

**Serie: „Biometrische Methoden in der medizinischen Forschung“**

In der DZZ-Ausgabe 8/2010 haben wir den ersten Teil der Artikelserie „Biometrische Methoden in der medizinischen Forschung“ veröffentlicht, die darauf abzielt, das Verständnis für wissenschaftliche Publikationen zu fördern. In dieser Serie werden wesentliche methodische Aspekte zu Studientypen und zum Studiendesign, zur Auswahl von Studientypen und zur Vermeidung von Verzerrungen und Fehlern dargelegt. Die Erstveröffentlichung der Reihe fand im *Deutschen Ärzteblatt* statt. Eine nähere Einleitung zu der Serie initiiert vom Institut für Medizinische Biometrie, Epidemiologie und Informatik (Direktorin Frau Prof. Dr. Maria Blettner) der Universitätsklinik Mainz sowie den ersten Teil „Kritisches Lesen wissenschaftlicher Artikel – Teil 1 der Serie zur Bewertung wissenschaftlicher Publikationen“ finden Sie im Internet unter <http://www.online-dzz.de/article/869>.

B. Röhrig<sup>1</sup>, J.B. du Prel<sup>2</sup>, M. Blettner<sup>3</sup>

# Studiendesign in der medizinischen Forschung – Teil 2 der Serie zur Bewertung wissenschaftlicher Publikationen<sup>\*</sup>

*Study design in medical research – part 2 of a series on evaluation of scientific publications*

**Einleitung:** Die wissenschaftliche Qualität und Aussagekraft einer medizinischen Studie wird entscheidend durch das Studiendesign bestimmt. Fehler in der Studienplanung sind später nicht mehr korrigierbar. Der vorliegende Artikel stellt unterschiedliche Aspekte des Studiendesigns vor.

**Methoden:** Anhand einer selektiven Auswahl wissenschaftlicher Artikel aus der internationalen Literatur und eigener wissenschaftlicher Expertise zum Thema Studiendesign werden sechs wesentliche Kriterien herausgearbeitet, die bei der Planung oder Beurteilung einer Studie unbedingt zu beachten sind.

**Ergebnisse:** Die sechs wesentlichen Kriterien des Studiendesigns sind Fragestellung, Studienpopulation, Beobachtungseinheit, Studientyp, Messverfahren und Fallzahlabschätzung.

**Diskussion:** Dem Leser soll ein Instrumentarium zur Beurteilung des Studiendesigns medizinischer Studien zur Verfügung gestellt werden. Damit kann er wissenschaftliche Studien einordnen und deren Qualität besser einschätzen. (Dtsch Zahnärztl Z 2010, 65: 488–493)

*Schlüsselwörter:* Studiendesign, Qualität, Studie, Studientyp, Messmethode

**Background:** The scientific value and informativeness of a medical study are determined to a major extent by the study design. Errors in study design cannot be corrected afterwards. Various aspects of study design are discussed in this article.

**Methods:** Six essential considerations in the planning and evaluation of medical research studies are presented and discussed in the light of selected scientific articles from the international literature as well as the authors' own scientific expertise with regard to study design.

**Results:** The six main considerations for study design are the question to be answered, the study population, the unit of observation, the type of study, the measuring technique, and the calculation of sample size.

**Conclusions:** This article is intended to give the reader guidance in evaluating the design of studies in medical research. This should enable the reader to categorize medical studies better and to assess their scientific quality more accurately.

*Keywords:* study design, quality, study, study type, measuring technique

<sup>\*</sup> Nachdruck aus: Dtsch Arztebl Int 2009; 106(11): 184–189; DOI: 10.3238/arztebl.2009.01184 © Deutscher Ärzte-Verlag GmbH Köln

<sup>1</sup> MDK Rheinland-Pfalz, Referat Rehabilitation/Biometrie, Alzey

<sup>2</sup> Institut für Epidemiologie, Universität Ulm

<sup>3</sup> Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI), Universitätsklinikum Mainz

**Peer reviewed article:** eingereicht: 30.11.2007, revidierte Fassung angenommen: 08.02.2008

**DOI:** 10.3238/dzz.2010.0488

## Einleitung

Studien in der Medizinischen Forschung können in fünf Phasen unterteilt werden. Diese sind Planung, Durchführung, Dokumentation, Analyse und Publikation [2, 23]. Neben finanziellen, organisatorischen, logistischen und personellen Fragen ist das Studiendesign unter wissenschaftlichen Gesichtspunkten der wichtigste Aspekt der Studienplanung. Häufig wird die Bedeutung des Studiendesigns hinsichtlich der späteren Qualität, Aussagekraft und letztendlich Publikationswürdigkeit einer Studie unterschätzt [2]. Lange vor der Probandenrekrutierung werden durch das Studienkonzept die Weichen für die Beantwortung der Fragestellung(en) gestellt. Fehler im Design können, im Gegensatz zur statistischen Auswertung, nach Beendigung einer Studie nicht mehr korrigiert werden. Deshalb muss das Studiendesign vor Beginn einer Studie sorgfältig festgelegt und in einem Studienprotokoll festgehalten werden.

