

**Medizinische Hochschule Hannover
Institut für Biometrie**

**Methodische Grundlagen
klinischer Studien**

H. Hecker
1991

Contents

1	Einführung	0
1.1	Elementare Grundbegriffe deskriptiver Statistik	0
2	Die Auswahl der Beobachtungseinheiten	3
2.1	Totalerhebung oder Stichprobe?	3
2.2	Grundgesamtheit definiert?	4
2.3	Zielpopulation definiert?	4
2.4	Repräsentativität der Untersuchungspopulation gewährleistet?	4
3	Die Auswahl der Merkmale	5
3.1	Die Hauptvariablen (die Variablen "der Fragestellung")	5
3.2	Die Basisvariablen einer Untersuchung	6
4	Die Anlage der Untersuchung	7
4.1	Die Beobachtungsstudie	7
4.1.1	Die Fall-Kontroll-Studie	8
4.1.2	Die Querschnitt-Studie	10
4.1.3	Die Kohorten-Studie	11
4.2	Die experimentelle Studie	11
4.2.1	Grundsätzliches zu experimentellen Studien	11
4.2.2	Besonderheiten bei Studien mit Zufallsdaten ("statistische Studien")	13
4.2.3	Besonderheiten bei Studien am Menschen	16
5	Das Studienprotokoll	18

1. Einführung

Zu Beginn einer Untersuchung sind in der Regel so viele und unterschiedliche Fragen zu klären und Faktoren zu beachten, daß man -insbesondere als Doktorand und damit häufig als Neuling- in Gefahr ist, Wichtiges außer acht zu lassen und damit das angestrebte Ziel gar nicht oder nur unter großem zusätzlichen Aufwand erreichen zu können. Bei Untersuchungen am Menschen -beobachtend oder experimentell- können Nachlässigkeiten und Fehler in der Planung und Durchführung einer Studie darüberhinaus für die Versuchspersonen bzw. Patienten weitreichende Konsequenzen haben, die vom Versuchsleiter und den weiteren Verantwortlichen (durchführende Ärzte, Biometriker) zu vertreten sind.

In der folgenden Übersicht werden daher die wichtigsten Punkte aufgelistet und erläutert, die bei der Planung und Durchführung einer Studie beachtet werden müssen. Die Darstellungen sollten den Leser dazu befähigen, allgemeine Grundsätze der Versuchsplanung und -durchführung kennenzulernen und auf die eigene Studie anzuwenden. Darüberhinaus soll die Bereitschaft gefördert werden, die Planung einer Studie sehr ausführlich zu gestalten, so daß auch solche Besonderheiten, die ganz spezifisch für die eigene Studie sind, frühzeitig erkannt und in den Planungsmaßnahmen berücksichtigt werden.

Zuvor werden aber noch elementare Begriffe der (deskriptiven) Statistik vorgestellt:

1.1. Elementare Grundbegriffe deskriptiver Statistik

Jede statistische Untersuchung beruht letztlich darauf, daß an mehreren *Beobachtungseinheiten* eine oder mehrere *Merkmale* gemessen bzw. aus Akten oder durch Befragung oder ähnliches erhoben werden. Die Ergebnisse der einzelnen Messungen sind die *Merkmalswerte*. Da die Merkmalswerte zwischen den Beobachtungseinheiten variieren, werden die Merkmale meist auch *Variable* genannt.

Beispiel 1:

Merkmale (Variable):	Variable 1:	Durchschnittlicher SO ₂ -Gehalt in Hannover
	Variable 2:	Anzahl gemessener Pseudo-Croup-Fälle in Hannover
Beobachtungseinheiten:		1.Jan. 91., . . . , 31.12.91
		2.Jan. 91
		⋮
		31.12.91

Zusammenfassende Darstellung mit den Merkmalswerten:

Beobachtungs- einheit	Merkmalswerte für	
	Variable1 (SO ₂ -Gehalt)	Variable 2 (Anz. Croup-Fälle)
1.Januar 91	0.34 mmHg	2 Fälle
2.Januar 91	0.34 mmHg	4 Fälle
3.Januar 91	0.34 mmHg	4 Fälle
4.Januar 91	0.34 mmHg	4 Fälle
⋮	⋮	⋮

Beispiel 2:

Merkmale (Variable): Variable 1: Therapie
Variable 2 Ergebnis der Behandlung
Beobachtungseinheiten: Patient A.A.
Patient G.S.
Patient B.S.
⋮

Zusammenfassende Darstellung mit den Merkmalswerten:

Beobachtungs- einheit	Merkmalswerte für	
	Variable1 (Behandlung)	Variable 2 (Ergebnis)
A.A.	BIOVIT	ohne Erfolg
G.S.	Placebo	mit Erfolg
B.S.	Placebo	ohne Erfolg
D.T.	BIOVIT	mit Erfolg
⋮	⋮	⋮

In dieser Form sind die Daten einer Untersuchung äußerlich als *Rechteck* oder Matrix zu erkennen, wobei jede Beobachtungseinheit eine *Zeile* und jede Variable eine *Spalte* dieser Matrix erzeugt. Dateien dieser Struktur werden daher auch als *Rechteckdateien* bezeichnet. Die Gesamtheit aller Meßergebnisse einer Untersuchung oder Teile hieraus werden in der deskriptiven Statistik in übersichtlicher Form und meistens komprimiert dargestellt.

In Beispiel 2 etwa würde es für den Vergleich der Behandlungen BIOVIT und PLACEBO nicht interessant sein zu wissen, *welche* der PLACEBO- bzw. BIOVIT-Patienten erfolgreich behandelt wurde; vielmehr reicht es zu wissen, *wieviele* der Patienten jeweils in den beiden Gruppen *mit* und *wieviele ohne* Therapieerfolg waren.

Die gebräuchlichsten Methoden solcher Komprimierung der Information sind die Bildung von *Kenngrößen eindimensionaler Verteilungen* von Variablen wie z.B. *Mittelwert, Standardabweichung, Minimum und Maximum* bei quantitativen Variablen (wie z.B. SO₂-Gehalt, systolischer Blutdruck etc.) sowie die Berechnung von *Prozentzahlen* oder *relativen Häufigkeiten* bei *qualitativen* Variablen (wie z.B. die Berechnung der Prozentzahl der untersuchten Patienten mit dem Therapieergebnis "mit Erfolg").

Bei gleichzeitiger Betrachtung der Werte jeweils *zweier* Variablen der Beobachtungseinheiten werden die Ergebnisse übersichtlich in Form von *Streudiagrammen* dargestellt, wenn beide Merkmale quantitativ sind wie im Beispiel 1, oder als *Kreuztabelle*, wenn beide Merkmale qualitativ sind wie in Beispiel 2.

Die Auswahl sowohl der Merkmale als auch der Art der Aufbereitung der Daten für die übersichtliche, komprimierte Darstellung hängt davon ab, *mit welcher Fragestellung* die Daten erhoben wurden und somit davon, *welche Art von Erkenntnissen aus der Untersuchung gezogen werden sollten*.

In Beispiel 1 etwa könnte das Ziel der Untersuchung gewesen sein, herauszufinden, ob -und wenn ja: welcher- Zusammenhang zwischen der Luftverunreinigung, gemessen als SO₂-Konzentration, und dem Risiko für die Entstehung von Pseudo-Croup-Anfällen, erfaßt über die täglichen Meldezahlen, besteht.

Oder es soll untersucht werden, ob der SO₂-Gehalt über das ganze Jahr hin bis auf Zufallsabweichungen konstant ist oder ob es typische jahreszeitliche Schwankungen gibt.

Vielleicht sollte aber auch die Frage beantwortet werden, ob sich jahreszeitlich bedingte Schwankungen im SO₂-Gehalt der Luft mit einer zeitlichen Verschiebung in der Anzahl gemeldeter Pseudo-Croup-Fälle wiederfinden.

Schließlich kann hinter der Untersuchung aber auch die Hypothese stehen, daß ein Zusammenhang zwischen SO₂ und Pseudo-Croup jedenfalls nicht so eng sein kann, daß er im Bereich relativ niedriger SO₂-Konzentration mit wenigen Spitzenwerten zu entdecken wäre.

