

Skript zur Veranstaltung

Epidemiologische Methoden und Anwendungen

Hochschule Magdeburg-Stendal
FB Sozial- und Gesundheitswesen

erstellt von:

Dr. Enno Swart, Diplom-Statistiker

Institut für Sozialmedizin und Gesundheitsökonomie (ISMHE)
Medizinische Fakultät der Otto-von-Guericke-Universität Magdeburg
Leipziger Str. 44, 39120 Magdeburg

☎ (0391) 67 24 306; Fax (0391) 67 24 310

✉ enno.swart@med.ovgu.de oder

✉ kontakt@enno-swart.de

Version 6, Januar 2008

Inhaltsverzeichnis

0. Vorbemerkung.....	4
1. Epidemiologie – zur Begriffsbestimmung	5
2. Kurze Anmerkungen zur Geschichte der Epidemiologie.....	7
2.1 John Snow und die Cholera in London	7
2.2 Semmelweiss und das Kindbettfieber	8
2.3 Die Entwicklung der Epidemiologie in Deutschland.....	10
3. Grundlagen der Statistik.....	13
3.1 Deskriptive Statistik	13
3.2 Lage- und Streuungsparameter von Verteilungen.....	14
3.3 Die wichtigsten Verteilungen der Statistik.....	18
3.3.1 Normalverteilung	18
3.3.2 Binomialverteilung.....	23
3.4 Statistische Testverfahren	25
3.4.1 Prinzip statistischer Testverfahren	25
3.4.2 t-Test auf Mittelwertunterschiede	27
3.4.3 Chi-Quadrat-Test.....	29
3.5 Analyse des Zusammenhangs zwischen Variablen.....	30
3.5.1 Korrelationsanalyse.....	30
3.5.2 Regressionsanalyse.....	33
4. Kurzer Abriss zu Methoden der Sozialwissenschaften	35
4.1 Gütekriterien eines Verfahrens.....	35
4.2 Aspekte der Befragung	36
4.3 Interview und Befragung.....	38
4.4 Beurteilungsfehler	39
5. Datenquellen für die Epidemiologie	40
6. Maßzahlen der Epidemiologie	45
6.1 Maßzahlen der Morbidität und Mortalität.....	45
6.2 Maßzahlen aus der Demographie	50
6.3 Standardisierung von Maßzahlen	56
7. Epidemiologische Studientypen.....	60
7.1 Experiment	60
7.2 Randomisierte Studie	60
7.3 Fall-Kontroll-Studie	61
7.4 Kohortenstudie	62
7.5 Querschnittstudie.....	63
7.6 Korrelationsstudien	64
7.7 Übersicht über die wichtigsten Studientypen.....	65
8. Zusammenhangsmaße	67
8.1 Relatives Risiko.....	67
8.2 Odds Ratio.....	69
8.3 Zusammenhang zwischen Relativem Risiko und Odds ratio.....	70
8.4 Attributables Risiko.....	70
9. Aspekte der Studienplanung.....	72
9.1 Verzerrungsmöglichkeiten	72
9.2 Confounder.....	73
9.3 Stichproben.....	74

9.4	Datenschutz	76
9.5	Arbeitsschritte in einer wissenschaftlichen Studie	77
9.6	Beispiel zur Studienplanung	78
10.	Eine Übung zur Studienanalyse	81
10.1	Fall-Studie mit Übungsaufgaben	81
10.2	Antworten auf die Fragen	84
11.	Grundzüge der klinischen Epidemiologie	86
11.1	Grundlagen	86
11.2	Ein Beispiel	90
12.	Assoziation oder Ursache – Das epidemiologische Kausalitätsprinzip	97
12.1	Die epidemiologische Bewertung von Assoziationen	97
12.2	Fall-Beispiel: Rauchen und Lungenkrebs	102
12.3	Antworten zur Fall-Studie	106
13.	Gute Epidemiologische Praxis	110
14.	Hinweise zum kritischen Lesen epidemiologischer Literatur	114
14.1	Kriterien der Literaturanalyse	114
14.2	Ein Beispiel zur Literaturanalyse	117
	Literatur	119
	Ergänzende Literatur zu einzelnen Kapiteln	122
	Anhang	125

0. Vorbemerkung

In der Regel biete ich im Sommersemester eine, im Wintersemester zwei Veranstaltungen zum Erwerb des Pflichtscheins Forschungs- und Planungsmethoden an. Sie behandeln mit unterschiedlichen Ansätzen die wichtigsten Methoden der Epidemiologie und verdeutlichen anhand konkreter Beispiele Probleme und Lösungsansätze bei der Durchführung und Auswertung von Studien. Aufgrund der begrenzten Zeitkapazität können nicht alle wichtigen Punkte im gebotenen Umfang besprochen werden. Dieses Skript dient daher in beiden Veranstaltungen der Vertiefung des Stoffes und enthält weiterführende Inhalte zur Abrundung der Thematik.

Beide Veranstaltungen - Seminar und Praktikum - wollen im einzelnen:

- die Bedeutung der Epidemiologie als grundlegendes und unverzichtbares Handwerkszeug der Gesundheitswissenschaften verdeutlichen,
- Verständnis und Interesse für Arbeitsweisen und Zielsetzungen der Epidemiologie wecken,
- grundlegende Methoden und Techniken der Epidemiologie vermitteln,
- praktische Aspekte und Probleme der Epidemiologie anhand ausgewählter Studien aufzeigen,
- die vielfältigen Anwendungsbereiche der Epidemiologie demonstrieren,
- die Fähigkeit für eine kritische Auseinandersetzung mit wissenschaftlichen Studien vermitteln,
- die Grundlagen für eine selbständige Anwendung epidemiologischer Methoden legen
- Gelegenheit zur Einübung des Erlernten geben.

In diesem Skript sind – neben einem ausführlichen Kapitel zu den wichtigsten statistischen Begriffen, das einer anderen Veranstaltung entnommen wurde - die wichtigsten Termini, Studententypen, Auswerteverfahren der Epidemiologie aufgeführt. Auch die praktischen Probleme der Studienplanung und Kriterien bei der kritischen Bewertung von epidemiologischen Studien werden angesprochen. Einige ausführlich dargestellte Beispiele runden die Darstellung ab. Einige Inhalte, die in der Veranstaltung aufgrund der knapp bemessenen Zeit nicht explizit behandelt werden konnten, wurden in das Skript aufgenommen, um einen ‚runden‘ Einblick in Arbeitsweise und Methoden der Epidemiologie als einem der wichtigsten, wenn nicht dem zentralen Fach der Gesundheitswissenschaften bzw. Public Health zu vermitteln.

Das Skript erhebt nicht den Anspruch eines Lehrbuches. Im Literaturverzeichnis sind daher zahlreiche weiterführende Quellen aufgezeichnet, die zum Teil auch in der Bibliothek der Hochschule Magdeburg-Stendal verfügbar sind. Für eine vertiefende und nachgehende Bearbeitung der in beiden Veranstaltung besprochenen Themen werden diese ausführlicheren, mit zahlreichen Beispielen illustrierten Monographien empfohlen (vor allem Beaglehole et al. 1997, Gordis 2000, Kreienbrock und Schach 2000).

Dieses Skript wird in Zukunft voraussichtlich weiterer Überarbeitungen, Korrekturen und Ergänzung erfahren (einschließlich der Ausmerzung verbliebener Schreibfehler). Kritik und Anregungen der Leser werden dankend entgegengenommen.

Für die parallele Einführung in das Statistik-Programm SPSS für Windows existiert ein separates Skript, das wie dieses Skript und sein Anhang auf meiner Webseite zum Herunterladen bereit steht.

E. Swart

im Januar 2008

1. Epidemiologie – zur Begriffsbestimmung

„Epidemiologie ist die Untersuchung der Verteilung und Determinanten von gesundheitsbezogenen Zuständen oder Ereignissen in umschriebenen Bevölkerungsgruppen und die Anwendung der Ergebnisse zur Steuerung von Gesundheitsproblemen.“

(Gordis 2001).

Epidemiologische Ergebnisse finden Anwendung

- ◆ in der Ursachenforschung,
- ◆ in der Analyse des naturgesetzlichen Krankheitsverlaufs,
- ◆ bei der Bewertung und Verbesserung der Gesundheitsdienste und
- ◆ in der (individualmedizinischen) Beurteilung von Risiken, Chancen und Prognosen.

Die Ziele der Epidemiologie im einzelnen sind:

1. Das Erkennen der Ätiologie oder der Ursache(n) einer Erkrankung sowie möglicher Risikofaktoren, also Einflussgrößen, die das Erkrankungsrisiko einer Person erhöhen,
2. Untersuchung des natürlichen Verlaufs und die Prognose von Krankheiten,
3. Bestimmung des Ausmaßes von Krankheiten in der Bevölkerung,
4. Bewertung neuer präventiver und therapeutischer Maßnahmen sowie von Änderungen in der medizinischen Versorgung (Evaluation),
5. Schaffung von Grundlagen für die Entwicklung der Gesundheitspolitik.

Epidemiologie ist nicht primär die Methodenlehre für klinische Forschung, obwohl es zahlreiche Verbindungen gibt (vgl. auch Kap 11). Zu den gesundheitsrelevanten Problemen und damit zu den erkenntnisrelevanten epidemiologischen Themen gehören die **6 D** (death, disease, discomfort, disability, dissatisfaction, destitution), also (vorzeitiger) **Tod, Krankheit** und **subjektive Krankheitserfahrung, Beschwerden** (wie Schmerz, Übelkeit, Atemnot usw.), **Verlust von Funktionen und sozialer Partizipation, Unzufriedenheit** (emotionale Reaktionen auf Krankheit und Versorgung) und **"Verelendung"**, d.h. Verlust von Geld oder sozialem Status durch Krankheit.

Zu den Determinanten dieser 6 D gehören biologische (neuerdings wesentlich genetische), soziale (einschl. der Funktion des Gesundheitswesens) und umweltbezogene Faktoren. Von der klinischen Individualmedizin unterscheidet sich die Epidemiologie vor allem durch ihren Bevölkerungsbezug (§1 der ärztlichen Berufsordnung: "Der Arzt dient der Gesundheit des Einzelnen und des ganzen Volkes.") und durch eine konsequente Berücksichtigung prä- und postklinischer Gesundheitsprobleme.

Die folgende Abbildung 1.1 verdeutlicht die ersten vier der genannten Ziele der Epidemiologie:

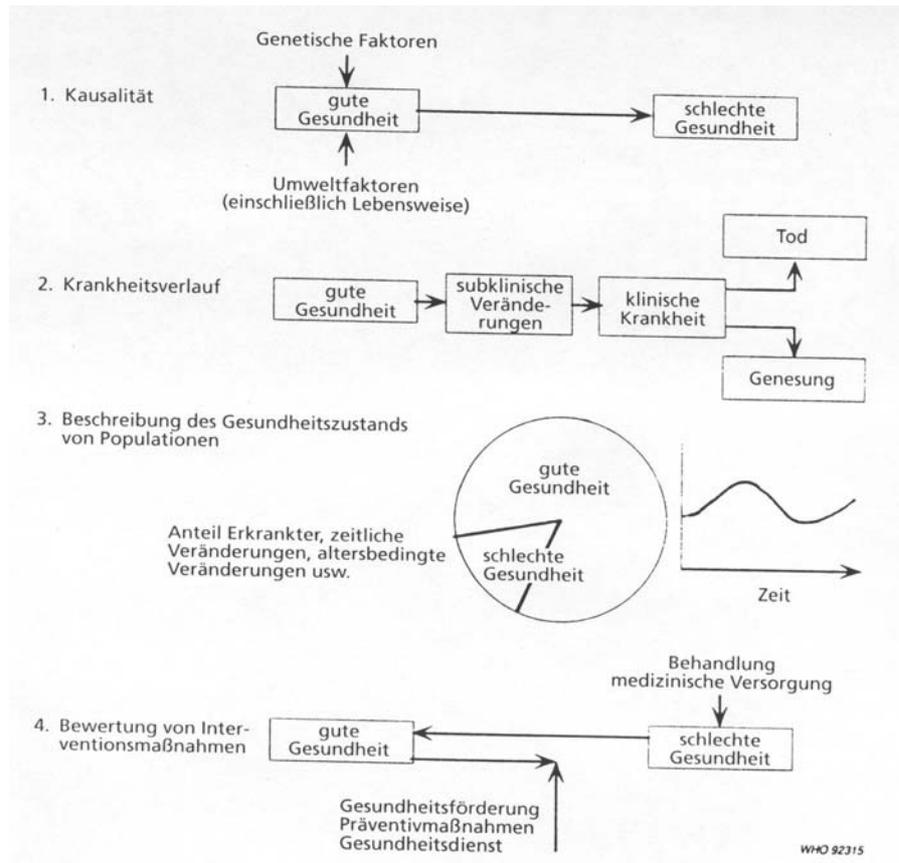


Abbildung 1.1: Anwendungsgebiete der Epidemiologie

(Beaglehole et al. 1997)

Herkömmlich unterscheidet man zwischen **deskriptiver Epidemiologie**, die die Verteilung von Krankheiten und physiologischen Variablen in einer definierten Bevölkerung beschreibt, **analytischer Epidemiologie**, die aufgrund gezielter Hypothesen Zusammenhänge und Determinanten von Krankheiten untersucht, und **experimenteller Epidemiologie**, die in Verbindung mit statistischer Versuchsplanung die Effektivität von Interventionsmaßnahmen (wie Prävention oder Versorgung) zu erforschen sucht. Besonders hier liegt der Anknüpfungspunkt zur patientenbezogenen klinischen Forschung, der Unterschied liegt in der epidemiologischen Orientierung auf die ('frei lebende') Bevölkerung und die Funktion der flächendeckenden medizinischen Versorgung.

2. Kurze Anmerkungen zur Geschichte der Epidemiologie

2.1 John Snow und die Cholera in London

Die Epidemiologie ist - wie andere Wissenschaften auch - durch bahnbrechende Untersuchungen herausragender Persönlichkeiten vorangetrieben worden, die auch in großem zeitlichen Abstand noch Beachtung verdienen. An erster Stelle ist hier John Snow zu nennen, der durch eine Art ‚Barfuß-Epidemiologie‘ den Ursachen der beiden großen Cholera-Epidemien in London Mitte des 19. Jahrhundert nachging. Durch die Aufzeichnung der Lage jedes Haushaltes mit einem Cholera-Toten entdeckte er eine augenfällige Assoziation zwischen der Quelle des Trinkwassers und der Zahl der Todesfälle. Er verglich die Todesfälle und die Mortalitätsraten pro 1.000 Einwohner verschiedener Londoner Distrikte und entdeckte dabei eine Häufung der Todesfälle im Bereich zweier Wasserversorgungsgesellschaften (Tabelle 2.1).

Tabelle 2.1: Bevölkerung, Cholera-Todesfälle (absolut und pro 1000 Einwohner) nach Versorgungsgebieten Londoner Wasserwerke während der Cholera-Epidemie, 1849

Water Company	Population	Deaths	Deaths/10,000 population
Southwark&Vauxhall	338,820	4,668	138
Lambeth	116,691	1,479	125
Kent	136,857	990	72
East London	443,915	2,312	52
Chelsea	192,236	876	46
New River	652,865	2,432	37
Hampstead	59,160	123	21
West Middlesex	249,950	473	19
Grand Junction	129,506	177	14

Quelle: Snow 1854 (<http://www.ph.ucla.edu/epi/snow.html>)

Weitere Beobachtungen erhärteten seinen Verdacht, dass kontaminiertes Wasser als Hauptursache für die Epidemie betrachtet werden muss. Dazu gehörte der Tatbestand, dass bei der zweiten Epidemie wenige Jahre später die Mortalitätsrate von Lambeth nach Verlagerung des Entnahmepunkte themseaufwärts nur noch 38 pro 10.000 Einwohner betrug und geringer war als in anderen Bezirken Londons (Tabelle 2.2).

Tabelle 2.2: Cholera-Todesfälle nach Versorgungsgebieten Londoner Wasserwerke während der Cholera-Epidemie, 1854

Water Company	Anzahl der Haushalte	Todesfälle	Todesfälle je 10.000 Haushalte
Southwark & Vauxhall	40.046	1.263	1.422
Lambeth	26.107	98	37
übriges London	256.423	1.422	59

Quelle: Snow 1854 (<http://www.ph.ucla.edu/epi/snow.html>)

Eine punktgenaue Aufzeichnung der Haushalte mit Cholera-Todesfällen während der zweiten Epidemie konnte als wesentliche Einzelquelle eine Wasserpumpe an der Broad Street identifizieren (Abbildung 2.1).



Abbildung 2.1: Karte John Snows mit der punktgenauen Einzeichnung der Cholera-Todesfälle

Weitere Bausteine lieferten Einzelbeobachtungen, etwa dass Arbeiter einer Brauerei kaum unter den Todesfällen zu beobachten waren, da sie besonders gereinigtes Wasser für ihr Bier verwendeten und dieses selbst in erheblichen Mengen zu sich nahmen. Oder der Todesfall einer Witwe, die als einzige ihres Blocks an Cholera starb. Der Grund dafür lag in der Tatsache, dass sie nach dem Tod ihres Mannes und einem nachfolgendem Umzug aus alter Gewohnheit weiter Wasser von der Pumpe an der Broad Street bezog, in deren unmittelbaren Umfeld mehrere hundert Personen an Cholera verstarben.

2.2 Semmelweis und das Kindbettfieber

Semmelweis war Assistenzarzt an der 1. Gebärklinik in Wien. Während seiner Dienstzeiten dort fiel ihm die hohe Rate an Todesfällen durch Kindbettfieber in der Abteilung für angehende Geburtshelfer (Ärzte) im Vergleich zur Abteilung für Hebammenschülerinnen auf (Tabelle 2.3a). Erklärungsversuche anderer Ärzte für diese Diskrepanz (z.B. Verletzung des Schamgefühls der Schwangeren bei der Untersuchung durch männliche Ärzte oder der rohe Umgang von ausländischen Studenten mit den Schwangeren) konnten die Unterschiede nicht erklären. Auffällig war auch der Unterschied in der Mortalität der Neugeborenen zwischen beiden Abteilungen oder die Tatsache, dass spontane, sogenannte Gassengeburt, bei denen keine gynäkologischen Untersuchungen mehr stattfanden, eine niedrige Sterblichkeit aufwiesen (Tabelle 2.3b).

Tabelle 2.3a: Todesfälle an Kindbettfieber in den Abteilungen der 1. Gebärklinik Wiens, 1841 bis 1846

Jahr	Abteilung für Ärzte			Abteilung für Hebammen		
	Anzahl der Entbundenen	Anzahl der Verstorbenen	Die Anzahl der Entbundenen verhält sich zur Anzahl der Verstorbenen wie 100 zu	Anzahl der Entbundenen	Anzahl der Verstorbenen	Die Anzahl der Entbundenen verhält sich zur Anzahl der Verstorbenen wie 100 zu
1841	3036	237	7.7	2442	86	3.5
1842	3287	518	15.8	2659	202	7.5
1843	3060	274	8.9	2739	164	5.9
1844	3257	260	8.2	2956	68	2.3
1845	3492	241	6.8	3241	66	2.03
1846	4010	459	11.4	3754	105	2.7
Summa	20042	1989	9.92	17791	691	3.38

Quelle: Semmelweis 1861, Tabelle Nr. I

Tabelle 2.3b: Säuglingssterblichkeit in den Abteilungen der 1. Gebärklinik Wiens, 1841 bis 1846

Jahr	I. Gebärklinik						II. Gebärklinik					
1841	Kinder	2813,	gest.	177,	Proz.	6.2,	Kinder	2252,	gest.	91,	Proz.	4.4
1842	„	3037,	„	279,	„	9.1,	„	2414,	„	118,	„	4.06
1843	„	2828,	„	195	„	6.8,	„	2570,	„	130,	„	5.05
1844	„	2917,	„	251	„	8.6,	„	2739,	„	100,	„	3.06
1845	„	3201,	„	260	„	8.1,	„	3017,	„	97,	„	3.2
1846	„	3533,	„	235	„	6.5,	„	3398,	„	86,	„	2.05

Quelle: Semmelweis 1861; Tabelle Nr. XIV

Auf die Spur des Erregers des Kindbettfiebers kam Semmelweiss durch den Tod eines Anatoms, der nach einer Verletzung durch ein Seziernmesser bei einer Sektion an den Symptomen des Kindbettfiebers verstarb. Außerdem fiel Semmelweiss auf, dass die angehenden Geburtshelfer ihre Untersuchungen häufig nach anatomischen Untersuchungen vornahmen, sich ihre Hände vorher aber lediglich mit Wasser und Seife reinigten.

Nachdem Semmelweiss vor der gynäkologischen Untersuchung Waschungen mit Chlorkalk eingeführt hatte (Juni 1847), sank die Kindbettsterblichkeit umgehend auf das Niveau der anderen Abteilung. Schon vorher war für wenige Monate (Dezember 1846 bis März 1847) die Sterblichkeit erheblich gesunken, als zwischenzeitlich von der Gewohnheit des morgendlichen Sezierens abgewichen wurde (Tabelle 2.4).

Tabelle 2.4: Sterblichkeit in der ersten Abteilung der Wiener Gebärklinik, 1846 bis 1847

Monat	Geburten	Todesfälle	Prozent	Geburten	Todesfälle	Prozent
1846				1847		
Januar	336	45	13,39	311	10	3,21
Februar	293	53	18,08	912	6	1,92
März	311	48	15,43	305	11	3,60
April	253	48	18,97	312	57	18,27
Mai	305	41	13,44	294	36	12,24
Juni	266	27	10,15	268	6	2,38
Juli	252	33	13,10	250	3	1,20
August	216	39	18,05	264	5	1,89
September	271	39	14,39	262	12	5,23
Oktober	254	38	13,98	278	11	3,95
November	297	32	10,77	246	11	4,47
Dezember	298	16	5,37	273	8	2,93

Quelle: Semmelweis, 1861

Beide Wegbereiter der Epidemiologie – Snow und Semmelweiss - haben sich damals schon grundlegender Techniken und Vorgehensweisen der Epidemiologie bemächtigt, die auch heute noch gültig sind:

- umfassende und sorgfältige Auswertung vorhandenen Datenmaterials,
- Prüfung, ob sich die Daten mit bisherigen Erklärungen für die beobachteten Phänomene in Einklang bringen lassen,
- Auseinandersetzung mit Erklärungsversuchen zu den Ursachen der Krankheiten und ihre systematische Widerlegung durch die empirischen Daten,
- gemeinsame Betrachtung klinischer Auffälligkeiten bei besonders gelagerten Einzelfällen und von Massendaten zur Aufstellung einer neuen ätiologischen Theorie,
- Prüfung, ob diese neue Theorie besser als die bisherigen mit den empirischen Daten zu vereinbaren ist,
- Ergreifung von Gegenmaßnahmen in Übereinstimmung mit diesen Theorien und die Evaluierung ihrer Wirkung durch Analyse neuer Daten,
- Heranziehung weiterer Datenquellen zur Überprüfung der neuen Theorie.

2.3 Die Entwicklung der Epidemiologie in Deutschland

Die Entwicklung der Epidemiologie im 19. Jahrhundert in Deutschland ist eng mit der Entwicklung der Sozialmedizin verbunden. Hier sind an herausragender Stelle Salomon Neumann und Rudolf Virchow zu nennen. Beide namhaften Ärzte kamen durch fundiertes Studium amtlicher Statistiken (Neumann) und durch empirische Studien zur ‚sozialen Not‘ in Oberschlesien und im Spessart (Virchow) zu der Erkenntnis, dass ein Großteil der damals vorherrschenden und vor allem die unteren sozialen Schichten betreffenden Krankheiten wie Typhus, Cholera, Tuberkulose in hohem Maße sozial bedingt sind und denen demzufolge vor allem durch Maßnahmen zur Verbesserung der sozialen Lage der ärmeren Bevölkerung begegnet werden kann.

Harte Zahlen liegen beispielsweise aus Hamburg für die Cholera-Epidemie 1892 vor. Danach gibt es einen eindeutigen Zusammenhang zwischen der Sozialschicht, ausgedrückt als durchschnittliches Einkommen in den Hamburger Stadtteilen oder der individuellen Steuerklasse, und der Sterblichkeit an Cholera (Tabellen 2.5a und 2.5b).

Tabelle 2.5a: Cholera-Epidemie in Hamburg 1892. I: Mortalität nach Stadtteilen

Stadtteil	durchschnittl. Pro-Kopf-Einkommen	Todesfälle pro 1000 Einwohner
Harvestehude	3156 M.	4,5
Rothenbaum	2311 M.	5,9
Hohenfelde	1321 M.	8,5
Uhlenhorst	1196 M.	14,8
Hamm	717 M.	16,8
Altstadt Süd	688 M.	14,1
Altstadt Nord	645 M.	17,5
St. Georg Süd	654 M.	16,2
St. Pauli	422 M.	12,6
Barmbek	336 M.	15,7
Neustadt Süd	332 M.	19,9
Billwerder.Ausschlag	271 M.	24,2

Quelle: Krasemann et. al. (1987)

Tabelle 2.5b: Cholera-Epidemie in Hamburg 1892. II: Erkrankungsfälle und Letalität nach Einkommen

Steuerliches Einkommen 1891	Erkrankungsfälle	darunter verstorben (%)
800 bis 1.000 M.	28.647	62
1.000 bis 2.000 M.	43.848	55
2.000 bis 3.500 M.	14.848	27
3.500 bis 5.000 M.	6.125	22
5.000 bis 10.000 M.	5.649	16
10.000 bis 25.000 M.	3.328	10
25.000 bis 50.000 M.	1.192	11
mehr als 50.000 M.	834	5

Quelle: Krasemann et. al. (1987)

Bekannt ist in diesem Zusammenhang der Ausspruch Neumanns „Die medizinische Wissenschaft ist in ihrem innersten Kern und Wesen eine soziale Wissenschaft, und solange ihr diese Bedeutung in der Wirklichkeit nicht indiziert sein wird, wird man auch die Früchte nicht genießen, sondern sich mit der Schale und dem Schein begnügen müssen. Die soziale Natur der Heilkunst steht über allem Zweifel“ (zitiert nach Blohmke et al. 1975). Im übrigen sei zur Geschichte der Sozialmedizin auf die entsprechenden Lehrbücher verwiesen.

Wegweisend für die (Sozial-)Epidemiologie und kennzeichnend für die Sozialmedizin in der 2. Hälfte des 19. und zu Beginn des 20. Jahrhundert, im Kaiserreich bis zur Weimarer Republik war das 1913 erschienene Werk von Mosse und Tugendreich ‚Krankheit und soziale Lage‘. Darin wird anhand amtlicher Statistiken ausführlich für verschiedene Krankheitsbilder und soziale Gruppen der Einfluss der sozialen Lage auf Gesundheit und Krankheit untersucht.

Durchgehend wird dabei das erheblich höhere Morbiditäts- und Mortalitätsrisiko sozial schwächerer Gruppen gegenüber Angehörigen der Mittel- und Oberschicht (ohne diese neueren Termini explizit zu verwenden) nachgewiesen. „Der Nachweis der engen Beziehungen zwischen Armut und Krankheit bedeutet eine schwere Anklage gegen die Kultur, gegen die Gesellschaft. Die Armut verurteilt den größten Teil der zivilisierten Menschheit zu einer unhygienischen Lebensweise mit ihren tödlichen Folgen. ... Diese Erkenntnis legt der Gesellschaft große und ernste Pflichten auf. Das Ziel ist, auch dem Armen ein hygienisch befriedigendes Dasein zu schaffen und so die gewaltigen Unterschiede auszugleichen, welche die Lebenserwartung der Armen und Reichen trennen.“

Die Sozialmedizin litt bis in die 60er Jahre an den Auswirkungen der Nazi Herrschaft und deren missbräuchlicher Verwendung des Begriffs Sozialhygiene. Erst mit Gründung der Deutschen Gesellschaft für Sozialmedizin im Jahr 1965 und der Errichtung der ersten Lehrstühle für Sozialmedizin in Heidelberg (Prof. H. Schaefer) und Hannover (Prof. M. Pflanz) begann

sich das Fach und damit ihre Kerndisziplin Epidemiologie wieder zu etablieren. Eine der ersten auch international beachteten deutschen Studien veröffentlichte Pflanz 1971, in der er ausführlich die Epidemiologie der Appendektomie und die im internationalen Vergleich hohe Eingriffshäufigkeit und Mortalität untersuchte. Dabei traten auch erhebliche Unterschiede in der Inzidenz nach Berufsstatus und bezüglich verschiedener Stadteile Hannovers zutage.

In diese Zeit fiel die Gründung des ersten deutschen bevölkerungsbezogenen Krebsregisters im Saarland. 1980 folgte mit dem Kinderkrebsregister Mainz das erste bundesweite Register, dessen epidemiologische Arbeit durch zahlreiche epidemiologische Studien zu möglichen Risiken durch Atomkraftwerke und Sondermülldeponien Beachtung fand.

Traditionelle Anwendungsgebiete der Epidemiologie sind neben der Krebs Epidemiologie die Herz-Kreislauf-, die Umweltepidemiologie und die Epidemiologie arbeitsbedingter Gesundheitsstörungen. Hier haben auch die großen epidemiologischen Studien nach dem 2. Weltkrieg im Ausland und mit Verzögerung in Deutschland ihren Ausgangspunkt. Die bislang größte Studie aus Deutschland in den letzten Jahren ist die Herz-Kreislauf-Präventionsstudie (DHP), die die Möglichkeiten gemeindebezogener Präventionsmaßnahmen untersuchte und in deren Verlauf in den 80er und frühen 90er-Jahren drei große repräsentative Befragungen zum Gesundheitszustand der deutschen Allgemeinbevölkerung stattfanden.

In den letzten 20 Jahren hat sich die Epidemiologie zahlreiche neue Gebiete erschlossen. Dazu gehören die Infektions-, Ernährungs- und genetische Epidemiologie oder die Versorgungs- und Evaluationsforschung. Durch die Etablierung von Public Health in Deutschland, zu dessen zentralen Methoden auch die Epidemiologie zählt, haben sich viele Anknüpfungspunkte an andere Disziplinen ergeben. Zu den aktuellen Fortschritten der Epidemiologie in Deutschland gehört die Herausgabe der ersten Gesundheitsberichte für Deutschland, die Differenzierung in Schwerpunkt- und Einzelthemen (beziehbar über www.gbe-bund.de und www.rki.de) und die Durchführung eines Bundesgesundheitsurvey (BGS) für Erwachsene im Alter zwischen 18 und 79 Jahren, die Folgeuntersuchung anhand einer telefonischen Befragung und die Ausweitung auf einen Kinder- und Jugendsurvey (2004 bis 2006), dessen erste Ergebnisse im Frühsommer 2007 publiziert wurden (Näheres unter www.kiggs.de). Die Bereitstellung der Daten dieser repräsentativen Bevölkerungsbefragungen als Public-Use-Files für die wissenschaftliche Forschung ist bereits erfolgt bzw. für die neueren Surveys vorgesehen

Nachdem Mitte der 90er Jahre des letzten Jahrhundert sich zunächst die Deutsche Arbeitsgemeinschaft für Epidemiologie (dae) als erste Fachgesellschaft der Epidemiologen in Deutschland gegründet hatte, allerdings noch unter dem Dach dreier Muttergesellschaften, verselbständigte sich die dae als eigenständige wissenschaftliche Gesellschaft seit dem 1. Januar 2006 unter dem Namen Deutschen Gesellschaft für Epidemiologie (DGEpi). Auf der Homepage der Gesellschaft (www.dgepi.de) sind zahlreiche Grundsatzpapiere zu methodischen und wissenschaftspolitische Aspekten zu finden (z.B. die Gute Praxis Epidemiologie; vgl. Kapitel 13)

Vertiefende Quellen zu diesen Bereichen am Ende dieses Skripts unter weiterführender Literatur.

3. Grundlagen der Statistik

Die folgenden Seiten entstammen einer früheren Lehrveranstaltung und wurden gekürzt in dieses Skript übernommen, um die wichtigsten Begriffe der Statistik in Erinnerung zu rufen. In der Veranstaltung selbst wird aus Zeitmangel dieses Kapitel nicht behandelt. Es wird davon ausgegangen, dass die wesentlichen Begriffe in anderen Lehrveranstaltungen besprochen wurden. Für weiterführende Überlegungen sei im übrigen auf die im Literaturverzeichnis aufgeführten Lehrbücher verwiesen.

3.1 Deskriptive Statistik

Die **deskriptive Statistik** dient der Strukturierung, Beschreibung und graphischen Darstellung unübersichtlichen und umfangreichen Datenmaterials.

Die befragten Personen bzw. die Objekte, an denen Messungen vorgenommen werden, nennt man **Untersuchungseinheiten**; die Größen, auf die sich Fragen oder Messungen beziehen, **Merkmale** oder **Variablen**.

Bei einer Erhebung wird an n Untersuchungseinheiten ein Merkmal X beobachtet. Jede Untersuchungseinheit erhält auf diese Weise einen Merkmalswert x_j . Die Urliste der n verschiedenen Merkmalswerte ist in der Regel sehr unübersichtlich. Man benötigt daher Maßzahlen zur knappen Beschreibung der Verteilung der einzelnen Merkmalswerte und graphische Verfahren zu ihrer anschaulichen Darstellung. Diese sind abhängig von der **Skalierung** der zu beschreibenden Merkmale.

Die einfachste Darstellung erfolgt durch eine **Häufigkeitstabelle**, in der die verschiedenen Merkmalsausprägungen und ihre absoluten bzw. relativen Häufigkeiten abgetragen werden. Die relativen Häufigkeiten, meistens in Prozentwerten dargestellt, erhält man, in dem man die Zahl der Beobachtungen mit einem bestimmten Merkmalswert durch die Gesamtzahl aller Beobachtungen, also n , dividiert (Tabelle 3.1).

Tabelle 3.1: Beispiel - Entlassungsanlässe bei Patienten einer chirurgischen Klinik

	absolute Häufigkeit	relative Häufigkeit (%)
regulär entlassen	358	84,2
intern verlegt	17	4,0
in anderes Krankenhaus verlegt	39	9,2
verstorben	8	1,9
auf eigenen Wunsch entlassen	3	0,7
gesamt	425	100,0

Zwei Möglichkeiten der graphischen Darstellung der in dieser Tabelle ausgeführten Ergebnisse bieten sich durch **Kreisdiagramme** - dort werden die Häufigkeiten durch Flächen dargestellt (vgl. Abbildung 3.1) – oder **Säulendiagramme**.

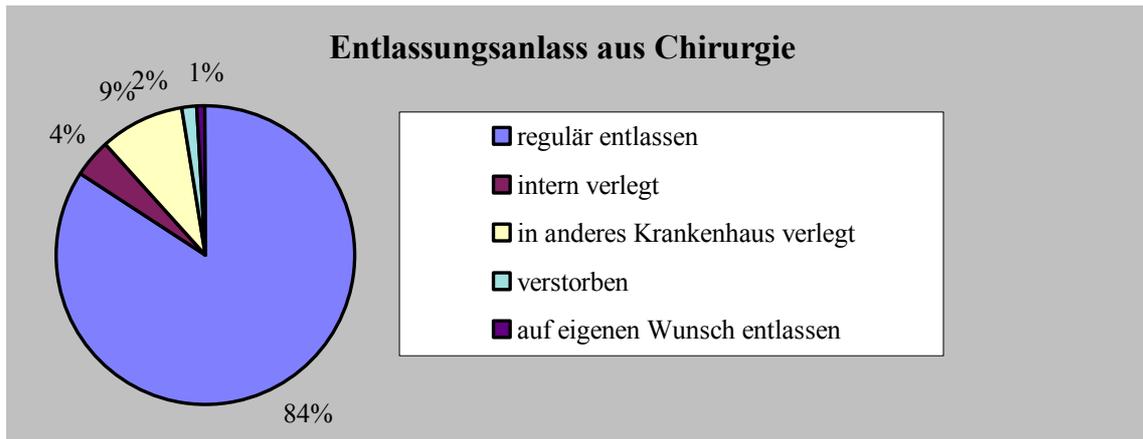


Abbildung 3.1: Beispiel für ein Kreisdiagramm

Da es sich beim Entlassungsanlass um ein **nominales Merkmal** handelt, lässt sich als zentraler Lageparameter nur der **Modus** als **häufigster Wert** (hier: reguläre Entlassung) angeben. Für die Beschreibung der Streuung in den Merkmalswerten kommt lediglich die Beschreibung der Gesamtverteilung in Frage, in unserem Beispiel also die geringe Häufigkeit aller anderen Entlassungsanlässe.

Zeitliche Verläufe lassen sich auch als Liniendiagramm darstellen (Abbildung 3.2).

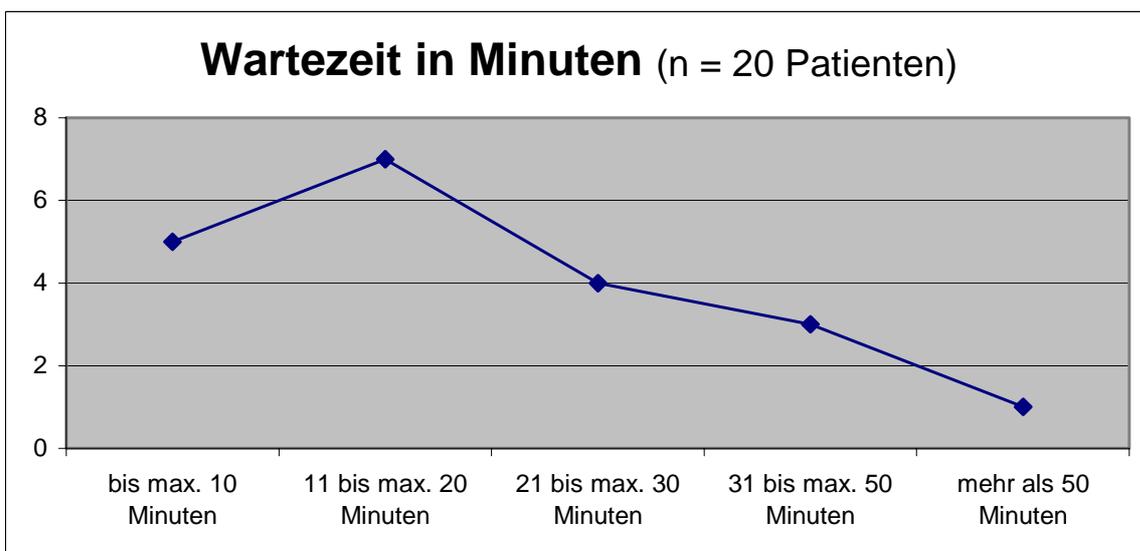


Abbildung 3.2: Beispiel für ein Liniendiagramm

3.2 Lage- und Streuungsparameter von Verteilungen

Die Wahl geeigneter Lage- und Streuungsparameter zur Beschreibung der empirischen Verteilung einer Variablen und später in der analytischen Statistik die Auswahl angemessener Schätz- und Testverfahren hängt vom **Messniveau** der zu untersuchenden Merkmale ab. Folgende Datentypen lassen sich unterscheiden:

Nominale Merkmale: Es gibt mehrere sich gegenseitig ausschließende Ausprägungen, die zusätzlich keine inhärente Ordnung aufweisen, wie Geschlecht, Farbe, Beruf.

Ordinale Merkmale: Im Gegensatz zu den nominalen Merkmalen kommt eine natürliche Ordnung hinzu. Beispiele: Schulnoten, soziale Klasse, Zufriedenheit auf einer mehrteiligen

Skala von sehr unzufrieden bis sehr zufrieden. Die Abstände zwischen den Ausprägungen können aber nicht quantifiziert werden. Der Leistungsunterschied zweier Schüler mit den Noten 3 und 4 bzw. 2 und 3 kann nicht gleich gesetzt werden.

Beide Merkmalstypen sind **qualitativ**. Bei den **quantitativen** Merkmalen lassen sich diskrete und stetige unterscheiden.

Diskrete Merkmale weisen nur eine begrenzte Zahl von möglichen Ausprägungen auf, etwa Alter, die Zahl der Schwangerschaften oder die Zahl von Nebendiagnosen bei stationär aufgenommenen Patienten. Der Abstand zwischen den Ausprägungen lässt sich quantitativ bewerten. Ein 50-Jähriger ist 10 Jahre älter ein 40-Jähriger und dieser Altersunterschied ist genauso groß wie zwischen einer 36-Jährigen und einer 26-Jährigen.

Bei **stetigen Merkmalen** ist die Zahl der Ausprägungen unbegrenzt, sie können zumindest theoretisch mit beliebiger Genauigkeit gemessen werden (es gibt keine ‚Lücke‘ zwischen den möglichen Werten), wie Körpergröße oder Gewicht. Das erlaubt sämtliche mathematischen Transformationen, was sich die statistische Test- und Schätztheorie zunutze macht.

Der Informationsgehalt und die Möglichkeiten der deskriptiven und analytischen Statistik nehmen kontinuierlich von nominalen bis stetigen Merkmalen zu.

Tabelle 3.2: Übersicht über die Merkmalstypen

		Quantitativ	
		stetig	diskret
		Blutdruck, Größe, Gewicht,	Alter, Zahl der Kinder, Asthmaattacken pro Woche
		Qualitativ	
		ordinal	nominal
		Schulnoten, Tumorstatus (I-IV) Zufriedenheit mit ärztlicher Behandlung	Geschlecht, Blutgruppe, 5-Jahrestüberleben nach Krebsoperation (ja/nein)

Bei **ordinalen Merkmalen** wie den Schulnoten, wo sich die Merkmalswerte zwar in eine Rangfolge bringen lassen, die Abstände zwischen den Merkmalswerten aber unterschiedlich oder gar nicht zu quantifizieren sind, kann neben dem Modus der **Median** als **mittlerer Wert** der der Größe nach geordneten Werte als Lagemaß herangezogen werden.

Der Median ist demnach bei einer ungeraden Anzahl von Merkmalswerten der mittlere Wert, bei einer geraden Anzahl von Beobachtungen liegt er in der Mitte der beiden mittleren Beobachtungen. Der Median hat gegenüber dem arithmetischen Mittel (s.u.) den Vorteil, dass er weniger empfindlich gegenüber Ausreißern ist. So würde sich in einer Reihe von 30 Verweildauern der Median sich nicht ändern, unabhängig davon, ob das Maximum bei 14, bei 50 oder gar 500 läge.

Diskrete oder Stetige Merkmale erlauben eine differenziertere Darstellung der Ergebnisse. Zur graphischen Präsentation müssen allerdings die Rohwerte geeignet in Klassen zusammengefasst werden, da bei einer großen Anzahl verschiedener Merkmalswerte die Darstellung einzelner relativer oder absoluter Häufigkeiten nicht sinnvoll ist.

Einem **Histogramm** liegt eine Klasseneinteilung der Rohwerte zugrunde. Dabei werden die Klassen auf der x-Achse, die absoluten Häufigkeiten auf der y-Achse, abgetragen. Die beobachteten Klassenhäufigkeiten werden dabei als aneinanderstoßende Rechtecke dargestellt, deren **Flächen**, nicht Höhe, proportional den Häufigkeiten sind.

In der folgenden Abbildung 3.3 sind Krankenhausverweildauern von 330 Fällen mit einer Nasenscheidewandverbiegung graphisch dargestellt. Die Klassenbreiten betragen vier Tage,

angegeben unter den Flächen ist jeweils die Klassenmitte. Die erste Klasse umfasst das Intervall von 0 bis unter 4 Tagen, das zweite von 4 bis unter 8 usw. bis zur letzten Klasse im Intervall von 24 bis unter 28 Tage.

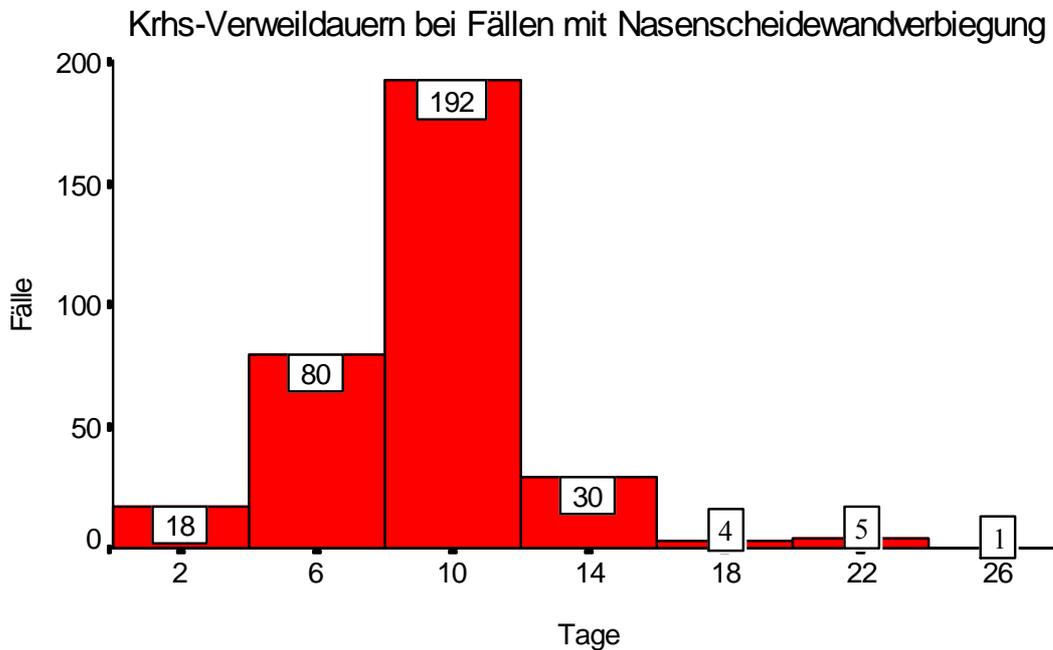


Abbildung 3.3: Beispiel für ein Histogramm

Die Aufstellung einer **Häufigkeitstabelle** macht bei diskreten oder stetigen Verteilungen nur Sinn, wenn die Zahl der tatsächlich realisierten Merkmalswerte überschaubar bleibt. Dann lassen sich aus der Häufigkeitstabelle eine Reihe von Verteilungsparametern ablesen.

Dazu muss vorher der Begriff der Summenhäufigkeit erläutert werden. Die (absolute oder relative) Summenhäufigkeit des Merkmalswerts x_i ist die Summe der (absoluten oder relativen) Häufigkeiten aller Merkmalswerte, die kleiner oder gleich dem Merkmalswert x_i sind.

In einer Häufigkeitstabelle der bekannten Art ergibt sich die Summenhäufigkeit, in dem die (absoluten oder relativen) Häufigkeiten der der Größe nach geordneten Merkmalswerte sukzessive aufsummiert werden. Aus der Häufigkeitstabelle können nun sofort der Modus und der Median abgelesen werden; der Median ist der Merkmalswert, bei dem die relative Summenhäufigkeit erstmalig die 50 %-Grenze überschreitet.

Ähnlich wie der Median können aus der Tabelle das untere und das obere Quartil als wichtige Lageparameter abgelesen werden. Das **untere Quartil** gibt die Obergrenze der 25 % kleinsten Merkmalswerte an, das **obere Quartil** entsprechend die untere Grenze der 25 % größten Merkmalswerte. Die Quartilswerte erhält man, in dem in der Tabelle die Werte ermittelt werden, bei denen die relative Summenhäufigkeit erstmalig den Wert 25 % bzw. 75 % überschreitet. Zwischen diesen beiden Werte liegen dann die mittleren 50 % aller Merkmalswerte (**Interquartilsabstand**).

Eine feinere Beschreibung des Datenmaterials ergibt sich, wenn man entsprechend geeignete **Perzentile** der Verteilung angibt. Üblicherweise sind die 5%-, 10%-, 90%- und 95%-Perzentile einer Verteilung von Interesse. Beispielsweise ist das 5%-Perzentil die obere Grenze für die 5% kleinsten Werte, das 90%-Perzentil die untere Grenze der 10% größten Werte.

Tabelle 3.3: Krankenhausverweildauern von 143 Fällen mit Hauptdiagnose Leistenhernie

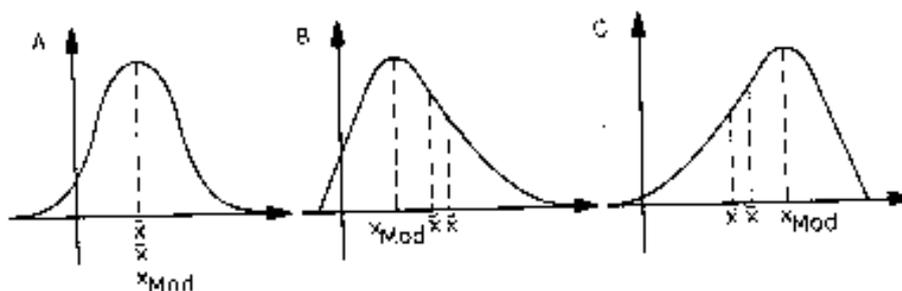
Tag	Fälle	rel. Häufigkeit (%)	relative Summenhäufigkeit	rel. Summenhäufigkeit (%)
1	2	1,4	2	1,4
2	4	2,8	6	4,1
3	12	8,3	18	12,4
4	8	5,5	26	17,9
5	2	1,4	28	19,3
6	8	5,5	36	24,8
7	19	13,1	55	37,9
8	22	15,2	77	53,1
9	25	17,2	102	70,3
10	13	9,0	115	79,3
11	7	4,8	122	84,1
12	9	6,2	131	90,3
13	4	2,8	135	93,1
14	4	2,8	139	95,9
15	3	2,1	142	97,9
17	2	1,4	144	99,3
36	1	0,7	143	100,0

Diskrete oder stetige Merkmale erlauben auch die Berechnung des **arithmetischen Mittelwerts**. Dieser ergibt sich, indem sämtliche Merkmalswerte aufsummiert und anschließend durch die Zahl aller Beobachtungen geteilt werden.

$$\bar{x} = \frac{1}{n} \sum x_i \quad (3:1)$$

Der Mittelwert ist anfällig gegenüber Ausreißern nach oben oder nach unten.

Häufigkeitserteilungen, die durch eine Konzentration der Merkmalswerte an einem Ende und wenige, aber auffällige Ausreißer am anderen Ende charakterisiert sind, werden als **schief (asymmetrisch)** bezeichnet. Der Vergleich der Parameter Modus, Median und Mittelwert erlaubt Aussagen über die Schiefe einer Verteilung. Verteilungen sind **rechtsschief** bei einer Häufung (=Schwerpunkt) von niedrigen Werten und Ausreißern hauptsächlich nach oben (Beispiel: Krankenhausverweildauer). Bei **linksschiefen** Verteilungen ist die Situation umgekehrt (Beispiel: Eisenwerte im Blut). Symmetrische Verteilungen haben eine Häufung der Werte in der Mitte des möglichen Merkmalbereichs und einen Abfall der Häufigkeiten gleichmäßig nach beiden Seiten. Bei schiefen Verteilungen sind die Mittelwerte stark von den Ausreißern beeinflusst.

**Abbildung 3.5: Lagebeziehungen zwischen Modalwert, Median und Mittelwert:**

A: symmetrische Verteilung (Modus \approx Median \approx Mittelwert); B: rechtsschiefe Verteilung Modus < Median < Mittelwert
C: linksschiefe Verteilung Modus < Median < Mittelwert

Zur Beschreibung der **Streuung** der Merkmalswerte haben wir oben bereits den Interquartilsabstand kennen gelernt. Ein weiteres Streuungsmaß ist die **Spannweite**, definiert als Differenz zwischen dem größten (**Maximum**) und dem kleinsten Wert (**Minimum**). Es liegt auf der Hand, dass die Spannweite stark von Ausreißern abhängt.

Das am häufigsten gebrauchte Streuungsmaß ist die **Varianz** der Merkmalswerte. Sie ist definiert als die Summe der quadrierten Abweichungen vom arithmetischen Mittelwert, dividiert durch die um Eins verringerte Zahl der Beobachtungen:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (3:2)$$

Da die Varianz aufgrund der Quadratur der Abweichungen nicht die gleiche Dimension wie die Merkmalswerte besitzt, wird häufig auch die **Standardabweichung** der Merkmalswerte angegeben. Sie ist einfach die Wurzel aus der Varianz und besitzt die gleiche Dimension wie die Beobachtungen.

Diese Eigenschaft erlaubt die Standardabweichung zum Mittelwert in Beziehung zu setzen. Man erhält den (dimensionslosen) **Variationskoeffizienten** durch Division der Standardabweichung durch den Mittelwert

$$CV = \frac{s}{\bar{x}} \quad (3:3)$$

Dieser erlaubt den Vergleich verschiedener Beobachtungsreihen.

3.3 Die wichtigsten Verteilungen der Statistik

3.3.1 Normalverteilung

Die **Normalverteilung** ist die mit Abstand wichtigste **Verteilung für stetige Zufallsgrößen**. Sie ist gekennzeichnet durch ihre symmetrische Form, die Konzentration der Werte in der Nähe des Erwartungswertes und die symmetrisch abnehmende Häufigkeit nach beiden Seiten mit zunehmender Entfernung vom Erwartungswert. Derartige Messwerte entstehen beispielsweise bei der Ermittlung von Körpermaßen (Tabelle 3.4), Gewichten, physiologischen Parametern, Laborwerten usw. Es handelt sich also in der Regel um physikalische, chemische, technische, biologische usw. Größen.¹

Tabelle 3.4: Körpergrößen von 40 Patienten (in cm):

167	182	183	177	161	171	180	177	179	186
176	179	173	174	176	171	169	170	174	173
174	183	176	176	171	173	169	168	172	179
165	181	171	179	167	180	185	173	174	170

Eine geeignete graphische Darstellung dieser Stichproben-Werte bietet das Histogramm. Mit zunehmendem Umfang von Stichproben nähert sich das Histogramm allmählich der typischen Form der Normalverteilung. Diese wird auch als **Gaußsche Glockenkurve** bezeichnet. Sie besitzt die beiden Verteilungsparameter (Modellparameter) μ (sprich: mü) und σ (sprich: sigma; Abbildung 3.5).