Der Begriff „Studiendesign“ wird in der Literatur uneinheitlich verwendet. Oft wird das Wort auf die Auswahl eines geeigneten Studientyps reduziert. Unter Studiendesign kann man jedoch auch das Gesamtkonzept aller Vorgehensweisen im Rahmen einer Studie verstehen. Durch die genaue Planung einer Studie, sollen Einflüsse, die das Ergebnis eines Testverfahrens verfälschen, vermindert werden [4, 7]. Dieser Artikel versteht den Begriff Studiendesign umfassend und stellt anhand einer selektiven Auswahl wissenschaftlicher Artikel aus der internationalen Literatur und eigener wissenschaftlicher Expertise sechs verschiedene Aspekte zum Thema „Studiendesign“ vor: Fragestellung, Studienpopulation, Studientyp, Beobachtungseinheit, Messverfahren und Fallzahlplanung. Anhand dieser soll der Leser die Ergebnisse von Publikationen besser einordnen und beurteilen können. Wer eigene Studien durchführen will, muss sich intensiver mit dem Thema Studiendesign beschäftigen.

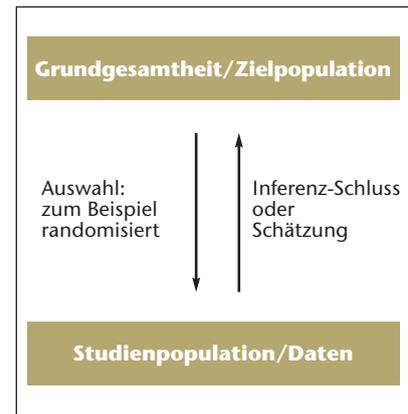
## Fragestellung

Entscheidend für die Studienplanung ist die forschungsleitende Fragestellung. Der Forscher muss sich über die Ziele der Studie im Klaren sein. Er muss

sich genau überlegen, auf welche Frage die geplante Studie eine Antwort geben will. Diese Fragestellung muss operationalisiert, also mess- und beurteilbar gemacht werden. Dazu müssen ein adäquates Design und geeignete Messgrößen gefunden werden. Hauptfragestellung(en) sind hierbei von Nebenfragestellung(en) zu unterscheiden. Als Ergebnis der Untersuchung sollen offene Fragen beantwortet oder aber neue Hypothesen generiert werden. Dabei sind die „sieben W's“ wichtig: weshalb, wer, was, wie, wann, wo, wie viele? Die Fragestellung impliziert somit auch die Zielgruppe und sollte sehr präzise formuliert werden. Statt etwa zu fragen „Wie ist die Lebensqualität von Patienten?“, ist genau festzulegen, bei welchen Patienten (zum Beispiel Alter), in welchem Gebiet (zum Beispiel Deutschland), bei welcher Krankheit (zum Beispiel Brustkrebs), in welchem Zustand (zum Beispiel Tumorstadium 3), gegebenenfalls nach welcher Intervention (zum Beispiel nach OP), welcher Endpunkt (hier: Lebensqualität) mit welcher Messmethode (zum Beispiel Fragebogen EORTC QLQ-C30) zu welchem Zeitpunkt bestimmt werden soll. Wissenschaftliche Fragestellungen beinhalten neben der reinen Deskription häufig Vergleiche, zum Beispiel zwischen zwei Gruppen beziehungsweise vor und nach einer Intervention. So interessiert beispielsweise die Lebensqualität von Brustkrebspatientinnen im Vergleich zu Frauen gleichen Alters ohne Krebs.

Der Forscher legt mit der Fragestellung bereits fest, ob er die Studie deskriptiv, explorativ oder konfirmatorisch auswerten will. Während in einer deskriptiven Auswertung Beobachtungseinheiten durch erhobene Variablen (zum Beispiel Blutparameter, Diagnose) beschrieben werden, verfolgt die explorative Analyse das Ziel, Zusammenhänge zwischen Variablen zu erkennen, zu bewerten und neue Hypothesen zu formulieren. Konfirmatorische Analysen sind dagegen auf statistische Beweisführung, also das Testen von festgelegten Studienhypothesen, angelegt.