Wir gehen nun in den folgenden Abschnitten die einzelnen Punkte der Versuchsplanung und -durchführung auf der Basis der daraus insgesamt resultierenden elementaren Grundsätze für statistische Analysen durch:

GRUNDSÄTZE statistischer Analysen

1. Jede empirische statistische Untersuchung basiert letztlich auf den Daten einer *Rechteckdatei*, in welcher jede Zeile eine *Beobachtungseinheit* und jede Spalte eine *Variable* oder ein *Merkmal* repräsentiert.
2. Die empirische Basis für eine Auswertung ist somit vollständig dadurch bestimmt,
 - - *welche Beobachtungseinheiten* für die Untersuchung ausgewählt wurden,
 - - *welche Merkmale* zur Messung an den Beobachtungseinheiten ausgesucht wurden und

- - nach welchem Plan und nach welcher Anordnung die Untersuchung (einschließlich Behandlungs- schema und Datenerhebung) die Messungen *durchgeführt* wurden. (Insbesondere auf diese drei Punkte **B**eobachtungseinheit, **M**erkmal und **D**urchführung wird in den nachfolgenden Abschnitten ausführlich eingegangen.)
3. Durch verschiedene Methoden der deskriptiven Statistik werden unterschiedliche Aspekte der Gesamtheit der zugrundeliegenden Merkmalswerte in komprimierter Form dargestellt.

2. Die Auswahl der Beobachtungseinheiten

In der Phase der Planung eines Versuchs steht ganz am Anfang die *Fragestellung* bzw. der *Problembereich*, zu dem die empirische Untersuchung einen *Erkenntnisgewinn* liefern soll. Die weitere Planung verläuft umso einfacher, je klarer und präziser die Fragestellung vorliegt. Tatsächlich ist die Festlegung aller Details, wie sie im Folgenden angesprochen wird, weitgehend mit der *Präzisierung der Fragestellung* identisch, denn jede Festlegung bezüglich einer der drei genannten Punkte (Beobachtungseinheit, Merkmal, Durchführung) legt auch die Aussagemöglichkeit der Daten in bestimmter Weise fest. Dieses wird deutlich, wenn wir die Punkte im einzelnen durchgehen.

2.1. Totalerhebung oder Stichprobe?

In einigen Studien ist die Gesamtheit aller Beobachtungseinheiten, die für die Untersuchung in Frage kämen, überschaubar und von der Anzahl her so begrenzt, daß die Möglichkeit besteht, diese total zu erfassen (z.B.: Befragung aller niedergelassenen Zahnärzte in Hamburg über die Anwendung der Röntgenverordnung). In anderen Fällen kann man für eine Totalerhebung auf bereits existierende Dokumentationssysteme zurückgreifen (z.B.: Klinische Dokumentation aller Geburten in einem Jahr in den Kliniken von München; Untersuchung aller Schulanfänger eines Jahrganges in Hannover).

Eine Totalerhebung hat offensichtlich zwei Vorteile:

1. Für die zu untersuchende Fragestellung wird das Optimum an Information bereitgestellt (so weit es die Beobachtungseinheiten betrifft).
2. Probleme bei der Extrapolation auf die *Grundgesamtheit* entfallen, *sofern man unter der Grundgesamtheit die Gesamtheit der in der Totalerhebung erfaßten Fälle versteht* (die in der Regel durch einen festen Zeitraum, eine festgelegte Region und durch Ein- und Ausschlußkriterien definiert ist, die sich an der Fragestellung orientieren).

Dieser zweite Punkt kommt allerdings nur dann zum Tragen, wenn die Fragestellung ausdrücklich auf die Untersuchung eben dieser Fälle begrenzt ist, ohne daß daraus Rückschlüsse oder Vorhersagen auf die Verhältnisse in anderen Regionen oder zu anderen Zeiten gezogen werden sollen. Die große Mehrzahl der klinischen Untersuchungen zielt aber nicht auf die Beschreibung eines solchen –in dieser Art jeweils einmaligen– Kollektivs; vielmehr stehen medizinische Zusammenhänge und physiologische Vorgänge im Mittelpunkt, die sich "im Prinzip" beliebig oft wiederholen können. In diesem Sinne stellen auch die Beobachtungen an einer Totalerhebung nur einen endlichen Ausschnitt aus der gesamten Wirklichkeit dar, so daß das Problem

der Extrapolation erhalten bleibt: Inferenzstatistische Methoden, die auf der Anwendung der Wahrscheinlichkeitsrechnung beruhen, also z.B. die Durchführung statistischer Tests oder die Bildung von Konfidenzbereichen, sind in diesem Fall also auch bei Totalerhebungen anzuwenden.

2.2. Grundgesamtheit definiert?

Unter der *Grundgesamtheit* versteht man die *Gesamtheit der Rahmenbedingungen, unter denen die Aussagen der Studie allgemeine Gültigkeit haben sollen*.

Diese Rahmenbedingungen beziehen sich einerseits auf die Charakterisierung der *Beobachtungseinheiten*, also z.B. der Patienten und eine genaue Beschreibung der Krankheit, deren Verlauf oder Behandlung untersucht wird; zum anderen auf alle "äußeren" *Rahmenbedingungen* wie z.B. die genaue Beschreibung der anzuwendenden Therapie und der Meßmethoden sowie der Zeitintervalle zwischen den Einzelmessungen.

Durch den ersten Teil, die Charakterisierung der *Beobachtungseinheiten*, ist die *Zielpopulation* festgelegt, also z.B. die Gesamtheit derjenigen *Patienten*, für die eine Aussage der Untersuchung allgemein Gültigkeit haben soll. Hierauf wird im nächsten Unterpunkt näher eingegangen.

Die nähere Charakterisierung der *äußeren Rahmenbedingungen* geschieht im einzelnen durch die genaue Beschreibung der *Merkmale*, die in die Untersuchung eingehen, der *Meßmethoden* sowie der gesamten *Versuchsordnung*. Diese Punkte werden unter 4. und 5. behandelt.

2.3. Zielpopulation definiert?

Die Festlegung der oben definierten *Zielpopulation* der Untersuchung erfolgt über die Angabe von *Ein- und Ausschlußkriterien*. In den *Einschlußkriterien* werden diejenigen Bedingungen aufgelistet, die von einer (potentiellen) Beobachtungseinheit (z.B. Patient) alle erfüllt sein müssen, damit diese für die Aufnahme in die Studie in Frage kommt. Ist aber eines der *Ausschlußkriterien* erfüllt, so wird eine Aufnahme in die Studie ausgeschlossen.

Handelt es sich bei den Beobachtungseinheiten um Patienten, so ist die Zielpopulation also durch die Gesamtheit derjenigen Patienten festgelegt, die alle Einschlußkriterien und keines der Ausschlußkriterien erfüllen. Dabei wird aber in der Regel an *zukünftige* Patienten ebenso gedacht wie an Patienten der gegenwärtigen Bevölkerung und zum gegenwärtigen Zeitpunkt. Sie ist in diesem Sinne auch nicht begrenzt.

Die Zielpopulation ist in dieser Weise "abstrakt" definiert und unabhängig von aktuell tatsächlich vorhandenen Beobachtungseinheiten. Die *Untersuchungspopulation* einer Studie hingegen setzt sich aus allen Beobachtungseinheiten zusammen, die zur Messung und Erhebung der Daten tatsächlich in die Studie aufgenommen wurden.

2.4. Repräsentativität der Untersuchungspopulation gewährleistet?

Nach den vorher beschriebenen Definitionen stellt die Untersuchungspopulation einer Studie einen Ausschnitt oder –wenn man daran denkt, wie abstrakt die Zielpopulation definiert ist– eine gewisse, endliche "Realisation" aus der Zielpopulation dar. Da die Ergebnisse aus einer Studie aber dazu dienen sollen, Erkenntnisse über die Zielpopulation insgesamt zu gewinnen,

sollte die *Studienpopulation*, anhand derer die Ergebnisse gewonnen werden, "*typisch*" sein für die *Zielpopulation*. Die genauere Formulierung dieser Forderung lautet:

Jeder Fall (=Beobachtungseinheit) der zur Zielpopulation gehört soll die gleiche Wahrscheinlichkeit haben in die Studie (d.h. in die Untersuchungspopulation) aufgenommen zu werden. Dieses ist durch geeignete Stichprobenverfahren zu realisieren.

Dieses ist es auch, was eigentlich unter der Forderung nach der "*Repräsentativität*" der Stichprobe zu verstehen ist. Häufig wird aber auch noch ein zweiter Punkt hierunter subsumiert: die Forderung nach einer ausreichenden *Größe* der Stichprobe. Beide Forderungen zusammen zielen auf ganz unterschiedliche Aspekte einer Stichprobe:

Ist eine Stichprobe *repräsentativ für die Zielpopulation*, so unterscheiden sich die Ergebnisse der Untersuchung höchstens *zufällig*, nicht aber *systematisch* von den tatsächlichen Verhältnissen in der Zielpopulation.

Die (zufallsbedingten) Unterschiede zwischen den Ergebnissen aus der Stichprobe und den tatsächlichen Verhältnissen in der Zielpopulation sind (Repräsentativität der Stichprobe vorausgesetzt) im Durchschnitt *um so kleiner*, je *größer der Stichprobenumfang*.

