten, teilweise bildgebenden, Untersuchungsmethoden, die noch vor wenigen Jahren für solche groß angelegten Studien nicht zur Verfügung standen [10]. Vergleichbare epidemiologische Forschungsinfrastrukturen wurden parallel zur NAKO in anderen europäischen Ländern initiiert, wobei jede andere Schwerpunkte setzt. Hierzu zählen z. B. die UK-Biobank [15], die mit 500.000 Freiwilligen im Alter von 40 bis 69 Jahren einen Schwerpunkt auf die Untersuchung von Bioproben legt, die niederländische Life-Lines-Studie [16], die mit ca. 170.000 Teilnehmern im Alter von 6 bis 93 Jahren ganze Familien untersucht, sowie die französische CONSTANCES-Studie [17], die sich mit 200.000 Teilnehmern im Alter von 18 bis 69 Jahren besonders auf soziale und berufliche Krankheitsursachen sowie die Erforschung von Alterungsprozessen konzentriert. Auch die schwedische LifeGene-Studie, die einen Gesamtumfang von 500.000 Studienteilnehmern anstrebt, bezieht Partner und sogar Kinder der eingeladenen 18- bis 45-jährigen Wohnbevölkerung ein, um dadurch auch die gesundheitliche Entwicklung von in die Kohorte hineingeborenen Kindern unter Berücksichtigung prä- und perinataler Daten erforschen zu können [18]. Die NAKO ist so angelegt, dass auch sie die Einbettung einer Geburtskohorte erlauben wird. Schon bei der Studienplanung wurde darauf geachtet, dass die genannten Studien teilweise ähnliche Untersuchungselemen-

te beinhalten. Dies eröffnet die Möglichkeit, zukünftig spezielle Fragestellungen und selten auftretende Erkrankungen, die nur mit extrem großen Fallzahlen untersuchbar sind, durch Datenpooling in gemeinsamer Kooperation zu erforschen.

Beitrag der NAKO zur Gesundheitsforschung

Ursachenforschung

Die genaue Erfassung von Lebensstilfaktoren wie Rauchen, Alkoholkonsum, körperlicher Aktivität und Ernährung in Verbindung mit chronischen Infektionen, Medikamentenkonsum, sozioökonomischen oder psychosozialen Faktoren sowie Expositionen aus der Umwelt und dem Arbeitsleben deckt ein breites Spektrum möglicher Ursachen für die im Fokus der Studie stehenden Erkrankungen ab. Auch wenn die Krankheitsverursachung durch einzelne dieser Risikofaktoren bereits bekannt ist, so mangelt es doch häufig an einer genauen Quantifizierung ihres Beitrags zur Krankheitsentstehung und der Ermittlung des attributablen Anteils der Erkrankungen, der auf eine bestimmte Einwirkung zurückzuführen ist. Auch das Zusammenwirken der verschiedenen Faktoren untereinander und mit der individuell unterschiedlichen genetischen Prädisposition kann im Rahmen der Kohorte untersucht werden, um auf diese Weise z. B. besonders durch eine Exposition gefährdete Menschengruppen zu erkennen, die von gezielten Präventionsmaßnahmen profitieren könnten. Gleichzeitig kann die simultane Betrachtung von medizinischen und psychosozialen Faktoren (Persönlichkeitsmerkmale, Stress, soziales Umfeld) sowie der sozialen Position (Bildungsstand, beruflicher Status, Verdienst) dabei helfen, die Gründe für regionale Unterschiede im Auftreten der Erkrankungen sowie die Gründe für die bestehende soziale Ungleichheit im Krankheitsgeschehen aufzudecken. Eine besondere Stärke der Studie besteht darin, dass zur gleichen Person mehrere (funktionale) Messungen und Bioproben im Zeitverlauf gewonnen werden, sodass nicht nur der natürliche Verlauf der Krankheitsgenese, sondern auch der kausale Pfad ihrer Entstehung besser verstanden werden kann.