Mit der Fragestellung ist Art und Umfang der erhobenen Daten ebenfalls determiniert. Mit ihr wird festgelegt, welche Daten zu welchem Zeitpunkt erhoben werden sollen. Dabei ist weniger oft mehr! Für die Fragestellung(en) irre-



**Grafik 1** Zusammenhang zwischen Grundgesamtheit/Population und Studienpopulation.

**Graphic 1** Correlation between basic population and study population.

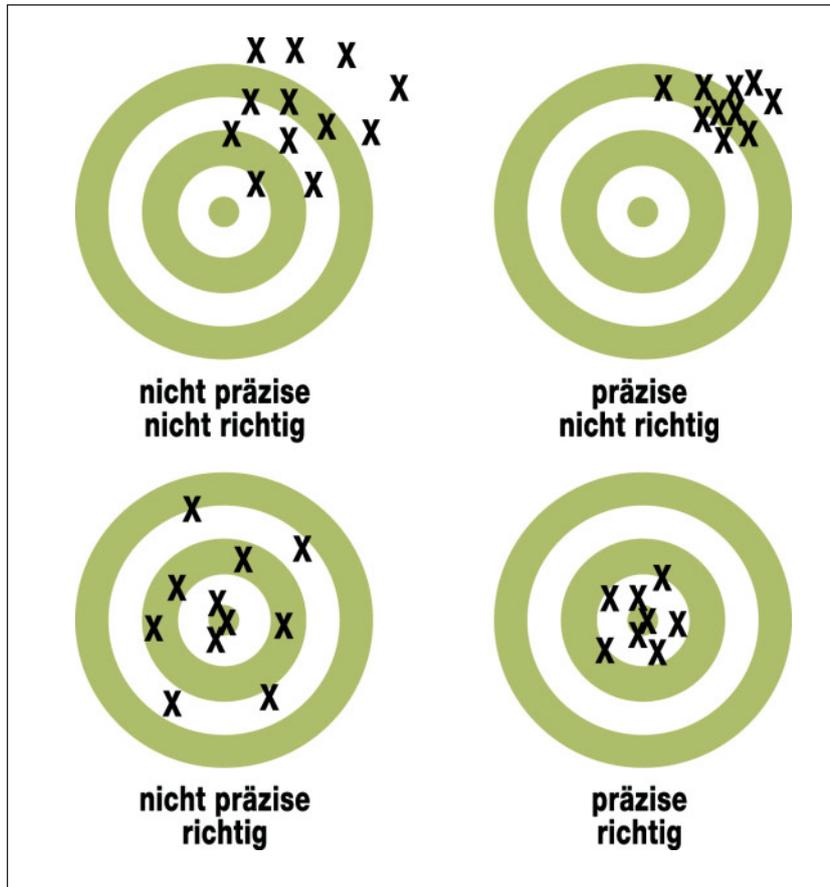
levante Daten sollten erst gar nicht erhoben werden. Die Erhebung zu vieler Variablen an zu vielen Messzeitpunkten kann eine niedrige Teilnehmerate, hohe „Drop-Out“-Raten und eine schlechte Compliance der Probanden zur Folge haben. Erfahrungsgemäß werden später doch nicht alle Daten ausgewertet.

Fragestellung und Auswertestrategie müssen vor Beginn einer Studie im Studienprotokoll festgehalten werden.

## Studienpopulation

Die Fragestellung der Studie impliziert die Zielgruppe, an der diese geklärt werden soll. Primär interessiert den Wissenschaftler jedoch nicht das konkrete Ergebnis der beobachteten Studienpopulation, sondern die Übertragbarkeit der ermittelten Ergebnisse auf die Zielpopulation. Dementsprechend sollen mittels statistischer Testverfahren Ergebnisse der Stichprobe auf die Grundgesamtheit verallgemeinert werden (Grafik 1).

Eine hohe Repräsentativität der Studienpopulation kann durch eine geeignete Auswahl der Stichprobe aus der Grundgesamtheit erreicht werden. Darüber können vorab definierte, selektive Ein- und Ausschlusskriterien entscheiden, beispielsweise Geschlecht, Alter und Tumorstadium. Die Auswahl der Studienteilnehmer kann zufällig, etwa durch zufällige Auswahl über Einwohnermeldeamtsregister, oder konsekutiv (nachfolgend), beispielsweise alle Pa-



**Grafik 2** Darstellung der Begriffe Reliabilität (= Präzision) und Validität (= Richtigkeit) anhand einer Zielscheibe.

**Graphic 2** Illustration of the terms reliability (= precision) and validity (= correctness) with the help of a target.

tienten einer Klinikabteilung in einem Jahr, erfolgen.

Bei selektiver Stichprobenauswahl kann nur eine Aussage über eine Grundgesamtheit, die eben diesen Auswahlkriterien entspricht, getroffen werden. Ob Patienten beispielsweise aus einer Spezialpraxis, einer Fachabteilung eines Krankenhauses oder von mehreren Praxen stammen, spielt für die Generalisierbarkeit der Ergebnisse eine wichtige Rolle.

