



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
Main Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2010

---

## **Wie viele Patienten brauche ich in meiner Studie? Fallzahlplanung für zwei typische Fragestellungen**

Held, U

Posted at the Zurich Open Repository and Archive, University of Zurich  
ZORA URL: <https://www.zora.uzh.ch/46202>  
Journal Article

Originally published at:

Held, U (2010). Wie viele Patienten brauche ich in meiner Studie? Fallzahlplanung für zwei typische Fragestellungen. Swiss Medical Forum, 10(40):685-686.

# Wie viele Patienten brauche ich in meiner Studie?

## Fallzahlplanung für zwei typische Fragestellungen

Ulrike Held

Horten-Zentrum, UniversitätsSpital, Zürich

Bei der Planung einer klinischen oder epidemiologischen Untersuchung ist die Festlegung des notwendigen Stichprobenumfangs von grosser Bedeutung. Regulatorische Instanzen und auch medizinische Fachzeitschriften verlangen immer häufiger eine detailliert dargestellte, nachvollziehbare und formale Berechnung des Stichprobenumfangs, die sog. Fallzahlplanung. In der Vergangenheit wurden sehr viele Studien mit zu wenigen Patienten durchgeführt, in denen dann der zu demonstrierende Effekt einer Behandlung nicht mit der angestrebten Präzision nachgewiesen werden konnte. Dieses Vorgehen ist als unethisch zu beurteilen, denn neben der Verschwendung von Geld und anderen Ressourcen setzen die Studienteilnehmer sich der Gefahr von Nebenwirkungen aus, ohne dass der Wirksamkeitsnachweis überhaupt zu erbringen ist. Leider sind häufig praktische Überlegungen wie auch Budgetrestriktionen eher die Basis der Fallzahlplanung als formale Berechnungen. Auf der anderen Seite ist ein zu grosser Stichprobenumfang ebenfalls unethisch, da mehr Menschen als nötig Placebo erhalten und die Einführung einer effektiven Therapie verzögert wird.

Bei der Formulierung der wissenschaftlichen Fragestellung muss man sich über drei wichtige Punkte Gedanken machen: (1.) Die Grösse des Effekts, den man nachweisen möchte. Man braucht eine kleinere Studie, um einen deutlichen Effekt nachzuweisen, und umgekehrt eine sehr grosse Studie, um einen kleinen Effekt aufzuzeigen. (2.) Das Signifikanzniveau, also die maximal «erlaubte» Wahrscheinlichkeit für einen Fehler der 1. Art. Je grösser die Evidenz sein soll, also je kleiner der p-Wert, desto grösser muss die Fallzahl sein. (3.) Die sog. *Power* oder Güte. Man wünscht sich eine hohe Wahrscheinlichkeit, dass bei Vorliegen eines Effekts dieser auch tatsächlich aufgedeckt wird.

Diese Überlegungen möchten wir nun für zwei Beispielsstudien konkret durchdenken und dann die Fallzahlen für beide Situationen berechnen.

Im ersten Beispiel soll ein den Blutdruck senkendes Medikament getestet werden. Zunächst muss das Studiendesign festgelegt werden, und wir entscheiden uns für eine randomisierte, plazebokontrollierte, *doppelt-verblindete* Studie. Dazu brauchen wir zwei Gruppen von Patienten mit erhöhtem Blutdruck: Die eine Gruppe soll das Medikament erhalten, die zweite Gruppe ein Placebo. Die Aufteilung auf die beiden Gruppen soll randomisiert erfolgen, das bedeutet also zufällig. Weder der behandelnde Arzt noch der Patient weiss, ob er Placebo oder Verum erhält (doppelte Verblindung). Im zweiten Schritt macht man sich darüber Gedanken, was man als einen relevanten Unterschied zwischen

der Placebo- und der Behandlungsgruppe ansieht. In diesem Fall soll der Unterschied mindestens 8 mm Hg betragen. Wir möchten also die Hypothese testen

$$H_0: \mu_0 - \mu_1 \geq 8$$

wobei  $\mu_1$  der Mittelwert in der Behandlungsgruppe ist und  $\mu_0$  der Mittelwert in der Placebogruppe. Des Weiteren muss man die Variabilität der Daten einbeziehen. Am besten ist es, wenn man eine andere Studie kennt, in der ein ähnlicher Zusammenhang schon einmal untersucht wurde und man daraus eine Schätzung für die Standardabweichung  $\sigma$  verwendet. Leider ist dieses Szenario eher unrealistisch. Hier ist aus einer Voruntersuchung bekannt, dass  $\sigma_0 = \sigma_1 = 14$  mm Hg ist. Wir möchten mit einer Wahrscheinlichkeit von höchstens 5% den Fehler 1. Art machen, also ist unser Signifikanzniveau  $\alpha = 0,05$ . Wir streben weiterhin eine Power von 90% an, das heisst, wir möchten die Nullhypothese bei einem tatsächlich vorliegenden Effekt mit einer Wahrscheinlichkeit von 90% verwerfen. Nun kann mit der folgenden Formel die nötige Fallzahl pro Gruppe berechnet werden:

$$n = \frac{(u + v)^2 (\sigma_1^2 + \sigma_0^2)}{(\mu_1 - \mu_0)^2}$$

In der Formel sind  $u$  und  $v$  noch nicht bekannt:  $u$  ergibt sich aus der angestrebten Power der Testentscheidung. In unserem Beispiel möchten wir eine Power von 90% erreichen, also brauchen wir das 90%-Quantil der Normalverteilung  $u_{0,9} = 1,28$ . Ähnlich hängt der Parameter  $v$  vom Signifikanzniveau  $\alpha$  ab, welches wir mit 0,05 festgelegt hatten. Dann entspricht  $v$  gerade dem  $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung, also bei uns  $v_{0,975} = 1,96$ . Wenn wir nun alles in die Formel einsetzen, erhalten wir die nötige Fallzahl pro Gruppe

$$n = \frac{(1,28 + 1,96)^2 (14^2 + 14^2)}{(-8)^2} = 64,3$$

Das heisst, man benötigt in jeder der beiden Gruppen 65 Patienten.