Prävention

Über die spezifische Identifizierung von kausalen Mechanismen der Krankheitsentstehung legt diese Ursachenforschung die Basis für gezielte und wirkungsvolle Präventionsmaßnahmen. Aufgrund der Kombination von Fragebogenangaben, medizinischen Untersuchungen und biologischen oder genetischen Markern aus Blut-, Urin- und Stuhlproben können hochkomplexe und umfassende Modelle zur Vorhersage des individuellen Erkrankungsrisikos entwickelt werden. Zukünftig könnten solche Modelle der personalisierten Prävention und – in der klinischen Praxis – der personalisierten Medizin dienen. Sofern entsprechende Interventionsmaßnahmen aus den Studienergebnissen abgeleitet werden, erfordern diese dann allerdings auch die weitere Evaluation auf Bevölkerungsniveau, um dem finalen Ziel einer evidenzbasierten Prävention zu dienen.

Früherkennung

Durch die Identifizierung von Krankheitsvorstufen und Prädiktoren für das spätere Auftreten von Erkrankungen kann die NAKO eine wertvolle Wissensbasis für die sekundäre Krankheitsprävention schaffen. Gerade die im Rahmen der Studie einzusetzenden modernen Untersuchungsmethoden, insbesondere die bildgebenden Verfahren wie MRT und Sonografie, sowie die heute schon angewendeten Technologien der molekularen Analyse von biologischen Materialien (Omics) eröffnen sehr vielversprechende Möglichkeiten, zukünftig bei den erkrankten Personen aus der Kohorte anhand der eingelagerten Proben und der zuvor erhobenen Messdaten nach Frühzeichen ihrer Erkrankung zu fahnden. Die auf diese Weise zu entdeckenden neuen Biomarker mit Vorhersagecharakter für spätere Erkrankungen könnten schnell Eingang in die klinische Praxis und das gezielte Screening von Risikogruppen in der Bevölkerung finden. Umgekehrt besteht aber auch die Möglichkeit, dass existierende oder potenzielle Früherkennungsmaßnahmen durch die Ergebnisse der Studie für obsolet erklärt werden und damit unnötige Kosten oder gar Gesundheitsschäden vermieden werden.

Gesundheitliche Versorgung

Durch die Erfassung der Inanspruchnahme medizinischer Leistungen in Kombination mit der Vielzahl der erhobenen medizinischen Variablen und Lebensstilfaktoren wird die NAKO die Versorgungsforschung in Deutschland bereichern, da sie durch die Erfassung von Confoundern vertiefende Analysen erlaubt, die in diesem Bereich bisher selten möglich waren. Diese Informationen werden durch Anreicherung mit Sekundärdaten, z. B. mit Abrechnungsdaten der Krankenkassen, die umfangreiche und qualitativ hochwertige Informationen zur Verschreibung von Medikamenten, zur stationären Therapie und der ambulanten Versorgung enthalten, zu einem besseren Verständnis gesundheitlicher Ungleichheit im Krankheitsgeschehen beitragen. Diese Daten werden aber auch dabei helfen, sowohl Versorgungsdefizite als auch Bereiche der Überversorgung zu identifizieren und so den Ressourceneinsatz im Gesundheitswesen zu verbessern.

Ethik und Datenschutz

Die Interaktion mit den Teilnehmern, aber auch die spätere Daten- und Probennutzung unterliegen höchsten ethischen Standards, die sich an den Empfehlungen des deutschen Ethikrats orientieren [19] und in einem eigenen Ethikkodex der NAKO festgehalten sind [13]. Die Studienunterlagen wurden den für die Studienzentren zuständigen ärztlichen Ethikkommissionen zur Prüfung vorgelegt und positiv bewertet. Darüber hinaus berät ein eigens für die NAKO ins Leben gerufener Ethikrat die Studie in allen ethischen Fragen. Zur Sicherung des Datenschutzes wurden umfangreiche Maßnahmen ergriffen, die für eine sichere und dauerhafte Trennung der Forschungsdaten von den personenidentifizierenden Daten gewährleisten. Zu diesem Zweck wurde eigens für die Studie eine unabhängige Treuhandstelle eingerichtet. Der Zugriff auf personenidenti-

fizierende Daten ist auf einen sehr engen Personenkreis beschränkt, und unbefugte Zugriffe werden durch modernste technische Sicherungen verhindert. Alle Datenflüsse und Schutzmaßnahmen wurden auf Grundlage des generischen Datenschutzkonzepts der Technologie- und Methodenplattform für die vernetzte medizinische Forschung (TMF) in enger Abstimmung mit der Bundesbeauftragten für den Datenschutz und die Informationsfreiheit (BfDI) entwickelt. Die BfDI wird die Studie auch zukünftig begleiten.