Zum Aspekt der Verallgemeinerung der Studienergebnisse gehört auch die Entscheidung, die Studie unizentrisch (nur an einer Institution) oder multizentrisch (an mehreren Institutionen) durchzuführen. Vorteile multizentrischer Studien sind, dass die benötigte Fallzahl in kürzerer Zeit erreicht werden kann und die größere Generalisierbarkeit der Ergebnisse aufgrund der unterschiedlichen Behandlungszentren. Damit nimmt die externe Validität zu.

## Studientyp

Vor der Festlegung des Studientyps muss der Forscher sich über die Forschungskategorie im Klaren sein. Prinzipiell wird zwischen Forschung an Primärdaten und Sekundärdaten unterschieden.

Unter Forschung an Primärdaten versteht man die Durchführung der eigentlichen wissenschaftlichen Studien mit Erhebung primärer Studiendaten. Damit sollen wissenschaftliche Fragestellungen beantwortet und neue Erkenntnisse gewonnen werden.

Demgegenüber werden in der Forschung an Sekundärdaten Ergebnisse bereits durchgeführter und publizierter Studien analysiert. Darunter fallen zum einen die (erneute) Analyse bereits erhobener Daten. Diese können von einem Register, aus der Bevölkerungsstatistik oder von Studien stammen. Zum anderen geht es darum, einen umfassenden

Überblick über den aktuellen Stand der Forschung zu geben und daraus geeignete Schlüsse zu ziehen. In der Sekundärdatenforschung werden Übersichtsartikel (narrative Reviews), systematische Reviews und Metaanalysen unterschieden.

Die zugrunde liegende Fragestellung entscheidet ebenfalls über die Wahl des Studientyps. In der Primärforschung wird zwischen experimenteller Forschung sowie klinischer und epidemiologischer Forschung unterschieden.

Unter experimentelle Forschung fallen zum einen angewandte Studien wie Tierversuche, Zellversuche, biochemische und physiologische Untersuchungen sowie Materialeigenschaftsstudien, zum anderen Methodenentwicklung wie die Entwicklung analytischer und biometrischer Verfahren.

Unter klinischer Forschung werden interventionelle und nicht interventionelle Studien zusammengefasst. Interventionelle klinische Studien (klinische Prüfungen) haben das Ziel, „klinische oder pharmakologische Wirkungen von Arzneimitteln zu erforschen oder nachzuweisen“ und „sich von der Unbedenklichkeit oder Wirksamkeit der Arzneimittel zu überzeugen“ (AMG § 4) [24]. Bei klinischen Prüfungen werden Patienten randomisiert den Therapiegruppen zugewiesen. Nicht interventionelle klinische Studien sind demgegenüber Beobachtungsstudien, in denen die Patienten eine individuell festgelegte Therapie erhalten [5, 20].

In der epidemiologischen Forschung interessieren die Verteilung und zeitliche Veränderung der Häufigkeiten sowie die Ursachen von Krankheiten. Experimentelle werden von beobachtenden Studien unterschieden [5, 11]. Interventionsstudien (zum Beispiel Impfen, Stoffzusatz in Lebensmittel, Trinkwasser-Fluoridierung) sind experimenteller Natur. Beispiele für beobachtende epidemiologische Studien sind Kohortenstudien, Fall-Kontroll-Studien, Querschnittstudien sowie ökologische Studien.

Auf die einzelnen Studientypen wird detailliert in einem Folgeartikel eingegangen.

## Beobachtungseinheit

Vor Beginn der medizinischen Untersuchung ist die Beobachtungseinheit

(Untersuchungseinheit, Merkmalsträger) festzulegen. In einer klinischen Studie ist typischerweise der Patient die Beobachtungseinheit. Beobachtungseinheit einer Studie kann jedoch auch ein technisches Modell, eine Erbinformation, eine Zelle, ein Zellverband, ein Organ, ein Organsystem, ein Einzelproband (Tier oder Mensch), ein Teilkollektiv sowie die Bevölkerung einer Region beziehungsweise eines Landes sein. Bei systematischen Reviews ist die Beobachtungseinheit eine Studie. Die Stichprobe umfasst die Gesamtzahl aller Beobachtungseinheiten. An den Merkmalsträgern werden die interessierenden Informationen beziehungsweise Daten (Beobachtungen, Variablen, Merkmale) erhoben. So wird an der Beobachtungseinheit Patient bei der Untersuchung des Herzens die Herzfrequenz als Merkmal der Leistungsfähigkeit bestimmt.