¹ In der Praxis streuen nicht alle Größen exakt normalverteilt. Es gibt oft mehr oder weniger große naturgegebene Abweichungen von der Normalverteilung. Relativ oft kann die Normalverteilung dann wenigstens näherungsweise eingesetzt oder durch eine geeignete Transformation der Zufallsvariablen eine Normalverteilung werden. Mitunter sind aber auch spezielle Methoden für andere Verteilungstypen erforderlich.

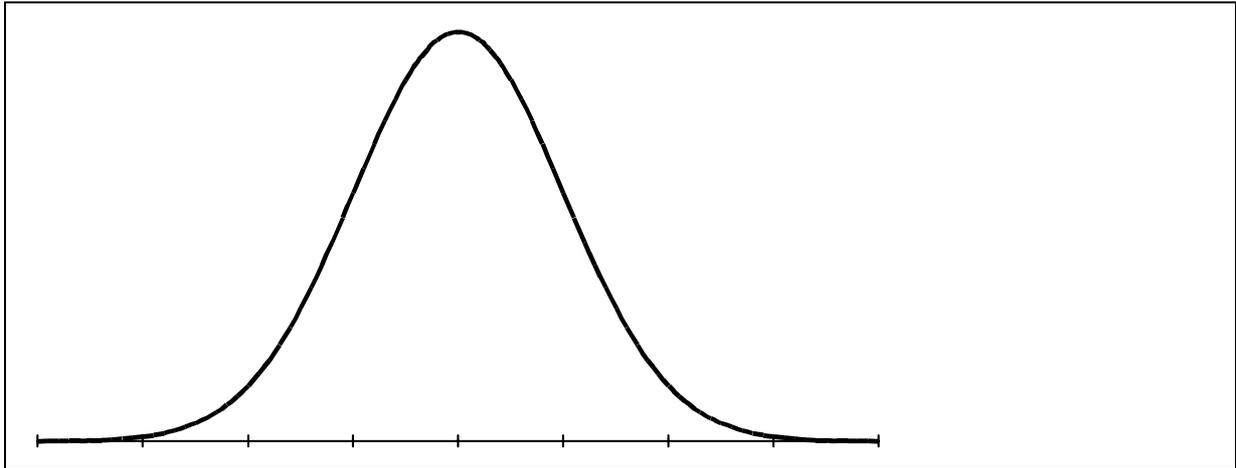


Abbildung 3.5: Gaußsche Glockenkurve

Die Normalverteilung ist durch zwei Parameter gekennzeichnet:

μ - Erwartungswert (mathematisch ungenauer: Mittelwert); dieser Verteilungsparameter charakterisiert die **Lage der Zufallsvariablen**. Er liegt an der Stelle der größten Häufigkeit (=Schwerpunkt) der Werte. Die Gaußsche Glockenkurve verläuft symmetrisch zu μ .

σ - Standardabweichung; sie charakterisiert, wie stark die einzelnen Messwerte x um den Schwerpunkt μ schwanken, ist also ein Maß für die **Streuung der Zufallsvariablen X**.

Die beiden Parameter μ und σ sind in der Regel nicht von vornherein bekannt, sondern müssen erst aus Stichproben geschätzt werden. Darauf wird später näher eingegangen. Die Gaußsche Glockenkurve bringt zum Ausdruck, dass die Wahrscheinlichkeit von Messwerten x_i in der Nähe des Erwartungswertes μ am größten ist und mit steigender Entfernung entsprechend dem Kurvenverlauf abnimmt.

Die Bedeutung der Kurve liegt darin, dass Flächenanteile unter der Kurve gleichbedeutend mit Wahrscheinlichkeiten sind. Bei deren Berechnung ergibt sich die Schwierigkeit, dass in der Praxis jede Zufallsvariable ihren eigenen Mittelwert μ und ihre eigene Standardabweichung σ hat. Es gibt also theoretisch eine unüberschaubar große Zahl von Normalverteilungen.

Die Normalverteilung hat nun aber die schöne Eigenschaft, dass durch die sogenannte **Standardisierung** jede konkret vorliegende Normalverteilung X auf eine einheitliche Form gebracht werden kann. Diese ist durch folgende Transformation gekennzeichnet:

$$\text{Originalwert } x_i \Rightarrow u_i = \frac{x_i - \mu}{\sigma} \quad (\text{standardisierter Wert}) \quad (3:3)$$

Die aus dieser Transformation hervorgegangene Zufallsvariable U ist ebenfalls normalverteilt mit Erwartungswert $\mu = 0$ und Standardabweichung $\sigma = 1$. Diese Normalverteilung wird als **Standardnormalverteilung** bezeichnet. Die Standardnormalverteilung, deren Werte in statistischen Tabellenwerken vertafelt sind, erlaubt die Berechnung von Wahrscheinlichkeiten dafür, dass Messwerte in einem bestimmten Intervall beobachtet werden.

Beispiel: Die Verteilung eines bestimmten physiologischen Parameters bei Patienten einer Krankenhausabteilung ist nach Auswertung von Patientenakten mit $\mu = 165$ und $\sigma = 13$ cm ermittelt worden. Es interessiert nun, mit welcher Wahrscheinlichkeit Messwerte unterhalb

von 145 und oberhalb von 200 erwartet werden können.² Dazu müssen diese Grenzwerte in einem ersten Schritt zunächst entsprechend (3:3) standardisiert werden:

$$\text{Standardisieren von 145: } u_{145} = (145 - 165) / 13 = -1,55$$

$$\text{Standardisieren von 200: } u_{200} = (200 - 165) / 13 = 2,69$$

Was sagen diese Werte aus? Der Wert u_{145} ist vom Betrag her gesehen kleiner als der u_{200} , liegt also näher am Erwartungswert Null der Standardnormalverteilung. Damit ist ein Wert von 145 oder in dessen Nähe wahrscheinlicher als ein Wert von rund 200.

Im zweiten Schritt lassen sich mit Hilfe statistischer **Tabellen zur standardisierten Normalverteilung** sich die Flächen $G(u)$ bzw. $Q(u)$ linksseitig bzw. rechtsseitig von einem vorgegebenen u -Wert unterhalb der Gaußschen Glocke der standardisierten Normalverteilung ablesen. Für negative u -Werte gilt infolge der Symmetrie der Verteilung:

$$G(-u) = 1 - G(u) = Q(u) \quad (3:4)$$

Auszüge aus derartigen Tabellen liefert die folgende Übersicht. Dabei gibt $p(x)$ die Wahrscheinlichkeit an, dass bei einer Standardnormalverteilung ein Wert jenseits von $u(x)$ auftritt.

Tabelle 3.5: Auszug aus einer Vertafelung der Standardnormalverteilung

$u(x)$	$p(x)$	$u(x)$	$p(x)$	$u(x)$	$p(x)$	$u(x)$	$p(x)$
0,0	0,5	1,0	0,841	2,0	0,977	3,0	0,999
0,1	0,540	1,1	0,864	2,1	0,982		
0,2	0,579	1,2	0,885	2,2	0,986		
0,3	0,618	1,3	0,903	2,3	0,989		
0,4	0,655	1,4	0,919	2,4	0,992		
0,5	0,692	1,5	0,933	2,5	0,994		
0,6	0,726	1,6	0,945	2,6	0,995		
0,7	0,758	1,7	0,955	2,7	0,997		
0,8	0,788	1,8	0,961	2,8	0,997		
0,9	0,816	1,9	0,971	2,9	0,998		

Mit den standardisierten Werten $u_{145} = -1,55$ und $u_{200} = 2,69$ lassen sich aus der vorangegangenen Übersicht die gesuchten Unterschreitungsanteile (für Werte unterhalb des Erwartungswertes) p_{145} und Überschreitungsanteile p_{200} (für Werte oberhalb des Erwartungswertes) als Flächen unter der standardisierten Normalverteilung ablesen:

$$u_{145} = -1,55 \quad \rightarrow \quad p_{145} = 0,06$$

$$u_{200} = 2,69 \quad \rightarrow \quad p_{200} = 0,003$$

Werte unterhalb von 145 treten mit einer Wahrscheinlichkeit von 6% auf, Werte oberhalb von 200 mit einer Wahrscheinlichkeit von lediglich 0,3%.

Bei vielen statistischen Anwendungen spielt auch das Ablesen des u -Wertes bei gegebenem $p(x)$ eine Rolle; für die wichtigsten u -Werte ist dies Ablesen mit der nachfolgenden kleinen Sonderaufstellung leicht möglich (Tabelle 3.6). Im übrigen sind derartige Tabellen in jedem Statistiklehrbuch zu finden.

² Streng genommen sind dies Schätzwerte, man sieht aber darüber in der Praxis oft hinweg und betrachtet diese als Parameter der Grundgesamtheit.

Tabelle 3.6: Auszug aus einer Vertafelung der Standardnormalverteilung (Forts.)

p(x)	u(x)	u(x)	p(x)	u(x)	p(x)
0,999	3,09	0,94	1,55	0,75	0,67
0,995	2,58	0,93	1,48	0,70	0,52
0,99	2,33	0,92	1,41	0,65	0,33
0,98	2,05	0,91	1,34	0,60	0,25
0,97	1,88	0,90	1,28	0,55	0,13
0,96	1,75	0,85	1,04	0,50	0,00
0,95	1,64	0,80	0,85		

In unserem Beispiel entspricht dies der Frage bedeutet, bis zu welchem Wert $x_{0,90}$ 90% der Werte des Laborparameters liegen. Es gilt dann

$p(x) = 0,90 \rightarrow u_{0,90} = 1,69$. Das ist der standardisierte Wert, bis zu dem 90 % der Messwerte liegen.

Wie lautet der entsprechende $x_{0,90}$ -Wert, bis zu dem 90 % der Parameterwerte liegen?

$$x_{0,90} = (u_{0,90} * \sigma + \mu) = x_{0,90} = (1,69 * 13 + 165) = 187$$

Bei einer Normalverteilung liegen innerhalb

$\mu \pm 1\sigma$	63 % aller Messwerte
$\mu \pm 2\sigma$	95,5 % aller Messwerte
$\mu \pm 3\sigma$	99,8 % aller Messwerte

sowie

$\mu \pm 1,65 \sigma$	90 % aller Messwerte
$\mu \pm 1,96\sigma$	95 % aller Messwerte
$\mu \pm 2,58 \sigma$	99 % aller Messwerte

Die Parameter μ und σ der Normalverteilung sind meist nicht von vornherein bekannt, sie müssen geschätzt werden. Dazu muss das entsprechende Merkmal durch eine **Stichprobe vom Umfang n** beobachtet werden; das sind die n Messwerte (auch 'Urwerte' genannt)

$$x_1, x_2, x_3, \dots, x_n,$$

die alle aus der gleichen Grundgesamtheit stammen sollen.

Aus einer Stichprobe vom Umfang n können verschiedene Kenngrößen berechnet werden. Die wichtigsten sind das **arithmetische Mittel** \bar{x} und die **Standardabweichung s**.

Beispiel: Bei den = 40 Körpergrößen des obigen ergibt sich als **Mittelwert** $\bar{x} = 174,6$ und als **Standardabweichung s** = 5,64

Wenn man die Körpergrößen aller Erwachsenen kennen würde (Grundgesamtheit!), dann ergäben sich daraus der **Mittelwert μ** und die **Standardabweichung σ** der Grundgesamtheit. Die aus der Stichprobe berechneten Werte \bar{x} und s sind **Schätzwerte für μ und σ** . Das arithmetische Mittel und die Standardabweichung sind im Falle einer Normalverteilung die besten Schätzungen für μ und σ , die man aus einer Stichprobe vom Umfang n berechnen kann.

Die gewonnenen Schätzwerte müssen dennoch mit einer gewissen Vorsicht interpretiert werden. Sie sind zwar die besten Schätzwerte, die aus einer Stichprobe abgeleitet werden können, sind aber andererseits selbst Zufallsgrößen und folglich zufälligen Einflüssen unterworfen! Infolge des Zufallscharakters einer Stichprobe könnten weitere Stichproben andere Schätzwerte ergeben, d.h. die Schätzung ist ungenau. Es ist unmittelbar einleuchtend, dass derartige Einflüsse mit steigendem Stichprobenumfang abnehmen.

Die Abschätzung der Ungenauigkeiten einer Schätzung ist nur mit dem entsprechenden **Vertrauensbereich** oder auch **Konfidenzintervall** möglich. Dazu muss vorher immer erst das **Vertrauensniveau p** festgelegt werden. Üblicherweise wird p nahe an eins gewählt, etwa p =

0,90, 0,95 oder sogar 0,99. Das Vertrauensniveau gibt die Wahrscheinlichkeit dafür an, dass der Vertrauensbereich für einen Schätzwert \bar{x} oder s einer (zukünftigen Stichprobe) den wahren, aber unbekanntem, Parameter μ bzw. σ überdeckt.

Wenn man nacheinander mehrere Stichproben aus der gleichen Grundgesamtheit ziehen würde, so würde man aus diesen Stichproben unterschiedliche Vertrauensbereiche ermitteln (Ursache dafür ist der Zufallscharakter der Stichproben - jede Stichprobe ist anders). Aufgrund zufälliger Einflüsse bei der Stichprobenziehung können dabei in seltenen Fällen extreme Intervalle entstehen, die den wahren Parameter **nicht** überdecken. Dies passiert allerdings mit einer Wahrscheinlichkeit von lediglich $\alpha = 1-p$ Prozent. Der Vertrauensbereich wird breiter, wenn das Vertrauensniveau erhöht wird. Der Vertrauensbereich wird schmaler, wenn der Stichprobenumfang erhöht wird.

Folgende **Vertrauensbereiche** sind bei der Normalverteilung von besonderem Interesse:

- für den Erwartungswert μ bei bekannter, meist aber unbekannter Standardabweichung σ :

$$\mu_{\text{un}} \leq \mu \leq \mu_{\text{ob}} \quad (3:5)$$

- für die Standardabweichung σ :

$$\sigma_{\text{un}} \leq \sigma \leq \sigma_{\text{ob}} \quad (3:6)$$

Beispiel: Bei einem Lieferanten medizinischer Produkte wird die Herstellung von Einwegspritzen mit einem Sollvolumen von 12ml überprüft. Aufgrund von Voruntersuchungen ist bekannt, dass die Werte einer Normalverteilung unterliegen. Zur Überprüfung der Einstellung der Maschine werden 30 Spritzen gemessen. Aus den Messwerten wurde $\bar{x}=12,25$ ml berechnet. Es besteht der Verdacht, dass die Produktion der Spritzen nicht mehr den Sollvorgaben entspricht, da das Volumen gegenüber dem Sollwert von 12 erhöht ist.

Der gewonnene Schätzwert $\bar{x} = 12,25$ ist zwar der beste Schätzwert, der aus einer Stichprobe angegeben werden kann, stellt aber andererseits eine Zufallsvariable dar und ist folglich, gerade bei einer kleinen Stichprobe, mit Ungenauigkeiten behaftet!

Weiterhin kann häufig die Standardabweichung σ der Zufallsvariable X nicht als bekannt vorausgesetzt werden, muss also ebenfalls aus der Stichprobe geschätzt werden. Diese Schätzung ergab einen Wert von $s = 0,25$.

Die Abschätzung der Ungenauigkeiten einer Schätzung ist nun mit dem entsprechenden **Vertrauensbereich** möglich. Der Vertrauensbereich für μ wird bei unbekannter Standardabweichung σ näherungsweise³ nach folgenden Beziehungen berechnet:

$$\mu_{\text{ob}} = \bar{x} + u_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad (3:7a)$$

$$\mu_{\text{un}} = \bar{x} - u_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad (3:7b)$$

Darin ist α die Irrtumswahrscheinlichkeit und $p = 1-\alpha$ das Vertrauensniveau.

Wie lautet der Vertrauensbereich für μ bei einem Vertrauensniveau von $p = 0,99$? Es muss zunächst der Wert $u_{1-\alpha/2}$ ermittelt werden

$$1. \quad u_{1-\alpha/2} = 2,58$$

$$2. \quad \mu_{\text{un}} = 12,25 - 2,58 \cdot 0,25 / \sqrt{30} = 12,37$$

$$\mu_{\text{ob}} = 12,25 + 2,58 \cdot 0,25 / \sqrt{30} = 12,11$$

³ Der exakte Vertrauensbereich ergibt sich bei Verwendung der t-Verteilung. Bei hinreichend großem n (ab ca. 25) sind die Unterschiede jedoch gering.

Der Prozess ist also tatsächlich außer Kontrolle, da der 99%-Vertrauensbereich den Sollparameter 12mm nicht überdeckt

Der **Vertrauensbereich** ist somit eine Erweiterung der Aussage des Schätzwertes: Der ermittelte Vertrauensbereich von $12,0375 \text{ mm} \leq \mu \leq 12,4625 \text{ mm}$ schließt den unbekannt wahren Wert des zu schätzenden Parameters μ (=Prozesslage) ein, wobei das Vertrauensniveau $p = 0,99$ bzw. die Irrtumswahrscheinlichkeit $\alpha = 0,01$ betragen.

Das bedeutet: An Stelle der ursprünglichen punktuellen Aussage des Schätzwertes von 12,25 mm (scheinbar genaue Aussage) tritt jetzt die erweiterte Aussage des Vertrauensbereiches, dass das Intervall von 12,0375 mm bis 12,4625 mm den unbekannt wahren Parameter μ mit einer Wahrscheinlichkeit von 0,99 überdeckt. Welchen Wert der Prozessmittelwert μ im Vertrauensbereich genau hat, darüber ist keine weitere Aussage möglich. Der beste vorliegende Schätzwert beträgt $\bar{x} = 12,25 \text{ mm}$.

Zum Verständnis der Vertrauensbereiche sei es noch einmal wiederholt: Es kommt es in seltenen Fällen vor, dass Vertrauensbereiche berechnet werden, die den tatsächlichen Parameterwert μ **nicht** überdecken. Die Häufigkeit einer derartigen Fehlaussage beträgt im vorliegenden Beispiel allerdings höchstens 1 % (= Irrtumswahrscheinlichkeit α).

Die Breite des Vertrauensbereiches beträgt

$$\Delta\mu = \mu_{ob} - \mu_{un} = 2 \cdot u_{1-\alpha/2} \cdot \frac{s}{\sqrt{n-1}} \quad (3:8)$$

Daraus folgt:

- Die Breite des Vertrauensbereiches ist proportional zu σ .
- Die Breite des Vertrauensbereiches ist umgekehrt proportional zu $\sqrt{n-1}$, d.h. für eine Halbierung des Vertrauensbereiches ist ein vierfacher Stichprobenumfang nötig !
- Eine Vergrößerung des Vertrauensniveaus p führt zu einer Vergrößerung des Vertrauensbereiches !

3.3.2 Binomialverteilung

Entnimmt man einer Grundgesamtheit Stichproben vom Umfang n , wobei die Stichprobe gegenüber der Grundgesamtheit sehr klein sein muss (Faustregel: die Grundgesamtheit muss mindestens 10 mal so groß sein wie die Stichprobe), und betrachtet allein, ob ein bestimmtes Ereignis E auftritt oder ein gewisser Tatbestand erfüllt ist (mit einer *unbekannten oder bekannten* Wahrscheinlichkeit p), so ist die Binomialverteilung als Modell geeignet. Verwendet wird die Binomialverteilung also bei Merkmalen, die nur zwei mögliche Ausprägungen haben können.

Binäre Merkmalswerte können z.B. sein:

- Erfolg bei einer Operation,
- Behandlungsfälle mit Überschreitung der Anhaltzahl,
- Patienten mit Begleiterkrankungen,
- Patienten über 65 Jahre,
- kranke Mitarbeiter

Beispiel: Eine Krankenhaus plant die Neugestaltung des zentralen Aufnahmebereichs, in dem 15 Mitarbeiter für die Patientenaufnahme und -beratung eingesetzt werden sollen. Sind mehr als drei Mitarbeiter an einem Tag krank, so sollen zur Sicherstellung dieses wichtigen Dienst-

leistungsprozesses Mitarbeiter aus anderen Bereichen angefordert werden. Der durchschnittliche Krankenstand bei allen Krankenhausmitarbeitern beträgt 8 %. Wie häufig ist damit zu rechnen, dass Aushilfskräfte angefordert werden müssen? Dabei entspricht der durchschnittliche Krankenstand in diesem Bereich dem Anteil kranker Mitarbeiter in der Grundgesamtheit, also allen Krankenhausmitarbeitern, mit $p = 0,08$.

Zur Berechnung dieser Wahrscheinlichkeiten benötigen wir die Binominalverteilung

Definitionen und Formeln

p Häufigkeit des Ereignisses E in der Grundgesamtheit

n Stichprobenumfang

x Anzahl der Ereignisse in der Stichprobe

$g(x)$ Wahrscheinlichkeitsfunktion

$G(x)$ Verteilungsfunktion

$\binom{n}{x}$ Binominalkoeffizient = $n! / x! (n-x)!$ mit

$n!$ = $n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1$ (sprich: n Fakultät)

$g(x)$ **Wahrscheinlichkeitsfunktion**, also die

Einzelwahrscheinlichkeit genau x Ereignisse in der Stichprobe zu finden.

$G(x)$ **Verteilungsfunktion** (Wahrscheinlichkeitssumme), d.h. die Wahrscheinlichkeit bis zu (höchstens) x Ereignisse in der Stichprobe zu finden.

$$g(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \quad (3:9a)$$

$$g(0) = (1-p)^n \quad (3:9b)$$

$$G(x) = \sum_{i=0}^x g(i) = g(0) + g(1) + \dots + g(x) \quad (3:9c)$$

Die Wahrscheinlichkeit, dass im Aufnahmebereich von den 15 Mitarbeitern an einem Tag keiner, einer, zwei, drei, vier oder fünf Personen erkrankt sind, betragen $0,286$; $0,0373$; $0,227$; $0,086$; $0,022$; $0,004$

Wie groß ist die Wahrscheinlichkeit, dass an einem Tag aus anderen Abteilungen Mitarbeiter aushelfen müssen, d.h. wie groß ist $P(x>3)$?

$$P(X>3) = 1 - (P \leq 3) = 1 - (0,0286 + 0,0373 + 0,227 + 0,086) = 1 - 0,972 = 0,028 = 2,8\%$$

Um die Wahrscheinlichkeit für höchstens drei kranke Mitarbeiter zu berechnen, werden die Einzelwahrscheinlichkeiten addiert und anschließend die Komplementärwahrscheinlichkeit gebildet. Bezogen auf die Problemstellung bedeutet dies, dass in ca. 2,8% der Fälle Hilfspersonal angefordert werden muss.

Beispiel: In der Allgemeinchirurgie eines Krankenhauses werden stichprobenartig 80 Fälle, die über Fallpauschalen abgerechnet wurden, auf die korrekte Zuordnung von Diagnose- und Operationsschlüssel aus insgesamt über 1.000 derartigen Operationen überprüft. Als Resultat dieser Stichprobenprüfung wurden 4 Fälle beanstandet. Wie kann der Anteil unkorrekter Zuordnungen geschätzt werden? *Der Schätzwert für p lautet $4 / 80 = 0,05$.*

Die Zielsetzung für das Unternehmen lautet, dass insgesamt 3% unkorrekte Zuordnungen der beiden Schlüssel akzeptiert werden können. Bei einem größeren Anteil sollten Maßnahmen für eine Verbesserung der Dokumentation eingeleitet werden (zusätzliche Anleitungen, spe-

zielles Personal usw.). Aufgrund des vorliegenden Stichprobenergebnisses wurden durch die Krankenhausleitung derartige Maßnahmen eingeleitet.

Wie ist diese Entscheidung aus statistischer Sicht zu werten? Der Schätzwert von 5% übertrifft zwar die Sollvorgabe; die Abweichung kann aber auch aus Zufallseinflüssen herrühren. Eine einzelne Maßzahl lässt darüber keine Aussage zu.

Der gewonnene Schätzwert muss wie bei der Normalverteilung mit Vorsicht interpretiert werden. Er ist zwar der beste Schätzwert, der aus einer Stichprobe angegeben werden kann, ist aber andererseits - wie oben schon erläutert - selbst eine Zufallsvariable und folglich zufälligen Einflüssen unterworfen! Diese nehmen mit steigendem Stichprobenumfang ab.

Die Abschätzung der Ungenauigkeiten einer Schätzung ist mit dem entsprechenden **Vertrauensbereich** möglich. Für das vorliegende Beispiel lautet der Vertrauensbereich

$$1,4 \% \leq p \leq 12,3 \%$$

Es wurde ebenfalls schon betont, dass vor Berechnung eines Vertrauensintervalls immer erst das **Vertrauensniveau** P festgelegt werden muss, in diesem Fall ist $p = 0,95$. Dieser **Vertrauensbereich** von $1,4 \% \leq p \leq 12,3 \%$ schließt den unbekanntem wahren Wert des zu schätzenden Parameters p (= Anteil unkorrekte Zuordnungen) mit einer Wahrscheinlichkeit von 95 % ein.

Der Vertrauensbereich schließt den Sollwert 3% mit ein. Insofern ist die Abweichung in der Stichprobe vermutlich auf Zufallseinflüsse zurückzuführen. Aus statistischer Sicht hätte es dieser Maßnahmen nicht bedurft.

3.4 Statistische Testverfahren

3.4.1 Prinzip statistischer Testverfahren

Die statistischen Tests sind von zentraler Bedeutung innerhalb der statistischen Analyseverfahren. Sie sind universell auf sehr viele verschiedene statistische Fragestellungen anwendbar. Viele Untersuchungen laufen darauf hinaus, Vergleiche durchzuführen oder festzustellen, ob Annahmen zutreffen oder nicht. Typische Fragestellungen sind

- Vergleich einer Grundgesamtheit mit einem Sollwert
- Vergleich von zwei oder mehr Grundgesamtheiten
- Untersuchung zum Zusammenhang von zwei oder mehr Variablen
- Untersuchung der Verteilungsform

Diese Untersuchungen werden auf der Basis von Stichprobenergebnissen durchgeführt. In der Literatur wird eine große Zahl möglicher Tests beschrieben. Im Rahmen dieses Kapitels können nur die wichtigsten Aspekte der statistische Testphilosophie aufgezeigt werden.

Jeder Hypothesentest folgt dem folgenden **Schema** (Abbildung 3.6).

- Feststellung der den Merkmalen zugrunde liegenden Verteilung
- H_0 Nullhypothese aufstellen
- H_1 Alternativhypothese ermitteln
- geeigneten Test auswählen
- zu tolerierendes Fehlerniveau (=Signifikanzniveau) festlegen
- erforderlichen Stichprobenumfang festlegen
- Prüfgröße aus den Daten der Stichprobe(n) errechnen
- Kritischen Wert aus der Tabelle (Verteilung) ablesen
- Entscheidung

Abbildung 3.6: Ablaufschema eines statistischen Tests

Ein statistischer Test untersucht die Frage, ob die zufallsbedingte Streuung ausreicht, die beobachteten Effekte zu erklären. Der statistische Test untersucht, mit welcher Wahrscheinlichkeit sich die vorliegenden Ergebnisse (oder noch extremere) unter der Nullhypothese zufällig hätten ergeben können, ob die Nullhypothese also mit den vorliegenden Daten vereinbar ist.

Als **Nullhypothese** H_0 wird üblicherweise die Übereinstimmung formuliert, d.h. das Fehlen eines Unterschieds zwischen zwei Grundgesamtheiten oder eines Zusammenhangs zwischen zwei Variablen. Die **Alternativhypothese** H_1 besagt, dass der beobachtete Unterschied oder der erkennbare Zusammenhang tatsächlich besteht, ohne ihn allerdings zu quantifizieren.

Beim statistischen Test können somit zwei Fehler auftreten. Der **Fehler 1. Art** besteht darin, die Nullhypothese fälschlicherweise abzulehnen. Diese **Irrtumswahrscheinlichkeit** α kann quantifiziert werden. Liegt sie unterhalb der üblichen Grenzen von 5%, 1% oder 0,1%, spricht man von einem signifikanten Ergebnis. Beim **Fehler 2. Art** β wird trotz eines bestehenden Unterschieds das Ergebnis als nicht signifikant, also noch mit dem Zufall erklärbar, bewertet. (Abbildung 3.7).

Testentscheidung	Tatsache	
	H_0 trifft zu	H_0 trifft nicht zu
H_0 wird verworfen	Fehler 1. Art Irrtumswahrscheinlichkeit α	$1-\beta$ (Power) richtige Entscheidung
H_0 wird nicht verworfen	richtige Entscheidung	Fehler 2. Art Wahrscheinlichkeit β

Abbildung 3.7: Mögliche Fehler bei Testentscheidungen, Fehler 1. und 2. Art

Bei statistischen Tests können nicht beide Fehlerarten kontrolliert werden, da sie wechselseitig voneinander abhängen. Ein kleiner Fehler 1. Art hat einen großen Fehler 2. Art zur Folge und umgekehrt. Bei statistischen Tests wird üblicherweise der Fehler 1. Art kontrolliert, da man an der Aufdeckung von Unterschieden zwischen zwei (oder mehreren) Population, Zeitpunkten oder Behandlungen interessiert ist. Bei einer Nullhypothese, die von der Gleichheit ausgeht zweier Parameter, kann die Wahrscheinlichkeit für die Beobachtung eines bestimmten Unterschieds bzgl. der Zielgröße bestimmt und damit auch die Fehlerwahrscheinlichkeit für die irrtümliche Verwerfung der Hypothese.

Da andererseits die Größe des tatsächlichen Effekts nicht bekannt ist, kann β , also der Fehler beim Festhalten an der Hypothese, nur grob geschätzt werden. Er hängt ab von der Fallzahl, der Art des Tests, der Irrtumswahrscheinlichkeit und natürlich der Größe des wahren Effekts.

Aus diesem Grund kann bei einem nicht-signifikanten Ergebnis auch nicht von einer Annahme von H_0 gesprochen werden, sondern nur davon, dass die Nullhypothese nicht abgelehnt werden konnte.

Achtung: Ein Nicht-Verwerfen der Nullhypothese H_0 darf keinesfalls als Beweis für deren Richtigkeit aufgefasst werden, es konnte lediglich aufgrund des Stichprobenergebnisses nicht das Gegenteil der Nullhypothese H_0 nachgewiesen werden. Dies ist darin begründet, dass der Fehler 1. Art, also die irrtümliche Verwerfung der Nullhypothese, genau quantifiziert werden kann, nicht jedoch der Fehler 2. Art, da wir keine Kenntnis von der ‚wahren‘ Alternative haben.

Der statistische Test kann als **zweiseitiger Test** (der interessierende Parameter ist größer oder kleiner als ein Sollwert bzw. die Unterschiede zweier Effekte sind größer oder kleiner als Null) oder – falls es Erkenntnis über die Richtung des zu erwartenden Effekts gibt - auch als **einseitiger Test** (größer (kleiner) als ein Sollwert bzw. Unterschiede zweier Effekte größer (kleiner) als Null) geplant werden. Der Stichprobenumfang bestimmt das Auflösungsvermögen des Tests. Diese Empfindlichkeit wird bereits vor Beginn eines Versuchs geplant.

Es ist zunehmend üblich geworden, einen statistischen Test unter Beachtung der Signifikanzniveaus $\alpha = 5\%$, 1% und $0,1\%$ durch folgende Bewertung der Prüfgröße durchzuführen:

- Nichtverwerfen H_0 bei $\alpha = 5\%$
 \Rightarrow **zufälliges Resultat** –
- Verwerfen H_0 bei $\alpha = 5\%$, Nichtverwerfen bei $\alpha = 1\%$
 \Rightarrow **signifikantes Resultat** *
- Verwerfen H_0 bei $\alpha = 1\%$, Nichtverwerfen bei $\alpha = 0,1\%$
 \Rightarrow **hochsignifikantes Resultat** **
- Verwerfen H_0 bei $\alpha = 0,1\%$
 \Rightarrow **höchstsignifikantes Resultat** ***

3.4.2 t-Test auf Mittelwertunterschiede

Nachfolgend wird das Grundprinzip eines Testablaufs anhand des einfachen, aber weit verbreiteten t-Tests dargelegt. Dazu wird die oben bereits mit dem Vertrauensbereich untersuchte betriebliche Situation jetzt nochmals mit einem Test analysiert.

Es wird die Einstellung einer Maschine, die Einwegspritzen fertigt, überprüft. Die Maschine soll mit einem Mittelwert $\mu_0 = 12,00$ mm arbeiten (= Sollwert der Prozesslage). Bei Verletzung der Forderung muss die Maschine neu eingestellt werden. Zur Überprüfung der Einstellung der Maschine wurden $n = 10$ Spritzen aus einer Lieferung herausgegriffen und gemessen. Aus den Messwerten wurde $\bar{x} = 12,25$ mm berechnet. Die Standardabweichung s wurde mit $0,025$ geschätzt.

Die **Fragestellung** lautet: Weicht der beobachtete Mittelwert $\bar{x} = 12,25$ mm nur **zufällig** oder weicht er **signifikant** (systematisch) von der Vorgabe $\mu_0 = 12,00$ mm ab?

Das im vorhergehenden Abschnitt bereits ermittelte Ergebnis, dass der Herstellungsprozess einer Überprüfung bedarf, kann auch mit dem folgenden Testablauf gewonnen werden:

Formulierung der Nullhypothese H_0

$$H_0: \mu = \mu_0 \quad \text{mit} \quad \mu_0 = 12,000 \text{ mm.} \quad (3:10a)$$

Die Alternativhypothese H_1 lautet:

$$\mu \diamond 12,000 \text{ mm} \quad (3:10b)$$

Diese übliche Schreibweise bringt in kurzer und prägnanter Form die Behauptung (es kann sich auch um eine Forderung, eine Vermutung oder eine Hoffnung handeln) zum Ausdruck, dass der tatsächliche Mittelwert μ genau mit dem geforderten Mittelwert $\mu_0 = 12,000 \text{ mm}$ übereinstimmt, d.h. dass die Maschineneinstellung den Sollvorgaben entspricht.

Der gesamte Test zielt auf die Entscheidung, ob diese Hypothese aufgrund des Stichprobenergebnisses zu verwerfen oder nicht zu verwerfen ist.

Berechnung der Prüfgröße

Beim vorliegenden t-Test auf Mittelwertabweichungen wird die Prüfgröße folgendermaßen berechnet:

$$t_{\text{Prüf}} = \frac{\bar{x} - \mu_0}{s} \sqrt{n} \quad (3:11)$$

Die Prüfgröße ist die mit der Wurzel des Stichprobenumfangs gewichtete Standardgröße des Stichprobenmittelwertes. Je größer der Wert dem Absolutbetrag nach ist, desto weniger wahrscheinlich ist eine nur zufällige Abweichung vom Sollwert.

In unserem Fall berechnet sich

$$T_{\text{prüf}} = (12,25 - 12,00) * 3,16 / 0,25 = 3,16.$$

Die **Entscheidungsregel** lautet:

$$\left| t_{\text{Prüf}} \right| > t_{\text{kritisch}} \quad \text{dann Verwerfen der Nullhypothese } H_0: \mu = \mu_0 \quad (3:12)$$

mit dem kritischen Wert $t_{\text{kritisch}} = t_{1-\alpha/2; n-1}$

α ist das '**Signifikanzniveau**' (auch Irrtumswahrscheinlichkeit genannt). Das ist die Wahrscheinlichkeit dafür, dass eine zutreffende Nullhypothese verworfen wird (=Fehler 1.Art). Die t-Verteilung ähnelt der Normalverteilung und nähert sich dieser mit zunehmendem n stark an.

Bei der Durchführung von Tests werden die kritischen Werte $t_{\text{kritisch}} = t_{1-\alpha/2; n-1}$ für $\alpha = 5 \%$, 1% und $0,1 \%$ angegeben und die Prüfgröße wird danach bewertet. Sie liegen bei 2,26; 3,25 bzw. 4,78.

Der Wert der Prüfgröße übertrifft in unserem Beispiel den kritischen Wert von 2,26. Mit einer Fehlerwahrscheinlichkeit von 5% ist die Hypothese, dass der Fertigungsprozess den Sollvorgaben entspricht, abzulehnen. Beide Ansätze kommen prinzipiell zum gleichen Ergebnis.

Der geringe Unterschied rührt daher, dass - aus hier nicht näher zu erläuternden statistischen Gründen - gegenüber der Berechnung des Konfidenzintervalls bei diesem die t-Verteilung als Annäherung an die Normalverteilung benutzt wurde

In vielen Veröffentlichungen und auch in den meisten Statistikprogrammen ist es üblich geworden, anstelle oder in Ergänzung der Teststatistik den sogenannten **p-Wert** anzugeben. Dieser gibt ebenso wie der Vergleich der Teststatistik mit einem Grenzwert darüber Auskunft, ob die Nullhypothese verworfen werden kann oder nicht.

Der p-Wert gibt bei Annahme der Gültigkeit der Nullhypothese die Wahrscheinlichkeit dafür an, dass die Teststatistik einen Wert größer oder gleich dem beobachteten Wert annimmt. Ist die Wahrscheinlichkeit gering, d.h. der Wert kleiner als üblicherweise 0,05, 0,01, dann ist das

beobachtete Ergebnis nur schwer mit der Nullhypothese zu vereinbaren, weil zufällige Effekte kaum ein Ergebnis dieser Größenordnung hervorgerufen haben können.

Kleine p-Werte unter einem vorgegebenen Wert für die Fehlerwahrscheinlichkeit α führen also zur Ablehnung der Nullhypothese.

3.4.3 Chi-Quadrat-Test

Ein weiterer sehr verbreiteter Test in der Statistik und der Epidemiologie ist der Chi-Quadrat-Test bei nominal oder ordinal verteilten Merkmalen.

Beispiel: Zur Unterstützung der Entscheidungsfindung über den verstärkten Ausbau ambulanter Leistungen am Krankenhaus sollen die Wiedereinweisungsraten bei ambulanten Operationen und vollstationärer Behandlung miteinander verglichen werden. Bei Leistenhernien (ICD 550) wurden von 75 ambulant operierten Fällen 15 innerhalb einer Woche wiederaufgenommen, bei vollstationärer Aufnahme waren es ebenfalls 15 bei insgesamt 125 Fällen.

Wiederaufnahme erforderlich	ambulante Operation	vollstationäre Aufnahme	gesamt
ja	15	15	30
nein	60	110	170
gesamt	75	125	200

Die Wiedereinweisungsraten bei beiden Operationsformen und insgesamt bei allen Patienten mit Leistenhernien betragen:

$$\text{Amb. Operation} = 15 / 75 = 20\%$$

$$\text{bei vollstationärer Aufnahme: } 15 / 125 = 12\%;$$

$$\text{insgesamt: } 30 / 200 = 15\%$$

Die Nullhypothese bei einem Test für oder gegen eine stärkere ambulante Erbringung dieser Operation lautet: *Die Wiedereinweisungsraten unterscheiden sich nicht.*

Beim Test auf statistisch signifikante Unterschiede zwischen den beiden Wiedereinweisungsraten - unter der Annahme, dass im übrigen weitere Einflussfaktoren auf diesen Qualitätsindikator wie Alter des Patienten, Komorbidität, Ausmaß sozialer Unterstützung, um nur einige zu nennen, bei Patienten beider Gruppen gleich seien - werden die beobachteten Zellhäufigkeiten in dieser **Vierfeldertafel** - allgemein bei beliebiger Spalten- und Reihenzahl: **Kontingenztafel** - mit den Zellhäufigkeiten verglichen, die bei einer gleichen (Durchschnitts-)Rate in beiden Gruppen zu erwarten gewesen wären.

Unter der Annahme, dass die oben ermittelte Durchschnittsrate für beide Gruppen (n=75 ambulante Operationen und n=125 vollstationäre Operationen) gegolten hätte, betragen die erwarteten Zellhäufigkeiten

$$\text{für ambulante. Operation: } 75 * 0,15 = 11,25; \text{ bei vollstationärer Aufnahme: } 18,75$$

Große Differenzen zwischen erwarteten und beobachteten Häufigkeiten sprechen nun gegen die Hypothesen gleicher Wiedereinweisungsraten und für einen unterschiedlichen Erfolg beider Operationsarten. Oder anders gesagt: Bei großen Unterschieden ist ein Zusammenhang bzw. eine Abhängigkeit der beiden Merkmale Operationsart und Wiedereinweisungshäufigkeit anzunehmen.

Der statistische Test prüft also - je nach Sichtweise - die **Gleichheit** von zwei Raten oder die **Unabhängigkeit** zweier Merkmale.

Nun werden die Differenzen zwischen beobachteten und erwarteten Zellhäufigkeiten quadriert und durch die erwartete Häufigkeit dividiert. Dies wird für alle vier Zellen gemacht und die Einzelergebnisse über alle vier Zellen aufsummiert. Die Zellhäufigkeiten für die möglichen Kombinationen der Merkmalswerte (jeweils 0 oder 1) zweier Merkmale X und Y werden allgemein wie folgt bezeichnet.

Merkmal X	Merkmal Y		gesamt
	0	1	
0	a	b	a+b
1	c	d	c+d
gesamt	a+c	b+d	n

Damit entsteht die **Chi-Quadrat-Teststatistik** (oder χ^2 -Teststatistik).

$$\chi^2 = \frac{n \cdot (a \cdot b - c \cdot d)^2}{(a+c) \cdot (b+d) \cdot (a+b) \cdot (c+d)} \quad (3:13)$$

Das Ergebnis dieser Berechnungen muss mit dem Grenzwert der zugehörigen Chi-Quadratverteilung mit einem Freiheitsgrad verglichen werden. Bei einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$ liegt er bei 3,84; bei einer Irrtumswahrscheinlichkeit von $\alpha = 1\%$ beträgt er 6,64.

Überschreitet der Wert der Teststatistik diesen Grenzwert, muss die Nullhypothese abgelehnt werden.

In unserem Beispiel berechnet sich die *Chi-Quadrat-Teststatistik* wie folgt:

$$200 \cdot (60 \cdot 15 - 15 \cdot 110)^2 / 60 \cdot 15 \cdot 15 \cdot 110 = 2,35 .$$

Dieser Wert ist kleiner als der Grenzwert 3,84. Also kann die Hypothese der Gleichheit der Wiedereinweisungsraten nicht verworfen werden.

Der Chi-Quadrat-Test kann sehr vielseitig bei nominalen oder ordinalen Merkmalen eingesetzt werden. Er ist auf jede andere Kontingenztafel beliebiger Größenordnung zu übertragen, also auch bei Variablen anwendbar, bei denen eine oder beide mehr als nur zwei mögliche Ausprägungen besitzen.

3.5 Analyse des Zusammenhangs zwischen Variablen

3.5.1 Korrelationsanalyse

Zahlreiche medizinische Prozesse hängen von einer großen Anzahl von Einflussgrößen ab. Zwar lassen sich diese jeweils einzeln bivariat unter Einbeziehung lediglich einer unabhängigen und einer abhängigen Variable statistisch untersuchen. Das komplexe Zusammenspiel mehrerer unabhängiger Größen und deren wechselseitige Abhängigkeiten lässt sich befriedigend aber nur unter einer simultanen, multivariaten Analyse unter Einbeziehung sämtlicher potentieller Einflussgrößen bewerten.

Hier ist nicht der Raum, diese Verfahren im einzelnen darzustellen. Vielmehr soll für die gebräuchlichsten nur deren grundsätzlicher Ansatz dargelegt werden. Da dieser nicht davon abhängt, wie viele Variablen einbezogen werden, werden zur anschaulichen Erläuterung hier allerdings nur zwei Variablen betrachtet. Eine Übertragung auf den Fall mehrerer Variablen ist dann leicht möglich.

Für die folgenden Ausführungen betrachten wir folgendes Beispiel. An 20 Patienten eines niedergelassenen Arztes wurden folgende Daten erhoben (Tabelle 3.7). Anhand dieser Messungen sollen Einflussfaktoren auf den Blutdruck bewertet werden.

Tabelle 3.7: Ergebnisse medizinischer Untersuchungen und Rauchstatus bei 20 Patienten

Geschlecht	Alter	rel. Gewicht	syst. Blutdruck	diast. Blutdruck	Raucher
männlich	25	88	170	95	ja
männlich	37	105	140	90	ja
männlich	39	79	145	90	ja
männlich	43	140	140	95	nein
männlich	48	96	100	70	nein
männlich	48	112	155	90	nein
männlich	49	125	160	95	ja
männlich	50	150	145	90	nein
männlich	50	99	160	95	ja
männlich	51	114	135	85	nein
männlich	52	132	140	90	nein
weiblich	52	128	155	95	ja
weiblich	54	130	150	90	nein
weiblich	55	164	190	100	ja
weiblich	56	112	165	95	ja
weiblich	57	96	175	100	ja
weiblich	58	115	155	90	nein
weiblich	58	123	170	95	ja
weiblich	59	97	155	110	nein
weiblich	59	104	140	85	nein

Gewicht in kg / (Größe in cm-100) * 100; Rauchstatus

Stellt man den diastolischen und den systolischen Blutdruck in einem 2-dimensionalen **Scatterplot** dar, erkennt man, dass sich durch die Punktwolke in guter Annäherung eine Gerade legen lässt, die Punkte allerdings streuen um die Gerade. Daraus ergeben sich weiterführende Fragen:

- Wie lässt sich die Stärke des Zusammenhangs quantifizieren?
- Wie groß ist die Steigung der gedachten Gerade durch die Punktwolke, d.h. wie stark nimmt der systolische Blutdruck mit Anstieg des diastolischen zu?

Zur Beantwortung dieser Fragen dienen die Korrelations- und Regressionsanalyse. Für zwei dichotome Merkmale (mit jeweils zwei Ausprägungen) haben wir den Chi-Quadrat-Test schon als Verfahren kennen gelernt, Zusammenhänge zwischen diesen zu untersuchen. Er gab Auskunft darüber, ob eine Abhängigkeit zwischen ihnen bestand oder nicht.

Bei zwei stetigen Merkmalen X und Y lässt die Berechnung eines **Korrelationskoeffizienten** r eine Aussage darüber, ob eine Assoziation zwischen ihnen besteht. Der Pearson'sche Korrelationskoeffizient ist definiert als

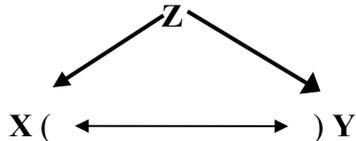
$$r_{X,Y} = \left(\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right) / \sqrt{\left(\sum_{i=1}^n x_i - n \bar{x} \right)^2 \left(\sum_{i=1}^n y_i - n \bar{y} \right)^2} \quad (3:14)$$

Der Korrelationskoeffizient hat einen theoretischen **Wertebereich** von -1 bis +1. Ein Wert von +1 bedeutet eine exakte positive (lineare) Beziehung zwischen X und Y, ein Wert von -1 eine eindeutige negative Beziehung.

In unserem Beispiel beträgt die Korrelation zwischen dem systolischen und diastolischen Blutdruck $r = 0,78$. Es gibt demnach einen starken Zusammenhang zwischen diesen beiden Variablen.

Aus einem hohen Wert des Korrelationskoeffizienten lässt sich aber keine ursächliche Abhängigkeit zwischen beiden Variablen ablesen. Eine derartige Aussage lässt sich nur aus der inhaltlichen Fragestellung ableiten. Beim Beispiel des Blutdrucks aus Kenntnis der physiologischen Zusammenhänge: ja. Allein aus Kenntnis der empirischen Daten: nein

Nun ist es denkbar, dass eine hohe Korrelation zwischen X und Y durch den Einfluss einer dritten Variablen Z zustande kommt. Ein häufig angeführtes Beispiel ist der scheinbare Zusammenhang zwischen der Zahl der Störche und der Zahl neugeborener Babies, wenn man jährliche oder monatliche Zeitreihen betrachtet. Diese unsinnige Assoziation wird allein durch eine parallel verlaufende zeitliche bzw. saisonale Entwicklung beider Zahlenreihen hervorgerufen. Der Störfaktor ist hier also die Zeit, die den eigentlichen interessierenden Zusammenhang überdeckt. Durch Berechnung **partieller Korrelationen** zwischen X und Y lässt sich der Einfluss einer dritten Variablen Z herausrechnen.



In unserem Beispiel beträgt die partielle Korrelation unter Berücksichtigung des Alters 0,775. Der Korrelationskoeffizient verändert sich also nur geringfügig. Das Alter spielt also für den Zusammenhang zwischen den beiden Variablen keine Rolle. Möglicherweise aber für die Höhe des Blutdrucks.

Sind die beiden Merkmale sehr schief verteilt oder gar nur diskret, dann sollte anstelle des Pearson'schen Korrelationskoeffizienten ein **Rangkorrelationskoeffizient** berechnet werden. Der Rangkorrelationskoeffizient nach Spearman wird auf die gleiche Weise wie der von Pearson berechnet, nur dass statt der Messwerte die **Ränge** genommen werden, die entstehen, wenn man die X- und die Y-Werte getrennt der Größe nach ordnet.

Einfacher ist die Berechnung, wenn man die äquivalente Gleichung

$$R_s = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \text{ mit } d_i = R(x_i) - R(y_i) \quad (3:15)$$

verwendet. Haben mehrere Beobachtungen den gleichen Merkmalswert ("**Bindungen**"), so wird ihnen allen als Rang das arithmetische Mittel der in Frage stehenden Ränge zugeteilt. In unserem Beispiel ändert sich die Höhe der Korrelation bei Übergang zum Rangkoeffizienten kaum etwas, er liegt bei 0,7928.

Es kann über die Berechnung eines Korrelationskoeffizienten eine Aussage darüber getroffen werden, ob eine Assoziation zwischen zwei Variablen X und Y besteht und wenn ja, ob sie sich gleichsinnig oder gegenläufig verändern.

Weitere Schlussfolgerungen sind aus dem Wert des Korrelationskoeffizienten nicht möglich. Es ist bei Vorliegen einer nennenswerten Korrelation noch keine funktionale Beziehung bestimmt, d.h. die Frage welcher Änderung Δx der Variablen X (Δ drückt in der Mathematik eine Änderung oder eine Differenz aus) eine bestimmte Änderung Δy der Variablen Y entspricht, oder konkret auf unser Beispiel bezogen: Wie ändert sich der systolische Blutdruck bei einem Anstieg des diastolischen um 10.

Insbesondere dürfen - wie schon erwähnt - allein aufgrund dieser Ergebnisse keine Aussagen über Ursache und Wirkung gemacht werden! Man hat zwar einen 'Zusammenhang' der Daten festgestellt (durch eine signifikante Korrelation), man weiß damit aber nicht, ob

- die Änderung der Variablen x die Änderung der Variablen y zur Folge hat,
- die Änderung der Variablen y die Änderung der Variablen x zur Folge hat,
- oder ob die Änderung von einer oder von mehreren anderen Variablen abhängt.

3.5.2 Regressionsanalyse

In der Korrelationsanalyse erhält man ein quantitatives Maß für den Zusammenhang zwischen zwei Merkmalen, in der Regressionsanalyse wird ein funktionaler Zusammenhang spezifiziert:

- Nachweis einer bekannten Beziehung,
- Schätzen der Parameter einer bekannten Beziehung,
- Erkennen eines funktionalen Zusammenhangs,
- empirische Repräsentation großer Datenmengen,
- Interpolation fehlender bzw. Prognose zukünftiger Werte.

Liegen die den beobachteten Wertepaaren (x_i, y_i) zugeordneten Punkte so in der Ebene (x, y) , dass man in guter Näherung eine Gerade durch diese Punkte legen kann und sind beide Variablen annähernd normal verteilt, so sind die Voraussetzung für die **lineare Regression** erfüllt. Die näher zu untersuchende lineare Beziehung zwischen X und Y hat die allgemeine Form

$$Y = \alpha + \beta * X \quad (3:16)$$

Es soll nun eine **Regressionsgerade** so durch die Punktwolke gelegt werden, dass durch sie eine "möglichst gute" Schätzung \hat{y} für die Ausprägung y des (abhängigen) Merkmals Y eines Objekts, dass die Ausprägung x des (unabhängigen) Merkmals X besitzt, gelingt. Als Kriterium für die Güte der Geradenanpassung wird die Summe der Abweichungsquadrate

$$S^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 \quad (3:17)$$

gewählt.

Die Schätzer a und b für α und β werden nun so gewählt, dass dieser Ausdruck minimiert wird ("**Kleinste-Quadrat-Schätzung**").

Für a und b ergeben sich in unserem Beispiel folgende Ausdrücke

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = 1,88$$

$$a = \bar{y} - b * \bar{x} = -22,1$$

Bei einem Anstieg des diastolischen Blutdrucks um 10, steigt der systolische Blutdruck um knapp 19 Punkte. Der Schätzer für a gibt den Punkt an, wo (theoretisch) die Gerade bei einem diastolischen Blutdruck von Null die y-Achse schneidet.

Die Regressionsgerade mit ihren beiden Schätzwerten a und b erlaubt es, für bekannte Merkmalsausprägungen X bzw. Y Schätzer für die entsprechenden Ausprägungen der Merkmale Y bzw. X zu berechnen. Bei einem diastolischen Druck von 80 ist ein systolischer Druck ungefähr von $-22,1 + 80 * 1,88 \approx 128$ zu erwarten.

Da beide Schätzwerte a und b ebenfalls Zufallsvariable sind, lassen sich zu Ihnen auch Standardabweichungen angeben. Auf diese Weise wird es möglich zu testen, ob sich die Schätzung b für den Steigungsparameter β signifikant von Null unterscheidet. Ist dies der Fall, kann von einer Assoziation, nicht unbedingt aber einer kausalen Beziehung, der beiden Merkmale X und Y gesprochen werden.

