P. Kranke, F. Schuster, R. Muellenbach, E.-M. Kranke[1], N. Roewer, T. Smul

Klinik und Poliklinik für Anästhesiologie, Universitätsklinikum Würzburg (Direktor: Prof. Dr. Norbert Roewer)

[2]Klinik und Poliklinik für Thorax-, Herz- und Thorakale Gefäßchirurgie, Universitätsklinikum Würzburg (Direktor: Prof. Dr. Rainer Leyh)

Grundlagen und Prinzipien klinischer Studien: Wie viele Patienten sollen (müssen) untersucht werden?

Fallzahlschätzung in klinischen Studien

Zusammenfassung

Prospektive klinische Untersuchungen von Medikamenten und Maßnahmen dienen dazu, aus einer Beobachtung eines bestimmten Patientenkollektivs Schlussfolgerungen für ein in der Regel größeres Patientenkollektiv zu treffen. Naturgemäß kann im Rahmen von klinischen Studien nur eine begrenzte Patientenzahl untersucht werden. Dennoch sollen die auf den Studienergebnissen basierenden Aussagen verlässlich sein und die tatsächliche Wirkung einer Intervention treffend beschreiben. Insbesondere sollen keine Überschätzungen oder falsch positive Ergebnisse (es wird ein statistisch signifikanter Unterschied entdeckt, obgleich die Therapie in Wirklichkeit gleich wirksam ist) und Unterschätzungen oder falsch negative Ergebnisse (es wird kein statistisch signifikantes Ergebnis ermittelt, obgleich in Wahrheit ein Unterschied einer definierten Größe vorliegt) erzielt werden.

Da Untersuchungen an Patienten oder Probanden aber in der Regel zeit- und personalintensiv und damit schlussendlich kostenintensiv sind, liegt es im verständlichen Interesse des Untersuchers oder Sponsors einer Studie, die Patienten-/Probandenzahl so groß wie nötig, aber eben gleichsam so gering wie möglich zu halten.

Um unter dieser Prämisse dennoch zu aussagekräftigen Ergebnissen zu gelangen, ist vor der Durchführung einer Studie in der Regel eine Abschätzung der erforderlichen Fallzahl empfehlenswert.

Der Artikel beschreibt das Prinzip und die Notwendigkeit der Fallzahlabschätzung und erläutert an Beispielen, wie eine Fallzahlabschätzung in einfachen Fällen durchgeführt werden kann.

Schlüsselwörter

Klinische Studien, Fallzahlschätzung, randomisierte klinische Studien, Power, Fehler 1. Art, Fehler 2. Art, evidenzbasierte Medizin

Abstract

Clinical trials usually aim at transferring results obtained with the investigation of a limited number of patients to a larger population of interest. Limitations with respect to personnel and financial resources as well as patients with a specific condition being eligible for inclusion lead to inherent restrictions regarding the number of patients that can be investigated in a clinical trial. Irrespective of these restraints the obtained results should be valid and reflect the true efficacy of an intervention. Any false-positive (detecting a statistically significant difference when the interventions are in reality equally effective) as well as false-negative (not detecting a statistically significant difference when a difference of a given magnitude in reality exists) results should be avoided or at least kept to a minimum.

The need to keep the number of subjects being investigated as small as possible but recruiting as many subjects as necessary to obtain meaningful results is one critical hurdle in the planning and conduct of clinical trials. Therefore, investigators should properly calculate sample size before they start with the conduct of a clinical trial.

This article describes the need for a sample size calculation as well as the general accepted principles. Based on a clinical scenario, the components and the conduct of a sample size calculation by means of a simple equation are explained. Finally frequent limitations and pitfalls of sample size calculation are discussed.