In einem zweiten Beispiel möchte man die Wirksamkeit eines neuen Multiple-Sklerose-Medikaments nachweisen, in welchem die Rate von Patienten mit mindestens einem Krankheitsschub innerhalb von 2 Jahren als primäre Zielgrösse betrachtet wird. Wieder vergleichen wir randomisiert, plazebokontrolliert und doppelt verblindet. Wir wissen aus einer anderen Studie, dass bei den Patienten in der Placebogruppe die Schubrate bei  $\mu_0 = 0,4$  im Zeitraum von 2 Jahren liegt, und wir erwarten bei der Verumgruppe eine Schubrate von nur



Ulrike Held

Die Autorin erklärt, dass sie keine Interessenkonflikte im Zusammenhang mit diesem Beitrag hat.

$\mu_1 = 0,25$ . Die Parameter  $u$  und  $v$  sind wie oben angegeben, und wenn wir wieder eine Power von 90% und ein Signifikanzniveau von 5% anstreben, erhalten wir

$$n = \frac{(u + v)^2 (\mu_1 + \mu_0)}{(\mu_1 - \mu_0)^2} = \frac{(1,28 + 1,96)^2 (0,25 + 0,4)}{(0,25 - 0,4)^2} = 303,3$$

Für diese Studie benötigen wir also 304 Patienten in der Verumgruppe und ebenso viele Patienten in der Plazebogruppe über einen Zeitraum von zwei Jahren, um die Wirksamkeit zu untersuchen.

Weitere typische Szenarien bei der Planung von klinischen Studien sind einfache Vergleiche einer Zielgrösse mit einem vorgegebenen Wert, also z.B. als Nullhypothese «der mittlere systolische Blutdruck in einer bestimmten Patientengruppe ist höchstens 110 mm Hg». Oder aber der Vergleich einer Rate oder eines Anteils mit einem festen vorgegebenen Wert. Ausserdem gibt es spezielle Formeln für die Fallzahlplanung in Fall-Kontroll-Studien. Eine gute Übersicht über die verschiedenen Methoden findet man in Kirkwood und Sterne (siehe Literaturangaben), ab Seite 420.

### Verlust von Patienten während der Studie

Man muss mit einem gewissen Verlust von Patienten über die Dauer der Studie rechnen oder auch mit dem Problem von einzelnen fehlenden Werten. Der Verlust von Patienten ist bestimmt höher, je länger die Studiendauer und je häufiger Nebenwirkungen des Medikaments auftreten. Andererseits werden bei einer langen Studiendauer möglicherweise auch Patienten aus der Plazebogruppe zu der Vermutung gelangen, sie erhielten das Medikament nicht, und aus diesem Grund ihre weitere Teilnahme verweigern. Aus all diesen Gründen sollte man sich Gedanken darüber machen, wie hoch die Ausfallrate in einer geplanten Studie sein wird, und die ursprüngliche Fallzahl um den nötigen Faktor erhöhen. Wenn wir schätzen, dass  $x\%$  der Patienten in der Abschlussanalyse nicht mit einbezogen werden können, dann sollte die geplante Fallzahl mit  $100/(100 - x)$  multipliziert werden. Wenn wir also z.B. von 15% ver-

lorenen Patienten ausgehen, dann sollte der Stichprobenumfang mit  $100/(100 - 15) = 1,18$  multipliziert werden bzw. um diesen Faktor erhöht werden.

## Glossar

### Einfach/doppelt-verblindete Studie

Man nennt eine Studie einfach-verblindet, wenn der Patient nicht weiss, in welcher Behandlungsgruppe er sich befindet. Bei doppelt-verblindeten Studien weiss auch der behandelnde Arzt es nicht. Mit einer Verblindung möchte man ausschliessen, dass der Patient und ggf. auch der Arzt in seiner Bewertung des Krankheitsbildes von seinem Wissen über die Behandlung beeinflusst wird.

### Power

Die Güte, Power oder Teststärke ist ein Mass für die Aussagekraft eines statistischen Tests. Es beschreibt, mit welcher Wahrscheinlichkeit ein Test (richtigerweise) die Nullhypothese verwirft, falls tatsächlich die Alternativhypothese gilt. Die Güte entspricht auch gerade dem Wert  $1 - \text{Wahrscheinlichkeit für Fehler 2. Art}$ . Man möchte normalerweise eine Güte von mindestens 80% erreichen.

---

### Korrespondenz:

Dr. Ulrike Held  
Horten-Zentrum  
UniversitätsSpital Zürich  
Bolleystrasse 40  
Postfach Nord  
CH-8091 Zürich  
[ulrike.held@usz.ch](mailto:ulrike.held@usz.ch)

---

### Empfohlene Literatur

- Held L, Rufibach C, Seifert B. Einführung in die Biostatistik. 4. Auflage. Zürich: Abteilung Biostatistik, Institut für Sozial- und Präventivmedizin der Universität Zürich; Juli 2009. <http://www.biostat.uzh.ch>.
- Hüsler J, Zimmermann H. Statistische Prinzipien für medizinische Projekte. 4. Auflage. Bern: Huber-Verlag; 2006.
- Kirkwood B, Sterne J. Essential Medical Statistics. 2. Auflage. Massachusetts: Blackwell Science; 2003.
- Kreienbrock L, Schach S. Epidemiologische Methoden. 4. Auflage. München: Elsevier-Verlag; 2005.