Die Nullhypothese lautet also $\beta = 0$. Diese wird mit einer Fehlerwahrscheinlichkeit von $\alpha = 5\%$ **abgelehnt**, wenn das Intervall

$$[b - 1,96 * \text{Standardabweichung von } b; b + 1,96 * \text{Standardabweichung von } b] \quad (3:18)$$

nicht den Wert 0 enthält

Unser Schätzer b für β hat die Standardabweichung 0,359. Das Intervall zur Überprüfung der Nullhypothese lautet $[1,18; 2,58]$. Die Steigung der Regressionsgeraden unterscheidet sich daher signifikant von Null.

Abschließend interessiert die Güte der Regressionsgeraden, konkret wie "eng" diese an den beobachteten Werten liegt. Dies wird ausgedrückt durch das **Bestimmtheitsmaß**. Im Fall einer linearen Regression ergibt sich dies als Quadrat der Pearson'schen Korrelationskoeffizienten

$$B_{Y,X} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3:19)$$

Das Bestimmtheitsmaß lässt sich bei allen Formen der Regression berechnen. Speziell bei der linearen Regression kann man sich die Eigenschaft zu Nutze machen, dass das Bestimmtheitsmaß das Quadrat des Regressionskoeffizienten ist.

$$B_{Y,X} = r_{x,y}^2 \quad (3:20)$$

Ein Bestimmtheitsmaß von Eins bedeutet, dass alle Punkte auf der Gerade liegen, also ein exakter funktionaler Zusammenhang zwischen beiden Variablen vorliegt. Eine Gerade, um die die Beobachtungspaare nur wenig schwanken, ist durch einen Wert nahe Eins gekennzeichnet, Werte nahe Null deuten auf eine fehlende Assoziation (oder eine nichtlineare Beziehung) hin. Im Beispiel ergibt sich $B_{Y,X} = (0,7784)^2 = 0,606$

4. Kurzer Abriss zu Methoden der Sozialwissenschaften

In diesem Kapitel kann wie schon im vorhergehenden Abschnitt nur auf die für die Epidemiologie, und dort speziell für die Studienplanung und die Phase der Datenerhebung, wesentlichen Punkte eingegangen werden. Im übrigen sei auf andere Lehrveranstaltungen und auf Lehrbücher zu Methoden der Sozialwissenschaften verwiesen.

4.1 Gütekriterien eines Verfahrens

Ein entscheidendes Problem bei jedem Erhebungsinstrument und jedem Test liegt in der Identifikation und Kontrolle nicht zufällig auftretender Mess- bzw. Erhebungsfehler, die die Datenerhebung beeinflussen und die "wahren" Werte der gemessenen Größen systematisch verfälschen. Zur Beschreibung der Qualität eines Verfahrens gibt es drei Kriterien.

Objektivität eines Verfahrens bedeutet, dass der Untersucher selbst keinen Einfluss auf die Art des Ergebnisses hat. Bei unterschiedlichen Untersuchungen müssen bei ansonst gleichen Bedingungen auch gleiche Ergebnisse auftreten. Die Objektivität eines Verfahrens lässt sich weiter unterscheiden in **Durchführungs-Objektivität** (betrifft den standardisierten Ablauf der Erhebung), **Interprete-Objektivität** (betrifft die Klassifizierung der Angaben aufgrund vorher festgelegter Regeln) und **Auswerte-Objektivität** (betr. die standardisierte Auswertung nach einem vorab festgelegten Ablauf).

Die **Reliabilität** oder **Zuverlässigkeit** eines Verfahrens gibt den Grad der Exaktheit der Messung an, unabhängig davon, was tatsächlich gemessen wird. Die Reliabilität beschreibt die Stabilität und Genauigkeit der Messungen sowie die Konstanz der Messbedingungen, z.B. bei der Erhebung der Patientenzufriedenheit über einen Fragebogen oder physiologische Messungen von Blutdruck oder Cholesterin. Die Wiederholung eines Tests soll bei stabilen Merkmalen gleiche Ergebnisse erbringen und damit weitgehend unabhängig von Zufallsschwankungen und Umweltbedingungen sein. Hohe Reliabilität ist ein Faktor hoher Objektivität.

Die Reliabilität eines Tests kann überprüft werden durch:

- einen **Retest**, der Aussagen über die Stabilität der Testergebnisse zu verschiedenen Zeitpunkten macht;
- eine **Testhalbierung** (Split – half) und die Messung der inneren Konsistenz einzelner Teile bzw. zweier Hälften eines Instruments;
- einen **Paralleltest**, also einer Erhebung gleichzeitig mit einem äquivalenten Instrument; z.B. strukturiertes Interview zur Erhebung der Patientenzufriedenheit.

Die **Validität** oder **Gültigkeit** eines Verfahrens schließlich gibt an, inwieweit es das misst, was es zu messen vorgibt. Es ist z.B. denkbar, dass die Zufriedenheit der Patienten nicht Ausdruck der Qualität der Behandlung ist, sondern ihres Gesundheitszustandes. Oder ein Intelligenztests misst statt der Intelligenz möglicherweise die Fähigkeit unter Zeitdruck ungewohnte Aufgaben zu bewältigen.

Die Validität kann überprüft werden durch:

- **Außenkriterium** (Übereinstimmung mit einem Standard; z.B. Gespräche mit Pflegekräften, Angehörigen),
- **Extremgruppen** (Vergleich der Ergebnisse bei stark unterschiedlichen Patientengruppen)
- **Vorhersage** (Ist die Wahl des Krankenhaus bei zukünftigen Behandlungen abhängig von der Patientenzufriedenheit bei vorhergehenden Behandlungen?)
- **Inhaltsvalidität** (Ist an der Art der Fragen der Zweck und inhaltliche Bezug erkennbar?)
- **Konstruktvalidität** (Geht der Fragebogen auf bekannte Aspekte der Patientenzufriedenheit ein?)

Validität und Reliabilität sind im übrigen unabhängig voneinander, wie die folgende Abbildung 4.1 verdeutlicht.

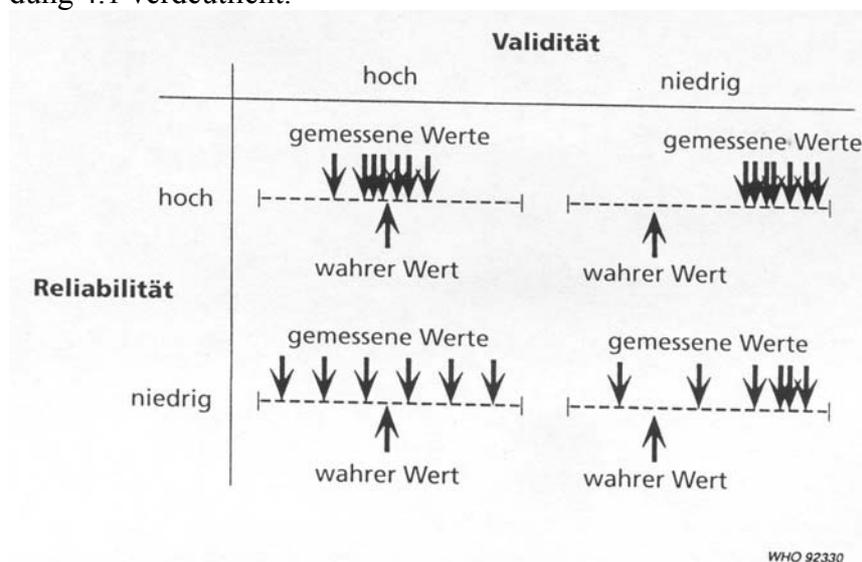


Abbildung 4.1: Bedeutung von Validität und Reliabilität

(Quelle: Beaglehole et al 1997)

4.2 Aspekte der Befragung

In den Sozialwissenschaften, aber auch in der Epidemiologie, bestehen viele Untersuchungen aus Befragungen der Untersuchungspersonen, zum Gesundheitszustand, zu Krankheiten, (gesundheitsbezogenen) Verhaltensweisen, Einstellungen u.v.m. Daher ist bei der Formulierung von Fragen und ihrer Anordnung in einem Interview oder einem Fragebogen auf verschiedene Aspekte zu achten. Die wichtigsten seien hier nur kurz angerissen.

Bei der **Formulierung von Fragen** sind drei Punkte zu klären:

- Warum wird die Frage gestellt?
- Welche Art von Frage (und Antwortvorgaben) ist angemessen?
- Wie ist die Frage zu formulieren?

Die Frage ist das Bindeglied zwischen den Variablen der Hypothesen und den Antworten. Von der Operationalisierung der Hypothesen und der Frage(n)formulierung hängt also entscheidend die spätere Überprüfung der Hypothesen ab. Die Hypothesen sollten sich in den Fragen widerspiegeln, andererseits sollte keine Frage ohne eine Verankerung in der inhaltlichen Fragestellung gestellt werden.

Fragen können grundsätzlich in offene und geschlossene unterteilt werden. **Offene Fragen** enthalten keine Antwortvorgaben, **geschlossene Fragen** zwei oder mehr Vorgaben. Je größer das Vorwissen über Meinungen, ihre Strukturierung und Informationsstand der Befragten ist, desto eher bieten sich geschlossene Fragen an.

Offene Fragen sollten Verwendung finden, wenn

- die Informationen über die Einstellungen von Personen zu einem gegebenen Problem gering ist,
- der Bezugsrahmen des Befragten ermittelt werden soll,
- differenzierte Einstellungen zur Formulierung von Hypothesen ermittelt werden sollen

Geschlossene Fragen bieten sich für eine Vielzahl von Situationen an, wie z.B. Fragen mit zwei oder mehr Alternativen, die Erhebung von Meinungen, Einstellungen oder Verhaltensweisen, die Bewertung hypothetischer Situationen, die Formulierung als Kontrollfragen oder als indirekte Frage. Schwierig können sich Fragen nach Meinungslosigkeit oder unangenehmen bzw. intimen Bereichen gestalten (Sexualität, Einkommen, Suchtverhalten, ...). Schließlich nehmen in den meisten Umfragen Fragen zu soziodemographischen Merkmalen (wie Schul- und Berufsausbildung, Berufstätigkeit) eine wichtige Bedeutung ein.

Bei der Klassifikation der Antwortmöglichkeiten geschlossener Fragen sollte darauf geachtet werden, dass diese eindeutig und vollständig sind, sich mögliche Antworten also nicht überschneiden und alle mögliche Antworten umfassen (incl. einer Kategorie ‚weiß nicht‘ o.ä.). Die Eigenschaft der Ausschließlichkeit, also nur eine Antwortkategorie ist zulässig, wird nicht bei allen geschlossenen Fragen sinnvoll sein (etwa bei Fragen nach Gründen für ein Verhalten). Als Grundsatz gilt: Jeder Befragte muss sich bei geschlossenen Fragen ohne Schwierigkeiten einordnen können.

Weitere **Aspekte der Frageformulierung** umfassen folgende Punkte:

- Fragen sollten kurz, einfach und auf den Bezugsrahmen des Befragten bezogen sein.
- Den Fragen oder zumindest Fragenkomplexen sollten Erklärungen als Satz vorangestellt werden.
- Gründe für eine Antwort können interessant sein (kann geschlossen sein, wenn die möglichen Gründe weitgehend bekannt sind).
- Es sollten leicht verständliche und eindeutige Begriffe verwendet werden. Der Informationsstand der Befragten muss berücksichtigt werden, ohne zuviel voraus zu setzen.
- Die Frage sollte nicht mehrere Dimensionen abbilden, sondern nur jeweils einzelne Sachverhalte
- Die Abfolge der Fragen spielt eine Rolle.
- Bei kritischen Fragen kann sich der Einbau von Kontrollfragen anbieten.
- Negativ belegte Sachverhalte sollten nicht direkt erfragt werden.
- Verzerrungen oder suggestive Fragen sind zu vermeiden.

Aus wissenschaftlichen Untersuchungen lassen sich folgende **Regeln des Fragens** ableiten:

- Je weiter ein Ereignis zurückliegt, desto ungenauer werden die Antworten.
- Je mehr sich eine Person für ein Thema interessiert, desto gültiger sind die Angaben.
- Je wichtiger ein Ereignis für eine Person ist, desto genauer werden die Angaben.
- Je bedrohlicher ein Ereignis für eine Person war, desto eher wird sie es vergessen.
- Je sozial missbilligter etwas ist (z.B. Geistes- oder Geschlechtskrankheiten), desto seltener wird darüber berichtet.

- Je sozial höher etwas bewertet ist (z.B. Arbeitsplatz, Einkommen, Spenden, gelesene Bücher), desto eher sind die Angaben zu hoch.
- Je eingehender jemand seine Gründe durchdacht hat, desto weniger wird die vorgegebene Liste bei einer geschlossenen Frage ausreichen.
- Je weniger jemand seine Gründe durchdacht hat, desto eher wird die Liste als Erleichterung angesehen werden. Bei sehr niedrigem Informations- und/oder Reflexionsstand besteht allerdings die Gefahr wahlloser Ankreuzungen.
- Unangenehme oder sozial missbilligte Gründe werden eher in geschlossenen Fragen genannt.

Häufig werden innerhalb einer Befragung aus mehreren Einzelfragen zusammengesetzte Indizes berechnet. Während die Einzelitems spezifische Dimensionen einer Eigenschaft oder Verhaltens abbilden sollen, versuchen Indizes die Vielschichtigkeit eben dieser Eigenschaft oder dieses Verhaltens in einer komprimierten Maßzahl abzubilden. Zum Beispiel kann in einem Index ‚Grad der Behinderung‘ die einzeln abzufragenden Fähigkeiten Einkaufen, sich selbstständig anzuziehen, allein Essen zuzubereiten usw. zusammengefasst werden.

Die Indexbildung versucht die Komplexität gewisser Merkmale über eindimensionale Einzelitems zu reduzieren und diese anschließend in einem komprimierten Indexwert anschaulich und für weitere Auswertungen praktikabel darzustellen. Dieses Vorgehen bedarf großer Sorgfalt, um einerseits alle Dimensionen des Merkmals zu erfassen, andererseits eine angemessene Gewichtung der verschiedenen Einzelitems vorzunehmen. Nach Möglichkeit sollten in der Sozialwissenschaft validierte und reliable Indizes Verwendung finden (unter Wahrung des Urheberrechts). Für eine Vielzahl von gesundheitsbezogenen Fragestellungen gibt es bereits bezüglich ihrer Testeigenschaften geprüfte Instrumente (z.B. für Hilfebedarf, Lebensqualität, psychische und seelische Befindlichkeit, soziale Integration u.v.m.).

4.3 Interview und Befragung

Eine Befragung kann als (standardisiertes, mündliches) Interview oder schriftlich über einen Fragebogen durchgeführt werden. Die Abgrenzungen werden fließend, wenn man an die zunehmend häufiger werdenden telefonischen Interviews denkt, bei denen ein persönlicher Kontakt zwischen Interviewer und Befragtem nur indirekt hergestellt wird. Welche von beiden Formen für eine Befragung die geeignetere ist, muss aus der Fragestellung und der Zielpopulation heraus entschieden werden.

Ein Interview erfordert eine umfassende Schulung der Interviewer und einen schriftlichen Interviewleitfaden, um eine hohe Akzeptanz bei den Befragten herzustellen und unverzerrte Antworten zu erhalten. Dafür ermöglicht ein Interview die Beantwortung von Rückfragen seitens der Befragten und die unmittelbare Klärung unklarer Angaben.

Eine schriftliche Befragung stellt hohe Anforderungen an die Verständlichkeit des Aufbaus und der Formulierung einzelner Fragen. Die Erreichung einer Mindestbeteiligung setzt intensivere Maßnahmen zur Sicherstellung der Akzeptanz und Vermeidung unnötigen Aufwands bei den Befragten voraus (Anschreiben, Darlegung der Gründe für die Befragung, Freiumschlag für Rücksendung, ...).

In der Epidemiologie werden vor allem in größeren Querschnitt- und Längsschnittstudien schriftliche Befragungen durchgeführt, da sie bei größeren Einzugsbereichen und/oder hohen Fallzahlen und bei Verwendung standardisierter und validierter Fragebögen einen geringeren Aufwand als ein Interview erfordern.

Auf die anderen (quantitativen und qualitativen) Erhebungsverfahren in den Sozialwissenschaften, kann an dieser Stelle nicht weiter eingegangen werden.

4.4 Beurteilungsfehler

Neben den nicht zu vermeidenden zufälligen Fehlern können in einer Befragung, sei es als Interview, als schriftliche Befragung oder in einem anderem Erhebungsverfahren der Sozialwissenschaften, eine Reihe von **systematischen Beurteilungsfehlern** auftreten. Die Gefahr für diese Fehler ist dabei in einem Interview oder einer direkten Beobachtung höher als in einer anonymisierten Erhebung wie einer schriftlichen Befragung. Zu den potenziellen Fehlern gehören im einzelnen gehören:

- **Rosenthal-(Pigmalion-)Effekt** beschreibt eine Veränderung der Leistung oder von Eigenschaften aufgrund der Einstellung des Beobachters oder des Versuchsleiters;
- **Hawthorne-Effekt** bezeichnet eine Veränderung allein aufgrund der Tatsache, dass man beobachtet wird bzw. an einer Studie teilnimmt;
- **Halo-Effekt**: die Beurteilung eines herausragenden Merkmals beeinflusst die Bewertung in anderen Dimensionen;
- **logischer Fehler** meint die fehlerhafte Übertragung einer Eigenschaft auf eine andere, z.B. Ordentlichkeit und Fleiß;
- **Kontrast-Fehler**: mangels validem Vergleichsmaßstab wird die eigene Person oder das eigene Umfeld als Standard betrachtet;
- **Strenge-/Mildefehler**: bekannte Personen werden möglicherweise systematisch milder oder auch strenger beurteilt;
- **Tendenz zur Mitte** bzw. zu Extremwerten ist häufig bei unsicheren bzw. selbstsicheren Personen anzutreffen;
- **Ja-Sage-Tendenz**: Fragen nach persönliche Eigenschaften werden eher bejaht als verneint;
- **Soziale Erwünschtheit** beschreibt ein Antwortverhalten im Sinne unterstellter sozialer Erwünschtheit;
- **Suggestion** kann durch den Interviewer oder Versuchsleiter erfolgen oder auch durch Autosuggestion;
- **Projektion** beschreibt eine Situationen, in der beim Beurteiler vorhandene Eigenschaften verstärkt beim Gegenüber wahrgenommen werden.

Diese potenziellen Untersuchungsfehler müssen bei der Planung einer Befragung gezielt überprüft und durch Beachtung der eingangs aufgestellten Regeln weitgehend vermieden werden. Dies kann zusätzlich auch im Rahmen eines **Pretests** erfolgen.

5. Datenquellen für die Epidemiologie

Epidemiologische Untersuchungen können sich bei der Datenerhebung grundsätzlich zweier Möglichkeiten bedienen. Es kann unterschieden werden zwischen:

Primärerhebung Für eine konkrete (epidemiologische, sozialwissenschaftliche) Fragestellung werden gezielt und speziell für diesen Zweck über ein geeignetes Erhebungsinstrument (z.B. Fragebogen) Daten gewonnen;

oder

Sekundärerhebung Bei der Untersuchung wird alternativ oder ergänzend zu einer Primärerhebung auf vorhandene Datenquellen, die zu einem anderen Zweck erhoben wurden, zurückgegriffen, z.B. auf Daten der amtliche Statistik oder der gesetzlichen Krankenkassen.

Gewissermaßen in der Mitte stehen Datensätze, die für unspezifische oder sehr breit angelegte Forschungszwecke erhoben wurden und nach Aufbereitung und dem Abschluss von gewissen Grundauswertungen der allgemeinen wissenschaftlichen Öffentlichkeit als sog. **Public-Use-Files** für weitergehende Forschungsprojekte zur Verfügung gestellt werden. Ein aktuelles Beispiel hierfür ist der Bundesgesundheitsurvey 1998, dessen geprüfter Datensatz durch das Robert-Koch-Institut (www.rki.de) für Wissenschaft und Lehre zur Verfügung gestellt wird. Auch die Daten des telefonischen Gesundheitsurvey aus dem Jahr 2003 liegen inzwischen als Public-Use-File vor; für den 2006 abgeschlossenen Kinder- und Jugendsurvey (www.kiggs.de) ist es vorgesehen

Bei der Datenbeschaffung und –interpretation muss weiter unterschieden werden zwischen

Individualdaten Die Daten werden an einzelnen Personen, den Untersuchungseinheiten, erhoben, z.B. zum Gesundheitszustand einzelner Personen oder zur Patientenzufriedenheit;

oder

Aggregatdaten Daten werden für Gruppen von Untersuchungseinheiten erhoben bzw. Individualdaten zur Auswertung zusammengefasst, z. B. Fragen zu einem Haushalt mit mehreren Personen (Größe, Einkommen) oder Kennzahlen eines einzelnen Krankenhauses (Patienten, Mitarbeiter). Bei der Mehrzahl der öffentlich zugänglichen Daten handelt es sich um Aggregatdaten.

Die deskriptive Epidemiologie verwendet zahlreiche heterogene Datenquellen für die Darstellung eines Gesamtbild des Gesundheitszustands der Bevölkerung ("Gesundheitsberichterstattung"). Verknüpfungen verschiedener Datenquellen können u.a. über die Aggregierungsmerkmale "6P" vorgenommen werden (period, place, provider, problem, procedure, person), also **Berichtsperiode** (,period'; Zeit), **Ort** (,place'), **Leistungserbringer** (,provider'; z.B. Arztgruppe, Krankenhaus), **Problem** (,problem'; Diagnose, "ICD"), **Maßnahmen** (,procedure'; Verfahren, Operation, "ICPM"), **Person** (,Person'; Individuum). Es ist daher wichtig, für diese 6 P in allen Bereichen des Gesundheitswesens möglichst gemeinsame Definitionen und Klassifikationen zu verwenden. Der Datenschutz ist zu berücksichtigen. Leider ist diese Forderung bislang nur teilweise umgesetzt.

Die Breite der für die Epidemiologie relevanten Daten verdeutlicht das Inhaltsverzeichnis des 1998 erschienenen ersten Gesundheitsberichts für Deutschland (Tabelle 5.1), der im übrigen auch im Internet auf der Homepage der Gesundheitsberichterstattung mit laufend aktualisierten Daten verfügbar ist (www.gbe-bund.de). Eine erste grundlegende Überarbeitung wurde 2006 publiziert (download über www.rki.de). Inhaltlich ähnliche, wenngleich nicht immer so

detaillierte, Gesundheitsberichte gibt es inzwischen für die meisten Bundesländer, die regelmäßig im Abstand von wenigen Jahren erscheinen. In Sachsen-Anhalt erschien der letzte ‚traditionelle‘ Gesundheitsbericht in der zweiten Jahreshälfte 2003. Im Jahr 2005 wurde der 5. Gesundheitsbericht auf der Basis eines umfangreichen Bevölkerungssurveys bei 4.000 Einwohnern Sachsen-Anhalts veröffentlicht (download über die Homepage des Landessozialministeriums; Adressen s.u.).

Tabelle 5.1: Inhaltsverzeichnis des Gesundheitsberichts für Deutschland (abgekürzt)

<p>Rahmenbedingungen des Gesundheitswesens: politisch und institutionell, demographisch und ökonomisch</p> <p>gesundheitliche Lage: Länge des Lebens (Lebenserwartung und Sterblichkeit), individuelle Aspekte von Gesundheit, Krankheitsfolgen (Behinderungen, Arbeitsunfähigkeit, Frühberentung, Pflegebedürftigkeit)</p> <p>Gesundheitsverhalten und Gesundheitsgefährdungen: gesundheitsbeeinflussende Lebensweisen (Ernährung, Rauchen, Alkohol, Drogen), Risikomerkmale der sozialen Lage, der Umwelt und der Arbeitswelt (z.B. Einkommen, Arbeitslosigkeit; Luft, Lärm; Berufskrankheiten, Arbeitsbelastungen)</p> <p>Krankheiten: Herzkreislauferkrankungen, Bösartige Neubildungen, Krankheiten des Muskel-Skelettsystems, der Atmungs-, Verdauungs-, Harn-, Geschlechtsorgane, psychiatrische Krankheiten, Allergien, Infektionskrankheiten, Diabetes</p> <p>Ressourcen der Gesundheitsversorgung: ambulante Versorgung (Praxen, Apotheken, Pflegeeinrichtungen), stationäre Versorgung (Krankenhäuser, Vorsorge- und Reha-Einrichtungen), Erwerbstätige</p> <p>Leistungen und Inanspruchnahme des Systems: Vorsorge, Früherkennung, ambulante Versorgung (incl. Arzneimittel), stationäre Versorgung, einzelne Bereiche (Krankenpflege, ambulantes Operieren)</p> <p>Ausgaben, Kosten, Finanzierung des Systems: Ausgaben und Kosten nach Leistungsarten, Sektoren, Einrichtungen, Krankheitsarten; Einkommen und Preise, Finanzierung</p>
--

Die Daten in diesen Gesundheitsberichten stammen aus verschiedenen Bereichen der amtlichen und nichtamtlichen Statistik (z.B. Daten der Leistungserbringer und Kostenträger) und aus epidemiologischen Untersuchungen. Derartige Daten können bei der Konzipierung einer Studie verwendet werden, etwa zur Abschätzung zu erwartender Krankheitshäufigkeiten, oder für die Einordnung eigener Ergebnisse. Außerdem dienen sie der Beschreibung gesundheitspolitischer Strukturen, vor deren Hintergrund sich Gesundheit, Krankheit und medizinische Inanspruchnahme abspielen.

Ergänzt werden diese breiten und allgemein gehaltenen Gesundheitsberichte durch Spezialberichte. So erschien 2001 der erste Frauengesundheitsbericht für Deutschland⁴ und das Robert-Koch-Institut Berlin gibt seit 1999 mit Hilfe externer Fachleute mehrmals jährlich Spezial- und Schwerpunktberichte zur Gesundheitsberichterstattung heraus, z.B. zu chronischen Schmerzen, medizinischen Behandlungsfehlern oder Schutzimpfungen, um nur einige Themen der letzten Jahre zu nennen (kostenloser download aller Hefte über www.rki.de). Inzwischen gibt es beinahe 40 dieser Hefte, die auch eine regelmäßige Überarbeitung erfahren.

Ausgangspunkt für die stetig wachsende Zahl an Gesundheitsberichten war die Kritik an den verfügbaren Daten und Analysen im Gesundheitsbereich:

- Fehlende Information zu wichtigen Krankheitsbereichen,
- Mangelnde systematische Verknüpfung des vorliegenden Datenmaterials,
- Willkürliche und nicht analytische Präsentation der Informationen,
- Nicht ausreichende Validität der Daten,
- Unzureichender Datenzugang für externe Nutzer.

⁴ Der erste Gesundheitsbericht für Männer wurde 1999 von der Stadt Wien herausgegeben. (kostenloser download über <http://www.wien.gv.at/who/manngb/99/>).

Zwar sind in den letzten Jahren einige dieser Defizite abgebaut worden, doch gibt es immer noch Lücken in der empirischen Datenbasis in wichtigen Versorgungsbereichen. So fehlen besonders für den ambulanten Bereich Informationen zum Leistungsgeschehen, die auch über Routinedaten der Kassen oder der Ärzte bislang nicht abrufbar sind. Hier herrscht weitgehend Unwissenheit über das, was täglich in den Praxen niedergelassener Ärzte geschieht. Mit der Umsetzung des § 295 SGB V ab 2004 wird zukünftig eine aussagefähige Datenbasis analog zum stationären Sektor zur Verfügung stehen. Weitere Defizite der Datenlage betreffen fehlende Schnittstellen zwischen den Leistungssektoren, oftmals geringe Disaggregation der Daten (nach Alter, Geschlecht, Region, Leistungserbringern oder weiteren soziodemographischen Merkmalen), uneinheitliche Klassifizierungssysteme und Auflösungen der Daten oder die ungenügende Aktualisierung der Daten (z.B. ambulante Versorgung, Rehabilitation) und in der ausreichenden Auflösung nach wichtigen Determinanten der Gesundheit, etwa nach Regionen oder sozialer Schicht. Die inhaltliche und methodische Weiterentwicklung der Gesundheitsberichterstattung bleibt daher auch zukünftig ein wichtiges Ziel der Versorgungsforschung.

Bei der Nutzung dieser und anderer Sekundärdaten muss immer berücksichtigt werden, dass sie für einen spezifischen Zweck und unter je eigenen Bedingungen erhoben worden sind. Es entsteht dadurch die Notwendigkeit, im Einzelfall die Qualität der Daten und die Relevanz für die eigene Fragestellung zu prüfen. Das betrifft u.a. die Herkunft der Daten, die Definition der Dateninhalte, die Ein- und Ausschlusskriterien, die Validität und Reliabilität der Angaben, die Grundgesamtheit bzw. Bezugspopulation, die Genauigkeit und Vollständigkeit der Erfassung, die Aktualisierung, die Auflösung nach Zeit und Raum und viele Aspekte mehr.

Es ist nicht der Raum, an dieser Stelle im einzelnen auf die aussagefähigen Datenquellen für das Gesundheitswesen einzugehen. Einige seien hier nur stichwortartig genannt (Tabelle 5.2). Genauere Informationen enthält man über die Gesundheitsberichterstattung des Bundes oder über die anderen Datenherren (vgl. Tabelle 5.3 am Ende; CAVE: die Adressen können sich schnell ändern, sind daher möglicherweise nicht auf dem neuesten Stand).

Tabelle 5.2: Häufig verwendete Datenquellen für die Epidemiologie

Art der Statistik	Inhalte	Datenherr
Todesursachenstatistik	Todesfälle nach Todesursache	Gesundheitsamt, Statistisches Landesamt
Krankenhausdiagnosestatistik	Krankenhausentlassungsfälle	Statistisches Landes- und Bundesamt
meldepflichtige Krankheiten	ausgewählte infektiöse Krankheiten	Gesundheitsamt
Schuleingangsuntersuchungen	Entwicklungsstörungen, behandlungsnotwendige Krankheiten	Gesundheitsamt
Musterungsuntersuchungen	Ergebnisse der Untersuchungen der Wehrpflichtigen	Kreiswehrrersatzämter
Vorsorge- und Früherkennungsuntersuchungen	Ergebnisse der U1 bis U9-Untersuchungen bei Klein- und Vorschulkindern	KBV
Daten zum Krebsgeschehen	Inzidenz, Mortalität	Krebsregister auf Landesebene, Dt. Kinderkrebsregister, Dach-Dokumentation Krebs
Routinedaten der Krankenkassen	Arbeitsunfähigkeit, stationäre Inanspruchnahme	einzelne Kassenkassen

Erwähnt seien neben den spezifischen Gesundheitsberichten auch eine Reihe anderer regelmäßig veröffentlichter Sozialberichte der zuständigen Bundes- und Landesministerien (Kinder- und Jugend, Sozial, Armut, Alten, Familien), die meist auch auf die gesundheitlichen

Auswirkungen der in den Berichten geschilderten sozialen Lagen eingehen. Ergänzend, zum Teil auch inhaltlich gegensätzlich, sind auch Sozialberichte paritätischer Organisation lesenswert (z.B. der Armutsbericht des DGB, der Hans-Böckler-Stiftung und des Paritätischen Wohlfahrtsverbandes)

Für Daten im internationalen Kontext bieten sich Links des statistischen Bundesamtes (www.destatis.de) zu verschiedenen Staaten an, der Informationsdienst der Organisation für wirtschaftliche Zusammenarbeit (www.oecd.org), die auch regelmäßig eine CD mit einer Vielzahl von Gesundheitsindikatoren anbietet, oder die Weltgesundheitsorganisation WHO an (für Europa unter www.who.dk).

Für die wissenschaftliche Sekundärnutzung stehen schließlich repräsentative Bevölkerungsumfragen wie aus der Deutschen-Herz-Kreislauf-Präventionsstudie und dem Bundesgesundheitsurvey 1998 und Ergebnisse der jährlichen Mikrozensuserhebungen für eine geringe Aufwandsentschädigung zur Verfügung (Kontakt über das Robert-Koch-Institut; www.rki.de).

Tabelle 5.3: Mögliche weiterführende Datenquellen für die Epidemiologie

Gesundheitsberichterstattung des Bundes, der Länder und der Kommunen	
Gesundheitsbericht für Deutschland (ab 1998)	http://www.gbe-bund.de/ - laufend aktualisiert
Bundesgesundheitsberichterstattung	mehrere thematische abgegrenzte Kurzberichte jährlich Frauengesundheitsbericht; erstmalig 2001
Robert-Koch-Institut, Berlin	http://www.rki.de/
Gesundheitsbericht Sachsen-Anhalt	alle 3-4 Jahre; http://www.sachsen-anhalt.de/
Gesundheitsbericht der Stadt Magdeburg	über http://www.magdeburg.de (Verwaltung/Politik)
Datenquellen der amtlichen Statistik auf Bundes-, Landes- und kommunaler Ebene: z.B.	
Statistisches Jahrbuch der Bundesrepublik Deutschland	und Sachsen-Anhalt (jeweils jährlich)
Statistische Blätter der Stadt Magdeburg	über http://www.magdeburg.de (Amt für Statistik)
Statistisches Bundesamt:	http://www.destatis.de/
Statistisches Landesamt	http://www.statistik.sachsen-anhalt.de/
Statistisches Amt der Stadt Magdeburg	über http://www.magdeburg.de (Verwaltung/Politik)
Kostenträger und Leistungserbringer:	
Spitzenverband der Krankenkassen	http://www.g-k-v.com/
Betriebskrankenkassen	http://www.bkk.de/
AOK	http://www.aok.de/bundesverband
Verb. d. Angestellten-, Arbeiter-Ersatzkassen (VdAK)	http://www.vdak.de/
Deutsches Ärztenetz (mit vielen Links)	http://www.arzt.de/
Kassenärztliche Bundesvereinigung	http://www.kbv.de/
Bundesärztekammer	http://www.bundesaerztekammer.de
Deutsche Krankenhausgesellschaft	http://www.dkgev.de/
Bei Behörden und Ministerien sind erfahrungsgemäß wenig aussagefähige Datenquellen verfügbar und dann häufig bereits mehrere Jahre alt.	
Bundesministerium für Gesundheit und Soziale Sicherung (BMGS)	http://www.bmg.bund.de
Sachverständigenrat des BMG	http://www.svr-gesundheit.de
Sozialministerium Sachsen-Anhalt	http://www.ms-sachsen-anhalt.de
Gesundheitsamt der Stadt Magdeburg	http://www.magdeburg.de/Stadt/AMT53.html

Zusammenfassend lässt sich der derzeitige Stand der Gesundheitsberichterstattung in Deutschland wie folgt charakterisieren:

- Es gibt eine Vielzahl von nebeneinander existierender Datenquellen verschiedener Datenherren, auch zum gleichen Thema. Eine ‚beste‘ oder umfassende Datenquelle gibt es nicht.

- Die Quellen unterscheiden sich im Umfang der dargestellten Indikatoren, im Grad der Kommentierung und in der Feinheit der Auflösung. Die Beurteilung der Datenqualität kann anhand bestimmter Kriterien beurteilt werden.
- Es gibt eine steigende Anzahl Landes- und kommunaler Gesundheitsberichte. Die meisten orientieren sich am 3. konsentierten Indikatorenkatalog der Arbeitsgemeinschaft der Obersten Landesgesundheitsbehörden (AOLG) aus dem Jahr 2003 mit insgesamt 297 Einzelindikatoren (download z.B. über die Seite des Landesinstituts für den öffentlichen Gesundheitsdienst NRW; www.loegd.de). Eine bessere Verknüpfung mit anderen Berichten wird angestrebt (z.B. Armut-, Sozial-, Alten-, Jugendberichterstattung).
- Es werden zunehmend spezielle Gesundheitsberichte für Teilgruppen der Bevölkerung und besondere Gesundheitsprobleme vorgelegt.

6. Maßzahlen der Epidemiologie

6.1 Maßzahlen der Morbidität und Mortalität

Epidemiologische Kennziffern lassen sich grundsätzlich in **Proportionen**, also Anteile, und **Raten** unterscheiden. Proportionen sind Brüche. In deren Zähler steht eine Teilmenge der Gesamtmenge, die wiederum den Nenner ausmacht. **Proportionen** nehmen damit Werte zwischen Null und Eins an und werden der besseren Anschaulichkeit wegen häufig in Prozent oder Promille angegeben. Beispiele für Proportionen sind:

- die **Säuglingssterblichkeit**, die die Zahl der im ersten Lebensjahr gestorbenen Säuglinge zu allen Lebensgeburten angibt,
- die **proportionale Mortalität**, die als der Anteil der an einer bestimmten Krankheit oder Krankheitsartengruppe Verstorbenen an allen Verstorbenen definiert ist,
- die **Sexualproportion**, die das Verhältnis männlicher Babies zu allen Neugeborenen darstellt.

In der Epidemiologie sind Proportionen weniger gebräuchlich, dass sie keinen direkten Bevölkerungsbezug haben. Raten im Gegensatz dazu haben eben diesen, sie geben Aufschluss über Häufigkeiten gesundheitsbezogener Ereignisse. Im *Zähler* enthalten sie das interessierende Ereignis, z.B. Arztkontakte, Neuerkrankungen, Todesfälle oder Geburten. Der *Nenner* enthält die Zahl der Personen unter ‚*Risiko*‘, bei denen das Ereignis eintreten kann. Das kann die Gesamtheit einer Bevölkerung sein oder nur eine Teilgruppe, z.B. alle Frauen im gebärfähigen Alter für das Ereignis ‚*Geburt*‘. Diese Abgrenzung ist im Einzelfall nicht immer einfach, so müssten für die Darstellung der Geburtenrate alle Frauen unter 45 Jahre mit Gebärmutterentfernungen (Hysterektomien) abgezogen werden, was in der Praxis aber nicht durchführbar ist. Wichtig ist zusätzlich die Angabe einer Zeiteinheit, auf die sich diese Angaben beziehen. Üblicherweise werden bei seltenen Ereignissen diese Brüche mit 100, 1000, 10000 oder ähnlichen Faktoren multipliziert, um aussagefähigere Zahlen zu erhalten.

Im folgenden werden zunächst die gebräuchlichsten epidemiologischen Maßzahlen für Morbidität und Mortalität, also die Krankheitshäufigkeit und die Sterblichkeit, vorgestellt. Die Definitionen der Maßzahlen und ihre Interpretationen können aber auf beliebige Ereignisse übertragen werden. Das können z.B. Kontaktanlässe in der ambulanten Versorgung, Krankenhausaufenthalte, Arbeitsunfähigkeitsepisoden, Reha-Maßnahmen o.ä. sein.

Die **Prävalenz** ist definiert als die Anzahl der Personen, die zu einem bestimmten Zeitpunkt erkrankt sind, geteilt durch die Gesamtzahl der Bevölkerung (**Punktprävalenz**). Wenn statt eines Stichtages ein begrenzter Zeitraum als Bezugsgröße genommen wird, z.B. ein Jahr, spricht man von **Periodenprävalenz** und zählt im Zähler alle Personen, die innerhalb diese Zeitraums mindestens einmal erkrankt waren.

Abbildung 6.1 zeigt, welche Faktoren die Prävalenzrate beeinflussen. Diese sind beim Vergleich von Prävalenzraten verschiedener Populationen oder verschiedener Zeiträume zu berücksichtigen. So steigt z.B. bei der AOK Sachsen-Anhalt die Krankenhaushäufigkeit bei chronischen Krankheiten, also die 1-Jahres-Prävalenz einer Krankenhausaufnahme, gemessen etwa als Fälle pro 1000 Versicherte, weil gesunde jüngere Versicherte (sog. ‚gute Risiken‘) verstärkt in andere Kassen oder die private Krankenversicherung wechseln. Andererseits kann es zu einer Abnahme der Krankenhaushäufigkeit durch Verlagerung stationärer Leistungen kommen, was einer Abnahme der Inzidenz entspricht.

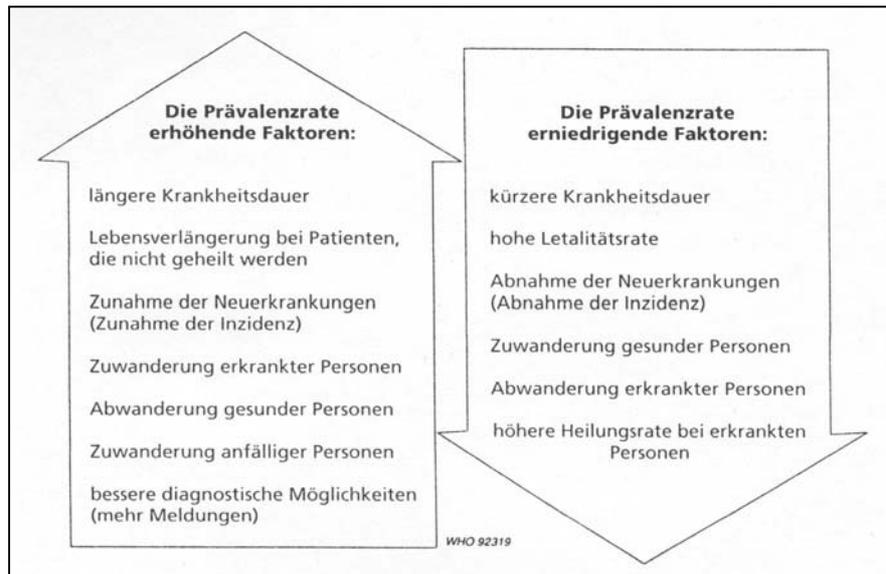


Abb. 6.1: Die Prävalenz beeinflussende Faktoren (Quelle: Beaglehole et al. 1997)

Die **Inzidenz** ist definiert als die Anzahl - in einem definierten Zeitraum - neu aufgetretener Krankheitsfälle, geteilt durch die Anzahl der Personen mit Erkrankungsrisiko. In der Epidemiologie stehen häufig nicht alle Personen im gesamten Zeitraum unter Risiko, z.B. weil sie aus der Beobachtung ausscheiden, sterben, wegziehen, die Krankenkasse wechseln o.ä. In solchen Fällen mit dynamischen und sich verändernden Risikobevölkerungen sollte im Nenner die Summe der exakten Zeiten unter Risiko stehen, sog. **Personenmonate** oder **Personenjahre**. Zwei Versicherte der AOK, die jeweils zur Mitte des Jahres die Kasse wechseln, tragen damit genauso ein Personenjahr im Nenner bei wie ein Versicherter, der ein ganzes Kalenderjahr bei der AOK krankenversichert war. Die entstehende Rate wird dann als **Inzidenzdichte** bezeichnet. Werden Personen über einen längeren Zeitraum beobachtet und die Ereigniswahrscheinlichkeit über die gesamte Zeitraum bestimmt, spricht man von kumulativer Inzidenz. So beträgt die kumulative Inzidenz für Mammakarzinom derzeit rund 7%, d.h. statistisch wird jede 15. Frau im Laufe ihres Lebens an Brustkrebs erkranken.

Bei einem Fließgleichgewicht, also einer annähernd gleichen Zahl von Neuerkrankungen und Zahl der Erkrankten, die gesund werden oder versterben, sowie einer geringen absoluten Prävalenz (< 10 %) gilt die Beziehung

$$\text{Prävalenz} = \text{Inzidenz} * \text{durchschnittliche Krankheitsdauer} \quad (6:1)$$

Bezüglich der Mortalität kann als einfachstes Maß der Anteil der Todesfälle an einer bestimmten Krankheit oder Krankheitsartengruppe an der Gesamtheit aller Todesfälle bestimmt werden (**proportionale Mortalität**, s.o.). Diese sagt aber nichts über das Mortalitätsrisiko aus. Daher sind auch für die Mortalität Raten gebräuchlich. Die **Mortalitätsziffer** (Sterblichkeitsziffer) teilt die Gesamtzahl der Verstorbenen eines Jahres durch die durchschnittliche Bevölkerungszahl oder Bevölkerung zur Jahresmitte.

Problematisch stellt sich für die meisten Krankheiten die Ermittlung von Morbiditätsziffern dar, da die Zahl der interessierenden Ereignisse nur in Ausnahmefällen einigermaßen exakt bestimmt werden kann. Dies ist möglich für die meldepflichtigen Infektions- und sexuell übertragbaren Krankheiten, die verschiedenen Krebslokalisationen aus den Meldungen der Landes- oder Bundeskrebsregister oder Missbildungen aus den wenigen deutschen Missbildungsregistern (eins davon befindet sich an der Otto-von-Guericke-Universität). Für diese Krankheiten ist eine Vollerhebung weitgehend gegeben und ein Bevölkerungsbezug herstellbar. Das gilt nicht mehr für einige Krankheitsregister (Herzinfarkt oder Schlaganfall) und

noch viel weniger für die Vielzahl von Krankheiten, die nicht ausschließlich stationär behandelt werden. Hier können nur spezielle Untersuchungen, wie etwa repräsentative Bevölkerungssurveys, valide Zahlen liefern.

Mortalitäts- und Morbiditätsraten können nach bestimmten Kriterien differenziert und für Teilgruppen der Bevölkerung getrennt ausgewiesen werden. In der Epidemiologie sind vor allem alters- und geschlechtsspezifische sowie krankheitsspezifische Raten verbreitet. Daneben finden sich Differenzierungen nach sozialen Merkmalen, Nationalität, Versicherungsstatus oder Regionen. Die Eingrenzungen erfolgt bei der Berechnung der Raten im Zähler und/oder im Nenner. **Altersspezifische Raten** beziehen die Zahl der Ereignisse von Personen einer bestimmten Altersgruppe auf deren Gesamtzahl. Das gleiche gilt für **geschlechtsspezifische Raten**. **Krankheitsspezifische Raten** zählen im Zähler nur Neuerkrankungen, erkrankte Personen oder Todesfälle bzgl. einer bestimmten Erkrankung; im Nenner steht aber weiterhin die gesamte Population unter Risiko.

Bei der Bestimmung von Raten ist auf eine möglichst valide und nachprüfbar Bestimmung der Angaben in Zähler und Nenner zu achten. Dazu gehört zum einen die Unterscheidung zwischen Personen- und Fallbezug und der Umgang mit wiederholten Ereignisse bei ein und derselben Person. Das gilt natürlich nicht für Todesfälle, aber für Krankheitsereignisse oder ärztliche Kontakte, denn Patienten können mehrmals in einem Berichtszeitraum erkranken oder ärztliche Hilfe in Anspruch nehmen. Die Krankenhausdiagnosestatistik ist z.B. eine Fall- und keine Patientenstatistik; Mehrfachaufnahmen eines Patienten werden je gesondert gezählt. Weiteres Prüfkriterium ist die Definition bzw. Bestätigung eines Ereignisses (z.B. berichtet Krankheitsereignisse vs. ärztlich bestätigte Diagnosen) oder die räumlich-zeitliche Eingrenzung (Wie werden Krankenhausfälle von Deutschen im Ausland oder Behandlungen von Ausländern in deutschen Krankenhäusern gezählt? Gilt als Einschlusskriterium das Aufnahme- oder das Entlassungsdatum?). Bezogen auf den Nenner kann es Schwierigkeiten bei der Bestimmung der Risikobevölkerung oder der Zeit unter Risiko geben (s.o.). Dazu zählt auch die Eingrenzung der Gesamtpopulation, z.B. der GKV-Versicherten in der Gesamtbevölkerung oder die Abgrenzung von Auszubildenden, Umschülern, Arbeitslosen etc.

Bei krankheitsspezifischen Mortalitätsanalysen ist die begrenzte Validität der Angaben auf dem Totenschein und der daraus resultierenden Todesursachenstatistik zu beachten. Die Angaben auf dem Totenschein bzgl. der unmittelbaren Todesursache und des für den ausstellenden Arzt oftmals verborgenen Grundleidens, besonders bei dem Arzt unbekanntem Verstorbenen und/oder bei chronisch-kranken, betagten oder multimorbiden Patienten, halten oftmals einer pathologischen Überprüfung nicht stand. Hinzu kommt die niedrige Autopsierate in Deutschland von deutlich unter 20 %. Gleichwohl ist die Todesursachenstatistik eine der wichtigsten Datenquelle für die Epidemiologie.

Besondere Bedeutung für die Beurteilung der Leistungsfähigkeit eines Gesundheitswesens haben die verschiedenen Kennziffern zur Sterblichkeit von Neugeborenen und Kleinkindern. Im einzelnen werden unterschieden:

perinatale Sterblichkeit:	Totgeburten und Frühgestorbene bis zum 7. Lebenstag
Frühsterblichkeit:	Todesfälle bis zum 7. Lebenstag
Spätsterblichkeit:	Todesfälle zwischen dem 7. Lebenstag und dem 28. Lebenstag
Nachsterblichkeit:	Todesfälle zwischen dem 28. Lebenstag und der Vollendung des ersten Lebensjahres

Als Nenner werden jeweils alle Lebendgeburten, bei der perinatalen Sterblichkeit auch die Totgeburten nach der 28. Schwangerschaftswoche gezählt. Die Raten werden üblicherweise pro 1000 Lebendgeborene ausgewiesen.

Die **Säuglingssterblichkeit** ergibt sich als Summe aus Früh-, Spät- und Nachsterblichkeit, damit aus allen Todesfällen bis zur Vollendung des ersten Lebensjahres. Sie ist einer der gebräuchlichsten Indikatoren beim Vergleich verschiedener Gesundheitssysteme. Abbildung 6.2 zeigt allein in Europa erhebliche Unterschiede bei der Säuglingssterblichkeit, sie liegt in einigen Ländern Osteuropas doppelt so hoch wie in West-, Mittel- und Nordeuropa. Selbst die hohen Werte für die Türkei liegen noch weit unter den Zahlen der ärmsten Länder der Welt.

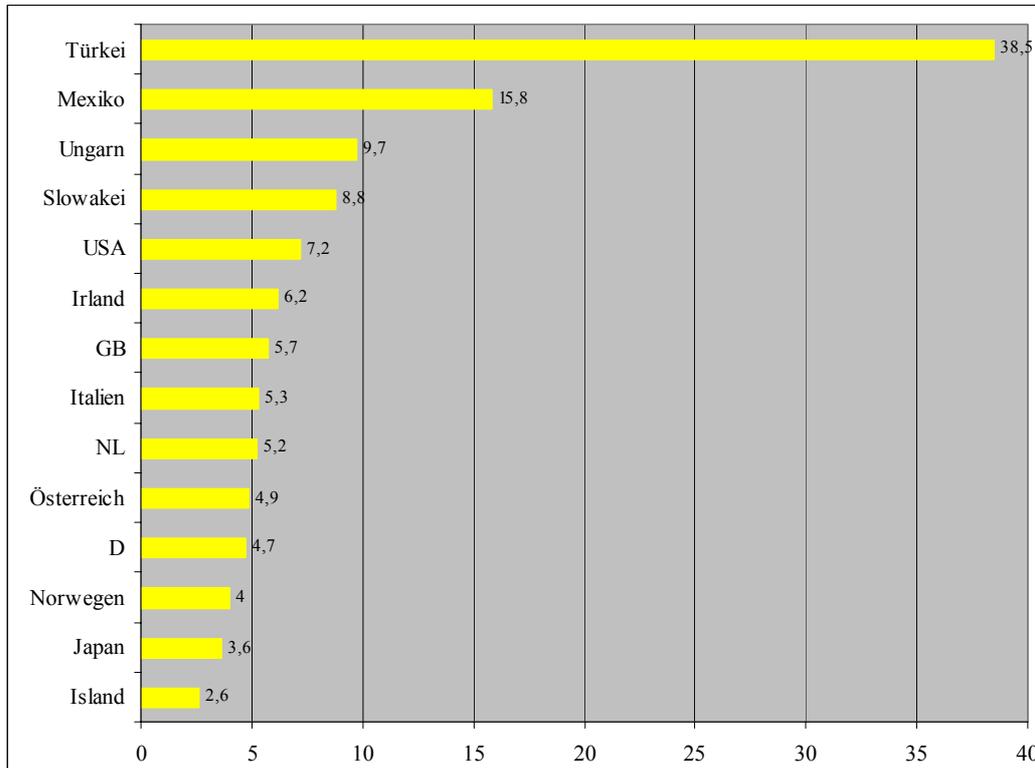


Abbildung 6.2: Säuglingssterblichkeit in verschiedenen Ländern der OECD, 1998 (Quelle: OECD 2001)

Die **Letalitätsrate** ist ein Gradmesser für die Schwere und Bedrohlichkeit einer Krankheit. Sie ist definiert als die Zahl der an einer bestimmten Krankheit Verstorbenen geteilt durch alle erkrankten Personen. Die Letalität ist damit keine Rate, sondern eine Proportion mit Werten zwischen 0 und 1 bzw. 0 und 100%. Die Letalitätsrate ist nur für lebensbedrohliche Krankheiten ein aussagefähiger Indikator (z. Neubildungen oder Myokardinfarkt), während man an den meisten chronischen Krankheiten nicht direkt stirbt (z.B. Diabetes, Rheuma). Bei der Angabe der Letalität ist ebenfalls der Zeitraum nach Diagnosestellung bzw. Erkrankung von Bedeutung. Diese kann kurz sein (z.B. 30 Tage nach akutem Myokardinfarkt) oder mehrere Jahre betragen (bei Krebserkrankungen).

Bei Neubildungen ist es üblich, **Überlebenswahrscheinlichkeiten** als Zielgröße bei klinischen oder epidemiologischen Studien zu verwenden. Zentral ist hierbei die 5-Jahres-Überlebenswahrscheinlichkeit, da bei vielen Lokalisation nach 5 Jahren vergleichsweise nur noch wenige Todesfälle auftreten und man bei vielen Patienten bei Überleben dann von ‚Heilung‘ sprechen kann.