Key Words

Clinical trials, sample size calculation, randomized clinical trials (RCT), power, type I error (α-error), type II error (β-error), evidence-based medicine

Einleitung

Klinische Studien, die eine Intervention im Vergleich zu einer Kontrollgruppe oder einer anderen Intervention untersuchen, beispielsweise die Verwendung einer Rollerpumpe im Vergleich zu einer Zentrifugalpumpe zur extrakorporalen Zirkulation, werden immer an einer begrenzten Anzahl an Patienten durchgeführt. Die Ergebnisse der Studie dienen dann vielfach dazu, das eine oder andere Verfahren aufgrund des nachgewiesenen Benefits zu präferieren (wenn ein Verfahren sich in Bezug auf eine wesentliche Zielgröße als vorteilhaft erwiesen hat) oder beide Verfahren in Bezug auf die betrachtete Zielgröße als gleichwertig zu betrachten. Mitunter kommen Autoren der Veröffentlichung aber auch zu dem Schluss, dass zwar „keine signifikanten Unterschiede“ zu beobachten waren, dass jedoch die Fallzahl der untersuchten Individuen zu klein gewesen sei, um sicher zu gehen, dass wirklich kein Unterschied zwischen den untersuchten Gruppen vorliegt. In solchen Fällen spricht man davon, dass „die statistische Power“ unzureichend war oder der Fehler 2. Art (= β-Fehler) – das ist die Wahrscheinlichkeit, einen statistisch signifikanten Unterschied nicht zu entdecken, wenn in Wahrheit ein Unterschied einer definierten Größe vorliegt – zu groß gewesen ist.

Hätte man die Studie im Vorfeld so planen und eine ausreichende Patientenzahl einschließen können, dass derlei unbefriedigende Aussagen gar nicht erst aufgetaucht wären?

Glaubt man den Empfehlungen der CONSORT-Leitlinien [1] mit ihren Vorgaben für die Darlegung von Fallzahlabschätzungen, ist man geneigt, die Frage mit „Ja“ zu beantworten.

Aufgrund von Kapazitätsproblemen und Ressourcenknappheit ist jedoch eine Fallzahlreduktion erwünscht oder vielleicht sogar notwendig. Bedenkt man ferner die zum Teil erheblichen Unsicherheiten mit Blick auf die Elemente der Fallzahlabschätzung kommen Zweifel auf, ob eine prospektive Abschätzung mit limitiertem Vorwissen jemals verlässlich sein kann.

Zweifelsohne gewährleistet eine sorgfältige Studienplanung unter Berücksichtigung relevanter Vorergebnisse oder klinischer Daten, dass auch „negative“ Ergebnisse (korrekt: Studien, die keine Unterschiede in Bezug auf die untersuchten Parameter zwischen den untersuchten Gruppen aufwiesen) mit einer hinreichenden Sicherheit der Aussage behaftet sind, doch sind alle Elemente der Fallzahlschätzung mit Unsicherheiten behaftet, so dass selbst bei gewissenhaftem Vorgehen nicht mit Sicherheit vermieden werden kann, dass die Schlussfolgerung lauten muss: „Studien mit einer größeren Patientenzahl sind zur Beurteilung der Wertigkeit der untersuchten Intervention erforderlich.“

Zur Beurteilung von Studienergebnissen wie zur Planung eigener Vorhaben ist es unerlässlich, die wesentlichen Elemente der Fallzahlberechnung zu kennen. Deshalb sollen im Folgenden die Komponenten der Fallzahlberechnung beschrieben und am Beispiel erläutert werden, wie eine Fallzahlabschätzung in einfachen Fällen durchgeführt werden kann.

Die Elemente der Fallzahlabschätzung

Gehen wir beispielhaft von einer Untersuchung aus, die einen dichotomen Endpunkt („vorhanden“ vs. „nicht vorhanden“) hat. Beispielsweise die eingangs beschriebene Prüfung, ob eine Rollerpumpe im Vergleich zu einer Zentrifugalpumpe zur extrakorporalen Zirkulation mit Blick auf das postoperative Auftreten (die Inzidenz) neurologischer Auffälligkeiten oder postoperativer kognitiver Defizite vorteilhaft ist.

