



Epidemiologische Studien

Auf Schatzsuche

3.1 Die ätiologische Fragestellung

Die analytische Epidemiologie kann als die Wissenschaft von der Erforschung der Ursachen von Erkrankungen in gesamten Populationen bezeichnet werden. Die wissenschaftliche Untersuchung basiert dabei auf einer ätiologischen Fragestellung, die die Formulierung von fachlichen Hypothesen beinhaltet und die Berechnung und Interpretation von vergleichenden epidemiologischen Maßzahlen zur Beantwortung dieser Frage nach sich zieht.

Ätiologische Fragestellungen werden bei der Hypothesenformulierung zunächst auf die Beschreibung eines Zusammenhangs einer Krankheit und einer bestimmten Exposition reduziert, wobei diese häufig als dichotom, d.h. vorhanden oder nicht, angenommen werden. Da aber im Regelfall davon auszugehen ist, dass eine Erkrankung multikausalen Ursprungs ist, sind bei der Untersuchung der interessierenden Ursache-Wirkungs-Beziehung weitere Faktoren zu berücksichtigen. Wird also z.B. nach den Ursachen des Auftretens des Lungenkarzinoms gefragt und hierbei etwa die Luftverunreinigung als eine mögliche Erkrankungsursache diskutiert, so ist immer auch zu fragen, ob weitere Risikofaktoren existieren, ob diese konkurrierend sind oder ob und in welcher Form auch Synergismen für die Entstehung dieses Krankheitsbildes verantwortlich gemacht werden können.

Daher reicht es im Regelfall nicht aus, Gruppen von Individuen durch einfache Vergleiche gegenüberzustellen und entsprechende epidemiologische Kennzahlen zu berechnen. Vielmehr ist für die Untersuchung der Ätiologie von Erkrankungen stets ein vertieftes Verständnis der biologischen Zusammenhänge sowie der medizinischen Grundlagen erforderlich. Im Folgenden werden hierzu einige grundlegende Überlegungen und Begriffsbildungen aufgeführt.

3.1.1 Modell der Ursache-Wirkungs-Beziehung und Kausalität

Als eine der grundlegenden Aufgaben der analytischen Epidemiologie kann, wie bereits mehrfach erwähnt, die Quantifizierung des Einflusses eines Risikofaktors auf eine definierte Krankheit angesehen werden.

Einfluss- und Zielvariable

Im Sinne eines Ursache-Wirkungs-Modells wird damit unterstellt, dass ein **Einflussfaktor (Risiko-, Ursachenfaktor oder Einflussvariable, -größe)** vorliegt, der die zu untersuchende Exposition beschreibt. Hierzu zählen z.B. die Anzahl gerauchter Zigaretten bei einer Fragestellung zur Krebsentstehung, die kumulativ aufgenommene Menge eines Schadstoffes am Arbeitsplatz bei einer arbeitsmedizinischen Untersuchung oder die Art und Menge aufgenommener Nahrung bei der Untersuchung des kindlichen Übergewichts.

Als **Wirkungs- oder Zielgröße** muss dann ein Krankheitsparameter definiert werden, der die zu untersuchende Wirkung der Einflussvariable angibt. Hier werden häufig das Vorliegen oder Fehlen eines Symptoms oder einer medizinischen Diagnose (qualitativ), aber auch quantitative medizinische Laborgrößen wie der Blutdruck, die Körpertemperatur oder Messwerte einer Lungenfunktionsuntersuchung zur Wirkungsbeschreibung herangezogen.

Im einfachsten Fall liegen Einfluss- und Zielgröße jeweils dichotom, d.h. mit nur zwei möglichen Ausprägungen, vor und die mathematisch-statistische *Formalisierung der ätiologischen Fragestellung* erfolgt, wie in Abschnitt 2.3.1 bereits beschrieben, in Form einer *Vierfeldertafel* gemäß Tab. 3.1.

Tab. 3.1: Vierfeldertafel zur Beschreibung der ätiologischen Fragestellung

Status	exponiert (Ex = 1)	nicht exponiert (Ex = 0)
krank (Kr = 1)	$P_{11} = \Pr(Kr = 1 \mid Ex = 1)$	$P_{10} = \Pr(Kr = 1 \mid Ex = 0)$
gesund (Kr = 0)	$P_{01} = \Pr(Kr = 0 \mid Ex = 1)$	$P_{00} = \Pr(Kr = 0 \mid Ex = 0)$

Diese Form der Darstellung kann natürlich erweitert werden, indem man mehrere Krankheits- bzw. Expositions-kategorien angibt (z.B. "gesund", "leicht erkrankt", "schwer erkrankt", "tot" oder "hoch", "mittel", "niedrig exponiert", siehe auch Tab. 6.3 und Tab. 6.4).

Innerhalb dieser Tafel können nun über die in Kapitel 2 definierten vergleichenden Maßzahlen inhaltliche Fragestellungen formalisiert werden. Die Frage, inwiefern überhaupt ein Zusammenhang zwischen einer Krankheit und einer Exposition vorliegt, führt dann z.B. zu der statistisch prüf-baren Hypothese, ob das Odds Ratio gleich eins ist oder nicht.

Dabei ist bei der Untersuchung des Einflusses der interessierenden Exposition auf die Zielvariable kritisch zu prüfen, ob eine beobachtete Beziehung durch dritte Variablen gestört, hervorgerufen oder verdeckt wird.

Eine **Störvariable**, auch als **Störfaktor**, **Störgröße** oder **Confounder** bezeichnet, liegt dann vor, wenn dieser Faktor, der nicht Ziel der Untersuchung ist, einerseits auf die Zielvariable der Krankheit kausal wirkt und andererseits gleichzeitig mit der interessierenden Exposition assoziiert ist (zur Veranschaulichung siehe auch Abb. 3.1 (b), (c)). Dieser Mechanismus, der unter dem Begriff des Confoundings (Vermengung, Vermischung) bekannt ist, hat eine zentrale Bedeutung in der analytischen Epidemiologie und muss daher stets besonders beachtet werden, da eine mangelnde Berücksichtigung zu falschen Schlüssen bzgl. des Vorliegens oder der Stärke von Zusammenhängen zwischen der interessierenden Exposition und einer bestimmten Krankheit führen kann.

Beispiel: Ein typisches Beispiel für einen Confoundingmechanismus findet sich bei der Betrachtung der Ätiologie des Speiseröhrenkarzinoms. Will man etwa eine Aussage dazu machen, inwiefern es einen Zusammenhang zwischen Alkoholkonsum und dem Auftreten dieses Karzinoms gibt, so muss stets auch der Risikobeitrag des Rauchens – ein bekannter Risikofaktor für das Auftreten des Speiseröhrenkarzinoms – berücksichtigt werden, denn Tabak- und Alkoholkonsum sind nicht unabhängig voneinander. Damit kann eine Vermengung der Effekte des Rauchens mit denen des Alkoholkonsums auftreten (siehe Abb. 3.1 (c)).

Variablen, die für unterschiedlichste Krankheiten als Confounder gelten, sind häufig das Alter, das Geschlecht sowie das Rauchen. Je nach betrachteter Krankheit kommt dann häufig eine Vielzahl weiterer (potenzieller) Störgrößen hinzu. So ist es z.B. bei der Betrachtung von Tierpopulationen üblich, die Größe von landwirtschaftlichen Betrieben als entsprechende Confounder zu berücksichtigen.

Confounder können sogar eine **Scheinassoziation** zwischen der interessierenden Exposition und der Zielvariable hervorrufen. In diesem Fall wirkt die Exposition nicht auf die Zielvariable, ist aber mit dem Confounder assoziiert und erscheint nur als mögliche Einflussvariable (siehe Abb. 3.1 (b)).

Beispiel: Ein Beispiel zum Auftreten einer Scheinassoziation zwischen der Säuglingssterblichkeit und der Dichte von Fernsehantennen in verschiedenen Wohngegenden Dublins berichten Skabane & McCormick (1995). Betrachtet man das Variablenpaar "Säuglingssterblichkeit" und "Dichte von Fernsehantennen", so lässt sich eine positive Assoziation nachweisen, jedoch ist diese offensichtlich nicht kausal. Ursächlich ist vielmehr die Variable "Sozialer Status", die sowohl die Säuglingssterblichkeit als auch die Wohnsituation beeinflusst, so dass der Schein einer Assoziation nur dadurch erweckt wird, dass dieser eigentlich verantwortliche Faktor nicht berücksichtigt wurde.

Das obige Beispiel verdeutlicht, dass eine Scheinassoziation dadurch charakterisiert ist, dass diese Assoziation nicht kausal sein kann. Eine indirekte Assoziation zwischen der interessierenden Exposition und der Zielvariable bedeutet somit, dass beiden ein bekannter oder unbekannter Faktor zugrunde liegt, der seinerseits eine künstliche Assoziation zwischen der interessierenden Exposition und der Zielgröße erzeugt.

Beispiel: Ein weiteres Beispiel ist die Assoziation von Tabakexposition und Leberzirrhose. Auch hier besteht kein kausaler Zusammenhang. Erzeugt wird diese Scheinassoziation vielmehr dadurch, dass eine gemeinsame Assoziation beider Variablen zum ursächlichen Faktor Alkoholkonsum besteht.

Um die Zusammenhänge der eingeführten Variablen darzustellen, hat es sich eingebürgert, diese mit Hilfe eines **ätiologischen Diagramms** zu veranschaulichen (siehe Abb. 3.1). Hierbei bedeutet ein einfacher Pfeil, dass zwischen den Faktoren ein kausaler Zusammenhang besteht, ein Doppelpfeil deutet an, dass eine ungerichtete Assoziation zwischen den Faktoren vorliegt. Nur die Kontrolle sämtlicher Faktoren, die auf diese Art und Weise mit einer Krankheit in Verbindung stehen, ergibt somit einen Aufschluss über die wahre epidemiologische Ursache-Wirkungs-Beziehung.

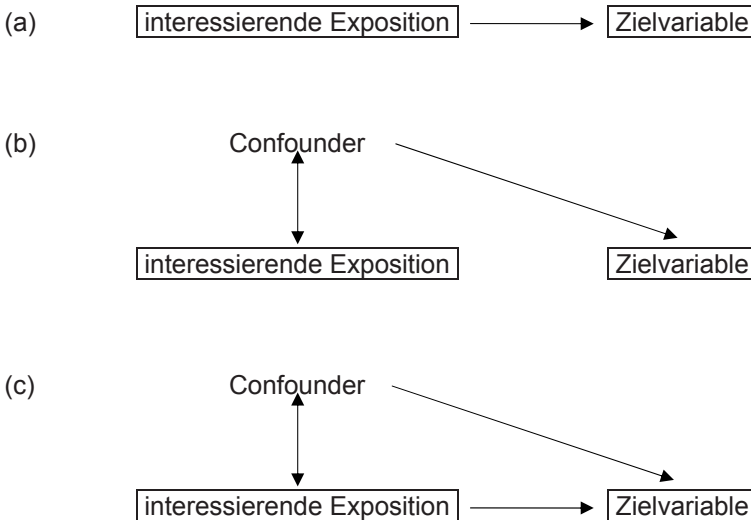


Abb. 3.1: Ätiologisches Diagramm: Schema der Ursache-Wirkungs-Beziehungen: (a) kausale Beziehung zwischen Exposition und Zielvariable; (b) durch einen Confounder hervorgerufene Scheinassoziation zwischen Exposition und Zielvariable; (c) durch einen Confounder beeinflusster Zusammenhang zwischen Exposition und Zielvariable
(—————> kausaler Zusammenhang, <————> ungerichtete Assoziation)

Typischerweise sind Zusammenhänge zwischen Expositionen und einer Krankheit nicht monokausal, sondern multikausal, d.h., es können so genannte **Zusatzvariablen** (engl. **Supplementary Risk Factors**) vorhanden sein. Diese Faktoren sind dadurch charakterisiert, dass sie ebenfalls auf die Zielvariable wirken, aber weder Ziel der Untersuchung sind noch mit der Einflussvariable zusammenhängen. Damit haben diese Faktoren keinen störenden Einfluss.

Das in Abb. 3.1 dargestellte ätiologische Diagramm kann nun um diese Zusatzvariablen erweitert werden. Es erweist sich in Studien als praktisches Hilfsmittel, um die Beziehungen zwischen den relevanten Einflussfaktoren darzustellen, wie das folgende Beispiel zeigt.

Beispiel: Ein mögliches Ursache-Wirkungs-Modell soll nachfolgend am Beispiel der Ätiologie des Lungenkrebses diskutiert werden. Hier wurde u.a. die Frage nach dem Zusammenhang zwischen dem Auftreten des Bronchialkarzinoms und einer Exposition mit dem radioaktiven Edelgas Radon in Wohnräumen diskutiert (siehe 1.2, vgl. auch Samet 1989, Wichmann et al. 1998). Ein Ursache-Wirkungs-Modell in obigem Sinne ist in Abb. 3.2 skizziert. Hierbei sind exemplarisch sowohl bekannte als auch mögliche Einflussgrößen in ihrem Zusammenhang mit dem Auftreten von Lungenkrebs dargestellt.

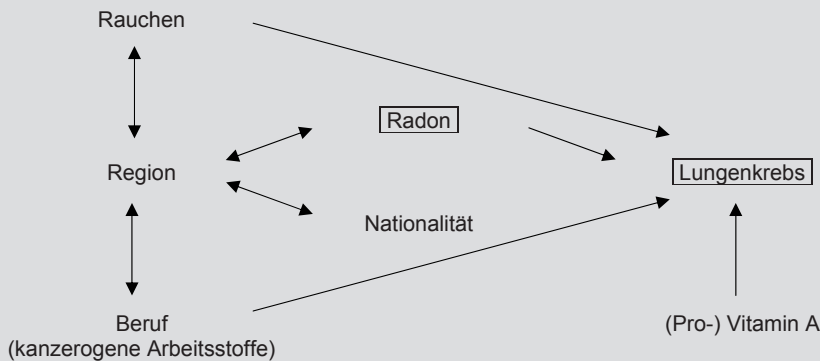


Abb. 3.2: Vereinfachtes Schema der Ursache-Wirkungs-Beziehungen zur Ätiologie des Lungenkrebses
($\longrightarrow$ kausaler Zusammenhang, $\longleftrightarrow$ ungerichtete Assoziation)

Kernpunkt des Modells ist die Frage, inwiefern die (kumulative) Exposition mit Radon ein Lungenkarzinom induziert. Diese Fragestellung wird von verschiedenen Faktoren beeinflusst:

Da das Auftreten von Radon unter anderem von der Gesteinsart (geologische Formation) abhängig ist, ist eine Assoziation zur Variable Region gegeben. Diese Variable Region steht nun in verschiedenartigen Beziehungen zu anderen Faktoren.

Sowohl das Rauchverhalten (höherer Anteil von Rauchern in Regionen mit höherer Besiedlungsdichte) als auch die Exposition mit krebserzeugenden Stoffen am Arbeitsplatz sind mit der Variable Region assoziiert. Da beide Variablen als ursächliche Risikofaktoren für die Entstehung des Bronchialkarzinoms angesehen werden können, sind diese somit als den Einfluss des Radon verändernde Confounder zu berücksichtigen.

Als nicht kausal kann beispielsweise die Variable Nationalität angesehen werden. Auch hier besteht eine Assoziation mit der Variable Region über den höheren Anteil ausländischer Bürger in Ballungsräumen, so dass eine Scheinassoziation zwischen Nationalität und Lungenkrebs ohne Berücksichtigung der übrigen Risikofaktoren entstehen könnte. In diesem Fall sind Rauchen, Beruf und Radon als Confounder zu berücksichtigen.

Protektiv bei der Entstehung des Lungenkarzinoms wirkt das (Pro-) Vitamin A. Unterstellt man, dass die Einnahme von Vitaminpräparaten keine regionale Abhängigkeit aufweist, so wird diese Variable als zusätzlicher Einflussfaktor wirken, ohne allerdings einen störenden Einfluss zu haben.

Das obige (immer noch stark vereinfachte) Beispiel deutet an, dass die Wechselbeziehungen zwischen ursächlichen und assoziierten Variablen innerhalb der ätiologischen Fragestellung

von großer Komplexität sein können. Dabei interessiert primär, welche Assoziationen eine Kausalbeziehung darstellen.

Kausalität

Um kausale Zusammenhänge zu entdecken, sind in vielen wissenschaftlichen Disziplinen experimentelle Versuche möglich, die den Kriterien der Wiederholbarkeit und der Randomisierung, d.h. der zufälligen Auswahl von Subjekten und zufälligen Zuweisung von Behandlungen, bei ansonsten konstanten Versuchsbedingungen genügen. Hierbei wird z.B. eine Exposition zufällig zugeordnet, um somit bis auf die Exposition vollkommen vergleichbare Gruppen zu erzeugen, wodurch ein direkter Wirkungsvergleich ermöglicht wird (vgl. z.B. Hartung et al. 2009).

Das grundlegende *Problem der analytischen Epidemiologie* ist nun, dass die "Versuchseinheiten" den "Behandlungen" nicht zufällig zugeteilt werden können, denn die epidemiologische Untersuchungsmethodik ist in der Regel die Beobachtung und *nicht das Experiment*. Daher ist die durch eine Randomisierung erreichbare direkte Vergleichbarkeit in der Epidemiologie nicht gegeben.

So gibt es in dem Beispiel zur Frage des durch Radon induzierten Lungenkrebses keine Möglichkeit, eine bestimmte Radonexposition einer Personengruppe zuzuordnen, denn dies würde bedeuten, Personen wissentlich dieser schädlichen Exposition auszusetzen.

Abgesehen von der Tatsache, dass sich eine solche Randomisierung beim Menschen aus ethischen Gründen verbietet, können die Versuchsbedingungen bei einer so langen Induktionszeit der Krankheit, wie bei Lungenkrebs, kaum konstant gehalten werden. Allein deshalb ist ein experimenteller Ansatz nicht sinnvoll.

Beobachtet man in einer nicht-experimentellen Untersuchung nun tatsächlich z.B. mehr Lungenkrebsfälle in einer Gruppe hoch Radonexponierter, so muss dieser Unterschied nicht die Folge der vermuteten Ursache "Radon" sein. Vielmehr können sich Exponierte und Nichtexponierte durch eine Reihe anderer spezifischer Eigenschaften unterscheiden, die in der Studie nicht berücksichtigt wurden und auch nicht – wie im Experiment – "herausrandomisiert" werden konnten.

Da eine statistische Auswertung nur in der Lage ist, eine Assoziation zwischen einer vermuteten Ursache und deren Wirkung aufzuzeigen, muss im Gegensatz zum Experiment für die Absicherung der Kausalitätsannahme innerhalb der epidemiologischen Forschung zusätzliche Evidenz erbracht werden.

Kausalität soll im Folgenden so verstanden werden (vgl. auch Rothman & Greenland 1998), dass ein Risikofaktor eine Krankheit dann verursacht, wenn folgende *drei Bedingungen* erfüllt sind:

- die Exposition geht der Erkrankung zeitlich voraus,

- eine Veränderung in der Exposition geht mit einer Veränderung in der Krankheitshäufigkeit einher,
- die Assoziation von Risikofaktor und Krankheit ist nicht die Folge einer Assoziation dieser Faktoren mit einem dritten Faktor.

Neben diesen Bedingungen wurden in der Literatur weitere wissenschaftliche Kriterien zur Abgrenzung einer Kausalbeziehung von einer indirekten Beziehung angeführt. Erste Überlegungen zur ursächlichen Beteiligung von Infektionserregern an einem Krankheitsgeschehen in der Bevölkerung findet man z.B. bereits in Arbeiten von Robert Koch. Wegweisend waren zudem die Kriterien, die von Austin Bradford Hill im Zusammenhang mit der Durchführung klinischer Studien aufgestellt wurden. Diese waren letztendlich auch Vorbild für die Definition von Kriterien, die bei der Bewertung von epidemiologischen Beobachtungsstudien zur Anwendung kommen (vgl. etwa Evans 1976, Mausner & Kramer 1985, Ackermann-Liebrich et al. 1986, Überla 1991 u.a.). Neben den oben genannten Bedingungen betreffen diese Kriterien die folgenden Punkte, die hier nur stichwortartig skizziert werden sollen:

- *Stärke der Assoziation*,
d.h. je stärker die Assoziation zwischen Risikofaktor und Zielgröße (z.B. ausgedrückt durch das relative Risiko oder das Odds Ratio), desto eher kann von einer Kausalbeziehung ausgegangen werden;
- *Vorliegen einer Dosis-Effekt-Beziehung*,
d.h., mit zunehmender Exposition erhöht sich das Risiko zu erkranken; die Reaktion auf die als ursächlich angesehene Exposition sollte bei den Individuen auftreten, die vor der Exposition diese Reaktion noch nicht gezeigt haben; eine Verminderung der Exposition sollte das Krankheitsrisiko wieder senken;
- *Konsistenz der Assoziation*,
d.h., die gefundene Assoziation soll auch in anderen Populationen und mit verschiedenen Studientypen reproduzierbar sein;
- *Spezifität der Assoziation*,
d.h., der Grad der Sicherheit, mit der beim Vorliegen eines Faktors die Krankheit vorhergesagt werden kann, sollte hoch sein; ideal wäre hier eine ein-eindeutige Beziehung, bei der ein Faktor notwendig und hinreichend für eine Krankheit ist; eine solche völlige Spezifität ist zwar ein starker Hinweis auf eine Kausalbeziehung, ist aber angesichts der Tatsache, dass ein Faktor verschiedene Erkrankungen hervorrufen kann und eine Krankheit durch verschiedene Faktoren hervorgerufen wird, nur selten realistisch;
- *Kohärenz mit bestehendem Wissen/biologische Plausibilität*,
d.h., die gefundene Assoziation sollte mit bestehendem Wissen übereinstimmen; dies gilt insbesondere nicht nur für epidemiologische Untersuchungen, sondern auch für tierexperimentelle Studien und generelle biologisch-medizinische Überlegungen; hierbei muss allerdings bedacht werden, dass ein wissenschaftlicher Fortschritt natürlich auch von Ergebnissen ausgeht, die mit gegenwärtigen Theorien nicht erklärbar sind.

Der Kausalitätsbegriff der Epidemiologie ist somit ein eher gradueller im Sinne von zunehmender Evidenz durch das Zusammentragen von Indizien. Gelegentlich werden die von Hill zusammengestellten Kriterien im Sinne eines aufsummierbaren "Kausalitätsindex" missverstanden. Sie sind aber keinesfalls so zu verstehen, dass eine Kausalbeziehung umso eher angenommen werden kann, desto mehr der Kriterien erfüllt sind. Sie stellen lediglich eine Bewertungshilfe dar. Inwieweit sich diese Kriterien zu einem interpretierbaren Gesamtergebnis zusammensetzen lassen, muss deshalb bei jeder Untersuchung z.B. gemäß obiger Kriterien neu hinterfragt werden, immer vorausgesetzt, dass die zuvor genannten drei Bedingungen erfüllt sind.

3.1.2 Populationsbegriffe, Zufall und Bias

Bei den bisherigen Betrachtungen waren wir stets davon ausgegangen, dass die epidemiologische Untersuchung einer Population unter Risiko gilt, d.h. einer wohl definierten Gruppe, in der die Individuen ein interessierendes Krankheitsbild entwickeln können oder nicht. Sämtliche Aussagen über epidemiologische Häufigkeitsmaßzahlen wie die Inzidenz oder vergleichende Parameter wie das Odds Ratio hatten dabei ihre Gültigkeit in Bezug auf die gesamte betrachtete Population.

Soll nun im konkreten Einzelfall eine Aussage über eine solche Population gemacht werden, so wird man in der Regel kaum die Möglichkeit haben, diese insgesamt zu untersuchen. Man wird sich stattdessen auf Stichproben, d.h. auf ausgewählte Teilpopulationen beschränken müssen. Da die Auswahl einer Teilpopulation eine Einschränkung darstellt, stellt sich die Frage, inwieweit sich dies auf die Interpretation von Ergebnissen auswirkt. Deshalb werden im Folgenden *verschiedene Populationsbegriffe* voneinander abgegrenzt.

Zielpopulation

Die im bisherigen Kontext als **Population unter Risiko** bezeichnete Gruppe wird in der Regel auch als **Ziel-, Bezugs- oder Grundpopulation** bzw. **-gesamtheit** bezeichnet. Sie gibt den Teil der Bevölkerung an, für den das Ergebnis der epidemiologischen Untersuchung gültig sein soll. So können etwa die gesamte Bevölkerung eines Landes, die Angehörigen eines bestimmten Alters oder Geschlechts oder sämtliche Nutztiere eines Landes – je nach epidemiologischer Fragestellung – als Grundgesamtheiten betrachtet werden.

Will man eine solche Zielgesamtheit empirisch untersuchen, um z.B. die Prävalenz einer Krankheit an einem Stichtag zu bestimmen, so ist es notwendig, für sämtliche Angehörigen dieser Population den Krankheitsstatus zu erfassen, um den Anteil der Erkrankten zu ermitteln.

Dass diese Vorgehensweise nur für kleine Populationen praktikabel ist, ist offensichtlich. Daher kann nur in wenigen Fällen eine so genannte **Total- oder Vollerhebung einer Zielpopulation** durchgeführt werden. Denkbar ist eine solche Vorgehensweise, wenn die Zielpopulation z.B. räumlich eingeschränkt werden kann, d.h., wenn beispielsweise die Auswirkung einer Industrieanlage auf die Gesundheit einer Bevölkerung in einem ländlichen Gebiet untersucht werden soll. Dieses Beispiel zeigt aber, dass solchen Totalerhebungen organisatio-

rische, technische und vor allem finanzielle Grenzen gesetzt sind. So muss eine *Vollerhebung* eher als *Ausnahme einer epidemiologischen Untersuchung* angesehen werden.

Untersuchungspopulation

Die *Regelform einer epidemiologischen Untersuchung* ist die *Stichprobenerhebung* einer Gruppe von Individuen. **Die Untersuchungs-, Studien- oder Stichprobenpopulation** ist die Teilpopulation, an der die eigentliche Untersuchung durchgeführt wird.

Um eine Studienpopulation zu wählen, sind verschiedenste Techniken denkbar. Die ausgewählte Untersuchungsgruppe sollte dabei möglichst repräsentativ für die Zielpopulation sein, also z.B. strukturell gut mit der Bevölkerung eines Landes übereinstimmen. Aber auch die praktische Realisierbarkeit spielt bei der Stichprobengewinnung eine Rolle.

Grundsätzlich kann man die Verfahren, Stichproben zu ermitteln, in drei verschiedene Prinzipien unterteilen. Man unterscheidet (1) die zufällige Auswahl von Individuen, (2) die Auswahl nach vorheriger Beurteilung (z.B. Auswahl einer für eine Kleinstadtbevölkerung als typisch geltenden Gemeinde oder Auswahl nach Quotenregel gemäß eines vorgegebenen Anteils von Männern und Frauen) sowie (3) die Auswahl aufs Geratewohl. Der Grad der Repräsentativität ist dabei umso höher, je näher das gesamte Auswahlverfahren dem Zufallsprinzip ist. Der an speziellen Techniken interessierte Leser sei z.B. auf Kreienbrock (1993), Dohoo et al. (2009) oder Kauermann & Küchenhoff (2011) verwiesen.

Da die Studienpopulation nur eine Auswahl darstellt, wird eine auf dieser Stichprobe basierende Aussage nicht exakt mit den Verhältnissen in der Zielgesamtheit übereinstimmen. Will man also etwa eine Prävalenz P_{Ziel} in einer Zielgesamtheit beschreiben und liegt in der Studienpopulation eine Prävalenz P_{Studie} vor, so muss aufgrund des Auswahlverfahrens davon ausgegangen werden, dass

$$\text{Prävalenz } P_{\text{Ziel}} \neq \text{Prävalenz } P_{\text{Studie}}$$

gilt. Der Grad dieser Abweichung kann ganz allgemein als Aussageungenauigkeit interpretiert werden und ist zentraler Bestandteil der epidemiologischen Untersuchungsplanung. Nur wenn diese Ungenauigkeit ausreichend gering ist, kann das Untersuchungsergebnis auch für die Zielgesamtheit interpretiert werden.

Auswahlpopulation

Dieses Phänomen weist auf eine besondere Abgrenzung in den Populationsbegriffen hin, denn häufig kann die Population, aus der die Stichprobe entnommen wird, sich aus praktischen Erwägungen heraus nur auf eine Untergruppe der Zielpopulation beziehen. In diesem Fall spricht man auch von einer **Auswahlpopulation** (engl. Study Base). Zur Erläuterung dieser Abgrenzung mag nachfolgendes Beispiel aus Abb. 3.3 dienen.