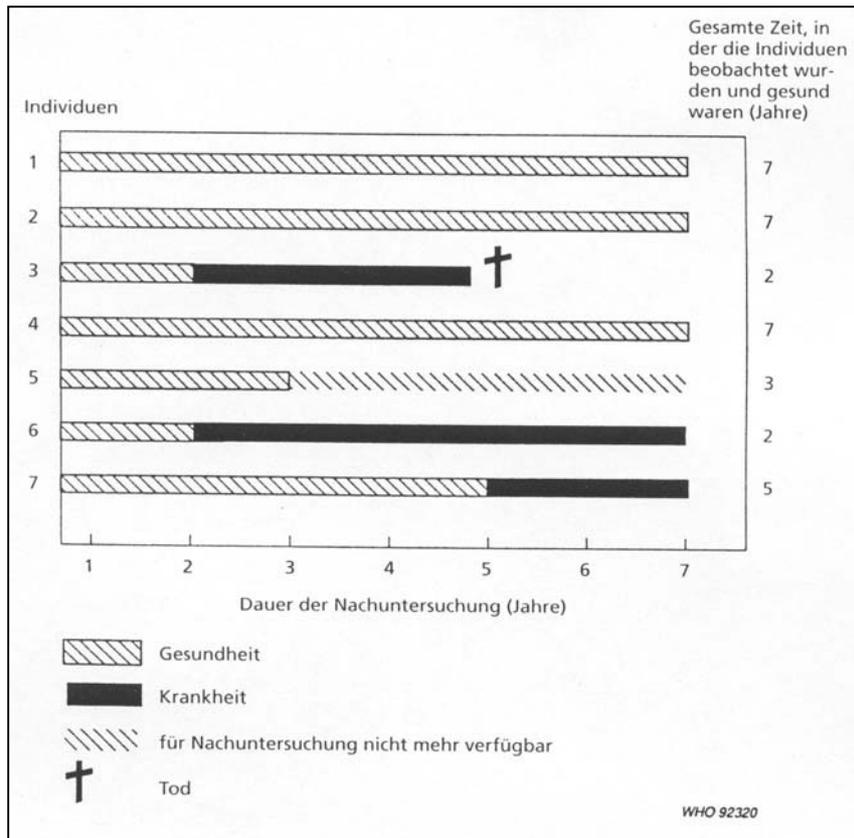


Abbildung 6.3: Verdeutlichung epidemiologischer Maßzahlen (Quelle: Beaglehole et al. 1997)

Anhand der Abbildung 6.3 lassen sich einige der vorgestellten Maßzahlen gut verdeutlichen:

- Die *Inzidenzrate* der Krankheit während der siebenjährigen Beobachtungszeit ist die Zahl der Ereignisse ($n=3$) geteilt durch die Gesamtzeiten der Risikoexposition (33 Personennjahre), also 9,1 Fälle pro 100 PJ.
- Die *kumulative Inzidenz* ist die Zahl der Ereignisse ($n=3$) geteilt durch die Zahl krankheitsfreier Personen ($n=7$) zu Beginn der Beobachtungszeit, also 43 Fälle pro 100 Personen).
- Die *durchschnittliche Krankheitsdauer* ergibt sich als die Gesamtdauer aller Krankheitsepisoden geteilt durch die Zahl der Episoden, also $10/3 = 3,3$ Jahre.
- Die *Prävalenzrate* hängt vom Messzeitpunkt ab. Zu Beginn der vierten Jahres ergibt sie sich beispielsweise als Quotient aus der Zahl der erkrankten Personen ($n=2$) durch alle Personen unter Beobachtung ($n=6$), also 33 Fälle pro 100 Personen.
- Mit der Näherungsformel für ein Fließgleichgewicht ergäbe sich als Prävalenz ein Wert von 30 ($9,1 \cdot 3,3$)
- Die *Letalität* der Krankheit beträgt ein Todesfall auf drei Erkrankungen, also 33 Prozent.

Bei der Morbidität und Mortalität sind **Kohorteneffekte** zu beachten. Das heißt, dass Personen in einer bestimmten Altersgruppe unterschiedlichen Sterblichkeitsrisiken ausgesetzt sind, je nachdem, in welchem Jahr/Zeitraum sie geboren sind. So haben nicht nur Säuglinge heute eine sehr viele höhere Lebenserwartung bzw. ein geringeres Mortalitätsrisiko als noch vor einhundert Jahren. Das gleiche gilt analog für alle anderen Lebensalter. Auch 40-, 60-, oder 80jährige haben ein geringeres Sterberisiko und damit eine höhere Lebenserwartung als Gleichaltrige zu jeder Zeit vor ihnen. Gleichzeitig werden sich Mortalitätsrisiken der heute geborenen Säuglinge in ihrem weiteren Leben mit dem zu erwartenden medizinischen Fortschritt verringern. Kohorteneffekte können sich aber auch in einer Erhöhung des (krankheitsspezifischen) Sterberisikos niederschlagen. So weisen Frauen heutzutage aufgrund des gestie-

genen Anteils von Raucherinnen eine deutlich höhere (krankheitsspezifische) Sterblichkeit beim Lungenkarzinom als ihre Müttergeneration auf.

Weitere Maßzahlen der Sterblichkeit, die zunehmend an Bedeutung gewinnen, auf die hier aber nicht näher eingegangen werden kann, sind die der vorzeitigen Sterblichkeit (Todesfälle vor dem vollendeten 65. Lebensjahr) und vermeidbare Todesfälle, worunter derzeit 29 Todesursachen subsummiert werden, die überwiegend durch medizinische und/oder präventive Maßnahmen vermieden werden könnten (Tab. 6.1). Es gibt Argumente, diese Liste durch Sterbefälle an AIDS, durch Alkohol- bzw. Drogenkonsum und Selbsttötung zu ergänzen.

Tabelle 6.1: Liste vermeidbarer Todesfälle

vermeidbar durch Impfungen (n=3)	Pertussis, Masern, Tetanus
vermeidbar durch Früherkennung (n=5)	Krens Gebärmutterhals, Brustkrebs, perinatale Sterbefälle, Müttersterblichkeit, angeborene Anomalien des Kreislaufsystems
vermeidbar durch Therapie (n=15)	Morbus Hodgkin, Asthma, Magen- und Duodenalgeschwüre, chron. rheum. Herzkrankheiten, Tuberkulose, Gebärmutterkrebs, Atemwegserkrankungen, Typhus, Osteomyelitis, Infektionen des Verdauungssystems, Hodenkrebs, Leukämie, Appendizitis, Eingeweidebrüche, Cholelithiasis/Cholezystis
vermeidbar durch med.-präventive Maßnahmen (n=3)	Ischämische Herzkrankheiten, Hypertonie und Schlaganfall, sonstige bösartige Neubildung der Haut
vermeidbar durch präventive Maßnahmen (n=3)	Lungenkrebs, Leberzirrhose, Kraftfahrzeugunfälle

6.2 Maßzahlen aus der Demographie

Für die Epidemiologie sind einige elementare Kenntnisse aus der Demographie von Bedeutung. Der Aufbau einer Bevölkerung und ihre absehbare Entwicklung sind wichtig bei der Einschätzung epidemiologischer Maßzahlen und deren Bedeutung für die medizinische Versorgung. Die Prozesse und Konsequenzen der sog. demographischen Alterung gehören mittlerweile zum Gegenstand jeder Diskussion um die Reformen der Sozialversicherungssysteme.

Der Aufbau einer Bevölkerung wird üblicherweise in Form einer **Bevölkerungspyramide** dargestellt, wobei die Pyramide nur die Struktur einer jungen wachsenden Bevölkerung wiedergibt, während die **Glockenform** stationäre, nicht wachsende und die **Urnenform** schrumpfende Bevölkerungen charakterisiert (Abbildung 6.4).

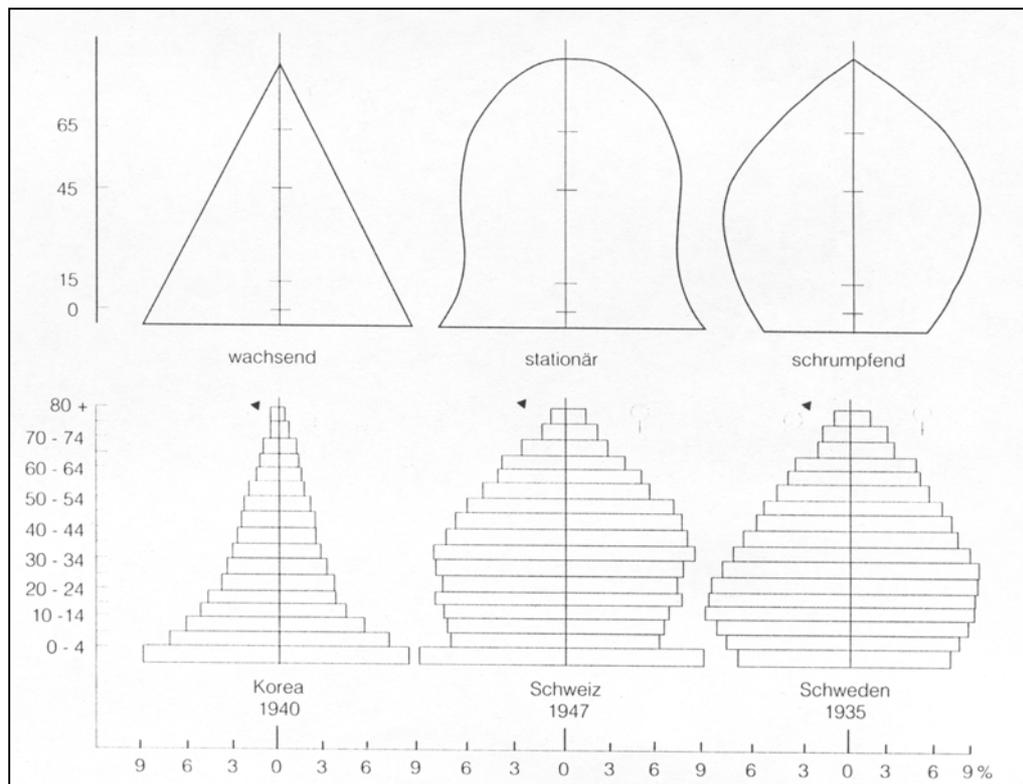


Abbildung 6.4: Charakteristische Formen einer wachsenden, nicht-wachsenden und schrumpfenden Bevölkerung (Hauser 1982)

Ein wichtige demographische Kennzahl ist die **Lebenserwartung bei Geburt**. Sie wird weltweit als Indikator für die Qualität der Gesundheitsversorgung bzw. für den Entwicklungsstand der Gesellschaft verwendet. So ging die Lebenserwartung in den Ländern der ehemaligen Sowjetunion Anfang der 90er Jahre aufgrund der gesellschaftlichen Umbrüche deutlich zurück, um erst jetzt wieder das Niveau Ende der 80er Jahre des vergangenen Jahrhunderts zu erreichen. Aus Abbildung 6.5 ist weiter erkennbar, dass sich zwar die Schere in der Lebenserwartung zwischen reichen und ärmeren Ländern schließt, der Unterschied in der Lebenserwartung aber immer noch mehrere Jahre beträgt. In einigen Ländern Schwarzafrikas liegt die Lebenserwartung sogar noch unter 60 Jahre.

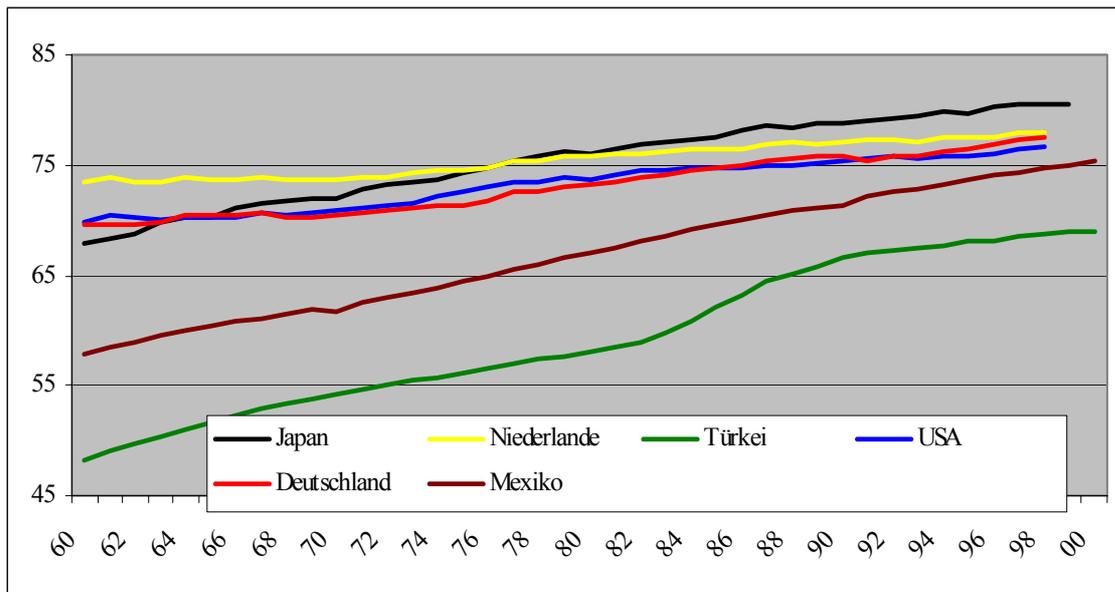


Abbildung 6.5: Lebenserwartung bei Geburt in ausgewählten Ländern der OECD, 1960 bis 2000 (Quelle: OECD 2001)

Aus **Sterbetafeln** können Lebenserwartungen zu Geburt und zu verschiedenen Lebensaltern bzw. Mortalitätsrisiken abgelesen werden. Ausgehend von einer Geburtskohorte von üblicherweise 100.000 Männern oder Frauen stellen Sterbetafeln den Verlauf der altersspezifischen Mortalität bis zum Alter von 90 oder 100 Jahren dar. Tabelle 6.2 zeigt eine (abgekürzte) Sterbetafel des statistischen Bundesamtes für deutsche Männer.

In der ersten Spalte sind die vollendeten Lebensjahre aufgetragen, wegen der überproportionalen Sterblichkeit in den ersten Lebenstagen und –monaten wird das erste Lebensjahr häufig noch in die ersten 4 Wochen und die weiteren 11 Monate unterteilt. In der zweiten und dritten Spalte sind die Sterbe- bzw. Überlebenslebenden der Ausgangskohorte nach der Vollendung des jeweiligen Lebensjahres aufgeführt. So sterben rund 0,5 % der Neugeborenen im ersten Lebensjahr, 2 % der 65jährigen vor der Vollendung des 66. Lebensjahres und 8% zwischen dem 80. und 81. Lebensjahr.

In der 4. Spalte sind die Überlebenden der Ausgangskohorte bis zum Alter x aufgeführt, in der fünften Spalte die Zahl der in diesem Lebensjahr bis zur Vollendung des nächsten Jahres Gestorbenen. In der sechsten Spalte schließlich ist die durchschnittliche **fernere Lebenserwartung** von Personen im vollendeten Alter x angegeben. In der ersten Zeile steht dort für Neugeborene die Lebenserwartung, derzeit 74,4 Jahre für Männer. Für die weiteren Jahrgänge stehen dort Restlebenserwartungen, bei 40jährigen Männern etwa 36,2 Jahre. Die Restlebenserwartung addiert zum aktuellen Alter ist größer als die Lebenserwartung bei Geburt, da bereits einige Risiken in jüngeren Lebensjahren überlebt wurden (bei Männern etwa Verkehrsunfalltod oder Herzinfarkt, bei Frauen die glücklicherweise nur noch geringe Müttersterblichkeit). Für weniger entwickelte Länder sind vor allem die Risiken durch akute Infektionskrankheiten zu nennen.

Sterbetafeln werden üblicherweise aus jeweils drei Jahren der Todesursachenstatistik berechnet (**Periodensterbetafeln**), da man schlechterdings nicht warten kann, bis eine ganze Kohorte nach 90 oder 100 Jahren abgestorben ist. In dem Maße, wie sich die Sterblichkeit im zeitlichen Verlauf ändert, stellt eine derartige Sterbetafel nicht das aktuelle Risiko der derzeitigen Bevölkerung dar. Bei weiterer Abnahme der Sterblichkeit liegen die Lebenserwartungen der aktuellen Bevölkerung über den in der Tabelle abgebildeten.

Tabelle 6.2: Abgekürzte Sterbetafel Deutschland, männlich 1997/1999

Vollend. Alter	Sterbe- wahrscheinlichkeit vom Alter x bis x+1	Überlebens- Überlebende im Alter x	Gestorbene im Alter x bis unter x+1	Durchschnittliche Lebenserwartung im Alter x	
(1)	(2)	(3)	(4)	(5)	
0	0,00517917	0,99482083	100 000	518	74,44
1	0,00046287	0,99953713	99 482	46	73,83
2	0,00030871	0,99969129	99 436	31	72,86
3	0,00021732	0,99978268	99 405	22	71,88
4	0,00019191	0,99980809	99 384	19	70,90
...
16	0,00045774	0,99954226	99 183	45	59,03
17	0,00061704	0,99938296	99 137	61	58,06
18	0,00106066	0,99893934	99 076	105	57,09
19	0,00101938	0,99898062	98 971	101	56,15
20	0,00098009	0,99901991	98 870	97	55,21
21	0,00096056	0,99903944	98 773	95	54,26
...
40	0,00209648	0,99790352	96 723	203	36,18
41	0,00231235	0,99768765	96 520	223	35,26
42	0,00263339	0,99736661	96 297	254	34,34
43	0,00291999	0,99708001	96 043	280	33,43
44	0,00305614	0,99694386	95 763	293	32,53
...
65	0,02095195	0,97904805	79 784	1 672	15,36
66	0,02292736	0,97707264	78 112	1 791	14,68
67	0,02576294	0,97423706	76 321	1 966	14,01
68	0,02853933	0,97146067	74 355	2 122	13,37
69	0,03091012	0,96908988	72 233	2 233	12,75
...
80	0,08052973	0,91947027	41 043	3 305	6,91
81	0,09126459	0,90873541	37 738	3 444	6,47
82	0,10223157	0,89776843	34 293	3 506	6,07
83	0,11272565	0,88727435	30 788	3 471	5,71
84	0,12404013	0,87595987	27 317	3 388	5,37
...
90	1,00000000	0,00000000	9 903	9 903	3,91

Statistisches Bundesamt 2002; aktuellere Sterbetafeln unter www.destatis.de

Tabelle 6.3 zeigt, wie sich die Lebenserwartung bei Geburt und zu gewissen Lebensaltern im 20. Jahrhundert entwickelt hat. Die größten prozentualen Gewinne sind bei Neugeborenen und bei Frauen über 80 Jahre zu verzeichnen. Die Entwicklung der Sterblichkeit ist bei Frauen günstiger als bei Männern, einerseits wegen der geringeren Risiken und andererseits wegen der Auswirkungen der beiden Weltkriege, die sich immer noch in den Sterbetafeln der derzeit lebenden männlichen Bevölkerung niederschlägt.

Tabelle 6.3: Veränderung der (ferneren) Lebenserwartung in verschiedenen deutschen Sterbetafeln

vollendetes Alter	1901/10	1932/34	1949/51	1992/94	1994/96	1997/99	Zuwachs vs. 1901/10 (in %)
bei Geburt							
m	44,8	59,9	64,6	72,8	73,3	74,4	+ 66,1
w	48,3	62,8	68,5	79,3	79,7	80,6	+ 66,7
20 Jahre							
m	42,6	48,2	50,3	53,7	54,1	55,2	+ 29,7
w	44,8	49,8	53,2	60,0	60,4	61,2	+ 36,4
40 Jahre							
m	26,6	30,8	32,3	34,9	35,3	36,2	+ 35,8
w	29,2	32,3	34,7	40,6	40,9	41,6	+ 42,7
60 Jahre							
m	13,1	15,1	16,2	18,0	18,3	19,0	+ 32,2
w	14,2	16,1	17,5	22,4	22,7	23,3	+ 64,4
80 Jahre							
m	4,4	4,8	5,2	6,3	6,5	6,9	+ 57,8
w	4,7	5,2	5,6	7,8	8,0	8,4	+ 80,0

Quelle: Statistisches Jahrbuch, verschiedene Jahrgänge

Eine weitere häufig verwendete demographische Maßzahl ist die **Fertilitätsrate**. Sie ist definiert als die Zahl der Geburten in einem gewissen Zeitraum, bezogen auf die Frauen im gebärfähigen Alter, üblicherweise auf die Altersgruppe von 15 bis 44 Jahre. Bekannt sind weiterhin die **Brutto- und Nettoreproduktionsrate** (BRR bzw. NRR). Die Bruttoreproduktionsrate gibt die Zahl der Mädchen an, die eine Frau - unter den jeweils gültigen altersspezifischen Fertilitätsziffern - während ihres Lebens zur Welt bringt. Bei der Nettoreproduktionsrate wird zusätzlich noch die Mortalität der Mädchen und Frauen bis zum Ende des gebärfähigen Zeitraums berücksichtigt. Eine NRR von 1 führt zu einer vollständigen Ersetzung der Mütter durch die Töchtergeneration, eine NRR < 1 bzw. eine Fertilitätsrate von weniger als 2,1 führt ohne positive Wanderungen zu einem Schrumpfen der Bevölkerung. Abbildung 6.6 zeigt, dass dies mit Ausnahme von Island und der Türkei für alle Länder Europas gilt.

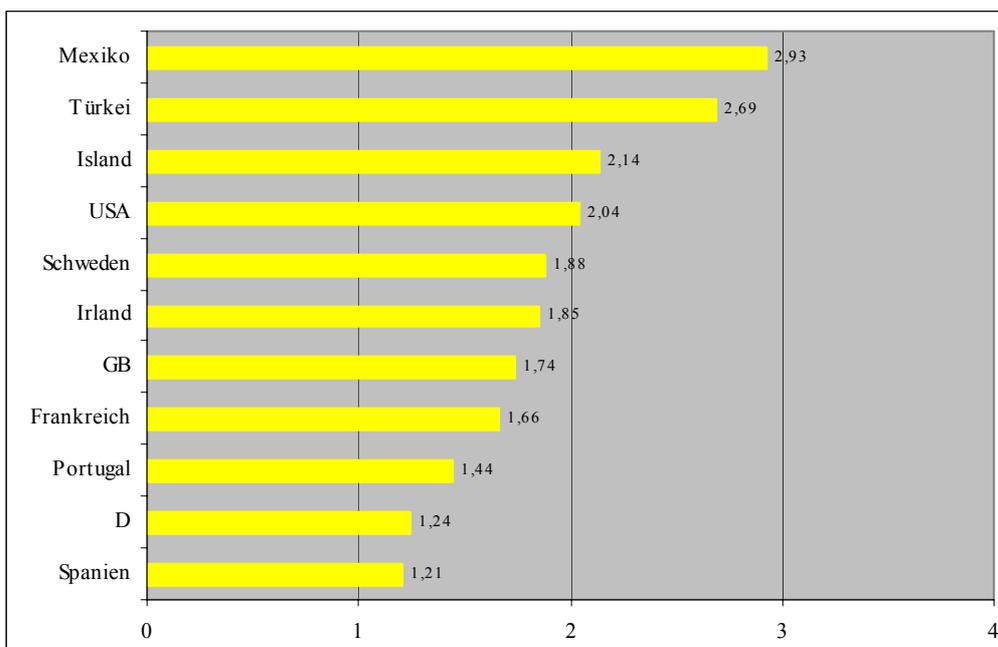


Abbildung 6.6: Fertilitätsrate in ausgewählten Ländern der OECD, 2000 (Quelle: OECD 2001)

Die gemeinsame Berücksichtigung der Mortalitäts- und Fertilitätsprozesse zusammen mit dem Wanderungsgeschehen erlaubt relativ genaue Vorhersagen über die Bevölkerungsent-

wicklung eines Landes über einen Zeitraum von mindestens zehn bis zwanzig Jahren, da sich zumindest Mortalitäts- und Fertilitätsraten nur langsam im zeitlichen Verlauf ändern. Das statistische Bundesamt veröffentlicht daher alle drei bis fünf Jahre sog. koordinierte Bevölkerungsprognosen, die derzeit letzte stammt aus dem Herbst 2006. In verschiedenen Szenarios wird die Bevölkerung bis zum Jahr 2050 prognostiziert. Die Gesamtzahl der Bevölkerung schwankt nach der vorletzten Prognose zwischen 81 Millionen (bei hoher Lebenserwartung und hohen Wanderungsüberschüssen von jährlich 300.000 Personen) und 67 Millionen (nur leicht ansteigende Lebenserwartung und geringe Wanderungsüberschüsse von jährlich 100.000 Personen). Die mittlere Variante rechnet um 2050 mit einer Bevölkerungszahl von 75 Millionen, das wäre eine Verringerung um 7 Millionen Einwohner gegenüber Ende 2001 (Abbildung 6.7; download über www.destatis.de).

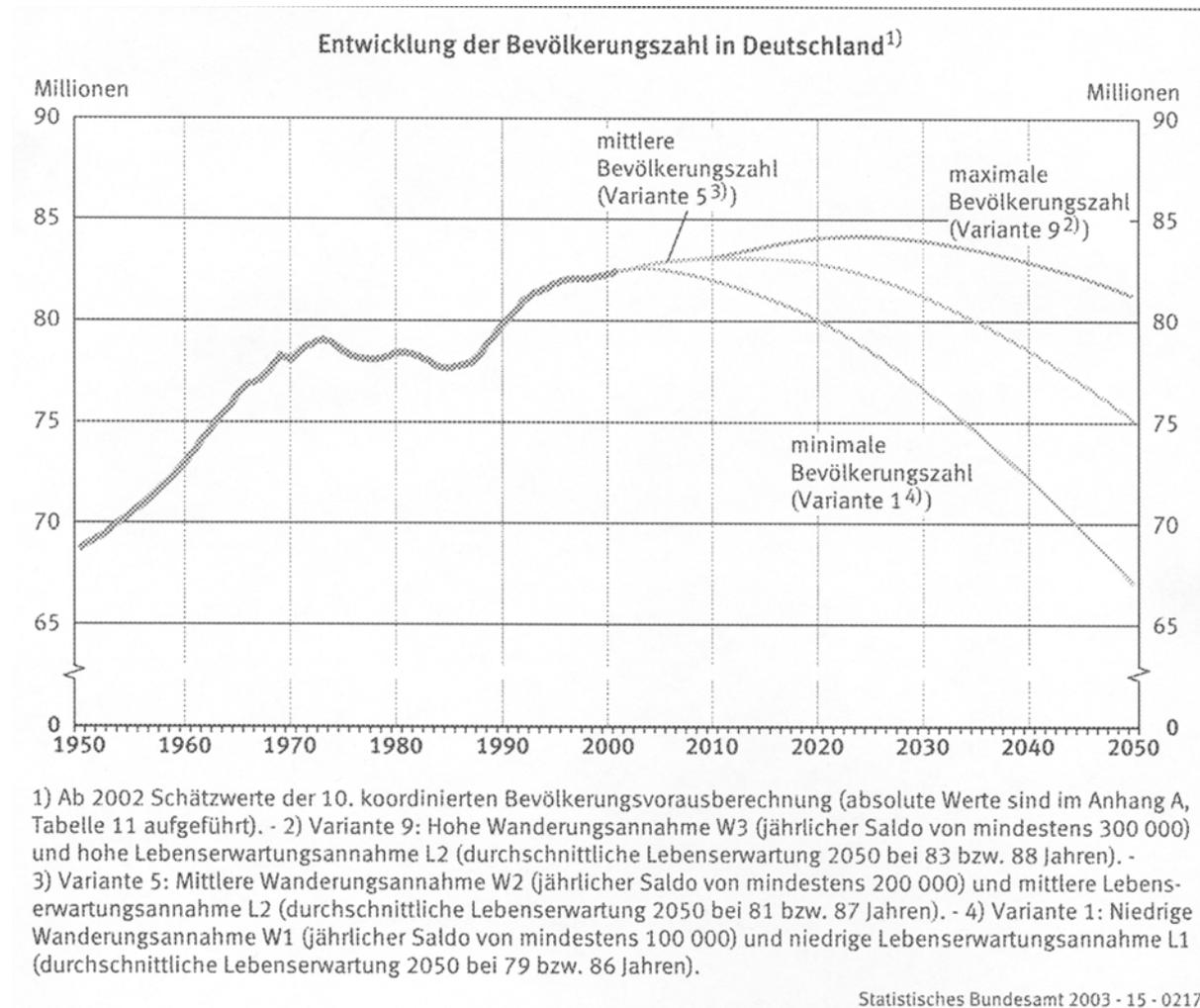


Abbildung 6.7: Entwicklung der Bevölkerungszahl in Deutschland nach der 10. koordinierten Bevölkerungsvorausberechnung

Wichtig bei der Bewertung derartiger Entwicklungen ist aber nicht nur die absolute Einwohnerzahl einer Gesellschaft, sondern auch die Veränderung in deren Altersstruktur. Nach der mittleren Variante wird sich der **Altenquotient**, also die Zahl der Einwohner im Rentenalter ab 65 Jahre bezogen auf 100 Personen im Erwerbsalter, von derzeit 27,5 auf 54,5 erhöhen, auf zwei Erwerbstätige käme damit ein Rentner, der nicht mehr oder nur wenig in die Sozialversicherungssysteme einzahlt. Diese gravierenden Veränderungen spiegeln sich auch graphisch

wider, der Bevölkerungsaufbau verschiebt sich von der Glocken- in die Urnenform (Abbildung 6.8).

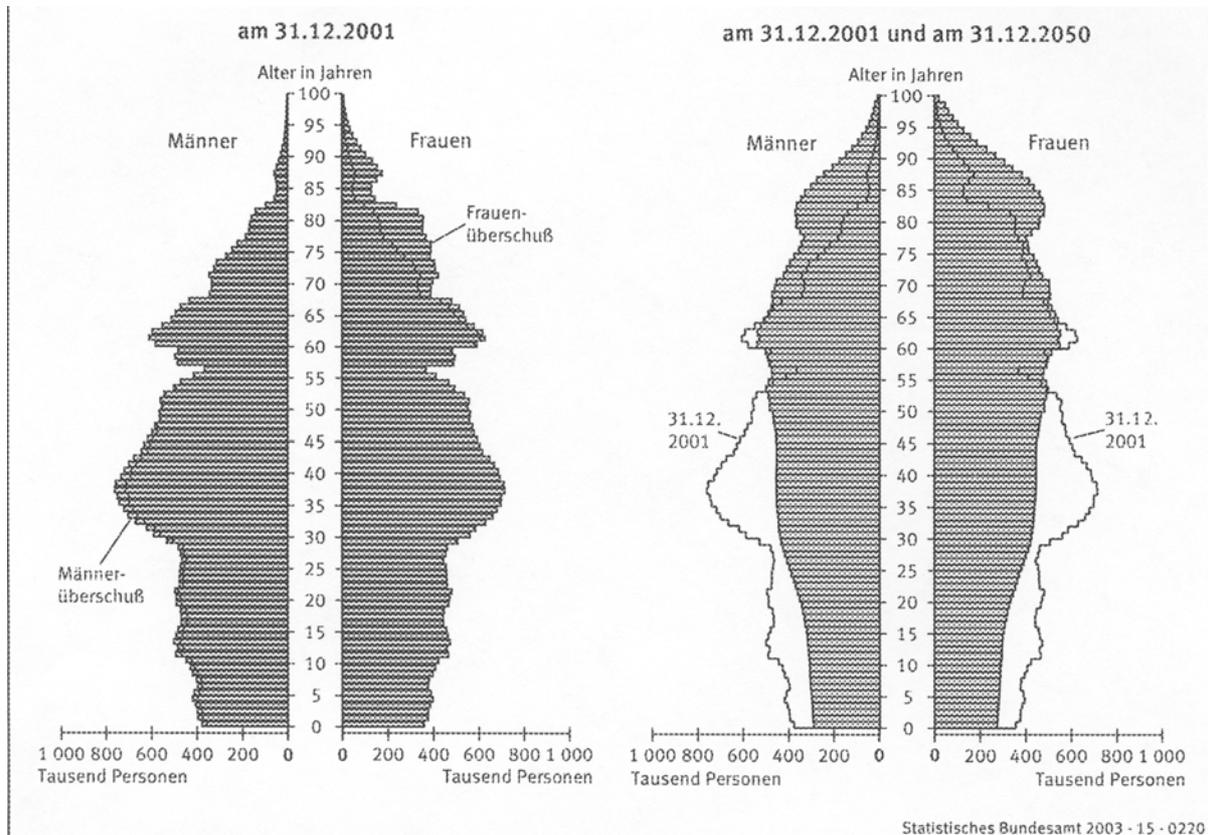


Abbildung 6.8: Altersaufbau der Bevölkerung in Deutschland Ende 2001 und Ende 2050 (mittlere Variante der nach der 10. koordinierten Bevölkerungsvorausberechnung)

6.3 Standardisierung von Maßzahlen

Der Vergleich von Maßzahlen der Morbidität und Mortalität, die in der Regel stark alters- und geschlechtsabhängig sind, ist erschwert oder gar unmöglich, wenn zwei Populationen unterschiedliche Alters- und Geschlechtsstrukturen aufweisen, wenn sich eine Population im zeitlichen Verlauf schnell ändert oder wenn die altersspezifischen Raten sehr unterschiedliche Entwicklungen aufweisen. Das Beispiel aus Tabelle 6.4 zeigt bei Schwarzen in allen Altersgruppen höhere altersspezifischen Sterberaten als bei Weißen und gleichwohl eine um 40% höhere rohe Mortalitätsrate bei Weißen. Da die Sterberaten bei beiden Gruppen mit dem Alter deutlich ansteigen, führt die unterschiedliche Alterstruktur in der weißen und der schwarzen Bevölkerung zu diesem scheinbar paradoxen Ergebnis. Man erkennt außerdem, dass der Unterschied in den Sterberaten zwischen Weißen und Schwarzen mit dem Alter abnimmt (,selektives Überleben‘, ,survival of the fittest‘).

Tabelle 6.4: Sterberaten nach Alter und Rasse, Baltimore 1965 (pro 1.000 Einwohner)

Altersgruppen	Weißer	Schwarze
<1 Jahr	23,9	31,3
1 bis 4 Jahre	0,7	1,6
5 bis 17 Jahre	0,4	0,6
18 bis 44 Jahre	2,5	4,8
45 bis 64 Jahre	15,2	22,6
65 Jahre und älter	69,3	75,9
gesamt	14,3	10,2

Quelle: Gordis (2001), S. 60

Zwar lassen sich die alters- und geschlechtsspezifische Raten miteinander vergleichen, doch der Vergleich der globalen Maßzahlen erfordert das ‚Herausrechnen‘ der Unterschiede in der Bevölkerungsstruktur. Dieses Verfahren nennt man Standardisierung. Tabelle 6.5 zeigt Unterschiede in der rohen und standardisierten Mortalität zwischen verschiedenen Ländern der OECD. Die hohe rohe Mortalität von Schweden relativiert sich bei Betrachtung standardisierter Raten, ebenso die geringe rohe Mortalität Mexikos wegen seiner jungen Bevölkerungsstruktur. Die Spanne der rohen Mortalitätsraten verringert sich etwa um ein Drittel bei Übergang zu den standardisierten Mortalitätsraten.

Tabelle 6.5: Rohe und altersstandardisierte Mortalität in verschiedenen Ländern, 1994

Land	Mortalität roh	Direkt stand.
Japan	700,6	539,2
Frankreich	901,8	601,5
Schweden	1041,5	614,3
Italien	972,5	652,6
Norwegen	1016,3	672,9
Niederlande	867,8	690,5
Großbritannien	1074,8	731,4
Vereinigte Staaten	875,4	736,3
Deutschland	1086,5	738,0
Irland	756,2	827,8
Mexiko	473,9	918,1
Korea	516,7	919,3
Tschechische Republik	1135,6	984,7
Slowakei	961,0	994,6
Ungarn	1431,5	1179,0
Spanne	474 bis 1432	539 bis 1179

Standard: OECD; Quelle OECD 2001

Die beiden geläufigen Verfahren der Standardisierung seien wegen ihrer zentralen Bedeutung in der deskriptiven Epidemiologie hier kurz erläutert. Bei der **direkten Standardisierung** wird eine Standardbevölkerung mit festgelegter Alters- und Geschlechtsstruktur zugrunde gelegt. Der Anteil der i-ten von insgesamt k Altersklassen an der Gesamtbevölkerung wird als w_{ij}^* bezeichnet, wobei noch nach Männern (j=1) und Frauen (j=2) differenziert wird. Die alters- und geschlechtsspezifischen Morbiditäts- oder Mortalitätsraten r_{ij} der Vergleichspopulation(en) werden nun mit der Alters- und Geschlechtsstruktur der Standardbevölkerung gewichtet:

$$r_{\text{dir. stand}} = \sum_{i=1,2; j=1, \dots, k} w_{ij}^* \cdot r_{ij} \quad (6:2)$$

Häufig liegen die Raten r_{ij} der Vergleichspopulation jedoch nicht vor (z.B. Mortalitätsraten nach Gemeinden) oder sie sind wegen kleiner absoluter Fallzahlen stark zufälligen Einflüssen unterworfen. Für diese Situation bietet sich die – häufiger verwendete - **indirekte Standardisierung** an. Dabei werden auf die bekannte Alters- und Geschlechtsstruktur W_{ij} (absolut) der

Vergleichspopulation die alters- und geschlechtsspezifischen Raten der Standardpopulation r_{ij}^* angewendet.

$$r_{\text{indir. stand}} = \sum_{i=1,2; j=1, \dots, k} W_{ij} \cdot r_{ij}^* \quad (6:3)$$

Damit ergeben sich die Fallzahlen, die zu erwarten gewesen wären, wenn die Morbiditäts- oder Mortalitätsverhältnisse der Standardbevölkerung in der Studienpopulation wirksam gewesen wären (= E, wie ‚expected‘). Diese lassen sich dann mit der beobachteten Fallzahl (= O wie ‚observed‘) vergleichen. Der Quotient aus O und E, auch SMR genannt⁵ gibt an, wie stark die Raten von der des Standards abweichen. Der Vorteil der indirekten Standardisierung liegt darin, dass nur die alters- und geschlechtsspezifischen Raten der Standardbevölkerung, die Struktur der Vergleichspopulation und die Gesamtzahl der Ereignisse in der Vergleichspopulation gebraucht werden.

Die Auswirkungen unterschiedlicher Altersstrukturen auf die rohen Sterberaten (und anderer epidemiologischer Kennziffern) und das Verfahren der **direkten Altersstandardisierung** kann an folgendem hypothetischem Beispiel nachvollzogen werden. Tabelle 6.6 zeigt, dass sich für die im Vergleich zur Standardbevölkerung ‚junge‘ Bevölkerung 1 die standardisierte Mortalitätsziffer gegenüber der rohen um 75% erhöht, bei der ‚älteren‘ Bevölkerung 2 liegt die standardisierte Mortalität um 12% niedriger als die rohe Mortalität.

Tabelle 6.6: Hypothetisches Beispiel für die direkte Standardisierung einer Mortalitätsrate

	Bevölkerung 1	Todesfälle	Sterberate je 100.000	Standard-Bevölk.	Anteil	Sterberate*Anteil
Altersgruppe	(1)	(2)	(3) = (2) / (1)	(4)	(5)	(6) = (5) * (3)
unter 18 Jahre	200.000	20	10	100.000	0,1	1
18 bis 44 Jahre	400.000	80	20	200.000	0,2	4
45 bis 64 Jahre	300.000	300	100	400.000	0,4	40
65 Jahre und älter	100.000	200	200	300.000	0,3	60
gesamt	1.000.000	600	60	1.000.000		
Standardisierte Sterberate (pro 100.000)						105
	Bevölkerung 2	Todesfälle	Sterberate je 100.000		Anteil	Sterberate*Anteil
Altersgruppe						
unter 18 Jahre	80.000	4	5		0,1	0,5
18 bis 44 Jahre	120.000	18	15		0,2	3
45 bis 64 Jahre	450.000	360	80		0,4	32
65 Jahre und älter	350.000	630	180		0,3	54
gesamt	1.000.000	1.012	101,2			
Standardisierte Sterberate (pro 100.000)						89,5

Tabelle 6.7 liefert ein reales Beispiel für die indirekte Standardisierung. Aus ihr ist ablesbar, dass bei Anlegung der Tbc-Mortalität aus der Allgemeinbevölkerung in der Population der weißen Minenarbeiter statt der beobachteten 436 Tbc-Todesfälle lediglich 181 zu erwarten gewesen wären. Damit ergibt sich für weiße Minenarbeit ein erheblich erhöhtes Mortalitätsrisiko (um 140% erhöht) gegenüber der weißen Allgemeinbevölkerung.

⁵ Standardisiertes Mortalitäts- bzw. Morbiditätsverhältnis (engl. ‚ratio‘), da die SMR keine Rate, sondern den Quotienten zweier Raten darstellt.

Tabelle 6.7: Beispiel für die Berechnung der SMR für Tuberkulose (Tbc) bei weißen Minenarbeitern im Alter von 20 bis 59 Jahre, 1950

	Population weißer Minenarbeiter	Tbc-Mortalität bei Männern der All- gemeinbevölkerung	Erwartete Tbc- Sterbefälle (=E) bei Allgemeinrisiko	Beobachtete Tbc- Sterbefälle (=O)
Alter	(1)	(2)	(3) = (1) * (2)	(4)
20 bis 24 Jahre	74.598	12,26	9,14	10
25 bis 29 Jahre	85.077	16,12	13,71	20
30 bis 34 Jahre	80.845	21,54	17,41	22
35 bis 44 Jahre	148.870	33,96	50,55	98
45 bis 54 Jahre	102.649	56,82	58,32	174
55 bis 59 Jahre	42.494	75,23	31,96	112
Gesamt	534.533		181,09	436

SMR = beobachtete Todesfälle / erwartete Todesfälle (O / E) *100 = 436 / 181,09= 241

Quelle: Gordis, S. 62

Es gibt verschiedene Standardbevölkerungen, die in der Epidemiologie Verbreitung finden. Dazu gehören eine deutsche, eine neue und alte Europa- und eine Weltstandardbevölkerung. Sie unterscheiden sich in ihrem Altersaufbau, wobei die deutsche die ‚älteste‘ und die Weltbevölkerung die ‚jüngste‘ ist. Welche für einen Vergleich am zweckmäßigsten ist, hängt von der inhaltlichen Fragestellung ab. Die standardisierten Raten schwanken stark in Abhängigkeit von der verwendeten Standardbevölkerung (Tabelle 6.8). Neben diesen weit verbreiteten Standardbevölkerungen ist grundsätzlich jede anderer Standard möglich. Bei Analyse der Inanspruchnahme stationärer Leistungen durch GKV-Versicherte bietet sich auch ein Standard auf der Basis der AOK- oder aller GKV-Versicherten an.

Tabelle 6.8: Direkt altstandardisierte Mortalitätsziffern, in Abhängigkeit vom verwendeten Standard Deutschland 1997

Todesursachen	Rohe Rate	Standardisierte Raten			Weltbevölkerung
		Deutschland 1987	Neue Europa- bevölkerung	Alte Europa- bevölkerung	
Alle Todesursachen	1.048,6	955,9	859,7	715,8	462,2
Bösartige Neubildungen	263,0	248,1	222,8	193,8	128,6
Krhn. Kreislaufsystem	506,9	450,3	401,6	317,9	187,3
Unfälle	27,4	26,4	24,8	22,4	18,8
dar. Verkehrsunfälle	10,5	11,3	10,8	10,3	10,2

Quelle: Gesundheitsberichterstattung des Bundes (<http://www.gbe-bund.de>)

Bei allen Ereignissen bis auf die Verkehrsunfälle liegen die rohen Raten über den standardisierten, weil die Krankheitshäufigkeiten mit dem Alter steigen und die hohen Raten der höheren Altersgruppen bei jüngeren Standardpopulation im Vergleich zu der aktuellen deutschen weniger stark gewichtet werden. Bei Verkehrsunfällen ist dies anders, weil daran auch vergleichsweise viele junge Menschen sterben.

7. Epidemiologische Studientypen

In diesem Kapitel werden die wichtigsten epidemiologische Studientypen in ihren Reinformen vorgestellt. Daneben gibt es noch verschiedene Mischformen, die unter Umgehung der Nachteile die jeweiligen Vorteile einzelner Typen zu nutzen versuchen. Darauf muss an dieser Stelle aus Gründen einer komprimierten Darstellung verzichtet werden. Auch in einigen nachfolgenden Kapiteln wird an mehreren Stellen auf die Studientypen eingegangen.

7.1 Experiment

Das **Experiment** stellt die exakteste Form der wissenschaftlichen Forschung dar. Es ist durch die Prinzipien **Willkürlichkeit**, **Variierbarkeit** und **Wiederholbarkeit** gekennzeichnet. Im einzelnen bedeutet diese:

- ◆ Kontrolle der für die Hypothese bedeutsamen Variablen,
- ◆ gezielte Manipulation der Versuchsbedingungen,
- ◆ Konstanzhaltung bzw. rechnerische Berücksichtigung der anderen Variablen und der Störgrößen,
- ◆ dadurch Wiederholbarkeit unter identischen Bedingungen.

Ein Experiment kann in seinem Idealtypus allerdings nur unter Laborbedingungen durchgeführt werden, ist daher primär in der klinischen Forschung (Stichwort **Einfach- und Doppelblindstudien**), weniger in der sozialwissenschaftlichen und epidemiologischen Forschung anzutreffen und hier nur der Vollständigkeit halber aufgeführt.

7.2 Randomisierte Studie

Randomisierte kontrollierte Studien sind gewissermaßen das epidemiologische Pendant zum klinischen Experiment und stellen den **Goldstandard** der Epidemiologie dar. Sie sind charakterisiert durch eine zufällige Aufteilung der Studienpopulation in Untersuchungs- und Kontrollgruppe. In der Epidemiologie werden diese RCT's („randomized controlled trials“) vor allem bei **Interventionsstudien** eingesetzt, in denen die Exposition gesteuert bzw. beeinflusst werden soll. Beispiel: Maßnahmen zur Schulung von Diabetes-Patienten ('Exposition') und Untersuchung ihres Einflusses auf die Zielgröße ‚Höhe und Konstanz des Blutzuckerspiegels‘. Größere Verbreitung haben randomisierte Studien in der klinischen Epidemiologie, etwa bei der Bewertung neuer Medikamente oder Therapieformen.

Die Randomisation kann unter allen Personen erfolgen, die gewisse Ein- und Ausschlusskriterien erfüllen, und sich zu einer Studienteilnahme bereit erklärt haben (vgl. Abbildung 7.1). Sie kann auch eine Stufe früher vor der Festlegung über eine Teilnahme erfolgen und dabei Effekte einer unvollständigen Umsetzung einer Intervention in der Praxis oder einer geringen Teilnahmebereitschaft berücksichtigen („**Intention-to treat-prinzip**“). So wurde beispielsweise bei großen bevölkerungsbezogenen Studien zum Mammographie-Screening verfahren, als das Angebot einer Brustkrebs-Früherkennung allen Frauen der Studiengruppe unterbreitet wurde, unabhängig von ihrer Bereitschaft zur Beteiligung an Früherkennungsmaßnahmen. Außerdem erfolgte die Analyse auf der Basis aller rauen aus Studien- und Kontrollgruppe, unabhängig von ihrer Mitwirkung am Programm

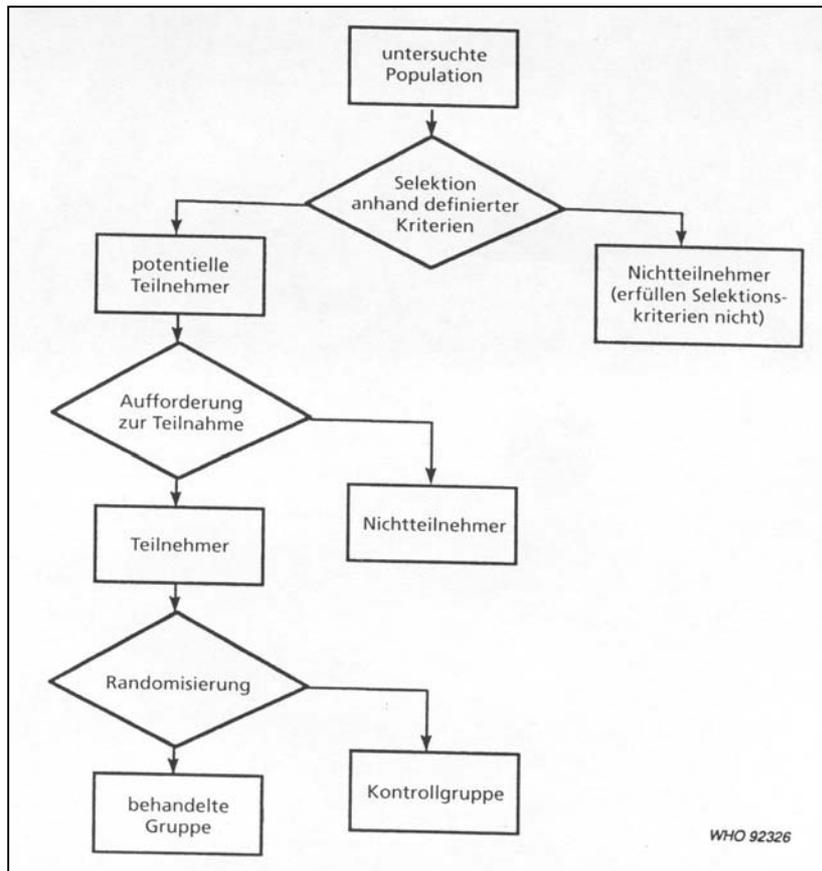


Abbildung 7.1: Design einer randomisierten kontrollierten Studie (Quelle: Beaglehole et al. 1997)

7.3 Fall-Kontroll-Studie

In diesem Studientyp werden die Studienteilnehmer nicht zufällig auf Studien- und Kontrollarm verteilt, sondern nach Vorliegen oder Nichtvorliegen einer Krankheit oder eines anderen definierbaren Tatbestandes. Das können Todesfälle oder Neuerkrankungen sein, aber beispielsweise auch die Inanspruchnahme bestimmter medizinischer Leistungen. **Fälle** sind Personen, bei denen das bestimmte Ereignis aufgetreten ist. Dazu wird eine geeignete **Kontrollgruppe** von Personen gebildet, bei denen diese Krankheit oder das entsprechende Ereignis nicht aufgetreten ist. Retrospektiv, also entgegen der Zeitrichtung, wird dann in beiden Gruppen das Vorhandensein möglicher Krankheitsursachen erhoben und miteinander verglichen. Das können Risikofaktoren, Umweltexpositionen, gesundheitsrelevante Verhaltensweisen o.ä. sein (Abbildung 7.2).

Die wichtigsten Vor- und Nachteile einer Fall-Kontroll-Studie sind in Tabelle 7.1 aufgeführt. Vor allem bei seltenen Erkrankungen wie Karzinomen werden primär Fall-Kontroll-Studien eingesetzt, weil die Rekrutierung über bereits manifeste Fälle erfolgt. Die Datenerhebung kann retrospektiv häufig vorhandene Datenquellen nutzen, derartige Studien sind schnell und kostengünstig durchzuführen, auch deshalb, weil sie relativ weniger Studien- und Kontrollpersonen bedürfen.

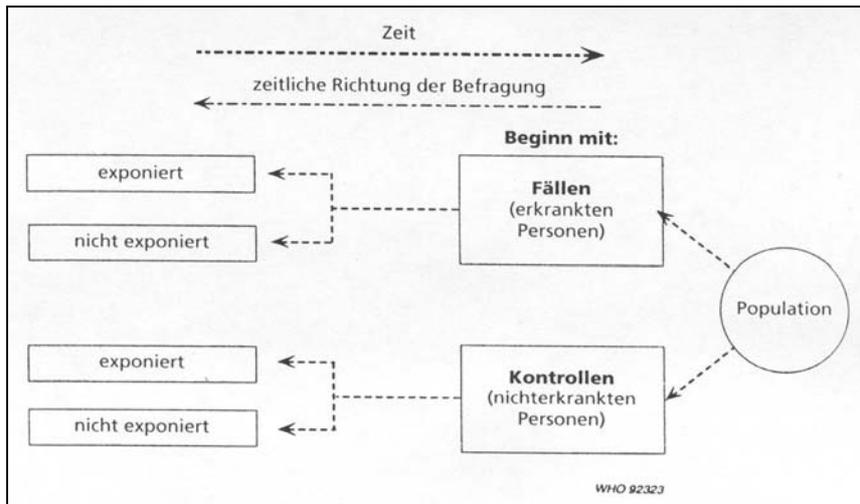


Abbildung 7.2: Design einer Fall-Kontroll-Studie (Quelle: Beaglehole et al. 1997)

Die Nachteile von Fall-Kontroll-Studien bestehen in der Gefahr von Verzerrungsmöglichkeiten. Diese liegen in der Schwierigkeit der Wahl geeigneter Kontrollen (die Verwendung von Krankenhauskontrollen ist verbreitet, nicht immer aber angemessen, vgl. Kapitel 12) und in der oftmals unklaren Validität retrospektiv, aus der Erinnerung erhobener Informationen (mit der Möglichkeit, dass sich Fälle genauer erinnern als Kontrollen; vgl. Kapitel 9). Hinzu kommt der Tatbestand, dass das Krankheitsrisiko bei Fällen und Kontrollen nicht exakt bestimmt werden kann, da das Einschlusskriterium der Status Fall oder Kontrolle und nicht Exposition oder Nicht-Exposition ist.