Genauere Kriterien zur Berechnung eines Stichprobenumfangs, der für die Erreichung einer angestrebten Genauigkeit bzw. Sicherheit von Ergebnissen benötigt wird, setzen Kenntnisse über die Konstruktion von Konfidenzbereichen und statistischen Tests voraus. (Eine Übersicht über entsprechende Literatur und Computerprogramme findet man in Band 28, Heft 2 (1997) der Zeitschrift "Informatik, Biometrie und Epidemiologie", Gustav Fischer Verlag)

3. Die Auswahl der Merkmale

Durch die Auswahl der Variablen (Merkmale), die in einer Studie erhoben werden, muß zunächst einmal sichergestellt werden, daß alle Fragestellungen, die ihr zugrunde liegen, dadurch vollständig abgedeckt sind. Darüberhinaus werden üblicherweise eine Reihe weiterer Variablen erhoben, die in der Fragestellung nicht unmittelbar erscheinen. Es ist nützlich, sich über die unterschiedlichen Rollen der einzelnen Variablen schon in der Planungsphase im Klaren zu sein.

3.1. Die Hauptvariablen (die Variablen "der Fragestellung")

Zu den *Hauptvariablen* sind alle die Merkmale zu rechnen, die in der Fragestellung unmittelbar erscheinen. Spätestens bei der Festlegung dieser Variablen ist man daher gezwungen, die Fragestellung zu präzisieren.

Beispiel: Soll der Zusammenhang zwischen Säuglingssterblichkeit und dem Vor-sorgeverhalten der Mutter untersucht werden, so ist zu präzisieren: die "Säuglingssterblichkeit" (werden Totgeburten dazugerechnet, Sterblichkeit unter der Geburt, innerhalb

der ersten Woche oder des ersten Jahres?) und das "Vorsorgeverhalten" (Zeit bis zum ersten Arztbesuch, Anzahl Arztbesuche im ersten Drittel der Schwangerschaft oder insgesamt, Teilnahme an geburtsvorbereitenden Kursen etc.?)

Dabei kann es deutlich werden, daß es zur Charakterisierung eines "Begriffes" (z.B. "Vorsorgeverhalten") *mehrerer* Merkmale bedarf, die jeweils unterschiedliche Aspekte dieses Begriffes abdecken. Häufig werden auch mehrere einzelne Merkmale (in Fragebögen oft "Items" genannt) zu einem Gesamt-"Score" oder zu einer Klassifizierung zusammengefaßt (z.B. die Einhorn-Klassifikation zur Einteilung der Metastasierung in die drei Kategorien: minimal, moderate, advanced disease). Es empfiehlt sich dann meistens, nicht (nur) den Score bzw. die Klassifikation in die Variablenliste aufzunehmen sondern (auch) die ihnen zugrundeliegenden Einzelbefunde. Meistens ist es in solchen Fällen im übrigen weniger fehleranfällig und auch weniger aufwendig, die Berechnung des Scores bzw. die Klassifikation erst für die Datenaufbereitung vorzusehen und nicht schon in die Liste der direkt zu erfassenden Variablen aufzunehmen.

3.2. Die Basisvariablen einer Untersuchung

Außer den Variablen, die direkt die Fragestellung betreffen, sind in einer Untersuchung üblicherweise noch eine Reihe weiterer Merkmale zu erheben. Hierfür gibt es zwei Gründe:

1. Die Verteilung der Variablen, die in der Studie untersucht werden sollen, oder die in Frage stehenden Zusammenhänge werden von weiteren, oft bereits bekannten Faktoren beeinflusst, ohne daß diese Beeinflussung selber Gegenstand der Untersuchung sein soll.

Beispiel: Es soll die Wirkung eines Medikamentes auf die Hemmung der Progression der Koronarsklerose untersucht werden. Es ist bekannt, daß die Progression durch das Rauchverhalten des Patienten beeinflusst wird.

Durch die Erfassung solcher Faktoren als Variable in der Studie erhält man die Möglichkeit, diese als *Kovariable* in die Auswertung einzubeziehen und etwaige Effekte der zu untersuchenden Variablen dadurch um so deutlicher herauszuarbeiten.

2. Jede Untersuchung sollte eine ausführliche Beschreibung der *Studienpopulation* enthalten. Hierzu ist die Erhebung aller für den Problembereich wichtigen Variablen nötig. Deren Häufigkeitsverteilungen stellt die Basis für eine solche Populationsbeschreibung dar.

Eine solche Populationsbeschreibung ist notwendig, obwohl durch die Ein- und Ausschlußkriterien die Untersuchungspopulation bereits eingegrenzt ist. Merkmale, die durch diese Kriterien nicht betroffen oder nur eingegrenzt sind (z.B. Alter zwischen 20 und 65 Jahre), können innerhalb dieser Grenzen bzw. ganz beliebig variieren. Dies kann dazu führen, daß verschiedene Studien mit derselben Fragestellung auch bei gleichen Ein- und Ausschlußkriterien nicht vergleichbar sind, weil sie sich in der Verteilung wesentlicher Variablen deutlich unterscheiden. (Beispiel: In Studie I sind überwiegend leichte, in Studie II überwiegend schwere Fälle einer Krankheit eingeschlossen). Durch eine ausführliche Beschreibung der Verteilung

solcher *Basisvariablen* wird die Grundlage für die Interpretation einer Studie und für den Vergleich mehrerer Studien mit gleicher oder ähnlicher Fragestellung gelegt.

Zu den Basisvariablen gehören bei Patienten-Studien in jedem Fall *Alter* und *Geschlecht* sowie *Anamnese* und *Eingangsdagnostik* der Patienten, für die Überwachung der Versuchsdurchführung zusätzlich das jeweilige *Datum aller Untersuchungen und Behandlungen*. Unbedingt ist bei jeder Studie für eine *eindeutige Identifizierung jeder Beobachtungseinheit* erforderlich.

4. Die Anlage der Untersuchung

Ganz wesentlich hängt die Aussagekraft einer Studie von der *Anlage der Untersuchung* ab. Dies gilt insbesondere dann, wenn –wie in der überwiegenden Mehrzahl der medizinischen Untersuchungen– die Fragestellung vom Typ

”Welchen Einfluß hat X auf Y”?

ist. In solchen Fällen wird X als *unabhängige* oder *prognostische* Variable bezeichnet (obwohl noch in Frage stehen kann, ob sie überhaupt einen Einfluß oder ”prognostischen” Wert hat) und Y als *abhängige* Variable oder *Zielgröße*. Unterschiede in der *Aussagekraft* von Studien äußern sich in diesem Fall darin, daß ein etwaiger statistisch gesicherter Zusammenhang zwischen X und Y als kausale Beziehung interpretiert werden kann oder nicht.

Beispiel: Wenn ein Zusammenhang zwischen perinataler Mortalität und der Nutzung der Vorsorgeuntersuchungen festgestellt wird, so muß die geringere Nutzung nicht ohne weiteres als Ursache für die erhöhte Mortalität interpretiert werden; vielmehr ist zu berücksichtigen, daß gerade Frauen mit höheren anamnestischen Risiken die Vorsorgeuntersuchungen weniger nutzen und der Zusammenhang daher auch (teilweise) durch die *Drittvariable* ”Anamnestische Belastung” erklärt werden kann.

Wir unterscheiden im wesentlichen zwei Studienformen, die sich in Hinblick auf die Aussagekraft in dieser Hinsicht grundsätzlich unterscheiden: die *Beobachtungsstudie* und die *experimentelle Studie*.

4.1. Die Beobachtungsstudie

Eine *Beobachtungsstudie* ist dadurch charakterisiert, daß -außer durch das Datenerheben und etwaige damit verbundene Untersuchungen oder Fragebogenaktionen- die Beobachtungseinheiten (z.B. Patienten oder Bevölkerungsgruppen) keine zusätzliche Beeinflussung oder Behandlung im Rahmen der Studie erhalten.

Bei Beobachtungsstudien liegt grundsätzlich das oben genannte Problem vor, wonach durch solche Studien festgestellte Beziehungen zwischen Variablen *nicht als kausale Beziehung* interpretiert werden können.

Man kann versuchen, das beschriebene Problem des Einflusses von *Drittvariablen* dadurch abzumildern, daß man

- nachträglich Untergruppen bildet, in denen die bekannten oder vermuteten einflußreichen Drittvariablen konstant sind (dies kann schnell zu sehr kleinen Untergruppen führen, die wegen der kleinen Fallzahlen dann keine gesicherten Aussagen mehr ermöglichen),

- - in multivariaten Auswertungen die bekannten oder vermuteten einflußreichen Drittvariablen mit in die Auswertung einbezieht, oder
- bereits durch die Art der Stichprobenerhebung den Einfluß von (ausgewählten) Drittvariablen auf die Zielgröße eliminiert

Zum letzten dieser drei Ansätze gehört insbesondere die *Paarbildung* ("matched pairs") in *Fall-Kontroll-Studien*.