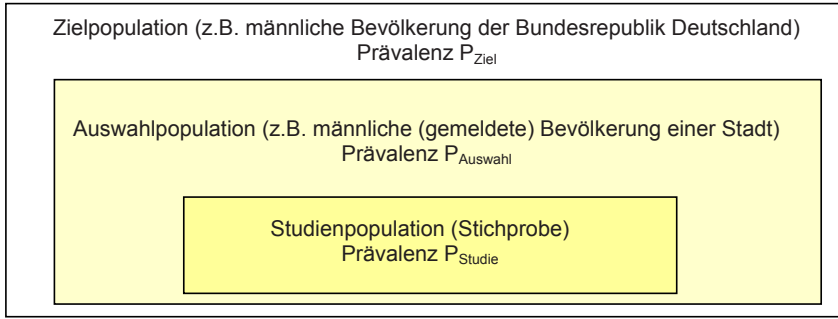


Abb. 3.3: Hierarchische Anordnung von Populationen und deren Prävalenzen

Beispiel: Die Zielgesamtheit einer epidemiologischen Untersuchung sei die männliche Gesamtbevölkerung Deutschlands, über die u.a. eine Prävalenzaussage, z.B. bzgl. des Anteils P_{Ziel} der Übergewichtigen, gemacht werden soll.

Beschränkt man sich nun aus Gründen der Praktikabilität bei der Durchführung der Untersuchung dieser Fragestellung darauf, eine Stichprobe von Männern aus der Einwohnermeldekartei einer einzelnen Großstadt zu ziehen, so ist diese Studienpopulation dadurch entstanden, dass die Auswahlpopulation gerade aus den männlichen Einwohnern nur dieser Stadt besteht.

Damit erhält man innerhalb der Untersuchung eine Prävalenzaussage P_{Studie} , die zunächst nur für die Prävalenz P_{Auswahl} in dieser Stadt gültig ist (siehe Abb. 3.3).

Das genannte Beispiel, das entsprechend auch für andere epidemiologische Maßzahlen Gültigkeit besitzt, zeigt, dass die exakte Abgrenzung der betrachteten Populationen von großer Bedeutung für die Interpretation einer epidemiologischen Untersuchung ist. Ein Studienergebnis P_{Studie} gilt zunächst ausschließlich nur für die Studienpopulation und kann in einem ersten Schritt auf den entsprechenden Parameter P_{Auswahl} der Auswahlpopulation übertragen werden, wenn Methoden der schließenden Statistik angewendet werden (siehe hierzu auch Anhang S.4.4). Kann man gewährleisten, dass die Auswahlpopulation eine gute Näherung der Zielpopulation darstellt, so ist dieser Schluss auch auf den Parameter P_{Ziel} der Zielgesamtheit selbst möglich.

Externe Population

Häufig ist allerdings die Situation gegeben, dass man mit einem erhaltenen Untersuchungsergebnis nicht nur auf die zugrunde liegende Zielgesamtheit schließen, sondern eine darüber hinausreichende Aussage treffen will. So ist in obigem Beispiel zur Prävalenz des Übergewichts unter Umständen auch die Frage interessant, inwiefern sich das Untersuchungsergebnis auf andere Länder oder auf Frauen übertragen lässt. Diese Fragestellungen führen zum Begriff der **externen Population**, also denjenigen Gruppen, die zunächst nicht in das Untersuchungskonzept eingebettet waren, für die aber dennoch eine Aussage abgeleitet werden soll.

Eine solche externe Induktion kann nicht mit Mitteln der Empirie oder der statistischen Schlussweise durchgeführt werden, sondern sollte nur unter Hinzuziehung soziodemographischer und fachwissenschaftlicher Information erfolgen.

Gesamtfehler einer epidemiologischen Untersuchung

Im Folgenden werden wir aus Gründen der Einfachheit stets voraussetzen, dass nur das Populationspaar Ziel- und Studienpopulation betrachtet wird, d.h., dass das Studienergebnis (z.B. die Studienprävalenz P_{Studie}) als repräsentativ für die Verhältnisse in der Zielgesamtheit (z.B. die Zielprävalenz P_{Ziel}) angesehen wird. Die Differenz zwischen beiden Größen bezeichnet man im Allgemeinen als **Gesamtfehler der Untersuchung**.

Charakteristisch für diesen Fehler ist, dass man seine Größe im Einzelfall nicht kennt, da nur die Studienprävalenz P_{Studie} als empirisches Ergebnis vorliegt und die Prävalenz P_{Ziel} in der Zielgesamtheit weiterhin unbekannt ist.

Will man somit ein Untersuchungsergebnis als repräsentativ für die entsprechenden Verhältnisse in der Zielgesamtheit werten und es entsprechend nutzen, ist es unumgänglich, sich mit der mutmaßlichen Größenordnung sowie der Struktur dieses Fehlers auseinanderzusetzen. Wie in allen empirischen Untersuchungen kann man auch im Rahmen der epidemiologischen Forschung zwei Fehlertypen unterscheiden: den zufälligen und den systematischen Fehler.

Zufallsfehler

Als **Zufallsfehler** bezeichnet man in der Regel eine Abweichung des Untersuchungsergebnisses von den wahren Verhältnissen der Zielgesamtheit, die dadurch entsteht, dass die Untersuchungsteilnehmer zufällig ausgewählt wurden.

Beispiel: Will man den Anteil von übergewichtigen Personen in einer Zielgesamtheit untersuchen und erhebt hierzu in einer Untersuchungspopulation die Körpergröße und das -gewicht, so wird jedes Individuum andere Größen- und Gewichtswerte haben, so dass bei jeder neuen Stichprobe jeweils andere individuelle Messergebnisse vorliegen werden. Die Schwankung dieser Werte um den wahren durchschnittlichen Wert in der Gesamtbevölkerung ist deshalb als zufällig anzusehen.

Das Beispiel zeigt, dass der Zufallsfehler allein dadurch entsteht, dass man sich im Rahmen einer Untersuchung auf ein Teilkollektiv beschränken muss. Der erwartete Zufallsfehler wird umso größer sein, je kleiner die ausgewählte Untersuchungspopulation ist. Daraus folgt, dass der zufällige Fehler einer epidemiologischen Untersuchung zwar nie ganz ausgeschlossen werden, eine Kontrolle dieses Fehlers aber durch die Größe der Stichprobenpopulation erfolgen kann. Die Bestimmung eines Stichprobenumfangs, der notwendig ist, um einen Zufallsfehler bestimmter Größenordnung nicht zu überschreiten, ist damit ein zentraler Aspekt epidemiologischer Untersuchungsplanung (siehe 4.3).

Systematische Fehler (Verzerrungen, Bias)

Im Gegensatz zum zufälligen Fehler lassen sich systematische Fehler nicht durch den Stichprobenumfang kontrollieren, denn sie sind dadurch charakterisiert, dass dieser Fehlertyp nicht vom ausgewählten Individuum abhängig ist. Von einem **systematischen Fehler**, einer **Verzerrung** oder einem **Bias** spricht man deshalb immer dann, wenn in der gesamten Untersuchungspopulation eine Abweichung des Ergebnisses zur Zielgesamtheit in eine bestimmte Richtung zu erwarten ist.

Beispiel: Im Beispiel zur Untersuchung des Übergewichts sind z.B. zwei Situationen denkbar. (A) Es könnte sein, dass die verwendete Waage z.B. wegen fehlerhafter Eichung geringere Werte ermittelt. So würde für jeden Untersuchungsteilnehmer in der Regel ein zu geringes Gewicht bestimmt und damit der wahre durchschnittliche Wert der Zielgesamtheit systematisch unterschätzt.

(B) Ein systematischer Fehler könnte auch dadurch entstehen, dass übergewichtige Personen weniger Bereitschaft zur Studienteilnahme zeigen als normal- und untergewichtige. Als Folge wären Übergewichtige in der Stichprobe unterrepräsentiert, so dass ihr Anteil geringer geschätzt wird, als er in der Zielpopulation ist.

Wegen der unbekannten Situation in der Grundgesamtheit kann in der Regel die Verzerrung nicht berechnet werden, jedoch ist wie in obigem Beispiel häufig eine Aussage über Richtung und Größenordnung einer Verzerrung möglich, wenn die Quellen der Verzerrungsmechanismen bekannt sind. Durch eine sorgfältige Planung und Durchführung von Studien kann dann einer Entstehung von Verzerrungseffekten vorgebeugt werden.

Üblicherweise unterscheidet man *drei Typen von Verzerrungen*, nämlich

- den Information Bias, d.h. die Verzerrung durch fehlerhafte Information (siehe obiges Beispiel A),
- den Selection Bias, d.h. die Verzerrung durch (nicht zufällige) Auswahl (siehe obiges Beispiel B) und
- den Confounding Bias, d.h. die Verzerrung durch mangelhafte Berücksichtigung von Störgrößen,

die im Wesentlichen durch die Art ihrer Entstehung gekennzeichnet sind.

Jede Art von Bias kann nun wiederum verschiedene Ursachen haben, die sich je nach Untersuchungsziel und -typ voneinander unterscheiden. Wegen der Bedeutung dieser Einflüsse werden wir bei der Beschreibung verschiedener epidemiologischer Studientypen (siehe Abschnitt 3.3) sowie der Studienplanung (siehe Kapitel 4) hierauf noch im Detail eingehen.

3.2 Datenquellen für epidemiologische Häufigkeitsmaße

Im vorherigen Abschnitt wurden grundlegende Konzepte bei der Betrachtung einer ätiologischen Fragestellung sowie bei der Abgrenzung von Populationen behandelt. Allerdings wurde noch nicht auf die Frage eingegangen, wie die zur Berechnung epidemiologischer Maße

zahlen notwendigen Informationen tatsächlich beschafft werden. Daher werden im Folgenden mögliche Informations- oder Datenquellen zur Beantwortung epidemiologischer Fragen vorgestellt.

3.2.1 Primär- und Sekundärdatenquellen

Werden zur Beantwortung einer epidemiologischen Fragestellung eigens Daten erhoben, spricht man von einer Primärdatenerhebung. In allen anderen Fällen wird versucht, auf die kostengünstigere und zeiteffizientere Alternative der Nutzung von bereits – in der Regel nicht zu Forschungszwecken – erhobenen Daten, den so genannten Sekundärdaten, zurückzugreifen. Beide Datenquellen werden zunächst unter grundsätzlichen Gesichtspunkten beschrieben, bevor in den folgenden Abschnitten ausführlicher auf konkrete Beispiele für Sekundärdatenquellen eingegangen wird.

Primärdaten

Primärdaten sind dadurch gekennzeichnet, dass der Datennutzer für das Erhebungskonzept und die Erhebungsmethodik der Informationen selbst verantwortlich ist und dass die Informationen gezielt für den jeweiligen Studienzweck gesammelt werden. Hierzu zählen sämtliche Daten, die im Rahmen von konkreten Studien an einzelnen Individuen erhoben werden (siehe hierzu insbesondere auch Abschnitt 3.3 sowie Kapitel 4 und 5).

Die prinzipiellen Vorteile einer eigenverantwortlichen Datenerhebung und Studiendurchführung sind offensichtlich, denn hier kann in der Regel eine in Methodik und Inhalt genau auf die jeweilige Fragestellung zugeschnittene Informationsbeschaffung erfolgen. Daher sind Primärdaten diejenigen mit der höchsten inhaltlichen Relevanz.

Sekundärdaten

Sekundärdaten sind dadurch charakterisiert, dass ihre Erhebung unabhängig von einer konkreten epidemiologischen Fragestellung erfolgt. Nach dieser Definition stellen z.B.

- Daten aus Primärerhebungen anderer,
- Daten aus speziellen Untersuchungsprogrammen (Monitoring),
- Daten aus speziellen Überwachungsprogrammen (Surveillance),
- Daten, die im Rahmen von Verwaltungsprozessen gesammelt werden,
- Daten aus der amtlichen Statistik

Sekundärdaten dar. Diese Daten werden heute in der Regel elektronisch in Datenbanken gespeichert, sind damit technisch verfügbar und stellen einen wertvollen Datenpool dar, der unter bestimmten Voraussetzungen die schnelle Beantwortung epidemiologischer Fragestellungen erlaubt, zumindest aber Basisinformationen für eine sorgfältig geplante epidemiologische Studie liefern kann.

Der *Vorteil der Nutzung von Sekundärdaten* liegt neben der *Kostengünstigkeit* vor allem in den in der Regel *großen Populationen*, die zur Verfügung stehen. Sollen z.B. Ursachen für sehr seltene Krankheiten untersucht werden, so können beispielsweise die administrativen Datenbanken der gesetzlichen Krankenkassen eine solche Untersuchung aufgrund der großen Anzahl an Versicherten ermöglichen. Darüber hinaus machen diese großen Datenbestände es auch möglich, regionale Unterschiede beispielsweise anhand der Daten der amtlichen Statistik zu untersuchen. Ein weiterer Vorteil ergibt sich dadurch, dass viele Sekundärdaten wie etwa aus der amtlichen Statistik für *lange Zeiträume* vorhanden sind. Diese erlauben somit die Analyse von Langzeitentwicklungen.

Darüber hinaus kann dank solcher Routinedatensammlungen oft von relativ *vollständigen Datengrundlagen* ausgegangen werden. Dies gilt besonders für amtliche Datensammlungen, die etwa für die Mortalitätsstatistik genutzt werden. Vor allem bei Daten, die in amtlichen Verwaltungsprozessen generiert werden, ist eine gesetzliche Grundlage vorhanden, die neben der Vollständigkeit auch Detailinformationen erlaubt, die wegen der großen *Vertraulichkeit* bei aktiven freiwilligen Datenerhebungen häufig vorenthalten werden.

Dennoch darf die Vielzahl verschiedenartiger Datenquellen, die möglicherweise epidemiologisch relevante Informationen enthalten, nicht darüber hinwegtäuschen, dass deren Aussagekraft häufig nur eingeschränkt ist. Im Detail wollen wir in Abschnitt 3.2.3 im Zusammenhang mit der amtlichen Mortalitätsstatistik als ein typisches Beispiel hierauf eingehen. Grundsätzlich sollten aber folgende *Nachteile von Sekundärdaten* beachtet werden: Mängel der Datenerhebung sind oft nicht leicht zu erkennen bzw. abzuschätzen. Die vom Primärnutzer verwendeten Begriffsdefinitionen und -abgrenzungen können unter Umständen schwer nachvollziehbar sein und müssen für die eigene Analyse übernommen werden. Der Entscheidung für die Verwendung von Sekundärdaten sollte deshalb die Frage vorausgehen, ob sich die Sekundärdatenquelle prinzipiell für den Untersuchungsgegenstand eignet.

Beispiel: Eine häufig benutzte Datenquelle ist die amtliche Mortalitätsstatistik (siehe auch Abschnitt 3.2.3). Wollte man diese etwa nutzen, um eine Aussage über die Inzidenz einer Erkrankung zu machen, so ist das nur dann sinnvoll, wenn die Erkrankung und der Todesfall kausal und zeitlich eng miteinander in Beziehung stehen und wenn die Erkrankung mit großer Sicherheit zum Tod führt.

So kann bei Unfällen mit Todesfolge oder bei Erkrankungen mit äußerst schlechter Prognose (z.B. Lungenkrebs) eine Nutzung der Mortalitätsstatistik durchaus vernünftig sein. Dagegen macht es wenig Sinn, sie bei Krankheiten mit guten Heilungschancen (z.B. Brustkrebs) zu verwenden, denn hier stehen Inzidenz und Mortalität nicht mehr in einem direkten Zusammenhang. Tatsächlich ließ sich in der Vergangenheit eine Abnahme der Brustkrebsmortalität beobachten, obwohl die Inzidenz zunahm.

Neben der Frage nach der grundsätzlichen Eignung können *drei Hauptnachteile bei der Nutzung von Sekundärdaten* identifiziert werden. Diese kann man unter den Schlagworten

- fehlende Nennerinformation,
- fehlendes Wissen über Risikofaktoren sowie
- unzureichende Standardisierung

zusammenfassen.

Die fehlende Nennerinformation, d.h. die *fehlenden Kenntnisse über die Zusammensetzung der Population unter Risiko*, macht in vielen Fällen die Nutzung von Routinedatensammlungen nur bedingt oder gar nicht möglich.

Beispiel: Ein klassisches Beispiel der Nutzung von Sekundärinformationen sind die Datenbestände aus Patientenkollektiven von Ärzten oder Kliniken. Bei Daten aus solchen Klinikregistern fehlt häufig der Bezug zur Risikopopulation, d.h., es ist zwar bekannt, welche Erkrankungen und Therapien vorliegen; man weiß allerdings meist nicht, welche Zielgesamtheit dem Arzt bzw. der Klinik eigentlich zuzuordnen ist. Durch die willkürliche Wahl einer ungeeigneten Risikopopulation ist dann bei Auswertungen auf dieser Datenbasis die Gefahr einer Verzerrung epidemiologischer Maßzahlen sehr hoch.

Werden beispielsweise in einer Spezialklinik für Krebserkrankungen in einem Jahr doppelt so viele Krebskranke eingeliefert wie im vorausgegangenen Jahr, so ist die Frage nach einer gestiegenen Inzidenzdichte nicht ohne Weiteres zu beantworten. Da anzunehmen ist, dass kompliziertere Fälle an eine große Klinik überwiesen werden, müsste dieser Klinik das Einzugsgebiet der Patienten bekannt sein, um die zugrunde liegende Risikopopulation bestimmen zu können. Im Beispiel wäre die gestiegene Zahl der Einlieferungen auch mit einer Vergrößerung des Einzugsgebietes der Klinik zu erklären, z.B. weil die Klinik die diesbezügliche Abteilung ausgebaut hat oder weil ein überregional als Koryphäe bekannter Chefarzt in die Klinik gewechselt hat.

In manchen Fällen ist der Mangel an Nennerinformation dadurch ausgleichbar, dass man eine Vergleichs- oder Kontrollgruppe in die Untersuchung mit einbezieht. Dann kann zwar keine repräsentative, für die interessierende Bezugspopulation gültige Aussage mehr gemacht werden, aber durch den Vergleich beider Gruppen können wichtige epidemiologische Erkenntnisse gewonnen werden. Zumindest muss dann natürlich die Vergleichbarkeit beider Gruppen gewährleistet sein, was u.U. schwer zu erreichen ist.

Immer dann, wenn Sekundärdaten dazu genutzt werden sollen, eine Ursache-Wirkungs-Beziehung zu beschreiben, ergibt sich als weiterer Hauptnachteil ihrer Nutzung die *mangelnden Kenntnisse zur Verteilung zugehöriger Risikofaktoren*, denn diese werden bei Routineuntersuchungen in der Regel nicht erfasst. Ein wichtiges Beispiel dafür sind die Meldungen gemäß Infektionsschutzgesetz.

Beispiel: Im Rahmen des Infektionsschutzgesetzes werden über ein standardisiertes Meldewesen eine vorliegende Infektionskrankheit bzw. der entsprechende Erreger, einige (pseudonymisierte) Personendaten wie Geschlecht, Alter, Wohnort und Umfeldsituation, der Verlauf der Erkrankung, die Diagnose, der Infektionsweg und ggf. die Verbindung zu anderen Erkrankungsfällen erfasst. Eine detaillierte Erfassung von Risikofaktoren erfolgt nicht (siehe auch Abschnitt "Datensammlung nach Infektionsschutzgesetz").

Will man nun das Auftreten von Erkrankungen wie z.B. der Giardiasis (eine Infektion des Dünndarms, die vom Parasiten *Giardia intestinalis* verursacht wird) nach Regionen erklären, so kann man das Krankheitsgeschehen etwa wie in Abb. 3.4 als Inzidenz nach Regierungsbezirken darstellen. Diese Darstellung hat allerdings keinerlei kausalen Bezug, denn das Hauptrisiko einer Infektion mit *Giardia intestinalis* besteht im Konsum entsprechend kontaminierter Lebensmittel oder Wasser. Informationen hierzu liegen aber weder in den Daten des Robert Koch-Instituts noch an anderer Stelle vor, so dass man das erkennbare räumliche Muster so nicht direkt erklären kann.

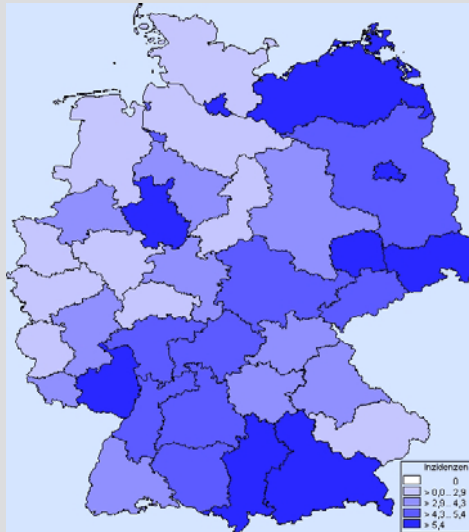


Abb. 3.4: Inzidenz von Giardiasis in den Regierungsbezirken in Deutschland im Jahr 2009 (Quelle: RKI, SurvStat@RKI)

Ein letzter wichtiger Nachteil bei der Nutzung von Sekundärdaten ist die *mangelnde Standardisierung* der Erhebungsmethodik. Es ist beispielsweise eine bekannte Alltagserfahrung, dass sich die Krankheitsdiagnose von Arzt zu Arzt unterscheidet und dass selbst Erhebungsfomulare nicht zu einer gleichen Diagnoseformulierung beitragen. Offiziell findet dieses Phänomen dann auch bei den ständig notwendigen Revisionen zu den ICD-Codes (International Statistical Classification of Diseases and Related Health Problems, WHO 1992) seinen Niederschlag, bei denen jeder Krankheit eine Zeichenfolge zugeordnet wird. Diagnostischer Fortschritt und ärztliche Freiheit führen somit häufig zu unterschiedlichen Bewertungen, so dass gerade bei Routineerhebungen kaum standardisierte Angaben zu erwarten und daher größere Ungenauigkeiten in Kauf zu nehmen sind.

3.2.2 Daten des Gesundheitswesens in Deutschland

Die Wahl der Datenquelle in der sekundärepidemiologischen Forschung ist neben der grundsätzlichen Frage nach der Relevanz für die zu untersuchende Problemstellung auch von Gesichtspunkten wie Datenzugänglichkeit und Verfügbarkeit, Aggregationsgrad, Dokumentationsgenauigkeit usw. abhängig. Während in der Vergangenheit diese technischen Fragen der Verfügbarkeit häufig entscheidend für ihre Nutzung waren, erscheint dies heute durch den vermehrten elektronischen und teilweise sogar öffentlichen Zugang zu diesen Datenquellen von geringerer Bedeutung zu sein.

Diese Frage soll daher nicht weiter diskutiert werden. Ohne Anspruch auf Vollständigkeit wird im Folgenden vielmehr ein Überblick über einige in Deutschland vorhandene Datenquellen gegeben und der Schwerpunkt dabei auf die fachlichen Fragen gelegt.

Fragebogenteil zur Gesundheit im Rahmen des Mikrozensus

Der gesetzlich verankerte Mikrozensus wird in der Regel jährlich durchgeführt. Für die zufällig ausgewählten Bürger besteht für den Hauptteil des Fragebogens Auskunftspflicht. Ein Muster des Fragebogens findet man beim Statistischen Bundesamt (2011). Die Erhebung erstreckt sich auf die gesamte Wohnbevölkerung in Deutschland. Dazu gehören alle in Privathaushalten und Gemeinschaftsunterkünften am Haupt- und Nebenwohnsitz in Deutschland lebenden Personen, unabhängig davon, ob sie die deutsche Staatsbürgerschaft besitzen oder nicht. Nicht zur Erhebungsgesamtheit gehören Angehörige ausländischer Streitkräfte sowie ausländischer diplomatischer Vertretungen mit ihren Familienangehörigen. Personen ohne Wohnung (Obdachlose) werden im Mikrozensus nicht erfasst. Mit einem Auswahlatz von 1% der bundesdeutschen Wohnbevölkerung ist der Mikrozensus eine sehr umfangreiche Stichprobe. Der Fragebogen zur Gesundheit wird dabei nur alle vier Jahre vorgelegt. Die Daten des Mikrozensus werden regelmäßig – tabellarisch zusammengefasst – publiziert.

Prinzipiell wird im Mikrozensus eine Periodenprävalenz erhoben, da sich die Fragen nach Krankheiten auf den Erhebungstag und die vier vorangegangenen Wochen beziehen. Da nicht nach einer Neuerkrankung, sondern nur nach Vorliegen einer Krankheit innerhalb der Periode gefragt wird, ist eine vierwöchige Inzidenzdichte nicht zu schätzen.

Neben dem Alter und dem Geschlecht wird das Rauchen als potenzieller Risikofaktor für viele Erkrankungen erhoben. Zudem kann die Angabe des Berufs einen gewissen Hinweis auf Risiken des Berufslebens geben.

Trotz dieser wesentlichen Basisinformation werfen die Mikrozensusdaten Probleme bei deren epidemiologischer Nutzung auf. So ist im Gegensatz zu anderen Fragebogenteilen die Auskunftserteilung zu den Gesundheitsfragen freiwillig, wird deshalb oftmals nicht ausgefüllt und ist folglich nicht uneingeschränkt als repräsentativ zu bezeichnen. Des Weiteren ist die Erfassung einer Multimorbidität nicht möglich, da bei Vorliegen mehrerer Krankheiten nur die subjektiv schwerste berichtet werden soll. Abgesehen von der problematischen Selbstdiagnose führt dies auch zu einer systematischen Untererfassung leichterer Krankheiten.

Datensammlung nach Infektionsschutzgesetz

Im Rahmen des Infektionsschutzgesetzes besteht in Deutschland bei 53 Erregern die Pflicht, dass Ärzte oder Untersuchungslabore eine Infektionserkrankung bzw. das Auftreten eines Erregers an das zuständige Gesundheitsamt bzw. das Robert Koch-Institut melden. Über ein standardisiertes Meldewesen werden dazu die Erkrankung bzw. der Erreger, einige (pseudonymisierte) Personendaten wie Geschlecht, Alter, Wohnort und Umfeldsituation, der Verlauf der Erkrankung, die Diagnose, der Infektionsweg und ggf. die Verbindung zu anderen Erkrankungsfällen erfasst. Eine detaillierte Erfassung von Risikofaktoren erfolgt nicht. Die Daten sind in aggregierter Form jederzeit öffentlich zugänglich (vgl. RKI. SurvStat@RKI 2010).

Da Ärzte und Untersuchungslabore jeden bekannt gewordenen Einzelfall den zuständigen Gesundheitsämtern melden, ist es prinzipiell möglich, eine Inzidenzmaßzahl anzugeben.

Problematisch ist hier allerdings eine mögliche Untererfassung der Fallzahlen aufgrund einer Nichtmeldung durch Ärzte bzw. durch Nicht-Arztbesuch. Man denke in diesem Zusammenhang an Salmonelleninfektionen, die bei leichterem Krankheitsverlauf häufig nicht erkannt werden, da ein Arzt gar nicht konsultiert wird. Diese Untererfassung ist je nach Infektionserreger erheblich. So berichtet das Robert Koch-Institut z.B. im Zusammenhang mit Infektionen durch Röteln von einer Unterfassung, so dass ggf. die wahre Anzahl von Erkrankungen um einen Faktor 10 höher liegen könnte (vgl. RKI-Ratgeber Infektionskrankheiten, 2010). Eine populationsbezogene epidemiologische Nutzung dieser Datenbestände ist nur sinnvoll, wenn wirklich von einer vollständigen Erfassung ausgegangen werden kann.

Datensammlungen der Krankenkassen

Für Verwaltungszwecke führen Krankenkassen Dateien über ihre Mitglieder mit einigen grundlegenden demographischen Angaben sowie über abgerechnete Leistungen (Leistungsdaten). Seit 2004 werden von den gesetzlichen Krankenkassen neben demographischen Angaben ambulante Diagnosen, abrechnungsfähige Verordnungen und stationäre Diagnosen personenbezogen zusammengeführt. Dabei erfolgt die Erfassung der abrechnungsfähigen Leistungen wie in Abb. 3.5 illustriert.

