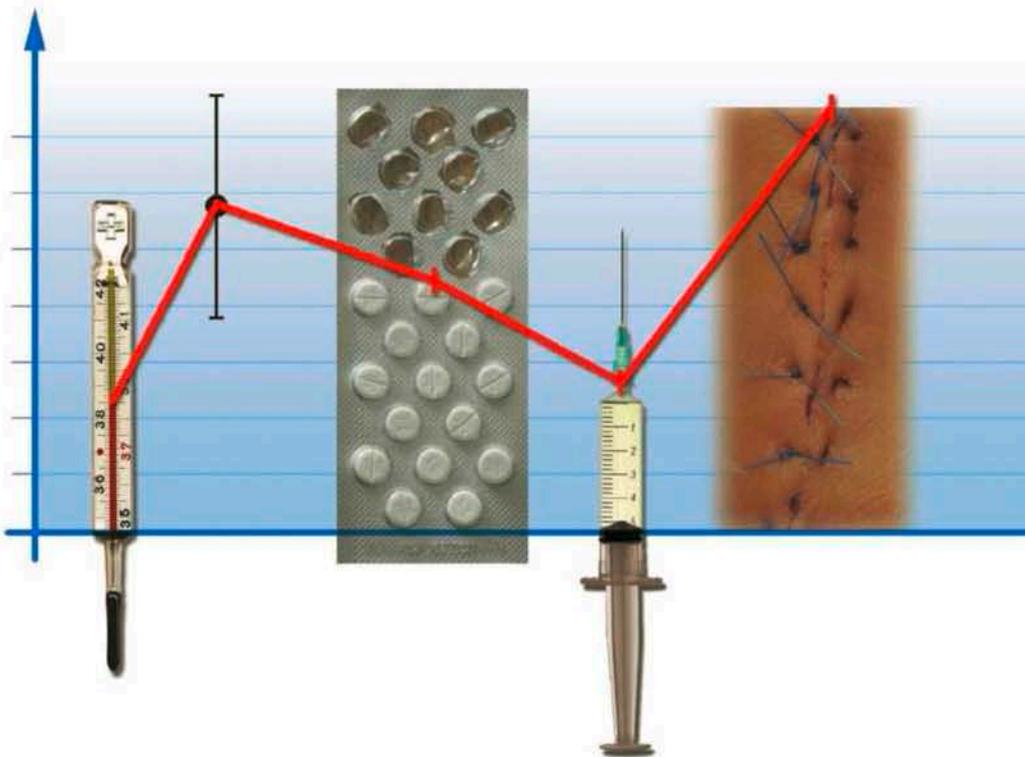


Wilhelm Gaus ■ Rainer Muche

# Medizinische Statistik

2. Auflage



Angewandte Biometrie  
für Ärzte und Gesundheitsberufe

Wilhelm Gaus ■ Rainer Mücke

# **Medizinische Statistik**

**2. Auflage**

This page intentionally left blank

**Wilhelm Gaus ■ Rainer Muche**

# **Medizinische Statistik**

**Angewandte Biometrie  
für Ärzte und Gesundheitsberufe**

**2., überarbeitete Auflage**

 **Schattauer**

## **Prof. Dr. phil. Wilhelm Gaus**

Universität Ulm, Institut für Epidemiologie und Medizinische Biometrie  
Schwabstraße 13, 89075 Ulm

E-Mail: wilhelm.gaus@uni-ulm.de

## **Prof. Dr. biol. hum. Dipl.-Statistiker Rainer Muche**

Universität Ulm, Institut für Epidemiologie und Medizinische Biometrie  
Schwabstraße 13, 89075 Ulm

E-Mail: rainer.muche@uni-ulm.de



Ihre Meinung zu diesem Werk ist uns wichtig!  
Wir freuen uns auf Ihr Feedback unter [www.schattauer.de/feedback](http://www.schattauer.de/feedback)  
oder direkt über QR-Code.

### **Bibliografische Information der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

### **Besonderer Hinweis:**

In diesem Buch sind eingetragene Warenzeichen (geschützte Warennamen) nicht besonders kenntlich gemacht. Es kann also aus dem Fehlen eines entsprechenden Hinweises nicht geschlossen werden, dass es sich um einen freien Warennamen handelt.

Das Werk mit allen seinen Teilen ist urheberrechtlich geschützt. Jede Verwertung außerhalb der Bestimmungen des Urheberrechtsgesetzes ist ohne schriftliche Zustimmung des Verlages unzulässig und strafbar. Kein Teil des Werkes darf in irgendeiner Form ohne schriftliche Genehmigung des Verlages reproduziert werden.

© 2014, 2017 by Schattauer GmbH, Hölderlinstraße 3, 70174 Stuttgart, Germany

E-Mail: [info@schattauer.de](mailto:info@schattauer.de)

Internet: [www.schattauer.de](http://www.schattauer.de)

Printed in Germany

Lektorat und redaktionelle Bearbeitung durch die Autoren

Umschlagabbildungen: Thermometer: © Otto Durst, [www.fotolia.de](http://www.fotolia.de); Spritze: © verdateo, [www.fotolia.de](http://www.fotolia.de); Operationsnaht: © Berchtesgaden, [www.fotolia.de](http://www.fotolia.de)

Satz: Reproduktionsfertige Vorlagen von den Autoren

Druck und Einband: Himmer GmbH Druckerei & Verlag, Augsburg

Auch als eBook erhältlich:

ISBN 978-3-7945-9072-8

ISBN 978-3-7945-3241-4

# Vorwort

**Jeder Mensch ist einmalig**, sowohl in seiner genetischen Veranlagung (eineiige Mehrlinge ausgenommen) als auch in seinem Lebenslauf. Deshalb können wir nicht erwarten, dass diagnostische Verfahren immer den richtigen Befund liefern und Therapien immer gleich wirken. Vielmehr sind die meisten Diagnosen richtig, aber nicht alle; häufig hilft eine Therapie, aber nicht immer, und gelegentlich tritt eine Komplikation ein. Damit sind wir schon mitten in der Statistik.

**Medizinische Statistik und Biometrie** ist mehr als die Anwendung von Computerprogrammen und den darin enthaltenen Formeln. Statistische Verfahren, Formeln und Programme sind Werkzeuge, die sachgerecht und fachkundig angewandt werden müssen. Zum Vergleich: Ein gut ausgestattetes Maleratelier ersetzt nicht den Künstler. Bevor man Daten auswerten kann, muss man die Fragestellung kennen, im Detail verstehen, wie die Daten entstanden sind, und die Datenqualität „diagnostizieren“. Nach der Auswertung müssen die Ergebnisse verstanden und interpretiert werden, außerdem ist die Aussagekraft der Daten und der daraus gewonnenen Ergebnisse zu beurteilen. Die Rechenarbeit erledigt heute der Computer, statistisches Denken und biometrische Strategien sind aber eher Philosophie als Rechnen.

**Kenntnisse in Statistik** braucht heute jede wissenschaftlich tätige Ärztin, jeder wissenschaftlich tätige Arzt. Aber auch die klinisch und praktisch tätigen Mediziner, die die wissenschaftlichen Ergebnisse verstehen und anwenden wollen, benötigen Statistik, um beispielsweise neue Therapien beurteilen zu können. Medizinische Statistik wird aber auch in vielen Gesundheitsberufen benötigt, z.B. von Medizinischen Dokumentaren, Medizininformatikern, medizinisch-technischem Personal, Psychologen und nicht zuletzt von Personen, die bei Krankenkassen, Kassenärztlichen Vereinigungen, Pharmafirmen oder im Bereich Public Health arbeiten. Spezielle medizinische Fachausdrücke werden erklärt, so dass auch Nichtmediziner von den Beispielen profitieren können. Für alle statistischen Fachausdrücke ist auch die englische Bezeichnung angegeben.

**Klinische Studien** und statistische Gedanken sind für die Medizin auf ihrem Weg von der Heilkunst zu einer evidenzbasierten Medizin unentbehrlich. Klinische Studien sind formalisierte Erfahrungsheilkunde. Sie sind aber nur nützlich, wenn sie statistisch korrekt sind.

**Beratung und Betreuung in Statistik.** Dieses Buch will Sie in das Gedankengut, die Werkzeuge und die Sprache der Statistik einführen. Es soll Ihnen Verständnis für die angewandte medizinische Biometrie geben und Ihnen eine Vorstellung vermitteln, was die Statistik leisten kann und was nicht. Es soll Ihnen helfen, Publikationen und deren Ergebnisse besser beurteilen zu können. Bei größeren experimentellen Arbeiten, bei Tierversuchen und bei klinischen Studien kann das Buch die Betreuung durch einen professionellen Biometriker nicht ersetzen. Aber Sie werden nach der Lektüre dieses Buches einen professionellen Statistiker besser verstehen und erheblich besser mit ihm zusammenarbeiten. Vor allem aber: Das Buch soll Ihnen die Scheu vor der Statistik nehmen, die näher am Puls des Lebens ist als der erste Eindruck vermuten lässt.

Bei der Lektüre werden Sie immer wieder dem Patienten **Max Moritz Schaufelstiel**, den Doktores **Emsig** und **Fleißig**, dem Statistiker **Pingelig** und ähnlichen Personen begegnen. Dies sind erfundene Namen. Eine mehr oder weniger gute Übereinstimmung mit einer wirklichen Person ist nicht beabsichtigt, bei der großen Anzahl wirklicher Personen aber durchaus rein zufällig möglich. Auch alle in diesem Buch gegebenen Beispiele haben – falls nicht ausdrücklich die Quelle genannt ist – erfundene Daten.

Die **Medizinische Statistik ist voller Tücken** wie folgendes vorgezogene Beispiel zeigt:

Die Beurteilung des Therapieerfolgs erfordert ein Erfolgskriterium, in der Statistik **Zielgröße** oder **Outcome-Variable** genannt. Beispiele für Zielgrößen sind Blutdrucksenkung, Anzahl der kardial bedingten Todesfälle innerhalb eines bestimmten Zeitraums, Überlebenszeit nach Diagnosestellung. Die **Wirksamkeit einer Therapie** ist nachgewiesen, wenn mit dieser Therapie die Zielgröße stärker verbessert wird als ohne diese Therapie. Im Bereich Public Health wird die Meinung vertreten, dass eine Therapie nicht nur wirksam, sondern auch nützlich sein muss. Der **Nutzen einer Therapie** ist definiert als die durch die Therapie erreichte Verbesserung der Lebensqualität und/oder eine Verlängerung der Lebenszeit. Wenn eine Therapie z.B. den Tumor verkleinert, aber dies nicht mit einer Verbesserung der Lebensqualität oder einer Lebensverlängerung verbunden ist, so ist diese Therapie zwar wirksam, aber nutzlos. Dies ist ein plausibles, vernünftiges und nachvollziehbares Konzept.