Für diesen Fall bestehen die Elemente der Fallzahlabschätzung aus:

1. Fehler 1. Art (α-Fehler),
2. Fehler 2. Art (β-Fehler bzw. die gewünschte Power, Berechnung: 1 minus Fehler 2. Art),
3. der erwarteten Ereignisrate in der Kontrollgruppe (hier: postoperative kognitive Defizite in der Gruppe „Rollerpumpe“) sowie
4. der erwarteten Ereignisrate in der Behandlungsgruppe (hier: postoperative kognitive Defizite in der Gruppe „Zentrifugalpumpe“) bzw. der erwartete Behandlungseffekt (Differenz zwischen Ereignisrate in der Kontroll- und der Behandlungsgruppe).

Ausgangspunkt im Rahmen dieser Prüfung könnte z. B. ein mutmaßlicher Nutzen aufgrund von biophysikalischen Überlegungen sein oder die Beobachtung anhand retrospektiver Analysen, dass Patienten nach Verwendung von Zentrifugalpumpen weniger verwirrt gewesen sind.

Die Nullhypothese

In der Statistik würde basierend auf diesen Überlegungen im Rahmen der Studienplanung die Annahme „Es besteht kein Unterschied in der Auftretenswahrscheinlichkeit von kognitiven Defiziten zwischen der Kontroll- (= Rollerpumpe) und der Behandlungsgruppe (= Zentrifugalpumpe)“ als Nullhypothese H_0 formuliert werden. Die Alternativhypothese im Fall einer Ablehnung der Nullhypothese H_0 würde dann lauten: „Die Verwendung einer Zentrifugalpumpe führt zu einer geringeren Auftretenswahrscheinlichkeit kognitiver Defizite im Vergleich zu Rollerpumpen.“

Zwei grundsätzliche Fehler oder Fehlannahmen können nun bei der Interpretation und damit der Verallgemeinerung von Studienergebnissen auftreten.

Fehler 1. Art oder α-Fehler

Zum einen kann aufgrund der Ergebnisse aus der Studie gefolgert werden, dass sich die Behandlungen (Rollerpumpe im Vergleich zu Zentrifugalpumpe) in Bezug auf die Inzidenz kognitiver Defizite unterscheiden, obgleich dies in Wirklichkeit gar nicht zutrifft.

Dieser Fehler 1. Art (oder α-Fehler) misst die Wahrscheinlichkeit, mit der eine falsch-positive Schlussfolgerung getroffen wird. Die bei einem Test bzw. einer Untersuchung akzeptierte Wahrscheinlichkeit, bei einer Entscheidung einen Fehler 1. Art zu begehen, nennt man auch das Signifikanzniveau. Obwohl das Signifikanzniveau frei wählbar ist, findet man in der Literatur häufig ein Niveau von 5 %. Das heißt, der Untersucher zielt darauf ab, mit einer Wahrscheinlichkeit von weniger als 5 % eine falsch-positive Schlussfolgerung zu treffen (hier: die Schlussfolgerung zu treffen, dass Zentrifugalpumpen besser als Rollerpumpen seien mit Blick auf die Vermeidung von kognitiven Defiziten, obgleich dies nicht zutrifft). Die Festlegung des Signifikanzniveaus bei einer Irrtumswahrscheinlichkeit von kleiner oder gleich 5 % bedeutet aber auch, dass eine von 20 Untersuchungen (5 %), bei denen die Nullhypothese richtig ist, zu dem Schluss kommt, sie sei falsch (also einen Behandlungseffekt proklamiert, obgleich dieser nicht vorliegt). Auch bei dieser Annahme wird also stets ein bestimmtes Maß an falsch-positiven Ergebnissen akzeptiert.

Ursächlich für die Wahl des Wertes 5 % ist unter anderem der Umstand, dass eine normalverteilte Zufallsgröße nur mit einer Wahrscheinlichkeit von weniger als 5 % einen Wert annimmt, der sich vom Erwartungswert um mehr als die zweifache Standardabweichung unterscheidet (siehe Memobox 1).