Ist die NAKO ein Beispiel für Big Data?

In der Öffentlichkeit werden die im Rahmen von Beobachtungsstudien gesammelten Daten gerne als „Datenkörper", „Datenbasis", „Datenbank" oder sogar als „Datenschatz" bezeichnet. Gerade die letzte Bezeichnung signalisiert, dass die Summe der in der Studie erhobenen Daten besonders wertvoll ist. Dabei umfassten epidemiologische Studien in der Vergangenheit manchmal nur einige Hundert und oft nur wenige Tausend Probanden. Durch das Zusammenfügen mehrerer Studien, Pooling genannt, wurden Umfänge von mehreren Zehntausend und in neuester Zeit auch mehreren Hunderttausend Personen erreicht. Mit einer Teilnehmerzahl von 200.000 Bewohnern aus Deutschland wird die NAKO zur „größten Gesundheitsstudie" in Deutschland, sodass es nicht verwundert, dass diese große Datenbasis die Assoziation „Big Data" mit sich bringt.

Was ist Big Data?

Eine Suche nach dem Begriff „Big Data" in „PubMed", der renommierten Literaturdatenbank für die Medizin, zeigt erste Einträge in 2003 (2) und 2004 (1) und dann erst wieder 2008 mit steigender Tendenz ab 2012, wobei von den 313 Literaturstellen allein 103 aus dem Jahr 2014 stammen (Stand 4. Mai 2014). Die zunehmende Auseinandersetzung mit diesem Thema folgte damit auf die Publikation einer speziell dem Thema „Big Data" gewidmeten Ausgabe der Zeitschrift *Nature* im September 2008. Der Großteil der Literaturstellen der Anfangszeit bezieht sich dabei auf Big Data im Sinne von einer Vielzahl von Messwerten, Genvarianten oder Genexpressionen. Die Epidemiologie steht hier nicht im Zentrum, nur 15 Referenzen weisen sowohl den Begriff „Big Data" als auch „Epidemiologie" auf. Die Planungen zur Nationalen Kohorte haben bereits in der zweiten Hälfte des vorigen Jahrzehnts begonnen, also deutlich bevor dieser Begriff in der wissenschaftlichen Diskussion eine Rolle spielte.

Erst in jüngster Zeit gelangte der Begriff in den deutschen Sprachraum. Nach Mayer-Schönberger und Cukier entstand der Begriff „Big Data" in Bereichen der Naturwissenschaften wie der Astronomie und der Genetik, die mit einer enormen Explosion an zu verarbeitender Datenmengen konfrontiert waren, die mit herkömmlichen Methoden der Datenanalyse nicht zu bewältigen sind [20]. Zunehmend wurden weitere Bereiche wie die Ökonomie davon erfasst:

Big Data bezeichnet Datenmengen, die zu groß sind, um sie mit händischen und klassischen Methoden der Datenverarbeitung auszuwerten. Die Daten können aus vielfältigen Quellen wie Sensoren, Kameras oder der Überwachung von Internetverkehr stammen. Es sind neue Technologien nötig, um Big Data zu erfassen, zu verteilen, zu speichern, zu durchsuchen,zu analysieren und zu visualisieren [21]

Des Weiteren wird darauf hingewiesen, dass die Begriffsdefinition unscharf ist und sich kontinuierlich wandelt, wobei auch auf die NSA-Affäre und die Diskussion um soziale Netzwerke wie Facebook hingewiesen wird. Allerdings haben sich nach Fasel drei Merkmale herauskristallisiert, die für Big Data charakteristisch sind und als die drei „V" bezeichnet werden: Volume, Velocity und Variety [22]. *Volume* bezieht sich auf die großen anfallenden Datenmengen und *Velocity* auf die zunehmende Geschwindigkeit ihrer Entstehung und Verwertung. Mit *Variety* ist die große Heterogenität und unterschiedliche Struktur und Granularität der zusammenzuführenden Daten gemeint, wie z. B. Daten mit flexiblen Formaten aus Webservices, die zur Auswertung völlig neue und andersartige Anforderungen stellen als herkömmliche Datenbanken.