Aber es gibt noch folgenden **Witz**: Max Moritz Schaufelstiel hat ganz schlimmen Durchfall, laufend geht ihm alles in die Hose. Er geht zum Arzt, der verordnet ihm ein Benzodiazepin. (Benzodiazepine sind eine Gruppe verwandter Substanzen, die als Psychopharmaka zugelassen sind, z.B. Valium®.) Schaufelstiel erzählt dies einem Freund – es ist ein echter Freund – und der ist verwundert: „Ein Benzodiazepin bei Diarrhöe?? Komisch!!“ und fragt Max Moritz: „Hilft es denn? Geht Dir nichts mehr in die Hose?“ „Doch“, ist die Antwort, „es ist wie bisher, aber es macht mir nichts mehr aus.“

**Fazit**: Wird Lebensqualität als Zielgröße genommen, so ist ein Benzodiazepin auch bei Diarrhöe wirksam und nützlich. Damit ist die Bedeutung der Lebensqualität als universelle Zielgröße und der damit definierte Nutzen einer Therapie erschüttert. Offensichtlich ist bei überwiegend somatischen, nicht psychischen Erkrankungen eine auf das Krankheitsbild zugeschnittene Zielgröße sinnvoller.

In Heidelberg steht über dem Eingang des Anatomischen Instituts: „**Hic gaudet mors succurrere vitae**“ (Hier freut sich der Tod, dem Leben zu helfen). Als die Studienplätze in Medizin per Los zugeteilt wurden, ersetzten Studenten das „m“ durch „s“. In dem geänderten Text „**Hic gaudet sors succurrere vitae**“ (Hier freut sich der Zufall, dem Leben zu helfen) steckt vielleicht mehr Wahrheit, als die Studierenden damals ahnten.

**Mathematische Formeln** sind nur angegeben wenn sie entweder für das Verständnis hilfreich sind (wie z.B. die Formel für die Varianz) oder wenn sie so einfach sind, dass sie bequem mit dem Taschenrechner gerechnet werden können (wie z.B. die partielle Korrelation). Formeln, die üblicherweise mit Computerprogramm gerechnet werden, oder für nicht-professionelle Statistiker schwierig zu verstehen sind, werden nicht angegeben, da weder hilfreich noch erforderlich.

**Dezimalpunkt**. In der Wissenschaft, auch in der Medizin wird für Dezimalzahlen zunehmend die anglo-amerikanische Schreibweise mit Dezimalpunkt verwendet. Dies gilt vom Taschenrechner bis zu statistischen Auswertesystemen. Im gesamten Buch wird der Dezimalpunkt, nicht das Dezimalkomma verwendet. Durch die anglo-amerikanische Schreibweise muss sich der Statistiker nicht laufend umstellen und es können z.B. Tabellen und Abbildungen leichter in englische Texte übernommen werden. Das Komma trennt zwei Zahlen.

**Rechengenauigkeit**. „In nichts zeigt sich der Mangel an mathematischer Bildung mehr als in einer übertrieben genauen Rechnung“ sagte **Carl Friedrich Gauß** (1777 bis 1855). Heute rechnet der Computer, der soll ruhig übertrieben genau rechnen, aber präsentieren Sie keine übertrieben genauen Ergebnisse. Im Buch finden Sie viele kleine, scheinbare Widersprüche, die trotz korrektem Runden entstanden sind. Allerdings, einige Beispiele haben übertrieben genaue Zahlenangaben um den Leser, der das Beispiel nachrechnet, nicht durch Rundungsfehler zu verunsichern.

**Computerprogramme.** Es gibt einfache und es gibt leistungsfähige statistische Auswertungssysteme, aber keine, die sowohl einfach als auch leistungsfähig sind. Die meisten Programme rechnen „richtig“, aber es kann verflucht schwierig sein,

- dem Computerprogramm das sachliche Problem beizubringen,
- zu erkennen, welches statistische Verfahren das Programm gerechnet hat, und
- die Ausgabe des Programms zu verstehen.

Das für Sie beste Auswertungssystem ist jenes, für das Sie einen kompetenten Ansprechpartner haben.

**Weibliche und männliche Formulierungen.** Bei allen Personen möchten wir stets Frauen und Männer gleichermaßen ansprechen. Dies lässt sich sprachlich nur umständlich und unästhetisch ausdrücken. Daher haben wir uns entschieden, mal die weibliche, mal die männliche Form zu verwenden, und bitten unsere Leserinnen und Leser, darunter stets beide Geschlechter zu verstehen.

**Übungsaufgaben.** Zu allen Themen werden Übungsaufgaben gestellt, die sich als Übungs-, Klausur- und Prüfungsaufgaben bei Medizinstudenten, Public-Health-Studenten und in der Ausbildung zum Medizinischen Dokumentar bewährt haben, insgesamt 174 Aufgaben. Diese zusätzlichen Beispiele ermöglichen Ihnen ein „learning by doing“. Vielleicht entspricht Ihr aktuelles Problem einer der Aufgaben. Mit den als Kapitel 32, S.517 beigefügten Lösungen können Sie sich selbst überprüfen und dazulernen.

**Umfang des Buches.** Interessierte sollten sich von den insgesamt 640 Seiten des Buches nicht abschrecken lassen. Die Titelei umfasst 18 Seiten, die Übungsaufgaben 43 Seiten und deren Lösungen 63 Seiten, Verzeichnisse, Literaturhinweise und Sachwortregister 70 Seiten, so dass der eigentliche Text „nur“ 446 Seiten belegt. Außerdem, der Text ist so lang, weil wir uns bemüht haben, zu erklären und verständlich zu sein, von verschiedenen Aspekten zu beleuchten und immer wieder an Beispielen zu veranschaulichen. Lesbarkeit und Verständlichkeit sind uns wichtiger als mathematische Kürze.

**Statistik und Medizin.** Die Geschichte der Medizin ist ein Wandel von der Heilkunst zu einer naturwissenschaftlichen Medizin und von qualitativen zu quantitativen Befunden. Im Mittelalter war z.B. die Harnschau eine ärztliche Kunst, dann kam die Beobachtung des Durchflusses (Diabetes), später der Honigsüße (mellitus), heute gilt die quantitative Bestimmung der Serumglukose und des HbA1c-Werts. Quantitative Methoden und statistische Denkweise waren lange Zeit verpönt. Auch beim Kongress der Pariser Académie Royal de Médecine im Jahre 1835 wurde argumentiert, dass der Patient durch statistische Methoden seine Individualität verlöre und mit der Wahrscheinlichkeitsrechnung die Medizin den Charakter einer Lotterie bekäme [Lit.88]. Das Problem der Mediziner mit der Statistik ist also nicht wirklich neu. Heute soll die medizinische Statistik den Arzt nicht bevormunden, sondern ihm helfen, für den einzelnen Patienten die richtige Diagnose und die optimale Therapie zu finden.

**Herzlichen Dank.** Bei der Erstellung eines solch dicken Buches schleichen sich – trotz mehrfacher Durchsicht – Fehler ein. Großen Dank an alle Leser der ersten Auflage, die uns auf Fehler und Mängel hingewiesen haben! Hoffentlich haben sich jetzt nicht allzu viele neue Fehler eingeschlichen.

**Willkommen.** In der Medizinstatistik gibt es – wie bei allen Fachgebieten – Höhen und Tiefen, einfache und schwierige Passagen, kurzweilige und zähe Strecken. Wir wünschen Ihnen viel Erfolg, aber auch etwas Vergnügen mit unserem Buch und heißen Sie herzlich willkommen im Land der Medizinstatistik!

Ulm, im Dezember 2016

**Wilhelm Gaus und Rainer Muche**



# Inhaltsdisplay

## Grundlagen

- Kap. 1 Statistische Vergleichbarkeit
- Kap. 6 Wahrscheinlichkeit
- Kap. 11 Normalverteilung und andere theoretische Verteilungen

## Unentbehrliches

- Kap. 4 Merkmalstypen und Skalen
- Kap. 7 Statistische und epidemiologische Kenngrößen
- Kap. 8 Gestalten von Tabellen, Abbildungen und Schemata

## Studien

- Kap. 2 Typen von Studien
- Kap. 3 Durchführen von Studien

## Schätzen

- Kap. 12 Schätzen
- Kap. 13 Konfidenzintervalle

## Testen

- Kap. 14 Prinzip des statistischen Tests
- Kap. 15 Chi-Quadrat Test und andere Tests für qualitative Zielgrößen
- Kap. 16 Logrank Test
- Kap. 17 Rangtests
- Kap. 18 t-Test
- Kap. 19 Einfache Varianzanalyse
- Kap. 20 Faktorielle Varianzanalyse
- Kap. 22 Multiples Testen
- Kap. 23 Allgemeine versus spezielle Hypothesen
- Kap. 24 Äquivalenzttests

## Spezielle Anwendungen

- Kap. 5 Häufigkeiten und Häufigkeitsverteilung
- Kap. 9 Korrelation und einfache lineare Regression
- Kap. 10 Überlebenszeitanalyse
- Kap. 21 Multiple Regression

## Übergordnetes

- Kap. 25 Fallzahlbestimmung
- Kap. 26 Fehlende Werte
- Kap. 27 Meta-Analysen
- Kap. 28 Medizinische Statistik – Mathematik oder Orakel?

## **Sieben Merkwürdigkeiten der Biometrie**

1. Biometrie ist Mathematik, aber nichts ist sicher.
2. Obwohl alles unsicher ist, ist es Wissenschaft.
3. Biometrie ist zwar Wissenschaft, aber die Datenlage ist unklar.
4. Obwohl die Datenlage unklar ist, gewinnen wir daraus Ergebnisse.
5. Die Ergebnisse sind klare Zahlen, aber diese sind vieldeutig.
6. Obwohl die Ergebnisse vieldeutig sind, werden sie publiziert.
7. Die publizierten Ergebnisse bringen Impact-Punkte.