Fehler 2. Art oder β-Fehler

Zum anderen können die Autoren der Publikation aufgrund der Studienergebnisse schlussfolgern, dass gar kein Behandlungseffekt (Unterschied zwischen Zentrifugal- und Rollerpumpe) vorliegt, die untersuchten Interventionen also unter den angenommenen Prämissen gleichwertig sind, obgleich sich die beiden Interventionen in Wirklichkeit unterscheiden. In diesem Fall läge eine falsch-negative Schlussfolgerung vor. Der Wert β (oder Fehler 2. Art) misst folglich die Wahrscheinlichkeit dieser falsch-negativen Schlussfolgerung.

Im Rahmen medizinischer Untersuchungen wird für das β-Fehler-Niveau in der Regel ein 4-mal so hoher Wert wie für das Signifikanzniveau α vorgeschlagen. Wenn, wie oben begründet, α bei 5 % liegt, sollte das β-Fehler-Niveau folglich 20 % betragen. Mit dieser Annahme ist die Intention verknüpft, dass die Wahrscheinlichkeit, eine falsch-negative Schlussfolgerung zu treffen, kleiner als 20 % ist. In der Tat wird bei Fallzahlabschätzungen häufig mit einem β-Fehler-Niveau von 20 % gearbeitet.

Liegt in einer Untersuchung die β-Fehler-Wahrscheinlichkeit (Wahrscheinlichkeit für einen Fehler 2. Art) unter dieser 20 %-Grenze, so ist die Teststärke oder auch „Power“ (sie wird berechnet als: 1 minus β) damit größer als 80 %. Die Power repräsentiert also die Wahrscheinlichkeit, eine falsch-negative Schlussfolgerung zu vermeiden. Positiv formuliert bedeutet dies, dass eine klinische Studie mit einer Power von 80 % mit einer Wahrscheinlichkeit von 80 % die Differenz zwischen zwei Behandlungsgruppen detektiert, sofern in Wirklichkeit ein Behandlungseffekt vorliegt (siehe ebenfalls Memobox 1).

Ereignisrate in der Kontrollgruppe

Die Ereignisrate in der Kontrollgruppe (hier: beobachtete Inzidenz von kognitiven Defiziten unter Einsatz der Rollerpumpe) ist zwar unschwer festzulegen, in der Praxis aber nicht immer einfach zu bestimmen. Zum einen würde eine präzise Festlegung der Ereignisrate in der Kontrollgruppe für die Fallzahlberechnung eine sorgfältige Voraberhebung voraussetzen, zum anderen können die noch so sorgfältig bestimmten

Fehler 1. Art oder α-Fehler
In der Statistik besteht beim Testen von Hypothesen ein Fehler 1. Art darin, eine Nullhypothese zurückzuweisen, obwohl sie wahr ist. Er spiegelt folglich die Wahrscheinlichkeit wider, einen Behandlungseffekt zu detektieren, obgleich kein Behandlungseffekt vorhanden ist (falsch-positives Ergebnis).

Fehler 2. Art oder β-Fehler
In der Statistik besteht beim Testen von Hypothesen ein Fehler 2. Art darin, dass eine Nullhypothese nicht abgelehnt wird, obwohl die Alternativhypothese korrekt ist. Er spiegelt folglich die Wahrscheinlichkeit wider, einen Behandlungseffekt zu übersehen, obgleich ein Behandlungseffekt mit definierter Größe vorhanden ist (falsch-negatives Ergebnis).

Memobox 1: Definition der Fehlermöglichkeiten

Daten im Rahmen der dann durchgeführten Studie durch definierte Ein- und Ausschlusskriterien (und damit die Selektion eines Patientenkollektivs, das sich von der „Normalpopulation" unterscheidet) erheblich davon abweichen.