Tabelle 7.1: Vor- und Nachteile einer Fall-Kontrollstudie

Vorteile	Nachteile
Gut geeignet für seltene Krankheiten oder solchen mit langen Latenzzeiten	Stützt sich auf Erinnerung oder Aufzeichnungen vergangene Expositionen
Relativ schnell und kostengünstig durchzuführen	Validität der Informationen ist schwer zu bestimmen oder unklar
Erfordert relativ wenige Versuchspersonen und Kontrollen	Kontrolle und Erfassung von Störgrößen oder Confoundern kann unvollständig sein
Vorhandene Datenquellen (z.B. Krankenakten) können genutzt werden	Auswahl einer geeigneten Kontrollgruppe kann unvollständig sein
Ohne Risiko für die Studienteilnehmer	Risiko bei Fällen und Kontrollen kann nicht bestimmt werden
Multiple potenzielle Krankheitsursachen können untersucht werden	Detaillierte Bestimmung des Ursache-Wirkungs-Mechanismus kaum möglich

Eine Fall-Kontroll-Studie reproduziert immer dann den Zusammenhang zwischen Exposition und Erkrankung in der Bezugspopulation, wenn die Auswahl(=Aufnahme-)Wahrscheinlichkeit für Fälle und Kontrollen unabhängig vom Expositionsstatus ist. Eine Verleugnung bzw. Überbewertung einer Exposition durch Erkrankte (oder eine Unter- bzw. Übererfassung) führt tendenziell zu einer Unter-(Über-)schätzung des Einflusses dieser Exposition. Zu geeigneten Risikomaßen bei Fall-Kontrollstudien vgl. Kapitel 8.

7.4 Kohortenstudie

Kohortenstudien werden auch als prospektive, Längsschnitt- oder Follow-up-Studien bezeichnet. Ausgangspunkt sind zwei Gruppen exponierter und nicht-exponierter Personen, die

alle krankheitsfrei sind. Ausgehend von diesem Startpunkt wird über einen festgelegten Zeitraum das Auftreten gewisser Zielereignisse (Inanspruchnahme ärztlicher Leistungen, Krankheit, Todesfälle) beobachtet und deren Häufigkeit (Inzidenz) verglichen.

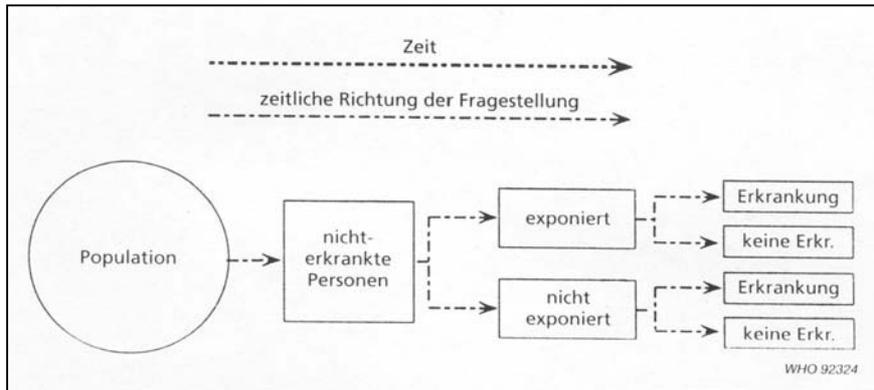


Abbildung 7.3: Design einer Längsschnittuntersuchung (Quelle: Beaglehole et al. 1997)

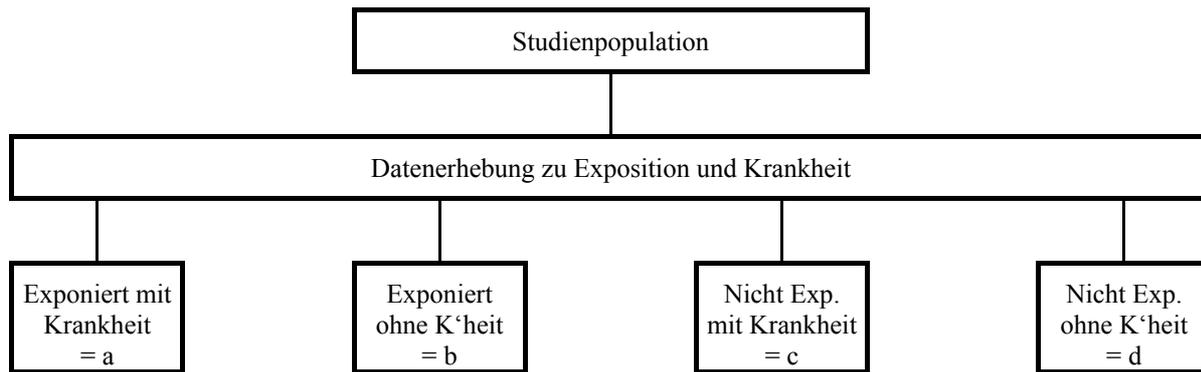
Kohortenstudien ermöglichen die Beobachtung der zeitlichen Entwicklung von Krankheiten und lassen eine exakte zeitliche Trennung von (potenzieller) Ursache und Wirkung zu. Werden Exponierte und Nicht-Exponierte repräsentativ ausgewählt, erlauben Kohortenstudien die Schätzung von absoluten und relativen Krankheitsrisiken. Diesen Vorteilen steht der hohe Aufwand an Zeit und Kosten gegenüber. Die unter Umständen mehrjährige Dauer derartiger Studien kann zum Ausscheiden von einer nennenswerten Anzahl von Personen führen. Schließlich eignen sich Kohortenstudien primär für häufig auftretende Zielereignisse, will man die Studien- und Kontrollgruppen nicht zu groß werden lassen (Tabelle 7.2)

Tabelle 7.2: Vor- und Nachteile einer Längsschnittstudie

Vorteile	Nachteile
Zeitliche Entwicklung der Krankheiten und der Ursache-Krankheits-Beziehung wird sichtbar	zeit- und kostenintensiv
Trennung zwischen Ursache und Wirkung	Erfordert viele Versuchspersonen bei seltenen Krankheiten. Bereitschaft zur Mitarbeit? Selektion?
Hohes Maß an Repräsentativität, daher gute Schätzbarkeit der Auswirkungen bei Exponierten und Kontrollen	Lange Laufzeit, verbunden mit der Gefahr des Ausfalls von Patienten ('drop-outs')
Beobachtbarkeit von mehreren Zielgrößen möglich	Kontrolle und Erfassung von Störgrößen oder Confoundern kann unvollständig sein
Maßnahmen zur Qualitätssicherung der Studiendurchführung sind möglich	Änderung der Rahmenbedingungen möglich (Inanspruchnahme, Versorgung, Exposition)

7.5 Querschnittstudie

In einer Querschnittstudie (,cross-sectional-study') werden zu einem festen Zeitpunkt oder einem umschriebenen Zeitfenster bei einer definierten Studienpopulation sowohl Information zum Gesundheitszustand und Krankheiten erhoben als auch potenzielle Einflussgrößen. Da die erhobenen Angaben zu bestehenden Krankheiten prävalent sind und i.d.R. keine Angaben zu Erkrankungsdatum und Krankheitsdauer erhoben werden, wird diese Studienform auch als **Prävalenzstudie** bezeichnet.



Da naturgemäß keine zeitlichen Abhängigkeiten zwischen Einflussfaktoren und Krankheit untersucht werden können (Problem Henne-Ei), dienen Querschnittstudien im Rahmen der deskriptiven Epidemiologie der Aufdeckung assoziativer Zusammenhänge und der Hypothesengenerierung. Querschnittstudien haben trotz dieser Beschränkung eine Reihe von Vorteilen, von denen in erster Linie die Praktikabilität, die Kostengünstigkeit und die schnelle Durchführbarkeit genannt werden sollten (Tabelle 7.3).

Tabelle 7.3: Vor- und Nachteile einer Querschnittstudie

Vorteile	Nachteile
Relativ schnell an großer Zahl von Personen durchführbar	Probanden können selektiert sein (z.B. durch Arztkontakte, KH-Aufnahme, Beruf, Tod)
Relativ billig	Krankheitsereignis selbst mag Variable beeinflusst haben (Bsp. Myokardinfarkt - Blutdruck)
Auch für seltener vorkommende Krankheiten (SP-Größe)	Anamnestische Angaben begrenzt zuverlässig
Population und Untersuchungsmethode vom Untersucher zu wählen	Bei seltenen Krankheiten hohe Probandenzahlen nötig
Standardisierte Befragungen und Untersuchungen diagnostischer Kriterien möglich	Problematisch bei kurzdauernden (akuten) Krankheiten
Häufigkeitsvergleiche zwischen verschiedenen Populationen und Populationsgruppen möglich	Keine Angaben zur Inzidenz oder zum Risiko
Aufstellung von Hypothesen möglich	Testen von Hypothesen nicht möglich.

Quelle: Heinemann und Sinnecker (1994)

Querschnittstudien werden erleichtert, wenn sie auf **Public-Use-Files** basieren, also der wissenschaftlichen Allgemeinheit zur Verfügung gestellten Originaldaten für eine Sekundärnutzung. Für Deutschland ist hier der Bundesgesundheitsurvey 1998 zu nennen, der für eine repräsentative Bevölkerungsstichprobe von mehr als 7000 Personen ein Vielzahl von Angaben zum Gesundheitszustand, ärztlicher Inanspruchnahme, gesundheitsrelevantem Verhalten und weiteren Einflussfaktoren auf die Gesundheit für eine wissenschaftliche Sekundärnutzung bereitstellt. Der Bundesgesundheitsurvey ist in kurzer Zeit Grundlage für eine Vielzahl epidemiologischer Studien geworden (vgl. www.rki.de).

7.6 Korrelationsstudien

Korrelationsstudien („ökologische Studien“) basieren nicht auf individuell erhobenen Merkmalen, sondern auf aggregierten Daten. Beobachtungseinheiten sind also nicht einzelne Personen, sondern Gruppen von Individuen und deren Gruppenmerkmale. Beobachtungen können beispielsweise für Staaten, Landkreise, Haushalte, Krankenhäuser o.ä. erhoben werden.

Der Vorteil von Korrelationsstudien liegt neben ihrem explorativen Charakter darin, dass sie meist schnell und kostengünstig durchgeführt werden können und sich dabei häufig auf vorhandene (amtliche) Routinedaten stützen können. Das Problem bei derartigen Studien hingegen ist, dass von Gruppenmerkmalen nicht auf Eigenschaften der Gruppenmitglieder geschlossen werden kann. Eine Assoziation zwischen Variablen auf Aggregatniveau sagt nicht automatisch etwas über die Beziehung bei einzelnen Individuen. Der Zusammenhang zwischen Lebenserwartung und Entwicklungsstand einer Gesellschaft bedeutet beispielsweise nicht, dass nicht auch in sehr armen Ländern hochbetagte Menschen leben. Dieser mögliche Trugschluss bei der Übertragung von Zusammenhängen auf Aggregatniveau auf die individuelle Ebene wird als **ökologischer Trugschluss** bezeichnet.

Ein weiteres Beispiel für Korrelationsstudien ist der positive Zusammenhang zwischen Krankenhausbettendichte und Krankenhausfallhäufigkeit. Daran ist auch erkennbar, dass Korrelationsstudien wie ihr Name schon sagt, nur Assoziationen zwischen Variablen entdecken und für die Formulierung von Hypothesen nutzen, aber keine Aussage über kausale Beziehungen treffen können. Denn es ist auf der Aggregatebene nicht zu entscheiden, ob überdurchschnittliche viele Krankenhausbetten Ausdruck einer überdurchschnittlichen Inanspruchnahme aufgrund hoher Morbidität sind oder ob ein Überangebot von Betten dazu führt, dass die Indikation zur stationären Aufnahme großzügiger gestellt wird und/oder Patienten länger als notwendig im Krankenhaus behandelt werden.

7.7 Übersicht über die wichtigsten Studientypen

Neben den genannten Reinformen gibt es noch wie schon kurz erwähnt Mischformen. Dazu gehört beispielsweise eine Fall-Kontrollstudie, in der für Fälle und Kontrollen Expositionen retrospektiv erhoben werden, die Studienteilnehmer aber über einen gewissen Zeitraum nachverfolgt werden. Gebräuchlich sind auch Querschnittstudien, die zu prospektiven Studien ausgebaut werden können, wenn sich die Studienteilnehmer für Nachbeobachtungen bereit erklären.

Für die spätere Berechnung von Risiken für eine Krankheit in Abhängigkeit von einer Exposition oder einem Risikofaktor empfiehlt sich bei allen Studienformen eine Darstellung der zentralen Ergebnisse in einer Vierfeldertafel. Zu beachten ist dabei, dass bei einer Fall-Kontroll-Studie vom Krankheitsstatus ausgegangen wird, also die Zeilensumme der Zahl der Fälle und der Kontrollen, zu Beginn festgelegt wird. Bei der Kohortenstudie liegen die Spaltensumme der Exponierten und Nicht-Exponierten zu Beginn fest, bei einer Querschnittstudie schließlich nur die Zahl aller zu befragenden Personen unabhängig von ihrer Exposition und Krankheitsstatus (Tabelle 7.4).

Tabelle 7.4: Darstellung der wichtigsten epidemiologischen Studientypen in einer Vierfeldertafel

Erkrankung	Exposition		gesamt
	nein	ja	
nein	a	b	a + b
ja	c	d	c + d
gesamt	a + c	b + d	a + b + c + d
Querschnittstudie	a + b + c + d	fest	
Fall-Kontroll-Studie	a + b, c + d	fest	
Kohortenstudie	a + c, b + d	fest	

Die nachfolgende Tabelle 7.5 stellt die Vor- und Nachteile der vier Hauptstudienformen bzgl. der wichtigsten Verzerrungsgefahren (vgl. auch Kapitel 9) und des Aufwandes gegenüber. Es zeigt sich, dass zwar Fall-Kontroll- und Kohortenstudien leistungsfähiger als ökologische und Querschnittstudien sind, es aber in der Epidemiologie keinen Königsweg gibt, sondern sich die Wahl der geeigneten Studienform nach der inhaltlichen Fragestellung und den organisatorischen Rahmenbedingungen richten muss. Die typischen Anwendungsgebiete für die Studientypen sind abschließend in Tab. 7.6 aufgeführt.

Tabelle 7.5: Vor- und Nachteile verschiedener Typen von beobachtenden Studien

	ökologische Studie	Querschnittstudie	Fall-Kontroll-St.	Kohortenstudie
<i>Wahrscheinlichkeit von:</i>				
Selektionsbedingter Verzerrung	<i>nicht geeignet</i>	mäßig	hoch	gering
Erinnerungsbedingter Verzerrung	<i>nicht geeignet</i>	hoch	hoch	gering
Verlust für Nachuntersuchung	<i>nicht geeignet</i>	<i>nicht geeignet</i>	gering	hoch
Verwechslungen	hoch	mäßig	mäßig	gering
Zeitaufwand	gering	mäßig	mäßig	hoch
Kosten	gering	mäßig	mäßig	hoch

Quelle: Beaglehole et al. (1997)

Tabelle 7.6: Anwendungen verschiedener Typen von beobachtenden Studien

Anwendung	ökologische Studie	Querschnittstudie	Fall-Kontroll-St.	Kohortenstudie
Erforschung einer seltenen Krankheit	++++	-	++++	-
Erforschung einer seltenen Krankheitsursache	++	-	-	++++
Test multipler Wirkungen der Ursache	+	++	-	+++++
Untersuchung multipler Expositionen und Determinanten	++	++	+++	+++
Messung zeitlicher Beziehungen	++	-	+ bei prospektiver Studie	++++
Direkte Messung der Inzidenz	-	-	+ bei populationsbez. Studie	++++
Untersuchung langer Latenzzeiten	-	-	+++	-

Quelle: Beaglehole et al. (1997); + ... +++ gibt den Grad der Eignung an; - Studientyp für die Anwendung nicht geeignet

8. Zusammenhangsmaße

8.1 Relatives Risiko

Eine wichtige Aufgabe der Epidemiologie ist die Aufdeckung und Bewertung möglicher Krankheitsursachen und damit verbundener Risiken. Eine häufige Fragestellung lautet: Wie hoch ist das Risiko, an einer bestimmten Krankheit zu erkranken, wenn man einer gewissen Exposition ausgesetzt ist. Beispiele sind Rauchen und Lungenkrebs (Zusammenhang eindeutig bewiesen) oder Strahlenbelastung durch Kernkraftwerke und kindliche Leukämien (keine eindeutige Evidenz bislang). Expositionen müssen nicht zwangsläufig eine Gesundheitsgefährdung darstellen, sondern beschreiben allgemein das Vorliegen gewisser Rahmenbedingungen oder Dispositionen, von denen ein Einfluss auf Gesundheit und Krankheit vermutet wird. Unter Exposition kann danach auch ein gesundheitsbewusstes Verhalten (Sport, Ernährung) oder die Inanspruchnahme präventiver Leistungen verstanden werden (z.B. Krebsfrüherkennung, U-Untersuchungen im Kindesalter). Um eine Exposition bewerten zu können, wird als Referenzgröße meist eine Kontrollgruppe ohne diese Exposition herangezogen und die Krankheitshäufigkeit bei Exponierten und Nicht-Exponierten verglichen.

Zunächst wird dazu das absolute Risiko für die Erkrankung bestimmt. Dies haben wir schon als Inzidenz, also die Zahl der Neuerkrankungen pro Zeiteinheit, kennen gelernt. Das absolute Risiko ist meist wenig aussagefähig, da häufig die Risiken absolut gesehen sehr gering sind. Selbst unter starken Rauchern erkranken bei weitem nicht alle an Lungenkrebs. Das Risiko wird daher erst in einem Vergleich mehrerer Risiken interpretierbar. Dazu bietet sich die *Risikodifferenz* und das *Risikoverhältnis (relatives Risiko)* an. Zur Erläuterung diene das folgende hypothetische Beispiel zur Untersuchung der Ursachen einer durch Lebensmittel hervorgerufenen Erkrankung. Wir sehen, dass die Risikodifferenz keine eindeutige Einschätzung der verschiedenen Lebensmittel erlaubt. Gleiche Risikodifferenzen führen aufgrund unterschiedlichen absoluten Niveaus der Risiken zu stark unterschiedlichen relativen Risiken. Tabelle 8.1 legt die Vermutung nahe, dass vom Thunfischsalat das höchste Risiko für die Vergiftung die ausging; aber auch der Eiersalats ist kritisch zu bewerten, da durch ihn viele zusätzliche Erkrankungen hervorgerufen werden. Jedoch ist dessen Wirkung weniger spezifisch, weil auch die Mehrheit der Nichtkonsumenten erkrankte.

Tabelle 8.1: Mögliche Ursachen einer Lebensmittelvergiftung

Lebensmittel	aßen (davon krank in %)	aßen nicht (davon krank in %)	Risikodifferenz	Risikoverhältnis
Eiersalat	93	63	30	1,48
Eis	78	64	14	1,21
Hüttenkäse	37	23	14	1,61
Thunfischsalat	36	6	30	6,00

Quelle: Gordis (2001), modifiziert

Aus diesem Grund ist das relative Risiko (=RR), also das Risiko für eine Erkrankung (=D) bei Exposition (=E) geteilt durch das Risiko bei Nicht-Exponierten oder

$$RR = P(D = 1 | E = 1) / P(D = 1 | E = 0) \quad (8:1a)$$

das zentrale Maß in der Epidemiologie zur Bewertung möglicher Krankheitsursachen. Die Deutung des relativen Risikos ergibt sich wie folgt:

- RR = 1: Das Risiko der Exponierten ist gleich groß wie das Risiko der Nicht-Exponierten (kein Zusammenhang)

- $RR > 1$: Das Risiko der Exponierten ist größer als das Risiko der Nicht-Exponierten (positiver Zusammenhang, Exposition möglicherweise krankheitsverursachend bzw. krankheitsfördernd)
- $RR < 1$: Das Risiko der Exponierten ist kleiner als das Risiko der Nicht-Exponierten (negativer Zusammenhang, Exposition möglicherweise protektiv)

Da sich das Risiko definitionsgemäß über die Zahl der Neuerkrankungen berechnet, kann es nur aus prospektiven Kohorten- bzw. Längsschnittstudien berechnet werden, in dem eine Studiengruppe unter Exposition und eine Kontrollgruppe ohne Exposition auf das Auftreten einer oder mehrerer Krankheiten beobachtet werden. Als Ergebnis ergibt sich Tabelle 8.2.

Tabelle 8.2: Berechnung des relativen Risikos in einer Kohortenstudie

Exposition	Erkrankung			Inzidenz der Erkrankung
	ja	nein	gesamt	
ja	a	b	a + b	a / (a + b)
nein	c	d	c + d	c / (c + d)
gesamt	a + c	b + d	a + b + c + d	

Das relative Risiko bei Vorliegen einer Exposition berechnet sich dann als

$$\text{Relatives Risiko (RR)} = \frac{a / (a + b)}{c / (c + d)} \quad (8:1b)$$

Es gibt statistische Verfahren, auf die hier nicht näher eingegangen werden kann, die eine Abweichungen des beobachteten relativen Risikos von 1 auf Signifikanz prüfen. Lediglich der Vollständigkeit halber sei die Testgröße erwähnt. Ausgehend von der inzwischen bekannten Vierfeldertafel ist sie wie folgt definiert:

$$\chi^2 = \frac{a / (a + b)}{c / (c + d)} \quad (8:2)$$

Wenn nicht alle Teilnehmer einer Kohortenstudie über die gesamte Laufzeit beobachtet werden können, werden für die Berechnung des relativen Risikos Inzidenzdichten bei exponierten Personen und Kontrollen benötigt. Das relative Risiko ergibt sich dann als Quotient dieser Inzidenzdichten (Tabelle 8.3).

Tabelle 8.3: Beziehung zwischen Serumcholesterinspiegel und KHK-Risiko nach Alter und Geschlecht (die ersten 12 Jahre der Framingham-Studie)

Chol (mg/100ml)	Männer		Frauen	
	30 bis 49 Jahre	50 bis 62 Jahre	30 bis 49 Jahre	50 bis 62 Jahre
	<i>Inzidenzraten (pro 1.000)</i>			
< 190	38,2	105,7	11,1	155,2
190 bis 219	44,1	187,5	9,1	88,9
220 bis 249	95,0	201,1	24,32	96,3
250 und höher	157,5	267,8	50,4	121,5
	<i>Relative Risiken (bez. auf Männer mit niedrigem Cholesterin, < 190 mg/100ml)</i>			
< 190	1,0	2,8	0,3	4,1
190 bis 219	1,2	4,9	0,2	2,3
220 bis 249	2,5	5,3	0,6	2,5
250 und höher	4,1	7,0	1,3	3,2

Quelle: Gordis (2001)

Selten kann in der Epidemiologie die Untersuchung des Zusammenhangs zwischen Exposition und Erkrankungshäufigkeit auf eine Vierfeldertafel begrenzt werden. Vielmehr müssen häufig weitere potenzielle Einflussfaktoren und Confounder wie Alter, Geschlecht oder andere Risikofaktoren berücksichtigt werden. Darauf wird in den Kapiteln 9 und 10 näher eingegangen.

8.2 Odds Ratio

Für die Berechnung des relativen Risikos werden Daten zur Krankheitsinzidenz bei exponierten und nicht-exponierten Personen benötigt. Bei einer Fall-Kontroll-Studie sind die Inzidenzen jedoch nicht bekannt, da von Erkrankten (Fällen) und Nicht-Erkrankten (Kontrollen) ausgegangen wird. Dennoch lässt sich unter bestimmten Bedingungen auch in einer Fall-Kontroll-Studie ein Näherungswert für das relative Risiko berechnen.

In einer Fall-Kontroll-Studie lässt sich bei Fällen und Kontrollen retrospektiv der Anteil der exponierten Personen bestimmen und miteinander in Beziehung setzen. Dazu verwenden wir **Odds** (Chancen). Die Chance für ein bestimmtes Ereignis ist allgemein der Quotient aus der Zahl der Möglichkeiten für ein bestimmtes Ereignis und der Zahl der Möglichkeiten für das Ausbleiben des Ereignisses. Bezogen auf das Würfeln mit einem Würfel beträgt die Chance für eine gerade Zahl 1, nämlich die drei Zahlen 2, 4 und 6 bezogen auf die ungeraden Zahlen 1, 3 und 5.

Übertragen auf eine Fall-Kontroll-Studie können wir sowohl für die Fälle als auch für die Kontrollen die Odds berechnen, dass eine Exposition vorlag. Für Fälle ist die Odds durch den Quotienten a/c gegeben, nämlich die Zahl der Fälle mit Exposition geteilt durch die Zahl der Fälle ohne Exposition. Für die Kontrollen ergibt sich analog die Chance als b/d (Tabelle 8.4; im formalen Aufbau gleich der Tabelle 8.2).

Tabelle 8.4: Berechnung des Odds Ration in einer Fall-Kontroll-Studie

Exposition /Risikofaktor	Erkrankung		gesamt
	ja	nein	
ja	a	b	a + b
nein	c	d	c + d
gesamt	a + c	b + d	a + b + c + d

Als Zusammenhangsmaß zwischen Erkrankung und Exposition bietet sich nun das Verhältnis (engl. ratio) der beiden Chancen an, die sogenannte Odds Ratio. Sie ist definiert als das Verhältnis der Odds, dass Fälle exponiert waren, zu den Odds, dass die Kontrollen exponiert waren. Sie berechnet sich danach als

$$\text{Odds Ratio (OR)} = \frac{a/c}{b/d} = \frac{a * d}{b * c} \quad (8:2)$$

Die Odds-Ratio ist also der Quotient des Produkts der Werte aus der Hauptdiagonalen und der Nebendiagonalen und wird deshalb auch als **Kreuzproduktverhältnis** bezeichnet.

Die Odds Ratio, die im übrigen auch in prospektiven Studien berechnet werden kann, wird ähnlich wie das relative Risiko interpretiert. Ohne einen Zusammenhang zwischen Exposition und Krankheit wird sich in Fall-Kontroll-Studien eine OR von ungefähr 1 ergeben. Je stärker die Exposition positiv mit der Krankheit assoziiert ist, also ein Risiko(-faktor) darstellt, desto größer ist der Quotient der beiden Chancen. Ein Wert kleiner 1 deutet dagegen auf einen protektiven Einfluss der Exposition hin. In Tabelle 8.5 wird die Berechnung der Odds Ratio für eine hypothetische Studie zum Zusammenhang zwischen Rauchen und Herz-Kreislauf-erkrankungen verdeutlicht.

Tabelle 8.5: Beispiel für die Berechnung des Odds Ratio in einer Fall-Kontroll-Studie

Exposition	KHK-Fälle	Kontrollen	gesamt
Raucher	112	176	288
Nichtraucher	88	224	312
gesamt	200	400	600
Anteil Raucher	56%	44%	
Odds Ratio = $112 \cdot 224 / 176 \cdot 88 = 1,62$			

8.3 Zusammenhang zwischen Relativem Risiko und Odds ratio

In einer Fall-Kontroll-Studie kann nur die Odds-Ratio als Zusammenhangsmaß berechnet werden, während sich in einer prospektiven Studie sowohl das relative Risiko als auch die Odds-Ratio bestimmen lassen. Unter gewissen Umständen kann aus der Odds Ratio einer Fall-Kontroll-Studie allerdings das relative Risiko geschätzt werden:

- wenn die Fälle bzgl. der Expositionsanamnese repräsentativ sind für alle Menschen mit dieser Krankheit aus der Bevölkerung;
- wenn die Kontrollen bzgl. der Expositionsanamnese repräsentativ sind für alle Menschen ohne diese Krankheit aus der Bevölkerung;
- wenn die untersuchte Krankheit selten auftritt.

Die letzte Voraussetzung verdeutlicht man sich mit den Bezeichnungen aus der Vierfeldertafel (vgl. Tabelle 8.2 und 8.4). Danach sind die beiden Maße definiert als

$$OR = (a/c) : (b/d)$$

und

$$RR = (a / (a+b)) : (c / (c+d)).$$

Das relative Risiko ist der Quotient aus den Inzidenzen bei Exponierten und Nicht-Exponierten, die Odds Ratio der Quotient der Chancenverhältnisse für eine Exposition bei Erkrankten und Nicht-Erkrankten. Bei seltenen Erkrankungen werden nur wenige der Exponierten auch tatsächlich erkranken. Daher kann näherungsweise $a+b$ mit b gleichgesetzt werden. Ebenso kann für die Nicht-Exponierten $c+d$ durch d approximiert werden. Daher kann bei seltenen Erkrankungen (Prävalenz geringer als 5 %) das relative Risiko durch den Ausdruck $a \cdot d / b \cdot c$ geschätzt werden, was der Formel für die Odds Ratio entspricht. Dieser Zusammenhang ermöglicht es, das relative Risiko auch in einer Fall-Kontroll-Studie (zumindest approximativ) zu schätzen. Aus diesem Grund werden beide Begriffe häufig synonym verwendet, auch wenn dies aus methodischen Gründen nicht zulässig ist.

8.4 Attributables Risiko

Das relative Risiko und das Odds Ratio beschreiben das Risikoverhältnis zwischen Exponierten und Nichtexponierten und damit etwas über die Stärke des Zusammenhangs zwischen Exposition und Krankheit. Das attributable Risiko gibt dagegen, an, welcher Anteil einer Krankheit auf die Exposition zurückzuführen ist und damit bei Verringerung oder Eliminierung des Risikos theoretisch vermieden werden könnte. Dabei kann unterschieden werden zwischen dem

- **attributables Risiko unter den Exponierten:** die Krankheitsrate unter den Exponierten, die auf die Exposition zurückgeführt werden kann und dem

- **bevölkerungsbezogenen attributablen Risiko:** die Krankheitsrate in der Gesamtbevölkerung, die auf die Exposition zurückgeführt werden kann

Ausgedrückt durch Inzidenzen sind die beiden attributablen Risiken definiert als

- **attributables Risiko unter den Exponierten:**
Inzidenz der Exponierten – Inzidenz der Nichtexponierten
- **bevölkerungsbezogenes attributables Risiko:**
Inzidenz in der Gesamtpopulation – Inzidenz der Nichtexponierten

Beide Maßzahlen werden häufig auch prozentual dargestellt (Tabelle 8.6).

Tabelle 8.6: Beispiel für die Berechnung des attributablen Risiko

Todesursache	Todesfälle je 1.000 Personen pro Jahr			
	Gesamt- population	Nichtraucher	alle Raucher	starke Raucher (> 24 Zig. / d)
alle Ursachen	14,05	12,06	16,32	19,67
Lungenkrebs	0,65	0,07	1,20	2,23
koronare Herzkrankheit	3,99	3,31	4,57	4,97
attributables Risiko für Lungenkrebs (der Raucher): pro 1.000 Personen und in %			1,13 / 94 %	2,16 / 97%
attributables Risiko für Lungenkrebs (bev.-bezogen)			0,58 / 89 %	
attributables Risiko für KHK (der Raucher): pro 1.000 Personen und in %			1,26 / 38 %	1,66 / 33%
attributables Risiko für KHK (bev.-bezogen)			0,68 / 17 %	

Quelle: Unwin et al. (1991)

9. Aspekte der Studienplanung

9.1 Verzerrungsmöglichkeiten

Epidemiologische Studien können durch eine Reihe systematischer Fehler zu verzerrten Ergebnissen führen. Bei der Planung von Studien muss daher besondere Sorgfalt auf die Identifizierung und Eliminierung möglicher Verzerrungsquellen geachtet. In der Epidemiologie werden diese **Bias** genannt, worunter systematische Abweichungen von Messungen oder Schlussfolgerungen vom wahren Sachverhalt verstanden werden. Diese können ihre Ursachen im Studiendesign, der Datenerhebung und –auswertung, der Ergebnisinterpretation oder der Publikation der Ergebnisse haben.

Es können verschiedene Formen des Bias unterschieden werden:

- **Detection/diagnosis bias:** Fehler bei der Erhebung der Daten oder der Diagnose bzw. Verifikation von Fällen gegenüber Kontrollen, z.B. durch Anwendung eines diagnostischen Tests bei Krankenhausfällen und nicht bei Bevölkerungskontrollen.
- **Information bias:** Unterschiedliche Genauigkeit in der Messung und Erhebung von Daten in verschiedenen Gruppen, z.B. intensivere Diagnostik in der Studiengruppe als bei Kontrollen.
- **Observer bias:** Verzerrung durch interpersonelle (zwischen mehreren Personen) oder intrapersonelle (zu verschiedenen Zeitpunkten bei der gleichen Person) Variation in der Beobachtung oder Erhebung von Daten, z.B. in der Bewertung diagnostischer Verfahren wie Röntgenaufnahmen oder pathologischen Befunden.
- **Interviewer bias:** Verzerrung, die durch die Person des Interviewers und sein Auftreten hervorgerufen wird.
- **Recall bias:** Verzerrung durch unterschiedliche Genauigkeit bei der retrospektiven Erhebung von Information aus der Erinnerung, etwa zwischen Eltern von leukämiekranken Kindern und Eltern von Kontrollen.
- **Reporting bias:** Verzerrung durch selektive Nicht-Berichterstattung bestimmter Ereignisse wie sexuell übertragbare Krankheiten oder Alkoholmissbrauch.
- **Response bias:** systematischer Fehler durch die Antworten von Personen, die freiwillig oder nicht an einer Erhebung teilnehmen, z.B. bei einer Volkszählung zwischen Befürwortern und Gegner einer derartigen Erhebung.
- **Selektionsbias:** Verzerrung aufgrund der Unterschiede zwischen Teilnehmern und Nichtteilnehmern an einer Studie, z.B. unterscheiden sich Teilnehmer an Früherkennungsuntersuchungen in vielen sozio-ökonomischen Merkmalen von Nichtteilnehmern.
- **Stichprobenbias:** Verzerrung durch Wahl eines Stichprobenverfahrens, bei dem nicht jedes Element der Grundgesamtheit die gleiche Wahrscheinlichkeit hat gezogen zu werden oder bei dem diese Wahrscheinlichkeit nicht bekannt ist. Das ist offensichtlich bei Straßenumfragen; aber auch der Fall, wenn bei Haushaltsbefragungen die Auswahl der Haushalte über die Bewohner erfolgt und dann große Haushalte eine größere Ziehungswahrscheinlichkeit als kleinere Haushalte besitzen.
- **Length sampling bias:** Verzerrung durch überproportionalen Anteil von langen Fällen in einer Studie, z.B. bei symptomloser Früherkennung oder bei Umfragen in Vertragsarztpraxen, bei denen Stammpatienten eher ausgewählt werden als neue Patienten bzw. Vertretungsfälle.
- **Lead time bias:** Verzerrung durch unterschiedlich lange Beobachtungszeiten, z.B. zwischen Karzinomen, die in der Früherkennung oder nach Beschwerden erkannt werden, auch wenn die Behandlung keinen Einfluss auf den Verlauf hat.

- **Withdrawal bias:** Unterschiede, die durch nicht neutrale Ausfälle aus einer Studie hervorgerufen werden, z.B. wenn aus einer Längsschnittstudie primär chronisch kranke Personen aufgrund ihres Gesundheitszustandes ausscheiden.

Vor allem aus der Sozialwissenschaft sind weitere Verzerrungsmöglichkeiten bekannt. Für die Epidemiologie seien hier die beiden wichtigsten genannt. Im übrigen sei auf Lehrbücher der Medizinischen Soziologie und der empirischen Sozialforschung verwiesen.

- **Healthy worker effect:** Positive Selektion in bestimmten Berufen, weil die gesundheitlich angeschlagenen aus der Tätigkeit oder aus dem Berufsleben insgesamt ausscheiden bzw. schon bei der Rekrutierung gesundheitliche Aspekte eine Rolle spielen. Das kann z.B. in Untersuchungen bei Pflegekräften aufgrund des verbreiteten Burn-out-Syndroms der Fall sein
- **Hawthorne-Effekt:** Die Kenntnis an einer Untersuchung teilzunehmen, kann für sich genommen schon das übliche Verhalten verändern, etwa im Hinblick auf sozial gewünschtes Verhalten oder weil sich Probanden bei physiologischen Tests (Hören, Sehen, Konzentration) mehr als gewohnt anstrengen.

9.2 Confounder

Auch wenn die oben genannten Verzerrungsmöglichkeiten geprüft und ausgeschlossen bzw. kontrolliert sind, kann die Interpretation von Studienergebnissen durch Confounder verfälscht werden. Ein Confounder ist eine Variable, die den beobachteten Zusammenhang zwischen einer unabhängigen Größe (in Abb. 9.1: Kaffeekonsum als Risikofaktor) und Zielgröße (hier: koronare Herzkrankheit) beeinflusst, da sie mit beiden Größen in Zusammenhang steht: Rauchen erhöht das Krankheitsrisiko und es gibt gleichzeitig eine positive Korrelation zwischen Rauchen und Kaffeetrinken.

Ein Confounder ist also ein Drittfaktor, der

- (1) ein Risikofaktor für die Studienkrankheit und
- (2) mit der zu untersuchenden Exposition assoziiert ist, ohne nur unmittelbare Konsequenz dieser Exposition zu sein.

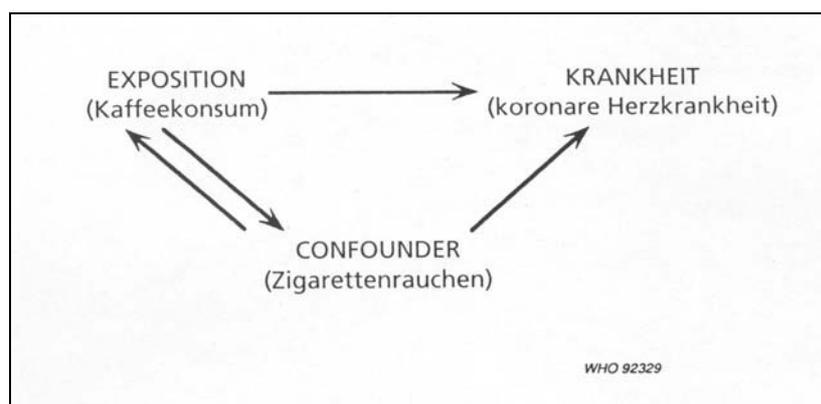


Abbildung 9.1: Beispiel eines Confounders (Quelle: Beaglehole et al. 1997)

Ob eine Variable ein Confounder ist, lässt sich nur anhand des unterstellten Erklärungsmodells und /oder der gestellten Frage beantworten. Es gibt hier keine allgemeingültigen Antworten. Relevante Confounder in der Epidemiologie sind jedoch in der Regel:

- Alter
- Geschlecht
- Rasse, Nationalität

- Soziale Schicht
- Komorbidität
- Ernährung
- Gesundheitsverhalten
- Weitere Risikofaktoren

Möglichkeiten der Behandlung von Confoundern sind:

- **Restriktion**, z.B. werden nur Personen einer bestimmten Altersgruppe untersucht;
- **Stratifizierung**: Der Zusammenhang zwischen Einfluss- und Zielgröße wird für die einzelnen Ausprägungen eines Confounders separat untersucht (vgl. in Kap. 10);
- **Matching** (individuell oder gruppenweise): Bei der Wahl der Studien- und Kontrollgruppe und der Rekrutierung der Studienteilnehmer wird auf gleiche Ausprägung der Confounder bei individuellen Beobachtungspaaren oder beiden Gruppen geachtet;
- **Standardisierung**: Wird meist nach Alter und Geschlecht vorgenommen, kann aber auch unter Einbeziehung anderer Merkmale erfolgen. Dabei werden gewisse Kennzahlen einer Untersuchungsgruppe auf eine hinsichtlich relevanter Merkmale bekannte Standardpopulation umgerechnet und damit vergleichbar gemacht (vgl. Kapitel 6 zur Standardisierung epidemiologischer Kennzahlen);
- **Multivariate statistische Verfahren**. Diese untersuchen den Zusammenhang zwischen potenziellen Einfluss- und Zielgrößen bei gleichzeitiger Berücksichtigung von Confoundern

9.3 Stichproben

Die Ziehung von Stichproben ist ein wichtiger Punkt bei der Planung und Durchführung epidemiologischer Studien. Zwar erfolgt in Fall-Kontroll-Studien häufig eine Vollerhebung aller relevanten Fälle nach gewissen Einschlusskriterien. Doch bei der Rekrutierung von Kontrollen, aber auch in den anderen wichtigen epidemiologischen Studientypen werden meistens Stichproben aus der Grundgesamtheit aller Personen gezogen, die gewissen Ein- und Ausschlusskriterien genügen. Eine Stichprobe ist eine Auswahl aus einer Grundgesamtheit, deren einzelne Elemente durch ein oder mehrere gleiche Eigenschaften gekennzeichnet sind. Es kann unterschieden werden in die

Erhebungseinheit	Einzelpersonen oder auch Haushalte (wie beim Mikrozensus) oder einzelne Krankenhaus-Stationen - und die
Untersuchungseinheit	Einzelpersonen, die Mitglieder eines Haushalts oder die Mitarbeiter einer Station.

Die Aussagen einer Untersuchung können sich sowohl auf die Untersuchungseinheiten beziehen - z.B. auf die Arbeitszufriedenheit der Mitarbeiter einer Station eines Krankenhauses - als auch auf die Erhebungseinheiten - z.B. Qualifikation, Qualitätssicherungsmaßnahmen, Anzahl der Überstunden pro Station.

Das Ziel einer Stichprobenerhebung liegt in

- Repräsentationsschluss (empirisch; z.B. Schätzung eines bestimmten Parameters)
- Inklusionsschluss (analytisch, Test einer bestimmten Hypothese)

Voraussetzungen für eine aussagefähige Stichprobe:

- I) Die Stichprobe muss sich hinsichtlich der Heterogenität ihrer Elemente und der Repräsentativität ihrer Variablen der Grundgesamtheit ähneln.

Beispiel: Eine Befragung unter Patienten der Magdeburger Universitätsklinik über ihre Zufriedenheit mit der Behandlung ist nicht repräsentativ für die Gesamtheit aller in Magdeburg versorgten Patienten (hinsichtlich Diagnoseverteilung, Krankheitsschweregrad, Herkunft etc)

II) Die Grundgesamtheit muss definiert werden können.

Beispiel: Bezieht sich die Befragung zur Patientenzufriedenheit auf alle behandelten Patienten? Sollen Patienten aus der Ambulanz oder ambulant Operierte ebenfalls befragt werden?

III) Die Elemente der Stichprobe müssen definiert werden können.

Beispiel: Patienten an einem bestimmten Tag auf festgelegten Stationen

IV) Das Auswahlverfahren muss klar sein und Forderung 1 genügen.

Beispiel: Es darf keine willkürliche Befragung zu einem beliebigen Tag in einer zufällig ausgewählten Station stattfinden, sondern es müssen die Tage, Auswahlätze (aus jeder Station ein gewisser Prozentsatz) etc. vorab festgelegt werden.

Die Stichprobenverfahren lassen sich im einzelnen wie folgt einteilen (Tabelle 9.1).

Tabelle 9.1: Systematik der Stichprobenauswahlverfahren (am Beispiel einer Befragung zur Patientenzufriedenheit)

willkürlich	Befragung von Patienten im Freigelände der Magdeburger Uniklinik
bewusst	nicht alle Elemente der Grundgesamtheit haben die Chance, in die SP zu gelangen, oder die Wahrscheinlichkeit hierfür ist nicht anzugeben z.B. nach: Verteilungen (z.B. ausgewählte Kliniken) Struktur (nur auf Intensivstationen behandelte Patienten) Zugehörigkeit (z.B. Patienten aus bestimmten Berufsgruppen) Sonderfall: Extremgruppen (z.B. ausländische Patienten) seltene Fälle (seltene Krankheitsbilder)
Quotaverfahren	Vorgabe einzelner (z.B. aus Vollerhebungen bekannter) Randverteilungen bei wichtigen unabhängigen Variablen z.B. Die Auswahl befragter Patienten soll bezüglich Alter, Geschlecht, Klinik bzw. Station der Gesamtheit aller an der Uniklinik behandelten Patienten entsprechen (diese Randverteilungen sind aus Leistungsstatistik bekannt)
Zufallsauswahl	Alle Elemente der Grundgesamtheit habe die gleiche oder zumindest bekannte Wahrscheinlichkeit, in die Stichprobe zu gelangen

Willkürliche Stichprobenverfahren entsprechen z.B. die in Massenmedien weit verbreiteten (TED-)Umfragen zu einem aktuellen Thema. Die daran teilnehmenden Personen sind aufgrund vielfältiger Verzerrungen in keiner Weise repräsentativ für die Grundgesamtheit. Daher verbieten sich auch darauf aufbauende Schlussfolgerungen. Systematische Stichprobenverfahren erlauben in der Regel ebenfalls keinen Schluss auf die allgemeine Grundgesamtheit, der sie entstammen. Dennoch können Sie angebracht sein, wenn Aussagen über besondere Teilgruppen der Grundgesamtheit getroffen werden sollen. Bei einem Quotaverfahren wird sichergestellt, dass die endgültige Stichprobe in zentralen Strukturparametern ein Abbild der Grundgesamtheit darstellt. Das gelingt aber nur, wenn zumindest am Ende der Stichprobenziehung vom Prinzip der Zufälligkeit abgegangen wird, wenn z.B. noch x Personen aus der Altersklasse über 65 Jahre oder y Männer rekrutiert werden müssen.

Am geeignetsten sind in der Regel Zufallsstichproben, bei denen ein rechnerischer Ausgleich für Abweichungen in der Zusammensetzungen der Stichprobe von der der Grundgesamtheit vorgenommen werden kann. Sie lassen sich weiter unterteilen (Tabelle 9.2)

Tabelle 9.2: Systematik zufälliger Stichprobenverfahren

Einfache Zufallsauswahl	gleiche Wahrscheinlichkeit für alle Elemente der Grundgesamtheit (GG) Bsp: Patienten, die an einem bestimmten Tag an der Uniklinik behandelt werden; Patienten mit einem bestimmten Anfangsbuchstaben oder Geburtstag Vorsicht: das Auswahlverfahren darf nicht mit Variablen der Erhebung korreliert sein (Patienten am fünften Tag des Aufenthalts)
Geschichtete Zufallsauswahl	Einteilung der GG in Schichten, dort gleiche Auswahlwahrscheinlichkeit Bsp: Die Schichten könnten durch die einzelnen Stationen oder Kliniken definiert sein; Zufallsauswahl in jeder Schicht (z.B. jeweils 5% der Patienten)
Mehrstufige Zufallsauswahl	auf der Grundlage einer multiplen Schichtung Bsp: Neben der Station könnte auch eine Schichtung nach Alter erfolgen (je Station und Altersgruppe 5%)
Klumpenauswahl	Die Elemente der GG sind in Klumpen zusammengefasst, die zufällig ausgewählt und dann vollständig erhoben werden. Bsp: Auswahl einzelner Stationen oder Krankenzimmer und Befragung aller dort behandelten Patienten

Verzerrungen in epidemiologischen Studien können einerseits aus dem Tatbestand resultieren, dass die Stichprobe kein Abbild der Grundgesamtheit darstellt, also auch kein Schätz- oder Testverfahren für Eigenschaften der Grundgesamtheit zulässt. In derartigen Fällen ist also die Stichprobenplanung und -ziehung mangelhaft.

Andererseits ist bei vergleichenden Studien der Fall denkbar, dass zwar sowohl Fälle als auch Kontrollen repräsentativ für ihre jeweilige Grundgesamtheit sind, diese Grundgesamtheiten sich aber unterscheiden, so dass Unterschiede in der Exposition nicht ohne weiteres verallgemeinert werden dürfen. Klassisches Beispiel dafür ist die Verwendung von Krankenhauskontrollen in Fall-Kontroll-Studien. So können bei einer schwerwiegenden Erkrankung, die ausnahmslos zu einer stationären Behandlung führt, zwar die im Krankenhaus rekrutierten Fälle repräsentativ für die Bezugsbevölkerung sein, möglicherweise aber nicht die Kontrollen. Etwa wenn die interessierende Exposition auch andere gesundheitliche Beeinträchtigungen mit nachfolgendem Krankenhausaufenthalt haben kann. Derartige Verzerrungen sind dann streng genommen nicht dem Stichprobenverfahren, sondern einer unangemessenen Abgrenzung der Grundgesamtheit geschuldet.

9.4 Datenschutz

Die Einhaltung der Datenschutzbestimmungen stellt häufig eine besondere Aufgabe bei der Planung und Durchführung von epidemiologischen Studien dar. Dies gilt vor allem dann, wenn personenbezogene Daten erhoben und für einen längeren Zeitraum für eine längsschnittliche Untersuchung gespeichert werden sollen. In solchen Fällen ist meist eine Einverständniserklärung der Studienteilnehmer unerlässlich. Aber auch die Sekundärnutzung von epidemiologisch relevanten Daten kann datenschutzrechtlich sensibel sein, etwa die Auswertung von Routinedaten der gesetzlichen Krankenkassen, auch wenn diese vorher anonymisiert worden sind.

Ohne auf Einzelheiten eingehen zu können, bedürfen folgende Punkte einer eingehenden Prüfung und Klärung zu Beginn einer epidemiologischen Untersuchung:

- Abstimmung des Procedere zur Erhebung und Datenhaltung mit dem Datenschutzbeauftragten und abschließende Genehmigung; ggf. Einbeziehung der zuständigen Ethikkommission

- Einverständniserklärung der Probanden, darin Aufklärung über die Studienziele, Hinweis auf Freiwilligkeit der Teilnahme, Hinweis auf Widerrufsrecht
- Gewährleistung der ärztlichen Schweigepflicht
- Anonymisierung gespeicherter Daten und deren anonymisierte Zusammenführung (z.B. Krankenhausdiagnose – Befragungsdaten)
- Gewährleistung der Datensicherheit, dazu gehören technische Vorkehrungen und Datenschutzverpflichtung der Mitarbeiter

Da Bundesdatenschutzgesetz kann unter www.juris.de heruntergeladen werden.

9.5 Arbeitsschritte in einer wissenschaftlichen Studie

In allen wissenschaftlichen Forschungsprozessen lassen sich im Prinzip sechs Phasen unterscheiden. Das gilt grundsätzlich auch für epidemiologische Untersuchungen. Die einzelnen Phasen mit den jeweiligen Arbeitsschritten seien hier ohne weitere Kommentare aufgeführt. Es bleibt bei spezifischen Studien zu prüfen, inwieweit die Schritte tatsächlich notwendig sind bzw. einer Modifizierung bedürfen.

1. Problemformulierung

vorläufige Formulierung der Fragestellung
wissenschaftliche Recherche zur Erschließung des Themas
Prüfung vorhandener (Sekundär-)Datenquellen
Abgrenzung des Themas

2. Festlegung Studiendesign

Festlegung der Zielgruppe der Untersuchung und ggf. von Kontrollpersonen
Formulierung detaillierter Fragestellungen und Hypothesen
Überlegungen zur Stichprobe, Erhebungsmethodik und Erhebungsinhalte
Festlegung des Auswertekonzepts
Planung und Organisation der Feldphase (und ggf. vorher eines Prätests)
Instrumentenentwicklung (Verwendung validierter Elemente, Verwendung von Standard-Schlüsselverzeichnissen (z.B. ICD)
Planung der Qualitätssicherung (Dokumentation der Arbeitsschritte, Interviewerschulung, Standards der Dateneingabe und Kodierung)

3. Datenerhebung

Durchführung des Prätests, falls notwendig
Durchführung der Hauptstudie und evtl. Auswertung von Sekundärquellen
Wahl der Untersuchungsobjekte, Ziehung der Stichprobe
Durchführung von Interviews und Beobachtungen, Versendung von Fragebögen
Physiologische Messungen
evtl. Mahnung, Nacherhebung
Maßnahmen zur Optimierung der Response, z.B. finanzielle Entschädigung, Erinnerungsschreiben, telefonische Ansprache

4. Datenkodierung, -eingabe, Fehlerkorrektur

Anpassung von Eingabe- und Datenhaltungssystemen
Kodierung und Klassifikation der Antworten
visuelles Editing eingehender Fragebögen
Validierung (z.B. durch Wiederholungsbefragung, Zwischenauswertungen)
Erfassung nicht-standardisierter Ergebnisse (z.B. freier Interviews)
Eingabe in EDV

Fehlersuche und Korrektur (z.B. Vollständigkeit, falsche Angaben, Inkonsistenzen Plausibilitätsprüfungen), Dokumentation der Änderungen
Zeitpunkte für Zwischen- und Endauswertungen festlegen
evtl. Nacherhebung

5. Datenanalyse

explorative, deskriptive und graphische Verfahren
analytische (qualitative und quantitative) Verfahren
Confounder-Kontrolle, multivariate Verfahren
Prüfung der Hypothesen entsprechend dem eingangs festgelegten Auswertepan
Unterscheidung von Hypothesentests und explorativen Auswertungen
Generierung neuer Hypothesen für spätere Untersuchungen

6. Interpretation, Bericht, Publikation

Darstellung der Ergebnisse, Schlussfolgerungen, Konsequenzen
kritische Bewertung der Methoden, Ergebnisse und Schlussfolgerungen
Bezug auf andere Studien mit gleicher oder ähnlicher Fragestellung
Ausblick auf neue Forschungsarbeiten

9.6 Beispiel zur Studienplanung

Anhand einer kleinen epidemiologischen Studie zur Gesundheit von Arbeitslosen soll auf zentrale Aspekte der Studienplanung beispielhaft eingegangen werden (vollständige Publikation im Anhang). Vorab muss man erwähnen, dass die referierte Studie im Rahmen eines einsemestrigen Praktikums mit Medizinstudenten im vorklinischen Studienabschnittes durchgeführt wurde. Daraus erklären sich einige der nachfolgend erläuterten und auch in der Publikation erwähnten Besonderheiten.