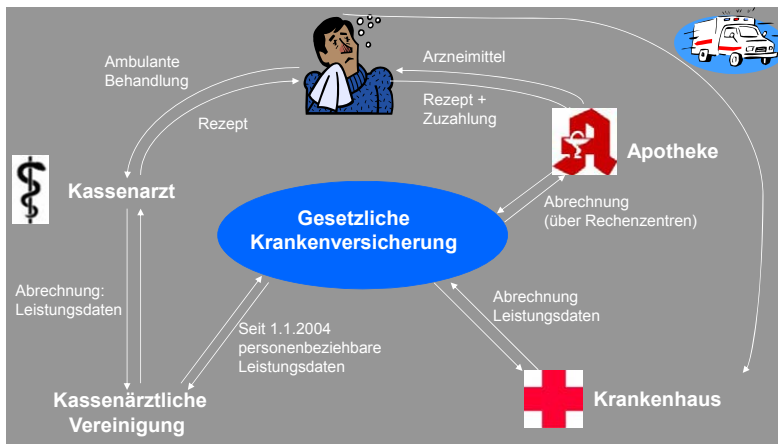


Abb. 3.5: Erfassungswege von abrechnungsfähigen Leistungen der gesetzlichen Krankenkassen

Zu den Daten, die von den Krankenkassen erfasst werden, gehören im Einzelnen

- Stammdaten:
Pseudonym des Versicherten, Geschlecht, Geburtsdatum, Wohnort (PLZ), Staatsangehörigkeit, Familienstand, ausgeübter Beruf des Mitglieds, Versichererstatus, Beitragsklasse, Versicherungsart (selbst/mitversichert), Eintritts- und Austrittsdatum, Austrittsgrund (z.B. Tod),

- Arzneimitteldaten:
Pseudonym des Versicherten, Abgabedatum, Abrechnungsmonat/-jahr, Ausstellungsdatum, Pharmazentralnummer, Anz. Packungen (Mengenfaktor), Kosten, Zuzahlung, Behandelnder Arzt (ID-Nr.),
- Stationäre Leistungsdaten:
Pseudonym des Versicherten, Art der stationären Maßnahme, Beginn und Ende der stationären Behandlung, Entbindungsdatum, Diagnosen (ICD-10), Aufnahmegewicht im 1. Lebensjahr, Aufnahmegrund, Entlassungsgrund, Operationen- und Prozeduren (OPS), Krankenhaus (ID-Nr.),
- Ambulante ärztliche Leistungsdaten:
Pseudonym des Versicherten, Diagnosen (ICD-10 vierstellig), Abrechnungsquartal/-jahr, Diagnosesicherheit, Seitenlokalisierung, Abgerechnete Gebührenposition (Mengenfaktor und Datum), Beginn und Ende der Behandlung, Entbindungsdatum, Behandelnder Arzt (ID-Nr.).

Da 85% der Bevölkerung in Deutschland in den gesetzlichen Krankenversicherungen versichert sind, liegt also für den größten Teil der Bevölkerung ein umfangreicher Datensatz zur Versorgungssituation mit einheitlicher Grundstruktur vor.

Diese Daten sind streng vertraulich und können nur unter äußerst restriktiven Bedingungen epidemiologisch genutzt werden. Inwieweit einzelne gesetzliche oder private Kassen repräsentativ sind oder etwa als eigenständige Populationen unter Risiko betrachtet werden sollten, muss dabei im Einzelnen überprüft werden. Es macht daher Sinn, die Daten verschiedener Kassen für epidemiologische Studien z.B. zur Untersuchung seltener Arzneimittelnebenwirkungen zusammenzuführen.

Mit der Nutzung dieser Daten sind auch Einschränkungen verbunden, die u.a. daraus resultieren, dass kaum Informationen zu Confoundern vorliegen. Ist eine Begleitmedikation als Confounder zu betrachten, so lässt sich dies in den Daten finden. Direkte Angaben z.B. zu Rauchen oder Übergewicht finden sich allerdings nicht. Weiterhin wird die Qualität ambulanter Diagnosen vor allem bei weniger schweren Erkrankungen angezweifelt, da z.B. die Diagnosen von Art und Umfang der abrechnungsfähigen Leistungen abhängen können. Erschwerend kommt hinzu, dass sich die Regelungen für abrechnungsfähige Leistungen im Zeitverlauf ändern. Weiterhin sind Abgabedatum und Menge von verordneten Arzneimitteln erfasst, allerdings nur von solchen, die abrechnungsfähig sind. Damit bleiben Selbstverordnungen von frei verkäuflichen Arzneimitteln unberücksichtigt. Nicht beantwortet werden kann die Frage, ob die abgegebenen Medikamente tatsächlich eingenommen wurden ("Compliance").

Trotz dieser Einschränkungen sind insbesondere die stationären Daten in Verbindung mit den demographischen und den Verordnungsdaten eine wertvolle Quelle zur Beantwortung von Fragen der medizinischen Versorgung und des Morbiditätsgeschehens. Gerade bei schwerwiegenden Erkrankungen wie z.B. kardiovaskulären Ereignissen, Krebserkrankungen und schweren Verletzungen, die praktisch immer zu einem stationären Aufenthalt führen, ist von einer weitgehend vollständigen und validen Erfassung auszugehen. Ein weiteres Plus dieser Daten ist die vollständige Abdeckung aller Altersgruppen und die Fortschreibung lon-

itudinaler Daten auf Individualniveau, die die Angabe von Inzidenzen und die Beobachtung von Langzeitfolgen ermöglichen.

Register

Eine systematische Erfassung von Krankheiten kann auch durch Registrierung erreicht werden. Um möglichst alle Neuerkrankungsfälle und unter Umständen auch Verläufe zu erfassen, besteht von Seiten eines solchen Registers Kontakt mit Kliniken, niedergelassenen Ärzten und Nachsorgeeinrichtungen. Pauschal sind *Inzidenzregister* bzw. *Morbiditätsregister* mit statistisch-epidemiologischer Zielrichtung und *Klinikregister* mit eher medizinisch-therapeutischer Zielsetzung zu unterscheiden. So genannte *Krankenhaus-Informationssysteme (KIS)* können dabei durchaus als Kombination beider Registerformen aufgefasst werden. Da alle diese Register personenbezogene Daten enthalten, ist die Nutzung dieser Daten strengen Datenschutzbestimmungen unterworfen und ein freier Zugriff auf Individualdaten nicht möglich. Etablierte Register wie z.B. Krebsregister publizieren diese aber im Rahmen der allgemeinen Gesundheitsberichterstattung, so dass zumindest aggregierte Daten öffentlich zugänglich sind (vgl. z.B. RKI 2009).

Grundsätzlich ist es möglich, mit Registerdaten Inzidenzen zu ermitteln. Wie genau Inzidenzen mit den Inzidenzregistern (z.B. Krebsregister) geschätzt werden, hängt entscheidend von der Bewältigung des Problems ab, Fehler durch doppelte, unvollständige oder fehlende Meldungen zu vermeiden. Dieses Problem ist in den Bundesländern unterschiedlich ausgeprägt, da die Gesetzgebung in den Bundesländern verschieden ist. So ist z.B. in einigen Ländern eine gesetzliche Grundlage für ein Melderecht, z.B. durch Pathologen, ohne die Notwendigkeit eines persönlichen Einverständnisses, gegeben.

Ein weiterer Nachteil einiger Register liegt in der Abgrenzung der Risikopopulation. Dies gilt vor allem für Daten aus Klinikregistern. Das in Abschnitt 3.2.1 beschriebene Problem zur Feststellung des Einzugsgebietes einer Klinik mag hierzu als typisches Beispiel dienen.

Erhebung bei Ärzten

Da niedergelassene Ärzte als diejenige Instanz des Gesundheitswesens gelten können, bei denen Erkrankungsfälle als Erstes dokumentiert werden, können auch Erhebungen bei Ärzten, wie sie von unterschiedlichen Trägern vorgenommen werden, als Sekundärdatenquelle betrachtet werden. Das Patientenkollektiv ausgewählter Arztpraxen wird dann analog zu Klinikregistern als ein so genanntes **Sentinel** aufgefasst und Krankheiten bzw. spezielle Symptome werden systematisch erfasst. Auch für diese Daten gelten strenge Datenschutzbestimmungen, und eine öffentliche Nutzung ist in der Regel nicht möglich.

Je nach Aufbau des Sentinels kann die Inzidenzrate theoretisch geschätzt werden, indem die Krankengeschichten der Patienten aller Ärzte in einer definierten Population erhoben werden. Beschränkt man sich sinnvollerweise auf eine Stichprobenerhebung, so ist allerdings auch hier von der gleichen Problematik wie bei Klinikregistern auszugehen. Auf der Basis der ausgewählten Praxen bekommt man eine Schätzung der Häufigkeit von Erkrankungen in

einer Population, die weitgehend unbekannt ist, da der genaue Einzugsbereich einzelner Arztpraxen in der Regel unbekannt ist.

Neben diesen Sentinels werden Daten der Kassenärztlichen Vereinigungen z.B. zu Diagnosen und Verordnungen im Zentralinstitut für die Kassenärztliche Versorgung in Deutschland zusammengeführt. Diese Daten unterliegen nicht den oben angesprochenen Selektionseffekten, da sie auf Bundesebene erfasst werden und in Bezug auf die gesetzlich Versicherten vollständig sind. Die zur Verfügung stehenden, umfangreichen Datenbestände aus der gesamten Bevölkerung erlauben u.a. eine Einschätzung der ärztlichen Versorgung in Deutschland. Darüber hinaus können z.B. Prävalenzen und Inzidenzen von Erkrankungen geschätzt werden. Allerdings dürfen die Verordnungsdaten und die ambulanten Daten nicht auf Individualbasis zusammengeführt werden. Ein weiterer Nachteil ergibt sich aus dem bereits oben angesprochenen Problem der zum Teil unzureichenden Qualität ambulanter Diagnosen.

Vorsorgeuntersuchungen

Neben der Erkrankung oder dem Symptom als eigentlichem epidemiologischen Endpunkt bieten auch Vorsorgeuntersuchungen wie z.B. das Mammographiescreening Möglichkeiten zur sekundärepidemiologischen Nutzung. Diese können einerseits als Teil der Erfassung der Leistungsdaten der Krankenkassen interpretiert werden (siehe oben), sind andererseits aber auch als eigenständige Datenquelle von Bedeutung, wobei sie in beiden Fällen als nicht öffentlich angesehen werden müssen und somit den bereits erwähnten Nutzungsrestriktionen unterworfen sind.

Die Qualität von Daten, die im Rahmen von Vorsorgeuntersuchungen anfallen, ist abhängig von der Fehlerquote der Vorsorgediagnostik, d.h. davon, wie viele Kranke fälschlicherweise als gesund bzw. Gesunde als krank klassifiziert werden (siehe auch Verzerrung durch Fehlklassifikation in Kapitel 4). Darüber hinaus stellt sich durch die freiwillige Inanspruchnahme einer Vorsorgeeinrichtung die Frage, inwieweit die Ergebnisse repräsentativ sind. So ist z.B. vorstellbar, dass Personen mit ersten Krankheitsanzeichen oder aber einem besonders ausgeprägten Gesundheitsbewusstsein die Vorsorgeeinrichtung häufiger in Anspruch nehmen. Dies würde somit zu einer Selektion bestimmter Personengruppen führen, was eine Verzerrung der Untersuchungsergebnisse zur Folge hätte.

Weitere Datenquellen

Weitere hier nicht näher erläuterte Quellen für sekundärstatistische Morbiditätsdaten in der Bundesrepublik Deutschland sind z.B. Daten aus *schulärztlichen Untersuchungen*, aus *Musterungen*, zur *Berufs- oder Erwerbsunfähigkeit*, aus *arbeitsmedizinischen Überwachungsuntersuchungen* und vieles mehr. Wie die Diskussion zu den obigen Datenquellen gezeigt hat, sind neben einigen spezifischen Problemen die in Abschnitt 3.2.1 aufgeführten grundsätzlichen Nachteile fast überall anzutreffen. Vor der Nutzung solcher Daten ist deshalb eine eingehende Prüfung ihrer Eignung für den Untersuchungszweck unabdingbar. Der an der Nutzung von Sekundärdaten im Detail interessierte Leser sei deshalb u.a. auf Schach et al. (1985) und Swart & Ihle (2005) verwiesen.

3.2.3 Nutzung von Sekundärdaten am Beispiel der Mortalitätsstatistik

Die beschriebenen Probleme bei der Nutzung von Sekundärdaten sollen am Beispiel der Mortalitätsstatistik weiter vertieft werden. Der zeitliche Verlauf und räumliche Vergleich von Mortalitätsdaten haben in der empirischen Gesundheitsforschung große Bedeutung. Im zeitlichen Verlauf können Hinweise über das Auftreten neuer oder das Verschwinden alter Risiken gegeben werden. Regional gegliederte Todesursachenstatistiken können Hypothesen für Krankheitsursachen generieren, sind aber auch ein Mittel zur Überwachung des Krankheitsstandes und der Umweltsituation. Hier sind alters- und geschlechtsspezifische Sterberaten wichtige Indikatoren.

Todesbescheinigungen

Alle der Weltgesundheitsorganisation WHO angeschlossenen Länder kennen die Meldepflicht der Sterbefälle, so dass in den meisten Ländern von einer Vollerhebung dieser ausgegangen werden kann. Todesbescheinigungen (Leichenschaucheine) haben in erster Linie juristische und demographische Bedeutung, dienen aber auch als Informationsgrundlage für die Todesursachenstatistik. Ein standardisiertes Formular (siehe z.B. Abb. 3.6) soll ein einheitliches Vorgehen beim Ausfüllen der Todesbescheinigungen gewährleisten.

In Deutschland sind die Bescheinigungen nach dem Gesetz über die Statistik der Bevölkerungsbewegung und die Fortschreibung des Bevölkerungsstandes sowie nach dem Personenstandsgesetz vorgeschrieben. Sie werden von Ärzten ausgefüllt, von Gesundheitsämtern gesammelt und in der Regel dort auch archiviert und über die Standesämter an die statistischen Landesämter gemeldet.

Als Todesursache wird eine Ursachenkette berichtet, bestehend aus

- der unmittelbaren Todesursache,
- dem dieser Todesursache zugrunde liegenden Ereignis/Zustand und möglichst auch
- dem dafür ursächlichen Grundleiden (siehe Abb. 3.6 (b), Ziffern 15–20).

Unter diesen Voraussetzungen ist anzunehmen, dass Todesbescheinigungen als gut nutzbare epidemiologische Datenquelle dienen können, denn *Vollzähligkeit* (Meldepflicht), *Standardisierung* (einheitliches Formblatt) und die dem Tod vorausgehende *medizinische Ursachenkette* können unterstellt werden. Dennoch ergibt sich bei der Nutzung von Mortalitätsdaten eine Vielzahl von Schwierigkeiten, die deren epidemiologische Interpretation einschränken. Diese Nachteile werden im Folgenden näher beschrieben.

Todesbescheinigung NRW - Nichtvertraulicher Teil -		Blatt 1	Untere Gesundheitsbehörde über Standesamt	Die Todesbescheinigung ist unverzüglich auszuhändigen.	Zutreffendes bitte ankreuzen und / oder ausfüllen														
1. Personalangaben			Wird vom Standesamt ausgefüllt Standesamt Sterbefall beurkundet, Sterbebuch-Nr.: Eingang vorgemerkt, Vorkenn-Liste-Nr.: Erdbestattung ☐ Feuerbestattung ☐																
1 Name (ggf. Geburtsname), Vorname(n) 2 Straße 3 Hausnummer 4 PLZ, Wohnort, Kreis 5 Geburtsdatum 6 Geburtsort, Kreis																			
7 Geschlecht ☐ männlich ☐ weiblich 8 Identifikation nach ☐ eigener Kenntnis ☐ Personalausweis/Reisepass ☐ Angaben Angehöriger/Dritter ☐ nicht möglich (kein Eintrag unter 1 - 6)																			
2. Feststellung des Todes/Sterbezeitpunkt			9 ☐ Nach eigenen Feststellungen ☐ Nach Angaben Angehöriger/Dritter am																
10 Falls Sterbezeitpunkt nicht bestimmbar: Leichenauffindung am			<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td>Tag</td><td>Monat</td><td>Jahr</td><td>um</td><td>Stunden</td><td>Minuten</td></tr> <tr> <td>Tag</td><td>Monat</td><td>Jahr</td><td>um</td><td>Stunden</td><td>Minuten</td></tr> </table>			Tag	Monat	Jahr	um	Stunden	Minuten	Tag	Monat	Jahr	um	Stunden	Minuten		
Tag	Monat	Jahr	um	Stunden	Minuten														
Tag	Monat	Jahr	um	Stunden	Minuten														
Ende des Durchschreibeverfahrens! Bitte die Blätter 2 ff. wegklappen und gesondert ausfüllen!																			
11 ☐ Sterbeort 12 ☐ Auffindeort, falls nicht Sterbeort 13 ☐ als tote Leibesfrucht geboren ☐ in der Geburt gestorben Zusatzangabe für totgeborene oder in der Geburt gestorbene Leibesfrüchte von mindestens 500 g (als Sterbezeitpunkt gilt der Geburtszeitpunkt): Name der Einrichtung (des Krankenhauses/Heimes o.ä.) Straße, Hausnummer PLZ, Ort oder Stempel der Einrichtung (falls vorhanden)																			
14 3. Todesart Gibt es Anhaltspunkte für äußere Einwirkungen, die den Tod zur Folge hatten? (z. B. Selbsttötung, Unfall, Tötungsdelikt, auch durch äußere Einwirkungen evtl. mitverursachte Todesfälle, Spättodesfälle nach Verletzung) ☐ nein wenn nein, Todesart ☐ natürlich oder ☐ ungeklärt, ob natürlich/nichtnatürlicher Tod ☐ ja (Wenn ja oder ungeklärt, im Vertraulichen Teil, Blätter 2 ff. Ziff. 20 [Epikrise] nähere Hinweise (falls möglich))																			
15 4. Warnhinweise Liegen Hinweise dafür vor, dass die/der Verstorbene an einer übertragbaren Krankheit nach § 6 oder § 7 Infektionsschutzgesetz (einschließlich HIV) erkrankt war? ☐ ja ☐ nein																			
16 Sind besondere Verhaltensmaßnahmen bei der Aufbewahrung, Einsargung, Beförderung und Bestattung zu beachten? ☐ nein ☐ ja, welche?																			
17 ☐ Sonstiges (z. B. Gefährdung durch Giftstoffe/Chemikalien):																			
Fortsetzung des Durchschreibeverfahrens!																			
18 Bescheinigt aufgrund meiner sorgfältigen Untersuchung am																			
Ich habe in meine Untersuchung die gesamte Körperoberfläche mit Rücken, Kopfhaut und allen Körperöffnungen einbezogen:																			
Ort, Datum Unterschrift <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td>Tag</td><td>Monat</td><td>Jahr</td><td>um</td><td>Stunden</td><td>Minuten</td><td>Uhr.</td></tr> <tr> <td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> </table> ☐ ja ☐ nein						Tag	Monat	Jahr	um	Stunden	Minuten	Uhr.							
Tag	Monat	Jahr	um	Stunden	Minuten	Uhr.													

Abb. 3.6: (a) Todesbescheinigung NRW, Blatt 1

Todesbescheinigung NRW - Vertraulicher Teil -		Blatt 5 Für den Arzt zur Dokumentation		Zutreffendes bitte ankreuzen und / oder ausfüllen ☒												
1. Personalangaben																
1	Name (ggf. Geburtsname), Vorname(n)															
2	Straße		3 Hausnummer													
4	PLZ, Wohnort, Kreis															
5	Geburtsdatum		6 Geburtsort, Kreis													
7 Geschlecht ☐ männlich ☐ weiblich 8 Identifikation nach ☐ eigener Kenntnis ☐ Personalausweis/Reisepass ☐ Angaben Angehöriger/Dritter ☐ nicht möglich (kein Eintrag unter 1 - 6)																
2. Feststellung des Todes/Sterbezeitpunkt																
9 ☐ Nach eigenen Feststellungen		☐ Nach Angaben Angehöriger/Dritter am		<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 10%;">Tag</td><td style="width: 10%;">Monat</td><td style="width: 10%;">Jahr</td><td style="width: 10%;">um</td><td style="width: 10%;">Stunden</td><td style="width: 10%;">Minuten</td> </tr> <tr> <td> </td><td> </td><td> </td><td> </td><td> </td><td> </td> </tr> </table>	Tag	Monat	Jahr	um	Stunden	Minuten						
Tag	Monat	Jahr	um	Stunden	Minuten											
10 Falls Sterbezeitpunkt nicht bestimmbar: Leichenauffindung am		<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 10%;">Tag</td><td style="width: 10%;">Monat</td><td style="width: 10%;">Jahr</td><td style="width: 10%;">um</td><td style="width: 10%;">Stunden</td><td style="width: 10%;">Minuten</td> </tr> <tr> <td> </td><td> </td><td> </td><td> </td><td> </td><td> </td> </tr> </table>		Tag	Monat	Jahr	um	Stunden	Minuten							
Tag	Monat	Jahr	um	Stunden	Minuten											
Sichere Zeichen des Todes 11 ☐ Totenflecke ☐ Totenstarre ☐ Fäulnis ☐ Hirntod ☐ Nicht mit dem Leben vereinbare Verletzungen 12 Reanimationsbehandlung durchgeführt ☐ ja ☐ nein Wer hat die Todesursache festgestellt? 13 ☐ Behandelnder Arzt ☐ Nicht behandelnder Arzt nach Angaben des behandelnden Arztes ☐ Nicht behandelnder Arzt ohne Angaben des behandelnden Arztes 14 Zuletzt behandelt durch Hausarzt/Krankenhaus (-abteilung) Name des Krankenhauses/Arztes o. ä. Straße, Hausnummer PLZ, Ort oder Stempel (falls vorhanden)																
Todesursache (nicht Endzustände wie Atemstillstand, Herz-Kreislaufversagen) ungefähre Zeitspanne vom Krankheitsbeginn bis Tod *)																
15 a) Unmittelbare Todesursache:																
16 b) Dies ist eine Folge von b1*)																
17 b2*)																
18 c) Hierfür ursächliche Grunderkrankungen: *)																
19 II Mit zum Tode führende Krankheiten ohne Zusammenhang mit dem Grunderkrankungen: *)																
*) ausfüllen, soweit dem Arzt möglich																
20 Epikrise Weitere Angaben zur Todesart (Blatt 1, Ziffer 14), falls erforderlich (z. B. Unfall, Vergiftung, Gewalteinwirkung, Selbsttötung sowie Komplikationen medizinischer Behandlung) Außen Ursache der Schädigung (Angaben über den Hergang); bei Vergiftung zusätzlich Angabe des Mittels																
21 Unfallkategorie (bitte nur Untergruppe ankreuzen) ☐ Schulunfall (ohne Wegeunfall) ☐ Sport- oder Spielunfall (nicht in Haus oder Schule) ☐ Wegeunfall ☐ Arbeits- oder Dienstunfall (ohne Wegeunfall) ☐ häuslicher Unfall ☐ sonstiger Unfall ☐ Verkehrsunfall ☐ unbekannt																
24 Diagnose durch Obduktion gesichert? ☐ nein ☐ ja																
25 Liegt der Obduktionsbefund bei? ☐ nein ☐ ja																
26 Bei ungeklärter Identität der Leiche: Bei nichtnatürlicher oder ungeklärter Todesart: Polizei unterrichtet? ☐ ja ☐ nein																
Bei Frauen, deren Alter eine Schwangerschaft nicht ausschließt 22 Liegt eine Schwangerschaft vor? ☐ nein ☐ ja ☐ unbekannt Bestehen Anzeichen für eine Schwangerschaft in den letzten 12 Monaten? ☐ ja ☐ nein																
Bei Kindern unter 1 Jahr und Totgeborenen 27 Wo wurde das Kind geboren? ☐ im Krankenhaus ☐ zu Hause ☐ sonstiger Ort Geburtsgröße ☐ cm ☐ g 28 Mehrlingsgeburt? ☐ nein ☐ ja 29 Bei in den ersten 24 Stunden gestorbenen Neugeborenen: ☐ Frühgeburt in der Schwangerschaftswoche ☐ Lebensdauer: ☐ volle Stunden ☐ unbekannt																
30 Bescheinigt aufgrund meiner sorgfältigen Untersuchung am Tag ☐ Monat ☐ Jahr ☐ um ☐ Stunden ☐ Minuten ☐ Uhr Ich habe in meine Untersuchung die gesamte Körperoberfläche mit Rücken, Kopfhaut und allen Körperöffnungen einbezogen: ☐ ja ☐ nein Ort, Datum Unterschrift																

Abb. 3.6: (b) Todesbescheinigung NRW, Blatt 5

ICD-Codierung

Die Verschlüsselung der Todesursache erfolgt in den statistischen Landesämtern nach dem so genannten ICD-Code (International Statistical Classification of Diseases and Related Health Problems, WHO 1992). Hier wird nach Hauptklassen der Krankheit gegliedert eine Zeichenfolge zugeordnet. Zurzeit wird die zehnte Revision dieses internationalen Schlüssels verwendet (Braun & Dickmann 1993, Klar et al. 1993). Für die Todesursachenstatistik wird die dem Tode zugrunde liegende Erkrankung aus den diesbezüglichen Klartextangaben ermittelt und verschlüsselt.

Die regelmäßigen Revisionen des ICD-Schlüssels erweisen sich bei der Beurteilung der Mortalitätsstatistik als äußerst problematisch. Dies gilt besonders dann, wenn die Mortalitätsdaten in ihrer zeitlichen Entwicklung betrachtet werden. Ein eindrucksvolles Beispiel dieser Problematik geben Wichmann & Molik (1984).

Beispiel: Bei der Umstellung von ICD-8 auf ICD-9 wurde erstmalig der Code 798.0 "plötzlicher Tod im frühen Kindesalter bei nicht näher bezeichneter Ursache" (engl. Sudden Infant Death Syndrome, SIDS) eingeführt. Von der Umstellung im Jahr 1979 bis zum Ende der 1980er Jahre stieg die berichtete Zahl dieser Todesursachen ständig an. Während des gleichen Zeitraums ist allerdings ein Rückgang in der Position 769 "Respiratory-Distress-Syndrome" zu beobachten, so dass die Summe beider Positionen ungefähr konstant geblieben ist (siehe Abb. 3.7).

Dennoch führte der scheinbar dramatische Anstieg berichteter Todesfälle zu Anfragen und Spekulationen, dass die zunehmende Umweltverschmutzung verantwortlich für die vermehrte Anzahl von Todesfällen sei (nach Wichmann & Molik 1984, Pesch 1992).

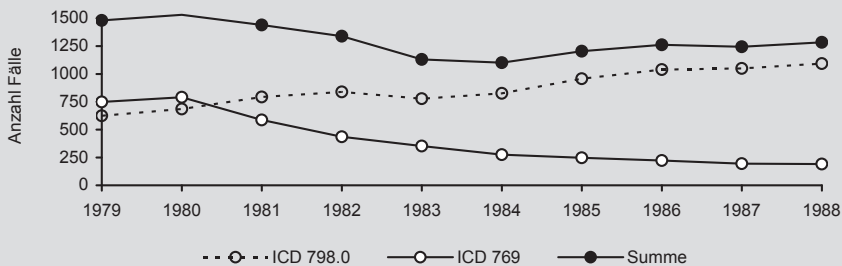


Abb. 3.7: Anzahl Fälle von plötzlichem Kindstod (ICD 798.0) und "Respiratory-Distress-Syndrome" (ICD 769) in Deutschland, von 1979 bis 1988 (Quelle: Statistisches Bundesamt, Einzelnachweis Pesch 1992)

Grundleiden und Kausalkette

Selbst wenn die Entwicklung einer Krankheit mittels der Klartextangaben auf den Todesbescheinigungen recht ausführlich dargestellt ist, kann die Todesursache nicht immer eindeutig

festgelegt werden. So kann auf der einen Seite nicht immer eindeutig entschieden werden, was Ursache und was Wirkung innerhalb der angegebenen Kausalkette ist.

Beispiel: Stirbt ein Diabetiker beispielsweise an Herzversagen, so stellt sich etwa die Frage, ob die Diabetes-Erkrankung zum Herzversagen geführt hat und damit die eigentliche Todesursache darstellt oder ob das Herzversagen an sich als Todesursache zu sehen ist und Diabetes nur als Begleiterkrankung zu werten ist.

Auf der anderen Seite ist in der Regel nicht nur eine Ursache für eine Krankheit verantwortlich, sondern ein Zusammenspiel mehrerer Faktoren. Um der Problematik, d.h. welche Erkrankung unmittelbar zum Tode geführt hat, zu begegnen, hat die WHO (1979) einen Katalog verfasst, der Zweifelsfragen regeln hilft. Eine gewisse Unsicherheit und insbesondere die fehlende Kenntnis solcher Spezialregeln in der Einzelfallroutine lassen allerdings vermuten, dass die Kausalkette hier häufig falsch dargestellt wird. Dies wird umso schwieriger, je älter die Verstorbenen sind, für die die Angabe eines Grundleidens oftmals nur noch theoretisch möglich ist. Vielmehr liegt bei älteren Menschen meist eine Multimorbidität vor, so dass eine einzelne Erkrankung nicht mehr als Grundleiden erkennbar ist. In einer solchen Situation führt dies häufig zu vollkommen unspezifischen Angaben wie Code 797 "Altersschwäche", was im Sinne der WHO-Regeln nur in den seltensten Fällen als richtige Codierung angesehen werden darf. Die Berichtspraxis zeigt allerdings, dass – wohl auch bedingt durch einen Rückgang der Diagnoseintensität bei zunehmendem Alter – diese Kategorie wesentlich stärker besetzt ist als erwartet werden darf.