Die Wahl der Beobachtungseinheit beeinflusst die Interpretation der Studienergebnisse. Aus statistischen Gründen ist darauf zu achten, ob die Beobachtungseinheiten hinsichtlich des Zielkriteriums unabhängig oder abhängig voneinander sind. Diese Unterscheidung ist nicht immer einfach. Sind beispielsweise die Zähne von Probanden die Beobachtungseinheit, so ist zu klären, ob diese hinsichtlich der Fragestellung unabhängig – also von verschiedenen Probanden – oder abhängig – also vom gleichen Probanden – sind. Zähne im Mund eines Patienten sind in aller Regel abhängig, da bestimmte Einflussfaktoren, wie beispielsweise die Ernährung und das Zahnputzverhalten, auf alle Zähne im Mund gleichartig wirken. Die Untersuchung extrahierter Zähne ist hingegen meist unabhängig, da es keine gemeinsamen Einflussfaktoren auf diese mehr gibt. Dies gilt insbesondere, wenn die Zähne noch weiter präpariert werden, beispielsweise durch Schneiden und Schleifen. Falls sich die Beobachtungen an den Zähnen jedoch auf Merkmale beziehen, die sich vor Extraktion der Zähne gebildet haben, dann sind diese als abhängig zu betrachten.

**Messverfahren**

Unter dem Begriff Messverfahren (= Messmethode) wird zum einen der Einsatz von Messinstrumenten, zum anderen die Messmethodik zusammengefasst.

Begriff	Synonym	englischer Ausdruck
Reliabilität (engl. „reliability“)	Präzision, Zuverlässigkeit, Wiederholbarkeit	precision
Validität (engl. „validity“)	Richtigkeit, Gültigkeit	trueness, accuracy of the mean
Genauigkeit	Güte, Zusammenfassung aus Reliabilität und Validität	accuracy

**Tabelle 1** Zusammenfassende Darstellung wichtiger Begriffe zur Validierung einer Messmethode.

**Table 1** Illustration of important terms concerning validity of a measuring method.

**Einsatz von Messinstrumenten**

Zu den Messinstrumenten zählen Geräte, die gezielt Messdaten erheben (zum Beispiel Blutdruck, Laborparameter) sowie standardisierte und selbst entworfene Fragebogenerhebungen (zum Beispiel Lebensqualität, Depression, Zufriedenheit).

Bei der Validierung eines Messinstruments werden dessen Güte und Einsatzfähigkeit mittels statistischer Kenngrößen beurteilt. Leider ist die Nomenklatur nicht immer einheitlich und unterscheidet sich zudem nach Fachdisziplinen (zum Beispiel chemische Analytik, psychologische Erhebungen mittels Fragebögen und Diagnosestudien). Immer gilt, dass ein Messinstrument mit hoher Güte eine hohe Reliabilität und eine hohe Validität haben sollte.

Reliabilität beschreibt, inwieweit ein Messverfahren bei Versuchswiederholung gleiche Resultate konsistent liefert [10]. Die Reliabilität (= Wiederholbarkeit) macht Angaben zur Präzision, also dem Auftreten zufälliger Fehler. Eine geringe Reliabilität der Messmethode führt zu kleineren Korrelationskoeffizienten, ungenaueren Messungen sowie zu höheren Fallzahlen [10]. Demgegenüber hat ein Messinstrument eine hohe Validität (= Richtigkeit), wenn es genau das misst, was es zu messen vorgibt. Die Validität macht folglich Angaben zum Auftreten systematischer Fehler [14]. Während die Reliabilität die Unterschiedlichkeit (= Varianz) von Wiederholungsmessungen beschreibt, handelt die Validität von der Abweichung zwischen gemessenem und wahren Messwert [14]. Grafik 2 stellt beide Begriffe anhand des Modells einer Zielscheibe gegenüber.

Reliabilität und Validität werden unter dem Begriff Genauigkeit zusammengefasst [6, 12]. Die Genauigkeit ist nur dann hoch, wenn sowohl Reliabilität als auch Validität hoch sind. Tabelle 1 fasst die Begriffe und häufig verwendete Synonyme zusammen.

Nicht nur die Messungen selbst können im Sinne einer fehlerhaften Messung nicht valide sein, sondern auch die Schlussfolgerungen, die aus ihnen, im Sinne der Fragestellung beziehungsweise des Konstrukts gezogen werden. Die externe kann von der internen Validität unterschieden werden [16]. Unter externer Validität versteht man die Verallgemeinerbarkeit der an der Studienpopulation gewonnenen Studienergebnisse auf die Zielpopulation. Die interne Validität beschreibt die Gültigkeit eines Ergebnisses für die eigentliche Fragestellung. Sie steigt mit einer detaillierten Planung, mit definierten Ein- und Ausschlusskriterien und der Reduktion äußerer Störeinflüsse.