**Wir bemühen uns, mit der Biometrie den Ärzten und Patienten zu helfen.**

# Inhaltsverzeichnis

<b>1</b>	<b>Statistische Vergleichbarkeit</b>	<b>19</b>
1.1	Wiederholen und Vergleichen	19
1.2	Unabhängigkeit der Beobachtungen	19
1.3	Systematische und zufällige Fehler	20
1.4	Eine Vergleichsgruppe ist notwendig	23
1.5	Vermengte Effekte	24
1.6	Struktur-, Behandlungs- und Beobachtungsgleichheit	25
1.7	Schichtung und matched Pairs	26
1.8	Randomisation	27
1.9	Verbundene Versuchsanordnung	30
1.10	Placebo und maskierte Gruppen	32
1.11	Übungsaufgaben	34
<b>2</b>	<b>Typen von Studien</b>	<b>37</b>
2.1	Einteilungskriterien für Studien	37
2.2	Retrospektive Auswertung von Krankenakten	38
2.3	Prospektive Befunddokumentation	41
2.4	Epidemiologische Krankheitsregister	42
2.5	Diagnostische Studien	42
2.6	Vorsorge und Früherkennung	43
2.7	Kohortenstudien (Cohort Studies)	45
2.8	Fall-Kontroll-Studien (Case Control Studies)	47
2.9	Kontrollierte, randomisierte Studien	49
2.10	Stufen der Arzneimittelentwicklung	50
2.11	Zusammenschau	51
2.12	Ethische Fragen	53
2.13	Übungsaufgaben	54
<b>3</b>	<b>Durchführung von Studien</b>	<b>57</b>
3.1	Studienplan	57
3.2	Randomisationspläne	62
3.3	Datenerhebungsbögen, Case Report Forms (CRFs)	63
3.4	Datenmanagement	68
3.5	Organisationen und Einrichtungen einer Studie	70
3.6	Multizentrische und kooperative Studien	73
3.7	Good Clinical Practice (GCP)	74
3.8	Monitoring	75
3.9	Auditing	76
3.10	Vorbereitung der Auswertung	76
3.11	Auswertung	78
3.12	Aussagekraft von Studien	80
3.13	Übungsaufgaben	80

<b>4</b>	<b>Merkmalstypen und Skalen</b>	<b>82</b>
4.1	Merkmale und Merkmalsausprägungen	82
4.2	Skalen	82
4.3	Merkmalstypen	83
4.4	Qualitative versus quantitative Merkmale	83
4.5	Diskrete versus stetige Merkmale	85
4.6	Ratingmerkmale versus Ränge	86
4.7	Visuelle Analogskala (VAS)	86
4.8	Informationsgehalt eines Merkmals und Umwandlungsmöglichkeiten	87
4.9	Einfluss-, Begleit- und Zielgrößen	88
4.10	Modellvorstellung	89
4.11	Übungsaufgaben	90
<b>5</b>	<b>Häufigkeiten und Häufigkeitsverteilung</b>	<b>91</b>
5.1	Häufigkeiten auszählen	91
5.2	Relative Häufigkeiten, Anteile und Prozentsätze	92
5.3	Häufigkeitsverteilung	96
5.4	Kumulierte Häufigkeiten	100
5.5	Beispiel einer empirischen Häufigkeitsverteilung	101
5.6	Kumulierte Einzelwerte	102
5.7	Kontingenztafel	105
5.8	Übungsaufgaben	106
<b>6</b>	<b>Wahrscheinlichkeit</b>	<b>108</b>
6.1	Begriff der Wahrscheinlichkeit	108
6.2	Bedingte Wahrscheinlichkeit	110
6.3	Formale Definitionen der Wahrscheinlichkeit	110
6.4	Zufällige Ereignisse, Erwartungswert	111
6.5	Wahrscheinlichkeitsverteilungen, Wahrscheinlichkeits- und Verteilungsfunktion	113
6.6	Additions- und Multiplikationssatz	117
6.7	Unabhängigkeit von Ereignissen	118
6.8	Satz von Bayes	119
6.9	Beispiel: Risiko für Spondylitis ankylosans	120
6.10	Subjektive Wahrnehmung von Chancen und Risiken	121
6.11	Elementare Kombinatorik	122
6.12	Übungsaufgaben	125
<b>7</b>	<b>Statistische und epidemiologische Kenngrößen</b>	<b>126</b>
7.1	Mittelwert, Median, Modus	126
7.2	Quantile (= Fraktile)	129
7.3	Streuemaße	131
7.4	Beispiel zu den Maßen der zentralen Tendenz und den Streumaßen	134
7.5	Genauigkeit von Messungen und Befunden	135
7.6	Kenngrößen für Krankheits- und Todesrisiken	135
7.7	Zuverlässigkeit einer diagnostischen Entscheidung	137
7.8	Güte einer Zufallsentscheidung	140
7.9	Beispiel einer Untersuchung zur Krankheitsfrüherkennung	141
7.10	ROC-Kurve	142
7.11	Risiken von Expositionen	144
7.12	Number Needed to Treat (NNT)	146
7.13	Beispiele zu den Kenngrößen	146
7.14	Index = Indexwert	148
7.15	Übungsaufgaben	148

<b>8</b>	<b>Gestalten von Tabellen, Abbildungen und Schemata</b>	<b>150</b>
8.1	Vor- und Nachteile von Tabellen, Grafiken und Schemata	150
8.2	Überschrift, Legende, Beschriftung einer Darstellung	151
8.3	Tabellen	152
8.4	Grafische Hilfsmittel	153
8.5	Darstellung von Kenngrößen	156
8.6	Darstellung von Anteilen	159
8.7	Zeitliche Verläufe	161
8.8	Schemata	163
8.9	Ablaufdiagramme	165
8.10	Kartogramme	165
8.11	Präsentationen für Vorträge	165
8.12	Übungsaufgaben	168
<b>9</b>	<b>Korrelation und einfache lineare Regression</b>	<b>170</b>
9.1	Univariate und bivariate Betrachtung	170
9.2	Kovarianz und Produkt-Moment-Korrelation	170
9.3	Rangkorrelation, biseriale Korrelation, Vierfelderkorrelation	176
9.4	Scheinkorrelation, Confounder	176
9.5	Partielle Korrelation	179
9.6	Einfache lineare Regression	180
9.7	Korrelation versus Regression	183
9.8	Übereinstimmungsmaße	184
9.9	Berechnung der Halbwertszeit	186
9.10	Übungsaufgaben	187
<b>10</b>	<b>Überlebenszeitanalyse</b>	<b>188</b>
10.1	Zeit bis ein bestimmtes Ereignis eintritt	188
10.2	Kenngrößen für Sterben und Überleben	189
10.3	Sterbetafel und Überlebenskurve nach der Aktuariatsmethode	191
10.4	Standardisierung	193
10.5	Beispiel zum Vergleich zweier Sterberisiken mit Altersstandardisierung	194
10.6	Zensurierungen (Censoring)	196
10.7	Produkt der Überlebenschancen nach Kaplan & Meier	197
10.8	Interpretation einer zensierten Überlebenskurve	200
10.9	Sterbe- und Überlebensfunktionen	202
10.10	Proportionale Hazards	205
10.11	Übungsaufgaben	209
<b>11</b>	<b>Normalverteilung und andere theoretische Verteilungen</b>	<b>210</b>
11.1	Theoretische Verteilungen versus empirische Häufigkeitsverteilungen	210
11.2	Diskrete Gleichverteilung	211
11.3	Zentraler Grenzwertsatz	211
11.4	Normalverteilung	213
11.5	Standard-Normalverteilung	215
11.6	Sind gewonnene Daten normalverteilt?	220
11.7	Beispiele für die Anwendung der Normalverteilung	222
11.8	Referenzbereiche	224
11.9	Binomialverteilung	225
11.10	Poisson-Verteilung	226
11.11	Übungsaufgaben	228

<b>12</b>	<b>Schätzen</b>	<b>229</b>
12.1	Grundgesamtheit und Stichprobe	229
12.2	Vor- und Nachteile von Stichproben	230
12.3	Stichprobengewinnung	231
12.4	Schätzfunktion und Schätzwert	234
12.5	Eigenschaften von Schätzern	235
12.6	Methoden zur Herleitung von Schätzfunktionen	236
12.7	Ausreißer und robuste Schätzer	237
12.8	Robuste Maße der zentralen Tendenz	239
12.9	Robuste Maße der Streuung	241
12.10	Beispiel zum gestutzten Mittelwert und zur robusten Standardabweichung	243
12.11	Masking, Swamping und Efficacy Loss	245
12.12	Übungsaufgaben	246
<b>13</b>	<b>Konfidenzintervalle</b>	<b>247</b>
13.1	Schätzgenauigkeit	247
13.2	Prinzip des Konfidenzintervalls	247
13.3	Konfidenzintervall des Mittelwerts	249
13.4	Konfidenzintervall des Medians	252
13.5	Konfidenzintervall der Standardabweichung	254
13.6	Konfidenzintervall eines Anteils	256
13.7	Konfidenzintervall des Korrelationskoeffizienten	258
13.8	Darstellung von Konfidenzintervallen	259
13.9	Übungsaufgaben	262
<b>14</b>	<b>Prinzip des statistischen Tests</b>	<b>264</b>
14.1	Der statistische Test ist ein Entscheidungsverfahren	264
14.2	Nullhypothese und die zugehörige Alternative	264
14.3	Der statistische Test ist eine Wahrscheinlichkeitsrechnung	265
14.4	Beispiel für einen einfachen statistischen Test (Einstichprobentest bei normalverteilter Zielgröße)	267
14.5	$\alpha$ -Fehler und $\beta$ -Fehler	268
14.6	Die Power eines statistischen Tests	269
14.7	Wahl des Signifikanzniveaus	272
14.8	Interpretation von Signifikanz und Nicht-Signifikanz	274
14.9	Rechnen eines Tests	275
14.10	Auswahl des Testverfahrens	277
14.11	Übungsaufgaben	279
<b>15</b>	<b>Chi-Quadrat-Test und andere Tests für qualitative Zielgrößen</b>	<b>281</b>
15.1	Allgemeines zum $\chi^2$ -Test	281
15.2	$\chi^2$ -Anpassungstest	282
15.3	Beispiel zum $\chi^2$ -Anpassungstest: Häufigkeit nosokomialer Infektionen	284
15.4	$\chi^2$ -Unabhängigkeitstest	287
15.5	Beispiel zum $\chi^2$ -Unabhängigkeitstest: Rezidivrate nach Operation einer Leistenhernie	288
15.6	Was tun bei kleinen Häufigkeiten?	290
15.7	Exakter Fisher-Test	291
15.8	Binomialtest, Poisson-Test	291
15.9	Übungsaufgaben	293