Die Qualität der Angaben auf einer Todesbescheinigung hängt zudem ganz entscheidend von den Umständen ab, unter denen sie ausgefüllt wurde. Erfolgt dies z.B. im Rahmen der stationären Aufnahme im Kontext umfangreicher medizinischer Diagnostik oder durch den Hausarzt, der die Krankenvorgeschichte des Verstorbenen gut kennt, so sind die Angaben als valide einzustufen. Ist der Leichen schauende Arzt jedoch ein eilig in der Nacht herbeigerufener Notarzt, dem der Patient und dessen Vorgeschichte unbekannt sind, so sind die gemachten Angaben zum Grundleiden und zur dem Tode zugrunde liegenden Erkrankung meist weniger vollständig und valide. Durch Obduktion ermittelte Todesursachen besitzen dagegen die höchste Validität.

Diagnosezuverlässigkeit

Trotz Standardisierung treten also falsche Diagnosen oder Kausalketten, unspezifische Angaben, die im eigentlichen Sinne zwar nicht falsch, aber zumindest wenig präzise sind, Auslassungen oder einfache Formfehler (Unlesbarkeit) in der Berichtspraxis in nicht unerheblichem Maße auf. So ist z.B. die *Validität für Todesursachen* wie Herz-Kreislauf-Erkrankungen oder Atemwegserkrankungen in der Praxis schlechter als für Krebserkrankungen, da Letzteren in der Regel eine umfangreiche klinische Diagnostik vorangeht.

Ein weiteres Problem tritt durch die so genannte "Diagnosefreiheit" auf, die dazu führt, dass die *Diagnosegewohnheiten* einzelner Ärzte oder auch zwischen einzelnen Regionen oder Ländern unter Umständen stark voneinander abweichen.

Beispiel: Ein besonders offensichtliches Beispiel hierzu findet man beim internationalen Vergleich der Säuglingssterblichkeit, der in der Regel durch kulturelle und gesellschaftspolitische Unterschiede in verschiedenen Ländern stark eingeschränkt ist. So ist z.B. beim Vergleich der Säuglingssterblichkeit in Deutschland und Italien darauf zu achten, dass der Anteil von Todesfällen vor und während der Geburt in Italien deutlich geringer ist als in Deutschland, da in Italien häufig eine Taufzeremonie durch die Eltern noch gewünscht wird.

Dass im internationalen Vergleich Abweichungen auftreten, ist ein bekanntes Phänomen, aber auch im nationalen Bereich sind solche Diagnoseabweichungen zu beachten. Neben wissenschaftlichen "Schulen" können hier vor allem administrativ bedingte Unterschiede, etwa zwischen den Bundesländern, aber natürlich auch die in der zeitlichen Betrachtung relevanten Verbesserungen in der Diagnosetechnik genannt werden. Diese ist in Deutschland als föderalem Land teilweise sogar gesetzlich oder durch entsprechende Verwaltungsvorschriften unterschiedlich zwischen den Bundesländern geregelt, denn die (Tier-) Gesundheitspolitik liegt im Verantwortungsbereich der Länder. So existieren in Deutschland z.B. keine einheitlichen Formulare für die Meldung der Todesursachen.

Möglichkeiten und Grenzen der Mortalitätsstatistik

Immer dann, wenn die epidemiologische Nutzung von Sekundärdaten auch der Ursachenforschung dienen soll, ist die Frage zu stellen, inwieweit solche Ursachen miterfasst worden sind. Eine solche Erfassung liegt für die Mortalitätsstatistik in der Regel nicht vor und somit ist eine Nutzung im strengen Sinne eigentlich nicht möglich.

Dennoch wird häufig versucht, beispielsweise im Zusammenhang mit Krebsatlanten, einen Kausalitätsbezug zwischen den aggregierten Daten der Mortalitätsstatistik und umweltbedingten Risikofaktoren herzustellen, wobei sowohl zeitliche als auch räumliche Bezüge unterstellt werden.

Auf diese Weise hergestellte Zusammenhänge müssen allerdings sehr zurückhaltend interpretiert werden, wie das obige Beispiel zum plötzlichen Tod im Kindesalter zeigt. Ein Anstieg oder Abfall der Mortalitätsrate ist somit nicht notwendigerweise mit dem Ansteigen oder Abfallen etwa einer Luftschadstoffkonzentration oder anderer im zeitlichen Trend veränderter Variablen in Verbindung zu bringen. Gleiches gilt natürlich auch für räumliche Vergleiche, wie das Beispiel zur Säuglingssterblichkeit in Deutschland und Italien demonstriert hat.

Der Grund für die Grenzen der Interpretierbarkeit von Mortalitätsdaten liegt also im Mangel an personenbezogenen Informationen zu Expositionen, die als mögliche Krankheitsursachen interessieren. Außer dem Geschlecht und dem Alter werden auf der Todesbescheinigung keine weiteren als potenzielle Risikofaktoren interpretierbaren Daten erhoben. So sind beispielsweise das Rauch- und Ernährungsverhalten für eine Vielzahl von Erkrankungen vom Herz-Kreislauf-Bereich über Atemwegserkrankungen bis hin zur Krebsmorbidity als Hauptrisiken anzusehen, aber nicht dokumentiert. Kausalitätsaussagen für Krankheiten, bei denen etwa die genannten Risikofaktoren als ursächlich anzusehen sind, sind folglich unzulässig (ein Beispiel für die ansonsten entstehende Fehlinterpretation der Mortalitätsstatistik wird im Abschnitt 3.3.1 ausführlich diskutiert).

Neben den nicht dokumentierten Risikofaktoren muss im Sinne der Ursachenforschung weiterhin berücksichtigt werden, dass die *Zeit zwischen dem ersten Auftreten* einer Erkrankung und dem *hierdurch bedingten Versterben* durchaus sehr *unterschiedlich* sein kann. Da zu krankheitsbezogenen Zeitunterschieden noch weitere durch die unterschiedlichen Wirksamkeiten von Therapien hinzukommen, ist der zeitliche Zusammenhang von Inzidenz und Mortalität bei manchen Krankheiten gering. Damit sind der zeitlichen Interpretation von Mortalitätsdaten zusätzliche Grenzen gesetzt.

Nicht nur wegen der genannten Einschränkungen sollten Mortalitätsdaten nicht einfach ungeplant genutzt werden. Die Auswertung von Mortalitäts- (und anderen sekundärepidemiologischen) Daten muss grundsätzlich einem formalem Studienprotokoll unterliegen. Wir werden auf entsprechende Leitlinien noch ausführlich in Kapitel 7 zurückkommen, wollen aber schon hier auf Swart & Ihle (2005) hinweisen, die neben den hier benannten Aspekten eine ausführliche Dokumentation der Grundlagen, Methoden und Perspektiven der Auswertung von Sekundärdaten vorgenommen haben.

Trotz dieser Probleme hat die Mortalität eine sehr wichtige Bedeutung, denn sie kann als einziger epidemiologischer Endpunkt aufgefasst werden, der zweifelsfrei und eindeutig feststellbar ist. Zudem wird die Mortalität bereits über Jahrhunderte dokumentiert, so dass sie eine hohe gesellschaftliche Akzeptanz besitzt. Es sollte daher zunächst geprüft werden, ob diese Daten bereits zur Beantwortung einer epidemiologischen Fragestellung ausreichend sind, bevor eine primärepidemiologische Studie durchgeführt wird. Insbesondere können Mortalitätsdaten z.B. zur Hypothesengenerierung genutzt werden.

Todesbescheinigungen stellen sogar eine zentrale Datenquelle für einen bestimmten Typ epidemiologischer Primärdatenerhebungen dar. Es handelt sich dabei um historische Kohortenstudien, bei denen die Sterblichkeit z.B. von Angehörigen einer bestimmten Berufsgruppe oder eines Industriezweigs mit der Allgemeinbevölkerung verglichen wird (siehe Abschnitt 3.3.4).

3.3 Typen epidemiologischer Studien

Kann das Untersuchungsziel nicht durch die Nutzung von Sekundärdaten erreicht werden, so ist zu prüfen, ob dies durch eine Primärdatenerhebung möglich ist. Je nach Untersuchungsziel können bzw. müssen verschiedene Studiendesigns zur Anwendung kommen, deren Aussagefähigkeit allerdings unterschiedlich ist. Dabei lassen sich generell experimentelle und beobachtende Ansätze differenzieren.

Zu den *experimentellen Untersuchungen* zählen neben klinischen *Therapiestudien* vor allem so genannte *Interventionsstudien*. Charakteristisch ist in beiden Ansätzen, dass zu vergleichende Gruppen der Untersuchungspopulation dem in der ätiologischen Fragestellung formulierten Expositionsfaktor geplant ausgesetzt werden. Dies gilt sowohl für den Vergleich von klinischen Therapien wie auch für die Intervention in der Bevölkerung, die z.B. aktiv eine Reduktion von Risikofaktoren anstreben (z.B. Ernährungsprogramme zur Vorbeugung von Übergewicht).

Bei *Beobachtungsstudien* wird dagegen kein gezielter Eingriff in die Exposition vorgenommen, sondern nur beobachtet, wie Krankheiten und Exposition in Beziehung stehen. Je nach zeitlicher Richtung der Betrachtung unterscheidet man dabei Querschnittsstudien sowie retrospektive bzw. prospektive Untersuchungen.

Als ein erster Schritt einer Ursachenanalyse werden allerdings häufig *Assoziationen von Maßzahlen* der Erkrankungshäufigkeit (basierend auf Sekundärdaten) zu aggregierten Daten von möglichen Einflussfaktoren (z.B. Pro-Kopf-Verbrauch bestimmter Nahrungsmittel oder geographische Charakteristika) berechnet. Solche Assoziationsrechnungen werden als ökologische Relationen bezeichnet.

3.3.1 Ökologische Relationen

Bei **ökologischen Relationen** handelt es sich um Auswertungstechniken, die als statistische Untersuchungseinheit nicht einzelne Individuen verwenden, sondern *Aggregationen*, d.h. Zusammenfassungen von Individuen einer Population. Übliche Aggregationsstufen sind im räumlichen Sinne administrative Zusammenfassungen wie Gemeinden, Landkreise, Regierungsbezirke, (Bundes-) Länder o.Ä. bzw. im zeitlichen Sinne über ein Jahr aufkumulierte Zahlen. Häufig werden auf Basis dieser Aggregationsstufen Assoziations-, Korrelations- und Regressionsrechnungen (siehe Anhang S) zu möglichen Risikofaktoren durchgeführt, die auf demselben Niveau aggregiert sind.

Die *Vorteile* einer solchen Vorgehensweise sind offensichtlich, denn häufig liegen aggregierte Daten für eine Vielfalt unterschiedlichster Krankheits- und Risikofaktoren bereits vor. Ökologische Relationen können daher als die statistische Auswertungstechnik der *sekundär-epidemiologischen Datenanalyse* bezeichnet werden. Da eine Aggregation die Bildung einer Summe bzw. eines Mittelwertes bedeutet (z.B. Anzahl aller Fälle, durchschnittliche Mortalität, durchschnittliche Schadstoffbelastung etc.), wird zudem die Variabilität der Faktoren stabilisiert (siehe Anhang S) und Zusammenhänge werden zum Teil erst hierdurch sichtbar.

Dass eine ökologische Relation leicht zu *Fehlinterpretationen* führen kann, hatten wir schon am Beispiel des plötzlichen Kindstods in Abschnitt 3.2.3 gesehen.

Beispiel: Klassische Beispiele für Fehlinterpretationen findet man vor allem bei der Betrachtung von räumlichen Assoziationen. Ein solches Beispiel hatten wir bereits in Abschnitt 3.2.1 kennen gelernt, in dem die räumliche Verteilung humaner Infektionen durch den Parasiten *Giardia intestinalis* dargestellt wurde. Abb. 3.4 zeigt ein deutliches räumliches Muster mit durchschnittlich höheren Inzidenzen im Süden und Osten.

Den Hauptrisikofaktor der Giardiasis stellt der Konsum von belasteten Lebensmitteln oder Trinkwasser dar. Diese Infektion ist durchaus auch in Deutschland möglich, jedoch ist ein erheblicher Anteil der Infektionen auf eine Exposition im Ausland, insbesondere in tropischen Ländern, zurückzuführen.

Will man somit das in Deutschland gefundene räumliche Muster in eine ökologische Relationsrechnung eingehen lassen, so wäre es erforderlich, die Verteilung der Konsumgewohnheiten und vor allem den Aufenthalt im Ausland in gleicher räumlicher Struktur zur Verfügung zu stellen. Solche Daten liegen in Deutschland aber nicht vor, so dass sich eine beobachtete Assoziation mit anderen Faktoren als Scheinassoziation herausstellen könnte (siehe Abschnitt 3.1.1).

Beispiel: Bevor zu Beginn der 1990er Jahre analytische Studien zu den Risikofaktoren des Lungenkrebses in Deutschland durchgeführt wurden, waren ökologischen Relationen Gegenstand einer wissenschaftlichen wie öffentlichen Diskussion. Auf den vier Karten von Abb. 3.8 sind jeweils aggregiert auf Ebene westdeutscher Regierungsbezirke folgende seinerzeit verfügbare Daten dargestellt: (a) Lungenkrebssterblichkeit bei Frauen, (b) Anteile rauchender Frauen, (c) Luftverschmutzung an SO_2 und (d) Radonbelastung in Wohnungen.

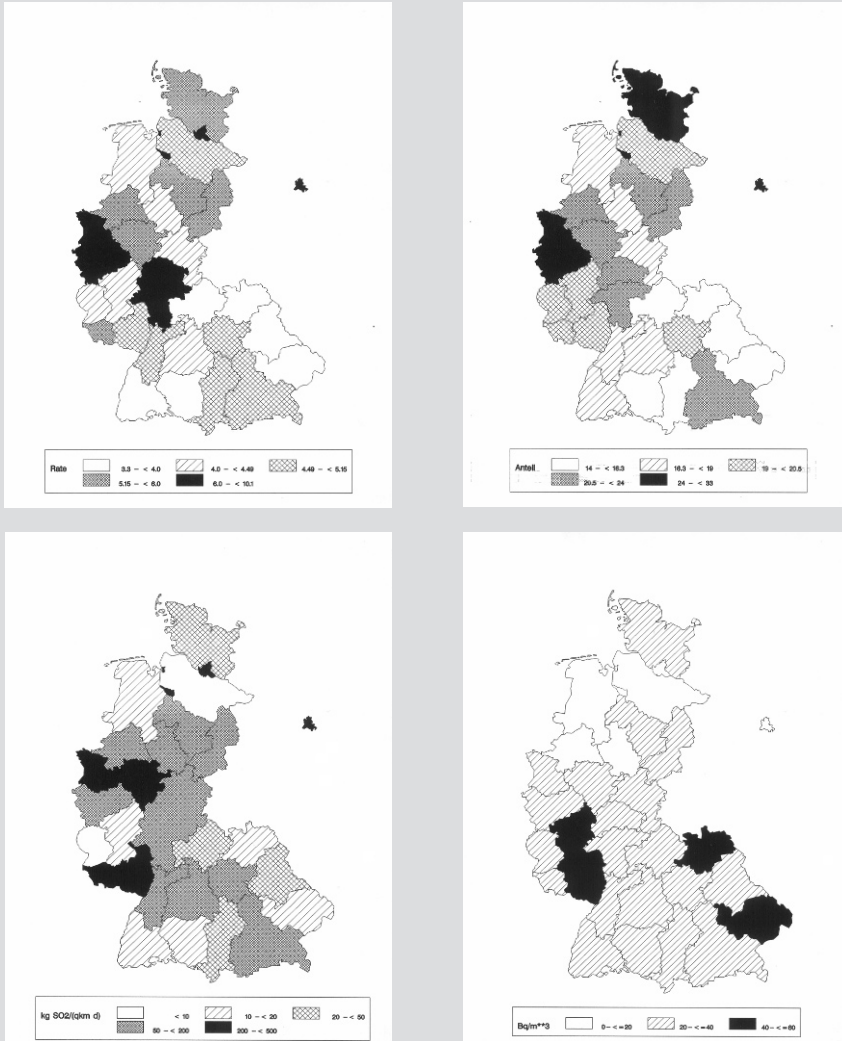


Abb. 3.8: Beispiel für verzerrte Schlussfolgerungen bei fehlender Berücksichtigung von Störvariablen.
 (a): Lungenkrebssterblichkeit bei Frauen je 100.000 (Quelle: Wichmann et al. 1991),
 (b): Raucheranteil bei Frauen in % (Quelle: Wichmann et al. 1991),
 (c): SO_2 -Emissionsdichte (Quelle: Wichmann et al. 1991),
 (d): Radon in Wohnungen (Quelle: Schmier 1984)

Die seinerzeit häufig gemachte Aussage, dass die Lungenkrebsmortalität durch die Luftverschmutzung verursacht sei, könnte beim Vergleich der Abbildungen (a) und (c) unterstellt werden. Eine andere Aussage ergäbe sich aus dem Vergleich der Abbildungen (a) und (d), die suggerieren, dass die Lungenkrebsmortalität nicht durch die Radonbelastung in Innenräumen verursacht sein kann.

Beide Interpretationen sind allerdings falsch, denn es besteht ein deutlicher Zusammenhang zur regionalen Verteilung des Zigarettenrauchens (b). Dieser ist, was viele Studien zeigen, kausal, so dass der Zusammenhang zur SO₂-Belastung bei Berücksichtigung des Rauchens stark abgeschwächt wird oder gar verschwindet. Der negative Zusammenhang zwischen der regionalen Verteilung der Radonbelastung in Wohnungen und der Lungenkrebssterblichkeit verschwindet oder wird umgekehrt, wenn das Rauchen in die Analyse einbezogen wird.

Abgesehen davon, dass statistische Assoziation für sich allein kein Kausalitätsnachweis ist (siehe hierzu insbesondere Abschnitt 3.1.1), führt eine einfache statistische Assoziations- oder Korrelationsrechnung somit leicht in die Irre, wenn eine ausreichende *Kenntnis über die Verteilung der wichtigsten Risikofaktoren* nicht vorliegt.

Wie das Beispiel verdeutlicht, können schwächere Risikofaktoren (Luftverschmutzung, Radon) falsch bewertet werden, wenn stärkere Risikofaktoren (Rauchen) unberücksichtigt bleiben. Im Beispiel entsteht diese Verzerrung dadurch, dass die Einflussgrößen ein ausgeprägtes regionales Muster aufweisen (Rauchen: In Industrieregionen und Städten wird mehr geraucht als auf dem Land; Luftverschmutzung: Die Luftqualität in Industrieregionen und Städten ist schlechter als auf dem Land; Radon: Die Radonbelastung ist in bergigen, dünner besiedelten Gebieten höher als in Ballungszentren).

Damit kann obige Form der ökologischen Relationsrechnung überhaupt nur dann sinnvoll interpretiert werden, wenn neben dem interessierenden Risikofaktor auch die Verteilung der wichtigsten anderen Risikofaktoren bekannt ist. Ansonsten besteht die Gefahr, einem *"ökologischen Trugschluss"* zu unterliegen. Dies ist der Grund, warum sich ökologische Relationen primär zur Hypothesengenerierung, nicht aber zur Prüfung von Hypothesen eignen. Neben den genannten Beispielen zum Auftreten von Infektionserkrankungen oder zur Auswirkung der Umweltbelastung auf die menschliche Gesundheit lassen sich zahlreiche weitere Beispiele finden, die zu einem ökologischen Trugschluss führen. Eine Auswahl von Beispielen aus epidemiologischer bzw. klinischer Sicht geben etwa Skabaneck & McCormick (1995).

3.3.2 Querschnittsstudien

Querschnittsstudien ("cross-sectional study") eignen sich dazu, die aktuelle Häufigkeit von Risikofaktoren und von Erkrankungen unabhängig vom Zeitpunkt ihres Eintritts – also die Prävalenz – in der Studienpopulation zu ermitteln. Daher wird dieser Studientyp auch als **Prävalenzstudie** bezeichnet.

Querschnittsstudien umfassen eine Auswahl von Individuen aus der *Zielpopulation zu einem festen Zeitpunkt (Stichtag)*. Für die ausgewählten Studienteilnehmer werden der *Krankheitsstatus* (Prävalenz) und die gegenwärtige oder auch frühere *Expositionsbelastung* (Expositionsprävalenz) *gleichzeitig* erhoben.

Da meist eine erhebliche Zeitspanne zwischen Exposition und Krankheitsausbruch liegt, ist diese Studienform zum *Kausalitätsnachweis von Risikofaktoren* im Rahmen der analytischen Epidemiologie nur geeignet bei *akuten Erkrankungen mit kurzer Induktionszeit* wie z.B. dem Auftreten einer Asthmasymptomatik bei Pollenflug. Querschnittsstudien sind deshalb vor allem ein Instrument der deskriptiven Epidemiologie. Sie dienen neben der Dokumentation eines Ist-Zustandes auch der Hypothesengenerierung.

Beispiel: Eine klassische Querschnittserhebung stellt der so genannte Bundesgesundheitsurvey dar (siehe Abschnitt 1.2). Hier werden in regelmäßigen zeitlichen Abständen mehrere Tausend Personen im Alter von 18 bis 79 Jahren durch das Robert Koch-Institut zu ihrem Gesundheitszustand befragt. Ihre Größe, ihr Gewicht und ihr Blutdruck werden gemessen sowie Blut und Urin untersucht. Da die Studie in einem relativ kurzen Zeitraum durchgeführt wird, kann von einer Erhebung zu einem Zeitpunkt ausgegangen werden, so dass die Prävalenz der erfassten Zielgrößen ermittelt werden kann.

Beispiel: Querschnittserhebungen finden vor allem auch in der Veterinärepidemiologie statt, z.B. bei der Erfassung von Zoonoseerregern in Lebensmittel liefernden Tieren (siehe Abschnitt 1.2). Untersucht man z.B. das Auftreten von bakteriellen Zoonoseerregern wie *Salmonella* spp, *Campylobacter coli* oder *Yersinia enterocolitica* in Beständen der Schweinemast, so ist eine Querschnittserhebung häufig die einzige mögliche Studienform, da die Mastzeiträume sehr kurz sind (das durchschnittliche Schlachtagter liegt bei etwa sechs Monaten). Daher werden z.B. Studien zum Zeitpunkt der Schlachtung (oder kurz davor) durchgeführt.

Auswertungsprinzipien

Das Konzept der Querschnittsstudie ist die klassische Form einer statistischen Stichprobenerhebung, da an einem festen Stichtag aus einer definierten Zielgesamtheit eine Untersuchungspopulation gewonnen wird. Im Idealfall wird diese Stichprobe mittels einer so genannten einfachen Zufallsauswahl gezogen. Die Studienpopulation kann in Erkrankte und Gesunde bzw. Exponierte und Nichtexponierte eingeteilt werden (siehe Abb. 3.9).

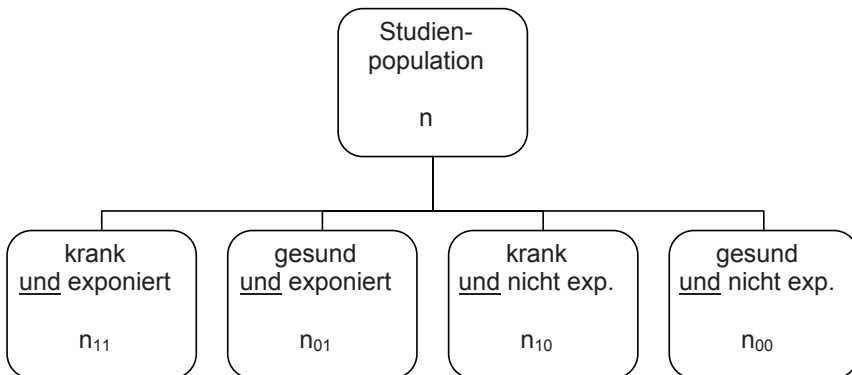


Abb. 3.9: Struktur einer Querschnittsstudie und Aufteilung der Studienteilnehmer

Durch diese Einteilung entsteht bei diesem Studiendesign somit die Vierfeldertafel aus Tab. 3.2.

Tab. 3.2: Beobachtete Anzahl von Kranken, Gesunden, Exponierten und Nichtexponierten in einer Querschnittsstudie

Status	exponiert (Ex = 1)	nicht exponiert (Ex = 0)	Summe
krank (Kr = 1)	n_{11}	n_{10}	$n_{1.} = n_{11} + n_{10}$
gesund (Kr = 0)	n_{01}	n_{00}	$n_{0.} = n_{01} + n_{00}$
Summe	$n_{.1} = n_{11} + n_{01}$	$n_{.0} = n_{10} + n_{00}$	$n = n_{..}$

In Tab. 3.2 stellt n_{ij} die *in der Studie beobachtete Anzahl* von Individuen im Krankheitsstatus i aus der Expositionsgruppe j , mit $i, j = 0, 1$, dar. Hierbei ist von *prävalenten beobachteten Anzahlen* an einem Stichtag auszugehen. Die Aufteilung dieser Anzahlen n_{ij} in der Tafel ist unter der Voraussetzung, dass eine einfache Zufallsauswahl realisiert wurde, zufällig; nur die Zahl n der Studienteilnehmer ist von vornherein festgelegt.

Damit ist es mit dieser Tafel möglich, **Schätzer für die Prävalenz** unter gleichzeitiger Angabe des Expositionsstatus anzugeben. Die Bezeichnung Schätzer soll dabei zum Ausdruck bringen, dass das Ergebnis, das aus der Studienpopulation berechnet wird, das (wahre, aber unbekannte) Ergebnis der Zielpopulation schätzt. Diese Schätzer für die beiden Expositionsgruppen lauten

$$\text{Schätzer Prävalenz (bei Exposition } j) = P(j)_{\text{Studie}} = \frac{n_{1j}}{n_{.j}}, j = 0, 1.$$

Da diesem Design jeglicher Inzidenzbezug fehlt, haben diese Maßzahlen keinen Risikobezug. Wenn man aber annehmen kann, dass bei der betrachteten Krankheit Inzidenz und Prävalenz sehr ähnlich sind, ist es üblich, einen Quotienten von Prävalenzen aus der Studie zu berechnen. Dieses Verhältnis definiert dann den Schätzer für den so genannten **Prävalenzquotienten** oder das **Prävalenzratio**

$$\text{Schätzer Prävalenzquotient} = PR_{\text{Studie}} = \frac{P(1)_{\text{Studie}}}{P(0)_{\text{Studie}}} = \frac{n_{11}/(n_{11} + n_{01})}{n_{10}/(n_{10} + n_{00})}.$$

Diese Größe ist kein relatives Risiko (Risk Ratio), sondern nur als eine *Approximation des relativen Risikos* zu verstehen, denn die Daten wurden zu einem Zeitpunkt erfasst und geben keine Neuerkrankungen in einem Zeitraum an.

Im Gegensatz zu diesem Quotienten erlaubt die Symmetrie der Vierfeldertafel die Angabe eines Schätzwerts für das Odds Ratio. Man erhält

$$\text{Schätzer Odds Ratio} = \text{OR}_{\text{Studie}} = \frac{n_{11}/n_{01}}{n_{10}/n_{00}} = \frac{n_{11} \cdot n_{00}}{n_{10} \cdot n_{01}}.$$

Beispiel: Auf Basis des Bundesgesundheits surveys berichten das RKI bzw. das Statistische Bundesamt regelmäßig über den Gesundheitszustand der deutschen Bevölkerung. Tab. 3.3. zeigt Ergebnisse des Bundesgesundheits surveys (BGS) 1998 bezüglich der Frage des Übergewichts. Die Weltgesundheitsorganisation spricht von Übergewicht, falls der Body-Mass-Index (Körpermasse in kg dividiert durch das Quadrat der Körperlänge in m²) größer als 25 ist.

Tab. 3.3: Beobachtete Anzahlen von Übergewicht nach Geschlecht im Bundesgesundheits survey 1998 (Quelle: RKI 1999)

Status	Männer	Frauen	Summe
übergewichtig (BMI ≥ 25)	2.299	1.919	4.218
normalgewichtig (BMI < 25)	1.148	1.702	2.850
Summe	3.447	3.621	7.068

Aus den Informationen aus Tab. 3.3 kann die Prävalenz des Übergewichts getrennt nach Männern und Frauen geschätzt werden. Die Werte lauten

$$P(\text{Männer})_{\text{BGS 1998}} = \frac{2.299}{3.447} = 0,667,$$

$$P(\text{Frauen})_{\text{BGS 1998}} = \frac{1.919}{3.621} = 0,530.$$

Hieraus ergibt sich als Quotient der beiden Studienprävalenzen

$$\text{PR}_{\text{BGS 1998}} = \frac{0,667}{0,530} = 1,258,$$

d.h., bei Männern ist die Prävalenz des Übergewichts 1,258-mal höher als bei Frauen. Drückt man diesen relativen Unterschied über die vergleichende Maßzahl des Odds Ratios der Studie aus, so erhält man den Wert

$$\text{OR}_{\text{BGS 1998}} = \frac{2.299 \cdot 1.702}{1.919 \cdot 1.148} = 1,776.$$

Auch diese Kennzahl zeigt somit das höhere Risiko zum Übergewicht bei Männern an, wobei das Odds Ratio wegen der hohen Prävalenz eine andere Größenordnung hat als der Prävalenzquotient (siehe hierzu auch Abschnitt 2.3.3).