Vorläufige Fragestellung: In den ersten Jahren nach der Wende wurden eine Vielzahl von Dauerarbeitslosen, häufig mehrfach, in ABM-Maßnahmen vermittelt. Dass diese Arbeiten die Chancen der Arbeitssuchenden auf dem ersten Arbeitsmarkt nicht wesentlich erhöhten, ist bekannt. Doch wie sieht es mit den gesundheitlichen Auswirkungen dieser Maßnahmen aus?

Felderschließung: Es gibt eine Vielzahl von empirischen Studien zum Zusammenhang zwischen Arbeitslosigkeit und Gesundheit. Eine gegenseitige Abhängigkeit ist offenkundig, wobei es gesicherte Evidenz sowohl für die Selektionshypothese (‚Krankheit verursacht Arbeitslosigkeit‘) als auch die Kausalitätshypothese (‚Arbeitslosigkeit macht krank‘) gibt. Die Recherche ergab weiterhin Hinweise auf positive gesundheitliche Auswirkungen einer Wiederaufnahme einer geregelten Tätigkeit. Dieser Befund deckt sich im übrigen mit beiden Theorien, denn gesunde Arbeitslose haben größere Chancen auf dem Arbeitsmarkt als kranke. Es wurden allerdings keine Studien gefunden, die die Auswirkung einer ABM-Tätigkeit zum Gegenstand hatten. Es konnten auch keine Sekundärdatenquellen für die Beantwortung der Problemstellung genutzt werden.

Abgrenzung des Themas: Unser endgültiges Thema lautete demnach: Hat eine ABM-Tätigkeit einen Einfluss auf die Gesundheit?

Studiengruppe: Die kleine Untersuchung hatte Pilotcharakter. Die Studiengruppe sollte daher auf leicht zugänglich ABM-Beschäftigte in und um Magdeburg begrenzt werden. In einer standardisierten schriftlichen Befragung sollte nach subjektiv empfundenen Auswirkungen der ABM-Tätigkeit auf die Gesundheit gefragt werden. Eine Kontrollgruppe war daher nicht vorgesehen. (Eine Kontrollgruppe hätte beispielsweise aus einer Gruppe von ehemaligen Arbeitslosen bestehen können, die in den ersten Arbeitsmarkt wieder integriert werden konnten.

Ob etwa hinsichtlich des Alters und anderer soziodemographischer Merkmale dabei hätte Strukturgleichheit hergestellt werden können, steht auf einem anderen Blatt.).

Fragestellungen und Hypothesen: Diese lassen sich im einzelnen auch dem Text entnehmen. Die zentrale Hypothese lautete, dass eine ABM-Tätigkeit trotz ihrer Befristung eine positive gesundheitliche Wirkung entfaltet. Weiterhin vermuteten wir, dass sich weitere positive Auswirkungen auf die psychosoziale Situation der ABM-Beschäftigten aufgrund der Wiederaufnahme eines geregelten Tagesablaufs und der Integration in eine neue Arbeitsgruppe zeigten. Wir gingen aber davon aus, dass die Auswirkungen einer ABM-Tätigkeit abhängig von deren konkreten Arbeitsbedingungen sind und dass weitere individuelle Faktoren die Auswirkungen als Confounder beeinflussen.

Zugang zur Zielgruppe: Der Zugang musste unter pragmatischen Gesichtspunkten erfolgen. Wir konnten eine große Magdeburger ABM-Gesellschaft (AQB) mit mehreren hundert ABM-Beschäftigten zur Unterstützung der Studie gewinnen. Dass diese Arbeitslosen sich strukturell von der Gesamtheit der Langzeitarbeitslosen unterschieden, kann möglich sein, konnte aber nicht überprüft werden. Zunächst war eine einmalige Vollerhebung unter allen Beschäftigten mit einem Fragebogen vorgesehen.

Erhebungsinstrument: Zur Erfassung der subjektiv empfundenen Gesundheit wurden Fragebogen geringfügig modifiziert, die bereits in Studien zur Gesundheit von Arbeitslosen im Raum Dresden eingesetzt wurden. Ergänzt durch die Studiengruppe wurden selbst entwickelte Fragen zur ABM-Tätigkeit und zu den Auswirkungen auf verschiedene Lebensbereiche. Nach ärztlich diagnostizierten Krankheiten wurde wegen fehlender Möglichkeit der Überprüfung verzichtet. Ein qualitativer Ansatz hätte den Rahmen eines Praktikums gesprengt und lag auch nicht in dessen Ziel.

Prätest: Ein Prätest sollte nach Möglichkeit immer durchgeführt werden. Sowohl um das Instrument bzgl. Validität und Reliabilität zu testen, als auch um Zugangswege und Akzeptanz zu prüfen. Aufgrund der knappen Zeit musste auf einen Prätest verzichtet werden.

Feldphase: Nach Absprache mit der Geschäftsführung der AQB und Information der Arbeitsgruppenleiter war eine einmalige Erhebung (für eine persönliche oder schriftliche Nachfassaktion fehlte die Zeit und Geld) durch Studenten während der Arbeitszeit an den Einsatzorten der ABM-Beschäftigten geplant. Die Teilnahme an der Befragung war natürlich freiwillig. Die Fragebögen sollten entweder an alle anwesenden Arbeiter verteilt, gleich ausgefüllt und wieder eingesammelt oder nach Ausfüllen von den Gruppenleitern nach einigen Tagen gesammelt an die Studienleiter geschickt werden. Es stellt sich bei den konkreten Absprachen heraus, dass wegen der bevorstehender Beendigung einiger Maßnahmen und wegen der Abwesenheit und schlechter Erreichbarkeit der Einsatzorte nicht alle Gruppen aufgesucht werden konnten. Außerdem konnten wir nur diejenigen Arbeiter befragen, die an dem Tag der Erhebung gerade arbeiten. Beschäftigte im Urlaub oder krank Gemeldete wurden nicht befragt und konnten auch nicht nacherfasst werden. Es kann vermutet, dass dies zu Verzerrungen im Sinne eines Healthy-Worker-Effekts geführt hat.

Studienbedingungen: Wenn ein Ausfüllen nicht direkt im Anschluss an die Aufklärung der ABM-Beschäftigten über Zielsetzung und Inhalte der Studie erfolgte, konnte kein Einfluss auf die Umgebungsbedingungen beim Ausfüllen des Fragebogens genommen werden. Insgesamt mussten wir den äußeren Rahmen der Studie so nehmen, wie er war. Das heißt konkret, dass die Mehrheit der Arbeitsgruppen im Freien tätig war und als Aufenthaltsraum häufig nur ein kleiner Bauwagen zur Verfügung stand. Ein individuelles und unbeeinflusstes Ausfüllen der Fragebögen war nicht möglich. Auch aus diesem Tatbestand kann eine Verzerrung der Ergebnisse resultieren, ohne dass diese in Richtung und Ausmaß bewertet werden könnte.

Dateneingabe: Abgesehen vom schwachen Rücklauf - nur die Hälfte der Beschäftigten laut Sollstärke der Arbeitsgruppen wurde erreicht, davon wiederum nahm etwa die Hälfte an der Befragung teil, so dass letztendlich rund ein Viertel der Beschäftigten der ABM-Gesellschaft befragt werden konnte – konnten bis auf 5 alle Fragebögen in die Analyse einbezogen werden.

Datenanalyse und weitere Schritte: Die Datenanalyse erfolgte durch den Studienleiter (E.S.), die Besprechung der Ergebnisse und ihre kritische Bewertung durch die gesamte Studiengruppe. Nach deskriptiven Auswertungen folgten zum Abschluss multivariate Analysen, um den Einfluss der ABM-Tätigkeit auf die Gesundheit von den potenziellen Confoundern zu trennen. Es zeigte sich nur tendenziell eine Bestätigung der Hypothese. Aufgrund zahlreicher Verzerrungen bedarf die initial formulierte Problemstellung weiterer Bearbeitung. Die Ergebnisse dieser Studie wurde der AQB mitgeteilt und auf wissenschaftlichen Kongressen vorgestellt. Wenige Jahre später fand im Rahmen einer Diplomarbeit an der Hochschule Magdeburg-Stendal eine qualitative Befragung von ABM-Beschäftigten statt, um einen Teil der offenen gebliebenen Fragen zu klären.

10. Eine Übung zur Studienanalyse

10.1 Fall-Studie mit Übungsaufgaben

An diesem ausgearbeiteten Übungsbeispiel⁶ soll ein idealtypisches epidemiologisches Vorgehen bei einer einfachen Fall-Kontroll-Studie dargestellt werden. Es umfasst die Skizzierung der Fragestellung, das Design der Studie, die Berechnung eines einfachen Effektmaßes und die Behandlung möglicher Confounder. Die einzelnen Arbeitsschritte sind als Fragen konzipiert, die anhand der Originalquelle auch im Eigenstudium bearbeitet werden können. Um die Übung zu verstehen und nachzuvollziehen, ist kein tiefgreifendes medizinisches Wissen erforderlich. Die kurz skizzierten Antworten auf die Fragen finden sich am Ende des Kapitels.

Einführung: Es gibt eine deutliche Beziehung zwischen der Einnahme exogener Östrogene und dem Auftreten von Endometrium(=Gebärmutter-schleimhaut-)karzinomen. Diese Assoziation ist biologisch plausibel, da Östrogene Wachstum (auch entarteter Zellen) stimulieren. Nach der Einführung oraler Kontrazeptiva (OK; ‚Pille‘) in den 60er Jahren wurde nach weiteren möglichen Beziehungen zwischen der Einnahme von Geschlechtshormonen und bösartigen Neubildungen der weiblichen Reproduktionsorganen gesucht.

Frage 1: Welche Indikationen für die Einnahme von Geschlechtshormonen sind bekannt? Zwischen welchen Präparatgruppen sollte man unterscheiden, wenn es um ein mögliches Karzinomrisiko geht?

Frage 2: Was wären Ihre ersten Schritte, wenn Sie seltene, aber gravierende unerwünschte Arzneimittelwirkungen suchen (UAW)? Diskutieren Sie die Vor- und Nachteile möglicher Zugangswege.

Frage 3: Aus welchen Gründen wird ein befürchtetes Krebsrisiko durch ein Arzneimittel nicht im Rahmen des Zulassungsverfahrens sicher ausgeschlossen?

Design: Aus dem Mitgliederverzeichnissen der Krankenschwesterverbände in 11 amerikanischen Bundesstaaten wurden alle Krankenschwestern herausgesucht, die 1976 zwischen 35 und 55 Jahre alt und verheiratet waren. Im gleichen Jahr erhielten sie mit der Post einen Fragebogen zur Gesundheit, in dem u.a. Angaben zur Art und zum Zeitpunkt einer eventuellen Krebsdiagnose erfragt und eine differenzierte Pillenanamnese erhoben wurde.

172 413 Krankenschwestern erhielten den Fragebogen, 71% (n=121964) schickten die Bögen ausgefüllt zurück. 98 Krankenschwestern teilten eine Ovarialkrebsdiagnose nach 1964 mit. Von 70 konnten – nach eingeholter Zustimmung und intensiver Nachsuche – die relevanten Krankenakten eingesehen werden. In 11 Fällen war eine andere Diagnose verzeichnet, in 47 Fällen handelte es sich um ein Ovarialkarzinom epithelialen Ursprungs und in 12 Fällen um Ovarialkarzinome anderen Gewebetyps. Die Auswertung beschränkte sich auf die 47 Patientinnen mit epithelialen Karzinomen.

Für jeden dieser Fälle wurden als Kontrolle 10 Frauen mit gleichem Geburtsjahr aus der Datenbasis gezogen, deren Ovarien zur Zeit der Krebsdiagnose des Falles intakt waren. Als ‚Pilleneinnahme‘ galt, wenn eine Ovarialkrebspatientin angab, irgendwann vor der Krebsdiagnose orale Kontrazeptive eingenommen zu haben.

Frage 4: Wie bewerten Sie die Grundgesamtheit der in dieser Studie befragten Personen? Welche Selektionsfaktoren können wirksam geworden sein? Stören Sie im Zusammenhang dieser Studie? Welche Vorteile kann eine Befragung von Krankenschwestern gegenüber einer Befragung von anderen Bevölkerungsgruppen haben?

⁶ ausgearbeitet durch Prof. Dr. med. Bernt-Peter Robra, M.P.H., Institut für Sozialmedizin und Gesundheitsökonomie der Otto-von-Guericke-Universität Magdeburg

Frage 5: Welche Pilleneinnahme(n) berücksichtigt man logischerweise bei den Kontrollpersonen?

Frage 6: Nach welchen Kriterien haben die Forscher vermutlich die o.g. Altersgruppe ausgewählt?

Roher und altersspezifischer Effektschätzer: Tabelle 10.1 fasst die Expositionsanamnese der Fälle und Kontrollen zusammen.

Tabelle 10.1: Einnahme oraler Kontrazeptive bei Fällen mit Ovarialkarzinomen und Kontrollen, geschichtet nach Alter

Alter bei Diagnose	Status	Pilleneinnahme			Odds Ratio (Konf.-intervall)
		gesamt	nein	ja	
unter 35	Fälle	11	9	2	0,2
	Kontrollen	109	55	54	(0,1 - 1,0)
35 bis 44	Fälle	19	13	6	1,0
	Kontrollen	185	128	57	(0,4 - 2,9)
45 und älter	Fälle	17	12	5	1,3
	Kontrollen	170	129	41	(0,4 - 3,9)
gesamt	Fälle	47	34	13	0,8
	Kontrollen	464	312	152	(0,4 - 1,5)

Quelle: Willett et al. (1981)

Frage 7: Was bedeutet das hier mitgeteilte Odds Ratio (als Schätzer für das relative Risiko)? Bitte rechnen Sie nach. Wie interpretieren Sie die mitgeteilten Konfidenzintervalle?

Frage 8: Warum teilen die Verfasser die Daten in Altersgruppen geschichtet mit? Wie beurteilen Sie das Muster der altersspezifischen ‚Effekte‘?

Frage 9: Wie könnten Sie die Expositionsanamnese ‚verfeinern‘?

Dosis-Wirkungs-Beziehung: Tabelle 10.2 zeigt die altersspezifischen Daten aufgeschlüsselt nach der Dauer der Pilleneinnahme.

Tabelle 10.2: Dauer der Einnahme oraler Kontrazeptive bei Fällen mit Ovarialkarzinomen und Kontrollen, geschichtet nach Alter bei Diagnose

Alter bei Diagnose	Status	nie	Dauer der Pilleneinnahme	
			unter 3 Jahre	3 Jahre und länger
unter 35	Fälle	9	2	0
	Kontrollen	55	42	12
35 bis 44	Fälle	13	2	4
	Kontrollen	127	27	30
45 und älter	Fälle	12	3	2
	Kontrollen	129	18	23
gesamt	Fälle	34	7	6
	Kontrollen	31	87	65
	rohes OR	1	0,7 (0,3 - 1,7)	0,8 (0,3 - 2,1)
	MH-OR*	1	0,7 (0,3 - 1,7)	0,8 (0,3 - 2,1)

* Mantel-Haenszel-Odds-Ratio; Quelle: Willett et al. (1981)

Frage 10: Wie kommen die Verfasser auf die beiden rohen Risikoschätzungen? Bitte rechnen Sie nach.

Frage 11: Kann es ein, dass die Dauer der Einnahme oraler Kontrazeptive vom Alter abhängt? Wenn ja, welches Problem müssen wir dann bewältigen? Wie können wir das machen?

Stratifizierung: Eine Störgröße kann man unter anderem durch Stratifizierung berücksichtigen. Hier kann man den gesuchten Zusammenhang zwischen der Einnahme oraler Kontrazeptiva und Ovarialkarzinomen innerhalb relativ schmaler Schichten gleichen Alters einzeln prüfen. Auf diese Weise stellt man sicher, dass Alterseinflüsse sich jedenfalls nicht mehr erheblich bemerkbar machen können.

Frage 12: In wie viele 4-Felder-Tafeln müssen Sie die Tabelle zerlegen, um die darin enthaltenen alters- und dauerspezifischen Odds ratios berechnen zu können?

Frage 13: Wie lauten die altersspezifischen Effektschätzer für die Einnahmekategorie unter 3 Jahre?

Frage 14: Wie würden sie die altersspezifischen Effektschätzer zusammenfassen?

Weitere Störgrößen: Die weitere Datenanalyse zeigt, dass Nulliparae (hier definiert als Frauen ohne eine Schwangerschaft von mindestens 6 Monaten Dauer) ein Odds Ratio von Ovarialkarziom von 2,2 (Konfidenzintervall 1,1 - 4,5) im Vergleich zu Frauen haben, die mindestens eine Schwangerschaft von 6 Monaten Dauer gehabt haben.

Frage 15: Es können ihnen mindestens zwei Gründe einfallen, wie dieser Zusammenhang hervorgerufen wird. Welche?

Frage 16: Wie auch immer, stört dieser Zusammenhang zwischen Krebsrisiko und Parität bei der Beurteilung des Pillenrisikos. Warum?

Frage 17: In welchem Detail müssten Sie die Daten bekommen, um den Faktor ‚Parität‘ zusätzlich zum Alter berücksichtigen? Bitte zeichnen Sie eine leere Tabelle mit entsprechenden Zeilen- und Spaltenüberschriften.

Es folgt die zugehörige Tabelle aus der Originalpublikation.

Tabelle 10.3: Einnahme oraler Kontrazeptive bei Fällen mit Ovarialkarzinomen und Kontrollen, geschichtet nach Alter und Parität

Alter bei Diagnose	Status	keine Schwangerschaft		Schwangerschaft		OR
		keine OK*	OK	keine OK	OK	
unter 35	Fälle	1	3	1	4	0,3
	Kontrollen	9	18	45	37	(0,1 - 1,3)
35 bis 44	Fälle	0	4	6	9	1,1
	Kontrollen	3	7	54	120	(0,4 - 3,2)
45 und älter	Fälle	0	1	5	11	1,3
	Kontrollen	3	16	38	113	(0,4 - 3,9)
gesamt	Fälle	1	10	12	24	0,8
	Kontrollen	15	41	137	270	(0,4 - 1,6)
rohes OR		0,3 (0,0 - 2,1)		1,0 (0,5 - 2,0)		

* OK: orale Kontrazeptiva; Quelle: Willett et al. (1981)

Frage 18: Was bedeuten hier die zusammengefassten Risiken nach Mantel-Haenszel am rechten Rand?

Frage 19: Wie beurteilen Sie nun die Beziehung zwischen Pilleneinnahme und Ovarialkrebs?

10.2 Antworten auf die Fragen

Frage 1: Indikationen für OK sind 1. die Schwangerschaftsverhütung und 2. die Behandlung von Menopausenbeschwerden bzw. die Hormonsubstitution. Ganz allgemein lassen sich die OK in Östrogen-, Gestagen- und Misch-Präparate unterscheiden. Östrogene sind die weiblichen Sexualhormone, die im Eierstock und der Nebenniere gebildet werden. Gestagene sind Hormone, die nach der Befruchtung und während der Schwangerschaft von der Plazenta und vom Gelbkörper gebildet werden.

Frage 2: Zu unterscheiden ist das Spontanberichtswesen über UAW's auf speziellen Meldebögen der Arzneimittelkommission der Deutschen Ärzteschaft und des Robert-Koch-Instituts sowie epidemiologische Fall-Kontroll- und prospektive Studien. Das Spontanberichtswesen wirkt nur bei gewisser ‚Grundaufmerksamkeit‘ der Ärzte, etwa aufgrund aktueller Publikationen (z.B. Lipobay im Jahr 2001), ist unsystematisch und eher zufällig. Prospektive Studien sind aufgrund der meist geringen Inzidenz der UAW's wenig erfolgversprechend. Nach erfolgter Markteinführung eines Medikaments sind Fall-Kontroll-Studien das am besten geeignete Instrument zur Aufdeckung und Bewertung der UAW's. Sie wurden beispielsweise eingesetzt zur Bewertung des Thromboserisikos bei der Einnahme der Minipille.

Frage 3: Das Zulassungsverfahren ist angesichts der meist längeren Zeit bis zur klinischen Manifestation eines Karzinoms zu kurz. Ebenso ist hier die geringe Inzidenz arzneimittelindizierter Karzinome zu beachten.

Frage 4: Aufgrund der geringen Inzidenz bedarf es einer großen Grundgesamtheit, um eine hinreichend große Anzahl von Fällen zu erhalten. Die Fall-Kontroll-Studie ist daher der geeignete Studientyp. Die Krankenschwestern unterscheiden sich vermutlich hinsichtlich der Einnahme von OK nicht von der Allgemeinbevölkerung. Das gleiche gilt für das Risiko von Ovarialkarzinomen. Die Beschränkung auf diese Berufsgruppe hat daher keine Auswirkungen auf die Verallgemeinerbarkeit der späteren Ergebnisse der Studie. Die Beschränkung auf Karzinome epithelialen (=nur aus Zellen bestehendes Gewebe) Ursprungs ergibt sich aus der zu untersuchenden karzinogenen Wirkung der OK und soll Verzerrungen durch andere Ursachen bei Karzinomen anderen Typs vermeiden. Fälle und Kontrollen sind aus einer Grundgesamtheit, daher sind Verzerrungen bei der Rekrutierung nicht zu vermuten und das Rekrutierungsverfahren vereinfacht sich. Krankenschwestern verfügen über ein erhebliches medizinisches Vorwissen, was die Validität der Angaben zu Karzinomerkrankungen erhöht. Auch ist die Motivation zur Teilnahme an der Studie möglicherweise überdurchschnittlich hoch.

Frage 5: Entscheidend ist die Exposition, d.h. die Einnahme OK, **vor** der Manifestation eines Ovarialkarzinoms. Das gleiche gilt für die Kontrolle. Es wird nur nach der Einnahme OK bis zum entsprechenden Alter bei der Diagnosestellung des zugehörigen Falles gefragt.

Frage 6: Die angeschriebenen Frauen waren bei der Markteinführung der OK zwischen 20 und 40 Jahre alt, also im reproduktionsfähigen Alter, aber jenseits der Pubertät. (Hinweis: Bis zum Abschluss der Pubertät gibt es erste Gipfel bei der Inzidenz und Mortalität des Ovarialkarzinoms, die aber genetische Ursachen haben).

Frage 7: Ein OR deutet auf einen protektiven Effekt der OK bzgl. des Ovarialkarzioms hin. Zum Nachrechnen: $OR = (13 \cdot 312) / (34 \cdot 152) = 0,78$. Allerdings ist der Effektschätzer nicht statistisch signifikant, da das Konfidenzintervall die Eins enthält.

Frage 8: Weil sowohl Inzidenz bzw. Mortalität als auch die Exposition stark altersabhängig ist. Die altersspezifischen OR deuten auf einen mit zunehmenden Alter abnehmenden protektiven Effekt der OK hin. Allerdings nehmen die Fallzahlen bei Altersdifferenzierung stark ab (Verbreiterung der Konfidenzintervalle!). Auch kann sich die Art der eingenommenen OK in den verschiedenen Altersgruppen verändert haben.

Frage 9: In dem man nach der Dauer der Einnahme (in Jahren) fragt.

Frage 10: Zum Nachrechnen: $(7 \cdot 311) / (34 \cdot 87) = 0,74$ (vorletzte Spalte) bzw. $(6 \cdot 311) / (34 \cdot 87) = 0,84$ (letzte Spalte)

Frage 11: Offensichtlich ja. Der Anteil der Frauen, die niemals OK eingenommen haben, steigt von 53% über 67% auf 75% in der Altersgruppe über 45 Jahre. Das Alter spielt in unserer Fragestellung die Rolle einer Störgröße, eines Confounders. Diesem Problem kann einerseits durch die Berechnung altersspezifischer Effektschätzer, andererseits durch die ‚Herausrechnung‘ des Alters und der Bestimmung eines altersbereinigten Effektmaßes begegnet werden.

Frage 12: In 6 Vierfeldertafeln bei 3 Altersklassen und 2 Kategorien bzgl. der Dauer der Einnahme (die ‚never-user‘ zählen jeweils als Referenzgruppe)

Frage 13: Zum Nachrechnen: $(2 \cdot 55) / (42 \cdot 9) = 0,29$ (unter 35 Jahre) bzw. $(2 \cdot 127) / (27 \cdot 13) = 0,72$ (35 bis 44 Jahre) bzw. $(3 \cdot 129) / (18 \cdot 12) = 1,79$ (45 Jahre und älter).

Frage 14: In der letzten Zeile der Tabelle 10.2 stehen die Mantel-Haenszel-Odds-ratios, die sich als gewichtete Effektmaße aus den altersspezifischen OR berechnen, wobei sich die Gewichtungsfaktoren aus der Besetzungszahl der einzelnen Altersklassen ergeben. In unserem Fall spielt das Alter keine Rolle, da die altersgewichteten OR gleich der rohen OR sind.

Frage 15: Frauen mit Kinderwunsch, aber ohne eingetretene Schwangerschaft, haben möglicherweise Probleme mit ihren Eierstöcken, die auch ein erhöhtes Krebsrisiko zur Folge haben könnten. Diese Frauen werden wegen ihres Kinderwunsches auch keine OK nehmen.

Frage 16: Ja, weil der Faktor Schwangerschaft sowohl mit der Einnahme von OK, als auch mit dem Karzinomrisiko assoziiert sein kann und damit wie das Alter als Confounder betrachtet werden muss.

Frage 17: Angaben zur Schwangerschaft müssten mindestens dichotom (jemals für mind. 6 Monate schwanger vs. niemals schwanger) vorliegen. Für jede der beiden Ausprägungen muss dann nach OK-Einnahme (ja/nein) differenziert werden. Die Tabelle muss dann den Aufbau von Tabelle 10.3 aufweisen.

Frage 18: Die OR am Rand sind um den Einfluss von Schwangerschaften bereinigte Effektschätzer. Sie unterscheiden sich nicht von denen aus Tabelle 1, also beeinflusst eine Schwangerschaft die bislang beobachtete Assoziation zwischen Ovarialkarzinomrisiko und OK-Einnahme nicht.

Frage 19: Auch nach Berücksichtigung möglicher Confounder (Alter, Schwangerschaften) scheint ein protektiver Effekt der Einnahme oraler Kontrazeptiva bzgl. des Ovarialkarzinoms zu bestehen. Die Hypothese der ‚incessant ovulation‘ besagt, dass durch das unterbrochene Wachstum des Epithels, sei es durch Schwangerschaft oder durch OK, auch das Tumorstadium verzögert wird.

11. Grundzüge der klinischen Epidemiologie

11.1 Grundlagen

Als klinische Epidemiologie bezeichnet man die Anwendung epidemiologischer Methoden auf klinische Fragestellungen oder klinisch definierte Bevölkerungsgruppen. Klinische Epidemiologie beschäftigt sich stark mit den Voraussetzungen und Folgen klinischer Entscheidungen. Zu ihren Voraussetzungen gehören daher Kenntnisse der Testtheorie. Ein einfaches Mittel zur Darstellung testtheoretischer Zusammenhänge ist die schon aus anderen Kapiteln bekannte Vierfeldertafel. Sie stellt sich damit als das zentrale Handwerkszeug des epidemiologen heraus.

Tabelle 11.1: Standardaufbau einer Vierfeldertafel der Testtheorie

		Krankheit		Summe
		vorhanden	nicht vorhanden	
diagnostischer Test	positiv	a	b	a + b
	negativ	c	d	c + d
Summe		a + c	b + d	a + b + c + d = n

Die Buchstaben stehen für absolute Häufigkeiten. Es werden also n Patienten untersucht oder einem diagnostischen Test unterzogen, von denen (a+b) testpositiv sind usw. Es lassen sich dann die folgenden Parameter kennzeichnen:

richtig Positive	a
richtig Negative	d
falsch Positive	b
falsch Negative	c
Prävalenz der Krankheit	$(a + c) / n$
Anteil der positiven Tests	$(a + b) / n$
Entdeckungsrate	a / n
Sensitivität	Wahrscheinlichkeit eines positiven Testbefundes bei erkrankten Personen $a / (a + c)$
Spezifität	Wahrscheinlichkeit eines negativen Testbefundes bei gesunden Personen $d / (b + d)$;
positiver Prädiktionswert (ppv)	Wahrscheinlichkeit, dass die Person bei positivem Testbefund erkrankt ist = $a / (a + b)$
negativer Prädiktionswert (npv)	Wahrscheinlichkeit, dass die Person bei negativem Testbefund nicht erkrankt ist = $d / (c + d)$

Sensitivität und Spezifität werden auch als Kennziffern der Testgüte bezeichnet. Man kann nicht generell sagen, dass die Sensitivität eines Tests wichtiger ist als seine Spezifität. Es kommt auf die diagnostische Fragestellung und die Prävalenz der Krankheit an sowie auf die relative Wertigkeit eines Fehlers in Richtung auf Überdiagnose oder Unterdiagnose. Ist die Sicherung der Diagnose z.B. nur mit großem Aufwand (invasiv) möglich, die Therapie eingreifend und/oder nur mäßig erfolgreich, wird man eine hohe Spezifität anstreben, um Nachteile der Abklärungsdiagnostik und/oder der überflüssigen Therapie bei Gesunden zu vermeiden. Ist die Diagnose dagegen leicht abzuklären und/oder die Therapie wenig eingreifend mit großer Aussicht auf Erfolg, wird man im Interesse einer hohen Sensitivität auch eine reduzierte Spezifität in Kauf nehmen. Eine streng formale Behandlung solcher Beziehungen lässt erkennen, dass je nach Kontext unterschiedliche Kombinationen von Sensitivität und Spezifität entschei-

dungsoptimal sein können. Damit sind auch Test-Grenzwerte vom Kontext abhängig, denn über die gesetzten Auffälligkeitsgrenzen bestimmt man Sensitivität und Spezifität eines Tests.

Das Krankheitsrisiko beträgt

bei positivem Test $a / (a + b)$ [das heißt = ppv] und
 bei negativem Test $c / (c + d)$ [das heißt = 1 - ppv]

$$\text{Relatives Risiko der Krankheit bei positivem Test} = \frac{a / (a + b)}{c / (c + d)} \quad (11:1)$$

Die Wahrscheinlichkeit für einen positiven Test beträgt

bei Krankheit $a / (a + c)$
 bei Nicht-Krankheit $b / (b + d)$

$$\text{Likelihood Ratio (LR)} = \frac{a / (a + c)}{b / (b + d)} = \frac{\text{Sensitivität}}{1 - \text{Spezifität}} \quad (11:2)$$

Die LR (genauer: die Likelihood Ratio eines positiven Tests, LR+) ist die Wahrscheinlichkeit, dass ein positiver Test bei Kranken zu erwarten ist, bezogen auf die Wahrscheinlichkeit, dass ein positiver Test bei Gesunden zu erwarten ist. In die LR gehen die Testgüteziffern Sensitivität und (das Komplement der) Spezifität ein. Die LR ändert sich, wenn man den Test verändert (testtheoretisch z.B. durch Wahl eines anderen Grenzwertes für die Kategorie "positiv", s.u.). Die LR beschreibt den diagnostischen Gewinn durch einen Testeinsatz. Dies ist (wie unten deutlich wird) die klinisch nützlichste Art, Sensitivität und Spezifität eines Tests in einem Parameter zusammenzufassen. Ein Wert von 1 bedeutet, dass der Test keine Informationen liefert, weil die Wahrscheinlichkeit für die Krankheit **nach** dem Test genauso groß wie **vor** dem Test.

Für die Interpretation eines Testergebnisses ist neben den beiden Parametern Sensitivität und Spezifität die Prävalenz der Zielerkrankung maßgeblich. Die Beziehung zwischen der Krankheitswahrscheinlichkeit vor einem Testeinsatz (d.h. der Prävalenz) und der Krankheitswahrscheinlichkeit nach einem Testeinsatz ist im **Satz von Bayes** zusammengefasst.

$$\text{ppv} = \frac{\text{Prävalenz} * \text{Sensitivität}}{(\text{Prävalenz} * \text{Sensitivität}) + (1 - \text{Prävalenz}) * (1 - \text{Sensitivität})} \quad (11:3)$$

Der Satz von Bayes ist vielleicht nicht anschaulich, aber er gibt die funktionelle Beziehung zwischen dem klinisch relevanten positiven Prädiktionswert einerseits und den Testgütekennziffern Sensitivität und Spezifität und der (a-priori-)Prävalenz der Krankheit an.

Typischerweise nimmt der ppv beim Übergang von sehr kleinen zu kleinen Prävalenzen sehr stark zu, während er beim Übergang von mittleren Prävalenzen zu hohen Prävalenzen nur noch relativ langsam wächst. Diese Beziehung kann durch Lösen der Gleichung für einen gegebenen Test (z.B. Se = 0,8; Sp = 0,9) bei zunehmender Prävalenz (z.B. 0,001; 0,01; 0,1; 0,3; 0,5; 0,7) leicht selbst geprüft werden (vgl. Abbildung 11.1). Der negative prädiktive Wert liegt dagegen bei niedrigen Prävalenzen in der Nähe von Eins und fällt dann bei mittleren Prävalenzen stark ab.

Der positive Prädiktionswert nach Testeinsatz (als differentialdiagnostisch relevanter Parameter) werden entsprechend umso größer sein

- je größer die Prävalenz
- je höher die Sensitivität und
- je höher die Spezifität (bzw. je kleiner das Komplement der Spezifität)

sind.

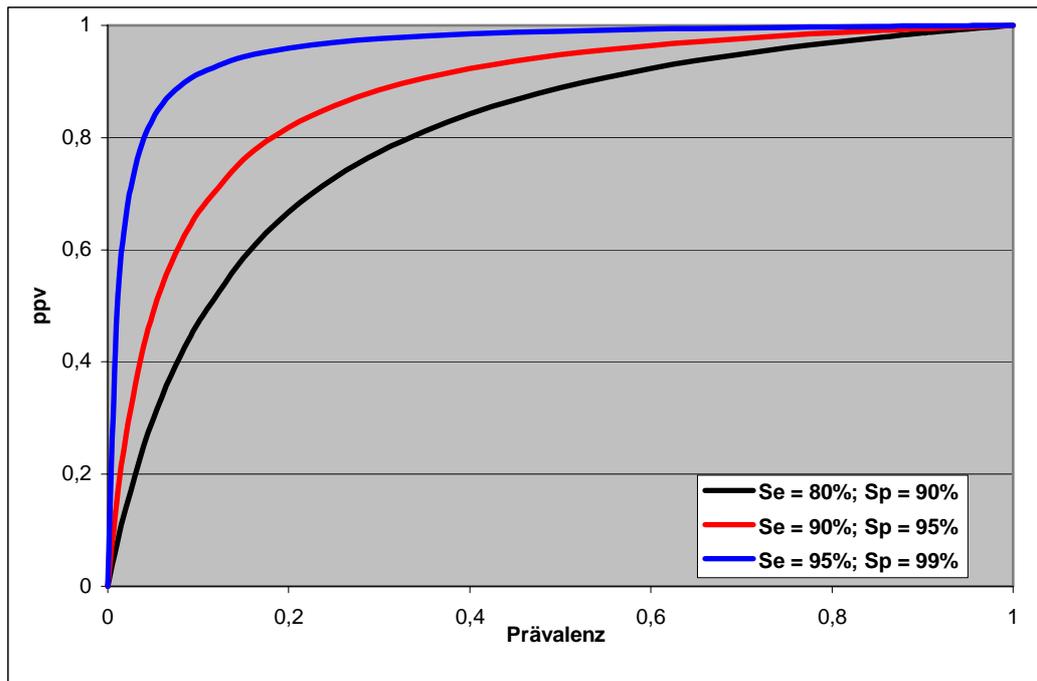


Abbildung 11.1: Abhängigkeit des positiven prädiktiven Werts eines diagnostischen Tests von der Prävalenz des Zielereignisses

Mit diesem Satz ist bei Einsetzen praktisch relevanter Werte auch erklärbar, warum in Screening-Situation in einer unausgelesenen Bevölkerung mit sehr niedriger Grundprävalenz sich viele, wenn nicht die Mehrheit der initial positiven Befunde bei weiterer Abklärung mit einem ‚Goldstandard‘ (d.h. einem nahezu perfekten Test) als negativ herausstellen (z.B. in der Krebsfrüherkennung).

Wer alles über die Testgüte weiß (Sensitivität, Spezifität, LR), aber die Rahmenbedingungen des Testeinsatzes nicht kennt (Prävalenz der Krankheit, z.B. Spezialambulanz gegen Allgemeinpraxis) weiß nicht, wie sehr er sich auf seinen Test verlassen kann (ppv). Empfehlungen über die weitere Abklärung von testpositiven Probanden können nur für einen bestimmten Prävalenzbereich gültig sein. Die Abklärungsstrategie eines positiven Tests in einer Patientengruppe mit hoher Prävalenz (z.B. Spezialambulanz, Unklinik) muss keineswegs identisch sein mit der Abklärungsstrategie desselben positiven Tests in einer Patientengruppe mit niedriger Prävalenz (z.B. in einer arbeitsmedizinischen Vorsorgeuntersuchung), weil der positive Prädiktionswert in beiden Situationen trotz gleicher Testgüte sehr unterschiedlich sein muss - und damit die Risiko-Nutzen-Bilanz des weiteren Vorgehens. Für ein Früherkennungsverfahren, also bei geringer Prävalenz des Zielereignisses, wird man beispielsweise eine hohe Spezifität fordern, um nicht zu viele falsch-positive Ergebnisse zu produzieren. Testpositive Fälle bedürfen dann einer weiteren Abklärung, während der npv nahe 1 liegt.

Bei vielen Tests, vor allem bei Laborwerten, handelt es sich nicht um binäre (ja/nein) Testergebnisse, sondern um stetig verteilte Messwerte. Lediglich die Verteilung der Messwerte bei Gesunden und Kranken unterscheiden sich. Die Verteilungen eines physiologischen Parameters werden sich in der Regel bei Gesunden und Kranken deutlich unterscheiden, partiell aber überlappen. So wird zwar von Hypertonie oder Hypercholesterinämie üblicherweise ab 160 oder 200 gesprochen, doch könnten diese Grenzwerte genauso gut bei 170 und 210 oder 150 und 190 liegen. Die Verschiebung des Grenzwerts verändert in solchen Situationen die Testparameter Sensitivität und Spezifität. Durch zunehmende Grenzwertsetzungen kann man beinahe jede beliebige Sensitivität mit der jeweils dazugehörigen Spezifität ‚einstellen‘. Werden

die vom Grenzwert abhängigen Testgüteziffern Sensitivität und Unspezifität (=1-Spezifität) gegeneinander aufgetragen, erhält man die sog. **ROC-Kurve**.

Im folgenden Beispiel (Abbildung 11.2) ist die Krankheit ein Harnblasenkarzinom. Als Test wird die Zahl sog. T-Zellen festgestellt, die bei Krebspatienten typischerweise niedriger ist als bei Gesunden. Allerdings überlappen sich die Verteilungen der Zellzahlen bei Gesunden und Krebspatienten. Setzt man den Grenzwert der Zellzahl schrittweise hoch (d.h. Patienten mit Werten unterhalb des Grenzwerts gelten als jeweils auffällig), nimmt die Sensitivität zu, die Spezifität ab. Bei sehr niedrigem Grenzwert bleiben viele Patienten noch unentdeckt (Sensitivität ist niedrig), dafür ist die Spezifität sehr groß. Mit zunehmendem Grenzwert steigt die Kurve, d.h. die Sensitivität steigt auf Kosten der Spezifität. Eine gleichzeitige Optimierung beider Gütekriterien ist bei stetigen Variablen demnach **nicht** möglich.

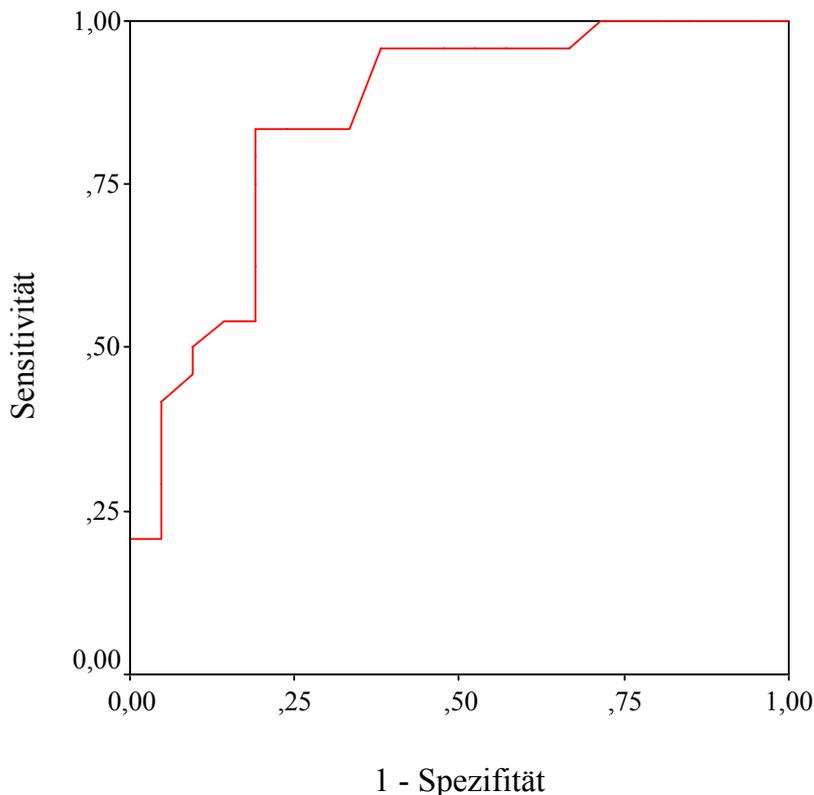


Abbildung 11.2: Beispiel für eine ROC-Kurve: Diagnose des Harnblasenkarzinoms über T-Zellbestimmung (Quelle: Bühl und Zöfel 2002)

Ein perfekter Test ordnet immer alle Patienten und alle Gesunden fehlerfrei zu. Die ROC-Kurve wäre in einem solchen Fall dann in die linke obere Ecke verlagert. Dies tritt in der Realität aber nicht ein. Die Fläche unter der Kurve ist dann ein Maß für den Informationsgewinn durch den Test auf dem ganzen Wertebereich der zugrunde liegenden Messgröße. Ihre Größe liegt zwischen 0 und 1, ein ‚Test‘ durch Raten besäße das Flächenmaß 0,5. Um einen optimalen Grenzwert für ein bestimmtes diagnostisches Entscheidungsproblem festzulegen, muss man falsch-positive und falsch-negative Fehler gegeneinander gewichten. Bei gleicher Bewertung falsch positiver und falsch negativer Befunde wird der optimale Grenzwert dadurch bestimmt, dass der Abstand der ROC-Kurve von der linken oberen Ecke minimiert wird. In der medizinischen Praxis ist diese Situation allerdings nicht typisch.

Diese testtheoretischen Konzepte der klinischen Epidemiologie lassen sich auf viele andere Fragestellungen der Gesundheitswissenschaften ausdehnen, in denen Tests zur Anwendung

gelangen wie z.B. Intelligenz- oder Entwicklungstests oder Instrumente zur Erhebung des Hilfebedarfs.

11.2 Ein Beispiel

(adaptiert nach Bhopal R (2002): Concepts of Epidemiology, New York: Oxford University)

In dieser kleinen Übung werden zentrale Begriffe der klinischen Epidemiologie am Beispiel eines hypothetischen Screening-Tests erläutert. Screeningverfahren sind dabei als Kontrapunkt zu diagnostischen Tests in einer klinischen Situation mit bereits beschwerdebehafteten Patienten zu sehen.

Screening ist allgemein definiert als 'the presumptive identification of unrecognized disease or defect by the application of tests, examinations or other procedures which can be applied rapidly. Screening tests sort out apparently well persons who probably have a disease from those who probably do not' (zitiert nach Last, A dictionary of epidemiology).

Screeningverfahren haben ganz allgemein **folgende Ziele** (Bhopal 2002):

- Bessere Versorgung und Prognose bei Patienten in frühem Krankheitsstadium
- Schutz der Bevölkerung vor ansteckenden Krankheiten
- Identifikation gesunder Individuen
- Rationale Zuweisung (begrenzter) Ressourcen
- Forschung zum natürlichen Verlauf von Krankheiten (durch Entdeckung und Beobachtung von Frühstadien)

Nicht alle Krankheiten eignen sich für ein Screening, sondern es sollten bestimmte Bedingungen erfüllt sein. Die bekanntesten Kriterien stammen von Wilson und Junger (1968):

1. Die Krankheit sollte ein wichtiges Gesundheitsproblem darstellen.
2. Es sollte eine anerkannte Behandlungsmöglichkeit für diese Krankheit geben.
3. Es sollten ausreichend Kapazitäten für Diagnostik und Behandlung dieser Krankheit vorhanden sein.
4. Die Krankheit sollte eine identifizierbare latente oder frühsymptomatische Phase besitzen.
5. Es sollte einen geeigneten Test zur Früherkennung dieser Phase geben.
6. Dieser Test sollte in der Zielpopulation auf Akzeptanz stoßen.
7. Der natürliche Verlauf der Krankheit, incl. der Latenzphasen, sollte ausreichend verstanden sein.
8. Es sollte Einvernehmen herrschen, welche Personen zu untersuchen und zu behandeln sind.
9. Die Kosten der Fallfindung (incl. der weiterführenden Diagnostik und Therapie) sollten in einem angemessenen Verhältnis zu den Ausgaben der gesamten Gesundheitsversorgung stehen.
10. Das Screeningverfahren sollte einen kontinuierlichen Prozess und nicht ein begrenztes Projekt darstellen.

Diese zehn Kriterien lassen sich auch in sechs Fragen an ein in Frage kommendes Screeningverfahren gießen:

1. Gibt es eine effektive Intervention im Falle einer Früherkennung?
2. Führt diese Intervention zu einem besseren Versorgungsergebnis gegenüber einer Behandlung bei Symptomauffälligkeit?

3. Gibt es einen effektiven Screeningtest, der die Krankheit früher als sonst üblich entdeckt?
4. Ist dieser Test bei der Zielpopulation akzeptiert und ist er für diese ausreichend verfügbar?
5. Verlangt die Zielkrankheit besondere Priorität?
6. Übertreffen die zu erwartenden Vorteile des Screenings deren Kosten?

Aufgabe 1: Überlegen Sie sich, inwieweit diese Kriterien für das Neugeborenen-Screening auf Phenylketonurie und das Mammographie-Screening erfüllt sind.

Ein Screening-Programm benötigt aufgrund seines meist hohen Aufwands, seiner oftmals hohen fachlichen und/oder technischen Anforderungen sowie einer längeren Latenzzeit, bis sich Auswirkungen der Früherkennung bemerkbar machen, einer systematischen Evaluation. Hinzu kommt der Tatbestand, dass die Auswirkungen eines Screening-Programms ganz entscheidend von dessen Akzeptanz in der Zielpopulation ab. Die Effektivität eines Screening-Programms lässt sich anhand von Prozess- und Ergebnisindikatoren beurteilen (Tabelle 11.2).

Tabelle 11.2: Mögliche Prozess- und Ergebnisindikatoren eines Screening-Programms

Prozessindikatoren	Ergebnisindikatoren
Anzahl der gescreenten Personen	Senkung der Mortalität in der Screening-Population
Anteil der gescreenten Zielgruppe und Häufigkeit der Screening-Untersuchungen	Senkung der Letalitätsrate bei gescreenten Personen
entdeckte Prävalenz der präklinischen Erkrankungen	proz. Zunahme der Fälle in einem frühen Stadium
Gesamtkosten des Programms	Senkung der Komplikationen
Kosten pro Fall	Vermeidung/Minderung der Rezidivrate
Anteil der Personen mit positivem Screening-Ergebnis, der einer abschließenden Diagnostik zugeführt wurde	Verbesserung der Lebensqualität von gescreenten Personen
Kosten pro neu diagnostiziertem Fall	
prädiktiver Wert eines positiven Tests in der gescreenten Bevölkerung	

Gordis (2002)

Bei der Bewertung eines Screening-Programms anhand dieser Indikatoren sind allerdings verschiedene methodische Probleme zu beachten. Diese können primär herrühren von:

Selektionsbias: Screening-Teilnehmer weisen häufig andere (meist günstigere) Risikomerkmale und gesundheitsrelevante Verhaltensweisen als Nichtteilnehmer auf.

Lead-Time-Bias: Durch eine frühzeitige Diagnose der Zielerkrankung wird die Prognose nicht notwendigerweise verbessert (vgl. Kriterien 2 und 3 von Wilson und Jungner). Dennoch verlängert sich scheinbar die Überlebenszeit, selbst wenn der Zeitpunkt des Versterbens unverändert bleibt.

Length-sampling-bias: In einem Screening mit festgelegten Intervallen werden überproportional viele langsam sich entwickelnde Krankheiten mit verhältnismäßig guter Prognose entdeckt. Fälle mit schneller Progredienz werden dagegen häufig zu sog. **Intervallfällen**, also Fällen, die zwischen zwei Screening-Untersuchungen klinisch manifest werden. Die Beachtung allein der im Screening entdeckten Fälle würde das Ergebnis daher positiv verzerren.

Die Ergebnisse eines Screeningverfahrens (und allgemein jedes diagnostischen Verfahrens) können in Form der bekannten Vierfeldertafel dargestellt werden (Tabelle 11.3):

Tabelle 11.3: Darstellung der Ergebnisse eines Screening-Tests in Abhängigkeit vom Krankheitsstatus

Screening-Test	Krankheit		gesamt
	vorhanden	nicht vorhanden	
positiv	a	b	a + b
negativ	c	d	c + d
gesamt	a + c	b + d	a + b + c + d

Frage 2: Durch welche Gütekriterien lässt sich ein diagnostischer Test beschreiben?

Beispiel: 500 Patienten mit einer bekannten Krankheit werden einem neuen diagnostischen Test unterzogen. Ebenso werden 500 Personen ohne diese Krankheit untersucht. Von den 500 Kranken haben 473 ein positives Testergebnis, von den Gesunden 7 Personen.

Frage 3a: Bringen sie diese Informationen in eine Vierfeldertafel.

Frage 3b: Berechnen Sie Sensitivität und Spezifität.

Frage 3c: Welches sind die Konsequenzen für diejenigen, die falsch klassifiziert wurden?

Die Gütekriterien sagen aber nichts über die Vorhersagekraft eines Testergebnisses aus, da im Screening oder einer klinischen Situation erst vom Ergebnis eines diagnostischen Tests auf das Vorliegen einer Erkrankung geschlossen werden soll.

Frage 4a: Welche Parameter geben Auskunft über die Aussagefähigkeit eines diagnostischen Tests?

Frage 4b: Berechnen Sie diese Parameter für unser obiges Beispiel.

Nun hängen ppv und npv nicht allein von Sensitivität und Spezifität eines Tests ab, sondern auch von der Prävalenz der zu identifizierenden Krankheit in der Zielpopulation. Stellen Sie sich dazu vor, der Test aus unserem ersten Beispiel mit den dort berechneten Werten für Sensitivität und Spezifität sei a) bei den Patienten einer allgemeinärztlichen Praxis mit einer Prävalenz von 10% und b) als Screeningverfahren in einer unausgelesenen Population mit einer Prävalenz von 1% angewendet worden. Die Zielpopulation habe beide Male eine Größe von 1.000 Personen.

Aufgabe 5: Erstellen Sie zwei Vierfeldertafeln für die beiden Szenarien und vergleichen Sie die prädiktiven Werte.

Im Gegensatz zu Sensitivität und Spezifität sind die prädiktiven Werte eines diagnostischen Verfahrens also von der Prävalenz der Zielkrankheit und damit von den Rahmenbedingungen ihrer Anwendung abhängig. In einer Spezialambulanz eines Krankenhauses müssen die Ergebnisse eines Tests anders bewertet werden als in einer allgemeinärztlichen Praxis oder in einer Screening-Situation. Diesen Tatbestand verdeutlicht Abbildung 11.3.

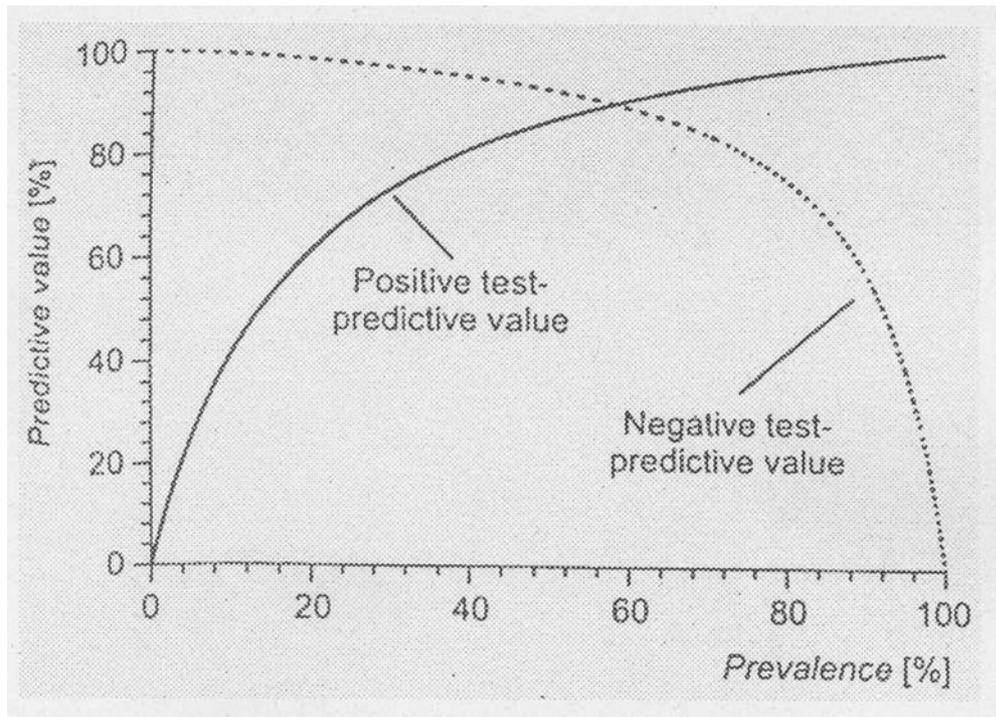


Abbildung 11.3: Abhängigkeit des positiven und negativen Prädiktionswerts eines diagnostischen Tests von der Prävalenz der Zielerkrankung (bei einer Sensitivität von 93% und einer Spezifität von 85%)

Aufgabe 6: Können Sie Regeln für die Gültigkeit positiver oder negativer Testergebnisse für die verschiedenen Situationen ableiten?