<b>16</b>	<b>Logrank-Test</b>	<b>296</b>
16.1	Voraussetzungen und Anwendungen	296
16.2	Rechengang	296
16.3	Beispiel: Postoperative Überlebenszeit	297
16.4	Übungsaufgaben	300
<b>17</b>	<b>Rangtests</b>	<b>302</b>
17.1	Voraussetzungen, Anwendungen und Bezeichnungen	302
17.2	Wilcoxon-Test für zwei Parallelgruppen	303
17.3	Wilcoxon-Test für Wertepaare	307
17.4	Kruskal-Wallis-Test	311
17.5	Friedman-Test	312
17.6	Rangtests sind Omnibustests	314
17.7	Exakte und approximative Tests	316
17.8	Bindungskorrektur	317
17.9	Wiederholung für alle Tests	318
17.10	Übungsaufgaben	319
<b>18</b>	<b>t-Test</b>	<b>321</b>
18.1	Normalverteilungsannahme versus verteilungsfreie Auswertung	321
18.2	Idee des t-Tests	321
18.3	t-Test für zwei Parallelgruppen	322
18.4	Beispiel zum t-Test für zwei Parallelgruppen: Geburtsgewicht von Mädchen und Jungen	323
18.5	t-Test für Wertepaare	324
18.6	t-Test für den Korrelationskoeffizienten	326
18.7	Übungsaufgaben	327
<b>19</b>	<b>Einfache Varianzanalyse</b>	<b>328</b>
19.1	Allgemeines zur Varianzanalyse	328
19.2	Einfache Varianzanalyse in anschaulicher Darstellung	330
19.3	Einfache Varianzanalyse in formaler Darstellung	333
19.4	Fixe und zufällige Einflussgrößen	338
19.5	Schätzung von Varianzkomponenten	340
19.6	Übungsaufgaben	343
<b>20</b>	<b>Faktorielle Varianzanalyse</b>	<b>345</b>
20.1	Zwei Einflussgrößen gleichzeitig betrachten	345
20.2	Wechselwirkungen (Interaktionen)	346
20.3	Zweifache Varianzanalyse	348
20.4	Beispiel für eine zweifache Varianzanalyse: Primär- und Erhaltungstherapie	352
20.5	Kombinierte und hierarchische Einflussgrößen	354
20.6	Wiederholungen als Einflussgröße	357
20.7	Verbundene Versuchsanordnung mit mehr als zwei Behandlungen	359
20.8	Drei- und mehrfache Varianzanalyse	361
20.9	Welche Angaben benötigt ein Computerprogramm?	362
20.10	Beispiel Entzündungsparameter nach Operation	363
20.11	Beispiel Schlagvolumen des Herzens vor und nach Bypass-Operation	366
20.12	Beispiel Schwelle des Stapediusreflexes	368
20.13	Beispiel Ringversuch von Laboratorien	370
20.14	Übungsaufgaben	372

<b>21</b>	<b>Multiple Regression</b>	<b>376</b>
21.1	Grundlagen	376
21.2	Multiple lineare Regression für eine stetige Zielgröße	379
21.3	Logistische Regression für eine alternative Zielgröße	381
21.4	Cox-Regression für Überlebenszeiten	384
21.5	Qualitative freie Variable und ihre Codierung	385
21.6	Wechselwirkungen	389
21.7	Multikollinearität	390
21.8	Variablenselektion	390
21.9	Residuen, Güte eines Regressionsmodells (goodness of fit)	392
21.10	Validierung von Regressionsmodellen	395
21.11	Übungsaufgaben	397
<b>22</b>	<b>Multiples Testen</b>	<b>399</b>
22.1	Gespräch zwischen Kliniker und Statistiker	399
22.2	Gedankenexperiment: Wahrscheinlichkeit für mindestens eine Signifikanz	399
22.3	Wie viele Tests sind an einem Datenmaterial möglich?	400
22.4	Wahrscheinlichkeit für mindestens eine Signifikanz bei explorativer Datenanalyse	401
22.5	Lokales versus multiples $\alpha$ -Risiko	403
22.6	Konfirmatorisches und orientierendes Testen	403
22.7	Aufteilen des Signifikanzniveaus nach Bonferroni	405
22.8	Holm-Prozedur	407
22.9	A priori geordnete Hypothesen	411
22.10	Prinzip des Abschluss-tests	412
22.11	Übungsaufgaben	415
<b>23</b>	<b>Allgemeine versus spezielle Hypothesen</b>	<b>418</b>
23.1	Mehrere Zielgrößen	418
23.2	Vergleiche bei mehr als zwei Gruppen	418
23.3	Dunnnett-t-Test, Tukey-Verfahren, Scheffé-Test	419
23.4	Unabhängige Vergleiche (orthogonale Kontraste) bei $\geq 3$ Gruppen oder Zeitpunkten	420
23.5	Beispiel zu orthogonalen Kontrasten	424
23.6	Zeitliche Verläufe	426
23.7	Anzahl der Tests versus Power oder die „Informations-Salami“	428
23.8	Möglichkeiten, die Anzahl der Tests klein zu halten	429
23.9	Anmerkung zu den Freiheitsgraden	430
23.10	Übungsaufgaben	431
<b>24</b>	<b>Äquivalenztests</b>	<b>433</b>
24.1	Fragestellung und zweiseitige Äquivalenz	433
24.2	Zerlegung der Äquivalenz in zwei einseitige Tests auf Unterschiede	435
24.3	Beispiel für einen zweiseitigen Äquivalenztest	437
24.4	Einseitige Äquivalenz, Test auf Nicht-Unterlegenheit	438
24.5	Beispiel für einen Test auf Nicht-Unterlegenheit	440
24.6	Fallzahl bei Äquivalenz	441
24.7	Übungsaufgaben	442
<b>25</b>	<b>Fallzahlbestimmung</b>	<b>444</b>
25.1	Notwendigkeit der Fallzahlplanung	444
25.2	Wovon hängt die erforderliche Fallzahl ab?	445
25.3	Welche Angaben werden benötigt?	447
25.4	Beispiele zur Fallzahlberechnung, Powerberechnung und zur Berechnung des erkennbaren Unterschieds	448

25.5	Sequenzielle Versuchspläne für laufende Auswertung .....	452
25.6	Gruppensequenzielle Auswertungen .....	457
25.7	Eingebaute Pilotstudie .....	459
25.8	Adaptive Fallzahlplanung .....	459
25.9	Fallzahl für Konfidenzintervalle .....	461
25.10	Übungsaufgaben .....	463
<b>26</b>	<b>Fehlende Werte .....</b>	<b>464</b>
26.1	Warum fehlen Beobachtungen und Werte? .....	464
26.2	Muster der fehlenden Werte .....	465
26.3	Konsequenzen fehlender Werte .....	465
26.4	Auswertung ohne Ersetzen der fehlenden Werte .....	467
26.5	Ersetzen fehlender Werte (Imputation) .....	467
26.6	Sensitivitätsanalyse (sensitivity analysis) .....	470
26.7	Übungsaufgaben .....	470
<b>27</b>	<b>Meta-Analysen .....</b>	<b>472</b>
27.1	Zweck .....	472
27.2	Arten von Meta-Analysen .....	473
27.3	Plan einer Meta-Analyse .....	474
27.4	Medizinische Kriterien für die in die Meta-Analyse aufzunehmenden Studien .....	475
27.5	Methodische Kriterien für die in die Meta-Analyse aufzunehmenden Studien .....	476
27.6	Literatursuche .....	476
27.7	Bewertung der Aussagekraft der selektierten Studien .....	477
27.8	Statistische Methoden zur Zusammenfassung der Studienergebnisse .....	478
27.9	Darstellung der Ergebnisse einer Meta-Analyse .....	482
27.10	Publication Bias .....	483
27.11	Interpretation, Aussagekraft und Grenzen von Meta-Analysen .....	486
27.12	Übungsaufgaben .....	487
<b>28</b>	<b>Medizinische Statistik — Mathematik oder Orakel? .....</b>	<b>488</b>
28.1	Verschiedene Wahrheiten? .....	488
28.2	Fehlerquellen .....	489
28.3	Fahrlässigkeit und Fälschung .....	490
28.4	Der gute Verkäufer: Anpreisen der Vorzüge, Verschleiern der Mängel .....	491
28.5	Simpsons Paradoxon .....	492
28.6	Beispiele für tückische Interpretation .....	495
28.7	Horoskope und Orakel .....	499
28.8	Medizin und Mathematik .....	500
28.9	Und die Moral von der Geschichte .....	501
<b>29</b>	<b>Weitere Aufgaben .....</b>	<b>502</b>
29.1	Infektionsalarm? .....	502
29.2	Kongressvortrag .....	502
29.3	Sex-Ratio in einem Isolat .....	503
29.4	Blutdrucksenkung mit drei Dosen .....	503
29.5	Zielgrößen bei rheumatoider Arthritis .....	504
29.6	Verzögerung des Wiederanstiegs des intraokularen Drucks .....	504
29.7	Zielgröße und Auswertung bei Morbus Huntington .....	505
29.8	Schädigung von Leberzellen durch Ischämie oder durch Re-Perfusion? .....	506

29.9	Kinder mit nur einer funktionsfähigen Niere .....	506
29.10	Nachhaltige Wirksamkeit einer Kneipp- und einer Schroth-Kur bei essentieller Hypertonie .....	506
29.11	Rauschgiftkonsum zu Lebzeiten .....	507
29.12	Denksport beim Würfelspiel .....	508
<b>30</b>	<b>Mathematische Fachausdrücke und Formeln verstehen .....</b>	<b>509</b>
<b>31</b>	<b>Statistiksoftware .....</b>	<b>515</b>
<b>32</b>	<b>Lösungen zu den Übungsaufgaben .....</b>	<b>517</b>
<b>33</b>	<b>Danksagung und persönliche Worte .....</b>	<b>580</b>
<b>34</b>	<b>Literaturhinweise .....</b>	<b>582</b>
<b>35</b>	<b>Sachwortregister .....</b>	<b>590</b>

# 1 Statistische Vergleichbarkeit

## 1.1 Wiederholen und Vergleichen

**Gegenstand der Statistik.** Die Statistik beschäftigt sich mit Dingen, die manchmal, gelegentlich, häufig oder selten eintreten. In der Statistik werden solche Ereignisse als **stochastische, als zufällige Ereignisse bezeichnet**. Zufällige Ereignisse müssen nicht rein zufällig sein, wie z.B. ein Münzwurf, sie müssen nur zumindest etwas Zufall enthalten. Weiterhin beschäftigt es die Statistik, wenn unter den gleichen Bedingungen mal größere, mal kleinere Werte erreicht werden. Entsteht unter den gleichen Bedingungen immer das gleiche Ergebnis, der gleiche Messwert, so sind das **determinierte Ereignisse** und die Statistik ist unnötig. Ebenso beschäftigt sich die Statistik nicht mit Dingen, die bei gegebenen Bedingungen niemals eintreten (**unmögliche Ereignisse**) oder immer eintreten (**sichere Ereignisse**).