4.1.1. Die Fall-Kontroll-Studie

Eine *Fall-Kontroll-Studie* (ob mit oder ohne Paarbildung) kann immer dann zur Anwendung kommen, wenn die Zielgröße *binär* ist, d.h. nur zwei mögliche Ausprägungen hat (wie z.B. "Komplikation während der Operation" mit den Ausprägungen "ja" und "nein"). Die Zielpopulation wird dann zunächst in die Population der "Fälle" (Zielgröße hat den Wert "ja") und der "Kontrollen" (Zielgröße = "nein") unterteilt. Man versucht dann herauszufinden, in welcher Hinsicht sich die Fälle von den Kontrollen unterscheiden, um auf diesem Wege *indirekt* Hinweise auf Zusammenhänge zwischen prognostischen Faktoren und dem Auftreten des untersuchten Ereignisses zu erhalten.

Ein naheliegenderes Vorgehen besteht dann darin, aus jeder der beiden Teilpopulationen (unabhängig voneinander) eine Totalerhebung oder eine repräsentative Stichprobe zu ziehen und anschließend die Häufigkeitsverteilungen der dabei erhobenen Variablen zwischen den beiden Gruppen miteinander zu vergleichen. Hierbei bleibt aber das Problem des Einflusses von Drittvariablen bestehen. Alternativ dazu wird daher gelegentlich so vorgegangen:

Fall-Kontroll-Studie mit Paarbildung (matched pairs):

Zu jedem "Fall" (zu jeder Beobachtungseinheit aus der Stichprobe bzw. aus der Totalerhebung der Population der "Fälle") wähle man aus der Kontrollgruppe (aus der Population der "Kontrollen") einen "statistischen Zwilling", d.h. eine Beobachtungseinheit, *die dem "Fall" in Hinblick auf die ausgesuchten Drittvariablen gleich oder ähnlich ist (matched pair)*. Gibt es dann deutliche Unterschiede zwischen beiden Gruppen bezüglich der Häufigkeitsverteilung einer prognostischen Variablen, so ist der dadurch angezeigte Zusammenhang mit der Zielgröße *jedenfalls nicht durch Einflüsse derjenigen Drittvariablen zurückzuführen, die bei der Paarbildung berücksichtigt wurden*.

Für Studien, die als matched-pair-Studien angelegt werden, gilt grundsätzlich:

- Durch die Paarbildung über die Gleichheit von Drittvariablen sind die jeweils zwei Subjekte aus Fall- und Kontrollgruppe nicht mehr im statistischen Sinn unabhängig. Für die statistische Analyse ist daher jeweils ein *Paar* als eine *Beobachtungseinheit* anzusehen.
- Etwaige Einflüsse derjenigen Drittvariablen, nach denen "gematched" wurde, auf die Zielgröße, können mit Hilfe der so erhobenen Daten selber *nicht untersucht werden*.
- Die Möglichkeit zur Paarbildung ist außer durch die Anzahl der verfügbaren Beobachtungseinheiten aus der Kontrollgruppe *wesentlich durch die Anzahl der Drittvariablen und die Anzahl der jeweils zugehörigen möglichen Werte (Kategorien) begrenzt*.

Diese Punkte sind unbedingt zu bedenken, wenn man in der Planungsphase einer Studie einen matched-pair-Ansatz erwägt.

Für Fall-Kontroll-Studien allgemein gilt, daß sie in zwei Situationen eingesetzt werden sollten:

1. Die *Zeit* zwischen dem Beginn einer eventuellen Wirkung von X auf die binäre Zielgröße Y und der Manifestation dieser Wirkung (Zeit zwischen Beginn der "Exposition" und dem Ergebnis) übersteigt die Dauer, die für eine Studie angesetzt werden kann (Beispiel: Zusammenhang zwischen erhöhtem Cholesterinwerten und dem Eintreten eines Herzinfarktes).
2. Das zu untersuchende "Ereignis" (Zielgröße = "ja") ist so *selten*, daß in einer vertretbar großen Stichprobe aus der Zielpopulation nur so wenige "Fälle" zu erwarten sind, daß diese für eine statistische Analyse nicht verwertbar sind (Beispiel: Etwa 5 - 12 perinatale Sterbefälle auf 1000 Geburten).

Insbesondere in der zweiten dieser beiden Situationen ist es oft am günstigsten, in der zur Verfügung stehenden *Fallgruppe* eine Totalerhebung durchzuführen, und aus der *Kontrollgruppe* eine *repräsentative Stichprobe* zu ziehen oder nach dem matched-pair Verfahren vorzugehen. Die "Fälle" sind dadurch in der Studie stark überrepräsentiert, so daß sich Risiken für das Eintreten des Ereignisses aus diesen Zahlen allein nicht berechnen lassen. (Eine Maßzahl, die den Zusammenhang zwischen X und Y aber auch in dieser Situation beschreibt ist die "relative Quote" oder "odds ratio")

Das Vorgehen in einer Fall-Kontroll-Studie ist in zweierlei Hinsicht "*retrospektiv*":

1. Die *Richtung der Fragestellung* wird umgekehrt: Das *Ergebnis Y* des zu untersuchenden Prozesses (X wirkt auf Y) wird als *gegeben* betrachtet, und es wird *in Abhängigkeit von diesem Ergebnis die Verteilung der prognostischen Variablen X* untersucht.
2. Die *Daten* werden *retrospektiv erhoben*. Dieses ist dadurch bedingt, daß erst nach Eintreten des Ergebnisses die Stichprobenerhebung und damit auch die Erhebung der Daten erfolgen kann.
Um diesen Aspekt gesondert zu kennzeichnen, wird hierfür zunehmend die Bezeichnung "*retroektiv*" verwendet, womit also das "retrospektive Datensammeln" beschrieben wird

Bei retrolektiven Studien ist man im wesentlichen auf solche Daten angewiesen, die im Rahmen üblicher Untersuchungen erhoben wurden und die z.B. in Krankenakten festgehalten sind, oder auf nachträgliche Befragungen. Beide Möglichkeiten weisen gewisse Nachteile auf: Daten aus früheren, nicht für die Studie angelegten Dokumentationen, weisen oft nicht den gewünschten Standard und die nötige Vollständigkeit auf. Bei nachträglichen Befragungen ist man auf *Erinnerungen* angewiesen, die lückenhaft sein können oder gar durch die inzwischen eingetretenen Ereignisse, insbesondere durch solche, die mit dem Untersuchungsgegenstand selbst zu tun haben, verändert sind. (Beispiel: In einer Perinatal-Studie kann die Auffassung der Mutter darüber, ob das Kind erwünscht war, durch die mit der Geburt erlebten Erfahrungen eine Veränderung erfahren, die von ihr selber im nachhinein nicht mehr gesehen wird.)

4.1.2. Die Querschnitt-Studie .

Ein Kennzeichen der Fall-Kontroll-Studie lag darin, daß die Zielpopulation in zwei Gruppen (oft auch "Strata" genannt) eingeteilt ("stratifiziert") wurde und die Stichprobenziehung sich dann mit den üblichen Regeln an diesen beiden Teilpopulationen (bzw. beim matched-pair-Ansatz primär an der Population der "Fälle") orientierte.

Eine *Querschnitt-Studie* ist nun dadurch charakterisiert, daß eine solche Stratifizierung *nicht* stattfindet und also die Repräsentativität der Stichprobe für die gesamte Zielpopulation (und nicht nur getrennt für die einzelnen Teilpopulationen) gefordert wird. Weiterhin ist mit dem Begriff der Querschnitt-Studie die Vorstellung verbunden, daß es sich hierbei um eine *einmalige* Datenerhebung handelt, jede Beobachtungseinheit also nur zu *einem* Zeitpunkt (bzw. engeren Zeitraum) untersucht wird, und nicht der Verlauf eines Prozess oder einer Entwicklung mit entsprechenden Wiederholungsmessungen erfaßt werden soll. **Beispiele** von Querschnitt-Studien:

1. Perinatal-Studie auf der Basis der Dokumentation aller oder eines Anteils aller Klinikgeburten in einem Jahr und in einer Region,
2. Untersuchung der bakteriellen Resistenz von Keimen gegen verschiedene Antibiotika durch Erfassung sämtlicher Antibiogramme in allen oder in ausgewählten Labors einer Region in einem Jahr,
3. Beschreibung und nähere Untersuchung des Verlaufs von Herztransplantationen in einer Klinik in definiertem Zeitraum,
4. Untersuchung des Gesundheitszustandes aller 50-jährigen Hannoveraner

Die Datenerfassung in einer Querschnitt-Studie kann *prospektiv* oder *retrospektiv* sein, d.h. die Studie ist "*prolektiv*" oder "*retrolektiv*", und es kann innerhalb einer Studie auch beide Formen der Datenerfassung geben. *Retrolektiv* ist die Studie nach der obigen Definition dann, wenn sie sich auf Daten stützt, die an den Beobachtungseinheiten *unabhängig von der Studie und vor der jeweiligen Aufnahme in die Studie* gemessen wurden, oder wenn über Befragung oder andere Techniken Informationen aus der Vergangenheit zur Datenerhebung benutzt werden. In *prolektiven* Studien dagegen werden die Daten nach einem Schema erhoben, das *vor* der eigentlichen Messung bzw. Erhebung *im Studienprotokoll festgelegt* ist und bei dem die Messung *den zur Zeit des Messung aktuellen Zustand* erfaßt.