Sofern es um Big Data im Zusammenhang mit Menschen geht, ergibt sich als Abgrenzungskriterium gegenüber gewöhnlichen „Datensammlungen", „Datenbasen", etc. neben der schieren Größe vor allen Dingen der folgende Aspekt: Das Bild des Menschen in Big Data ist ein Extrakt seiner digitalen Hinterlassenschaften (oder einer Teilmenge davon), die

- zumeist für andere Zwecke gesammelt wurden und
- häufig aus unterschiedlichen Bereichen kommen, wobei den Betroffenen in der Regel gar nicht klar ist, dass diese Daten zusammengeführt werden,

und deren Auswertung

- für einen Zweck verwendet wird, über den das Individuum nicht informiert wird und von dem es in der Regel auch keinen Nutzen hat,
- den Menschen als Merkmalsträger begreift und ihn – sofern er individuelle Konsequenzen (egal ob positiv oder negativ) spürt – als Mitglied einer Gruppe mit genau diesen Eigenschaften behandelt,
- auf rein statistischen Modellen beruht, deren Gültigkeit weder plausibilisiert noch falsifiziert werden kann.

Was unterscheidet die Nationale Kohorte von Big Data?

Die Nationale Kohorte hat einen klar definierten Zweck. Sie soll helfen, die Ursachen von Volkskrankheiten aufzuklären, Risikofaktoren zu identifizieren, Wege einer wirksamen Vorbeugung aufzuzeigen und Möglichkeiten der Früherkennung von Krankheiten zu identifizieren. Sie stellt eine Forschungsressource dar, die von der Wissenschaft nur zu diesem Zweck und nur im Rahmen genau definierter Regeln genutzt werden kann.

Ein Mensch, der zufällig über das Einwohnermelderegister in einer der Studienregionen ausgewählt wurde, wird persönlich angeschrieben und über den Zweck der Datengewinnung, die extra für diesen Zweck erfolgt, informiert und aufgeklärt. Er wird nicht auf seinen Datensatz reduziert, sondern steht als Person in unmittelbarem Kontakt mit dem Studien-

personal. Die Person kann zu ihren Daten jederzeit Auskunft verlangen und hat es in der Hand, zuzustimmen oder abzulehnen, ob weitere Daten personalisierter Art über sie eingeholt werden. Kohortenmitglieder sind somit nicht auf ihre digitale Hinterlassenschaft reduziert. Es ist selbstverständlich, dass die teilnehmenden Personen jederzeit ihr Einverständnis zurücknehmen können.

Jede teilnehmende Person bekommt eine unmittelbare Rückmeldung zu ihren Untersuchungsergebnissen und kann sicher sein, dass die wissenschaftliche Nutzung ihrer Daten dem Gemeinwohl dient. Das Prinzip der Transparenz und der Nachvollziehbarkeit ist für die Nationale Kohorte leitend. Hierzu gehört, dass die Forschungsergebnisse allgemein zugänglich veröffentlicht werden und damit für eine kritische Diskussion allen Interessierten – auch außerhalb der wissenschaftlichen Gemeinschaft – zur Verfügung stehen.