Vor- und Nachteile von Querschnittsstudien

Der größte Vorteil von Querschnittsuntersuchungen besteht darin, dass man bei repräsentativer Zufallsauswahl von der Studienpopulation *auf die Zielpopulation schließen* und die Häufigkeit der Risikofaktoren ermitteln kann. Zudem lassen sich Querschnittsstudien im Gegensatz zu anderen primärepidemiologischen Untersuchungen je nach Untersuchungsumfang in (relativ) *kurzen Zeiträumen* durchführen.

Dagegen sind Querschnittsstudien zur Ursachenforschung von Krankheiten nur bedingt geeignet, da wegen der Prävalenzerhebung die *zeitliche Abfolge von Exposition und Krankheit unklar* bleiben kann, wobei das Problem der Krankheitsdauer und deren Auswirkung auf Prävalenz und Inzidenz eine besondere Rolle spielt (siehe Abschnitt 2.1.1).

Beispiel: Ein Beispiel ist die Wirkung von Fettleibigkeit (Body-Mass-Index ≥ 30) auf das Auftreten von Herzerkrankungen. Da adipöse Personen vorzeitig an anderen Todesursachen versterben, sind diese stark übergewichtigen Personen in einer Querschnittsstudie unterrepräsentiert: Damit kann bei der Betrachtung eines Prävalenzquotienten ein Unterschied unter Umständen verborgen bleiben.

Neben dieser Anfälligkeit der Prävalenz gegenüber Faktoren, die die Krankheitsdauer beeinflussen (Mortalität, Genesung), gibt es noch weitere Verzerrungsmechanismen, die typisch für diesen Studientyp sind. Darauf werden wir bei der Diskussion systematischer Fehler genauer eingehen (siehe Abschnitt 4.6).

Zusammenfassend lässt sich sagen: Bei *nicht zu seltenen, lang andauernden Krankheiten* (z.B. chronischer Bronchitis, Rheuma, chronischen Infektionen etc.) und *Dauergewohnheiten als Risikofaktoren* (z.B. Rauchen, Wohnen, Arbeitsplatz etc.) sind Querschnittsstudien durchaus sinnvoll und werden entsprechend häufig durchgeführt (siehe auch Abschnitt 1.2).

Bei *akuten Erkrankungen* eignen sich Querschnittsstudien sogar für die Bearbeitung ätiologischer Fragestellungen, sofern diese Erkrankungen in der Zielpopulation häufig sind und die in Frage kommenden Krankheitsursachen in engem zeitlichen Bezug zum Auftreten der Erkrankung stehen.

Für *seltene Krankheiten* sind Querschnittsstudien dagegen *wenig geeignet*, denn es müssten extrem große Studienkollektive in eine Untersuchung eingeschlossen werden, damit aus der Studie eine Aussage gewonnen werden kann. Daher wird in diesen Fällen meist ein anderer Studientyp favorisiert.

3.3.3 Fall-Kontroll-Studien

In einer **Fall-Kontroll-Studie** vergleicht man eine Gruppe von Erkrankten, die Fälle, mit einer Gruppe von nicht Erkrankten, den Kontrollen, hinsichtlich einer zeitlich vorausgegangenen Exposition. Von der Wirkung ausgehend wird somit nach potenziellen Ursachen geforscht, die dem Einsetzen der Wirkung um viele Jahre vorausgegangen sein können.

Grundsätzlich gilt somit für Fall-Kontroll-Studien, dass die Individuen einer definierten Population, die innerhalb eines bestimmten Zeitraums die Zielerkrankung bekommen haben, als Fälle auswählbar sind. Als Kontrollen sind Individuen derselben Population auswählbar, sofern sie im betrachteten Zeitraum unter Risiko standen, diese Erkrankung zu bekommen. Dabei muss die Auswahl der Studienteilnehmer unabhängig von ihrem jeweiligen Expositionsstatus erfolgen. Fall-Kontroll-Studien beginnen also mit der Erkrankung und ermitteln davon ausgehend retrospektiv die der Erkrankung vorausgehenden Expositionen.

Sind nun Fälle und Kontrollen in Bezug auf bekannte Risikofaktoren miteinander vergleichbar und unterscheiden sich beide Gruppen bezüglich der interessierenden Exposition, so interpretiert man dies als einen Hinweis auf eine ätiologische Beziehung zwischen der Exposition und der Krankheit. Die zeitliche Abfolge von vermuteter Ursache und Wirkung wird bei diesem Studientyp somit "rückblickend" oder "retrospektiv" ermittelt (siehe Abb. 3.10).

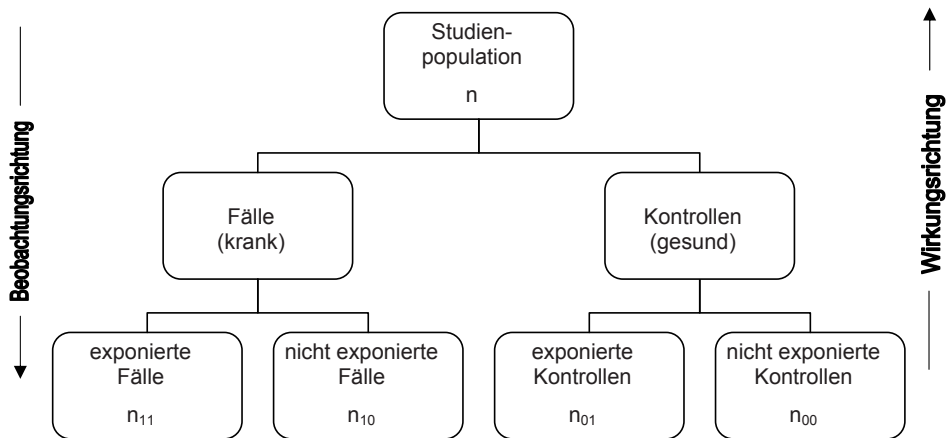


Abb. 3.10: Struktur einer Fall-Kontroll-Studie und Aufteilung der Studienteilnehmer

Die Entwicklung des Designs der Fall-Kontroll-Studie erfolgte insbesondere unter dem Aspekt, dass bei *Krankheiten, die selten sind und/oder lange Induktionszeiten aufweisen*, eine einfache Zufallsauswahl etwa im Rahmen einer Querschnittsstudie nur wenige oder gar keine Krankheitsfälle enthalten würde. Dadurch ist die inhaltliche Entwicklung von Fall-Kontroll-Studien eng mit der Epidemiologie von Krebserkrankungen verknüpft. Entsprechend finden sich dort viele Beispiele für diesen Studientyp.

Beispiel: Das Bronchialkarzinom ist in der Bundesrepublik Deutschland zwar mit einer Mortalitätsrate von ca. $MR = 80$ von 100.000 die mit Abstand häufigste Krebstodesursache bei Männern (Robert Koch-Institut 2009), doch ist trotz dieser für Krebserkrankungen extrem hohen Fallzahl die Inzidenz als gering anzusehen. So müssten bei den zu erwartenden Fallzahlen in einer Querschnittsstudie mehrere tausend Personen eingeschlossen werden, um eine hinreichende Anzahl von Lungenkrebskrankungen hierin zu beobachten.

Deshalb werden zur Ätiologie des Bronchialkarzinoms in der Regel Fall-Kontroll-Studien durchgeführt, da dadurch etwa über Spezialkliniken oder über Krebsregister eine ausreichende Zahl Erkrankter in die Studie eingeschlossen werden können.

Beispiel: Eine weitere Anwendung des Prinzips der Fall-Kontroll-Studie ergibt sich dann, wenn man mit einem vollkommen neuen, zunächst sporadisch auftretenden Krankheitsbild konfrontiert wird. So war es nahe liegend, als Ende der 1980er Jahre in Großbritannien die ersten Fälle der bovinen spongiformen Enzephalopathie (BSE) aufgetreten sind, diese neuen Fälle mit Kontrollen zu vergleichen, um Hinweise auf potenzielle Ursachen zu erhalten.

Als eine intuitive Möglichkeit zur Auswahl von Fällen bietet sich ein Prävalenzkonzept an, das an einem Stichtag Fälle definiert, egal wie lange der Beginn der Erkrankung zurückliegt. Diese Vorgehensweise ist allerdings aus zweierlei Gründen nicht sinnvoll. So sind einerseits insbesondere bei lang andauernden Krankheiten die langlebigen Fälle überrepräsentiert. Andererseits ist es gerade bei seltenen Krankheiten erforderlich, die Fälle über einen gewissen Zeitraum zu sammeln, um eine ausreichend große Fallzahl zu erreichen. Aus diesem Grund ist es empfehlenswert, nur inzidente, d.h. neu erkrankte Fälle in Fall-Kontroll-Studien einzuschließen. Dabei sollte das Zeitfenster zwischen Diagnose und Einschluss in die Studie umso kürzer sein, je kürzer die mittlere Überlebenszeit der Patienten mit der interessierenden Erkrankung ist.

Bei einer Vielzahl potenzieller Einflussfaktoren für die Entstehung einer Krankheit ist die zentrale Annahme des Fall-Kontroll-Designs die Vergleichbarkeit der Fall- mit der Kontrollgruppe, um den Einfluss der in der ätiologischen Hypothese betrachteten Exposition bestimmen zu können. Theoretisch wird diese Vergleichbarkeit dadurch erreicht, dass beide Studiengruppen aus derselben Zielgesamtheit gewonnen werden, so dass beim Fall-Kontroll-Design die besondere Schwierigkeit in der Definition der Zielpopulation und der adäquaten Auswahl der Studienpopulation liegt.

Vergleichbarkeit von Fall- und Kontrollgruppe

Die Frage, welche Individuen als Fälle und welche als Kontrollen dienen können, ist theoretisch eindeutig zu beantworten: Als Kontrollen kommen nur Individuen in Betracht, die der Population unter Risiko (Zielpopulation) angehören, aus der die Fälle stammen. Damit repräsentieren Kontrollen bezüglich der Verteilung der Expositionsvariablen die Population, aus der die Fälle stammen. Anders ausgedrückt bedeutet dies, dass Kontrollen diejenigen Individuen repräsentieren sollen, die, hätten sie die Krankheit entwickelt, auch als Fälle in die Studie eingeschlossen werden können.

Je nach Definition der Zielpopulation unterscheidet man populationsbezogene und auswahlbezogene Fall-Kontroll-Studien.

Populationsbezogene Fall-Kontroll-Studien

Eine **populationsbezogene Studie** erhält man dann, wenn sowohl die Fall- als auch die Kontrollgruppe repräsentativ für eine geographisch klar definierte Population sind, wie z.B. die Einwohner einer Stadt oder eines Landkreises, d.h. jeweils eine repräsentative Stichprobe aus der kranken bzw. krankheitsfreien Teilgruppe dieser Population darstellen.

Eine **populationsbezogene Fallgruppe** kann man nur dadurch erhalten, dass man die Gesamtheit aller Krankheitsfälle ermittelt und diese insgesamt oder eine repräsentative Stichprobe in die Studie einbezieht. Theoretisch benötigt man hierzu ein Register sämtlicher Erkrankungen, so dass diese häufig auch als **Registerstudie** bezeichnet wird. Ob ein solches Register zur Verfügung steht, hängt neben gesetzlichen Regularien von der entsprechenden Erkrankung ab.

Beispiel: Das so genannte Kinderkrebsregister der Universität Mainz registriert seit 1980 für Deutschland insgesamt alle Krebserkrankungen von Kindern unter 15 Jahren. Daneben gibt es verschiedene Krebsregister in Ländern und Regionen wie etwa das Krebsregister des Saarlandes, das als eines der Ersten in Deutschland seit 1970 alle Neuerkrankungen an Krebs innerhalb des Bundeslandes erfasst und dokumentiert. Neben Registern, die eine freiwillige Sammlung entsprechender Informationen darstellen, sind zudem bei einigen Erkrankungen Meldepflichten etabliert. Dies gilt z.B. für Infektionserkrankungen nach dem Infektionsschutzgesetz, die Deutschland weit beim Robert Koch-Institut gesammelt werden oder auch bei den meldepflichtigen Tierseuchen, die im System TierSeuchenNachrichten (TSN) des Friedrich-Loeffler-Instituts in einer zentralen Datenbank gespeichert werden.

Eine **populationsbezogene Kontrollgruppe** erhält man als repräsentative Stichprobe aus der definierten Zielpopulation, aus der die Fälle entstammen, unter Verwendung der in Abschnitt 3.1.2 beschriebenen Auswahlverfahren. In Deutschland hat es sich bewährt, die Zielpopulation geographisch über die Population von Kreisen und Gemeinden zu definieren, über die zum Teil umfangreiche soziodemographische Daten bei den statistischen Landesämtern verfügbar sind und aus denen für Forschungszwecke über die zuständigen Einwohnermeldeämter nach einem Zufallsverfahren eine nach Alter und Geschlecht stratifizierte Stichprobe der Wohnbevölkerung gezogen werden kann (siehe auch Abschnitt 4.2).

Abgesehen von dem meist hohen Aufwand solcher Erhebungen ist man zudem häufig mit dem Problem einer mangelnden Teilnahmebereitschaft konfrontiert, da epidemiologische Studien für die Teilnehmer stets auch einen gewissen Aufwand bedeuten (Befragung, u. U. körperliche Untersuchung, Probennahme, Messung etc.). Dies kann eine starke Selektion der Studienteilnehmer und eine mangelnde Repräsentativität zur Folge haben, was ggf. die Verallgemeinerbarkeit der gewonnenen Studienergebnisse einschränkt. Das heißt, trotz einer sorgfältigen Studienplanung und der Durchführung einer Stichprobenerhebung kann es sein, dass der Populationsbezug nicht mehr gegeben ist. Da dieses so genannte Non-Response-Problem erhebliche Auswirkungen auf die Qualität einer Studie haben kann, werden wir diesen Aspekt in Abschnitt 4.6.1 vertiefend behandeln.

Repräsentieren aber sowohl die Fall- als auch die Kontrollgruppe die entsprechenden Teilgruppen aus der Zielpopulation, so ist das Fall-Kontroll-Design sehr gut für die ätiologische Forschung einsetzbar.

Auswahlbezogene Fall-Kontroll-Studien

Eine **auswahlbezogene Fall-Kontroll-Studie** wird stark von pragmatischen Gesichtspunkten bestimmt. Ist die vollständige Erfassung aller Fälle einer definierten Population gar nicht oder nicht mit vertretbarem Aufwand möglich, so wählt man eine nach definierten Kriterien

gut erreichbare Gruppe von Fällen und definiert dann die Population, der diese Fälle entstammen. Aus dieser "post hoc" definierten Population sind die Kontrollen auszuwählen.

Praktisch gut durchführbar ist es, Patienten eines Krankenhauses mit einer diagnostizierten Krankheit als Fälle zu betrachten. Die Zeitspanne für die Rekrutierung der Fälle ist hier gewöhnlich relativ kurz und die Kooperationsbereitschaft der Personen groß. Die Schwierigkeit liegt dann in der Definition einer Kontrollgruppe, denn die Fälle eines einzelnen Krankenhauses stellen in der Regel keine zufällige Stichprobe aller Fälle aus der Bevölkerung einer definierten Region dar und die Einweisungsstrukturen in eine Klinik sind zudem typischerweise nicht bekannt. In der Regel wird deshalb von der vereinfachenden Annahme ausgegangen, dass sich eine fiktive Population definieren lässt, die den Einzugsbereich der für die Studie ausgewählten Klinik ausmacht. Sofern dieser Einzugsbereich für die Patienten mit der interessierenden Erkrankung (die Fälle) identisch mit dem der übrigen Patienten ist, sind damit die übrigen Patienten grundsätzlich als Kontrollpersonen geeignet.

Selbst wenn man von der unrealistischen Annahme ausgehen könnte, dass die Einzugspopulation einer einzelnen Klinik repräsentativ für die Allgemeinbevölkerung ist, tritt ein zusätzliches Problem auf. Man muss nämlich davon ausgehen, dass das Patientenkollektiv einer Klinik die Population, die unter Risiko steht, die Zielerkrankung zu bekommen, nur verzerrt widerspiegeln kann.

Beispiel: Bei der Untersuchung des Risikofaktors "Rauchen" auf die Entstehung von Lungenkrebs können Patienten einer Klinik, die an Bronchialkrebs leiden, die Fallgruppe bilden. Um die Vorteile der Krankenhaus-erhebung (z.B. leichter Informationszugang, gute Kooperationsbereitschaft der Patienten) auch für die Kontrollgruppe zu nutzen, könnten die Kontrollen aus anderen Stationen, wie z.B. der "Unfallchirurgie", rekrutiert werden.

Da Rauchen aber nicht nur Lungenkrebs, sondern auch andere Erkrankungen wie Arteriosklerose oder Koronarerkrankungen fördert, kann die Wahl der "falschen" Krankenhausstation zu einer systematischen Verzerrung der Ergebnisse führen. Dies ist im Beispiel dann gegeben, falls Kontrollen aus einer anderen Station kommen, in der Raucher überrepräsentiert sind, weil diese infolge ihres Rauchverhaltens dort behandelt werden müssen (z.B. aus der inneren Abteilung, in der u.a. Herz-Kreislauf-Erkrankungen und andere durch Rauchen bedingte Krebserkrankungen häufig sind).

Das nahe liegende Prinzip, dass man Krankenhausfällen auch **Krankenhauskontrollen** gegenüberstellt, ist deshalb anfällig für Verzerrungen. Hat eine Erkrankung den gleichen Risikofaktor wie die Zielerkrankung, wird unter diesen Patienten die Prävalenz des Risikofaktors größer sein als bei den übrigen Personen, die unter Risiko stehen, die Zielerkrankung zu erleiden. Ein Vergleich dieser Patienten mit den Fällen führt dann zu einer *Unterschätzung* des mit diesem Risikofaktor verbundenen *Erkrankungsrisikos*.

Als Kontrollpatienten kommen also nur Personen in Betracht, die nicht wegen einer Erkrankung stationär behandelt wurden, für die die gleichen Risikofaktoren ursächlich verantwortlich sind wie für die Zielerkrankung. Hier besteht natürlich das Problem, dass die entsprechenden Ausschlusskriterien nur für bereits bekannte Zusammenhänge zwischen Risikofaktoren bzw. der interessierenden Exposition und Erkrankungen definierbar sind. Um dieses Problem zu minimieren, wird bei krankenhausbasierten Fall-Kontroll-Studien häufig eine Mischung verschiedener – nicht mit bekannten Risikofaktoren der Zielerkrankung assoziiert –

ten – Erkrankungen in der Kontrollgruppe angestrebt. Selbst wenn, wie oben geschildert, die Kontrollgruppe von einer Krankenhausstation rekrutiert wird, auf der keine Patienten mit expositionsassoziierten Erkrankungen liegen, so kann dennoch die Vergleichbarkeit gestört sein. So ist es z.B. denkbar, dass das Einzugsgebiet der Klinik für Fälle und Kontrollen unterschiedlich ist.

Da Schlussfolgerungen auf einem Vergleich der Verteilung der interessierenden Exposition in Fall- und Kontrollgruppe beruhen, ist es aber denkbar, **mehrere Kontrollgruppen** zu bilden, um die Konsistenz der zu beobachtenden Assoziation beurteilen zu können.

Die Vorstellung, dass Fälle und Kontrollen in Bezug auf sämtliche Faktoren, ausgenommen die Exposition, absolut identisch sein müssen, stellt eine Übertragung "prospektiver" Denkweisen auf die Situation der Fall-Kontroll-Studie dar und ist in der Praxis nur selten gültig. Entscheidend ist, dass Fälle und Kontrollen für die Faktoren vergleichbar sind, die mit der Krankheit und der Exposition assoziiert sind. Für Faktoren, die nur mit der Exposition, nicht aber mit der Krankheit assoziiert sind, ist dies nicht zwingend erforderlich.

Das oben beschriebene allgemeine Prinzip, demzufolge die Restriktion auf inzidente Fälle möglichen Selektionseffekten vorbeugt, gilt natürlich auch für krankenhausbasierte Fall-Kontroll-Studien. Bei diesem Studientyp ist eine analoge Überlegung allerdings auch für Kontrollpatienten anzustellen: Würde man alle Patienten als Kontrollen zulassen, die an einem Stichtag in der Klinik angetroffen werden (und die nach sorgfältiger Restriktion der als auswählbar klassifizierten Erkrankungen überhaupt als Kontrollen für die Zielerkrankung in Frage kommen), so wären in der entsprechenden Stichprobe zwangsläufig die Patienten überrepräsentiert, deren Zustand einen längeren Krankenhausaufenthalt zur Folge hatte. In diesem Fall wäre also ein Selektionseffekt zu befürchten, der nur vermieden werden kann, wenn die Auswahlwahrscheinlichkeit eines Kontrollpatienten unabhängig von seiner Aufenthaltsdauer in der Klinik ist. In der Praxis wäre dies z.B. dadurch erreichbar, dass man zu jedem inzidenten Fall den ersten Patienten als Kontrolle zieht, der mit einer der Auswahl Diagnosen stationär aufgenommen wird.

Auswertungsprinzipien

Im Gegensatz zu einer Querschnittsstudie, in der prävalente Fälle an einem Zeitpunkt beobachtet werden, werden bei einer Fall-Kontroll-Studie von einem festen Zeitpunkt retrospektiv Beobachtungen durchgeführt. Will man im Sinne einer Vierfeldertafel Kranke und Gesunde in Exponierte und Nichtexponierte klassifizieren, so ist zunächst genauer zu definieren, in welchem zeitlichen Bezug die Aufteilung in den Krankheitsstatus erfolgt.

Aufgrund der eingangs diskutierten Überlegungen zur Vermeidung von Selektionseffekten ist es sinnvoll, bei Fall-Kontroll-Studien ein Inzidenzkonzept zugrunde zu legen. Dabei werden nur diejenigen Fälle in die Studie eingeschlossen, die innerhalb eines genau definierten Zeitraums – dem Rekrutierungszeitraum – neu erkrankt sind. Als Kontrollen kommen nur die Individuen in Betracht, die innerhalb dieses Zeitraums der Zielpopulation angehörten und dabei unter Risiko standen, die Zielerkrankung zu bekommen. Formal lässt sich dies wie folgt beschreiben: Es sei für die Zielpopulation vorausgesetzt, dass

- jedes Individuum über ein Rekrutierungsintervall $\Delta = (t_0, t_\delta)$ vollständig beobachtbar ist und zum Zeitpunkt t_δ entweder als Fall ($Kr = 1$) oder als Kontrolle ($Kr = 0$) erkennbar ist und
- jedes Individuum eindeutig als exponiert ($Ex = 1$) oder nicht exponiert ($Ex = 0$) klassifizierbar ist.

Sind am Ende der Periode Δ insgesamt n_1 Fälle und n_0 Kontrollen rekrutiert, so kann nach Feststellung des Expositionsstatus die Vierfeldertafel aus Tab. 3.4 beobachtet werden.

Tab. 3.4: Beobachtete Anzahl von Fällen, Kontrollen, Exponierten und Nichtexponierten in einer Fall-Kontroll-Studie

Status	exponiert ($Ex = 1$)	nicht exponiert ($Ex = 0$)	Summe
Fälle ($Kr = 1$)	n_{11}	n_{10}	$n_{1.} = n_{11} + n_{10}$
Kontrollen ($Kr = 0$)	n_{01}	n_{00}	$n_{0.} = n_{01} + n_{00}$
Summe	$n_{.1} = n_{11} + n_{01}$	$n_{.0} = n_{10} + n_{00}$	$n = n_{..}$

In Tab. 3.4 bezeichnet n_{ij} allgemein die beobachtete Anzahl von Individuen der Studie im Krankheitsstatus i ($i = 0, 1$) aus der Expositionsgruppe j ($j = 0, 1$) am Ende der Rekrutierungsperiode Δ . Auch die Struktur dieser Tafel ist zu der in Tab. 3.1 ähnlich. Allerdings sind die entsprechenden Wahrscheinlichkeiten hier nicht die Risiken wie aus Tab. 3.1, d.h. z.B. $\Pr(\text{krank}|\text{exponiert})$, sondern gerade die Wahrscheinlichkeiten für das Vorliegen einer Exposition, wenn das Individuum zudem ein Fall oder eine Kontrolle ist, also z.B. $\Pr(\text{exponiert}|\text{Fall})$. Charakteristisch ist nämlich für Fall-Kontroll-Studien, dass die Anzahl der Fälle und Kontrollen, also die *Zeilensummen* $n_{1.} = n_{11} + n_{10}$ bzw. $n_{0.} = n_{01} + n_{00}$, *vorgegeben* ist. Durch die Erhebung des Expositionsstatus bei den Fällen und Kontrollen ergibt sich deren Aufteilung auf die Exponierten und die Nichtexponierten. Insgesamt bedeutet dies, dass es mit dieser Tafel *nicht möglich* ist, einen sinnvollen *Schätzer für ein Risiko* anzugeben, so dass insbesondere auch das *relative Risiko* hieraus *nicht geschätzt* werden kann.

Aus diesem Grund wurde ursprünglich in Fall-Kontroll-Studien nur überprüft, ob ein Unterschied zwischen dem Anteil der exponierten Fälle und dem Anteil der exponierten Kontrollen besteht. Dabei wurde keinerlei Versuch unternommen, eine Risikoerhöhung durch die Exposition zu quantifizieren.

Da das Odds Ratio OR aber, wie in Abschnitt 2.3.3 erläutert, auch als Verhältnis der Expositionschancen interpretiert werden kann, ist es möglich, aus einem Fall-Kontroll-Design eine Schätzgröße hierfür zu ermitteln und es gilt

$$\text{Schätzer Odds Ratio} = \text{OR}_{\text{Studie}} = \frac{n_{11}/n_{10}}{n_{01}/n_{00}} = \frac{n_{11} \cdot n_{00}}{n_{10} \cdot n_{01}}.$$

Damit ist auch aus einer Fall-Kontroll-Studie eine vergleichende epidemiologische Maßzahl schätzbar, die für seltene Erkrankungen dem relativen Risiko sehr nahe kommt (siehe Abschnitt 2.3.3).

Beispiel: BSE wurde erstmals 1986 in Großbritannien diagnostiziert, so dass Wilesmith et al. (1988) erste epidemiologische Studien durchführten, um Risikofaktoren zu identifizieren. Dabei fanden sie heraus, dass alle bis dahin erfassten Rinder mit BSE (Fälle) mit kommerziellem Kraftfutter gefüttert worden waren. Zur weiterführenden Betrachtung der Risikofaktoren wurde in der Fall-Kontroll-Studie im Bereich der Kälberfütterung von Wilesmith et al. (1992) insbesondere der Einsatz von Fleisch-Knochen-Mehl in zugekauften Futtermitteln betrachtet. Nach Dahms (1997) konnten in dieser Untersuchung die Daten von 264 Herden gemäß Tab. 3.5 ausgewertet werden.

Tab. 3.5: BSE-Fälle und Kontrollen (Anzahl Herden) und Fütterung mit Fleisch-Knochen-Mehlen in der Fall-Kontroll-Studie von Wilesmith (1992) (Quelle: Dahms 1997)

Status	Fütterung mit Fleisch-Knochen-Mehl		Summe
	ja	nein	
Fälle (BSE in der Herde)	56	112	168
Kontrollen (kein BSE in der Herde)	7	89	96
Summe	63	201	264

Insgesamt wurden 168 BSE-Herden als Fälle und 96 Kontrollherden untersucht. Hier ist es somit nicht möglich, eine Inzidenz oder Prävalenz von BSE zu ermitteln, denn die Gesamtzahl der Fälle und Kontrollen ist vorgegeben und eine Berechnung eines Anteils Erkrankter in der Studienpopulation macht somit keinen Sinn.

Aus den Informationen aus Tab. 3.5 kann aber das Odds Ratio der Studie ermittelt werden. Man erhält

$$\text{OR}_{\text{Dahms}} = \frac{56 \cdot 89}{7 \cdot 112} = 6,36,$$

was auf eine erhebliche Risikoerhöhung durch entsprechende Fütterung hinweist. Diese und andere Indizien führten seinerzeit daher zur Anordnung eines entsprechenden Fütterungsverbots.

Insbesondere für nicht seltene Erkrankungen sind weitere Auswertungskonzepte von Fall-Kontroll-Studien dokumentiert, um dem Mangel zu begegnen, dass ein Risiko nicht schätzbar ist. Da dies über den Rahmen dieser Darstellung hinausgehen würde, sei z.B. auf Breslow & Day (1980), Miettinen (1985) und Rothman et al. (2008) verwiesen.