**Fälle und Fallzahl.** Die Statistik beschäftigt sich also mit Dingen, die variieren. Um trotz dieser Variation eine Aussage über Vorgänge oder Beobachtungen machen zu können, genügt es nicht, einmalig zu beobachten oder einmalig zu messen; vielmehr müssen unter den möglichst gleichen Bedingungen bei möglichst vielen Fällen Befunde erhoben und Messungen durchgeführt werden. Fälle können **Patienten, Probanden** (das sind gesunde **Versuchspersonen**), **Versuchstiere** oder sonstige **Beobachtungseinheiten** oder **Versuchseinheiten** sein. Die Anzahl der Fälle, bei denen diese gleichen Untersuchungen durchgeführt wurden, heißt statistische Fallzahl. Die Termini „Fall“ und „Fallzahl“ sind historisch gewachsen, aus Höflichkeit sprechen wir meist von Patienten, meinen damit aber alle Arten von Fällen.

**Gruppe.** Die Fälle, die unter den gleichen Bedingungen beobachtet oder gemessen worden sind, bilden eine Gruppe. Als Ergebnis ist interessant, wo die meisten Daten liegen, was durch den Mittelwert ausgedrückt werden kann. Interessant ist aber auch, wie sehr die Beobachtungen oder Messungen streuen, d.h. wie sehr sie innerhalb einer Gruppe voneinander abweichen.

**Mehrere Gruppen.** Es kann durchaus interessant sein, eine einzelne Gruppe zu beobachten und die gewonnenen Daten auszuwerten. Meist wird es aber erst dann richtig spannend, wenn zwei oder mehr Gruppen beobachtet und die Ergebnisse der Gruppen miteinander verglichen werden. Meist interessiert besonders die Gruppe, die einer speziellen Behandlung unterzogen wurde, sie wird als **Untersuchungsgruppe** bezeichnet. Die andere Gruppe ohne die spezielle Behandlung ist dann die **Vergleichsgruppe**. Natürlich können auch mehr als zwei Gruppen miteinander verglichen werden. Gelegentlich ist die Einteilung in Untersuchungs- und Vergleichsgruppe unzweckmäßig, weil alle Gruppen gleichermaßen interessieren (mehrere Untersuchungsgruppen).

## 1.2 Unabhängigkeit der Beobachtungen

**Der statistische Fall.** Für jede Statistik ist die Unabhängigkeit der Beobachtungen von entscheidender Bedeutung. In der medizinischen Statistik ist typischerweise jeder „statistische Fall“ ein eigener, separater Patient beziehungsweise ein eigenes, separates Versuchstier. Dann ist die Unabhängigkeit der Beobachtungen gewährleistet. Statistiken, bei denen die Fallzahl größer ist als die Anzahl der Patienten oder der Versuchstiere – weil der gleiche Patient oder das gleiche Versuchstier mehrmals in der Statistik eingeht – sind heikel und mit Vorsicht zu betrachten.

**Vorher-nachher-Vergleiche und zeitliche Verläufe.** Natürlich kann es sehr wohl sinnvoll sein, denselben Patienten (oder dasselbe Versuchstier) mehrmals zu beobachten und zu untersuchen, um den zeitlichen Verlauf zu ermitteln, um Werte vor und nach der Therapie miteinander zu vergleichen, um Veränderungen festzustellen und vieles andere mehr. Trotzdem: Die Anzahl der Fälle ist die Anzahl der beobachteten Patienten. **Beispiel Wachstumskurve:** An 200 neugeborenen, gesunden Mädchen wird im ersten Lebensjahr monatlich, in den folgenden 5 Jahren alle 3 Monate das Körpergewicht bestimmt, um daraus Wachstumskurven zu gewinnen. Damit wird jedes Mädchen 32-mal gewogen. Insgesamt werden somit bei 200 Mädchen je 32 Werte = 6 400 Werte gewonnen. Die Fallzahl ist aber 200.

**Merkmale je Fall.** Ebenso selbstverständlich können wir von jedem Studienpatienten (oder jedem Versuchstier) vielerlei Merkmale erheben, z.B. Geschlecht, Alter, Körpergröße, Körpergewicht, Diagnose und vieles mehr. Die Anzahl der Daten kann weit größer sein als die Fallzahl.

**Beispiel Körpergröße.** Zu ermitteln ist die derzeitige Körpergröße gesunder Frauen im Alter von 18 Jahren. Offensichtlich ist es völlig unsinnig, bei der 18-jährigen Karin Gernegroß die Körpergröße eintausendmal zu messen (also an einer einzigen Person 1000 statistische Fälle zu erzeugen) und den daraus resultierenden Mittelwert mit  $n = 1000$  zu beschriften. Um etwas über die Körpergröße heutiger 18-jähriger Frauen zu erfahren, muss man schon 1000 Frauen im Alter von 18 Jahren je einmal messen.

**Beispiel Körpergröße morgens und abends.** Will man dagegen untersuchen, ob die Körpergröße abends kleiner ist als morgens (weil die Last des Tages auf jedem ruht, und aus den Knorpeln und Bandscheiben die Flüssigkeit entweicht), so wird an z.B.  $n = 100$  verschiedenen Personen sowohl morgens als auch abends die Körpergröße gemessen (jede Person, jeder Fall liefert also ein Wertepaar) und daraus die intraindividuelle Differenz bestimmt. Die geeignete Statistik ist der Mittelwert der Differenzen. Dieser Mittelwert der Abend-Morgen-Differenzen der Körpergröße beruht auf 100 statistischen Fällen (Personen) – einhundert Mal wird die gleiche Messmethode wiederholt, obwohl insgesamt 200 Werte vorliegen.

**Beispiel Blendungsempfindlichkeit.** Ein drittes Beispiel zur Fallzahl ist die Blendungsempfindlichkeit des rechten und linken Auges, die an z.B. 50 Medizinstudenten bestimmt wurde. Dabei wurde die Blendungsempfindlichkeit nur nach „stark“, „mäßig“ und „gering“ bewertet. Siehe **Box 1.1**.

### 1.3 Systematische und zufällige Fehler

**Ungenauigkeit bei Messwerten.** Messwerte können – mehr oder weniger – von Messung zu Messung schwanken, d.h. einen zufälligen Fehler haben. **Zufallsfehler** entstehen weil die Messbedingungen nicht völlig konstant gehalten werden können.

**Unsicherheit bei medizinischen Befunden.** Auch medizinische Befunde, der erreichte Therapieerfolg, also nahezu alle medizinische Daten sind mit einer gewissen Unsicherheit behaftet. Diese Unsicherheit wird **Zufallskomponente** genannt. Diese Zufallskomponente entsteht durch biologische Variabilität und weil viele medizinische Größen nur indirekt bestimmt werden können. Hinzu kommen aber auch die Tagesform des Patienten, der Einfluss der Situation, in der der Befund erhoben wird, nicht konstante Messbedingungen, Subjektivität des Arztes usw. bis hin zum Lächeln der Krankenschwester.

**Einmaligkeit jedes Menschen.** Das Lehrbuchwissen der Medizin gilt für bestimmte Kollektive von Patienten, z.B. für Typ-2-Diabetiker. Für den einzelnen Patienten, auch wenn er zu dem Kollektiv gehört, gilt es nur mehr oder weniger, einfach weil jeder Mensch biologisch einmalig ist und auch einen ganz persönlichen, einmaligen Lebensweg hat. Die dadurch bedingten Abweichungen des einzelnen Individuums vom Kollektiv werden in der Statistik als **Zufallsfehler** bezeichnet und zusammengefasst.

Ein **systematischer Fehler** entsteht, wenn ein Messfehler eine bestimmte Tendenz hat, z.B. wenn das Messgerät falsch eingestellt ist und generell etwas zu große Werte oder generell etwas zu kleine Werte liefert. Systematische Fehler können durch fehlende Eichung, falsche Kalibrierung, ungeeignete Geräte, falsche Materialien und viele andere Ursachen auftreten. Ein systematischer Fehler entsteht z.B. auch, wenn die Körpertemperatur unter der Achsel gemessen und als Kerntemperatur betrachtet wird, weil unter der Achsel eine etwas geringere Temperatur herrscht als im Rektum. Ist die Größe eines systematischen Fehlers bekannt, so kann man ihn korrigieren. Die Beobachtung und Untersuchung bestimmter, selektierter Patienten kann ebenfalls zu systematischen Fehlern führen.

**Zufällige versus systematische Fehler.** Zufallsfehler haben keine Tendenz. Auf lange Sicht und bei vielen Messungen gleicht sich der Zufallsfehler aus, im Mittel ist der Zufallsfehler null. Nicht dagegen systematische Fehler. **Ein systematischer Fehler verfälscht den Mittelwert, ein zufälliger Fehler erhöht die Streuung der Daten (Box 1.2).** Zufallsfehler sind im Prinzip unvermeidlich, man kann sich nur bemühen, sie möglichst klein zu halten. Im Gegensatz sollten systematische Fehler möglichst vermieden werden, weil sie die Ergebnisse verfälschen.

### Box 1.1: Blendungsempfindlichkeit bei Medizinstudenten

#### a) unvollständige Darstellung und irreführende Fallzahl

	Rechtes Auge	Linkes Auge	Summe
<b>Stark</b>	2	2	4
<b>Mäßig</b>	9	7	16
<b>Gering</b>	39	41	80
<b>Summe</b>	50	50	100

Diese **Darstellung a)** liefert nur einen Teil der gewonnenen Information und beruht auf Augen, nicht jedoch auf Personen. Vermutlich ist die Blendungsempfindlichkeit des rechten und linken Auges einer Person nicht unabhängig voneinander. Deshalb wird mit  $n = 100$  Augen eine zu große Fallzahl vorgetäuscht.

#### b) vollständige Darstellung mit korrekter Fallzahl

Blendungsempfindlichkeit		Rechtes Auge			Summe
		Stark	Mäßig	Gering	
Linkes Auge	Stark	1	1	0	2
	Mäßig	0	4	3	7
	Gering	1	4	36	41
Summe		2	9	39	50

Die **Darstellung b)** liefert zusätzlich die Information, wie gut rechtes und linkes Auge übereinstimmen. Außerdem beruht diese Darstellung auf Personen, nicht auf Augen und ist somit bei der Fallzahl ehrlich.

Die **Statistik** ist die Wissenschaft, die sich mit den Zufallsfehlern beschäftigt. Mit **Wiederholungen** und statistischen Verfahren kann man die wahren Werte schätzen und auch die Genauigkeit dieser Schätzungen abschätzen. Schließlich kann man mit statistischen Tests entscheiden, ob die beobachteten Unterschiede zwischen den Gruppen zufallsbedingt sein können (nicht aber, ob sie zufallsbedingt sind). All das wird noch ausführlich behandelt. Während die Statistiker sehr geübt im Umgang mit Zufallsfehlern sind, können systematische Fehler durch statistische Verfahren nicht (oder fast nicht) erkannt und eliminiert werden. Deshalb muss größter Wert darauf gelegt werden, mit Sachverstand systematische Fehler von vornherein zu vermeiden.