Die bisherigen Überlegungen zu diagnostischen Verfahren gelten für Tests mit nur zwei möglichen Ergebnissen 'positiv' oder 'negativ'. Häufig liegen klinische Informationen aber auf einem Kontinuum möglicher Werte vor (z.B. Blutdruck, Blutzuckerspiegel). Dann muss anhand eines **Schwellenwertes** eine Klassifizierung in testpositive und testnegative Personen vorgenommen werden. Eine eindeutige Trennung in Kranke und Gesunde würde man nur erreichen, wenn sich die Verteilungen des klinischen Parameters bei gesunden und kranken Personen nicht überschneiden. Dies ist in der Regel nicht der Fall und sei es aufgrund von Messfehlern oder intraindividuellen Schwankungen. So führt jede Festlegung eines Schwellenwertes dazu, dass ein gewisser Anteil der Gesunden fälschlicherweise als krank eingestuft wird, ebenso wie ein Teil der Kranken als falsch-negativ klassifiziert wird.

Aufgabe 7: Stellen Sie sich vor, Sie verwenden den Blutzuckerspiegel (mg/dl) zwei Stunden nach Nahrungsaufnahme als diagnostischen Test auf Diabetes und nehmen den Wert 140 als Grenzwert für ein positives Ergebnis. a) Wie schätzen Sie die Sensitivität und Spezifität eines Tests mit diesem Schwellenwert ein? b) Wie verändern sich tendenziell Sensitivität und Spezifität, wenn Sie den Schwellenwert auf 180 erhöhen bzw. auf 100 senken?

Die in diesem Beispiel aufgezeigten Zusammenhänge zwischen Sensitivität und Spezifität gelten für alle diagnostischen Verfahren, die auf der Messung stetiger klinischer Parameter beruhen. Es können nicht gleichzeitig beide Testparameter optimiert werden, sondern die Verbesserung des einen geht immer zu Lasten des anderen. Eine anschauliche Möglichkeit zur Darstellung der Abhängigkeit von Sensitivität und Spezifität von der Wahl des Schwellenwertes bietet die sog. **ROC-Kurve**, bei der Sensitivität und (1-Spezifität) für die verschiedenen möglichen Schwellenwerte gegeneinander abgetragen werden (Abbildung 11.4). Die Kurve gibt damit den Informationsgewinn des diagnostischen Tests in Form der **Likelihood Ratio** ($LR) = Se / (1 - Sp)$ wider. Je mehr die Kurve in die linke obere Ecke gewölbt ist, desto besser diskriminiert er zwischen Gesunden und Kranken. Bei gleicher Gewichtung falsch-

positiver und falsch-negativer Befunde ergibt sich als bester Schwellenwert derjenige Punkt auf der Kurve, der der linken oberen Ecke am nächsten liegt. Die eingezeichnete Gerade charakterisiert einen 'Test' ohne Informationsgewinn, etwas das Werfen einer Münze.

Frage 8: Welcher Schwellenwert eignet sich laut Abbildung 11.4 am besten für die Diagnose von Diabetes? Würde Sie für eine Screening-Situation einen anderen (größeren oder kleineren) Schwellenwert wählen?

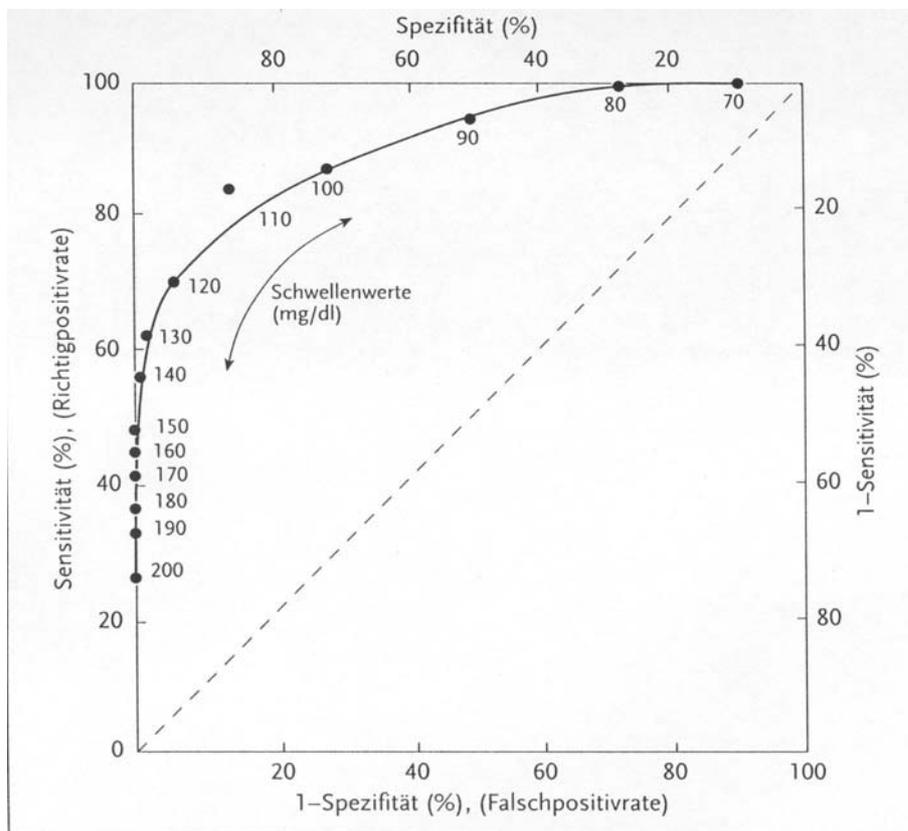


Abb. 3.4: Eine ROC-Kurve. Die Genauigkeit des Blutzuckerwertes 2 h postprandial als diagnostischer Test auf Diabetes mellitus. (Daten aus: Public Health Service. Diabetes program guide. Publication no. 506. Washington, DC: U.S. Government Printing Office, 1960.)

Abbildung 11.4: Abhängigkeit der Sensitivität und der Spezifität vom Grenzwert des diagnostischen Tests auf Diabetes (entnommen aus Fletscher et al. 2007)

Antworten:

Frage 1: Für das Neugeborenen-Screening können alle Kriterien als erfüllt betrachtet werden, da die Zielkrankheit zwar selten ist, aber schwerwiegende Konsequenzen für die Betroffenen hat, der Test einfach und billig ist, auf hohe Resonanz stößt und die Rahmenbedingungen günstig sind.

Beim Mammographie-Screening hingegen sind die Kriterien 3 (*Gibt es genügend Einrichtungen und Ärzte mit ausreichender Erfahrung?*), 6 (*Ist die Akzeptanz bei den Frauen wegen der Art der Untersuchung und der Strahlenbelastung gegeben?*) und 9 (*Sind die hohen Kosten für ein flächendeckendes Screening angesichts der knappen Kassen der GKV vertretbar?*) kritisch zu hinterfragen.

Frage 2: **Sensitivität** $a / (a + c)$
Spezifität: $d / (b + d)$

Antwort 3a: Siehe nachfolgende Tabelle

Screening-Test	Krankheit		Gesamt
	Vorhanden	Nicht vorhanden	
Positiv	473	7	480
Negativ	27	493	520
Gesamt	500	500	1.000

Frage 3b: Sensitivität: $473 / 500 = 94,6\%$;
Spezifität: $493 / 500 = 98,6\%$;

Frage 3c: **Falsch-Testpositive** müssen zeitweise mit der Befürchtung einer unter Umständen bedrohlichen Krankheit leben. Sie werden (u.U. kostenaufwändig) diagnostisch weiter abgeklärt; **falsch-Testnegative** haben mit den Konsequenzen einer unerkannten Krankheit zu leben (verschlechterte Prognose, falsche Sicherheit).

Frage 4a: **positiver prädiktiver Wert (ppv)** $= a / (a + b)$
negativer prädiktiver Wert (npv) $= d / (c + d)$

Frage 4b: $ppv = 473 / 480 = 98,5\%$;
 $npv = 493 / 520 = 94,8\%$

Frage 5a: Prävalenz 10%

Screening-Test	Krankheit		Gesamt
	Vorhanden	Nicht vorhanden	
Positiv	94,6	12,6	107,2
Negativ	5,4	887,4	892,8
Gesamt	100	900	1000

$ppv = 94,6 / 107,2 = 88,2\%$; $npv = 887,4 / 892,8 = 99,4\%$

Frage 5b: Prävalenz 1%

Screening-Test	Krankheit		Gesamt
	Vorhanden	Nicht vorhanden	
Positiv	9,46	13,86	23,32
Negativ	0,54	976,14	976,88
Gesamt	10	990	1000

$ppv = 9,46 / 23,32 = 40,6\%$; $npv = 976,14 / 976,68 = 99,9\%$

Frage 6: Zur Vorhersagekraft diagnostischer Tests gibt es folgende Regeln:

12. □ Bei **niedriger Prävalenz** (z.B. Screening) kann bei einem Tests mit hoher **Sensitivität** ein **negatives** Testergebnis zu **Ausschluss (out)** der Krankheit führen => **SNNout**.

b) Bei **hoher Prävalenz** (in einem durch vorhergehende medizinische Maßnahmen selektierten Patientenklientel) kann bei einem Tests mit hoher **Spezifität** ein **positives** Testergebnis zu **Bestätigung** der Krankheit (**in**) führen => **SPPin**.

c) Dagegen bedürfen testpositive Personen im Screening und testnegative Patienten in einem selektierten Patientengut weiterer Abklärung.

Aufgabe 7: Einem Lehrbuch zur klinischen Epidemiologie (Fletscher et al. 1999) können folgende Werte entnommen werden:

Blutzucker-Serumkonzentration (mg/dl)	Sensitivität	Spezifität
70	98,6	8,8
80	97,1	25,5
90	94,3	47,6
100	88,6	69,8
110	85,7	84,1
120	71,4	92,5
140	57,1	99,4
130	64,3	96,9
160	47,1	99,8
180	38,6	100,0
200	27,1	100,0

Danach beträgt die Sensitivität des Tests bei einem Grenzwert von 140 mg/dl 57%, die Spezifität 99,4%. Bei einer **Erhöhung** des Grenzwertwerts **sinkt die Sensitivität** und **steigt die Spezifität**, bei einer Absenkung des Grenzwertes dagegen **steigt die Sensitivität** und **sinkt die Spezifität**.

Frage 8: Bei gleicher Gewichtung von Sensitivität und Spezifität liegt der beste Schwellenwert etwa bei 100 mg/dl. Im Screening bei niedriger Prävalenz der Zielerkrankung, aber der Gefahr von Komplikationen bei übersehenem Diabetes könnte der Schwellenwert abgesenkt werden, um bei negativem Ergebnis Diabetes relativ sicher ausschließen zu können, da gleichzeitig eine weitere Abklärung der initial Testpositiven relativ einfach ist.

12. Assoziation oder Ursache – Das epidemiologische Kausalitätsprinzip

12.1 Die epidemiologische Bewertung von Assoziationen

In diesem Kapitel soll diskutiert werden, welche Evidenz, also belegbare Hinweise und Erkenntnisse, sich für einen kausalen Zusammenhang zwischen einer potenziellen Ursache und einer Krankheit aus epidemiologischen Studien ziehen lässt. Die Schwierigkeit dabei liegt in der Abgrenzung einer Ursache-Wirkungs-Beziehung von einer rein statistisch auffälligen Assoziation zweier Variablen.

Allgemein lassen sich verschiedene **Typen von Evidenz** unterscheiden (Raspe 1996):

- Klassiker, Autoritäten, klinische Schule, Lehrer
- durchschnittliches Verhalten, überkommene Routinen
- Deduktionen aus patho-physiologischen Theorien
- erinnerte Erfahrung
- Kasuistik, kontrollierte Erfahrung
- professioneller Konsensus, Konsensuskonferenzen
- kontrollierte Studien
- kontrollierte Anwendungsbeobachtung
- systematische Übersicht, Metaanalysen, Entscheidungsanalyse

Zu dieser Auflistung kommen nun die verschiedenen epidemiologischen Methoden und Studientypen hinzu, die im folgenden auf ihre Fähigkeit zur Identifizierung einer Ursache-Wirkungs-Beziehung untersucht werden sollen.

Als **Ursache** einer Krankheit wird ein Ereignis, eine Bedingung, eine Eigenschaft oder eine Kombination dieser Faktoren verstanden, die eine bedeutsame Rolle in der Entstehung dieser Krankheit spielt. Logischerweise muss dieser Grund der Krankheit **zeitlich vorausgehen**.

Eine Ursache ist **hinreichend** für die Krankheit, wenn sie zwangsläufig die Krankheit hervorruft oder auslöst. Sie ist **notwendig**, wenn sich die Krankheit nicht ohne sie entwickeln kann.

hinreichend:	Ursache	->	Krankheit
notwendig:	Krankheit	->	Ursache

A) Ist ein Faktor hinreichend und notwendig, bedeutet dies ein 1:1-Verhältnis zwischen Ursache und Krankheit, was in der Realität äußerst selten auftritt.

B) Bei notwendigen, aber nicht hinreichenden, Faktoren, müssen weitere Bedingungen erfüllt sein, dass die Krankheit entsteht. So löst das Tuberkelbakterium nicht in jedem Infizierten zwangsläufig Tuberkulose aus.

C) Hinreichende, aber nicht notwendige Faktoren, können zwar die Krankheit auslösen, doch sind dazu auch andere Faktoren allein in der Lage. Das trifft z.B. auf Rauchen und Lungenkrebs oder radioaktive Strahlung und Leukämie zu (vereinfacht, denn natürlich erkranken nicht alle Raucher an Lungenkrebs).

D) Bei vielen Krankheiten, vor allem chronischer Natur, wirken Faktoren, die weder hinreichend noch notwendig sind, in einem komplexen Modell der Krankheitsentstehung.

Häufig wird in der Epidemiologie von der Krankheit ausgegangen und nach ihren Ursachen gesucht (seltener wird umgekehrt verfahren: ausgehend von einem definierten Risiko wird nach dessen Folgen gesucht.). Die Epidemiologie beschäftigt sich mit **komplexen Ursachen-**

Wirkungs-Beziehungen, unterscheidet sich insofern von den Ansätzen der Naturwissenschaften. Denn **Robert Kochs Regeln** zur Verursachung einer Krankheit durch lebende Organismen:

- der Organismus muss immer zusammen mit der Krankheit auftreten,
- der Organismus kann isoliert werden und wächst in Reinkultur,
- der Organismus muss die Krankheit auslösen, wenn er in ein sukzessibles Tier verpflanzt werden,
- der Organismus muss im Tier identifiziert werden können,

sind für die meisten Krankheiten (Massenphänomene) der bevölkerungsbezogenen Epidemiologie offensichtlich nicht angemessen.

Vier **Typen von Krankheitsfaktoren** lassen sich unterscheiden, die notwendig sein können, aber nur selten hinreichend sind:

- **prädisponierende Faktoren** (Alter, Geschlecht, vorhergehende Krankheiten) können zur Anfälligkeit gegenüber Krankheitserregern führen;
- **ermöglichende Faktoren** (geringes Einkommen, ungenügende medizinische Versorgung, schlechte Wohnverhältnisse) können die Entwicklung von Krankheiten fördern;
- **beschleunigende Faktoren** (Exposition gegenüber einer spezifischen Noxe) können mit dem Beginn einer Krankheit assoziiert sein;
- **verstärkende Faktoren** (z.B. anhaltende Exposition) können eine bestehende Krankheit verstärken.

Risikofaktoren sind Faktoren, die positiv mit dem Risiko der Krankheitsentstehung korreliert, aber nicht hinreichend für das Auftreten der Krankheit sind. Risikofaktoren können mit mehreren Krankheiten assoziiert sein (z.B. Rauchen), gleichermaßen kann eine Krankheit mehrere Risikofaktoren besitzen (z.B. koronare Herzkrankheiten).

Mit Hilfe epidemiologischer Methoden kann die **relative Bedeutung** einzelner Faktoren bei der Krankheitsentstehung berechnet werden, ebenso wie das Potenzial der Krankheitsverhütung bei Ausschaltung eines einzelnen Faktors (vgl. Kapitel 8).

Synergismus liegt vor, wenn sich zwei Faktoren in ihrer Wirkung auf die Krankheitsentwicklung verstärken oder abschwächen (z.B. das extrem erhöhte Lungenkrebsrisiko bei Rauchen und gleichzeitiger Asbestexposition; Tabelle 12.1).

Tabelle 12.1: Altersstandardisierte Mortalitätsraten an Lungenkrebs (Fälle pro 100.000 Einw.) in Abhängigkeit vom Raucherstatus und Asbestexposition

Asbestexposition	jemals geraucht	Lungen-Ca-Todesfälle	Faktor (rel. Risiko)
nein	nein	11	1
ja	nein	58	5,3
nein	ja	123	11,2
ja	ja	602	54,7

Quelle: Beaglehole et al. 1997

Bevor eine beobachtete Assoziation zwischen einer Krankheit und einer möglichen Ursache auf die Ursächlichkeit dieses Faktors untersucht wird, müssen erst mögliche Fehlerquellen (Selektion, Confounding, Zufall) überprüft werden. Dazu ist in den vorangegangenen Kapitel schon einiges gesagt worden.

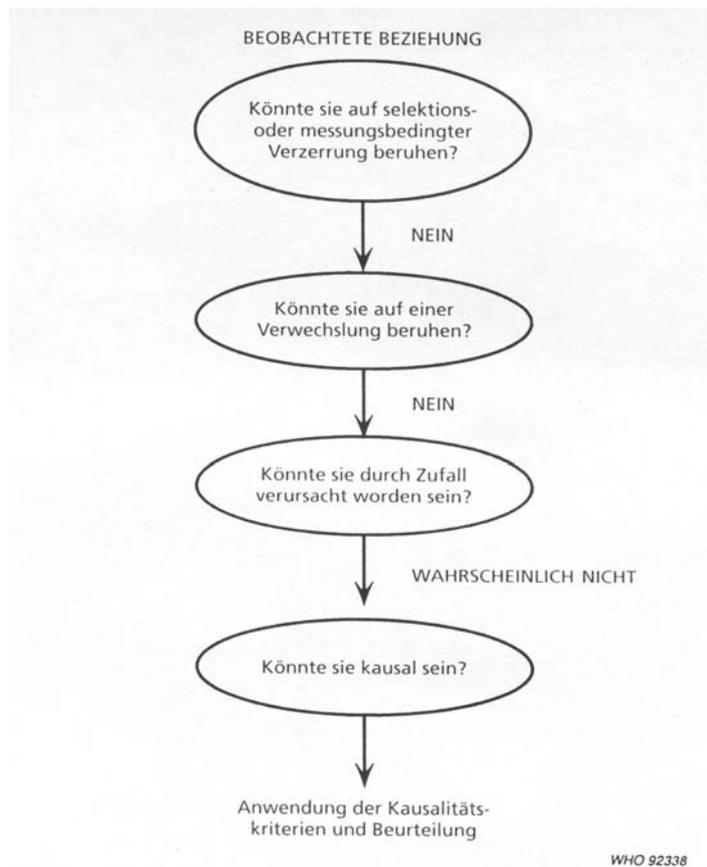


Abbildung 12.1: Bewertung des Zusammenhangs zwischen einem möglichen ursächlichen Faktor und einem Resultatkriterium (Quelle: Beaglehole et al. 1997)

Kriterien für die Bewertung einer Assoziation wurden von Hill 1965 aufgestellt.⁷ Diese (hier modifizierten) Kriterien sollen in der nachstehenden Reihenfolge überprüft werden:

Tabelle 12.2: Kausalitätskriterien von Hill

Zeitliche Beziehung	Geht die Ursache der Wirkung voraus? (unverzichtbares Kriterium)
Plausibilität	Stimmt die Assoziation mit anderen Erkenntnissen überein? (Wirkungsmechanismen, Ergebnisse von Tierversuchen)
Konsistenz	Führen andere Studien zum gleichen Ergebnis?
Stärke	Wie stark ist der Zusammenhang zwischen der Ursache und der Wirkung? (relatives Risiko)
Dosis-Wirkungs-Beziehung	Steigt das Ausmaß des Zusammenhangs mit zunehmender Exposition?
Reversibilität	Führt die Beseitigung einer vermuteten Ursache zu einem verringerten Krankheitsrisiko?
Studienplan	Basieren die Ergebnisse auf einem überzeugenden Studienplan?
Beurteilung der Evidenz	Auf wie vielen Beweisen beruhen die Ergebnisse/Folgerungen?

nach Beaglehole et al. (1997)

Die **zeitliche Beziehung** zwischen Ursache und Wirkung ist von entscheidender Bedeutung: Der Faktor muss **vor** der Krankheit auftreten (vgl. Abbildung 12.2). Die Überprüfung dieses Sachverhalt ist bei Querschnitts- oder Fall-Kontrollstudien schwer bzw. gar nicht möglich.

⁷ Die Originalkriterien nach Hill lauten in dieser Reihenfolge:

1: Strength - 2: Consistency - 3: Specificity - 4: Temporality - 5: Biological gradient - 6: Plausibility - 7: Coherence - 8: Experiment - 9: Analogy

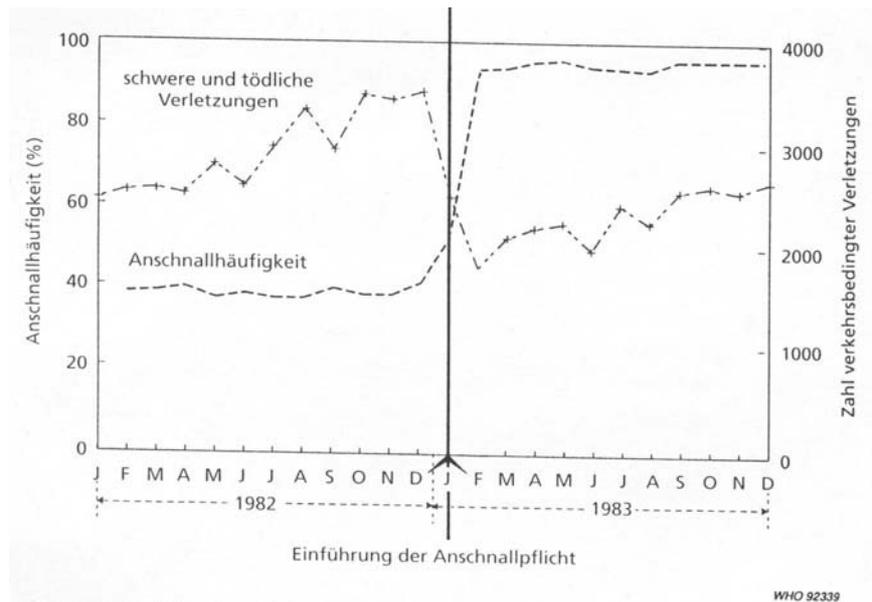


Abbildung 12.2: Anschnall- und Unfallhäufigkeit (Beaglehole et al 1997)

Plausibilität in der Ursache-Wirkungs-Beziehung liegt vor, wenn die beobachtete Assoziation sich plausibel in andere (medizinische, naturwissenschaftliche) Erkenntnissen fügt. Hier ist allerdings Vorsicht geboten, denn epidemiologische Auffälligkeiten können auch vor den entsprechenden naturwissenschaftlichen Erkenntnissen beobachtet werden. Beispiel hierfür sind Snow's bahnbrechende Arbeiten zu den Ursachen der Cholera-Epidemien in London und die Untersuchungen von Semmelweis zur Ätiologie des Kindbettfiebers (vgl. Kapitel 1).

Konsistenz in der Epidemiologie liegt vor, wenn inhaltlich ähnliche Studien zu vergleichbaren Ergebnissen kommen. Mit Hilfe von **Meta-Analysen** kann unter bestimmten Voraussetzungen eine zusammenfassende Bewertung dieser Studien vorgenommen werden (vgl. Abbildung 12.3). Dabei ist aber besondere Vorsicht angebracht: Die Studien müssen gewisse Qualitätsstandards erfüllen, hinsichtlich ihrer Designs, Zielgruppen, Ein- und Ausschlusskriterien, Definition von Fällen und Kontrollen, Erhebungsverfahren zur Bestimmung von Status, Exposition und anderen Variablen sowie anderer Studiencharakteristika vergleichbar sein und die Erfassung aller potenziell relevanter Studien muss vollständig sein.

Eine **starke** Assoziation (relatives Risiko > 2) deutet eher auf eine Ursache-Wirkungs-Beziehung hin als eine schwache. Allerdings ist auch hier Vorsicht geboten: Die Stärke der Assoziation hängt von der Prävalenz anderer potentieller Einflussfaktoren und von ihrem wechselseitigen Zusammenwirken ab. Bei einer hohen Prävalenz (z.B. Übergewicht) kann bevölkerungsbezogen auch ein niedriges relatives Risiko (z.B. bzgl. Herz-Kreislaufkrankungen) hohe versorgungsrelevante Auswirkungen haben. Mittels des attributablen Risikos (vgl. Kapitel 8) kann der Anteil der Krankheits- oder Todesfälle bestimmt werden, die auf einen bestimmten Risikofaktor zurückgeführt werden können. Viele der aktuellen Fragen der Epidemiologie beschäftigen sich mit vergleichsweise niedrigen Risiken, was Probleme hinsichtlich eines aussagefähigen Studiendesigns aufwirft, etwa die Frage nach notwendigen Fallzahlen für ausreichende Power oder die Validität und Genauigkeit der Messinstrumente.

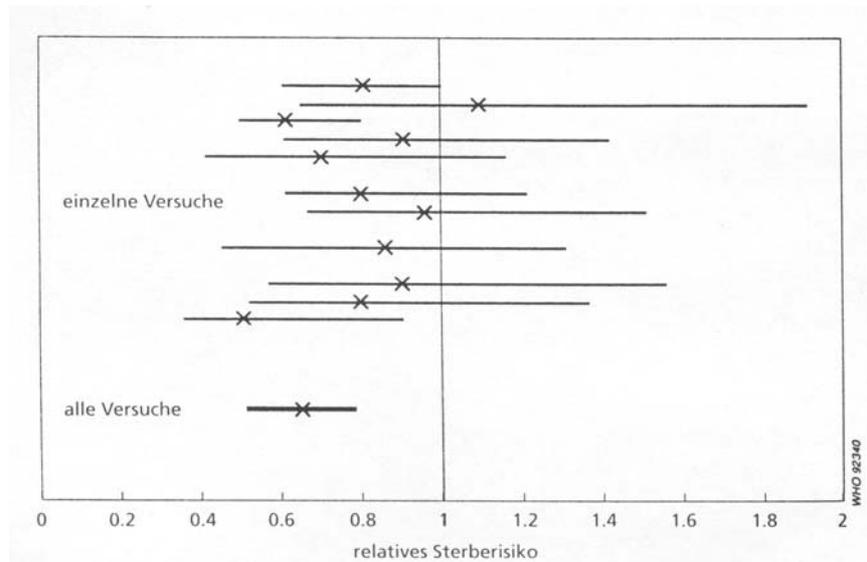


Abbildung 12.3: Metaanalysen kontrollierter randomisierter Studien zum Evaluation von Betablockern zur Prävention von Todesfällen nach Herzinfarkt (Beaglehole et al. 1997)

Eine **Dosis-Wirkungs-Beziehung** liegt vor, wenn Veränderungen im Niveau des Faktors mit Änderungen in der Krankheitshäufigkeit verbunden sind. So steigt im nachfolgenden Beispiel (vgl. Tabelle 12.3) die Prävalenz der Ertaubung sowohl mit der Expositionsdauer als auch mit der Stärke der Exposition. Negative synergistische Effekte der Lärmexposition sind ebenfalls zu beobachten.

Tabelle 12.3: Prozentanteil ertaubter Personen in Abhängigkeit von Expositionsdauer und -intensität

Durchschnittlicher Lärmpegel während eines achtstündigen Arbeitstages (Dezibel)	Expositionsdauer (Jahre)		
	5	10	40
< 80	0	0	0
85	1	3	10
90	4	10	21
95	7	17	29
100	12	29	41
105	18	42	54
110	26	55	62
115	36	71	64

Quelle: Beaglehole et al. 1997

Ursache-Wirkungs-Beziehungen können durch **Reversibilität** geprägt sein. In diesem Fall nimmt die Krankheitshäufigkeit wieder ab, falls der verursachende Faktor nicht mehr wirkt. Zum Beispiel ist das Lungenkrebsrisiko von Ex-Rauchern gegenüber weiterhin Rauchenden verringert (vgl. nachfolgendes Studienbeispiel). Reversibilität lässt sich aber nur dann beobachten, falls noch keine irreversiblen Gesundheitsschädigungen durch die Exposition eingetreten sind wie etwa bei der HIV-Infektion.

Schließlich haben die verschiedenen epidemiologischen **Studientypen** von randomisierten kontrollierten Studien bis zu ökologischen Studien unterschiedliche Aussagekraft, eine ursächliche Beziehung zwischen potenzieller Ursache und Wirkung nachzuweisen. Bei positivem Ausgang der Studie im Sinne einer statistisch signifikanten Assoziation nimmt die Evidenz für eine Ursache-Wirkungs-Beziehung ab (Tabelle 12.4; vgl. auch Kapitel 8).

Tabelle 12.4: Relative Eignung verschiedener Studientypen zum ‚Beweis‘ einer kausalen Beziehung

Studientyp	Eignung	Aussagefähigkeit
Randomisierte kontrollierte Untersuchung	gut	Unter kontrollierten Bedingungen ist die Ursache mit dem Krankheitsauftreten verbunden
Interventionsstudie	gut	Schluss von der Veränderbarkeit der Krankheit durch Veränderung des Risikofaktors auf Kausalität (sogar Reversibilität)
Kohortenstudien	mäßig	Schluss vom erhöhten Risiko für Exponierte auf eine Ursache-Wirkungs-Beziehung
Fall-Kontroll-Studien	mäßig	Schluss vom erhöhten Auftreten des Risikofaktors bei den Erkrankten auf eine Ursache-Wirkungs-Beziehung (‚Wo Rauch ist, da ist auch Feuer‘)
Querschnittstudien	schlecht	Assoziation vorhanden / nicht vorhanden auf individueller Ebene
Ökologische Studien	kaum	Assoziation vorhanden / nicht vorhanden auf Ebene aggregierter Daten (Bevölkerungen)

Quelle: Beaglehole et al. 1997, ergänzt

12.2 Fall-Beispiel: Rauchen und Lungenkrebs

(entnommen aus Gerstmann (1998): Epidemiology kept simple, New York: Wiley)

Eine kausale Beziehung zwischen Rauchen und Lungenkrebs wurde zuerst in den 20er Jahren aufgrund klinischer Beobachtungen vermutet. Um diese Beziehung näher zu untersuchen, wurden zwischen 1930 und 1960 zahlreiche epidemiologische Studien unternommen. Zwei davon wurden von Doll und Hill in Großbritannien durchgeführt. Die erste war eine Fall-Kontroll-Studie (ab 1947), die das Rauchverhalten von Lungenkrebspatienten mit dem anderer Patienten verglich. Bei der zweiten handelte es sich um eine prospektive Studie, die 1951 begann und Todesfälle unter britischen Ärzten in Zusammenhang mit ihrem Rauchverhalten untersuchte.

Die Daten der Fall-Kontroll-Studie wurden in Krankenhäusern Londons und seiner Umgebung gewonnen. Anfangs 20 und später mehr Krankenhäuser wurden zwischen April 1947 und Februar 1952 gebeten, alle neu aufgetretenen Lungenkrebsfälle zu melden. Die Patienten wurden anschließend in einem Interview zu ihrem Rauchverhalten befragt, ebenso die Kontrollen. Die Kontrollen bestanden aus Patienten ohne Krebserkrankung, die zur gleichen Zeit im gleichen Krankenhaus behandelt wurden.

Die Teilnehmer der prospektiven Studie rekrutierte man über das Verzeichnis aller Ärzte, die Oktober 1951 im British Medical Register für England und Wales geführt wurden. Über einen Fragebogen wurden bei ihnen gegenwärtige und vergangene Rauchgewohnheiten erhoben. Informationen über Lungenkrebs und anderen Todesursachen erhielt man über Todesbescheinigungen in den nachfolgenden Jahren.

Frage 1a: Warum handelt es sich in Studie 1 um eine Fall-Kontroll-Studie?

Frage 1b: Warum handelt es sich in Studie 2 um eine prospektive Studie?

Frage 1c: Wie sollte sich die vermutete Assoziation zwischen Rauchen und Lungenkrebs in den Ergebnissen beider Studien niederschlagen?

Im folgenden geht es zunächst um die Fall-Kontroll-Studie.

Frage 2: Warum wurden Krankenhäuser als Ort für die Rekrutierung der Fälle gewählt? Welche anderen Möglichkeiten zur Gewinnung von Fälle und Kontrollen wären denkbar gewesen?

Frage 3: Worin liegen die Vorteile der Gewinnung von Kontrollen aus den gleichen Krankenhäusern wie die Fälle?

Frage 4a: Wie repräsentativ für alle Personen mit Lungenkrebs sind die Patienten aus den Krankenhäusern?

Frage 4a: Wie repräsentativ für alle Personen ohne Lungenkrebs sind die Krankenhaus-Kontrollen ohne Lungenkrebs?

Frage 4c: Wie können diese Überlegungen zur Repräsentativität beider Gruppen die Interpretation der Studienergebnisse beeinflussen?

Mehr als 1700 Lungenkrebsfälle, alle unter 75 Jahre alt, bildeten die Grundgesamtheit der FK-Studie. Ungefähr 15% von ihnen wurden nicht interviewt aufgrund von Tod, Entlassung, Schwere der Krankheit oder mangelhaften Sprachkenntnissen. Einige der bereits interviewten Patienten wurden später ausgeschlossen, als sich die Diagnose Lungenkrebs nicht bestätigte. Die endgültige Fallgruppe enthielt 1.465 Personen (1.357 Männer, 108 Frauen). Die Befragung der Fälle und Kontrollen ergab für die Männer folgendes Ergebnis (Tabelle 12.5).

Tabelle 12.5: Raucherstatus und Lungenkrebs

Studiengruppe	Raucher	Nichtraucher	gesamt
Fälle	1350	7	1.357
Kontrollen	1296	61	1.357
Gesamt	2.646	68	2.714

zitiert nach Gerstman (1998)

Frage 5a: Wie hoch ist der Anteil der Exposition bei Fällen und Kontrollen?

Frage 5b: Berechnen Sie die Odds (Chance) für Rauchen bei Fällen und Kontrollen. Berechnen Sie die Odds Ratio und ein 95%-Konfidenzintervall.

Frage 5c: Unter welchen Bedingungen ist die Odds ratio ein guter Schätzwert für das relative Risiko?

Frage 6: Welche Schlussfolgerungen ziehen Sie aus diesen Ergebnissen?

Die folgende Tabelle 12.6 zeigt die Verteilung von Fällen und Kontrollen auf verschiedene Expositionsklassen bei Männern.

Tabelle 12.6: Lungenkrebs und Stärke der Exposition

	1-14 Zig. tägl.	15-24 Zig. tägl.	25+ Zig. tägl.	Nichtraucher	gesamt
Fälle	565	445	340	7	1.357
Kontrollen	706	408	182	61	1.357
Gesamt	1.271	853	522	67	2.714

zitiert nach Gerstman (1998)

Frage 7: Berechnen Sie die Odds Ratio für jede Expositionsklasse gegenüber jeweils den Nichtrauchern. Interpretieren Sie die Ergebnisse.

Frage 8: Welche anderen Erklärungen für die beobachtete Assoziation gibt es?

Zurück zur prospektiven Studie (Originalpublikation im Anhang). Im Oktober 1951 wurden Fragebögen an 59.600 Ärzte verschickt. Die Ärzte sollten sich selbst in eine von drei Klassen einteilen: (1) aktueller Raucher, (2) Ex-Raucher, (3) Nichtraucher. Raucher und Ex-Raucher wurden nach Art und Umfang des Rauchens, Alter bei Aufnahme des Rauchens und ggf. der

Zeit seit Beendigung des Rauchens gefragt. Als Nichtraucher wurden alle Personen klassifiziert, die nicht mehr als maximal eine Zigarette pro Tag für maximal ein Jahr geraucht haben. Auswertbare Antworten gab es von 40.637 Ärzten (Responserate 68%), von denen 34.445 Männer und 6.192 Frauen waren.

Frage 9a: Welche Gründe mag es für die Durchführung der Studie bei Ärzten gegeben haben?

Frage 9b: In welchem Ausmaß kann die Responserate die Studienergebnisse beeinflusst haben?

Der Rest der Fragen betrifft die männlichen Teilnehmer der Studie im Alter von 35 Jahren und älter. Das Auftreten von Lungenkrebs wurde über 10 Jahre anhand des zentralen Todesbescheinigungsregisters in Großbritannien und der Liste der gestorbenen Ärzte der britischen Ärzteorganisation verfolgt (November 1951 bis Oktober 1961). Jeder Fall von Lungenkrebs wurde anhand von Krankenakten verifiziert.

Lungenkrebs wurde entsprechend der damals verfügbaren Möglichkeiten diagnostiziert, 70% über Biopsie, Autopsie oder Sputum-Zytologie (kombiniert mit Bronchoskopie oder Röntgen), 29% allein über Zytologie oder Bronchoskopie oder Röntgen, nur 1% allein über Fallgeschichten, ärztliche Untersuchung oder Todesbescheinigung.

Von 4.597 Todesfällen insgesamt entfielen 157 auf Lungenkrebs. Abzüglich 4 nicht verifizierbarer Fälle bleiben 153 Fälle für die Auswertung. Tabelle 12.7 zeigt die Zahl der Verstorbenen an Lungenkrebs nach Zahl der täglich gerauchten Zigaretten (ohne die Ex-Raucher). Für jede Kategorie ist die altersstandardisierte Zahl an Personenjahren unter Risiko aufgeführt. Die Angabe der Stärke des Rauchens war für 136 der 153 Fälle verfügbar.

Tabelle 12.7: Lungenkrebs-Sterbefälle nach Zahl der täglich gerauchten Zigaretten

	Nichtraucher	1-14 Zig. tägl.	15-24 Zig. tägl.	25+ Zig. tägl.
Todesfälle	3	22	54	57
Personenjahre (PJ) unter Risiko	42.800	38.600	38.900	25.100
Inzidenzdichte (pro 1.000 PJ)				
Verhältnis der Inzidenzdichten	1,0			
Differenz der Inzidenzdichten	0,0			

zitiert nach Gerstman (1998)

Frage 10a: Was bedeutet das Konzept der Personenjahre? Berechnen Sie Inzidenzdichten, das Verhältnis und die Differenz der Inzidenzdichten für jede Expositionsstufe. Was bedeutet jede der Maßzahlen?

Frage 10b: Tabelle 12.8 zeigt die Anzahl der Lungenkrebstodesfälle bei Rauchern und Nichtrauchern. Berechnen Sie die Inzidenzdichten bei Rauchern, Nichtrauchern und in der Gesamtpopulation. Wenn keiner der Teilnehmer geraucht hätte, welcher Anteil der Lungenkrebsfälle hätte vermieden werden können? Wie wird dieser Anteil genannt?

Frage 10c: Welcher Anteil der Lungenkrebsfälle bei Rauchern ist dem Rauchen zuzuschreiben? Wie wird dieser Anteil genannt?

Tabelle 12.8: Lungenkrebstodesfälle bei Rauchern und Nichtrauchern

	Raucher	Nichtraucher	gesamt
Todesfälle	133	3	136
Personenjahre unter Risiko	102.600	42.800	145.400

zitiert nach Gerstman (1998)

Tabelle 12.9 untersucht den Zusammenhang zwischen Rauchen und Lungenkrebs unter Einbeziehung der Exraucher mit Berücksichtigung der Zeit seit Beendigung des Rauchens.

Tabelle 12.9: Lungenkrebs-Sterbefälle bei Rauchern, Ex-Rauchern und Nichtrauchern

Rauchstatus	Zahl der Fälle	Inzidenzdichte (pro 1.000 PJ)	Verhältnis der Inzidenzdichten
Raucher	133	1,30	18,5
Ex-Raucher, seit:			
weniger als 5 Jahre	5	0,67	9,6
5 bis 9 Jahren	7	0,49	7,0
10 bis 19 Jahren	3	0,18	2,6
20 und mehr Jahren	2	0,19	2,7
Nichtraucher	3	0,07	1,0 (Referenz)

zitiert nach Gerstman (1998)

Frage 11: Welche Schlussfolgerungen legen diese Daten nahe in Bezug auf Public Health und präventive Maßnahmen?

Die Studie untersuchte auch Mortalitätsraten bei kardiovaskulären Krankheiten unter Rauchern und Nichtrauchern. Tabelle 12.10 liefert Informationen zum Vergleich der Mortalität bzgl. Lungenkrebs und kardiovaskulären Krankheiten.

Tabelle 12.10: Mortalität an Lungenkrebs und kardiovaskulären Krankheiten bei Rauchern, Nichtrauchern und in der Allgemeinbevölkerung

	Lungenkrebs	Kardiovaskuläre Krankheiten
Inzidenzdichte insgesamt (pro 1.000 PJ)	0,94	8,87
Inzidenzdichte bei Nichtrauchern	0,07	7,32
Inzidenzdichte bei Rauchern	1,30	9,51
Verhältnis der Inzidenzdichten	18,5	1,3
Differenz der Inzidenzdichten	1,23	2,19
Attributabler Anteil (Bevölkerung)	95%	23%

zitiert nach Gerstman (1998)

Frage 12a: Welche Krankheit hat eine stärkere Assoziation mit dem Rauchen? Warum?

Frage 12b: Wenn die Anzahl der Exzess-Todesfälle durch Rauchen als Maß für die Public-Health-Relevanz genommen wird, welche Krankheit ist dann von größerer Bedeutung?

Wie erwähnt begannen Doll und Hill ihre Fall-Kontroll-Studie 1947, ihre prospektive Studie 1951. Die Ergebnisse der beiden Studien sind nachfolgend vergleichend aufgeführt (Tabelle 12.11).

Tabelle 12.11: Inzidenzdichtenverhältnisse und Odds Ratios der beiden Doll-Hill-Studien

Zahl täglich gerauchter Zigaretten	Inzidenzdichtenverhältnis der Kohortenstudie	Odds Ratios der Fall-Kontroll-Studie
Keine (Nichtraucher)	1,0 (Referenz)	1,0 (Referenz)
1 bis 14	8,1	6,5
15 bis 24	19,8	9,5
25 und mehr	32,4	16,3
Alle Raucher zusammen	18,5	9,1

zitiert nach Gerstman (1998)

Frage 13: Vergleichen Sie die Ergebnisse der beiden Studien. Wo stecken die Gemeinsamkeiten und die Unterschiede in den Maßzahlen der Assoziation?

Frage 14a: Welches sind die Vorteile und Nachteile von Fall-Kontroll- und Kohortenstudie? Füllen Sie die nachstehende Tabelle aus.

Frage 14b: Welchen Studientyp hätten Sie zuerst angewendet? Warum? Warum war eine zweite Studie sinnvoll? Warum hat man einen anderen Studienansatz gewählt?

Tabelle 12.12: Vergleich von Fall-Kontroll- und Kohorten-Studie

	Fall-Kontroll-Studie	Kohortenstudie
Stichprobenumfang		
Kosten		
Studiendauer		
Seltene Krankheiten		
Seltene Expositionen		
Multiple Expositionen		
Multiple Zielereignisse		
Natürlicher Verlauf		
Krankheitsraten		
Selektionsbias		
Recall Bias		
Verlust von Teilnehmern		

zitiert nach Gerstman (1998)

Frage 15: Welcher der folgenden Kriterien für Kausalität (nach Hill) werden durch die Ergebnisse der beiden referierten Studien erfüllt?

- Stärke des Zusammenhangs
- Konsistenz mit anderen Studien
- Zeitliche Abfolge
- Biologischer Gradient
- Spezifität des Effekts
- Biologische Plausibilität

12.3 Antworten zur Fall-Studie

Antwort 1a: In einer Fall-Kontroll-Studie werden Personen mit einer bestimmten Krankheit (hier: Lungenkrebs) mit Personen ohne diese Krankheit verglichen (Kontrollen). Das Ziel ist die Untersuchung der Antwort, ob sich die beiden Gruppen bzgl. der Exposition gegenüber einem spezifischen Faktor (hier: Rauchen) unterscheiden.

Antwort 1b: Eine prospektive Studie (Kohortenstudie) erstreckt sich auf eine gut definierte Gruppe von Personen, von denen einige gegenüber einem bestimmte Risikofaktor exponiert sind und andere nicht. Über einen bestimmten Zeitraum werden Exponierte und Nichtexponierte auf das Auftreten einer bestimmten Krankheit hin beobachtet.

Antwort 1c: Im Sinne der Hypothese müsste in der FK-Studie der Anteil der Raucher unter den Lungenkrebsfällen höher als in der Kontrollgruppe liegen. In der prospektiven Studie sollte die Lungenkrebssterblichkeit unter Rauchern höher als unter Nichtrauchern sein.

Antwort 2: Krankenhäuser haben eine hohe Wahrscheinlichkeit Fälle zu finden, diese genau zu diagnostizieren und anschließend zu befragen. Mögliche andere Quellen für Fälle sind: Krebsregister, Todesfallregister, Versicherungsdateien, Praxen, berufsbezogene Dateien; Kontrollen: Nachbarschaften, Bekanntschaften, Bevölkerungskontrollen.

Antwort 3: Vorteile sind: guter Zugang, zeitliche Übereinstimmung, Kontrolle für sozioökonomischen Status und Umwelteinflüsse, Wohnort, Zugang zur Versorgung, Art und Genauigkeit der Diagnosestellung.

Antwort 4a: Repräsentativ, weil fast alle Fälle irgendwann im Krankenhaus behandelt werden. Dennoch mögen sie kränker sein, ihre Krankheit in fortgeschrittenerem Stadium sein oder mehr Komplikationen bzw. andere Krankheiten haben gegenüber nicht stationär behandelten Lungenkrebspatienten.

Antwort 4b: Meistens sind Krankenhauskontrollen nicht repräsentativ für die Allgemeinbevölkerung. allein durch den Tatbestand, dass sie stationär behandelt wurden.

Antwort 4c: Über die Kontrollgruppe soll die Prävalenz der Exposition in der Allgemeinbevölkerung geschätzt werden. Krankenhauskontrollen können auch wegen anderer rauchenassoziierter Krankheiten behandelt worden sein. Dann ist der Anteil der Raucher unter den Kontrollen größer als in der Allgemeinbevölkerung und das wahre Risiko des Rauchens für Lungenkrebs würde unterschätzt. Dies ist ein Beispiel für Selektionsbias.

Antwort 5a: Anteil bei Fällen: $1350/1357 = 0,995$; Anteil bei Kontrollen: $1296/1357 = 0,955$

Antwort 5b: Odds für Rauchen bei Fällen: $1350/7 = 192,9$; Odds für Rauchen bei Kontrollen: $1286/61 = 21,2$; Odds ratio (Kreuzproduktverhältnis): $(1350*61)/(1296*7) = 9,1$; Konfidenzintervall (4,0; 21,8)

Antwort 5c: Wenn a) die untersuchte Krankheit selten auftritt und b) die Fälle bzw. Kontrollen hinsichtlich ihrer Expositionsanamnese jeweils repräsentativ für alle Erkrankten bzw. Nichterkrankten sind. [Über die vermutliche Selektion durch die Wahl der Kontrollen in dieser Studie haben wir in 4c reflektiert.]

Antwort 6: Die Odds für Rauchen sind bei Lungenkrebspatienten 9fach gegenüber der Chance bei Kontrollen erhöht. Da Lungenkrebs eine seltene Krankheit ist, ist dieser Wert eine gute Näherung für das relative Risiko der Raucher. Zu beachten ist die hohe Gesamtprävalenz der Raucher bei Fällen und Kontrollen (> 95%).

Antwort 7: Die drei OR ergeben sich als $(565*61)/(706*7) = 6,5$ für Expositionsklasse unter 15 Zigaretten; $(445*61)/(408*7) = 9,5$ für Exposition 15 bis 24 Zigaretten pro Tag; $(340*61)/(182*7) = 16,1$ für Exposition von mehr als 24 Zigaretten pro Tag. Die OR nimmt stetig mit Stärke der Exposition zu, zeigt also eine Dosis-Wirkungs-Beziehung zwischen der Zahl täglich gerauchter Zigaretten und der Stärke der Assoziation.

Antwort 8: Selektionsbias, Informationsbias, Confounding und Zufall (trotz statistischer Signifikanz) sind mögliche Erklärungen. Der mögliche Selektionsbias ist in Antwort 4c bereits thematisiert worden; er würde das wahre RR jedoch unterschätzen. Ein Informationsbias würde auftreten, wenn Lungenkrebspatienten sich ihrer Rauchgewohnheiten in der Vergangenheit

besser erinnerten als die Kontrollen. Angesichts des damals noch weitgehend unbekanntem Zusammenhangs zwischen Rauchen und Lungenkrebs ist dies nicht sehr wahrscheinlich. Alter mag ein Confounder sein. Ein Confounder ist mit der Exposition assoziiert und gleichzeitig ein unabhängiger Risikofaktor des Zielereignisses. Wenn Lungenkrebs unter älteren Personen häufiger auftritt und gleichzeitig der Anteil der Raucher altersabhängig ist, dann mag der beobachtete Zusammenhang durch das Alter (mit-)verursacht worden sein.

Antwort 9a: Aufgrund des beruflichen Interesses konnte eine hohe Akzeptanz der Studie und Validität der Daten erwartet werden. Ärzte gehören vorwiegend der Mittel- und Oberschicht an, so dass mögliche Verzerrungen aus der vorherigen Studie aufgrund der sozialen Herkunft der Teilnehmer überprüft werden können.

Antwort 9b: Allgemein werden Responseraten von mindestens 80% angestrebt. Allerdings stellt der realisierte Wert von 68% schon ein gutes Ergebnis dar. Wenn die Teilnahme an der Studie weder mit der Exposition noch mit dem Auftreten des Zielereignisses assoziiert ist, mindert eine geringe Teilnahmerate nur die Power der Studie, also die Fähigkeit, Unterschiede zwischen Exponierten und Nichtexponierten bzgl. des Zielereignisses aufzudecken, führt aber nicht zu einem Bias. Anderenfalls ist ein Selektionsbias ein Problem. Daher sollten die Nichtteilnehmer anhand verfügbarer Merkmale wie Alter, Geschlecht u.ä. so gut wie möglich charakterisiert und mit den Teilnehmern verglichen werden.

Antwort 10a: siehe nachstehende Tabelle 12.13. Beim Konzept der Personenjahre (PJ) wird individuell die Zeit unter Risiko ermittelt und dann für diese verschiedenen Klassen aufsummiert. Ein Teilnehmer, der bis zum Ende der Studie überlebt, steuerte 10 PJ bei, ein 1958 an Lungenkrebs Verstorbener 7 PJ, ein 1954 ins Ausland verzogener Teilnehmer noch 3 PJ. Inzidenzdichte ist ein Maß für das Risiko an Lungenkrebs zu sterben. Das Verhältnis der Inzidenzdichten gibt das **relative Risiko** wieder, die Differenz der Inzidenzdichten den absoluten Unterschied der Risiken.

Tabelle 12.13: Lungenkrebs-Sterbefälle nach Zahl der täglich gerauchten Zigaretten

	Nichtraucher	1-14 Zig. tägl.	15-24 Zig. tägl.	25+ Zig. tägl.
Todesfälle	3	22	54	57
Personenjahre unter Risiko	42.800	38.600	38.900	25.100
Inzidenzdichte (pro 1.000 PJ)	0,07	0,57	1,39	2,27
Verhältnis der Inzidenzdichten	1,0	8,1	19,8	32,4
Differenz der Inzidenzdichten	0,0	0,50	1,32	2,20

zitiert nach Gerstman (1998)

Antwort 10b: Inzidenzdichte bei Rauchern: $133 / 102,600 = 1,296$ pro 1.000 PJ; bei Nichtrauchern: $3 / 42,8 = 0,07$ pro 1.000 PJ; insgesamt: $136 / 145,4 = 0,935$ pro 1.000 PJ; Verhältnis der Inzidenzdichten: $1,296 / 0,07 = 18,5$ (Tabelle 12.14).

Anteil der vermeidbaren Todesfälle ohne Raucher: $(0,935 - 0,07) / 0,07 = 0,925$. Dieser Anteil heißt attributabler Anteil in der Bevölkerung

Antwort 10c: Anteil der zuschreibbaren Todesfälle unter Rauchern: $(18,5 - 1) / 18,5 = 0,95$ (attributabler Anteil unter den Exponierten, entspricht 126 von 133 Fällen)

Tabelle 12.14: Lungenkrebstodesfälle bei Rauchern und Nichtrauchern

	Raucher	Nichtraucher	gesamt
Todesfälle	133	3	136
Personenjahre unter Risiko	102.600	42.800	145.400
Inzidenzdichte (pro 1.000 PJ)	1,296	0,07	0,935

zitiert nach Gerstman (1998)

Antwort 11: Das niedrigste Risiko ist bei Nichtrauchern zu beobachten. Darüber hinaus sinkt das relative Risiko für Lungenkrebs mit zunehmender Zeit seit Beendigung des Rauchens, ist aber auch nach zehn Jahren noch um den Faktor 3 gegenüber Nichtrauchern erhöht. Daher sind Maßnahmen zur Aufgabe des Rauchens zwar sinnvoll, der größere Effekt ist aber von einer Primärprävention des Rauchens zu erwarten.