Vorteile von Fall-Kontroll-Studien

Obwohl eine direkte Quantifizierung von Risiken in Fall-Kontroll-Studien nicht möglich ist, ist dieses Design wegen seiner Vorteile weit verbreitet. Diese Vorteile, insbesondere im Gegensatz zum Prinzip der Kohortenstudie (siehe Abschnitt 3.3.4), stellen sich zusammengefasst wie folgt dar:

Prinzipiell kann dem Fall-Kontroll-Design eine (relativ) *kurze Studiendauer* unterstellt werden. Das Studiendesign ermöglicht die *Untersuchung seltener Krankheiten*, da die Erkrankten gezielt über Spezialkliniken oder Register identifiziert werden können. Gleichzeitig eignet es sich besonders für die Untersuchung von *Krankheiten mit einer langen Induktionszeit*, denn diese muss in Fall-Kontroll-Studien nicht abgewartet werden, und zwar im Gegensatz zu einer Studie, die das neue Auftreten eines Krankheitsereignisses aktiv beobachten muss.

Weiterhin ist das Fall-Kontroll-Design besonders für die (simultane) *Untersuchung mehrerer Risikofaktoren* für die ausgewählte Krankheit geeignet, denn die Exposition wird im Nachhinein festgestellt und so können verschiedene Expositionen bestimmt werden. Dies ist fast immer dann unabdingbar, wenn ein vollständig neues Krankheitsphänomen auftritt (siehe z.B. die Ausführungen zu BSE) oder wenn die Ursachen für ein konkretes Ausbruchsgeschehen zu untersuchen sind (z.B. bei Ausbrüchen von Infektionserkrankungen mit unbekannter Ursache).

Den genannten Vorteilen steht allerdings auch eine Reihe von Nachteilen gegenüber.

Nachteile von Fall-Kontroll-Studien

Neben den mangelnden Quantifizierungsmöglichkeiten von Inzidenzen haben Fall-Kontroll-Studien den Nachteil, dass es zwar theoretisch möglich ist zu bestimmen, ob der Einflussfaktor der Krankheit vorausgeht, jedoch ist die zuverlässige Ermittlung zeitlich zurückliegender Expositionen fehleranfällig, da sie in der Regel auf Selbstangaben der Studienteilnehmer basiert. Die zweite Hauptschwierigkeit besteht in der adäquaten Auswahl einer geeigneten Vergleichsgruppe, den Kontrollen, und darin, die Datenerhebung bei Fällen und Kontrollen gleichermaßen zu standardisieren. Diesen Schwierigkeiten kann nur durch ein hohes Maß an sorgfältiger Planung, Durchführung und Auswertung einer Fall-Kontroll-Studie begegnet werden.

Ein zentrales Problem bei Fall-Kontroll-Studien ist die *retrospektive Expositionsbestimmung*. Die Exposition durch den Risikofaktor kann dem Krankheitsausbruch um viele Jahre vorausgegangen sein. In solchen Fällen muss die Exposition mittels Gedächtnis, Aussagen Dritter oder Aufzeichnungen rekonstruiert werden. Die Verlässlichkeit solcher Eigenangaben muss stets äußerst kritisch betrachtet werden, insbesondere dann, wenn den Studienteilnehmern die Assoziation von Risikofaktor und Krankheit bekannt ist. Kumulative Expositionsmessungen (lebenslange Expositionsbelastung) sind für manche Variablen schwierig und nur ungenau nachvollziehbar. Man denke hierbei z.B. an Passivrauchexpositionen oder Belastungen mit Noxen am Arbeitsplatz, die in der Regel nur sehr schwer berichtet werden können.

Wie eingangs erläutert, ist zentrale Voraussetzung der ätiologischen Interpretation einer Fall-Kontroll-Studie die Vergleichbarkeit von Fall- und Kontrollgruppe. Selbst wenn es gelungen ist, vollständig vergleichbare Fall- und Kontrollgruppen aus derselben Zielbevölkerung zu rekrutieren, kann ein *Beobachtungsunterschied* den Vergleich wieder zunichte machen. Ein solcher Unterschied kann auf zwei Weisen entstehen. Zum einen kann sich die Erwartungshaltung des Untersuchers, dass z.B. eine betrachtete Exposition für die Zielerkrankung ursächlich ist, unbewusst darauf auswirken, wie er Fragen formuliert oder non-

verbal unterstützt. Wenn dadurch dann z.B. Fälle im Vergleich zu Kontrollen suggestiv befragt werden, so ist zu befürchten, dass sich das Antwortverhalten beider Gruppen allein aufgrund der unterschiedlichen Befragung unterscheiden wird. Ein extremes Szenario, das fast zwingend einen Untersucherbias nach sich zieht, wäre eine Studie, in der die Fälle von anderen Personen (z.B. ihren behandelnden Ärzten) befragt werden als die Kontrollen (z.B. von geschulten Interviewern). Zum anderen muss aber auch befürchtet werden, dass die Erkrankung selbst einen Einfluss auf das Antwortverhalten der Studienteilnehmer hat.

Beispiel: Es wird oft unterstellt, dass Fälle, vor allem, wenn sie an einer schweren Erkrankung wie Krebs leiden, nach Erklärungen für ihre Krankheit suchen und infolgedessen bestimmte Expositionen häufiger erinnern als Kontrollpersonen. Aber auch ein umgekehrter Effekt ist denkbar, bei dem z.B. gerade Krebspatienten, die ihre Erkrankung verdrängen, weniger detaillierte Angaben zu ihren Expositionen machen als Kontrollpersonen.

In Experimenten können solche Einflüsse durch Doppel-Blind-Versuche vermieden werden. Blindbefragungen sind in der Epidemiologie aber nur selten möglich. Somit kann nur durch ein hohes Maß an Standardisierung, etwa durch standardisierte Befragungen, zentrale Schulung der Mitarbeiter etc., eine Beobachtungsgleichheit erreicht werden. Übrigens liegt hier ein weiterer Vorteil krankenhausbasierter Studien, da sowohl das Setting (Krankenhaus) als auch die Tatsache, dass beide Studiengruppen aus Patienten bestehen, die Vergleichbarkeit der Beobachtungssituation erhöhen.

Bei der Auswahl von Fällen und Kontrollen und der Schätzung der Exposition ist die retrospektive Studie somit sehr anfällig für eine Reihe unterschiedlichster Verzerrungsmechanismen, die wir im Detail in Abschnitt 4.6 behandeln werden.

3.3.4 Kohortenstudien

Der Studientyp der **Kohortenstudie**, auch **Longitudinal-**, **Follow-up-** oder einfach **prospektive Studie**, wird im Wesentlichen dadurch charakterisiert, dass man eine (Studien-) Population über eine vorgegebene **Beobachtungs-** oder **Follow-up-Periode** Δ von einem Zeitpunkt t_0 bis zu einem Zeitpunkt t_1 hinsichtlich des Eintretens interessierender Ereignisse, wie Erkrankungen oder Todesfälle, beobachtet.

Grundsätzlich gilt somit für Kohortenstudien, dass ihre Mitglieder zu Beginn des Beobachtungszeitraums unter Risiko stehen müssen, die Zielerkrankung(en) zu bekommen. Individuen, die bereits daran erkrankt sind, sind von der Kohorte auszuschließen. Kohortenstudien beginnen mit der Exposition und beobachten dann das Auftreten von (unterschiedlichen) Erkrankungen bzw. Sterbefällen.

Im einfachen Fall ist für jedes Individuum der Studienpopulation der Expositionsstatus zu Beginn der Follow-up-Periode Δ bekannt und Individuen können im zeitlichen Verlauf den Expositionsstatus nicht ändern. Bei einer dichotomen Exposition unterteilt man die Studienpopulation dann in eine exponierte Risikogruppe und eine nicht exponierte Vergleichsgruppe.

Am Ende der Beobachtungsperiode vergleicht man beide Gruppen hinsichtlich des Auftretens der Krankheit im Beobachtungszeitraum Δ . Zeigt die Risikogruppe eine höhere Inzidenz (gemessen als Inzidenzrate oder Risiko) als die Vergleichsgruppe, so wird der Unterschied auf das Wirken des Risikofaktors zurückgeführt, sofern beide Gruppen vergleichbar sind. Somit hat eine Kohortenstudie Ähnlichkeit mit einer experimentellen Studie mit dem Unterschied, dass die zu vergleichenden Gruppen nicht randomisiert sind. Durch den prospektiven Charakter von Kohortenstudien ist sichergestellt, dass die Exposition der Erkrankung vorausgeht und dass die Expositionsänderung unabhängig von der Erkrankung erfolgt (siehe Abb. 3.11). Aus diesem Grund wird diesem Studientyp ein hoher Evidenzgrad in Bezug auf die Bewertung kausaler Zusammenhänge beigemessen.

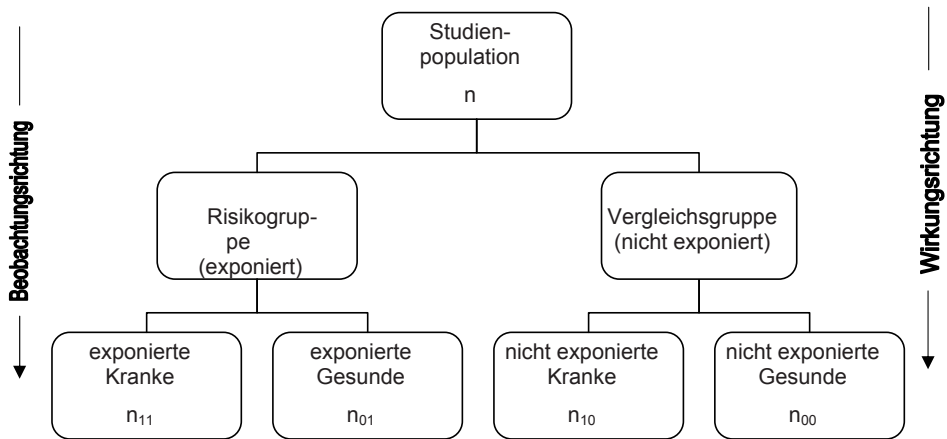


Abb. 3.11: Struktur einer Kohortenstudie und Aufteilung der Studienteilnehmer

Beispiel: Ein Beispiel für ein solches Studiendesign ist die Bitumen-Kohortenstudie (BICOS) bei männlichen Asphaltarbeitern (siehe Abschnitt 1.2). Zur Frage, ob die Sterblichkeit an Lungenkrebs und anderen Erkrankungen bei Arbeitern, die gegenüber Bitumendämpfen bei Asphaltarbeiten exponiert waren, erhöht ist, wurde (in verschiedenen europäischen Ländern) eine Kohortenstudie vom Anfangszeitpunkt 1975 bis zum Endzeitpunkt 2005 durchgeführt. Die Mitglieder der Kohorte wurden in über 100 Straßenbau- und Asphaltmischbetrieben in Deutschland rekrutiert. Dabei wurden alle ehemaligen und bis zum Studienbeginn aktiven Betriebsmitarbeiter unter Berücksichtigung bestimmter Einschlusskriterien erfasst und anhand betrieblicher Unterlagen in Bitumenexponierte und Nichtbitumenexponierte aufgeteilt. Die deutsche Kohorte umfasste mehr als 10.000 Männer.

Innerhalb dieses allgemeinen Studiendesigns werden nun verschiedene Typen von Kohortenstudien unterschieden.

Kohortenstudien mit internen und externen Vergleichsgruppen

Zur genaueren Beschreibung der Kohortenstudie lässt sich diese zunächst unterteilen in eine Follow-up-Studie mit einer internen Vergleichsgruppe bzw. in eine Studie mit einer externen Vergleichsgruppe.

Bei einer **Follow-up-Studie mit einer internen Vergleichsgruppe** wird die gesamte Studienpopulation in unterschiedlich stark exponierte Teilkollektive zerlegt. Bei einer **Follow-up-Studie mit einer externen Vergleichsgruppe** stellt die Studienpopulation dagegen eine exponierte Gruppe dar, die mit der Krankheitshäufigkeit einer externen Population verglichen wird.

Beispiel: Bei der Bitumen-Kohortenstudie (BICOS) wurden die Mitarbeiter von 100 Straßenbau- und Asphaltmischbetrieben als Kohorte definiert. Da jeder einzelne Mitarbeiter einen anderen Arbeitsplatz hat bzw. andere Arbeitstätigkeiten durchführt, können diese in Arbeiter mit (erheblicher) Bitumenexposition und ohne Bitumenexposition unterschieden werden. Damit ist BICOS ein klassisches Beispiel für eine Follow-up-Studie mit interner Vergleichsgruppe.

Darüber hinaus wurde die Sterblichkeit innerhalb der Kohorte auch mit der Sterblichkeit der deutschen Gesamtbevölkerung verglichen. Hierzu wurden die Mortalitätsdaten des Statistischen Bundesamts genutzt (siehe Abschnitt 3.2.3). In diesem Fall diente also die deutsche Gesamtbevölkerung als externe Vergleichsgruppe.

Geschlossene und offene Kohorten

Ein anderes, mit Blick auf die zu verwendenden Häufigkeitsmaße wichtiges Unterscheidungsmerkmal von Kohortenstudien ergibt sich dadurch, ob Individuen zu einem Zeitpunkt nach Follow-up-Beginn noch der Studienpopulation beitreten können oder nicht.

Beispiel: Eine denkbare Definitionsmöglichkeit einer Kohorte wäre, alle an einem festen Zeitpunkt t_0 (z.B. $t_0 = 1.$ Januar 1975) auf exponierten Arbeitsplätzen beschäftigten Personen über einen vorgegebenen Zeitraum Δ (z.B. $\Delta = 30$ Jahre) zu beobachten. Da im Verlauf der Follow-up-Periode Δ neue Beschäftigte eingestellt werden bzw. Arbeiter den Arbeitsplatz auch verlassen können, muss festgelegt werden, wie mit diesen Individuen umzugehen ist. Dabei ist auch zu überlegen, ob Arbeiter, die vor dem Stichtag bereits auf exponierten Arbeitsplätzen tätig waren, sogar ausgeschlossen werden sollten.

Dies führt zu der Unterscheidungsmöglichkeit für Follow-up-Studien in Studien mit **geschlossenen (festen)** und mit **offenen (dynamischen) Kohorten**. Im ersten Fall wird eine zu Beginn der Beobachtungsperiode definierte Gruppe von Individuen als feste Kohorte bestimmt und über die Follow-up-Periode beobachtet. Hier werden nach Beobachtungsbeginn keine weiteren Mitglieder in die Kohorte aufgenommen. Bei offenen Kohorten ist dagegen der Eintritt in die Kohorte auch nach Beginn der Beobachtungsperiode möglich.

Bei Kohorten von beruflich exponierten Personen besteht das Problem, dass für Personen, die bereits vor dem Beobachtungsbeginn unter Exposition standen, die Expositionsdauer unbekannt ist. Sofern die Exposition die Mortalität oder die Dauer der Betriebszugehörigkeit beeinflusst, ist sogar zu befürchten, dass die Arbeiter, die zum Stichtag noch an der entsprechenden Stelle beschäftigt sind, eine selektierte Gruppe darstellen. Dieses Problem umgeht man bei so genannten **Inzeptionskohorten**. Dabei werden Personen erst ab Beginn ihrer Exposition in die Kohorte eingeschlossen. In dem Beispiel der beruflichen Exposition bedeutet dies, dass nur Arbeiter eingeschlossen werden, die die interessierende Arbeit nach dem Stichtag (Beginn der Beobachtungsperiode) begonnen haben. Dieser Stichtag kann so festgelegt werden, dass eine Mindestdauer einer Betriebszugehörigkeit oder Exposition erforderlich ist, um überhaupt in die Kohorte aufgenommen zu werden. Bei der Berechnung

der Zeit unter Risiko muss man dann beachten, dass Zeiten, die ein Individuum leben musste, um überhaupt in die Kohorte eingeschlossen zu werden, nicht in die Berechnung der Zeit unter Risiko eingehen dürfen. Diese nicht zu berücksichtigende Zeit nennt man **"immortal Person-Time"**.

Beispiel: Falls eine Kohortenstudie durchgeführt werden soll, um den Einfluss von Expositionen am Arbeitsplatz in einer bestimmten Fabrik auf die Entstehung von Lungenkrebs zu untersuchen, so könnte man eine Kohorte von Arbeitern bilden, die mindestens fünf Jahre in dieser Fabrik beschäftigt waren. Ereignisse, die in den ersten fünf Jahren der Beschäftigungszeit aufgetreten sind, dürfen dann nicht gezählt werden, da diese Jahre allein zur Definition der Kohorte dienen und somit ein Kohortenmitglied mindestens diese fünf Jahre gelebt haben muss. Dementsprechend sind diese fünf Jahre auch bei der Berechnung der Zeit unter Risiko auszuschließen.

Diese unterschiedlichen Aspekte bei der Ermittlung der Risikozeiten in dynamischen Kohorten werden ausführlich in Breslow & Day (1987) behandelt.

Prospektive und historisch-prospektive Kohortenstudien

Für das Studiendesign der Follow-up-Studie wurde u.a. auch die Bezeichnung *prospektive Studie* benutzt. Dabei ist es sogar möglich, prospektive Studien vollständig in der Vergangenheit durchzuführen, d.h., zum Zeitpunkt der Studiendurchführung liegt der Zeitraum, auf den sich die Follow-up-Periode bezieht, bereits in der Vergangenheit.

Um dieser Tatsache gerecht zu werden, wird gewöhnlich unterschieden zwischen den eigentlichen **prospektiven Follow-up-Studien** (engl. **concurrent Follow-Up-Study**), bei denen die Datenerhebung in der Gegenwart beginnt und den **historischen Follow-up-Studien** (auch **Kohortenstudie mit zurückverlegtem Ausgangspunkt**).

Historische Kohortenstudien können dann durchgeführt werden, wenn Daten aus der Vergangenheit vorhanden sind, die es gestatten, ab einem bestimmten Zeitpunkt eine Kohorte zu rekonstruieren, in der einige Personen die Exposition aufweisen und einige nicht. Beispiele für diesen Studientyp findet man vor allem im Bereich der Arbeitswelt (z.B. auch bei BICOS), in der man zur Ermittlung von Kohortenmitgliedern und ihren Expositionen oft auf betriebliche Unterlagen zurückgreifen kann.

Vielzweckkohorten

Im klassischen Fall einer Kohortenstudie werden die gesundheitlichen Folgen einer einzelnen Exposition untersucht. Als Beispiel hierfür kann die Asphaltarbeiterkohorte (BICOS) gelten, aber auch Kohortenstudien, in denen Umweltexpositionen wie z.B. durch die industrielle Dioxinfreisetzung in Seveso oder Erkrankungsrisiken der Wohnbevölkerung in der Nähe von Atomkraftwerken untersucht wurden. Oftmals wird für diese Kohorten, die nur einem einzelnen spezifischen Zweck dienen, das Design einer historisch-prospektiven Kohorte gewählt, um in vertretbar kurzer Zeit eine belastbare Aussage zu möglichen Krankheitsrisiken treffen zu können.

Wird eine Kohortenstudie jedoch prospektiv geplant und durchgeführt, so wird sie oft breit angelegt, um den relativ großen Studienaufwand in ein günstiges Verhältnis zum wissenschaftlichen Nutzen zu bringen. Solche Kohorten nennt man **Vielzweckkohorten** (engl. **multi-purpose Cohorts**), weil nicht einzelne, sondern verschiedene Einflussfaktoren parallel untersucht werden. Beispiele hierfür sind die berühmte Framingham-Studie (FHS 2011), die Nurses-Health Studie (NHS 2011) und die EPIC-Studie (siehe Abschnitt 1.2).

Typischerweise beginnt eine Studie mit einer umfangreichen Basis-Untersuchung aller Kohortenmitglieder, der dann in Abständen weitere Untersuchungen folgen. Dabei können biologische Proben (Blut, Urin, Speichel) zur Bestimmung von Expositions- und Krankheitsmarkern mit medizinischen Untersuchungen und der Befragung zu Lebensstilfaktoren kombiniert werden. Durch die systematische Einbeziehung von bekannten Risikofaktoren für die häufig auftretenden Erkrankungen, die im Rahmen der jeweiligen Studie untersucht werden können, bietet dieses Design den großen Vorteil, dass eine effektive Kontrolle von Störgrößen (Confoundern) möglich ist.

Eingebettete Fall-Kontroll-Studien

Beobachtet man im Verlaufe des Follow-Up innerhalb einer Kohortenstudie Erkrankungsfälle, so können diese wiederum zur Grundlage einer eigenständigen Studie werden. Man spricht von einer **eingebetteten Fall-Kontroll-Studie**, wenn diese auf einer realen Kohorte aufbaut, deren einzelne Mitglieder zuvor in die Kohorte eingeschlossen und über eine definierte Zeitperiode beobachtet wurden. Alle Individuen aus dieser Kohorte, die während des Beobachtungszeitraums die Zielerkrankung bekommen haben, sind der Fallgruppe zuzuordnen und alle übrigen Personen kommen als Kontrollen in Betracht, solange sie unter Risiko gestanden haben.

Diese Vorgehensweise kann zu einer tiefer gehenden Betrachtung der ätiologischen Zusammenhänge führen.

Beispiel: Der Vergleich der Sterblichkeit unter Asphaltarbeitern mit der Sterblichkeit der Allgemeinbevölkerung ergab eine ungefähr doppelt so hohe Lungenkrebsmortalität der gegenüber Bitumen exponierten Arbeitern als erwartet (vgl. Behrens et al. 2009). Da die Berechnung dieses Unterschiedes auf der indirekten Standardisierung beruhte, scheiden mögliche Altersunterschiede als Erklärung aus. Auch Periodeneffekte können diesen Unterschied nicht erklären, da bei der Berechnung des SMR die über die Jahrzehnte der Beobachtungszeit sich verringernde Mortalität in Deutschland berücksichtigt wurde.

Dennoch bestanden Zweifel daran, ob die beobachtete Risikoerhöhung ursächlich auf Bitumenexpositionen zurückgeführt werden kann. So konnte der interne Vergleich mit nicht gegenüber Bitumen exponierten Arbeitern das erhöhte Risiko nicht bestätigen, was aber z.T. auch daran liegen könnte, dass die Arbeiter der Vergleichskohorte gegenüber potenziellen Karzinogenen der Lunge wie Dieselruß, Quarzstaub oder Steinkohlenteer exponiert waren. Weitere Zweifel wurden durch die Annahme genährt, dass Asphaltarbeiter häufiger rauchen als die Allgemeinbevölkerung. Ein erhöhtes Lungenkrebsrisiko in dieser Berufsgruppe könnte also durch diesen Confounder bedingt sein. Als weitere Komplikation war in Betracht zu ziehen, dass ein großer Teil der Arbeiter bereits vor Eintritt in die Kohorte berufstätig und dort möglicherweise karzinogenen Arbeitsstoffen ausgesetzt war. Diese Expositionen waren im Rahmen der Bitumen-Kohortenstudie nicht beobachtbar, da sie vor der festgelegten Beobachtungszeit der Kohorte lagen. Das besondere Problem dabei war, dass in früheren Jahren sehr häufig Steinkohlenteer in den Asphaltmischungen eingesetzt wurde. Dieser war aber in der Beobachtungszeit ganz überwiegend durch Bitumen als Bindemittel ersetzt worden. Steinkohlenteer ist ein bekannt-

tes starkes Karzinogen der Lunge. Es war daher zu befürchten, dass gerade die Arbeiter, die während der Beobachtungszeit gegenüber Bitumen exponiert waren, an früheren Arbeitsplätzen gegenüber Steinkohlenteer exponiert waren und damit ein massives Confounding durch diese früheren Expositionen bedingt sein könnte.

Damit war klar, dass die bestehenden Zweifel nur durch eine detaillierte Erhebung zusätzlicher Informationen ausgeräumt werden konnten. Die benötigten Daten nachträglich bei allen 10.000 Kohortenmitgliedern zu ermitteln wäre nicht nur logistisch kaum zu bewältigen, es ist auch nicht notwendig. Stattdessen wurde eine eingebettete Fall-Kontroll-Studie mit allen aufgetretenen Lungenkrebsfällen und einer Zufallsauswahl der übrigen Kohortenmitglieder als Kontrollen durchgeführt (vgl. Olsson et al. 2010).

Im Rahmen einer telefonischen Befragung wurden diese Fälle und Kontrollen zu ihrem früheren Rauchverhalten und zu ihrer gesamten beruflichen Vorgeschichte interviewt. Bei inzwischen verstorbenen Personen erfolgte diese Befragung bei einem näheren Angehörigen, bevorzugt der Ehefrau. Auf diese Weise wurde die fehlende Confounderinformation zusammengetragen und gemeinsam mit den zuvor gesammelten betrieblichen Expositionsdaten analysiert, wobei die Informationen zu Intensität und Aufnahmeweg der Bitumenexposition zusätzlich noch genauer quantifiziert wurden. Diese Analyse ergab dann keinen überzeugenden Hinweis mehr auf ein durch Bitumen bedingtes Lungenkrebsrisiko. Allerdings blieb ein Verdacht, dass dermale Bitumenexposition mit einem Risiko behaftet sein könnte.

Dieses Beispiel zeigt, dass eine eingebettete Fall-Kontroll-Studie ein ökonomisches Instrument ist, zusätzlich benötigte Informationen zu gewinnen, die für die Bewertung eines beobachteten Zusammenhanges benötigt werden und die nicht mit vertretbarem Aufwand in der gesamten Kohorte zu ermitteln sind.

Dabei ist die direkte Durchführung einer Fall-Kontroll-Studie in der Regel nicht möglich, denn meist ist die untersuchte Exposition zu selten, als dass man sie im Rahmen einer populationsbasierten Studie mit ausreichender Häufigkeit beobachten könnte. Die Untersuchung seltener Expositionen und ihre Erfassung vor Eintritt der Erkrankung sind eine besondere Stärke von Kohortenstudien. Die Untersuchung mehrerer Expositionen (einschließlich Confounderinformationen) ist dagegen die Stärke von Fall-Kontroll-Studien. In der Kombination aus Kohorten- und Fall-Kontroll-Studie kommen die Stärken beider Designs zusammen. Hinzu kommt die Tatsache, dass die sonst schwierige Auswahl geeigneter Kontrollen im Rahmen einer eingebetteten Studie fast problemlos möglich ist, da die Population unter Risiko genau bekannt ist.

Es ist ein besonderer *Vorteil eingebetteter Fall-Kontroll-Studien*, dass sie prospektiv sind, d.h. vor Eintritt der Erkrankung, erhobene Expositionsdaten nutzen können. Ein häufig genutzter Anwendungsbereich hierfür sind die z.B. in Vielzweckkohorten gewonnenen und dann eingelagerten biologischen Proben. Die Analyse spezifischer biologischer Marker in diesen Proben kann sehr teuer und aufwändig sein. Ein ökonomischer Umgang mit diesen Proben ist durch eingebettete Studien möglich, bei denen die Laboranalyse auf eine kleine Gruppe interessierender Fälle und eine Zufallsauswahl an Kontrollpersonen beschränkt wird.

Dies erfolgt z.B. im Rahmen von Kohortenstudien wie EPIC oder IDEFICS (siehe Kapitel 1.2), bei der das vor Erkrankungseintritt gesammelte Material über viele Jahre tiefgefroren wird und später nach Eintritt der Erkrankung aufgetaut und analysiert wird. Würde das biologische Material erst im Rahmen der Fall-Kontroll-Studie gesammelt, ließe sich die zeitliche Abfolge zwischen Exposition und Erkrankung nicht mehr eindeutig herstellen und es bliebe die Frage, inwiefern die untersuchten Parameter durch die Erkrankung beeinflusst sein könnten.

Auswertungsprinzipien

Bei Kohortenstudien hängen die Maße zur Beschreibung der Erkrankungshäufigkeit in der Studienpopulation von der Art des Follow-up-Designs ab. Im Folgenden soll zunächst davon ausgegangen werden, dass eine feste Kohorte (= Studienpopulation) vorliegt, in der

- jedes Individuum über das Zeitintervall $\Delta = (t_0, t_\delta)$ vollständig beobachtet wird und zum Zeitpunkt t_δ entweder erkrankt ($Kr = 1$) oder aber krankheitsfrei ($Kr = 0$) ist und
- jedes Individuum zum Zeitpunkt t_0 eindeutig als exponiert ($Ex = 1$) oder nicht exponiert ($Ex = 0$) klassifizierbar ist.

Sind zu Beginn der Beobachtungsperiode t_0 insgesamt $n_{.1}$ Personen in der Expositionsgruppe und $n_{.0}$ Personen in der Vergleichsgruppe, so kann am Ende des Follow-up die Vierfeldertafel Tab. 3.6 erstellt werden.