**Sachliche Fehler.** Neben dem systematischen und dem zufälligen Fehler gibt es noch falsche Entscheidungen, fehlerhafte Rechnungen (wie z.B.  $2 + 2 = 5$ ) usw., die als sachliche Fehler oder im Spaß als „falsche Fehler“ bezeichnet werden können. Umgangssprachlich bezieht sich Fehler meist auf diese sachlichen Fehler.

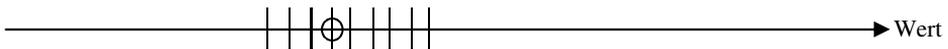
**Error, bias, mistake.** Systematische, zufällige und sachliche Fehler sind sorgfältig zu unterscheiden. Im Englischen wird diese Unterscheidung schon durch die Sprache erzwungen:

Error = zufälliger Fehler

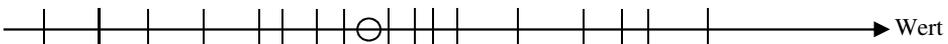
Bias = systematischer Fehler

Mistake = sachlicher Fehler.

### Box 1.2: Systematischer und zufälliger Fehler



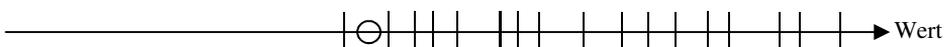
ohne systematischen Fehler, kleiner Zufallsfehler  
= richtig und präzise



ohne systematischen Fehler, großer Zufallsfehler  
= richtig, aber wenig präzise



mit systematischem Fehler, kleiner Zufallsfehler  
= nicht richtig, aber präzise



mit systematischen Fehler, großer Zufallsfehler  
= weder richtig, noch präzise

## 1.4 Eine Vergleichsgruppe ist notwendig

**Therapieerfolg unter Placebo.** Placebo ist ein Arzneimittel, das keine wirksame Substanz enthält. Trotzdem kann unter Placebobehandlung ein nennenswerter Therapieerfolg erreicht werden. Dies kann vielerlei Ursachen haben, die wichtigsten davon sind:

- Die **Selbtheilungskräfte** des Körpers erreichen eine Besserung oder Spontanheilung.
- Die Krankheit hatte bei Therapiebeginn ihren **Höhepunkt schon erreicht oder überschritten**.
- Durch den **Wegfall des Krankheitsauslösers** (der **Noxe**) kann ohne spezifische Arzneimitteltherapie ein guter Therapieerfolg erreicht werden. Wird z.B. bei neu aufgetretener Hypertonie (Bluthochdruck) ein Medikament verordnet und die Patientin versöhnt sich kurz nach Beginn der Behandlung mit ihrem Mann, so kann sich auch bei Placebothherapie ein guter Therapieerfolg einstellen.
- **Wegfall der Arbeits- und Alltagslast.** Durch den Status als Patient muss man nicht zur Arbeit oder zur Schule gehen; damit entfallen auch der Ärger mit dem Chef und die Angst vor der Klassenarbeit. Stattdessen erhält der Patient **menschliche Zuwendung** des Pflegepersonals, des Arztes, seiner Familie und seiner Freunde.
- Eine Therapie besteht auch aus vielerlei **unspezifischen Maßnahmen**, wie z.B. Bettruhe, Pflege, Umsetzung in eine schöne und erholsame Umwelt.
- Schließlich trägt auch der Glaube des Patienten an das Medikament, an den Arzt und an die Medizin zum Therapieerfolg bei. Dies nennt man den **Placeboeffekt**. Bei subjektiven oder subjektiv beeinflussbaren Zielgrößen kann ein erheblicher Placeboeffekt auftreten.

Spitzfindig betrachtet und bissig formuliert weiß man im Einzelfall nicht immer, ob der Patient wegen oder trotz ärztlicher Behandlung geheilt wurde (oder verstorben ist).

„**Wer heilt hat Recht**“ wird oft als Nachweis der Wirksamkeit einer Therapie benutzt. Allerdings ist der Spruch irreführend, weil ein Erfolg aus vielerlei Gründen eingetreten sein kann. Nur wer besser heilt als Placebo hat ein wirksames Arzneimittel oder wer besser heilt als die Standardtherapie hat eine überlegene Behandlung.

**Beispiel Therapieerfolg bei Pollinosis** (Heuschnupfen). Der Doktorand Franz Unbesorgt will ein phytomedizinisches (aus Pflanzen hergestelltes) Arzneimittel gegen Heuschnupfen erproben. Er entwickelt einen Erhebungsbogen und gewinnt für die Durchführung zehn niedergelassene Allgemeinärzte. In diesen zehn Praxen erhalten alle Patienten mit der klaren Diagnose Pollinosis das zu beurteilende Mittel. Die Studienpatienten werden 14 Tage später zum Therapieerfolg befragt. Bei 60 der 83 Patienten, das sind 72%, zeigt sich ein guter Therapieerfolg. Der Doktorand Unbesorgt berichtet: „Bei Heuschnupfen hilft das phytomedizinische Arzneimittel bei fast  $\frac{3}{4}$  der Patienten.“ Der böse Statistiker „schießt den Doktoranden ab“ und behauptet, es sei völlig unklar, welchen Anteil die Studienmedikation an dem beobachteten Therapieerfolg habe. Es könne ja sein, dass während der Behandlung ein Landregen die Pollen aus der Luft ausgewaschen habe oder die Blütezeit vieler Allergene zu Ende gegangen sei. Er empfiehlt eine neue Studie, die sowohl eine Verum- als auch eine Placebogruppe umfasst.

**Wirksamkeit und Überlegenheit.** Für den Nachweis der Wirksamkeit oder Überlegenheit einer Therapie ist eine **Vergleichsgruppe** unabdingbar. Erst wenn der Therapieerfolg unter der zu beurteilenden Medikation den Therapieerfolg unter Placebo übersteigt, ist die Wirksamkeit der Studienmedikation nachgewiesen. Der Nachweis der Überlegenheit einer neuen Therapie ist erst gegeben, wenn der Therapieerfolg in der Gruppe mit der neuen Therapie den Erfolg in der mit der bisherigen Standardtherapie behandelten Vergleichsgruppe übersteigt.

## 1.5 Vermengte Effekte

**Beispiel Schwangerschaftsgymnastik und Vitalstatus des Neugeborenen.** Oberarzt Dr. Gründlich will sich über den Nutzen der Schwangerschaftsgymnastik habilitieren. Er erhebt bei allen Geburten eines Jahres, ob die Gebärende an einer Schwangerschaftsgymnastik teilgenommen hat oder nicht; weiter erhebt er das Alter, die Parität (die wievielte Schwangerschaft) und einige andere Daten der Mutter sowie den Vitalstatus des Neugeborenen. Der Vitalstatus der Neugeborenen wird ohnehin erhoben nach einer von **Virginia Apgar** (1909 bis 1974) eingeführten Skala. Der **Apgar-Score** variiert zwischen 0 = Todgeburt und 10 = voll vital. Für 1 191 Erstgebärende ergibt die Auswertung:

		Apgar-Score 10 oder 9
Schwangerschafts- gymnastik	ja	405 von 473 = 85.6%
	nein	533 von 718 = 74.2%

Dr. Gründlich interpretiert: „Bei erstgebärenden Frauen verbessert die Teilnahme an einer Schwangerschaftsgymnastik den Vitalstatus des Neugeborenen.“ Der Statistiker kontert: „Vielleicht!“

**Begründung des Statistikers.** Wenn sich Gruppen in mehr als einer Hinsicht unterscheiden, so nennt man das vermengte Effekte. Im Beispiel der Schwangerschaftsgymnastik und dem Vitalstatus des Neugeborenen sind viele Effekte vermengt. Frauen, die spontan an einer Schwangerschaftsgymnastik teilnehmen, sind meist gesundheitsbewusster, freuen sich besonders auf ihr Kind, rauchen weniger, trinken weniger Alkohol, ernähren sich vernünftiger und nehmen mehr Rücksicht auf ihre Schwangerschaft. Der von Dr. Gründlich gefundene Unterschied im Vitalstatus der Neugeborenen wird nicht bestritten, er kann aber viele Gründe haben und nur die Götter wissen, ob die Teilnahme an einer Schwangerschaftsgymnastik den Vitalstatus des Neugeborenen wirklich verbessert oder nicht.

**Interpretation eines Effekts.** Bei vermengten Effekten ist eine Aussage, worauf der beim Erfolg beobachtete Unterschied beruht, nicht möglich. Wird z.B. in einer Patientengruppe A, die mit der Therapie A behandelt worden ist, ein besserer Therapieerfolg beobachtet als in der Patientengruppe B, die mit der Therapie B behandelt worden ist, so nützt diese Beobachtung nichts, wenn sich die Gruppen außer in der Therapie auch noch z.B. in der Altersstruktur, dem Schweregrad der Erkrankung oder in der Dauer der Anamnese (Vorgeschichte der Krankheit) unterscheiden. Der bessere Therapieerfolg in der Gruppe A kann durch Therapie A bedingt sein, aber auch durch den geringeren Schweregrad der Erkrankung bei diesen Patienten oder durch eine kürzere oder längere Krankheitsdauer (**Box 1.3**). Ein unterschiedliches Ergebnis kann nur interpretiert werden, wenn sich die Gruppen nur in einer einzigen Hinsicht unterscheiden, d.h. wenn keine vermengten Effekte vorliegen.

**Interpretation kein Effekt.** Wenn bei Gruppen mit verschiedenen Therapien und weiteren Unterschieden, d.h. bei vermengten Effekten kein Unterschied im Erfolg beobachtet wird, so ist nicht gewährleistet, dass die Therapien gleichwertig sind. Es ist durchaus denkbar, dass z.B. Therapie A besser ist, aber die Patienten, die diese Therapie erhalten haben, im Mittel älter waren, und die bessere Wirksamkeit der Therapie A durch das höhere Alter kompensiert wird, d.h. sich die Wirkung beider Effekte auf den Therapieerfolg weitgehend ausgleicht. Bei vermengten Effekten kann weder ein unterschiedlicher Erfolg in den Gruppen noch ein gleichartiger Erfolg in den Gruppen sachlich interpretiert werden. Eine Interpretation des Ergebnisses ist nur dann möglich, wenn die Gruppen vergleichbar sind, d.h. wenn sie im statistischen Sinne gleich sind und damit keine vermengten Effekte enthalten.