Ereignisrate in der Behandlungsgruppe oder erwarteter Behandlungseffekt
Wird beispielsweise im skizzierten Beispiel von einer Inzidenz postoperativer kognitiver Defizite von 10 % unter Verwendung einer Rollerpumpe ausgegangen, muss als letztes zu definierendes Element im Rahmen der Fallzahlanalyse die erwartete Inzidenz in der Behandlungsgruppe definiert werden. Unter der Annahme (z. B. basierend auf retrospektiven Analysen bereits behandelter Patienten oder einer kleinen Vorstudie) einer absoluten Reduktion der Ereignisrate von 3 % (30%ige relative Risikoreduktion) würde die erwartete Inzidenz postoperativer kognitiver Beeinträchtigungen bei Verwendung der Zentrifugalpumpe dann bei 7 % liegen.

Ausgangssituation:
Es soll eine bereits bekannte Intervention (z. B. Rollerpumpe), die mit einer geschätzten Inzidenz eines Ereignisses (z. B. postoperative kognitive Defizite, wie im Text beschrieben) von 10 % assoziiert ist, gegen eine „neue" Intervention (z. B. Zentrifugalpumpe) getestet werden. Ein Behandlungseffekt im Sinne einer absoluten Risikoreduktion von 3 % wird als klinisch relevant angesehen. Die resultierende Inzidenz des beobachteten Ereignisses in der Behandlungsgruppe läge damit bei 7 %.

Weitere Annahmen:
α-Fehler = 0,05 (5 %)
β-Fehler = 0,2 (20 %), damit liegt die Power bei 80 % (1 minus β-Fehler)
n = erforderliche Fallzahl für die beiden untersuchten Gruppen
p_1 = Ereignisrate in der Behandlungsgruppe, hier: 7 %
p_2 = Ereignisrate in der Kontrollgruppe, hier 10 %
R = relatives Risiko (p_1 / p_2), hier: 7 % / 10 % = 0,7
K = Faktor, abhängig von der gewünschten Power und dem β-Fehler (Tabelle 1)

Die vereinfachte Formel lautet:

$$n = \frac{K\,[(R+1) - p_2\,(R^2+1)]}{p_2\,(1-R)^2}$$

Das Einsetzen der angenommenen Werte ergibt:

$$n = \frac{7{,}85\,[(0{,}7+1) - 0{,}1\,(0{,}7^2+1)]}{0{,}1\,(1-0{,}7)^2}$$

Im Ergebnis bedeutet dies:
n = 1352,81667 ≈ 1353 Patienten pro Gruppe.
Bei Verwendung des Statistikprogramms GraphPad InStat™ ergibt die Fallzahlberechnung mit 1422 Patienten pro Gruppe einen vergleichbaren Wert.

α-Fehler	Power (1 – β)		
	0,8	0,9	0,95
0,05	7,85	10,51	13,00
0,01	11,68	14,88	17,82

Tab. 1: Faktor (K) in Abhängigkeit von der gewünschten Power und dem α-Fehler

Memobox 2: Fallzahlabschätzung per einfacher Formel für binäre Ergebnisse

Abschätzung der erforderlichen Fallzahl

Mit den genannten Annahmen, die das Ziel der Studie, die zugrunde liegende Inzidenz unter herkömmlicher Therapie (Rollerpumpe) und den erwarteten Therapieeffekt (Zentrifugalpumpe) widerspiegeln, lässt sich durch einfache Statistikprogramme rasch die erforderliche Fallzahl pro Gruppe ermitteln.

Auch per Hand unter Zuhilfenahme komplexer Formeln lassen sich Fallzahlen berechnen. Ein einfacherer Rechenweg ist in Memobox 2 gezeigt.

Einfluss des α-Fehlers, der Power sowie der Ausgangs- und Behandlungsinzidenz auf die erforderliche Fallzahl

Die in Tabelle 1 gezeigten Faktoren zeigen unmissverständlich, was intuitiv verständlich ist: Je niedriger der α-Fehler sein soll und je größer die Power einer Studie sein soll, desto größer muss die Fallzahl sein.