Fazit

- **Mit der Nationalen Kohorte wird eine Ressource für die deutsche Gesundheitsforschung aufgebaut, die nicht nur wegen ihrer, Größe sondern insbesondere auch wegen des Einsatzes modernster Untersuchungsmethoden und des großen Umfangs an Untersuchungen bisher einmalig ist. Die Alterung unserer Bevölkerung macht einen Paradigmenwechsel erforderlich, der neben der Verbesserung von Diagnostik und Therapie von Erkrankten die Vorbeugung und Früherkennung dieser Krankheiten in das Zentrum stellt, um die Voraussetzungen für ein gesundes Altern zu schaffen. Dieses auch international der Spitzenforschung zuzuordnende Projekt wird es ermöglichen, neue Präventionsstrategien zu entwickeln und die Auswirkungen chronischer Erkrankungen durch Früherkennung abzumildern oder sogar zu vermeiden.**
- **Die Nationale Kohorte unterscheidet sich deutlich von dem, was im Allgemeinen unter Big Data verstanden wird. Sie besteht aus Menschen, die nach ausführlicher Aufklärung über die Studienziele zugestimmt haben, dass über sie Daten erhoben und für die Gesundheitsforschung genutzt werden dürfen. Der verantwortungsvolle Umgang mit diesen Daten nach strengsten Datenschutzregelungen und neuesten technischen Standards ist die Voraussetzung dafür, dass die Bereitschaft zur weiteren Teilnahme an der Nationalen Kohorte bestehen bleibt. Nur so wird die Voraussetzung dafür geschaffen, dass die lange Beobachtungsdauer erreicht wird, ohne die die Erforschung der interessierenden Erkrankungen nicht möglich ist.**

Korrespondenzadresse

W. Ahrens
Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS, Bremen
Achterstraße 30, 28359 Bremen
ahrens@bips.uni-bremen.de

Acknowledgement. Die Nationale Kohorte wird gefördert vom Bund, den Ländern (Förderkennzeichen des Bundesministeriums für Bildung und Forschung (BMBF 01ER1301A), der Helmholtz-Gemeinschaft sowie durch Eigenleistungen der beteiligten Institutionen.

Einhaltung ethischer Richtlinien

Interessenskonflikt. W. Ahrens und K.-H. Jöckel geben an, dass kein Interessenskonflikt besteht.

Alle beschriebenen Untersuchungen am Menschen wurden mit Zustimmung der zuständigen Ethik-Kommission, im Einklang mit nationalem Recht sowie gemäß der Deklaration von Helsinki von 1975 (in der aktuellen, überarbeiteten Fassung) durchgeführt. Von allen Studienteilnehmern liegt eine Einverständniserklärung vor.