**Messmethodik**

In der Messmethodik werden Anzahl und zeitlicher Ablauf der durchzuführenden Messungen beschrieben. Um vergleichbare Messergebnisse zu erhalten, sollten die Messbedingungen standardisiert sein – Stichwort: Objektivität. Klinische Messungen im Rahmen von Studien wie eine Blutdruckmessung sollten beispielsweise immer zur gleichen Zeit, im gleichen Raum, in der gleichen Position, mit demselben Messinstrument und von derselben Untersuchungsperson durchgeführt werden. Werden verschiedene Messmethoden angewandt, beispielsweise Untersucher, Messinstrumente, Analyselabors oder

Item	Inhalt/Angaben
Fragestellung	Ist die Fragestellung klar definiert?
Studienpopulation	<ul style="list-style-type: none"> <li>• Angaben zu <ul style="list-style-type: none"> <li>– Rekrutierung (Art, Gebiet, Zeit)</li> <li>– Soziodemografische Angaben zu den Probanden (zum Beispiel Alter, Geschlecht, Krankheit)</li> <li>– Ein- und Ausschlusskriterien</li> <li>– Zeitraum der Nachbeobachtung</li> </ul> </li> </ul>
Studientyp	<ul style="list-style-type: none"> <li>• Forschung an Sekundärdaten</li> <li>• Forschung an Primärdaten (= eigentliche Studien) <ul style="list-style-type: none"> <li>– experimentelle Studien</li> <li>– klinische Studien</li> <li>– epidemiologische Studien</li> </ul> </li> </ul>
Beobachtungseinheit	<ul style="list-style-type: none"> <li>• Technisches Modell (zum Beispiel eine Prothese, Werkstoff der Zahnheilkunde, eine Blutprobe)</li> <li>• Erbinformation</li> <li>• Zelle</li> <li>• Zellsystem</li> <li>• Organ (zum Beispiel Herz, Lunge)</li> <li>• Organsystem (zum Beispiel Herz-Kreislauf-System)</li> <li>• Einzelproband (Tier oder Mensch)</li> <li>• Teilkollektiv (zum Beispiel Krankenhauskollektiv, Risikogruppe)</li> <li>• Bevölkerung (zum Beispiel aus einer Region)</li> </ul>
Messverfahren	<ul style="list-style-type: none"> <li>• Einsatz von Messinstrumenten (= Validierung) <ul style="list-style-type: none"> <li>– Reliabilität</li> <li>– Validität</li> </ul> </li> <li>• Messmethodik <ul style="list-style-type: none"> <li>– zeitlicher Ablauf</li> <li>– Anzahl der Untersucher</li> <li>– Standardisierung der Messbedingungen</li> <li>– Festlegung des Skalenniveaus</li> </ul> </li> </ul>
Fallzahlplanung	<ul style="list-style-type: none"> <li>• Wurde eine Fallzahlplanung durchgeführt?</li> <li>• wenn ja . . . wie waren die Bedingungen: <ul style="list-style-type: none"> <li>– Art des Tests</li> <li>– Signifikanzniveau</li> <li>– Power</li> <li>– klinisch relevanter Unterschied</li> <li>– Streuung/Varianz</li> </ul> </li> </ul>

**Tabelle 2** Checkliste zur Bewertung des Studiendesigns

**Table 2** Checklist to evaluate the study design.

Erhebungszeiten, muss die Übereinstimmung der Messungen abgeklärt werden [14, 16].

Außerdem ist das Skalenniveau der Erhebungsmerkmale von entscheidender Bedeutung. Vereinfacht können Merkmale nach ihrem Niveau mit zunehmender Wertigkeit in nominale, ordinale und metrische Skalen unterteilt werden. Das Skalenniveau ist deshalb so wichtig, weil hiervon sowohl die deskriptive (= statistische Maßzahlen) als auch die schließende Statistik (= statistische Testverfahren) abhängen. Die Transformation von einer höherwertigen zu einer niedrigerwertigen Skala ist prinzipiell möglich, nicht jedoch umgekehrt. Wird beispielsweise der Hämoglobingehalt von Patienten auf einer metrischen Skala bestimmt (zum Beispiel in

g/dL), kann eine Einteilung in eine ordinale Skala (zum Beispiel niedriger, normaler und hoher Hämoglobin-Status) erfolgen, nicht aber umgekehrt.