Aus dieser Definition geht auch hervor, daß man bei Daten aus *prolektiven* Studien (bzw. aus den prolektiven Teilen einer Studie) im allgemeinen eine *höhere Qualität* erwarten darf. Damit sind *weniger fehlende Angaben* gemeint, eine *bessere Standardisierung des Meßvorganges*, größere *Präzision* der Messungen (d.h. kleinere Meßfehlerstreuungen) und größere *Validität* (d.h. die Auswahl eines Meßverfahrens, das auch tatsächlich das intendierte Merkmal erfaßt und nicht ein damit allenfalls statistisch korrelierendes Merkmal; extremes Beispiel: Klassifizierung der koronaren Herzkrankheit durch Koronarangiographie statt durch Befragung des Patienten über Herzbeschwerden).

4.1.3. Die Kohorten-Studie

Eine *Kohorten-Studie* ist eine Beobachtungsstudie, in welcher die *Zielgrößen mit ihren Veränderungen* sowohl *im Zeitverlauf* als auch *in Abhängigkeit von den prognostischen Variablen* (und deren Veränderungen) untersucht werden.

Die Studienpopulation ist nach Möglichkeit eine repräsentative Stichprobe der Zielpopulation. Stratifizierungen nach Werten der *Zielgrößen* sind *nicht erlaubt* (und i.d. Regel auch nicht möglich, da das Ergebnis, der Wert der Zielgröße, erst zukünftig, also erst nach der Stichprobenziehung, bekannt sein wird; eine gewisse Ausnahme bilden hier Kohortenstudien mit zurückverlegtem Ausgang). Es ist aber möglich, nach den Werten von *prognostischen Variablen Strata zu bilden* (z.B. die Zielpopulation nach Rauchern und Nichtrauchern zu trennen) und innerhalb jedes Stratums eine repräsentative Stichprobe zu bilden. Dadurch können Untergruppen, die in der Zielpopulation nur einen kleinen Teil ausmachen, in der Studienpopulation auf gleiche Fallzahlen gebracht werden. Dies ist dann sinnvoll, wenn die Genauigkeit der statistischen Aussagen für alle Untergruppen auf ein gleichmäßig hohes Niveau gebracht werden soll. Aussagen, die sich auf die Zielpopulation als Ganzes beziehen, sind bei diesem Vorgehen dann aber nicht mehr möglich, wohl aber getrennte Aussagen über die einzelnen Strata und der direkte Vergleich dieser Strata *in Bezug auf die Zielgröße* (beispielsweise die 1-Jahres-Überlebenswahrscheinlichkeit nach Infarkt bei Frauen (20% aller Infarktpatienten) und Männern sowie ihr Vergleich).

Die *Richtung der Fragestellung* einer Kohortenstudie ist nach diesen Ausführungen in jedem Fall *prospektiv*, d.h. in Abhängigkeit von den *gegebenen Werten der prognostischen Variablen* wird die *Verteilung der Zielgröße* untersucht.

Kohortenstudien sind außerdem stets *prolektiv*, es sei denn, daß der Ausgang der Studie zurückverlegt ist.

Beispiel für eine Kohortenstudie mit zurückverlegtem Ausgang: Untersuchung des Verlaufs der instabilen Angina: Von allen Patienten, die in den vergangenen drei Jahren mit instabiler Angina in die Uni-Klinik aufgenommen wurden, werden die für die Fragestellung benötigten Daten aus den Krankenakten in einen Dokumentationsbogen übertragen; zur Untersuchung des weiteren Verlaufs wird zu Beginn der Studie und in zwei aufeinanderfolgenden Jahren jeweils eine Nachbefragung durchgeführt.

4.2. Die experimentelle Studie .

4.2.1. Grundsätzliches zu experimentellen Studien.

Aus den vorhergehenden Ausführungen geht hervor, daß in einer *Beobachtungsstudie* an den Beobachtungseinheiten im Rahmen der Studie nicht mehr als folgende "Aktionen" vorgenommen werden:

- Auswahl aus der Zielpopulation in die Studienpopulation unter Beachtung von Ein- und Ausschlußkriterien und Stichprobenplan,
- - Messung der für die Studie vorgesehenen Merkmale (also Datenerfassung oder Datenerhebung, durch Messen, Befragen oder durch Rekonstruktion aus Akten) nach vorher festgelegtem Schema.

Bei *experimentellen* Studien kommt demgegenüber mindestens eine weitere "Aktion", also ein Eingriff in den "natürlichen" Verlauf hinzu, der durch die Fragestellung der Studie motiviert ist.

Beispiel einer Studie mit "experimentellen Zügen": Im Rahmen einer klinisch indizierten Koronarangiographie werden in vorgegebenen Zeitabständen bis zum Zeitpunkt "30 min nach Beginn" zusätzliche Bilder der Koronarien erzeugt und gleichzeitig Blutproben entnommen. Die anschließende Vermessung soll Aufschluß über die Plasmakonzentration des vor der Angiographie gegebenen Medikamentes im Verlauf der Zeit ergeben und diese pharmakokinetischen Daten mit der Dilatation der Koronarien in Zusammenhang bringen.

Der experimentelle Aspekt ist in diesem Beispiel nicht sehr deutlich ausgeprägt: er besteht nach der Beschreibung allein darin, daß die übliche Angiographie den Erfordernissen des Studienplanes angepaßt wird und zusätzliche Meßvorgänge initiiert werden. Man könnte die Untersuchung insofern auch als reine Beobachtungsstudie ansehen. Dabei ist allerdings zu bedenken, daß hier die Messung selber -als invasives Verfahren- bereits als massiver "Eingriff" zu sehen ist und jede Änderung, insbesondere Verlängerung oder Verkomplizierung, auch wenn es "nur" der Messung dient, selber auch als Äktion anzusehen ist. Damit sind zumindest alle rechtlichen und ethischen Aspekte berührt und zu beachten, die mit der "biometrischen Forschung am Menschen" verbunden sind (siehe auch Abschnitt 5.2.3).

Die in diesem Beispiel skizzierte Versuchsanordnung weist aber noch nicht die Züge auf, die einen Versuch unter *methodischen* Aspekten als ein *Experiment* ausweisen: Ein Experiment, welches die Fragestellung

Welchen Einfluß hat X auf Y?

behandelt, sollte im Grundsatz folgende Struktur haben:

1. Die äußeren *Rahmenbedingungen des Experimentes* sind zu definieren. (Hierunter ist die gesamte Versuchsanordnung zu verstehen. Durch die Rahmenbedingungen werden diejenigen Merkmale festgelegt, die für die gesamte Dauer des Experimentes konstant sein sollen. Dazu können z.B. die technischen Geräte eines Laborversuchs gehören und ihre Handhabung, Materialien, Temperatur, diagnostischer und meßtechnischer Standard, aber auch die Dosierung bei einer Medikamentenstudie (wenn nicht die Dosis selber als prognostischer Faktor untersucht werden soll), die Einhaltung des vorgesehenen Zeitintervalles zwischen Medikamenteneinnahme und Messung einer Wirkung, Standardisierung der Eingangsinformationen und -fragen bei Interviews, IQ-Tests oder diagnostischen Tests und anderes mehr.)
2. Die Werte der prognostischen Variablen X (das können auch mehr als *ein* Merkmal sein) werden durch die experimentelle Anordnung *kontrolliert* und *systematisch variiert*. Damit ist gemeint, daß bei jeder Beobachtungseinheit *die Werte der prognostischen Variablen* nach einem vorher aufgestellten Plan festgelegt und im *Experiment selber* entsprechend realisiert ("eingestellt") werden. Die Zuordnung dieser Werte zu den einzelnen Beobachtungseinheiten erfolgt *unabhängig von der Beobachtungseinheit selber*, also unabhängig davon, *welche Werte die anderen Merkmale der Beobachtungseinheit haben*.

3. Auch die Werte von "Drittvariablen" können in diesem Sinne kontrolliert und systematisch variiert werden. Es ist dann aber zu bedenken, welchen Effekt von X auf Y man erfassen möchte. (Beispiel: X wirkt auf Y dadurch, daß X die "Zwischenvariable" Z verändert und diese ihrerseits Y beeinflusst. Der Gesamteffekt von X auf Y wird in diesem Fall nicht sichtbar, wenn bei der Veränderung von X die Variable Z durch die experimentelle Anordnung konstant gehalten wird.)
4. Zu jeder vorgesehenen Konstellation der durch die Versuchsanordnung kontrollierten Variablen (X und evtl. einige Drittvariable) werden die *experimentell nicht gesteuerten* Merkmale (insbesondere die Zielgrößen Y) in einer oder mehreren Einzeldurchführungen der Versuchsserie *gemessen*.