## 1.6 Struktur-, Behandlungs- und Beobachtungsgleichheit

**Zufällig** bedeutet im allgemeinen Sprachgebrauch, dass sich irgendetwas ungeplant ereignet hat oder dass man ungezielt etwas getan hat und sich dabei nichts Besonderes gedacht hat. Beispiele: Zufällig habe ich Moritz in der Stadt getroffen. Mir ist zufällig ein altes Foto von Max in die Hände gefallen.

**Streng zufällig.** In der Statistik wird „zufällig“ sehr viel strenger gebraucht. Werden z.B. aus einer Herde von 30 Schafen ohne einen besonderen Plan 5 Tiere eingefangen, so ist dieses keine zufällige Auswahl im statistischen Sinne. Möglicherweise hat man bevorzugt die kranken, langsamen und zutraulichen Schafe eingefangen und die gesunden, flinken und ängstlichen Tiere in der Herde belassen. Streng zufällig bedeutet, dass ein **Zufallsmechanismus** angewandt wird. Gängige Zufallsmechanismen sind der **Münzwurf** mit den Zufallsereignissen „Wappen“ und „Zahl“ und das **Würfeln** der Zufallszahlen 1 bis 6. In der ernsthaften Statistik wird jedoch mit Computerprogrammen gearbeitet, die Zufallszahlen erzeugen (Kap.1.8, S.27).

**Statistische Gleichheit.** Das Wort „gleich“ hat in der Mathematik und in der Statistik unterschiedliche Bedeutung. Gleich bedeutet in der Mathematik „**genau gleich**“. In der Statistik jedoch heißt gleich „**möglichst wenig, jedoch höchstens zufällig**“. Dieses Buch ist ein Statistik-Buch, daher wird „gleich“ im statistischen Sinn verwendet. Die mathematische Gleichheit wird als „**genau gleich**“ bezeichnet. Die statistische Gleichheit, dieses „möglichst wenig, jedoch höchstens zufällig“ gilt in dreifacher Hinsicht: Strukturgleichheit, Behandlungsgleichheit und Beobachtungsgleichheit.

**Strukturgleichheit** bedeutet, dass sich die Gruppen in ihrer Zusammensetzung möglichst wenig, jedoch höchstens zufällig unterscheiden. Alle möglichen Störfaktoren wie z.B. Alter, Geschlecht, genaue Diagnose, Krankheitsdauer, Schweregrad der Krankheit usw. unterscheiden sich bei Studienbeginn in den Gruppen höchstens zufällig.

### Box 1.3: Vermengte Effekte z.B. Schweregrad und Therapie

Schweregrad einer Verletzung:	leicht	mittel	schwer
	↓	↓	↓
Therapie:	A	B	C
	↓	↓	↓
Überlebenszeit	lang	mittel	kurz

Bei einer bestimmten Art der Verletzung wird bei geringem Schweregrad Therapie A angewandt, bei mittlerem Schweregrad die Therapie B und bei hohem Schweregrad die Therapie C. Stellt sich bei der Nachbeobachtung heraus, dass die Patienten mit der Therapie A lange überleben, Patienten mit Therapie B mäßig lang überleben und Patienten mit Therapie C nur kurz überleben, so ist nicht nachgewiesen, dass Therapie A die beste und Therapie C die schlechteste ist. Bei diesem Vorgehen sind Schweregrad und Therapie vermengt. Die lange Überlebenszeit kann an der Therapie A, aber auch an dem geringen Schweregrad liegen. Ebenso kann die kurze Überlebenszeit durch Therapie C, aber auch durch den hohen Schweregrad bedingt sein. **Fazit:** Es ist unklar, ob ein langes Überleben mehr durch die Therapie A oder mehr durch den geringen Schweregrad bedingt ist.

**Behandlungsgleichheit.** Nicht nur alle Fälle einer Gruppe, auch die Fälle der verschiedenen Gruppen müssen im statistischen Sinne gleich behandelt werden, ausgenommen die Einflussgröße, die untersucht werden soll. **Beispiel:** Wird die Wirkung von zwei Arzneimittelvarianten vergleichend beurteilt, so erhalten alle Patienten in allen Gruppen die gleiche supportive Therapie, die gleiche Pflege, die gleiche Zuwendung usw. Alle diese Komponenten bis hin zum Lächeln der Krankenschwester – wir übertreiben nur geringfügig – dürfen sich zwischen den Gruppen höchstens zufällig unterscheiden.

**Beobachtungsgleichheit.** Alle Patienten, Probanden, Versuchstiere müssen gleich intensiv beobachtet werden. Sowohl Therapieerfolg als auch Therapiemisserfolg als auch unerwünschte Ereignisse müssen mit gleichem Ernst und gleicher Sorgfalt in allen Gruppen erhoben werden. Dies gilt insbesondere dann, wenn die Vergleichsgruppe keine ernsthafte Therapie erhält. Naheliegend ist, dass in der Gruppe, die intensiver beobachtet wird, auch mehr unerwünschte Ereignisse erfasst und dokumentiert werden, einfach weil die Patienten häufiger die Möglichkeit haben, über unerwünschte Ereignisse zu berichten.

**Notwendigkeit der Versuchsplanung.** Nur wenn in einer Studie Struktur-, Behandlungs- und Beobachtungsgleichheit erreicht wird, lassen sich die Ergebnisse schlüssig interpretieren. Durch Versuchsplanung wird diese Struktur-, Behandlungs- und Beobachtungsgleichheit erreicht, und es werden damit vermengte Effekte vermieden.

**Wie wird Strukturgleichheit erreicht?** Die vier wichtigsten Maßnahmen, um Strukturgleichheit zu erreichen, sind Schichtung, matched Pairs, verbundene Versuchsanordnung und Randomisation. Dabei werden die drei Maßnahmen Schichtung, matched Pairs und verbundene Versuchsanordnung zusammen genommen auch als **Blockbildung** bezeichnet. Blockbildung bewirkt, dass sich die Gruppen möglichst wenig unterscheiden. Randomisation bewirkt, dass sich die Gruppen höchstens zufällig unterscheiden.

## 1.7 Schichtung und matched Pairs

**Schichtung (stratification).** Wenn bekannt ist, dass eine bestimmte messbare Störgröße den Therapieerfolg erheblich beeinflusst, dann will man es nicht dem Zufall überlassen, ob diese Störgröße in allen Gruppen gleich stark auftritt. Vielmehr wird man die Zuteilung der Patienten zu den Gruppen so steuern, dass diese Störgröße (Begleitvariable) alle Gruppen möglichst gleichmäßig betrifft.

**Beispiel für Schichtung nach Geschlecht.** Es ist bekannt, dass auf bestimmte Naturheilverfahren Frauen besser ansprechen als Männer. Der Therapieerfolg in einer Gruppe hängt also erheblich vom Anteil der Frauen in der Gruppe ab. Nun wird man den Anteil der Frauen in den Gruppen nicht dem Zufall überlassen, sondern die Zuteilung der Patienten in die Gruppen so steuern, dass der Anteil der Frauen in den Gruppen genau gleich, z.B. 75% ist. Der Statistiker sagt: „Es wurde nach Geschlecht geschichtet.“ Angenommen für eine Studie mit den drei Gruppen A, B, C stehen 90 Frauen und 30 Männer zur Verfügung, so teilt man die 90 Frauen auf die drei Gruppen auf und ebenso die 30 Männer, so dass dann jede Gruppe genau 30 Frauen und genau 10 Männer = 40 Fälle hat.

**Schichtungskriterien.** Durch Schichten wird eine wichtige und bekannte Störgröße gleichmäßig auf die Gruppen geteilt. Eine Studie kann auch nach mehreren Kriterien geschichtet werden. Schichtungskriterien sind häufig das Geschlecht, die Altersklasse, der Schweregrad oder das Stadium der Erkrankung, die genaue Diagnose, die behandelnde Klinik.

**Beispiel für zwei Schichtungskriterien.** In einer Studie zur akuten Leukämie sollen die Patienten nach den drei Diagnosen akute lymphatische Leukämie (ALL), akute myeloische Leukämie (AML) und akute undifferenzierte Leukämie (AUL) geschichtet werden, ebenso ob sie aktuell eine Infektion haben oder nicht. Aus den drei Diagnosen und dem Kriterium mit/ohne Infektion entstehen insgesamt  $3 * 2 = 6$  Schichten: ALL mit Infektion, ALL ohne Infektion, AML mit Infektion, AML ohne Infektion, AUL mit Infektion, AUL ohne Infektion.

**Anzahl der Schichten.** Die Gesamtzahl der Schichten ist das Produkt aus der Anzahl der Ausprägungen der Schichtungskriterien. Bei mehreren Schichtungskriterien entsteht somit schnell eine große Anzahl von Schichten. Die Gesamtzahl der Schichten muss deutlich kleiner sein als die Anzahl der Patienten je Gruppe, weil sonst die Schichtung nicht mehr wirkt. Beispiel: Bei 3 Gruppen mit je 25 Patienten und 6 Schichten sind im Mittel von jeder Gruppe  $25/6 = \text{ca. } 4$  Patienten in einer Schicht. Die am schwächsten besetzte Schicht sollte mindestens einen Patienten von jeder Gruppe haben.

**Matched Pairs.** Ideal wären Studien mit eineiigen Zwillingen mit gleichem Lebensweg und in der gleichen Umwelt, wobei der eine Zwilling die eine, der andere Zwilling die andere Therapie erhält. Diese Situation lässt sich nachbilden, indem zwei Personen, die in vielerlei Aspekten ähnlich sind, als ein **statistisches Zwillingpaar** betrachtet werden. Solche matched Pairs sind hilfreich, wenn es nur wenige spezielle Fälle, aber viele Vergleichspersonen gibt. Dann sucht man zu jedem „Fall“ eine passende „Kontrollperson“. Die Kontrollen werden nach vorab festgesetzten **Zuordnungskriterien** ausgesucht. Der „Fall“ und der im zugeordnete „Zwilling“ bilden ein „statistisches Paar“. Manchmal werden einem speziellen Fall auch mehrere „Kontrollen“ zugeordnet. Matched Pairs werden in Fall-Kontroll-Studien (Kap.2.8, S.47) angewandt und um eine Vergleichsgruppe zu bilden.