### Fallzahlab-schätzung

Zu jedem Studiendesign muss vor Studienbeginn die benötigte Zahl der Beobachtungseinheiten (zum Beispiel Patienten) abgeschätzt werden, um die Hauptfragestellung zu beantworten [3, 8, 17]. Dies erfolgt mittels Fallzahlplanung. Hierfür sind jedoch Kenntnisse über den zu erwartenden Effekt (zum Beispiel der klinisch relevante Unterschied) und dessen Streuung (zum Beispiel die Standardabweichung) nötig. Sie können in Voruntersuchungen oder durch Anga-

ben in der Literatur ermittelt werden. Allgemein gilt, dass man eine große Fallzahl braucht, um einen kleinen Unterschied zu entdecken. Die Fallzahl wird ebenfalls groß, wenn die Streuung innerhalb der Studiengruppen hinsichtlich des Zielkriteriums groß ist. Die Fallzahlplanung hilft, dass eine Studie groß genug aber nicht zu umfassend wird. Häufig begrenzt der zeitliche Rahmen und/oder das Budget die Fallzahl. Das entspricht nicht einer guten wissenschaftlichen Praxis. Eine kleine Stichprobengröße beinhaltet eine zu geringe Power und birgt die Gefahr, echte Unterschiede nicht nachweisen zu können [8, 13]. Die Durchführung einer Studie mit zu kleiner oder auch mit zu großer Fallzahl ist ethisch (Belastung der Patienten, eventuell randomisierte Therapiezuweisung) aber auch ökonomisch (finanziell, strukturell und personell) nicht zu rechtfertigen [1, 8, 13, 15]. Der Forscher sollte sich überlegen, ob alternative Vorgehensweisen, wie das Aufstocken der zeitlichen, personellen und finanziellen Kapazitäten oder die Durchführung einer multizentrischen Studie in Kooperation mit Kollegen möglich sind.

### Diskussion

Bei medizinischen Studien werden Planung, Durchführung, Dokumentation, Analyse und Publikation unterschieden [2, 23]. In der Planung kommt dem Studiendesign entscheidende Bedeutung zu. Es legt nicht nur die statistische Analyse fest, sondern letztendlich auch Aussagekraft, Bedeutung und Umsetzung der Studienergebnisse [23]. Für die rasche Beurteilung des Studiendesigns anhand von sechs Punkten wurde eine Checkliste erstellt (Tab. 2).

Nach *Sackett* beziehen sich von 56 typischen Fehlern in Studien etwa zwei Drittel auf Fehler im Design und in der Durchführung [22]. Diese sind nach der Datenerhebung nicht mehr korrigierbar. Die Studie ist somit wenig aussagekräftig. Deshalb muss das Design vor Beginn einer Studie genau geplant und im Studienprotokoll festgelegt werden. Den Zeitaufwand hierfür darf man nicht unterschätzen.

Letztendlich sind wissenschaftliche Studien mit schlechtem Studiendesign unethisch. Probanden (oder Tiere) werden unnötig belastet und Forschungs-

kapazitäten verschwendet [18, 21]. In medizinischen Studien muss sowohl die individuelle Ethik (Schutz des Individuums) als auch die kollektive Ethik (Nutzen für die Gesellschaft) berücksichtigt werden [21]. Ein oft festzustellender Mangel medizinischer Studien ist eine zu geringe Fallzahl, was zu einer zu geringen Power führt [19]. Deshalb kann oft ein real vorhandener Unterschied – zum Beispiel zwischen der Wirkung zweier Therapien – nicht entdeckt oder nur ungenau beschrieben werden [9]. Eine niedrige Power entsteht durch eine zu geringe Fallzahl, einen zu kleinen Unterschied zwischen den Studiengruppen sowie eine zu große Streuung der Messwerte. *Sterne* fordert, die Qualität von Studien durch eine höhere Fallzahl und eine bessere Messgenauigkeit zu verbessern [25]. Andererseits führt eine zu große Fallzahl dazu, dass unnötig viele Probanden (oder Tiere) Belastungen ausgesetzt werden und Ressourcen (zum Beispiel Personal, Finanzen) verschwendet werden. In der Planungsphase muss deshalb durch Fallzahl-

abschätzung die Machbarkeit einer Studie beurteilt werden. Gegebenenfalls sind geeignete Maßnahmen zur Sicherung einer ausreichend hohen Power zu ergreifen. Der Verweis auf zu wenig Zeit beziehungsweise zu geringe finanzielle Kapazitäten ist deplaziert. Verringerung der Heterogenität durch Einschlussbedingungen, Verbesserung der Messgenauigkeit und Kooperationen in multizentrischen Studien können zur Steigerung der Power beitragen. Der Erkenntnisgewinn durch eine einzige, akkurat durchgeführte Studie mit sinnvollem Design und ausreichender Stichprobengröße ist deutlich größer als die Durchführung mehrerer diesbezüglich unzureichender Studien.

Nur eine sorgfältige Studienplanung führt zu Ergebnissen, die in hochrangigen Journalen publiziert werden. Nach Ablauf der Studie sind Fehler und Unzulänglichkeiten in der Planung nicht mehr zu korrigieren. Deshalb empfiehlt sich bereits in der Planungsphase einer Studie die Konsultation eines erfahrenen Biometrikers [2, 8, 13, 15]. 