Bei Einhaltung dieser Grundsätze ist gewährleistet:

Ist nach den empirischen Ergebnissen eine Veränderung der Werte der prognostischen Variablen mit einer Veränderung der Werte der Zielgrößen Y verbunden so werden die Veränderungen in Y auch tatsächlich durch die Veränderungen in X hervorgerufen.

In diesem Sinne ist dann die Beziehung zwischen X und Y *kausal*. Das heißt nicht, daß die Wirkung von X auf Y *unmittelbar* sein muß; X kann auch in der Mitte oder am Anfang einer längeren Wirkungskette stehen, die möglicherweise noch gar nicht oder nur in Ansätzen bekannt ist. Eine Interpretation der folgenden Art ist dann aber in jedem Fall gerechtfertigt:

"Eine Änderung in X von x1 auf x2 bewirkt in Y eine Änderung von y1 auf y2, sofern die Rahmenbedingungen des Experimentes eingehalten werden."

Man beachte noch einmal: dieser Schluß ist bei Beobachtungsstudien *nicht* zulässig: Untersucht man z.B. sämtliche Unfälle einer Stadt und stellt dabei fest, daß der Gesundheitszustand der Unfallbeteiligten nach einer Woche insgesamt um so schlechter ist, je mehr Sanitäter, Ärzte und Notarztwagen am Unfallort eingesetzt wurden, so ist daraus nicht zu schließen, daß eine Erhöhung dieses Einsatzes zu einem höheren Ausmaß an Verletzungen bei Verkehrsunfällen führt. (Die Beziehung zwischen Einsatz und Gesundheitszustand wird durch die Drittvariable "Schwere des Unfalls" erklärt: Bei schwereren Unfällen sind erstens mehr und schwerere Verletzungen zu erwarten und werden zweitens mehr Ärzte etc. eingesetzt.) Eine experimentelle Studie zu dieser Fragestellung müßte so angesetzt werden, daß der Einsatz von Personal und Geräten von Unfall zu Unfall nach einem zuvor festgelegten Plan variiert wird, und zwar unabhängig von der Schwere des Unfalls. Es ist offenbar, daß eine solche Studie aus ethischen Gründen nie durchgeführt werden darf. Dies ist ein zweiter Hinweis darauf, daß methodische Erfordernisse –insbesondere bei experimentellen Studien– stets der Kontrolle durch ethische und rechtliche Überlegungen bedarf.

4.2.2. Besonderheiten bei Studien mit Zufallsdaten ("statistische Studien")

In klinischen Studien (wie in vielen anderen Bereichen der Forschung) muß man in der Regel davon ausgehen, daß das Ergebnis einer Einzelmessung eines Merkmales an einer Beobachtungseinheit *nie mit vollständiger Sicherheit vorhergesagt werden kann*. Dies gilt auch, wenn

eine Reihe von Merkmalen, die das Meßergebnis in bekannter Weise beeinflussen, im Experiment kontrolliert oder vorher gemessen wurden.

Beispiel: Die Wahrscheinlichkeit, daß die Chemotherapie PVBI bei Hodentumorpatienten zur kompletten Remission führt, beträgt 65% (die Vorhersage "Therapieerfolg" wäre also in 35% der Fälle falsch) Ist bei einem Patienten zusätzlich bekannt, daß die Metastasierung in die Kategorie "minimal disease" fällt, so beträgt die Erfolgswahrscheinlichkeit für die Therapie 95% (für diese Gruppe von Patienten wäre die Vorhersage "Therapieerfolg" also noch in 5% der Fälle falsch).

Dieser "Rest an Unsicherheit" führt dazu, daß man das Ergebnis jeder einzelnen Messung als –in gewissem Umfang– "zufallsabhängig" und das entsprechende Merkmal dementsprechend als "Zufallsvariable" betrachtet.

Bei mehrfacher Wiederholung des Experimentes können sich die Werte einer Variablen (z.B. einer Zielgrößen Y) von mal zu mal ändern, *obwohl die Rahmenbedingungen und alle zuvor deklarierten prognostischen Faktoren konstant gehalten werden*. Konstante Rahmenbedingungen und konstante prognostische Faktoren sind also nicht mit nur *einem* Wert der Zielgröße verbunden sondern mit *verschiedenen* Werten, die wiederum *mit unterschiedlichen Häufigkeiten auftreten*. Jede Konstellation von prognostischen Faktoren erzeugt also bei definierten Rahmenbedingungen eine bestimmte "Häufigkeitsverteilung" der Zielgröße Y (und nicht nur einen bestimmten Wert für Y).

Hieraus ergeben sich insgesamt mehrere *Folgerungen*:

1. Das Ziel der Studie ist -formal gesehen- die Untersuchung von *Häufigkeitsverteilungen* der Hauptvariablen. Dabei interessieren oft nur bestimmte *Parameter der Verteilungen* wie z.B. der Mittelwert (in der Grundgesamtheit), der Median oder die Streuung einer Zielgröße und *die Abhängigkeit dieses Parameters von prognostischen Faktoren*.

Beispiel: Die *mediane Überlebenszeit* und die *1-Jahres-Überlebenswahrscheinlichkeit* sind Beispiele für zwei verschiedene Parameter der *Verteilung der Variablen "Überlebenszeit"*. Häufig wird in diesem Zusammenhang aber die *gesamte Verteilung* der Überlebenszeit (so weit wie beobachtbar) und ihre Abhängigkeit von prognostischen Faktoren untersucht, beispielsweise also durch Vergleich der Überlebenskurven von unterschiedlich behandelten Patienten.

2. Die Kenntnisse über die *Verteilung* einer Variablen wird um so genauer, je häufiger der Einzelversuch unter gleichen Rahmenbedingungen wiederholt und die Variable dabei jeweils gemessen wird. (Man sagt auch: je größer die Anzahl der "Realisierungen" der Zufallsvariablen.)

Daher sind vor der Studie Überlegungen zur Größe des *Stichprobenumfangs* erforderlich.

3. Alle Ergebnisse einer Studie, z.B. die empirische Verteilung einer Variablen, sind -auch bei großem Stichprobenumfang- immer noch in einem gewissen Ausmaß zufallsabhängig. Demgegenüber geht man davon aus, daß es eine Verteilung der Zufallsvariablen "*in der Grundgesamtheit*" gibt, die *nicht zufallsabhängig* ist. Damit ist die *Grenzverteilung* gemeint, an die die empirische Verteilung mit wachsenden Stichprobenumfang *beliebig nahe* herankommt.

Beispiel: Anzahl der Würfe vor der ersten "6" beim Würfeln mit einem unverfälschten Würfel. Jeder Einzelversuch besteht darin, daß bis zum ersten Erscheinen der "6" gewürfelt wird und die Anzahl der Würfe bis dahin (die Wartezeit) notiert wird. Dieses wird 100 mal durchgeführt. Zu jeder *möglichen* Anzahl $k=0,1,2,..$ wird die *relative Häufigkeit* der Einzelversuche aufgetragen, bei denen die Wartezeit höchstens gleich k ist. Diese relativen Häufigkeiten werden den *theoretisch berechneten Wahrscheinlichkeiten* (der Verteilung in der Grundgesamtheit) gegenübergestellt. Man stellt dann in der regel eine "sehr gute" Übereinstimmung zwischen beiden Kurven fest.

4. Die Beziehungen zwischen den empirischen (zufallsabhängigen) Ergebnissen und der "wahren", *nicht zufallsabhängigen* Verteilungen zu quantifizieren, ist die Aufgabe der "*analytischen Statistik*" oder "*Inferenzstatistik*". Sie bedient sich der Hilfsmittel der *Wahrscheinlichkeitsrechnung*.