**Beispiel für die Bildung von matched Pairs.** Zuordnungskriterien sind gleiches Geschlecht, gleiche Altersklasse, gleiche Diagnose und ein möglichst gleichzeitiger Behandlungsbeginn. Ein Mann der Altersklasse  $> 40$  bis  $\leq 50$  Jahre mit der Diagnose geschlossene Oberschenkelfraktur ist ein „Fall“ von verzögerter Frakturheilung. Zu diesem wird eine „Kontrolle“ gesucht. Kontrollen sind Patienten mit gleichartiger Fraktur, aber nicht verzögerter Frakturheilung. Deshalb wird von allen männlichen Patienten der Altersklasse  $> 40$  bis  $\leq 50$  Jahre und geschlossener Oberschenkelfraktur und ohne verzögerte Frakturheilung derjenige als Kontrolle ausgewählt, dessen Behandlungsbeginn dem Fall am nächsten kommt. Bei der Auswertung wird gesucht, in welcher Hinsicht sich Fälle und Kontrollen sonst noch unterscheiden, um Hinweise auf die Ursachen für verzögerte Frakturheilung zu finden.

## 1.8 Randomisation

**Zufallszahlengeneratoren.** Der Wurf einer Münze und eines Würfels oder das Ziehen einer Kugel aus einer Urne sind zwar Zufallszahlengeneratoren, sie sind aber schlecht nachprüfbar. Außerdem kann die Münze oder der Würfel unterschiedlich gerundete Kanten oder sonstige Abweichungen vom Ideal haben. Deshalb verwendet man als Zufallszahlengenerator ein Computerprogramm, das Zufallszahlen liefert. Die meisten Zufallszahlengeneratoren liefern **reelle Zahlen** (Zahlen mit Dezimalpunkt) zwischen null und eins. Will man Zufallszahlen z.B. zwischen 90 und 110 haben, so wird jede vom Zufallszahlengenerator gelieferte Zahl mit 20 multipliziert und dann noch 90 addiert.

**Für mathematisch Interessierte.** Zufallszahlengeneratoren erzeugen nur Pseudo-Zufallszahlen. Kennt man das Programm und die Startzahl, so lassen sich die Pseudo-Zufallszahlen vorhersagen, echte Zufallszahlen lassen sich nicht vorhersagen. Außerdem kommt nach einer bestimmten Anzahl von

Zufallszahlen, die bei guten Generatoren allerdings extrem groß ist, die gleiche Zahlenfolge erneut. Diese formalen Einschränkungen sind aber für die Praxis der klinischen Studien bedeutungslos.

**Zufallsauswahl.** Aus einer Menge werden Individuen streng zufällig ausgewählt (= selektiert) und dann befragt oder untersucht. Bei dem in Kapitel 1.6, S.25 erwähnten Beispiel mit der Schafherde werden alle 30 Schafe nummeriert, für jedes Tier wird ein Zettel mit seiner Nummer in eine Urne gegeben und dann 5 Zettel und damit Tiere gezogen. Anstatt dieser „Mini-Lotterie“ kann für jedes Tier eine Zufallszahl gezogen werden und die 5 Tiere mit der kleinsten Zufallszahl haben „den Kürzeren gezogen“ und werden herausgenommen. Allerdings ist bei klinischen Studien selten eine Zufallsauswahl möglich, häufig dagegen eine Zufallszuteilung.

**Zufallszuteilung.** Randomisieren ist streng zufällig zuteilen und bedeutet, eine Anzahl von Patienten oder Versuchstieren (Fälle) streng zufällig in zwei oder mehr Gruppen einzuteilen [Lit.29].

**Randomisation in zwei Gruppen.** Für jeden Patienten wird eine Zufallszahl zwischen 0 und 1 gezogen. Patienten mit einer Zufallszahl  $< 0.5$  bilden die Gruppe A, die anderen Patienten die Gruppe B. Die Anzahl der Patienten je Gruppe kann sich rein zufällig unterscheiden.

**Randomisation in zwei genau gleich große Gruppen.** Meist ist es zweckmäßig, die Randomisation so zu steuern, dass genau gleich große Gruppen entstehen. Dazu werden zunächst die Identifikationen der zu randomisierenden Patienten (z.B. die Initialen) aufgeschrieben. Dann wird für jede Person eine Zufallszahl gezogen. Wird für zwei Personen die gleiche Zufallszahl gezogen, so wird für diese Personen nochmals gezogen. Anschließend werden die Personen nach der Größe der für sie gezogenen Zufallszahlen sortiert, die Hälfte der Personen mit den großen Zufallszahlen bilden die eine Gruppe, die anderen Personen, die die kleinen Zufallszahlen gezogen haben, die andere Gruppe.

**Randomisation zukünftiger Patienten.** Meist ist vor Studienbeginn weder die genaue Fallzahl noch die Identifikation von Patienten bekannt. Die Patienten erhalten jedoch streng in der Reihenfolge der Aufnahme in die Studie eine laufende studieninterne Patientennummer. Dann werden anstatt der Patienten die Patientennummern randomisiert.

**Geschichtete Randomisation.** Bei den meisten Studien wird sowohl geschichtet als auch randomisiert. Die Randomisation erfolgt dann innerhalb jeder Schicht (**Box 1.4**). Allerdings ist die Zugehörigkeit der Patienten zu den Schichten vorab nicht bekannt. Deshalb können sich die Schichten ungleichmäßig füllen. Während sich bei der Schichtung die Schichtungskriterien genau gleich auf die Gruppen verteilen (soweit das mit ganzen Zahlen möglich ist), gewährleistet die Randomisation, dass die Unterschiede zwischen den Gruppen – auch hinsichtlich aller bekannten und unbekanntem Störgrößen – höchstens zufällig sind. Die Randomisation ist noch wichtiger als die Schichtung.

**Concealment.** Bei der Randomisation ist sehr wichtig: **Zuerst wird entschieden, ob ein Patient in die Studie aufgenommen wird, und dies wird festgestellt und dokumentiert. Dann erst darf das Ergebnis der Randomisation bekannt werden.** Der Verschluss der Randomisationsliste heißt Concealment. Ohne Concealment besteht die Gefahr, dass bei einem Patienten, der in die Studie aufgenommen werden kann, zuerst nachgeschaut wird, in welche Gruppe er randomisiert wird und dann, wenn das Ergebnis der Randomisation gefällt, der Patient tatsächlich aufgenommen wird. Gefällt dagegen das Ergebnis der Randomisation nicht, so wird der Patient nicht aufgenommen. Durch diese Selektion wird der Effekt der Randomisation aufgehoben.

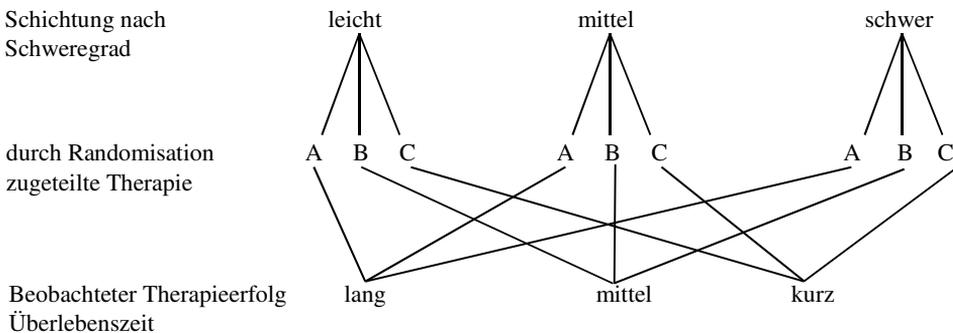
**Randomisation per Telefon, Telefax oder Computerzugriff.** Der Kliniker, der einen Patienten in eine Studie aufnehmen möchte, schickt die Daten des aufzunehmenden Patienten an die randomisierende Stelle und erhält von dort das Ergebnis der Randomisation. Die randomisierende Stelle

dokumentiert jede Randomisation. Heute gibt es Computerprogramme, die eine Randomisation durchführen und dokumentieren. Programme haben den Vorteil, dass sie rund um die Uhr verfügbar sind. Bei all diesen Randomisationsverfahren ist das Concealment gewährleistet.

**Randomisation mit verschlossenen Umschlägen.** Für jede Schicht erstellt der Statistiker einen Stapel verschlossener Umschläge. Außen auf dem Umschlag stehen die Bezeichnung der Studie, die Schichtzugehörigkeit und die Patientenummer. Wird ein Patient in die Studie aufgenommen, so wird seine Schichtzugehörigkeit ermittelt, dann aus den Umschlägen dieser Schicht der Umschlag mit der kleinsten, noch nicht vergebenen Patientenummer herausgesucht und aufgerissen. Innerhalb des Umschlags stehen die gleichen Daten, zusätzlich noch die Behandlungsgruppe. Damit ist der Patient einer Behandlungsgruppe zugeteilt. Der Umschlag kann auch Klebeetiketten zum Beschriften der Studienformulare enthalten. Die Randomisation mit verschlossenen Umschlägen ist bequem. Allerdings ist das Concealment nicht sicher gewährleistet: Wenn die Zuteilung eines Patienten nicht gefällt, ist es möglich, den Patienten nicht in die Studie aufzunehmen und den bereits geöffneten Umschlag für den nächsten Patienten zu verwenden.

**Randomisationsersatz.** Neben der Randomisation gibt es eine Reihe anderer Zuteilungsverfahren, die jedoch alle erhebliche Nachteile haben und nicht angewandt werden sollten. Werden alle Patienten, die z.B. an einem geraden Tag in die Klinik aufgenommen werden, der Untersuchungsgruppe, alle Patienten, die an einem ungeraden Tag aufgenommen werden, der Vergleichsgruppe zugeteilt, so besteht der Verdacht, dass die Aufnahme eines Patienten so gelegt wird, dass er in die erwünschte Gruppe kommt. Erfolgt die Zuteilung nach Geburtsdatum, etwa in der Form, dass alle Patienten, die in einem ungeraden Monat (Januar, März, usw.) geboren worden sind, in die Untersuchungsgruppe, alle Patienten, die in einem geraden Monat geboren worden sind (Februar, April, usw.) in die Vergleichsgruppe kommen, so besteht der Verdacht, dass wegen des fehlenden Concealments die Patienten, die in eine unerwünschte Gruppe kämen, nicht in die Studie aufgenommen werden.

#### Box 1.4: Nach Schweregrad geschichtete Randomisation



Jeder Patient wird zuerst nach seinem Schweregrad der Schicht „leicht“, „mittel“ oder „schwer“ zugeordnet. Dann werden die Patienten, die den gleichen Schweregrad haben, streng zufällig einer der drei Therapiegruppen A, B oder C zugeordnet (randomisiert). Im Gegensatz zu **Box 1.3** ist jetzt der Anteil der Patienten mit leichtem Schweregrad in allen drei Therapiegruppen gleich, ebenso der Anteil mit hohem Schweregrad. **Fazit:** Das längere Überleben ist durch die Therapie A bedingt.