Etwas versteckter, aber mit keineswegs geringerem Einfluss, wirkt sich die Ereignisrate in der Kontrollgruppe auf die Fallzahlabschätzung aus. Unter der Annahme eines gleichen relativen Risikos (R) hätte die Annahme eines doppelt so hohen Ausgangsrisikos (20 % statt 10 % in der Kontrollgruppe) – unter sonst gleichen Annahmen – eine deutlich niedrigere Fallzahlabschätzung von 612 Patienten pro Gruppe zur Folge.

Gleiches gilt für die Ereignisrate in der Behandlungsgruppe und damit den als relevant erachteten Behandlungseffekt. Die Annahme eines deutlicheren Behandlungseffektes (in unserem Beispiel: die Annahme einer noch niedrigeren Inzidenz kognitiver Dysfunktionen in der Behandlungsgruppe) würde zu einer drastischen Senkung der erforderlichen Patientenzahl führen. Für ein relatives Risiko von 0,35, also eine Reduktion der Ereignisrate von 10 % auf 3,5 %, sind gemäß der beschriebenen Formel lediglich 230 Patienten (bei Verwendung

des Statistikprogramms GraphPad InStat: 263 Patienten) pro Gruppe erforderlich.

Für eine grobe Abschätzung kann gelten, dass die erforderliche Fallzahl in etwa im umgekehrten Verhältnis zum Quadrat der Verringerung der Effektgröße wächst (d. h. eine Halbierung der Effektgröße hat eine mehr als vierfache Zunahme der Studiengröße zur Folge).

Diskussion und Ausblick

Die Bedeutung der Fallzahlanalyse im Vorfeld einer klinischen Studie lässt sich aus der Vielzahl an Studien erklären, die selbst in hochrangigen Journalen eine zu geringe Anzahl an untersuchten Patienten oder Probanden aufwiesen, gerade um negative Ergebnisse hinreichend zu untermauern [2]. Das Konzept der Stichprobenberechnung wurde in den folgenden Jahren von vielen Journalen aufgegriffen, die eine Anmerkung zur vorab getätigten, plausiblen Fallzahlberechnung als Annahmevoraussetzung definierten. Im bekannten CONSORT-Statement, das als Grundlage für die Planung, Durchführung und den Bericht klinischer Studien von vielen wissenschaftlichen Journalen in die „Hinweise für Autoren“ aufgenommen wurde, finden sich dezidiert Ausführungen zur Darlegung der Fallzahlabschätzung.

Die Intention ist gut und richtig: Es soll kein Patient an einer Untersuchung teilnehmen, ohne dass diese Bereitschaft zur Teilnahme an der Untersuchung nicht mit einem Nutzengewinn für die Wissenschaft oder die nachfolgenden Patienten in gleicher Situation verknüpft wäre. Letzteres wäre der Fall, wenn durch unzureichende Power im Falle fehlender Unterschiede zwischen den untersuchten Gruppen das Statement im Ergebnis- oder Diskussionsteil heißen muss: „Größere Studien an umfangreicheren Patientengruppen sind zur Beantwortung der bearbeiteten Fragestellung erforderlich.“

Die Ausführungen zu den Elementen der Fallzahlabschätzung machen aber vor allem eines deutlich: So gewissenhaft man auch an die Fragestellung herangehen mag, gerade in Bezug auf die Faktoren „Ereignisrate in der Kontrollgruppe“ und „erwarteter / als relevant erachteter Behandlungseffekt“ (bzw. Ereignisrate in der Behandlungsgruppe) kommen Fehlschätzungen naturgemäß sehr häufig vor. Selbst bei sehr konservativem Vorgehen und Berücksichtigung mutmaßlicher Einflussvariablen (Überschätzung von Therapieeffekten) und Beeinflussung der Ausgangsinzidenz durch Selektionsmechanismen oder Störgrößen kommen Fehleinschätzungen vor.