Damit diese angewendet werden kann, ist die Untersuchung so durchzuführen, daß die jeweiligen *Voraussetzungen für die Berechnung von Wahrscheinlichkeiten* erfüllt sind. Dazu gehört in der Regel:

5. Die Einzelversuche müssen unabhängig voneinander durchgeführt werden. Dies bedeutet z.B. bei (diagnostischen) Tests, daß sich die Probanden nicht gegenseitig beeinflussen, und daß bei Patientenstudien gegenseitiges Beeinflussungen durch Ansteckung oder anderes ausgeschlossen werden.
6. Abgesehen von systematischen Veränderungen der prognostischen Faktoren müssen alle anderen Merkmale *entweder konstant sein oder rein zufällig variieren*. Dies ist noch einmal die Forderung nach Konstanz der Rahmenbedingungen und nach der Repräsentativität der Stichprobe.
7. *Randomisierung*: Bei experimentellen Studien bedeutet die vorige Forderung (6.) zusätzlich, daß die Beobachtungseinheiten der Studienpopulation den einzelnen Behandlungsgruppen (den Konstellationen der prognostischen Faktoren) nach Kriterien zugeordnet werden, die *unabhängig von den Merkmalen der Beobachtungseinheit* sind. Dies wird durch eine randomisierte Zuteilung erreicht. Üblicherweise mit Hilfe von *Zufallszahlen* wird den einzelnen Beobachtungseinheiten (meist: Patienten) der Studienpopulation eine der im Studienplan vorgesehenen Behandlung zugeordnet. Dazu wird meist schon vor Beginn eine *Randomliste* erstellt, in welcher die Reihenfolge der anzuwendenden Behandlungen festgelegt ist (z.B. A A B A B B ...). Die Zuordnung entsprechend dieser Liste erfolgt dann gemäß der Reihenfolge der Aufnahme der Patienten in die Studie (1. Patient erhält Behandlung A, 2. Patient A, 3. Patient B etc.). Damit daran keine willkürlichen (und von Merkmalen des Patienten abhängige) Veränderungen vorgenommen werden können, ist mit Aufnahme des Patienten in die Studie zunächst seine Randomnummer festzulegen (z.B. "3. Patient", also Randomnummer 3) und zu dokumentieren, und erst *danach* in der Randomliste die zugehörige Behandlung (z.B. Behandlung B) offenzulegen. Dies wird z.B. dadurch realisiert, daß die zuzuteilende Behandlung jeweils in verschlossenem Umschlag hinterlegt ist, auf welchem außen die Randomnummer vermerkt ist. Erst nach Aufnahme in die Studie ist dann der Umschlag zu öffnen.

8. Die Forderung (6.) nach "rein zufälliger Variation von Drittvariablen" bezieht sich auch darauf, daß *durch den Untersucher selber* (trotz Standardisierung) ein Einfluß auf die *Meßergebnisse* und -bei Patientenstudien- auch auf die Ergebnisse der *Behandlung* ausgeübt werden kann. Dieser Effekt ist nur dann *unabhängig von der Konstellation der prognostischen Faktoren, wenn der Untersucher den jeweiligen Wert dieser Faktoren gar nicht kennt*. Wenn es also sowohl möglich als auch ethisch vertretbar ist, sollte daher der Untersucher *ohne Kenntnis der prognostischen Faktoren* (d.h. in der Regel: ohne Kenntnis der zugeteilten Medikation bzw. Therapie), in diesem Sinne also "*blind*", die Messungen durchführen bzw. den Patienten behandeln.
9. Die Folgerungen des vorigen Punktes (8.) sind bei Patientenstudien auch auf den Patienten anzuwenden: Das Wissen des Patienten, eine neue Therapie bzw. eine Standardtherapie zu erhalten, kann das Ergebnis der Therapie ebenfalls beeinflussen ("*Placebo-Effekt*"). Sofern es ethisch vertretbar ist, sollte der Einfluß dieser "Drittvariablen" daher ausgeschaltet werden. Das ist dadurch erreichbar, daß auch *der Patient nicht weiß, welche Behandlung er erhält*. Sind in diesem Sinne sowohl der behandelnde Arzt als auch der Patient "blind", so spricht man von einer "*Doppelblind-Studie*". Ist nur der Arzt, nicht aber der Patient über die wahre Behandlung informiert, so handelt es sich um eine "*Einfachblind-Studie*". Eine "Blindbehandlung" ist üblicherweise nur bei Medikamentenstudien möglich. Sie wird in diesem Falle dadurch realisiert, daß die verschiedenen Medikamente so aufbereitet werden, daß sie sich äußerlich und geschmacklich nicht voneinander unterscheiden lassen.
- Die in Punkt 7 beschriebene *Randomisierung* ist bei Blindstudien so zu modifizieren, daß die Zuteilung der Patienten zu den Randomnummern mit der Medikation von vornherein gekoppelt wird: Beispielsweise erhält der Patient mit der Randomnummer 3 die Medikamentenpackung mit der Nummer 3, die -nicht erkennbar- bereits als Medikament mit der Substanz B vorbereitet ist.

Über die hier dargestellten Grundsätze hinaus gibt es in der Versuchsplanung eine Reihe weiterer Überlegungen, die an die Zufallsabhängigkeit, also an die statistische Natur der zu untersuchenden Phänomene anknüpfen. Sie führen zu speziellen Versuchsplänen wie *Blockanlagen, Anlagen für mehrere Störfaktoren, faktoriellen Anlagen, Periodenversuchsplänen (Cross-over) und sequentiellen Versuchsplänen*.

4.2.3. Besonderheiten bei Studien am Menschen

Es ist bereits mehrfach deutlich geworden, daß methodische Forderungen insbesondere bei Studien an Menschen (Versuchspersonen oder Patienten), aber auch an Versuchstieren, durch ethische Überlegungen zu "kontrollieren" sind.

Dies ist nun nicht eine Aufgabe, die ein *Einzelner* nach *individuellen Maßstäben* oder eine *Studiengruppe* (Leiter, behandelnde Ärzte, Pharmakologen, Biometriker etc.) *frei nach eigenem Ermessen* durchzuführen hätte. Vielmehr gilt folgendes:

1. *Es gibt rechtliche und ethische Normen, an die alle Beteiligten gebunden sind.*

2. Bei klinischen Prüfungen (das ist die "Anwendung eines Arzneimittels am Menschen zum dem Zweck, über den einzelnen Anwendungsfall hinaus Erkenntnisse über den therapeutischen oder diagnostischen Wert eines Arzneimittels, insbesondere über seine Wirksamkeit und Unbedenklichkeit, zu gewinnen") muß vor der Aufnahme der Prüfung eine "unabhängige und sachkundige Ethik-Kommission gehört werden". (Zitiert nach der Bekanntmachung von Grundsätzen für die ordnungsgemäße Durchführung der klinischen Prüfung von Arzneimitteln, BAnz. S. 16617 vom 30.Dez. 1987.)

Die *rechtlichen und ethischen Normen* sind im wesentlichen in folgenden Texten niedergelegt:

1. Das Gesetz über den Verkehr mit Arzneimitteln (Arzneimittelgesetz (AMG) vom 24.8.1976; letzte Änderung am 11.4.1990), hierin insbesondere §40 -§42.
2. "Bekanntmachung von Grundsätzen für die ordnungsgemäße Durchführung der klinischen Prüfung von Arzneimitteln", BAnz. S. 16617 vom 30.Dez. 1987.
3. Die "revidierte Deklaration von Helsinki": Bekanntmachung von Beschlüssen des Weltärztebundes vom 26. Mai 1987 (BAnz. 108: 7109 vom 13. Juni 1987)
4. Die EG-Note for Guidance "Good Clinical Practice for Trials on Medicinal Products in the European Community" des "Committee for Proprietary Medicinal Products" von 1990.

Bis auf den Vierten sind alle genannten Texte in dem Buch

R.TOELLNER (Hrsg.): Die Ethik-Kommission in der Medizin. Stuttgart, New York (1990)

abgedruckt. Weiterhin informiert dieses Buch über "Problemgeschichte, Aufgabenstellung, Arbeitsweise, Rechtsstellung und Organisationsformen Medizinischer Ethik-Kommissionen". Primär methodische Aspekte, aber auch Hinweise auf weitere rechtliche Grundlagen, enthält das Buch

N.VICTOR, H.SCHÄFER, H.NOWAK: Arzneimittelforschung nach der Zulassung. Berlin, Heidelberg (1991)

Mit diesen Hinweisen soll ein Hilfe für den Einstieg in diesen wichtigen Problembereich gegeben werden, der hier in diesem Skript nun nicht ausführlicher behandelt wird. Als ganz wesentliche Leitsatz seien jedoch hervorgehoben:

"Wer eine klinische Prüfung plant oder durchführt, muß sich bewußt sein, daß es zwischen der Fürsorgepflicht gegenüber dem einzelnen Patienten beziehungsweise Probanden und dem allgemeinen Verlangen nach therapeutischem Fortschritt abzuwägen gilt. Gemessen an der voraussichtlichen Bedeutung des Arzneimittels für die Heilkunde müssen die Risiken für die teilnehmenden Personen ärztlich vertretbar sein."

"Vor Aufnahme in die Prüfung müssen die Patienten beziehungsweise Probanden in die Teilnahme an der Prüfung eingewilligt haben, nachdem sie über Wesen, Bedeutung und Tragweite in verständlicher Form aufgeklärt worden sind." ("Informed consent")

(Zitiert aus den "Grundsätzen für die ordnungsgemäße Durchführung der Prüfung von Arzneimitteln").