Literatur

1. Peters E, Pritzkuleit R, Beske F, Katalinic A (2010) Demografischer Wandel und Krankheitshäufigkeiten. Eine Projektion bis 2050 [Demographic change and disease rates: a projection until 2050]. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 53(5):417–426. doi:10.1007/s00103-010-1050-y
2. Schmidt S, Hendricks V, Griebenow R, Riedel R (2013) Demographic change and its impact on the health-care budget for heart failure inpatients in Germany during 1995–2025. Herz 38(8):862–867. doi:10.1007/s00059-013-3955-3
3. Raspe H, Stumpf S (Hrsg) (2012) Priorisierung im Gesundheitswesen 2012 – zum aktuellen Stand der Diskussion. Z Evid Fortbild Qual Gesundhwes 106(6):377–474
4. Schwartz SM, Pomana L, Hyde-Nolan ME, Carter EW (2014) Sustained economic value of a wellness and disease prevention program: an 8-year longitudinal evaluation. Popul Health Manag 17(2):90–99. doi:10.1089/pop.2013.0042
5. Neumann A, Schwarz P, Lindholm L (2011) Estimating the cost-effectiveness of lifestyle intervention programmes to prevent diabetes based on an example from Germany: markov modelling. Cost Eff Resour Alloc 9(1):17. doi:10.1186/1478-7547-9-17
6. Weintraub WS, Daniels SR, Burke LE, Franklin BA, Goff DC Jr, Hayman LL, Lloyd-Jones D, Pandey DK, Sanchez EJ, Schram AP, Whitsel LP, American Heart Association Advocacy Coordinating Committee, Council on Cardiovascular Disease in the Young, Council on the Kidney in Cardiovascular Disease, Council on Epidemiology and Prevention, Council on Cardiovascular Nursing, Council on Arteriosclerosis, Thrombosis and Vascular Biology, Council on Clinical Cardiology, and Stroke Council (2011) Value of primordial and primary prevention for cardiovascular disease: a policy statement from the American Heart Association. Circulation 124(8):967–990. doi:10.1161/CIR.0b013e3182285a81
7. Kreienbrock L, Pigeot I, Ahrens W (2012) Epidemiologische Methoden, 5. Aufl. Springer, Berlin (http://dx.doi.org/10.1007/978-3-8274-2334-4)
8. Ray WA (2003) Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol 158(9):915–920
9. Wichmann H-E, Kaaks R, Hoffmann W, Jöckel K-H, Greiser KH, Linseisen K-H (2012) Die Nationale Kohorte. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 55:781–789
10. German National Cohort (GNC) Consortium (2014) The German National Cohort: aims, study design and organization. Eur J Epidemiol 29(5):371–382. doi:10.1007/s10654-014-9890-7
11. Ahrens W, Greiser H, Linseisen J, Kluttig A, Schipf S, Schmidt B, Günther K (2014) Das Design der Machbarkeitsstudien für eine bundesweite Kohortenstudie in Deutschland: Die Pretests der Nationalen Kohorte (NAKO) [The design of a nationwide cohort study in Germany: the pretest studies of the German National Cohort (GNC)]. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 57(11):1246–1254. doi:10.1007/s00103-014-2042-0
12. Starkbaum J, Gottweis H, Gottweis U, Kleiser C, Linseisen J, Meisinger C, Kamtsiuris P, Moebus S, Jöckel KH, Börm S, Wichmann HE (2014) Public perceptions of cohort studies and biobanks in Germany. Biopreserv Biobank 12(2):121–130
13. Nationale Kohorte e. V. Nationale Kohorte – Gemeinsam forschen für eine gesündere Zukunft. http://www.nationale-kohorte.de. Zugegriffen: 11. Jan. 2015
14. Ahrens W, Pigeot I (2012) Internationale Kohortenstudien [International cohort studies]. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 55(6–7):756–766. doi:10.1007/s00103-012-1495-2
15. Ollier W, Sprosen T, Peakman T (2005) UK Biobank: from concept to reality. Pharmacogenomics 6(6):639–646
16. Scholtens S, Smidt N, Swertz MA, Bakker SJ, Dotinga A, Vonk JM, van Dijk F, van Zon SK, Wijmenga C, Wolffenbuttel BH, Stolk RP (2014) Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol first published online December 14, 2014 doi:10.1093/ije/dyu229
17. Zins M, Coeuret-Pellicer M, Guéguen A, Gourmelen J, Nachtigal M, Ozguler A, Quesnot A, Ribet C, Rodrigues G, Serrano A, Sitta R, Brigand A, Henny J, Goldberg M (2010) The CONSTANCES cohort: an open epidemiological laboratory. BMC Public Health 10:479. doi:10.1186/1471-2458-10-479

18. Almqvist C, Adami HO, Franks PW, Groop L, Ingelsson E, Kere J, Lissner L, Litton JE, Maeurer M, Michaëlsson K, Palmgren J, Pershagen G, Ploner A, Sullivan PF, Tybring G, Pedersen NL (2011) LifeGene–a large prospective population-based study of global relevance. Eur J Epidemiol 26(1):67–77. doi:10.1007/s10654-010-9521-x
19. Deutscher Ethikrat (2010) Humanbiobanken für die Forschung – Stellungnahme. http://www.ethikrat.org/dateien/pdf/stellungnahme-humanbiobanken-fuer-die-forschung.pdf. Zugegriffen: 11. Jan. 2015
20. Mayer-Schönberger V, Cukier K (2013) Big Data – Die Revolution, die unser Leben verändern wird. Redline Wirtschaftsverlag, München
21. Big Data – Wikipedia. http://de.wikipedia.org/wiki/Big_Data. Zugegriffen: 11. Jan. 2015
22. Fasel D (2014) Big Data – Eine Einführung. HMD Praxis Wirtschaftsinformatik 51(4):386–400