Antwort 12a: Das Verhältnis der Inzidenzdichten ist das primäre Maß für die Stärke der Assoziation. Demnach besteht ein stärkerer Zusammenhang zwischen Rauchen und Lungenkrebs als zwischen Rauchen und kardiovaskulären Krankheiten.

Antwort 12b: Die Differenz der Inzidenzdichten pro 1.000 Personenjahre ist bei kardiovaskulären Krankheiten größer als bei Lungenkrebs, obwohl das Verhältnis der Inzidenzdichten vergleichsweise niedrig ist. Ohne Rauchen würden demnach mehr kardiovaskuläre Todesfälle als Todesfälle durch Lungenkrebs vermieden werden.

Antwort 13: Das Odds Ratio in der Fall-Kontroll-Studie unterschätzt konsistent das Verhältnis der Inzidenzdichten, wahrscheinlich wegen der Verwendung von Krankenhauskontrollen. Dennoch liefern beide Studien konsistente Ergebnisse, incl. einer Dosis-Wirkungs-Beziehung.

Antwort 14a: siehe die gefüllte Tabelle 12.14.

Antwort 14b: Allgemein ist eine Fall-Kontroll-Studie schneller und kostengünstiger durchzuführen als eine Kohortenstudie. Wenn eine FK-Studie Ergebnisse liefert, die weiterer Bestätigung bedürfen, bietet sich eine zweite Studie an. Eine prospektive Studie kann eine Bestätigung der Ergebnisse liefern, gibt darüber hinaus tiefere Einblicke in die zeitliche Abfolge von der Exposition bis zum Ausbruch der Krankheit und ihrem weiteren Verlauf, erlaubt die Berechnung von Krankheitsraten und kann meist leichter verallgemeinert werden.

Tabelle 12.15: Vergleich von Fall-Kontroll- und Kohorten-Studie

	Fall-Kontroll-Studie	Kohortenstudie
Stichprobenumfang	klein	groß
Kosten	gering	hoch
Studiendauer	kurz	lang
Seltene Krankheiten	günstig	ungünstig
Seltene Expositionen	ungünstig	günstig
Multiple Expositionen	günstig	ungünstig
Multiple Zielereignisse	ungünstig	günstig
Natürlicher Verlauf	ungünstig	günstig
Krankheitsraten	nicht messbar	günstig
Selektionsbias	potenziell problematisch	weniger problematisch
Recall Bias	potenziell problematisch	weniger problematisch
Verlust von Teilnehmern	weniger problematisch	potenziell problematisch

zitiert nach Gerstman (1998)

Antwort 15: Die Ergebnisse dieser Studien erfüllen die Kriterien Stärke, Konsistenz, zeitliche Abfolge und Dosis-Wirkungs-Beziehung. Demgegenüber können zu den Kriterien Spezifität des Effekts und biologische Plausibilität keine Aussagen getroffen werden.

13. Gute Epidemiologische Praxis

Die in den vorhergehenden Kapiteln dargelegten methodischen und praktischen Aspekte der Epidemiologie werden leider nicht in allen Studien beachtet. Mängel in der Studienplanung, -durchführung und -auswertung werfen nicht nur ein schlechtes Licht auf die betreffenden Studien, sondern unter Umständen auch auf die gesamte Epidemiologie, die sich in Deutschland noch nicht überall als ernstzunehmende Wissenschaft etabliert hat. Nicht zu vergessen sind mehrere Skandale um manipulierte Studienergebnisse in der klinischen Forschung in den letzten Jahren, zuletzt in Südkorea Ende 2005, wengleich auch auf einem Forschungsgebiet außerhalb der Epidemiologie.

Aus diesen Gründen hat eine Arbeitsgruppe der (ehemaligen) Deutschen Arbeitsgemeinschaft für Epidemiologie (da epi), die Dachorganisation der in Deutschland tätigen Epidemiologen, in einem aufwändigen wissenschaftlichen Konsensprozess Grundsätze Guter Epidemiologischer Praxis (GEP) formuliert, so wie es sie in anderen Bereichen bereits seit längerem gibt (GCP – Good Clinical Practice; GLP – Good Laboratory Practice). Diese Grundsätze verstehen sich als externer Qualitätsmaßstab für die Bewertung epidemiologischer Studien, ebenso wie als handlungsleitende Orientierung für eigene Studien planende Epidemiologen.

Die GEP wurden im Jahr 2000 veröffentlicht und umfassen **11 Leitlinien**, die wiederum in mehrere **Empfehlungen** untergliedert sind. An dieser Stelle seien nur kommentarlos die Leitlinien und Empfehlungen aufgeführt. Für die detaillierten Erläuterungen sei auf die Homepage der Deutsche Gesellschaft für Epidemiologie verwiesen (www.dgepi.de).⁸ Dort stehen die GEP auch zum Download bereit. Die GEP wurden im Sommer 2003 einer ersten Evaluierung und Überarbeitung unterzogen. Die revidierte Fassung wurde im Jahr 2005 veröffentlicht.

Leitlinie 1 (Ethik): Epidemiologische Untersuchungen müssen in Einklang mit ethischen Prinzipien durchgeführt werden und Menschenwürde sowie Menschenrechte respektieren.

Empfehlung 1.1: Vor der Durchführung einer epidemiologischen Studie soll die Stellungnahme einer Ethikkommission eingeholt werden.

Leitlinie 2 (Forschungsfrage): Die Planung einer epidemiologischen Studie erfordert explizite und operationalisierbare Fragestellungen, die spezifisch und so präzise wie möglich formuliert sein müssen. Die Auswahl der zu untersuchenden Bevölkerungsgruppen muss im Hinblick auf die Forschungsfrage begründet werden.

Empfehlung 2.1: Wenn in einer Studie Hypothesen konfirmatorisch geprüft werden sollen, müssen diese vor Beginn der Studie formuliert werden.

Leitlinie 3 (Studienplan): Grundlage einer epidemiologischen Studie ist ein detaillierter und verbindlicher Studienplan, in dem die Studiencharakteristika schriftlich festgelegt werden.

Empfehlung 3.1: Der Studientyp soll beschrieben und seine Wahl angemessen begründet werden.

Empfehlung 3.2: Die Studienbasis und das Auswahlverfahren der Studienteilnehmer sollen beschrieben und angemessen begründet werden.

Empfehlung 3.3: Bereits bei der Planung epidemiologischer Studien soll möglichen Verzerrungen der Ergebnisse entgegen gewirkt werden.

Empfehlung 3.4: Das Konzept der Minimierung und Kontrolle potenzieller Selektionsverzerrungen auf Grund von Nichtteilnahme und Nichtverfügbarkeit der Daten zu ausgewählten Studienteilnehmern soll im Studiendesign festgehalten werden.

⁸ Die dae hat sich im Herbst 2005 neu konstituiert. Die Nachfolgeorganisation heißt Deutsche Gesellschaft für Epidemiologie (www.dgepi.de).

Empfehlung 3.5: Alle interessierenden Variablen sollen präzise definiert und möglichst standardisiert operationalisiert werden. Für die Bestimmung sind möglichst valide und reliable Mess- und Erhebungsinstrumente einzusetzen.

Empfehlung 3.6: Im Studienplan ist eine Begründung und quantitative Abschätzung des Studienumfangs anzugeben.

Empfehlung 3.7: Ergänzend zum Studienplan sollten in einem Operationshandbuch sämtliche organisatorischen Festlegungen zur Vorbereitung und Durchführung der Studie einschließlich der Erhebungsinstrumente dokumentiert werden.

Empfehlung 3.8: Für die Auswertungsphase der Studie sind ausreichende zeitliche und personelle Ressourcen vorzusehen.

Leitlinie 4 (Probenbanken): In vielen epidemiologischen Studien ist die Anlage einer biologischen Probenbank notwendig bzw. sinnvoll. Hierfür und für die aktuelle und vorgesehene zukünftige Nutzung der Proben ist die dokumentierte Einwilligung aller Probanden erforderlich.

Empfehlung 4.1: Die verantwortliche Institution und die verantwortlichen Personen, die für die Führung der Probenbank zuständig sind, sollten den Probanden gegenüber benannt werden. Dabei sollen Art und Menge des entnommenen Materials zusammen mit Lagerungsform, -ort und -dauer beschrieben werden. Die Probanden sind über die Eigentumsverhältnisse an den entnommenen Proben aufzuklären.

Empfehlung 4.2: Bei der Nutzung von in Probenbanken asserviertem Material für primär nicht geplante Fragestellungen sind die Leitlinien für GEP erneut zu berücksichtigen.

Leitlinie 5 (Qualitätssicherung): In epidemiologischen Studien ist eine begleitende Qualitätssicherung aller relevanten Instrumente und Verfahren sicherzustellen.

Empfehlung 5.1: In jeder epidemiologischen Studie, bei der Primärdaten erhoben werden, ist zu prüfen, ob vor Beginn der Hauptstudie eine separate Pilotstudie erforderlich ist.

Empfehlung 5.2: Ergibt sich während der Durchführung der Studie die Notwendigkeit, die dort festgelegten Verfahrensweisen zu verändern, so sind diese zu begründen, zu dokumentieren und allen Studienmitarbeitern rechtzeitig bekannt zu geben.

Empfehlung 5.3: Vor Beginn der Feldarbeit sollen die an der Datenerhebung beteiligten Personen ausführlich geschult und ausgebildet werden.

Empfehlung 5.4: Regeln und Erläuterungen für die Durchführung der Erhebung sollen in Form eines Erhebungshandbuches schriftlich fixiert werden und dem Erhebungspersonal zur Verfügung stehen. Das Erhebungshandbuch wird Bestandteil des Operationshandbuches.

Empfehlung 5.5: Insbesondere bei großen, zeitlich lang dauernden und multizentrischen Untersuchungen ist zu überprüfen, ob eine Qualitätssicherung der Verfahren über eine externe Person oder Institution erfolgen sollte.

Leitlinie 6 (Datenhaltung und -dokumentation): Für die Erfassung und Haltung aller während der Studie erhobenen Daten sowie für die Aufbereitung, Plausibilitätsprüfung, Kodierung und Bereitstellung der Daten ist vorab ein detailliertes Konzept zu erstellen.

Empfehlung 6.1: Alle während der Studie erhobenen Daten (Dokumentationsbögen, Fragebögen, Mess- und Laborwerte etc.) sollen zeitnah in eine Datenbank überführt werden, die eine sichere Erfassung und Haltung der Daten gewährleistet.

Empfehlung 6.2: Der nach Prüfeingabe erhaltene Rohdatensatz soll in unveränderter Form aufbewahrt werden.

Empfehlung 6.3: Eine Kodierung der Daten hat stets unabhängig zu erfolgen, d.h. blind für den jeweiligen Status bzw. die Gruppenzugehörigkeit der betreffenden Person.

Empfehlung 6.4: Plausibilitätskontrollen erfolgen prinzipiell auf der Grundlage des prüfeingebenden Rohdatensatzes. Eventuell erforderliche Änderungen der Variablenwerte oder die Bildung neuer Variablen sind in jedem Fall schriftlich zu dokumentieren.

Empfehlung 6.5: Der nach Plausibilitätsprüfung und Datenkorrektur überarbeitete Datensatz ist als Auswertungsdatensatz zu kennzeichnen und unabhängig vom Rohdatensatz zu speichern.

Leitlinie 7 (Auswertung): Die Auswertung epidemiologischer Studien soll unter Verwendung adäquater Methoden und ohne unangemessene Verzögerung erfolgen. Die den Ergebnissen zugrunde liegenden Daten sind in vollständig reproduzierbarer Form für mindestens 10 Jahr aufzubewahren.

Empfehlung 7.1: Die Auswertung zu den einzelnen Fragestellungen soll nach einem zuvor schriftlich erstellten Analyseverfahren erfolgen.

Empfehlung 7.2: Zwischenauswertungen sollen nur begründet durchgeführt werden.

Empfehlung 7.3: Die Auswertungen epidemiologischer Studiendaten sollen vor der Publikation der Gegenprüfung unterzogen werden. Die ihnen zugrunde liegenden Daten und Programme sollen anschließend in vollständig reproduzierbarer Form archiviert werden.

Leitlinie 8 (Datenschutz): Bei der Planung und Durchführung epidemiologischer Studien ist auf die Einhaltung der geltenden Datenschutzvorschriften zum Schutz der informationellen Selbstbestimmung zu achten.

Leitlinie 9 (vertragliche Rahmenbedingungen): Die Durchführung einer epidemiologischen Studie setzt definierte rechtliche und finanzielle Rahmenbedingungen voraus. Hierzu sind rechtswirksame Vereinbarungen zwischen Auftraggeber und Auftragnehmer sowie zwischen Partnern von Forschungsk Kooperationen anzustreben.

Empfehlung 9.1: Mit dem Auftraggeber sollen transparente und realistische Vereinbarungen getroffen werden. Bei der Vielfältigkeit der speziellen Konstellationen sind unterschiedliche Vertragsformen möglich.

Empfehlung 9.2: Die Publikation der Ergebnisse einer Auftragsforschung darf weder durch den Auftraggeber noch durch die Studienleitung verhindert, behindert oder unzumutbar verzögert werden.

Empfehlung 9.3: Schriftliche Vereinbarungen sollen grundsätzlich mit allen Kooperationspartnern erfolgen. Dies gilt unabhängig davon, ob es sich um gleichberechtigte Studienzentren im Rahmen einer multizentrischen Studie handelt oder ein Kooperationspartner im Sinne eines Auftragnehmers ein oder mehrere Arbeitspakete eines größeren Studienprojektes bearbeitet.

Leitlinie 10 (Interpretation): Die Interpretation der Forschungsergebnisse einer epidemiologischen Studie ist Aufgabe des/der Autor/en einer Publikation. Grundlage jeder Interpretation ist eine kritische Diskussion der Methoden, Daten und Ergebnisse der eigenen Untersuchung im Kontext der vorhandenen Evidenz. Alle Publikationen sollten einem externen Review unterworfen werden.

Leitlinie 11 (Kommunikation und Public Health): Epidemiologische Studien, deren Anliegen die Umsetzung von Ergebnissen in gesundheitswirksame Maßnahmen ist, sollten die betroffenen Bevölkerungsgruppen angemessen einbeziehen und eine qualifizierte Risikokommunikation mit der interessierten Öffentlichkeit anstreben.

Empfehlung 11.1: Ergibt sich nach dem professionellen Urteil des Autors aus den Forschungsergebnissen einer epidemiologischen Studie die Notwendigkeit von Konsequenzen, sollen diese, beispielsweise in Form einer Empfehlung, explizit formuliert werden. Dabei müssen sich Epidemiologen bei Bedarf auch für eine effektive Risikokommunikation mit Nicht-Epidemiologen verantwortlich fühlen.

Empfehlung 11.2: Die in einer Studie eingesetzten Instrumente sollen Interessierten offen gelegt werden.

Empfehlung 11.3: Bei jeder Studie sollte geprüft werden, ob und inwieweit der Datensatz der Erhebung der wissenschaftlichen Öffentlichkeit für Forschungszwecke angeboten wird.

Die GEP beziehen sich auf alle Formen epidemiologischer Studien. Für die wichtigen Bevölkerungsumfragen (vgl. den Bundesgesundheitsurvey 1998) gibt es ergänzend und in seinen Erläuterungen z.T. ausführlicher die ‚Best Practice for Survey and Public Opinion Research‘ der American Association for Public Opinion Research (AAPOR; www.aapor.org):

1. Umfragen sollten spezifischen Zielen dienen.
2. Anstelle von Umfragen sollten alternative Formen der Datenerhebung in Erwägung gezogen werden.
3. Stichproben sollten so gewählt werden, dass sie die zu untersuchende Population gut repräsentieren.
4. Forschungsdesigns sollten so gewählt werden, dass sie Kosten und Fehler im Gleichgewicht halten.
5. Fragen sollten so formuliert werden, dass sie zu den messenden Konzepten und der zu untersuchenden Population passen.
6. Fragebogen und das gesamte Verfahren sollten einem Prätest unterzogen werden, um Probleme noch im Vorfeld der Umfrage aufzudecken.
7. Interviewer sollten sorgfältig mit den Befragungstechniken und den Inhalten einer Umfrage vertraut gemacht werden.
8. In jeder Phase der Umfrage sollten Qualitätsprüfungen erfolgen.
9. Die Mitarbeit der Befragungspersonen und die Ausschöpfungsrate sollen innerhalb der durch Ethik gesetzten Grenzen maximiert werden.
10. Statistische Analysen und Berichtstechniken sollen den Daten angemessen sein.
11. Vertraulichkeitsversprechen an die Befragten sollen sorgfältig geplant und eingehalten werden.
12. Alle Methoden und Verfahren einer Umfrage sollen veröffentlicht werden, um Evaluation und Replikation zu ermöglichen.

Neben den im folgenden Kapitel niedergelegten Gesichtspunkten für das kritische Lesen epidemiologischer Studien dienen die GEP und die Guidelines nicht nur der Strukturierung und Planung eigener Studien, sondern auch der Bewertung fremder Untersuchungen sowie der Gestaltung der formalen Beziehungen zwischen den Studienbeteiligten.

Angesichts der steigenden Nutzung von Sekundärdaten in der (Versorgungs-)Epidemiologie, dazu zählen in erster Linie Routinedaten der gesetzlichen Krankenversicherung, Renten- und Unfallversicherung sowie Daten der Bundesagentur für Arbeit, aber auch Daten der Krebs- und Krankheitsregister, hat die Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS) der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSM) im Jahr 2005 als Ergänzung der GEP eine Gute Praxis Sekundärdatendatenanalyse veröffentlicht, die im Jahr 2007 einer ersten Revision unterzogen wurde, wobei eine formale Anbindung an die GEP erfolgte (download über www.dgsmp.de).

14. Hinweise zum kritischen Lesen epidemiologischer Literatur

14.1 Kriterien der Literaturanalyse

Ein wesentlicher Aspekt der täglichen Arbeit eines Gesundheitswissenschaftlers oder Epidemiologen ist das Studium epidemiologischer, sozialmedizinischer oder sozialwissenschaftlicher Studien und Veröffentlichungen, sei es bei der Planung einer neuen Studie oder zur Einordnung und Bewertung eigener Forschungsergebnisse. Angesichts des großen Umfangs relevanter Literatur und der unterschiedlichen Qualität der Zeitschriften und Beiträge sollte man sich grundsätzlich um eine kritische Distanz zu Veröffentlichungen und um deren systematische Bewertung bemühen.

Das kritische Lesen epidemiologischer Literatur bezieht sich auf alle Teile eines wissenschaftlichen Artikels, beginnend mit der Einleitung und der Abgrenzung des Forschungsthemas sowie der Formulierung der Fragestellung. Es folgt das Design der Untersuchung mit Beschreibung der Population, der Erhebungsinstrumente und der Analyseverfahren (Material und Methode, M+M), die Darstellung und Interpretation der Ergebnisse sowie die Überprüfung der Hypothesen bis zu den Schlussfolgerungen und Konsequenzen aus der Studie.

Im einzelnen können folgende Fragen an den Text gestellt werden. Dabei sind die Kriterien nicht immer eindeutig voneinander zu trennen. Ebenso sind nicht alle der hier aufgeführten Kriterien für jede Studie relevant. So entfällt bei einer Querschnittsstudie häufig die Wahl einer externen Kontrollgruppe und bei explorativen Studien ist die Angaben der Hypothesen und des Signifikanzniveaus nicht zwingend erforderlich.

I) Ziel oder Hypothese der Untersuchung

1. Auf welchen theoretischen Überlegungen basieren die Fragestellungen?
2. Was sind die Ziele der Studie und welche Fragen sollen durch sie beantwortet werden?
3. Sind die Hypothesen genannt? Sind sie plausibel?
4. Auf welche Grundgesamtheit (Bevölkerung) sollen sich die erwarteten Ergebnisse beziehen?

IV) Design der Untersuchung

5. Handelt es sich um ein Experiment (Interventionsstudie), wurde eine Analyse klinischer Daten vorgenommen oder wurden Beobachtungen gemacht?
6. Welcher epidemiologischer Studientyp liegt vor? Handelt es sich um eine prospektive Kohortenstudie, um eine retrospektive Fall-Kontroll-Studie oder um eine Querschnittsstudie?
7. Passt das Studiendesign zur epidemiologischen Fragestellung und den Hypothesen?
8. Wäre ein anderes Design denkbar gewesen? Wenn ja, welches?
9. Wie groß ist der Umfang der Studie?
10. Wurde eine Vollerhebung durchgeführt oder eine Stichprobe gezogen?
11. Wie wurde die Stichprobe gezogen? Ist sie repräsentativ oder gibt es Gründe zu vermuten, dass sie für die Grundgesamtheit nicht typisch ist? Wie groß sind Brutto- und Nettostichprobe?
12. Wie wurden die Studienmitglieder identifiziert? Sind die Ein- und Ausschlusskriterien genannt? Sind damit alle Zielpersonen erfasst?
13. Ist eine Vergleichsgruppe in die Studie einbezogen? Wären andere Kontrollen denkbar gewesen? Sind Selektionseffekte bei der Wahl der Kontrollen möglich? Beeinflussen sie die Aussagefähigkeit der Studie?
14. Welche Software wurde benutzt?

III) Beobachtungen

15. Sind die Hypothesen (vgl. 3.) in geeigneter Weise operationalisiert?
16. Welche Erhebungsinstrumente wurden eingesetzt? Wie wurde die Exposition ermittelt? Wie wurden die Diagnosen und andere Zielgrößen bestimmt?
17. Sind die Erhebungsinstrumente hinsichtlich Inhalten, Validität, Reliabilität klar beschrieben?
18. Sind die Erhebungsinstrumente für die Ziele der Studie geeignet?
19. Sind Voraussetzungen wie diagnostische Kriterien, Messungen, Endpunktkriterien klar definiert und benannt?
20. Sind die Methoden der Klassifizierung und Messungen einheitlich bei allen Versuchspersonen gewesen?
21. Sind die Beobachtungen valide und reliabel (reproduzierbar)? Gibt es möglicherweise Verzerrungen? Wenn ja, wurden sie berücksichtigt?
22. Sind potenzielle Confounder benannt und erhoben worden?
23. Sind die angewandten statistischen Verfahren benannt und begründet? Ist der Modellbildungsprozess und die Auswertestrategie beschrieben?

IV) Darstellung der Ergebnisse

24. Sind die Ergebnisse klar, objektiv und in ausreichendem Detail dargestellt, so dass die Leser sie selbst beurteilen können?
25. Sind die Ergebnisse schlüssig? Stimmen die Kalkulationen?
26. Sind die Tabellen und Abbildungen nachvollziehbar und verständlich?
27. Eignen sich die Daten zur statistischen Auswertung? Sind die statistischen Methoden für die Art der Daten angemessen? Ist die Analyse korrekt durchgeführt und interpretiert worden?
28. Welches sind die Hauptresultate?
29. Sind die Befunde konsistent?
30. Welches Beziehungsmaß wurde gewählt?
31. Wurde ein adäquates Signifikanzniveau gewählt?
32. Werden auch nicht-signifikante Ergebnisse berichtet?
33. Gibt es eine Dosis-Wirkungs-Beziehung?
34. Können die Befunde durch Verzerrungen oder Missklassifikationen erklärt werden?
35. Wurde für Confounder adäquat kontrolliert?
36. Wurde potenzieller Bias untersucht und ggf. quantifiziert (z.B. durch eine Sensitivitätsanalyse)?

V) Diskussion

37. Sind die eingangs gestellten Hypothesen befriedigend untersucht worden?
38. Gibt eine angemessene Trennung zwischen Hypothesentestung und Generierung neuer Hypothesen bzw. ad-hoc-Auswertungen?
39. Sind die Schlussfolgerungen gerechtfertigt? Lassen sich die Ergebnisse auf die Grundgesamtheit übertragen?
40. Gibt es andere Erklärungsansätze für die beobachteten Ergebnisse? Gibt es eine kritische Bewertung des eigenen Vorgehens? Werden die Begrenzungen der Studie erkannt und diskutiert?
41. Werden mögliche Verzerrungen diskutiert und ihren Auswirkungen abgeschätzt?
42. Gibt es eine adäquate Auseinandersetzung mit anderen Studien?
43. Sind die Ergebnisse mit denen anderer Studien vergleichbar? Wie werden Unterschiede erklärt?

44. Erfolgt eine Bewertung der Exposition und des damit verbundenen Risikos für die untersuchte Population und die Gesamtbevölkerung?
45. Werden Schlussfolgerungen für praktische Aspekte der Gesundheitsversorgung gezogen?
46. Gibt es einen Ausblick auf offene Forschungsfragen? Werden Vorschläge für weitere Untersuchungen unterbreitet?

VI) Zusammenfassung

47. Gibt es eine Zusammenfassung?
48. Gibt die Zusammenfassung das Ziel, die Methoden, die Ergebnisse und die Schlussfolgerungen korrekt wieder?
49. Fehlen in der Zusammenfassung wichtige Aspekte der Untersuchung?
50. Sind Teile der Zusammenfassung nicht durch entsprechende Aussagen im Text belegt?

Sollte die Publikation zu einer oder mehrere dieser Fragen keine befriedigende Antwort geben, ist sie besonders kritisch zu hinterfragen.

Zentrale Bedeutung beim Literaturstudium hat die Bewertung des Studientyps (vgl. Kapitel 7). Nicht alle epidemiologischen Studienformen sind gleichermaßen für die verschiedenen Zielsetzungen einer Studie geeignet (Tabelle 14.1).

Tabelle 14.1: geeignete Studienformen für verschiedene epidemiologische Fragestellungen

Zielsetzung der Untersuchung	geeigneter Studientyp
Untersuchung seltener Krankheiten (z.B. Krebs)	Fall-Kontroll-Studie
Untersuchung seltener Expositionen (z.B. chemische Stoffe)	Kohortenstudie in exponierten Populationen
Untersuchung multipler Expositionen (Bsp. orale Kontrazeptiva und Rauchen)	Fall-Kontroll-Studie, prospektive Studie
Untersuchung multipler Zielereignisse wie Risiken bzgl. verschiedener Krankheiten	Kohortenstudie
Schätzung von Inzidenzraten in exponierten Populationen	Kohortenstudie
Untersuchung von Kovariaten, die über die Zeit variieren	bevorzugt in Kohortenstudien

Quelle: Blettner et al. (2001)

Es gibt schließlich eine Reihe weiterer Kriterien, die auf spezifische Studientypen zugeschnitten sind. Die Listen finden sich allesamt im Internet (vgl. zusätzliche Literatur)

14.2 Ein Beispiel zur Literaturanalyse

An einem Beispiel soll die Anwendung dieser Kriterien veranschaulicht werden. Es handelt sich um die Studie ‚Gesundheitsrelevante Verhaltensweisen und Problembewältigung von Jugendlichen unter den Bedingungen des sozialen Wandels in den neuen Bundesländern‘ von Tomaszewski et al. (1997; Kopie des Artikels im Anhang). Die kritische Durchsicht erbringt – in Stichworten – folgende Ergebnisse.

Nr.	Kriterium	Ergebnisse der Bewertung
Einleitung		
1	theoretische Überlegungen	doppelte Dynamik des Übergangsprozesses, vielfältige Anforderungen der Jugendlichen nach der Wende
2	Ziele der Studie	Überprüfung des Zusammenhangs zwischen Gesundheitsverhalten und Belastungssituation und Selbstkonzept zur Problembewältigung
3	Hypothesen	positive Zukunftssicht und angemessenes Selbstkonzept zur Problembewältigung können protektiv wirken
4	Grundgesamtheit	Jugendliche in Ostdeutschland
Material und Methode		
5,6	Studientyp	standardisierte Querschnitterhebung unter Schülern und Auszubildenden; Befragung zur Situation im Jahr vor der Studie
7	Design angemessen?	Querschnitterhebung für explorative Studien gut geeignet.
8	anderes Studiendesign denkbar?	Für Hypothesentestung sind grundsätzlich Fall-Kontroll- oder prospektive Studie besser geeignet, aber mit höherem Aufwand verbunden
9	Umfang	1.764 Jugendliche, nach Alter und Geschlecht aufgeschlüsselt
10	Vollerhebung / SP	Stichprobe; Größe der Grundgesamtheit nicht erwähnt
11	Art der SP-Ziehung	nicht spezifiziert; Repräsentativität wird nur behauptet, aber nicht belegt
12	Ein- und Ausschlusskriterien	Alter 14 bis 19 Jahre (ohne weitere Begründung; Pubertät begonnen bei Wende?); alle Schulformen; Lehrlinge ausreichend vertreten?
13	Vergleichsgruppe	nein; die Untersuchung war nicht vergleichend angelegt
14	benutzte Software	k.A. (aber wenig kritisch, da keine speziellen Verfahren angewendet)
15	Operationalisierung der Hypothesen	FASP; Konsumverhalten; Erhebung von Belastungssituationen; Tätigkeitsänderung der Eltern; soziales Umfeld nicht erwähnt
16, 17	Erhebungsinstrumente	Fragebogen nicht näher spezifiziert; Verweis auf FSAP
18, 19	Instrumente geeignet	bekannte (vermutlich valide) Instrumente, auch in anderen Jugendstudien eingesetzt; aber nicht belegt
20	Confounder	soziale Schicht über Tätigkeit der Eltern; Alter; Schulform
21	Messungen	Selbstangaben der Teilnehmer; Fragebogen zum Selbstauffüllen; unklar bleibt, wo der Fragebogen ausgefüllt wurde (in der Schule oder zu Haus)
22	Validität, Reliabilität	Angaben zum Alkohol-, Zigaretten- und Medikamentenkonsum valide?
23	statistische Verfahren	nicht erwähnt

Ergebnisse		
24	Ergebnisdarstellung	bleibt deskriptiv, Werte des FSAP nicht anschaulich
25	Ergebnisse schlüssig	Ergebnisse passen in das theoretische Konstrukt, bleiben aber auf deskriptiver Ebene; gering ausgeprägte Unterschiede im FASP von fraglicher Relevanz; Plausibilität der Angaben zum Alkohol-, Tabak- und Medikamentenkonsum?
26	Tabellen, Abbildungen	übersichtlich und verständlich; Legende und Erläuterungen ausreichend
27	statistische Auswertung	keine Tests oder Modelle durchgeführt
28	Hauptresultate	beträchtlicher Konsum von Alkohol, Tabak und Medikamenten; schwach ausgeprägte Assoziation von Gesundheitsverhalten und Belastungen
29	Befunde konsistent	ja; Tabellen, Abbildungen und Text bergen keine abweichenden Ergebnisse; insgesamt aber nur geringe Assoziationen
30	Beziehungsmaß	Korrelationen (vgl. Tabelle 7)
31	Signifikanzniveau	1% (nicht vorab festgelegt) bei Korrelationskoeffizienten; Signifikanz ist nicht gleichbedeutend mit inhaltlicher Relevanz
32	nicht-signif. Ergebnisse?	die meisten Darstellungen bleiben auf deskriptivem Niveau
33	Dosis-Wirkungs-Beziehung	schwach erkennbar ist eine positive Korrelation der Zahl der sozialen Belastungen, Gesundheitsverhalten und Fähigkeit zur Selbstbewältigung
34	Verzerrungen möglich	soziale Einflussfaktoren werden nicht thematisiert; keine Differenzierung nach Art der sozialen Belastungen
35, 36	Adjustierung für Confounder; Bias untersucht	differenzierte Darstellung nach Alter, Geschlecht
Diskussion		
37	Hypothesen untersucht	Korrelationen zwar signifikant, aber sehr gering; Relevanz?
38	Hypothesentestung vs. Exploration	Studie hat vom Ansatz her einen explorativen Charakter
39	Schlussfolgerungen gerechtfertigt?	ja, keine übertriebene Belastung durch Wende; Werte für Gesundheitsverhalten im erwarteten Rahmen? nicht alle in Einleitung genannte Fragestellungen wurden thematisiert (z.B. soziales Umfeld)
40	Bewertung des eigenen Vorgehens	Diskussion, warum die Hypothese nicht bestätigt wurde (Spalte 3, S. 273); keine Diskussion, ob das eigene Ergebnis im erwarteten Rahmen bleibt
41	Diskussion von Verzerrungen	hier nicht relevant
42	Auseinandersetzung mit anderen Studien	Die Ergebnisse von drei Jugendstudien werden knapp angerissen, aber nicht ausführlich reflektiert; vor allem werden keine Vergleichszahlen genannt
43	Vergleichbarkeit mit anderen Ergebnissen	auch andere Studien zeigen präventives Potenzial von Strategien zur Bewältigung von Stresssituationen, nur kurz erwähnt
44	Risiko für Gesamtbevölkerung	hier nicht relevant, Evaluation eines besonderen Ereignisses
45	praktische Aspekte der Studie	Bedeutung gesundheitsfördernder Maßnahmen und Kompetenz beim Umgang mit belastenden Situationen
46	offene Fragestellungen	kurzer Ausblick auf die Notwendigkeit prospektiver Untersuchungen
Zusammenfassung		
47	Zusammenfassung	ja
48, 49	Vollständigkeit	Hypothese nicht spezifiziert, keine einzelnen Ergebnisse
50	nicht belegte Inhalte	nein

Literatur

Epidemiologie – Lehrbücher

- Ahlbom A, Norell (1991): Einführung in die moderne Epidemiologie. München: MMV-Verlag
- Ahrens W, Pigeot I (2005): Handbook of Epidemiology. Berlin, Heidelberg: Springer
- Beaglehole R, Bonita R, Kjellström T (1997): Einführung in die Epidemiologie. Bern: Hans Huber
- Bhopal R (2002): Concepts of epidemiology. Oxford: Oxford University Press
- Fletcher RH, Fletcher SW (2007): Klinische Epidemiologie. Grundlagen und Anwendungen. Bern: Hans Huber, 2. vollst. überarb. dt. Ausgabe
- Friedman GD (1994): Primer of Epidemiology. New York: McGraw-Hill, 4th ed.
- Gerstman BB (1998): Epidemiology kept simple. New York: Wiley
- Gordis L (2001): Epidemiologie. Marburg: Kilian
- Greenberg RS, Daniels SR, Flanders WD, Eley JW, Boring JR (1996): Medical epidemiology. New York: McGraw-Hill, 3rd ed.
- Heinemann L, Sinnecker H (1994): Epidemiologische Arbeitsmethoden. Stuttgart: Gustav Fischer Verlag
- Hennekens CH, Buring JE (1987): Epidemiology in Medicine. Philadelphia: Lippincott, Williams & Wilkins
- Kohlmann T (2000): Grundlagen der Epidemiologie. In: Gostomzyk JG (Hrsg.): Angewandte Sozialmedizin. Landsberg: ecomed (Loseblattsammlung, Kap. I-2.1)
- Jekel JF, Katz DL, Elmore (2001): Epidemiology, biostatistics, and preventive medicine, 2nd ed. Philadelphia: Saunders
- Kleinbaum DG, Kupper LL, Morgenstern H (1982): Epidemiologic Research. New York: Van Nostrand Reinhold
- Kleinbaum DG, Sullivan KM, Barker ND (2007): A Pocket Guide to Epidemiology. New York: Springer
- Kreienbrock L, Schach S (2000): Epidemiologische Methoden. Heidelberg: Spektrum Akademischer Verlag, 3. Aufl.
- Last JM (1988): A dictionary of epidemiology, 2nd ed. New York: Oxford University Press
- Rothman KJ, Greenland S (1998): Modern epidemiology; 2nd ed. Philadelphia: Lippincott-Raven
- Sackett DL (1991): Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston: Little, Brown
- Sackett DL (1999): Evidenzbasierte Medizin: EBM-Umsetzung und –Vermittlung. München-Zuckschwerdt
- Silman AJ (2001): Epidemiological studies: a practical guide. Cambridge: University Press, 2nd ed
- Unwin N, Carr S, Leeson J (1997): An introductory study guide to public health and epidemiology. Buckingham: Open University Press.

Epidemiologie – im Internet

<http://www.pitt.edu/~super1/>

<http://bmj.com/epidem/>

<http://www.poems.msu.edu/InfoMastery/>

zahlreiche Lectures on epidemiology

Coggon D, Rose G, Barker DJP (1997): Epidemiology for the Uninitiated

Einführung in die Evidenzbasierte Medizin

Statistik – Lehr- und Übungsbücher

- Bortz (2004): Statistik für Human- und Sozialwissenschaftler. Berlin: Springer, 6. Aufl.
- Coggon D (2007): Statistik. Einführung für Gesundheitsberufe. Bern: Hans Huber
- Clauss G, Finze F-R, Partzsch L (1995): Statistik für Soziologen, Pädagogen, Psychologen und Mediziner. Bd. 1. Grundlagen. Thun: Verlag Harri Deutsch 2. überarb. Aufl.
- Glantz SA (1998): Biostatistik (incl. PC-Programm). Frankfurt: McGraw Hill
- Harms V (1998): Biomathematik, Statistik und Dokumentation. Kiel: Harms Verlag, 7. neu-bearb. Aufl.
- Hartung J (2002): Statistik. München: Oldenbourg, 13. Aufl.
- Kohlmann T (2000): Grundbegriffe der Statistik. In: Gostomzyk JG (Hrsg.): Angewandte Sozialmedizin. Landsberg: ecomed (Loseblattsammlung, Kap. I-2.2)
- van der Lippe P (1993): Deskriptive Statistik. Stuttgart: G. Fischer Verlag (UTB 1632)
- Ostermann R, Wolf-Ostermann K (2005): Statistik in Sozialer Arbeit und Pflege. München: Oldenbourg. 3. überarb. Auflage
- Sachs L, Hedderich J (2006): Angewandte Statistik. Berlin: Springer, vollst. neu bearb. Aufl.
- Swinscow TDV (1997): Statistics at Square One. London: BMJ-Verlag, 9th ed.
- Werner J (1984): Medizinische Statistik. München und Wien: Urban&Schwarzenberg

Sozialmedizin, Gesundheitswissenschaften, Medizinische Soziologie – Grundlagen

- Brennecke R (Hrsg.) (2004): Sozialmedizin. Bern: Hans Huber
- Buser K, Schneller T, Wildgrube K (2003): Kurzlehrbuch Medizinische Psychologie, Medizinische Soziologie. Stuttgart: Urban&Fischer, 6. Aufl.
- Geyer S (2003): Forschungsmethoden in den Gesundheitswissenschaften. Weinheim: Juventa
- Gostomzyk JG (Hrsg.) (2000): Angewandte Sozialmedizin. Landsberg: ecomed (Loseblattsammlung)
- Gutzwiller F, Jeanneret O (Hrsg.) (1999): Sozial- und Präventivmedizin. Bern: Hans Huber, 2. Auflage
- Kasten E, Sabel B (2005): GK 1 - Medizinische Soziologie, Medizinische Psychologie. Stuttgart: Thieme, 13. Aufl.
- Niehoff JU (1995): Sozialmedizin systematisch: Lorch: Uni-med
- Niehoff JU, Braun B (2003): Handwörterbuch Sozialmedizin und Public Health. Baden-Baden: Nomos Verlagsgesellschaft.
- Hurrelmann K (Hrsg.) (1999): Gesundheitswissenschaften. Berlin, Heidelberg: Springer
- Hurrelmann K, Laaser U (Hrsg.) (1998): Handbuch Gesundheitswissenschaften. Weinheim: Juventa, 2. Auflage
- Schaefer H, Blohmke M (1978): Sozialmedizin. Stuttgart: Thieme, 2. Aufl.
- Schwartz FW, Badura B, Leidl R, Raspe H, Siegrist J (Hrsg.) (2002): Das Public Health Buch. München und Wien: Urban&Schwarzenberg, 2. Aufl.
- Siegrist J (2005): Medizinische Soziologie. München: Urban&Fischer, 6. Aufl.
- Strauss B, Berger U, Troschke J von, Brähler (2004): Lehrbuch Medizinische Psychologie und Medizinische Soziologie. Göttingen: Hogrefe
- Waller H (2002): Sozialmedizin: Grundlagen und Praxis. Stuttgart: Kohlhammer, 5. Aufl.

Demographie - Lehrbücher

- Hauser J (1982): Bevölkerungslehre. Bern, Stuttgart: Haupt
- Müller U (2000): Maßzahlen der Bevölkerungsstatistik. In: Müller U, Nauck B, Diekmann A (Hrsg.): Handbuch der Demographie Bd. 1, Berlin, Heidelberg: Springer S. 1-91
- Esenwein-Rothe I (1992): Einführung in die Demographie: Bevölkerungsstruktur und Bevölkerungsprozess aus der Sicht der Statistik. Wiesbaden: Steiner

Empirische Sozialforschung – Lehrbücher

- Atteslander P (2006): Methoden der empirischen Sozialforschung. Berlin, New York: Walter de Gruyter, 11. Aufl.
- Bortz J, Döring N (2001): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Berlin, Heidelberg: Springer, 3. Aufl.
- Diekmann (200x): Empirische Sozialforschung. Grundlagen, methoden, Anwendungen. Reinbek: rororo, 10. Aufl.
- Flick U (2004): Qualitative Forschung. Reinbek: Rowohlt, 3. Aufl.
- Friedrichs J (1990): Methoden der empirischen Sozialforschung, 14. Auflage. Wiesbaden: Vs Verlag für Sozialwissenschaften
- Holm K (1991): Die Befragung. Tübingen: Francke. 4. Aufl. (UTB 372)
- König R (Hrsg.) (1974): Handbuch der empirischen Sozialforschung in 4 Bänden. Stuttgart: Enke. Bd. 2: Grundlegende Methoden und Techniken der empirischen Sozialforschung, 1. Teil; Bd. 3a: Grundlegende Methoden und Techniken der empirischen Sozialforschung, 2. Teil
- Kromray H. (2002): Empirische Sozialforschung. Opladen: Leske und Budrich
- Lamnek S (2005): Qualitative Sozialforschung. Weinheim: Beltz, 4. Aufl.
- Mayring P (2002): Einführung in die qualitative Sozialforschung. Weinheim: Beltz, 5. Aufl.
- Mayring P, Gläser-Zikuda M (Hrsg.) (2005): Die Praxis der Qualitativen Inhaltsanalyse. Weinheim: Beltz
- Schnell R (2005): Methoden der empirischen Sozialforschung. München, Wien: Oldenbourg

Ergänzende Literatur zu einzelnen Kapiteln

Kapitel 2

- Becker N, Wahrendorf (1998): Krebsatlas der Bundesrepublik Deutschland 1981-1990. Berlin: Springer
- Blohmke M, V. Ferber C, Kisker KP, Schaefer H (1975): Handbuch der Sozialmedizin. Bd 1: Grundlagen und Methoden der Sozialmedizin. Stuttgart: Enke Verlag
- Forschungsverbund DHP (1998): Die deutsche Herz-Kreislauf-Präventionsstudie: Design und Ergebnisse. Bern: Hans Huber
- Janßen C, Borgetto B, Heller G (Hrsg.) (2007): Medizinsoziologische Beiträge zur Versorgungsforschung: Theoretische Ansätze, Methoden und Instrumente sowie ausgewählte Ergebnisse. Weinheim: Juventa
- Krasemann EO, Laaser U, Schach E (1987): Sozialmedizin. Schwerpunkte Rheuma und Krebs. Heidelberg: Springer
- Kreuter HH, Klaes L, Hoffmeister H, Laaser U (1995): Prävention von Herz-Kreislauf-Krankheiten. Ergebnisse und Konsequenzen der Deutschen Herz-Kreislauf-Präventionsstudie. Weinheim: Juventa
- Lichtner S, Pflanz M (1971): Appendectomy in the Federal Republic of Germany: epidemiology and medical care pattern. Medical Care 9: 311-330
- Mosse M, Tugendreich G (1994): Soziale Lage und Gesundheit. Göttingen: Cromm, 3. Aufl. (reprint von 1913)
- Pesch B, Halekoh U, Ranft U, Richter M, Pott F (1994): Atlas zur Krebssterblichkeit in Nordrhein-Westfalen. Düsseldorf
- Pfaff H
- Schipperges H (1994): Rudolf Virchow. Reinbek:rororo
- Semmelweis I Ph (1861): Ätiologie, Begriff und Prophylaxis des Kindbettfiebers (Nachdruck Leipzig 1912).
- Vandenbrouke JP (2001): Changing images of John Snow in the history of epidemiology. Sozial- und Präventivmedizin 46: 288-293
- Informationen über John Snow: <http://www.ph.ucla.edu/epi/snow.html>
- Krebsregister Saarland: <http://www.krebsregister.saarland.de>
- Informationen zum Bundesgesundheitsurvey: <http://www.rki.de>
- Informationen zum Kinder- und Jugendsurvey: <http://www.kiggs.de>

Kapitel 4

- Ahrens W, Bellach B-M, Jöckel K-H (1998): Messung soziodemographischer Merkmale in der Epidemiologie. München: MMV-Verlag
- Latza U, Hoffmann W, teschüren C, Chang-Claude J, Kreuzer M, Schaffrath RA, Kropp S, Stang A, Ahrens W, Lampert T (2005): Erhebung, Quantifizierung und Analyse der Rauchexposition in epidemiologischen Studien. Berlin. Robert-Koch-Institut

Kapitel 5

- Bardehle D, Annuß R, Hermann S, Ziese T, Böhm (2004). Der neue Länderindikatorensetz für die Gesundheitsberichterstattung. Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz 47: 762-770
- Bundesministerium für Familie, Senioren, Frauen und Jugend (Hrsg.) (2001): Bericht zur gesundheitlichen Situation von Frauen in Deutschland. Stuttgart: Kohlhammer

- Forschungsgruppe Gesundheitsberichterstattung (1990): Aufbau einer Gesundheitsberichterstattung. Bestandsaufnahme und Konzeptvorschlag. St. Augustin: Asgard Verlag, 49-69
- Hanesch W, Krause P, Bäcker G. Armut und Ungleichheit in Deutschland. Armutsbericht der Hans-Böckler-Stiftung, des DGB und des Paritätischen Wohlfahrtsverbandes. Reinbek: Rowohlt Taschenbuch Verlag 2000
- Hey M, Maschewsky-Schneider U (Hrsg.) (2006): Kursbuch Versorgungsforschung. München: Med. Wiss. Verlagsanstalt
- Jacob R (2006): Sozial- und Gesundheitsberichterstattung. Hintergründe, Defizite, Möglichkeiten. Frankfurt/M: Peter Lang
- Kuhn J, Wildner M (2006): Gesundheitsdaten verstehen. Bern: Huber-Verlag
- Ministerium für Arbeit, Soziales und Gesundheit Sachsen-Anhalt (Hrsg.) (2003): 4. Landesgesundheitsbericht Sachsen-Anhalt. Magdeburg [download über www.sachsen-anhalt.de]
- Ministerium für Arbeit, Soziales und Gesundheit Sachsen-Anhalt (Hrsg.) (2005): 5. Landesgesundheitsbericht Sachsen-Anhalt. Magdeburg [download über www.sachsen-anhalt.de]
- Pfaff H, Schrappe M, Lauterbach KW, Engelmann U, Halber M (Hrsg.) (2003): Gesundheitsversorgung und Disease Management. Grundlagen und Anwendungen der Versorgungsforschung. Bern: Hans Huber
- Reintjes R, Klein S (2007): Gesundheitsberichterstattung und Surveillance. Bern: Hans Huber
- Robert-Koch-Institut (Hrsg.) (2001): Auszug aus dem Autorenhandbuch zur GBE. Berlin
- Robert-Koch-Institut (Hrsg.) (versch. Jahre): Spezialberichte zur Gesundheitsberichterstattung. Berlin [download über www.rki.de]
- Statistisches Bundesamt (Hrsg.) (1998): Gesundheitsbericht für Deutschland. Stuttgart: Metzler-Poeschel
- Swart E, Ihle P (Hrsg.) (2005): Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden, Perspektiven. Bern: Hans Huber
- Swart E, Heller G (2007): Nutzung und Bedeutung von GKV-Routinedaten. In: Janßen C, Borgetto B, Heller G (Hrsg.): Medizinsoziologische Beiträge zur Versorgungsforschung: Theoretische Ansätze, Methoden und Instrumente sowie ausgewählte Ergebnisse. Weinheim: Juventa, 93-112
- Süß W, Möller H, Trojan A, Fehr R (2004): Integrierte Basis-Berichterstattung für gesündere Städte und Kommunen. Bielefeld: Landesamt für den öffentlichen Gesundheitsdienst (<http://www.loegd.de>)

Kapitel 6

- Statistisches Bundesamt (2006): Bevölkerungsentwicklung Deutschlands bis zum Jahr 2050 – Ergebnisse der 11. koordinierten Bevölkerungsvorausberechnung [download über <http://www.destatis.de>]

Kapitel 9

- Wichmann H-E (1999): Epidemiologie und Datenschutz – Wege zur partnerschaftlichen Zusammenarbeit. Informatik, Biometrie und Epidemiologie in Medizin und Biologie 39: 35-44
- Lauenroth N, Swart E (2004): ‚Perspektivlosigkeit ist meine Krankheit.‘ Subjektive Empfindungen von langzeitarbeitslosen. Das Gesundheitswesen 66:
- Lauenroth N, Swart E (2004): Chancen von Gesundheitsförderungs- und Präventionsstrategien in ABM-Gesellschaften. Prävention 27: 116-119

Kapitel 10

- Willett WC, Brain C, Hennekens CH, Rosner B, Speizer FE (1981): Oral Contraceptives and Risk of Ovarian Cancer. Cancer 48: 1684-1687

Kapitel 11

Bühl A, Zöfel P (2002): SPSS, Version 11. Einführung in die moderne Datenanalyse unter Windows. Bonn: Addison-Wesley

Institut für Sozialmedizin und Gesundheitsökonomie (2002): Unterlagen zur Veranstaltung Sozialmedizin, Epidemiologie und Gesundheitsökonomie (<http://www.med.uni-magdeburg.de/fme/institute/ism/>)

Wilson JM, Jungner (1968). Principles and practice of screening for disease. Genf: Weltgesundheitsorganisation

Kapitel 12

Beaglehole R, Bonita R, Kjellström T. Einführung in die Epidemiologie. Bern: Hans Huber 1997 (Kapitel 5: Kausalität in der Epidemiologie, S. 71-81)

Doll R, Hill AB (1952): A study of the aetiology of carcinoma of the lung. Brit Med J, 2, 1271-1286

Doll R, Hill AB (1964): Mortality in relation to smoking: ten years' observations of british doctors. Brit Med J, 1, 1399-1410 und 1460-1467

Gordis L (2001): Epidemiologie: Marburg: Kilian (Kap. 13: Vor Zusammenhängen und Ursache: Aus statistischen Studien Schlüsse ziehen, S. 219-241)

Hill AB (1965): The environment and disease: association or causation? Proceedings of the Royal Society of Medicine: 58: 295-300

Raspe H (1996): Evidence based medicine: Modischer Unsinn, alter Wein in neuen Schläuchen oder aktuelle Notwendigkeit? Zeitschrift für Ärztliche Fortbildung und Qualitätsmanagement 90: 553-562

Kapitel 13

Hoffmann W, Latza U, Terschüren C (2005): Leitlinien und Empfehlungen zur Sicherung Guter Epidemiologischer Praxis (GEP – überarbeitete Fassung nach Evaluation. Gesundheitswesen 67: 217-225 [download über www.dgepi.de]

Swart E, Ihle P, Geyer S, Grobe T, Hofmann W (2005): Gute Praxis Sekundärdatenanalyse (GPS). Gesundheitswesen 67: 416-421 [download der Revisionsfassung über www.dgsmp.de und FB Epidemiologie/AGENS]

Kapitel 14

Blettner M, Heuer C, Razum O (2001): Critical reading of epidemiological papers. Eur J Public Health 11: 97:101

Oxnan AD (1994): Systematic review: checklists for review articles. Brit Med J 309: 648-651

Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker B, Sipe TA, Thacker SB (2000): Metaanalysis of observational studies in epidemiology. A proposal for reporting. JAMA 283: 2008-2012

Tomaszewski K, Adam I, Hinze L (1997): Gesundheitsrelevante Verhaltensweisen und Problembewältigung von Jugendlichen unter den Bedingungen des sozialen Wandels in den neuen Bundesländern. Sozial- und Präventivmedizin 42: 268-275

Tooth L, Ware R, Brain C, Purdie DM, Dobson A (2005): Quality of reporting of observational longitudinal research. Amer J Epid 161: 280-288

zusätzliche Netzseiten mit Kriterienlisten:

STROBE: Checklist for essentials Items, vol. 3 (www.strobe-statement.org)

CONSORT: Improving the quality of reports of metaanalysis of randomised controlled trials: the QUORUM statement checklist (www.consort-statement.org)

Anhang

Anhang 1:

Swart E Mächler H (2000): Verbessert sich der Gesundheitszustand nach Wiederaufnahme einer ABM-Tätigkeit? Das Gesundheitswesen 62: 335-341

Anhang 2:

Doll R, Hill AB (2004): Mortality in relation to smoking: 50 years' observations of british male doctors. Brit Med J 328: 1519-1527

Anhang 3:

Tomaschewski K, Adam I, Hinze L (1997): Gesundheitsrelevante Verhaltensweisen und Problembewältigung von Jugendlichen unter den Bedingungen des sozialen Wandels in den neuen Bundesländern. Sozial- und Präventivmedizin 42: 268-275

Alle drei Texte in einer gesonderten pdf-Datei zum Download