Die Forderung nach der Abwägung von voraussichtlichem Nutzen für die Heilkunde und dem Risiko für den Patienten impliziert z.B. auch für viele Studien, daß sie *sequentiell* (d.h. im Verlaufe der Studie in mehreren Schritten bei wachsender Patientenzahl und/oder Beobachtungsdauer) ausgewertet werden müssen. Andernfalls besteht die Gefahr, daß ein weniger wirksames Medikament oder eine Behandlung mit stärkeren Nebenwirkungen weiterhin einem Teil der Probanden gegeben wird, obwohl dieser negative Effekt aus den bereitliegenden Zwischenergebnissen schon deutlich abzulesen wäre.

Eine weitere Folgerung besteht darin, daß *jeder einzelne Patient kontinuierlich zu beobachten ist*, damit eventuelle individuelle Risiken rechtzeitig erkannt werden.

Die Kriterien für den *individuellen Studienabbruch* (frühzeitige Herausnahme eines Patienten aus der Studienmedikation) und für den frühzeitigen *Abbruch der gesamten Studie* sind vor Beginn der Studie festzulegen.

5. Das Studienprotokoll

Die methodischen Überlegungen entsprechend den oben dargestellten Richtlinien und unter der Berücksichtigung weiterer, spezieller Erfordernisse sind möglichst gründlich und bis in die Einzelheiten vor Beginn der Studie anzustellen. Sie führen dann dazu, daß man sich in der gesamten Anlage der Studie sowie in vielen Einzelheiten während dieser Planungsphase festlegt. Alles dies ist *schriftlich* festzuhalten und insgesamt zu einem *Studienprotokoll* zusammenzufassen.

Dies ist *grundsätzlich für jede Art von Studie* unbedingt zu empfehlen. Andernfalls besteht die Gefahr, daß sich anfänglich festgelegte Standards für Diagnostik, Messung und Behandlung im Verlaufe der Studie ändern, daß -insbesondere bei multizentrischen Studien- verschiedene Untersucher unterschiedliche Standards haben oder daß man anfänglich getroffenen Regelungen für die Durchführung ganz einfach vergißt. Bei länger dauernden Studien mit wechselnden Untersuchern wird das Problem der "Konstanz" besonders schwierig. Schließlich führt auch häufig erst die schriftliche Fixierung dazu, daß man Probleme überhaupt erkennt und damit "behandelbar" macht.

Für die *klinische Prüfung* ist das Studienprotokoll auch *rechtlich vorgeschrieben* und bildet den wesentlichen Bestandteil in einem entsprechenden Antrag an die Ethik-Kommission ("Grundsätze für die ordnungsgemäße Durchführung der klinischen Prüfung von Arzneimitteln").

Dies ist eine Forderung, die auch der Weltärztebund in der revidierten Deklaration von Helsinki stellt:

"Die Planung und Durchführung eines jeden Versuches am Menschen sollte eindeutig in einem Versuchsprotokoll niedergelegt werden; dieses sollte einem besonders berufenen unabhängigen Ausschuß zur Beratung, Stellungnahme und Orientierung zugeleitet werden."

Als gewisse Konkretisierung und Ergänzung der dargestellten Überlegungen zur Methodik medizinischer Studien geben wir zum Abschluß die Liste derjenigen Punkte wieder, zu denen

bei klinischen Prüfungen der "Prüfplan" (entspricht "Versuchsprotokoll" bzw. "Studienprotokoll") Angaben enthalten soll:

"Grundsätze für die ordnungsgemäße Durchführung der klinischen Prüfung von Arzneimitteln"

Abschnitt 2.5 und 2.6

(BAnz.243: 16617, vom 30.Dez.1987):

2 Planung der klinischen Prüfung

2.5 Vor Beginn der Prüfung ist ein Prüfplan aufzustellen. Er soll Angaben zu folgenden Punkten haben:

2.5.1 Zielsetzung und Begründung der Prüfung; Festlegung des Hauptzielkriteriums und Begründung seiner Eignung für die Erreichung des Prüfziels,

2.5.2 Charakterisierung des zu prüfenden Arzneimittels; die Zusammensetzung und die pharmazeutische Qualität müssen über eine eindeutige Identifizierung (Chargenbezeichnung) zurückverfolgt werden können,

2.5.3 Beschreibung des Prüfdesigns und gegebenenfalls Definition der Beobachtungseinheit,

2.5.4 Definition der Zielpopulation durch Ein- und Ausschlußkriterien,

2.5.5 Methodik der Personenauswahl,

2.5.6 Handhabung des Randomisierungsverfahrens und Beschreibung der Dekodierung bei Doppelblindstudien,

2.5.7 begründete Angaben über die Zahl der Patienten beziehungsweise Probanden unter Berücksichtigung der geschätzten Ausfallraten,

2.5.8 bei multizentrischen Prüfungen: Anzahl der Zentren und Anzahl der Personen pro Zentrum,

2.5.9 Behandlung (Art, Dosis, Dauer, Art der Anwendung des Arzneimittels, ambulante/stationäre Durchführung) in den einzelnen Gruppen,

2.5.10 zulässige und unzulässige Begleittherapie,

2.5.11 Auflistung aller Ziel- und Begleitvariablen,

2.5.12 die verwendeten Meßverfahren und deren Validierung. Bei multizentrischen Prüfungen müssen die entscheidenden Meßmethoden standardisiert sein.

2.5.13 Ermittlung, Bewertung und Dokumentation unerwünschter Begleiterscheinungen,

2.5.14 ausführliche Beschreibung des Prüfungsablaufs einschließlich des Zeitplans für die Untersuchungstermine,

2.5.15 Überprüfung der Compliance,

2.5.16 vorgesehene Gesamtdauer der Prüfung

2.5.17 biometrische Auswertungsmethoden mit Festlegung der Arbeitshypothesen und der Irrtumswahrscheinlichkeiten sowie Zeitpunkt und Umfang vorgesehener Zwischenauswertungen,

2.5.18 eventuell notwendige Vorsichtsmaßnahmen einschließlich Handlungsanweisungen, wie etwa Veränderungen der Dosierung,

2.5.19 Kriterien für den Abbruch der klinischen Prüfung sowohl im Einzelfall als auch für die gesamte Prüfung,

2.5.20 Verfahren zur Kontrolle der Einhaltung des Prüfplans,

2.5.21 Anleitung zur Dokumentation der Befunde,

- 2.5.22 Quellenangaben der verwendeten Informationen, insbesondere der benutzten oder zu benutzenden historischen und bibliographischen Daten,
 - 2.5.23 der Ort (die Orte) der Prüfung sowie die Art der Einrichtung, wo die Prüfung stattfindet,
 - 2.5.24 Name, Qualifikation und Verantwortungsbereich des jeweiligen Arztes für die einzelnen Abschnitte der klinischen Prüfung.
 - 2.6 Zur Erfassung und Dokumentation der Befunde bei den einzelnen Patienten ist ein Prüfbogen zu verwenden, der alle Angaben enthalten muß, die zur fundierten Beantwortung der im Prüfplan formulierten Fragestellungen notwendig sind. Hierzu gehören mindestens Angaben
 - 2.6.1 zur Identifizierung unter Berücksichtigung des Datenschutzes,
 - 2.6.2 Alter, Größe und Gewicht, Geschlecht, wichtige prognostische Faktoren (z.B. Raucher, Diät, bisherige Krankheitsdauer),
 - 2.6.3 eine etwaige Schwangerschaft bei Frauen im gebärfähigen Alter,
 - 2.6.4 Erfüllung der Einschlußkriterien und Nichtvorliegen von Ausschlußkriterien,
 - 2.6.5 Diagnose und Begründung für die Anwendung des Arzneimittels, Zeitpunkt der Diagnosestellung, Begleitdiagnosen, sowie Zeitpunkt der Stellung der Begleitdiagnosen.
 - 2.6.6 Einzeldosis, Tagesdosis, Dosierungsschema und Art der Anwendung des Arzneimittels,
 - 2.6.7 Beginn und Ende (Datumsangaben) der Behandlung und des Beobachtungszeitraums,
 - 2.6.8 alle Begleittherapien und relevante Vortherapien,
 - 2.6.9 Ergebnisse der Messung der Ziel- und Begleitvariablen mit Angabe der Meßzeitpunkte,
 - 2.6.10 unerwünschte Begleiterscheinungen (Art, Zeitpunkt des Auftretens, Dauer, Intensität, Maßnahmen/Folgen, Zusammenhang),
 - 2.6.11 zur Compliance
 - 2.6.12 Gründe für einen Abbruch
 - 2.6.13 Gesamtbeurteilung (Wirksamkeit und Verträglichkeit),
 - 2.6.14 Name und Adresse des prüfenden Arztes.
- Ein Muster des Prüfbogens ist Bestandteil des Prüfplans.