**Interessenkonflikt:** Die Autoren erklären, dass kein Interessenkonflikt im Sinne der Richtlinien des International Committee of Medical Journal Editors besteht.

#### Korrespondenzadressen

Dr. rer. nat. Bernd Röhrig  
MDK Rheinland-Pfalz  
Referat Rehabilitation/Biometrie  
Albiger Straße 19d  
55232 Alzey  
E-Mail: Bernd.Roehrig@mdk-rip.de

Prof. Dr. Maria Blettner  
Universitätsmedizin der Johannes  
Gutenberg-Universität Mainz  
Institut für Medizinische Biometrie,  
Epidemiologie und Informatik  
Obere Zahlbacher Straße 69  
55131 Mainz  
Tel.: 0 61 31 / 17 – 32 52  
Fax: 0 61 31 / 17 – 29 68  
E-Mail:  
maria.blettner@unimedizin-mainz.de  
www.imbei.uni-mainz.de

## Literatur

- Altman DG: Statistics and ethics in medical research, misuse of statistics is unethical. *BM* 281, 1182–1184 (1980)
- Altman DG, Gore SM, Gardner MJ, Pocock SJ: Statistical guidelines for contributors to medical journals. *BMJ* 286, 1489–1493 (1983)
- Altman DG: Practical Statistics for Medical research. Chapman and Hall, London 1991, 1–9
- Altman DG, Machin D, Bryant TN, Gardner MJ: Statistics with confidence. 2nd edition. *BMJ Books*, Bristol 2000, 173
- Beaglehole R, Bonita R, Kjellström T: Einführung in die Epidemiologie. Verlag Hans Huber, Bern 1997, 53–84
- Brüggemann L: Bewertung von Richtigkeit und Präzision bei Analysenverfahren, *GIT Labor-Fachzeitschrift* 2, 153–156 (2002)
- DocCheck-Flexikon: Thema: Studiendesign. <http://flexikon.doccheck.com/Studiendesign>
- Eng J: Sample size estimation: how many individuals should be studied? *Radiology* 227, 309–313 (2003)
- Faller H: Signifikanz, Effektstärke und Konfidenzintervall. *Rehabilitation* 43, 174–178 (2004)
- Fleiss JL: The design and analysis of clinical experiments. John Wiley & Sons, New York 1986, 1–32
- Fletcher RH, Fletcher SW, Wagner EH, Heering J: Klinische Epidemiologie. Grundlagen und Anwendung. Verlag Hans Huber, Bern 2007, 1–24 und 349–378
- Funk W, Dammann V, Donnevert G: Qualitätssicherung in der Analytischen Chemie: Anwendungen in der Umwelt-, Lebensmittel- und Werkstoffanalytik, Biotechnologie und Medizintechnik. 2. Aufl., Wiley-VCH, Weinheim, New York 2005, 1–100
- Halpern SD, Karlawish JHT, Berlin JA: The continuing unethical conduct of underpowered clinical trials. *JAMA* 288, 358–362 (2002)
- Hüttner M, Schwarting U: Grundzüge der Marktforschung. 7. Aufl., Oldenburg Verlag, München 2002, 1–600
- Krummenauer F, Kauczor HU: Fallzahlplanung in referenzkontrollierten Diagnosestudien. *Fortschr Röntgenstr* 174, 1438–1444 (2002)
- Lienert GA, Raatz U: Testaufbau und Testanalyse. 2. Aufl., Psychologie Verlags Union, Weinheim 1998, 220–271
- Machin D, Campbell MJ, Fayers PM, Pinol APY: Sample size tables for clinical studies. 2. Aufl., Blackwell Science Ltd., Oxford, London, Berlin 1987, 296–299
- May WW: The composition and function of ethical committees. *J Med Ethics* 1, 23–29 (1975)
- Moher D, Dulberg CS, Wells GA: Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 272, 122–124 (1994)
- Moher D, Schulz KF, Altman D, for the CONSORT Group: The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomized Trials. *Ann Intern Med* 134, 657–662 (2001)
- Palmer CR: Ethics and statistical methodology in clinical trials. *JME* 19, 219–222 (1993)
- Sackett DL: Bias in analytic research. *J Chronic Dis* 32, 51–63 (1979)
- Schäfer H, Berger J, Biebler KE, et al.: Empfehlungen für die Erstellung von Studienprotokollen (Studienplänen) für klinische Studien. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* 30, 141–154 (1999)
- Schumacher M, Schulgen G: Methodik klinischer Studien, Methodische Grundlagen der Planung, Durchführung und Auswertung. 2. Aufl., Springer, Berlin, Heidelberg, New York 2007, 1–28
- Sterne JAC, Smith GD: Sifting the evidence – what’s wrong with significance tests? *BMJ* 322, 226–231 (2001)