Tab. 3.6: Beobachtete Anzahlen von Exponierten, Nichtexponierten, Gesunden und Kranken in einer Kohortenstudie

Status	Risikogruppe ($Ex = 1$)	Vergleichsgruppe ($Ex = 0$)	Summe
krank ($Kr = 1$)	n_{11}	n_{10}	$n_{1.} = n_{11} + n_{10}$
gesund ($Kr = 0$)	n_{01}	n_{00}	$n_{0.} = n_{01} + n_{00}$
Summe	$n_{.1} = n_{11} + n_{01}$	$n_{.0} = n_{10} + n_{00}$	$n = n_{..}$

In Tab. 3.6 bezeichnet n_{ij} die beobachtete Anzahl von Individuen im Krankheitsstatus i ($i = 0, 1$) aus der Expositionsgruppe j ($j = 0, 1$) am Ende der Periode Δ . Die Tafel stimmt in ihrer Struktur mit Tab. 3.1 überein. Charakteristisch ist, dass aufgrund des Kohortendesigns die *Spaltensummen* $n_{.1} = n_{11} + n_{01}$ bzw. $n_{.0} = n_{10} + n_{00}$ als Gruppen *vorgegeben* sind und sich im Verlaufe der Studie eine Aufteilung dieser Individuen auf die Erkrankten und Gesunden ergibt.

Gemäß der Definition der kumulativen Inzidenz, die die Anzahl der Neuerkrankungen in der Periode Δ auf die Gesamtzahl der unter Risiko stehenden Individuen bezieht (siehe Abschnitt 2.1.1), können die **Erkrankungsrisiken** aus Tab. 3.1 damit sinnvollerweise mit nachfolgenden Größen aus der Kohortenstudie **geschätzt** werden.

$$\text{Schätzer kumulative Inzidenz (bei Exposition } j) = CI(j)_{\text{Studie}} = \frac{n_{1j}}{n_{\cdot j}}, j = 0, 1.$$

Damit ergibt sich als **Schätzwert** für das **relative Risiko** RR aus der Kohortenstudie die Größe

$$\text{Schätzer relatives Risiko} = RR_{\text{Studie}} = \frac{CI(1)_{\text{Studie}}}{CI(0)_{\text{Studie}}} = \frac{n_{11}/(n_{11} + n_{01})}{n_{10}/(n_{10} + n_{00})},$$

bzw. als **Schätzer** für das **Odds Ratio** OR

$$\text{Schätzer Odds Ratio} = OR_{\text{Studie}} = \frac{n_{11}/n_{01}}{n_{10}/n_{00}} = \frac{n_{11} \cdot n_{00}}{n_{10} \cdot n_{01}}.$$

Analog zu diesen Effektmaßen sind auch die Risikodifferenz und die verschiedenen attributablen Risikomaße (siehe Abschnitt 2.4) schätzbar, denn die geschlossene Kohortenstudie zeichnet sich gerade dadurch aus, dass das Risiko im Sinne der kumulativen Inzidenz beschrieben wird.

Hat man allerdings eine offene Kohorte definiert, so sind die Häufigkeiten von Erkrankten und Gesunden nicht wie in Tab. 3.6 gegeben. Es ist zwar möglich, die Anzahlen von Kranken in der Risikogruppe n_{11} bzw. in der Vergleichsgruppe n_{10} anzugeben, die Anzahl der Kohortenmitglieder ist aber nicht definiert bzw. sollte nach den Ausführungen aus Abschnitt 2.1.2 durch das Konzept der Zeit unter Risiko ersetzt werden. Dann ergibt sich allerdings die Möglichkeit, aus der Studie auch je eine Inzidenzdichte für die zu vergleichenden Gruppen zu ermitteln. Man erhält somit als **Schätzer** für das **relative Risiko** bei **einer offenen Kohorte**

$$\text{Schätzer relatives Risiko} = RR_{\text{offen}} = \frac{ID(1)_{\text{Studie}}}{ID(0)_{\text{Studie}}} = \frac{n_{11}/P\Delta(1)_{\text{Studie}}}{n_{10}/P\Delta(0)_{\text{Studie}}}.$$

Hierbei stellen die Größen $P\Delta(1)_{\text{Studie}}$ und $P\Delta(0)_{\text{Studie}}$ die aus der Studie ermittelten Gesamttrisikozeiten in den beiden Expositionsgruppen dar (siehe Abschnitt 2.1.2).

Dieses Konzept der Schätzung des relativen Risikos bei offenen Kohorten kann dann auch auf die Situation einer Kohortenstudie mit externer Vergleichsgruppe erweitert werden, indem man bei der Berechnung die Inzidenzdichte der Vergleichsgruppe durch die (erwartete) Inzidenzdichte in der externen Vergleichspopulation ID_{erw} ersetzt. Dann erhält man als **Schätzer** für das **relative Risiko** bei **einer offenen Kohorte und externer Vergleichsgruppe**

$$\text{Schätzer relatives Risiko} = RR_{\text{ext}} = \frac{ID_{\text{Studie}}}{ID_{\text{erw}}}.$$

Diese Größe ist identisch mit dem im Zusammenhang mit der Standardisierung bereits eingeführten Standardisierten Mortalitätsratio SMR (siehe Abschnitt 2.2.2.2). Werden statt Sterbefällen Neuerkrankungen betrachtet, wird dieser Quotient als **Standardisiertes Inzidenz Ratio SIR** bezeichnet.

Beispiel: Im Rahmen der Bitumenkohorte wurde neben einer internen Vergleichskohorte die alters- und kalenderzeitspezifische Sterblichkeit der deutschen Wohnbevölkerung zur Berechnung des SMR herangezogen. Dabei wurde die beobachtete Lungenkrebssterblichkeit in verschiedenen, bezüglich ihrer Bitumenexposition definierten Teilkohorten mit der auf Basis der deutschen Sterblichkeitsziffern erwarteten Sterblichkeit verglichen (Tab. 3.7).

Tab. 3.7: Beobachtete und (in der Allgemeinbevölkerung) erwartete Lungenkrebssterblichkeit in expositionsspezifischen Teilkohorten in der Bitumenkohorte (Quelle: Behrens et al. 2009)

Teilkohorte	Beobachtet (n)	Erwartet (n _{erw})	RR _{erw}
Bitumenexponiert, keine Teerexposition	33	15,87	2,08
Potenziell Teerexponiert	15	8,72	1,72
Exposition unbekannt	14	10,94	1,28
Nicht exponiert	39	21,43	1,82
Gesamtkohorte	101	57,07	1,77

Diese Daten zeigen z.B. mit

$$RR_{erw} = SMR = \frac{n}{n_{erw}} = \frac{33}{15,87} = 2,08$$

eine Verdoppelung des Lungenkrebsrisikos der gegenüber Bitumen exponierten Asphaltarbeiter im Vergleich zur deutschen Wohnbevölkerung. Interessanterweise ist auch das Risiko der Nichtexponierten erhöht, so dass sich auch für die Gesamtkohorte eine erwartete Mortalität ergibt, die um 77% über der erwarteten Lungenkrebssterblichkeit liegt.

Vorteile von Kohortenstudien

Der Hauptvorteil des Kohortendesigns ist, wie obige Ausführungen zu den Auswertungsprinzipien zeigen, die direkte *Schätzbarkeit der Risiken*, in der Beobachtungsperiode zu erkranken. Damit besitzt dieses Studiendesign ein hohes Maß an Evidenz bei der Überprüfung einer ätiologischen Hypothese.

Die Tatsache, dass Kohortenstudien von der Exposition ausgehen, ist ihre besondere Stärke. So ist im Gegensatz zu anderen Studienformen für die gesamte Beobachtungsperiode eine genaue Erfassung der Exposition (zumindest theoretisch) möglich. Die mitunter zeitlich variablen Expositionsbelastungen können prinzipiell genau dokumentiert werden. Damit sind die angestrebte *Schätzung* einer (kontinuierlichen oder zumindest mehrstufig kategoriellen) *Expositions-Wirkungs-Beziehung* und die Untersuchung zeitlicher Muster sehr verlässlich.

Werden die Mitglieder der Kohorte nicht nur einmal zum Zeitpunkt t₀, sondern zu mehreren Zeitpunkten innerhalb der Periode Δ untersucht, so ist zudem auch der natürliche Verlauf einer Erkrankung in der zeitlichen Entwicklung beobachtbar. Dies ermöglicht einen Einblick

in die Pathogenese und ist besonders interessant bei der Untersuchung heterograde Krankheitsbilder.

Dadurch, dass Kohorten von der Exposition ausgehen, ergibt sich der Vorteil, dass für eine Exposition die *Wirkung auf verschiedene Erkrankungs- bzw. Todesursachen* untersucht werden kann. Damit ist eine detaillierte, umfassende Beurteilung des Gesundheitsrisikos durch eine interessierende Exposition möglich. Das gilt insbesondere für solche Faktoren, die auf Populationsebene selten sind und auf dem Niveau der Gesamtpopulation nur zu einem geringen Teil eine besonders interessierende Krankheit verursachen.

Nachteile von Kohortenstudien

Als ein Nachteil von Follow-up-Studien kann zunächst festgestellt werden, dass Kohortenstudien sich *weniger für das Studium sehr seltener Krankheiten eignen*. Wegen der geringen Inzidenzen sind entweder sehr große Teilkohorten in den Expositionsgruppen und/oder lange Beobachtungszeiten notwendig, um eine ausreichende Zahl von Krankheitsfällen in der Studie zu beobachten.

Darüber hinaus können Ausfälle von Mitgliedern der Studienpopulation (*Follow-up-Verluste*) dazu führen, dass eine systematische Verzerrung der Studienergebnisse auftritt, falls diese Ausfälle nicht zufällig erfolgen und mit dem Untersuchungsgegenstand assoziiert sind. Die potenziellen Fehlerquellen für diese Art von Bias sind sehr verschieden. Es können hierbei expositions- und krankheitsbedingte Ausfälle unterschieden werden.

Beispiel: Beispiele für solche Follow-up-Verluste findet man insbesondere in berufsepidemiologischen Studien wie etwa BICOS. So ist es vorstellbar, dass insbesondere ausländische Arbeiter an stark exponierten Arbeitsplätzen eingesetzt wurden. Wenn diese Arbeiter am Ende ihres Berufslebens in ihre Heimat zurückkehren, dann sind sie in der Regel dem Follow-up entzogen, so dass eventuell auftretende Folgeerkrankungen unbeobachtet bleiben. Darüber hinaus kann aber auch das Auftreten von ersten Symptomen einer Erkrankung den Arbeiter veranlassen, den Arbeitsplatz zu wechseln. In beiden Fällen ergibt sich eine Selektion des Kollektivs der Kohortenmitglieder, die das Ergebnis der Studie verzerren kann.

Ein wesentlicher Nachteil von prospektiv durchgeführten Kohortenstudien ist dadurch gegeben, dass sie in der Regel extrem *zeit- und kostenintensiv* sind. Dieser Nachteil wirkt besonders gravierend bei der Untersuchung von Krankheiten mit einer langen Induktionszeit, denn mit der damit verbundenen langen Beobachtungsperiode erhöhen sich u.a. Kosten und logistische Schwierigkeiten bei der Betreuung der Studienpopulation.

Ein besonderer Nachteil von historischen Kohortenstudien ist in der Regel der *Mangel an Confounderinformation auf Individualniveau*. Exemplarisch seien hier Industriekohorten wie die Asphaltarbeiterkohorte genannt, für die zwar umfangreiche Expositionsdaten vorliegen, jedoch keine Information über Lebensstilfaktoren wie Rauchen.

Ein anderer Störmechanismus kann die Einführung *neuer Diagnosetechniken* mit exakteren Messmethoden sein. Dies kann zu Problemen bei der Abgrenzung des Krankheitsbildes ins-

besondere bei heterograden Krankheitsbildern führen, da eine Klassifikation nach den anfänglichen Kriterien nicht mehr möglich ist (Problem der konstanten Messbedingungen).

Auch die interessierenden Einflussgrößen selbst können einem Wandel unterzogen sein. Wird die Studie nur für eine sehr begrenzte Anzahl von Einflussfaktoren zu Beginn der Beobachtungsperiode angelegt, so ist Evidenz bezüglich der unberücksichtigten Einflussfaktoren später kaum noch zu generieren. Bei einer extrem langen Studiendauer kann es sogar passieren, dass die Ausgangshypothese an Relevanz verliert. Ebenso ist es denkbar, dass die Dringlichkeit einer zu überprüfenden Hypothese nicht mit der für die prospektive Follow-up-Studie erforderlichen Zeitspanne zu vereinbaren ist.

3.3.5 Interventionsstudien

Studiendesigns, die zum Ziel haben, Individuen oder ganze Kollektive wie z.B. Gemeinden geplant und kontrolliert einer Behandlung oder allgemeiner einem Faktor auszusetzen, werden unter dem Begriff **Interventionsstudien** zusammengefasst. Die Voraussetzung für jede Intervention ist, dass ausreichende Evidenz dafür vorliegt, dass die Interventionsmaßnahme einen gesundheitlichen Nutzen hat, und dass dieser Nutzen einem möglichen Schaden durch die Intervention bei Weitem überwiegt. Es lassen sich im Wesentlichen drei Typen von Interventionsstudien unterscheiden: die *randomisierte kontrollierte Studie*, die *Präventionsstudie auf Individualebene* und die *Präventionsstudie auf kommunaler Ebene*.

Randomisierte kontrollierte Studien

Das Prinzip einer Interventionsstudie lässt sich am besten anhand einer **randomisierten kontrollierten Studie** veranschaulichen. Dieser Studientyp findet insbesondere in der klinischen Forschung z.B. im Bereich der Arzneimittelzulassung oder auch in Tierexperimenten seine Anwendung. In solchen **klinischen Studien** werden anhand eines festgelegten Zufallsprinzips Patienten, nachdem sie der Teilnahme mit der Unterschrift auf ihrer Einverständniserklärung zugestimmt haben, verschiedenen Behandlungen zugewiesen. Dabei kann es sich um z.B. ein neues Präparat im Vergleich zu einem Standardpräparat oder Placebo handeln. In diesem Fall spricht man von einem **zweiarmigen Versuchsaufbau**. Natürlich ist es auch denkbar, mehr als zwei Gruppen zu vergleichen, wofür jedoch andere Verfahren in der Planung und der Auswertung benötigt werden. Wir beschränken uns im Folgenden auf den Zwei-Gruppen-Vergleich.

Beispiel: Um in einer Interventionsstudie ethisch rechtfertigen zu können, Personen bewusst einer Behandlung oder einem bestimmten Faktor auszusetzen, muss zunächst geklärt sein, ob die jeweiligen Behandlungen einen gesundheitlichen Nutzen haben.

In einer klinischen Studie der Arzneimittelzulassung wird die nötige Evidenz wie folgt erbracht: In der so genannten Phase-III-Studie, der eigentlichen Zulassungsstudie, kommen nur Substanzen zum Einsatz, die sich zunächst im Labor und im Tierversuch als potenziell wirksame Substanzen herauskristallisiert haben. In der Phase I der klinischen Entwicklung wird diese Substanz an gesunden Probanden und schließlich in der Phase II an kleinen Fallzahlen von Patienten getestet, um die geeignete Dosis zu ermitteln, bevor sie abschließend in größeren Phase-III-Studien eingesetzt werden.

Die *Randomisierung* ist ein wesentliches Merkmal einer kontrollierten Studie, wobei darauf zu achten ist, dass damit keine willkürliche Auswahl der Patienten gemeint ist – dies ist aus ethischen Gründen nicht möglich –, sondern eine zufällige Zuweisung der Behandlungen an die Patienten. Die Auswahl der Patienten erfolgt dabei nach wohl definierten *Ein- bzw. Ausschlusskriterien*. Durch die Randomisierung soll eine möglichst große Vergleichbarkeit der Gruppen erreicht werden, weil man davon ausgeht, dass hierdurch sonstige Störgrößen kontrolliert werden, da sie in beiden Gruppen gleichermaßen auftreten. Weiterhin wird in klinischen Studien, sofern die Behandlung dies zulässt, noch dadurch für Vergleichbarkeit der Gruppen gesorgt, dass die Studien typischerweise *doppel-blind* durchgeführt werden. Darunter versteht man, dass weder Arzt noch Patient wissen, zu welcher Gruppe der Patient zugeordnet wurde. Damit soll verhindert werden, dass alleine durch die Erwartungshaltung – sei es des Arztes oder des Patienten – Effekte erzielt werden.

Beispiel: Wird in einer klinischen Studie zur Arzneimittelzulassung zum Beispiel ein neues Medikament zur Blutdrucksenkung gegen ein herkömmliches getestet, die beide oral verabreicht werden, lässt sich ein Doppel-Blind-Versuch realisieren.

Soll jedoch eine Studie durchgeführt werden, bei der zwar derselbe Wirkstoff in jedem Studienarm eingesetzt wird, aber dessen subkutane Verabreichung gegen die Standardtherapie getestet werden soll, bei der der Wirkstoff intravenös verabreicht wird, so ist ein Doppel-Blind-Versuch offensichtlich nicht möglich.

Ein weiteres wesentliches Merkmal ist der experimentelle Charakter, d.h., die Patienten werden bewusst und kontrolliert einer Behandlung ausgesetzt. Aufgrund dieser wesentlichen Merkmale wird diesem Design der höchste Evidenzgrad bei der Bewertung kausaler Zusammenhänge zugesprochen. Das bedeutet, dass wenn am Ende der Beobachtungsperiode ein Unterschied im Therapieerfolg bei den beiden Gruppen beobachtet wird, so lässt sich dieser auf die erfolgte Behandlung zurückführen. Das Konzept einer klinischen Studie ist in Abb. 3.12 dargestellt.

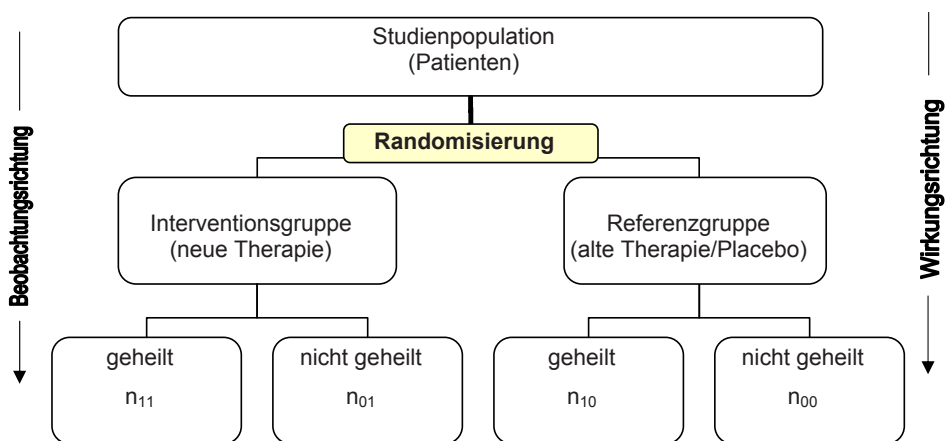


Abb. 3.12: Struktur einer klinischen Studie und Randomisierung der Studienteilnehmer

Die Durchführung randomisierter kontrollierter Studien setzt voraus, dass man diese in einer definierten Umgebung, z.B. in einer Klinik, durchführt, da so am besten eine Kontrolle der durchzuführenden Vorgehensweisen möglich ist. Grundsätzlich ist die Durchführung aber auch außerhalb eines klinischen Settings möglich. Praktisch stößt man aufgrund der logistischen Probleme aber schnell an Grenzen.

In solchen Fällen bieten sich so genannte **Cluster-randomisierte Studien** an, bei denen entsprechende Interventionsmaßnahmen organisch zusammengehörenden Clustern, wie z.B. Schulen oder Gemeinden, zufällig zugewiesen und diese Cluster dann als Ganzes in die Studie eingeschlossen werden. Die zwei Behandlungsarme ergeben sich hier dadurch, dass einige Cluster als Interventionsgruppe dienen und andere als Kontrollgruppe. Cluster-randomisierte Studien bedürfen spezieller Auswertungsverfahren, da die Korrelation innerhalb eines Clusters bei der Analyse berücksichtigt werden muss (Campbell 2012).

Präventionsstudien

Bei Interventionen außerhalb der klinischen Forschung sind sowohl Studien denkbar, die sich an bereits erkrankte Individuen richten, als auch Studien, die gesunde Individuen als Zielgruppe haben. Im letzteren Fall spricht man von Studien der **Primärprävention**.

Beispiel: In Interventionsstudien mit dem Ziel der Primärprävention werden (noch) gesunde Personen einer Präventionsmaßnahme ausgesetzt, wie etwa einer Raucherentwöhnung, um die Entstehung einer Krankheit, wie z.B. Lungenkrebs, zu verhindern. Dazu muss ausreichende Evidenz für einen kausalen Zusammenhang zwischen dem Risikofaktor, hier Rauchen, und der Krankheit, hier Lungenkrebs, vorliegen. Diese Evidenz kann etwa aus epidemiologischen Studien gewonnen werden.

Ein Präventionsprogramm allein macht jedoch noch keine Studie aus. Dazu ist eine wissenschaftliche Überprüfung der Wirksamkeit erforderlich, die einem stringenten zuvor festgelegten Protokoll folgt, in dem die Erfolgskriterien definiert sind. Eine solche Evaluation vollzieht sich üblicherweise auf den drei Ebenen:

- Struktur (u.a. Kosten, Zeitaufwand, praktische Probleme),
- Prozess (u.a. Teilnahme, Akzeptanz und Nachhaltigkeit) und
- Ergebnis (Verhaltensänderungen und gesundheitliche Endpunkte).

Studien der Primärprävention können zum einen das Ziel verfolgen, das Verhalten von Personen zu verändern. Zum anderen kann es aber auch das Ziel sein, die Lebensumwelt so zu gestalten, dass die neu geschaffenen Verhältnisse gesundheitsfördernd sind, z.B. durch fahrradfreundlichen Städtebau oder durch Rauchverbote im öffentlichen Raum. Häufig wird in **Präventionsstudien** eine Kombination aus **Verhältnis-** und **Verhaltensprävention** angestrebt.

Beispiel: Berühmte Beispiele zur Verhältnisprävention sind (a) die Einführung des Rauchverbots in öffentlichen Gebäuden sowie in Restaurants und Gaststätten, (b) die Jodierung von Trinkwasser oder Salz zur Redu-

zierung von Schilddrüsenerkrankungen oder (c) die Fluoridierung von Trinkwasser zur Vermeidung von Zahnkaries.

Beispiele für Verhaltensprävention sind (a) Die Kampagne "Fünf am Tag", mit der erreicht werden sollte, dass jeder Mensch pro Tag fünf Portionen frisches Obst oder Gemüse verzehrt, (b) der "Lauf zum Mond" zur Steigerung der körperlichen Aktivität im Rahmen der Deutschen Herz-Kreislauf-Präventionsstudie (DHP) oder (c) Raucherentwöhnungsprogramme.

Präventionsstudien können so angelegt werden, dass sie sich etwa an ganze Gemeinden richten. Diese Ebene ist insbesondere für die *Verhältnisprävention* geeignet. Bei solchen **Präventionsstudien auf kommunaler Ebene** bedient man sich häufig regionaler Zeitungen, Rundfunk- und Fernsehprogramme, um die Interventionsziele der Bevölkerung zu kommunizieren. Der Erfolg der damit verbundenen Maßnahmen hängt stark davon ab, wie gut es gelingt, lokale Akteure zur Unterstützung ihrer Umsetzung und ihrer Verstärkung zu gewinnen. Bei diesem Studientyp werden die Mitglieder der Zielbevölkerung nicht individuell angesprochen, selbst wenn dabei individuelle *Verhaltensänderungen* angestrebt werden. Entsprechend wird der Erfolg der Maßnahmen auf Gruppenniveau und nicht auf Individualniveau evaluiert. Zu diesem Zweck kann zum Beispiel in zwei aufeinander folgenden Querschnittsstudien, zwischen denen die Interventionsmaßnahmen stattgefunden haben, geprüft werden, ob die Prävalenzen für die interessierenden Merkmale sich in die gewünschte Richtung verändert haben. Eine andere Möglichkeit besteht darin, anhand von Sekundärdaten Veränderungen der Mortalität oder der Inzidenz zu untersuchen.

Im Gegensatz dazu wenden sich **Präventionsstudien auf Individualebene** einzelnen Individuen zu, die individuell für die Teilnahme an der Studie gewonnen werden müssen, um sie im Längsschnitt zu untersuchen. Diese Studienteilnehmer werden in einer so genannten **Grundlagenstudie** oder **Baseline-Survey** hinsichtlich der für die Studienziele entscheidenden Merkmale untersucht und befragt. Jeder einzelne Studienteilnehmer wird dann über die Studiendauer beobachtet, was verschiedene Zwischen- und Abschlussuntersuchungen einschließen kann. Idealerweise wird die so gebildete Kohorte nach Interventions- und Referenzgruppe aufgeteilt, wobei die Art der Aufteilung durch logistische Notwendigkeiten beeinflusst wird. Dabei wird der Interventionsgruppe ein umfangreiches Maßnahmenpaket angeboten, während in der Referenzgruppe aus ethischen Gründen statt keiner Intervention oft ein Minimalangebot gemacht wird. Dieses Minimalangebot kann z.B. eine einmalige Informationsveranstaltung oder nur einfaches Informationsmaterial beinhalten, während das umfangreiche Maßnahmenpaket zusätzlich intensive Schulungen und ggf. ein individualisiertes Feedback umfassen kann. Der Erfolg dieser Interventionsmaßnahmen kann auf individueller Ebene evaluiert werden, und zwar auf der einen Seite durch einen Vorher-Nachher-Vergleich der interessierenden Merkmale und auf der anderen Seite durch den Vergleich der entsprechenden Veränderungen zwischen Interventions- und Referenzgruppe.

Beispiel: In der IDEFICS-Studie wurden mehr als 16.000 Kinder aus acht europäischen Ländern für eine Studie zur Primärprävention von Übergewicht und Fettleibigkeit rekrutiert. Dazu wurden in jedem Land eine Interventions- und eine Vergleichsregion ausgewählt. Pro Region wurden je ca. 500 Kinder aus Grundschulen und Kindergärten in die Studie eingeschlossen. Alle Kinder durchliefen eine umfangreiche medizinische Basisuntersuchung, während ihre Eltern zu Essgewohnheiten, sozialen und Lebensstilfaktoren befragt wurden. Das Interventionsprogramm umfasste vier Ebenen: Ausgehend von Schule bzw. Kindergarten, auf die sich die Maßnahmen konzentrierten, erfolgten Angebote zur Verhaltensänderung für die Kinder (Individuen) und ihre Familien, während auf Gemeindeebene und bei strukturellen Veränderungen wie z.B. sichere Radwege und

bewegungsfördernde Schulhöfe initiiert wurden. Damit wurden im Rahmen der vielschichtigen Intervention der IDEFICS-Studie Verhaltens- und Verhältnisprävention miteinander kombiniert. Dabei standen die Lebensbereiche Ernährung, körperliche Aktivität und Stressbewältigung im Mittelpunkt.

Den Kontrollregionen wurde angeboten, dass sie nach Evaluation der Interventionsmaßnahmen das Maßnahmenpaket zur Verfügung gestellt bekommen und in seiner Umsetzung geschult werden.

Zwei Jahre nach der Basisuntersuchung und nach Einführung des Interventionsprogramms wurden allen Studienteilnehmer zu einer Folgeuntersuchung eingeladen, der mehr als 70% der Teilnehmer folgten. Die Evaluation der Interventionseffekte erfolgte durch einen Vergleich von Basis- und Folgeuntersuchung. Um beobachtete Veränderungen tatsächlich den Interventionsmaßnahmen zuschreiben zu können und von generellen zeitlichen Trends zu unterscheiden, erfolgte ein gleichsinniger Vergleich in der Referenzgruppe.

Beispiel: Auch im Anwendungsbereich des Veterinary Public Health sind Interventionsstudien ein wesentliches Instrument der Verbesserung der Qualität der Tiergesundheit und der Lebensmittelsicherheit. Ein Beispiel für die Prävention auf Individualebene geben Beyerbach et al. (2001). Hier wird das Problem der Infektion von Milchviehherden mit dem potenziellen Zoonoseerreger *mycobacterium paratuberculosis* betrachtet, der in den betroffenen Beständen erhebliche wirtschaftliche Verluste verursacht. Die Erkrankung hat eine lange, bis zu mehreren Jahren währende, Inkubationszeit und zeigt eine langsame Ausbreitungstendenz innerhalb der betroffenen Herden. Die Diagnose infizierter Tiere wird durch eine intermittierende Erregerausscheidung, eine sich nur langsam aufbauende humorale Immunantwort und eine wenig konstant auftretende zelluläre Immunantwort erschwert. Zudem wird ein Zusammenhang mit der Morbus-Crohn-Erkrankung des Menschen diskutiert.

Daher ist es zur Sanierung von Betrieben erforderlich, zunächst einen Status quo der Erregerprävalenz zu bestimmen. Anschließend werden gezielte Interventionen durchgeführt, wie etwa die Selektion von Tieren, Verkaufseinschränkungen und spezifische hygienische Maßnahmen, die in einer nachfolgenden Erhebung evaluiert werden.