Nicht bestritten werden kann ferner die Incentivierung im Nachhinein durchgeführter und manipulierter Fallzahlabschätzungen, ist es doch einfach darzulegen, dass die tatsächlich beobachtete Inzidenz in der Behandlungsgruppe genau mit der angenommenen Inzidenz für die Fallzahlabschätzung übereinstimmt, als eine erhebliche Diskrepanz zu begründen.

Aus diesen Ausführungen wird deutlich, dass die Autoren der Fallzahlabschätzung grundsätzlich positiv, jedoch mit der nötigen Skepsis gegenüberstehen.

Manipulationen an Fallzahlabschätzungen lassen sich wohl am ehesten durch Registrierung klinischer Forschungsvorhaben vermeiden, für die bereits Foren verfügbar sind (www.clinicaltrials.org). Gleichwohl darf nicht vergessen werden, dass die Durchführung klinischer Studien ohnehin einen erheblichen Kraftakt bedeutet, der nicht durch weitere administrative Hürden verkompliziert werden sollte.

Die stringente Umsetzung der Anforderung des Arzneimittelgesetzes, der Good-Clinical-Practice-Empfehlungen und der gültigen EU-Direktiven auch für sogenannte „Investigator initiated trials“ (von Ärzten oder Forschern initiierte klinische Prüfungen) hat zur Folge, dass die administrativen Aufgaben zu sehr in den Vordergrund rücken. Dies hat unter Umständen zur Folge, dass eine aufschlussreiche Überprüfung getätigter Aussagen von publizierten Studien künftig sicher nicht mehr so zahlreich durchgeführt wird. Gerade deshalb ist zu fordern, dass die kritischen Überprüfungen von Verfahren und Interventionen nicht noch weiter mit Stolpersteinen versehen werden.

Da Untersuchungen an Patienten oder Probanden in der Regel zeit- und personalintensiv und damit schlussendlich kostenintensiv sind, liegt es im verständlichen Interesse des Untersuchers oder Sponsors einer Studie, die Patienten-/Probandenzahl so groß wie nötig, aber eben gleichzeitig so gering wie möglich zu halten. Für den Leser von Berichten über klinische Studien ist es unumgänglich, die Grundprinzipien der Fallzahlanalyse verstanden zu haben, um eine Wertung darüber zu treffen, ob z. B. eine Gleichwertigkeit der Interventionen aufgrund der Studienergebnisse tatsächlich vorliegt oder nur aufgrund einer zu niedrigen Power der präsentierten Untersuchung suggeriert wird.

Viel bedeutsamer als die Fallzahlplanung erscheint jedoch, dass alle erlangten Ergebnisse der Öffentlichkeit zugänglich gemacht werden, um ein verzerrtes Bild einer Intervention (positiv wie negativ) zu vermeiden. So gesehen ist die Politik vieler Journale, nur Studien mit „adäquater Power“ zu veröffentlichen, zwar gut gemeint (Ansporn, eine ausreichende Fallzahl zu berücksichtigen), im Sinne der „Wahrheitsfindung“ kurz- und mittelfristig jedoch nicht gut gemacht und keineswegs förderlich. Kleine Studien ohne „signifikanten Unterschied“, anderweitig möglicherweise sorgfältig geplant und durchgeführt, haben eine geringere Wahrscheinlichkeit, Eingang in die ausgewogene objektive Wertung zur Nützlichkeit einer Intervention zu finden, was schlussendlich die richtige Wahrnehmung zum Nutzen oder Schaden einer Intervention erschwert.

LITERATUR:

[1] Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF: Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA. 1996; 276: 637–639

[2] Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR: The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 „negative“ trials. N Engl J Med 1978; 299: 690–694

Weiterführende Literatur

Schulz K, Grimes D: Sample size calculations in randomized trials: mandatory and mystical. Lancet 2005; 365: 1348–1353

Priv.-Doz. Dr. Peter Kranke, MBA
Oberarzt
Klinik und Poliklinik für Anästhesiologie
Universitätsklinikum Würzburg
Zentrum Operative Medizin
Oberdürrbacher Str. 6
97080 Würzburg
E-Mail: peter.kranke@t-online.de