Auswertungsprinzipien

Bei Interventionsstudien gehen wir davon aus, dass alle Mitglieder der Kohorte (= Studienpopulation) über ein festes Zeitintervall beobachtet werden. Im Folgenden werden wir die Auswertungsstrategie am Beispiel einer klinischen Studie formalisieren, d.h., wir gehen davon aus, dass

- jedes Individuum über das Zeitintervall $\Delta = (t_0, t_s)$ vollständig beobachtet wird und zum Zeitpunkt t_s entweder geheilt oder aber nicht geheilt ist und
- jedes Individuum zum Zeitpunkt t_0 eindeutig als therapiert ($Ex = 1$) oder nicht therapiert ($Ex = 0$) klassifizierbar ist.

Handelt es sich bei der Interventionsstudie um eine Präventionsstudie, so wird am Ende der Beobachtungsperiode im einfachsten Fall die Anzahl der Interventionserfolge gezählt, also beispielsweise die Anzahl der Individuen, die ihr Verhalten geändert haben, oder die Anzahl an Individuen, bei denen das Auftreten einer Krankheit verhindert werden konnte. Entsprechend müssen die nachfolgenden Ausführungen auf den interessierenden Fall hinsichtlich ihrer Interpretation übertragen werden.

Kommen wir also auf die Situation einer klinischen Studie zurück. Sind zu Beginn der Beobachtungsperiode t_0 insgesamt $n_{1.}$ Individuen in der Behandlungsgruppe und $n_{0.}$ Individuen in der Placebogruppe, so kann am Ende des Follow-up die Vierfeldertafel Tab. 3.8 erstellt werden.

Tab. 3.8: Beobachtete Anzahlen in der Behandlungsgruppe und in der Referenzgruppe (Placebo), aufgeteilt nach Therapieerfolg (geheilt, nicht geheilt) in einer klinischen Studie

Status	Behandlungsgruppe (Ex = 1)	Placebogruppe (Ex = 0)	Summe
geheilt	n_{11}	n_{10}	$n_{1.} = n_{11} + n_{10}$
nicht geheilt	n_{01}	n_{00}	$n_{0.} = n_{01} + n_{00}$
Summe	$n_{1.} = n_{11} + n_{01}$	$n_{0.} = n_{10} + n_{00}$	$n = n_{..}$

In Tab. 3.8 bezeichnet n_{1j} die beobachtete Anzahl von geheilten Individuen und n_{0j} die beobachtete Anzahl von nicht geheilten Individuen aus der Behandlungsgruppe j ($j = 0, 1$) am Ende der Periode Δ . Tab. 3.8 stimmt im Wesentlichen mit Tab. 3.6 überein, da eine Interventionsstudie durch das prospektive Design mit einer Kohortenstudie übereinstimmt. Das heißt, dass auch hier charakteristisch ist, dass die *Spaltensummen* $n_{1.} = n_{11} + n_{01}$ bzw. $n_{0.} = n_{10} + n_{00}$ als die Umfänge der Interventionsgruppe und der Referenz(Placebo-)gruppe *vorgegeben* sind und sich erst nach der erfolgten Therapie dieser Individuen eine Aufteilung in geheilte und nicht geheilte Personen ergibt.

Die kumulative Inzidenz bezieht bei diesem Studientyp die Anzahl der Geheilten in der Periode Δ auf die Gesamtzahl der mit der neuen Behandlung bzw. mit Placebo behandelten Individuen (siehe Abschnitt 2.1.1). Analog zur Kohortenstudie können die **"Risiken" (Wahrscheinlichkeiten) für die Heilung** aus Tab. 3.1 damit sinnvollerweise mit nachfolgenden Größen aus der Interventionsstudie **geschätzt** werden.

$$\text{Schätzer kumulative Inzidenz (bei Exposition } j) = CI(j)_{\text{Intervention}} = \frac{n_{1j}}{n_{.j}}, j = 0, 1.$$

Analog zur Kohortenstudie lässt sich daraus auch ein **Schätzwert** für das **relative Risiko** RR für Heilung berechnen als

$$\text{Schätzer relatives Risiko} = RR_{\text{Intervention}} = \frac{CI(1)_{\text{Intervention}}}{CI(0)_{\text{Intervention}}} = \frac{n_{11}/(n_{11} + n_{01})}{n_{10}/(n_{10} + n_{00})},$$

bzw. für das **Odds Ratio** OR

$$\text{Schätzer Odds Ratio} = \text{OR}_{\text{Intervention}} = \frac{n_{11}/n_{01}}{n_{10}/n_{00}} = \frac{n_{11} \cdot n_{00}}{n_{10} \cdot n_{01}}.$$

Analog zu diesen Effektmaßen sind auch die Risikodifferenz und die verschiedenen attributablen Risikomaße (siehe Abschnitt 2.4) schätzbar.

Für den Fall, dass in einer Interventionsstudie die Anzahl der Teilnehmer nicht fest ist, also eine zur offenen Kohorte analoge Situation eintritt, so liegen die Häufigkeitsanzahlen aus Tab. 3.8 nicht in dieser Form vor. Es ist zwar möglich, die Anzahlen der geheilten Individuen in der Behandlungsgruppe n_{11} bzw. in der Placebogruppe n_{10} anzugeben, die Anzahl der Kohortenmitglieder sollte aber gemäß Abschnitt 2.1.2 durch die Zeit unter Risiko ersetzt werden. Dann kann aus der Studie jeweils eine Inzidenzdichte für die zu vergleichenden Gruppen ermittelt werden. Man erhält somit als **Schätzer** für das **relative Risiko** bei einer **offenen Interventionsstudie**

$$\text{Schätzer relatives Risiko} = \text{RR}_{\text{offen}} = \frac{\text{ID}(1)_{\text{Intervention}}}{\text{ID}(0)_{\text{Intervention}}} = \frac{n_{11}/\text{P}\Delta(1)_{\text{Intervention}}}{n_{10}/\text{P}\Delta(0)_{\text{Intervention}}}.$$

Hierbei stellen die Größen $\text{P}\Delta(1)_{\text{Intervention}}$ und $\text{P}\Delta(0)_{\text{Intervention}}$ die aus der Interventionsstudie ermittelten Gesamttrisikozeiten in den beiden Behandlungsgruppen dar (siehe Abschnitt 2.1.2).

Beispiel: Im Rahmen der US-amerikanischen Women’s Health Initiative (WHI) wurden zwischen 1993 und 1998 161.809 Frauen im Alter von 50 bis 79 Jahren durch 40 klinische Zentren in eine Serie klinischer Studien (Fettreduktion, Supplementierung mit Vitamin C und D, postmenopausale Hormontherapie) und in eine Beobachtungsstudie eingeschlossen. In den auf 8,5 Jahre angelegten Studienteil zur Hormontherapie mit einer Kombination aus Östrogen und Progesteron wurden im genannten Zeitraum 16.608 Frauen mit zur Basisuntersuchung intaktem Uterus randomisiert. Die Teilnehmerinnen erhielten entweder das Hormonpräparat (n = 8.506) oder Placebo (n = 8.102) (siehe Abb. 3.13). Die Verabreichung erfolgte doppelt verblindet.

Das Ziel der Studie bestand darin, die wichtigsten gesundheitlichen Nutzeffekte und Risiken der in den USA gebräuchlichsten Hormontherapie zu ermitteln (Rossouw et al. 2002). Die primär interessierenden Endpunkte waren auf der Nutzenseite kardiovaskuläre Erkrankungen, insbesondere Herzinfarkt (wofür ein protektiver Effekt der Therapie erwartet wurde) und bezüglich möglicher Risiken das Auftreten von Brustkrebs (wofür ein Risiko durch die Therapie erwartet wurde). Die Studie musste aufgrund des Überwiegens der beobachteten Gesundheitsrisiken zu möglichen Nutzeffekten nach durchschnittlich 5,2 Jahren Beobachtungszeit gestoppt werden. Bezogen auf die beiden zentralen Endpunkte ergab die Studie das in Tab. 3.9 dargestellte Ergebnis.

Tab. 3.9: Inzidenz von Brustkrebs und Herzinfarkt (Beobachtungszeit ca. 5,2 Jahre) in der Hormontherapie-Interventionsstudie der WHI (Quelle: Rossouw et al. 2002)

	Hormone (n = 8.506)	Placebo (n = 8.102)	RR _{Intervention}
Brustkrebsfälle	166	124	1,28
Herzinfarktfälle	164	122	1,28

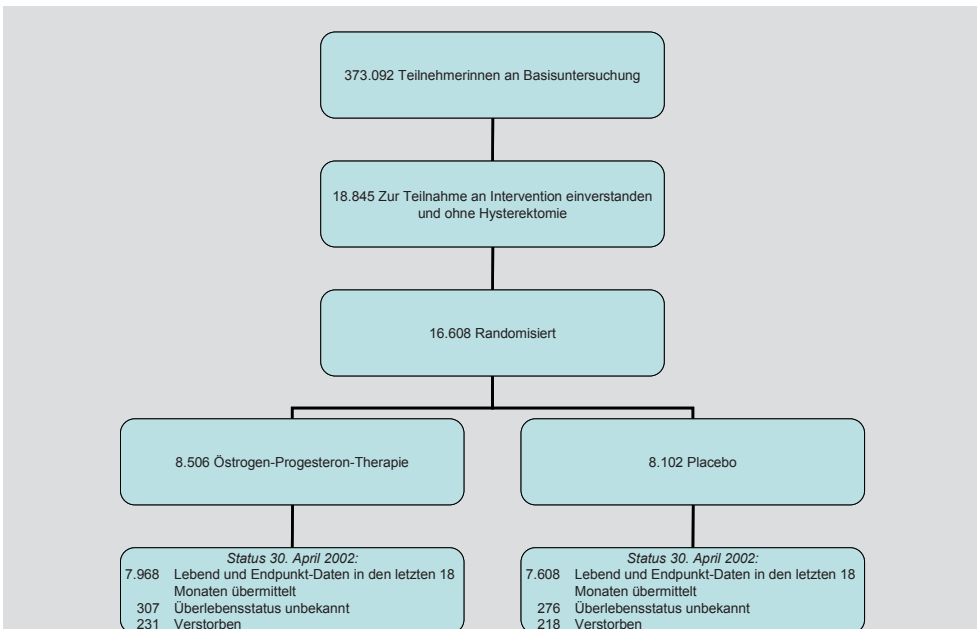


Abb. 3.13: Rekrutierungsschema und Randomisierung der Studienteilnehmerinnen der Women's Health Initiative in der Interventionsstudie zur Hormontherapie nach der Menopause

Näherungsweise lässt sich das relative Risiko für das Auftreten von Brustkrebs in dieser Studie berechnen mit

$$\text{Schätzer relatives Risiko} = \text{RR}_{\text{Intervention}} = \frac{\text{CI(I)}_{\text{Intervention}}}{\text{CI(0)}_{\text{Intervention}}} = \frac{166 / 8.506}{124 / 8.102} = \frac{0,0195}{0,0153} = 1,28.$$

Dies entspricht einem um 28% erhöhten Risiko für Brustkrebs bei einer postmenopausalen Östrogen-Progesteron-Therapie gegenüber Placebo. Auch für das Herzinfarktrisiko ergibt sich mit $\text{RR}_{\text{Intervention}} = 1,28$ ein erhöhter Wert. Einschränkend ist hier zu erwähnen, dass diese Art der Berechnung davon ausgeht, dass es sich um eine geschlossene Kohorte handelt. Tatsächlich handelte es sich aber um eine offene Kohorte, in die Frauen während der Beobachtungszeit ein- und aus der sie wieder austreten konnten. Die Autoren publizierten daher die Ergebnisse der korrekteren Analyse auf Basis der Personenzzeit, von der jedoch die hier gezeigte grobe Berechnung nur geringfügig abweicht.

Nachdem vorangegangene Beobachtungsstudien bereits auf das erhöhte Brustkrebsrisiko durch postmenopausale Hormontherapie hingewiesen hatten, gab das Ergebnis dieser Interventionsstudie schließlich den Anlass, diese Therapie nicht weiter zu empfehlen und stattdessen vor ihren Risiken zu warnen.

Vorteile von Interventionsstudien

Der größte Vorteil einer Interventionsstudie besteht darin, dass sie, insbesondere wenn sie randomisiert durchgeführt wird, den *höchsten Evidenzgrad für eine kausale Assoziation* besitzt. Das liegt einerseits an ihrem longitudinalen Design und andererseits an dem konkreten Vergleich. Beides ermöglicht es, Veränderungen in der Interventionsgruppe tatsächlich auf die erfolgte Intervention zurückzuführen. Durch die Randomisierung wird zudem eine hohe

Vergleichbarkeit der beiden Gruppen erreicht, da hierdurch eine gleiche Verteilung von Störfaktoren auf die zu vergleichenden Gruppen erreicht wird.

Dementsprechend sind Präventionsaktivitäten, die nicht durch eine Vergleichsgruppe kontrolliert werden, mit äußerster Skepsis zu betrachten. Dazu gehört auch, dass solchen Aktivitäten in der Regel kein formales Studienprotokoll zugrunde liegt und ihre Evaluation – wenn überhaupt – nur durch die Versendung eines qualitativen Fragebogens erfolgt, der sich z.B. nur auf die Akzeptanz der Maßnahme durch die Teilnehmer bezieht.

Im Übrigen gelten natürlich alle Vorteile einer Kohortenstudie für eine Interventionsstudie entsprechend.

Nachteile von Interventionsstudien

Auch wenn Interventionsstudien der höchste Evidenzgrad beigemessen wird, sind mit ihnen auch Nachteile verknüpft. Neben dem mit einer solchen Studie verbundenen *hohen Kosten- und Zeitaufwand* sind dabei folgende Aspekte zu nennen.

Bei einer klinischen Studie liegt ein großer Nachteil darin, dass ein *hoch selektiertes Kollektiv* in die Studie eingeschlossen wird. So sind im Allgemeinen multimorbide Patienten, alte Menschen, und Kinder ausgeschlossen. Dadurch wird die externe Validität eingeschränkt. Hinzu kommt, dass die Studienteilnehmer einem sehr stark reglementierten Regime unterliegen, das sowohl die Einnahme zusätzlicher Präparate als auch Lebensstile wie z.B. Ernährungsweisen einschränkt. Dadurch werden Wechselwirkungen mit anderen Präparaten häufig erst nach der Zulassung eines neuen Medikaments entdeckt, wenn die Patienten das neue Medikament unter Alltagsbedingungen einnehmen. Auch seltene Nebenwirkungen können in der Regel erst nach der Zulassung beobachtet werden, da dazu eine wesentlich größere Anzahl von Personen erforderlich sind, die das Präparat einnehmen. Langzeiteffekte – sowohl im positiven als auch im negativen Sinn – können im Rahmen zeitlich begrenzter klinischer Studien nicht beobachtet werden.

Beispiel: Bei der Behandlung von chronischen Schmerzpatienten, z.B. bedingt durch Arthritis, werden typischerweise nicht-steroidale Anti-Rheumatika (NSAR) verabreicht, die jedoch bei Einnahmen über längere Zeiträume zu ernsthaften unerwünschten Arzneimittelwirkungen führen können. U.a. können gastrointestinale Blutungen auftreten.

Daher wurde nach Alternativen gesucht, bei denen diese unerwünschten Arzneimittelwirkungen nicht auftreten. Entwickelt und auf dem Markt zugelassen wurden so genannte Cox-2-Inhibitoren, unter denen gastrointestinale Blutungen nicht beobachtet wurden. Jedoch traten nach längerer Einnahmezeit vermehrt unerwünschte kardiovaskuläre Wirkungen auf wie z.B. Myokardinfarkt oder Schlaganfall. Diese Komplikationen waren im Rahmen der üblichen Laufzeit einer klinischen Studie nicht zu entdecken. Sie führten zur Marktrücknahme von VIOXX und anderen Cox-2-Inhibitoren.

Randomisierte kontrollierte Studien werden auch mit dem Ziel durchgeführt, der *Entstehung von Krankheiten durch eine geeignete Interventionsmaßnahme* entgegenzuwirken. So wird derzeit versucht, der Entstehung von Gebärmutterhalskrebs durch Impfung gegen den Human Papilloma Virus (HPV) entgegenzuwirken. Solche Studien haben mit mehreren Prob-

lemen zu kämpfen. Da von einem Erfolg der Maßnahme nur gesprochen werden kann, wenn die Krankheit tatsächlich später oder gar nicht aufgetreten ist, stellen sich zwei Fragen, und zwar einerseits hinsichtlich der *erforderlichen Laufzeit der Studie* und andererseits hinsichtlich der *ethischen Vertretbarkeit*.

Viele Studien beschäftigen sich mit den großen Volkskrankheiten, Krebs und Herz-Kreislaufkrankungen, die typischerweise eine lange Latenzzeit haben (siehe obiges Beispiel zur WHI). Derartige *späte Endpunkte* erfordern ggf. Studien, die über mehrere Jahre oder Jahrzehnte laufen müssten, was mit einer *enormen Logistik*, extrem *hohen Kosten* und möglicherweise einer *hohen Ausfallquote* der Studienteilnehmer verbunden ist. Insbesondere Letzteres kann zu erheblichen Verzerrungen der Studienergebnisse führen, wenn der Ausfallgrund mit der Intervention oder dem Interventionserfolg in Zusammenhang steht (siehe auch Kapitel 4). Ist die Interventionsmaßnahme erfolgreich, kommt als *ethisches Problem* hinzu, dass man sie besser schon früher der Zielpopulation angeboten hätte. Außerdem ist es kaum ethisch vertretbar, in den Studienarmen auf das Auftreten von Krebs zu warten, wenn man dem Krebs schon vorher durch entsprechende frühzeitige Behandlung hätte entgegenwirken können. Daher wird nach alternativen Endpunkten gesucht, die bereits frühzeitig sichtbar werden und als *Surrogatvariablen* für den eigentlichen Endpunkt dienen können.

Beispiel: Ist z.B. Gebärmutterhalskrebs der eigentliche interessierende Endpunkt, so könnte der Erfolg der Maßnahme auch an Präkanzerosen evaluiert werden wie z.B. zervikaler intraepithelialer Neoplasie oder Adenomakarzinom in situ, die auf dem Krankheitsentstehungspfad dem Zervixkarzinom selbst vorgelagert sind.

Eine andere Möglichkeit besteht darin zu überprüfen, ob sich durch die Maßnahme die Verteilung der primären Risikofaktoren verändert hat.

Beispiel: So wurde im Rahmen der Deutschen Herzkreislauf-Präventionsstudie (DHP) überprüft, ob sich durch das Präventionsprogramm z.B. das Ernährungsverhalten und das Rauchverhalten positiv verändert haben.

Dies leitet zu dem großen Problem von Interventionsstudien über, die auf Verhaltensänderungen abzielen. Typischerweise ist es sehr schwierig, Lebensstile von Personen durch Intervention von außen zu verändern, so dass derartige Studien nur selten erfolgreich sind. Im Gegenteil sind die erzielten Effekte, wenn überhaupt, eher klein und häufig auch nicht nachhaltig, d.h., sobald das aktive Interventionsprogramm mit entsprechender wissenschaftlicher Begleitung beendet ist, versanden viele positive Ansätze hin zu einem gesundheitsfördernden Verhalten wieder. Nachhaltigkeit kann z.B. durch entsprechende gesetzliche Regelungen erreicht werden, wie das Rauchverbot in öffentlichen Gebäuden, Restaurants oder Gaststätten zeigt.

3.3.6 Bewertung epidemiologischer Studiendesigns

In den obigen Abschnitten haben wir die verschiedenen Studiendesigns so eingeführt, dass mit jedem neu eingeführten Design ein höherer Evidenzgrad für kausale Zusammenhänge erreicht werden kann. In Tab. 3.10 wird noch einmal zusammenfassend deutlich, inwieweit

sich die unterschiedlichen Studientypen für die Bewertung ätiologischer Zusammenhänge eignen und welchen Schluss die jeweiligen Studientypen zulassen. Dabei bleibt unberücksichtigt, welche anderen Zwecke mit dem jeweiligen Design zusätzlich verfolgt werden können, wie z.B. die Gesundheitsberichterstattung mittels Querschnittsstudien.

Tab. 3.10: Aussagekraft hinsichtlich ätiologischer Zusammenhänge bei unterschiedlichen epidemiologischen Studientypen

Studientyp	Aussagekraft
Ökologische Korrelation	Assoziation <i>auf Gruppen-Niveau</i> : Hypothesengenerierung
Querschnittsstudie	<i>Individuelle</i> Assoziation: überwiegend Hypothesengenerierung
Fall-Kontroll-Studie	Schluss von vermehrter Exposition bei den Erkrankten auf eine Expositions-Effekt-Beziehung
Kohortenstudie	Schluss vom erhöhten Erkrankungsrisiko der Exponierten auf Expositions-Effekt-Beziehung
Interventionsstudie	Schluss von der Veränderbarkeit (Reversibilität) der Krankheitsinzidenz durch Veränderung des Risikofaktors auf Kausalität

Im Folgenden werden wir die Grundprinzipien, auf denen die verschiedenen Studientypen basieren, bewertend gegenüberstellen.

Bei randomisierten kontrollierten Studien wird die aus der Auswahlpopulation gezogene Studienpopulation nach dem Zufallsprinzip (Randomisierung) auf zwei Behandlungsarme aufgeteilt. Jede Gruppe wird ab Behandlungsbeginn über einen angemessenen Zeitraum beobachtet, um zu ermitteln, bei wie vielen der Patienten ein Therapieerfolg eingetreten ist. Man spricht diesem Studientyp den höchsten Evidenzgrad hinsichtlich der Bewertung kausaler Zusammenhänge zu, da er einen *Randomisierungsschritt* beinhaltet, der den Einfluss von Störfaktoren eliminieren soll, und da er die Interventionsgruppe der Studienpopulation bewusst (experimentell) gegenüber einem Faktor (Behandlung, Präventionsmaßnahme) exponiert und den *nachfolgenden Effekt* im Vergleich zur Referenzgruppe ermittelt.

Eine Kohortenstudie folgt grundsätzlich dem gleichen Prinzip wie eine randomisierte Studie, allerdings mit dem entscheidenden Unterschied, dass bei ihr die Zuordnung zu den beiden Vergleichsgruppen nicht randomisiert erfolgen kann, da die Exposition aus ethischen oder praktischen Gründen nicht experimentell zugewiesen werden kann. Dieses Design ist daher nur beobachtend: *Ausgehend vom Expositionsstatus* wird über den Beobachtungszeitraum (Follow-up-Periode) die *Krankheitsinzidenz* zwischen Exponierten und Nichtexponierten

miteinander verglichen. Dieses Design wird als *prospektiv* bezeichnet, weil die Beobachtungsrichtung synchron zur zeitlichen Abfolge von Exposition und Erkrankung erfolgt.

Eine Fall-Kontroll-Studie nimmt ihren *Ausgangspunkt* dort, wo eine Kohortenstudie ihren Endpunkt hat: *bei der Erkrankung* (den Fällen). Von dort wird rückblickend ermittelt, welche Expositionen der Erkrankung vorausgegangen sind. Um zu beurteilen, ob eine der interessierenden Expositionen bei Fällen tatsächlich gehäuft vorgekommen ist, wird letztlich die *Expositionsprävalenz* von Fällen und Kontrollen miteinander verglichen. Die Richtung der Beobachtung verläuft also bei Fall-Kontroll-Studien umgekehrt zur zeitlichen Abfolge von Exposition und Erkrankung. Deshalb charakterisiert man diesen Studientyp als *retrospektiv*.

Bei Querschnittsstudien erfolgt im Prinzip eine Momentaufnahme der Zielpopulation, bei der die *aktuelle Morbidität gleichzeitig mit aktuell vorliegenden Expositionen* ermittelt wird. Im Rahmen eines solchen Surveys können die Prävalenzen sowohl für bestehende Erkrankungen als auch für bestehende Expositionen bestimmt werden (Prävalenzstudie).

Eine Zusammenfassung der Begrifflichkeiten der in diesem Kapitel beschriebenen Studiendesigns wird in Tab. 3.11 gegeben. In der ersten Spalte sind die in diesem Buch primär verwendeten Bezeichnungen genannt; gebräuchliche Alternativbezeichnungen finden sich – ohne Anspruch auf Vollständigkeit – in der mittleren Spalte. Teilweise sind die verwendeten Begriffe nicht eindeutig definiert oder werden in der Literatur uneinheitlich verwendet. Jedes dieser Designs lässt sich durch die jeweils betrachtete Untersuchungseinheit charakterisieren (Spalte 3).

Tab. 3.11: Systematik der unterschiedlichen Studientypen in der Epidemiologie

Studientyp	Alternativbezeichnung	Untersuchungseinheit
<i>Beobachtungsstudien</i>		
Ökologische Studie	Korrelationsstudie	Populationen
Querschnittsstudie	Prävalenzstudie; Survey	Individuen
Fall-Kontroll-Studie	Case-Referent Study	Individuen
Kohortenstudie	Follow-up; Longitudinal	Individuen
<i>Experimentelle Studien</i>		
Präventionsstudie auf Gemeindeebene	Gemeindeintervention	Gruppen (Gemeinden)
Präventionsstudie auf Individualebene	Field Intervention	Gesunde Individuen
Randomisierte kontrollierte Studie	Klinische Studie	Individuelle Patienten

Häufig werden auch die Begriffe "*analytische Studie*" und "*deskriptive Studie*" zur Charakterisierung epidemiologischer Beobachtungsstudien verwendet. Als deskriptiv werden Studien charakterisiert, die die Verteilung oder zeitliche Entwicklung von Gesundheitszuständen und ihren Determinanten in einer Population beschreiben. In analytischen Studien werden Endpunkte in Abhängigkeit von Einflussfaktoren gesetzt, um z.B. ätiologische Zusammenhänge zwischen Risikofaktoren und Erkrankungen zu quantifizieren. Mit dem Begriffspaar deskriptiv/analytisch lässt sich keines der Designs eindeutig definieren, jedoch haben Querschnittsstudien eher deskriptiven und Kohortenstudien eher analytischen Charakter.

Um die Stärken und Schwächen der jeweiligen Studiendesigns einander gegenüber zu stellen, haben wir einige wesentliche Kriterien ausgewählt und eine Bewertung der vier beobachtenden Studiendesigns vorgenommen, inwieweit sie diese Kriterien erfüllen (siehe Tab. 3.12). Im Detail wurden diese bereits in den vorangegangenen Abschnitten zu dem jeweiligen Design diskutiert.

Tab. 3.12: Vor- und Nachteile verschiedener Typen von Beobachtungsstudien

Bewertungskriterium	Studientyp			
	ökologisch	Querschnitt	Fall-Kontroll	Kohorte
Seltene Krankheiten	😊😊😊	😞	😊😊😊😊😊	😞
Seltene Ursachen	😊	😞	😞	😊😊😊😊😊
Multiple Endpunkte	😊	😊	😞	😊😊😊😊😊
Multiple Expositionen	😊	😊	😊😊😊	😊 ¹⁾
Zeitliche Abfolge	😊	😞	😊 ²⁾	😊😊😊😊😊
Direkte Ermittlung der Inzidenz	😞	😞	😊 ¹⁾	😊😊😊😊😊
Lange Induktionszeit	😞	😞	😊😊	😊 ³⁾

¹⁾ Wenn populationsbasiert ²⁾ Wenn in Kohorte eingebettet ³⁾ Wenn historisch

Fall-Kontroll-Studien sind demnach besonders gut geeignet, um seltene Erkrankungen zu untersuchen, wobei sie bezogen auf die ausgewählte Erkrankung eine Vielzahl von Expositionen betrachten können. Dabei können sie auch zeitlich lang zurückliegende Expositionen einbeziehen.

Im Gegensatz dazu eignen sich Kohortenstudien besonders gut zur Untersuchung seltener Expositionen (man denke z.B. an Industriekohorten, die gegenüber bestimmten Gefahrstoffen exponiert sind), können aber aufgrund des Designs jeweils mehrere Endpunkte zu einer gegebenen Exposition in Beziehung setzen. Multiple Expositionen sind nur in so genannten Vielweckkohorten untersuchbar, die z.B. einen populationsbasierten Survey als Ausgangspunkt haben können. Eine besondere Stärke von Kohortenstudien ist die eindeutige zeitliche Abfolge zwischen Exposition und Erkrankung, bei der die Expositionsdaten unabhängig

vom Erkrankungsstatus bzw. vor Eintritt der Erkrankung gewonnen werden. Letzteres gilt für Fall-Kontroll-Studien nur dann, wenn sie in eine Kohortenstudie eingebettet sind.

Querschnittsstudien eignen sich sowohl für die Untersuchung verschiedener Erkrankungen als auch verschiedener Expositionen. Auch wenn sich ökologische Korrelationen gut für die Betrachtung seltener Erkrankungen eignen, ist ihre Aussagekraft dadurch beschränkt, dass die Expositions- und Krankheitsdaten nur auf Gruppenniveau zueinander in Beziehung gesetzt werden, was diesen Studientyp anfällig für den so genannten ökologischen Trugschluss macht.

Epidemiologische Methoden

Kreienbrock, L.; Pigeot, I.; Ahrens, W.

2012, XIV, 498 S. 70 Abb., Softcover

ISBN: 978-3-8274-2333-7