

# Drei Fallgruben bei Nicht-Unterlegenheitsstudien

Peter Kleist

## Quintessenz

- Nicht-Unterlegenheitsstudien bieten eine Reihe von Fallgruben. Die wichtigsten sind: Unwirksamkeit der Kontrolltherapie, Fehler bei der Festlegung des Grenzwerts für eine gerade noch akzeptable Unterlegenheit der Prüfbehandlung sowie Schwächen in Design und Durchführung der Studie.
- Die Wirksamkeit der Kontrollbehandlung wird nicht direkt gemessen, sondern vorausgesetzt. Häufig fehlen für die Annahme, dass die Kontrollbehandlung wirksam ist, ausreichende Grundlagen.
- Eine zu grosszügige Festlegung des Nicht-Unterlegenheitsgrenzwerts kann zur Schlussfolgerung von Nicht-Unterlegenheit führen, obwohl dem Unterschied faktisch eine klinische Relevanz zukommt.
- Ein schlechtes Studiendesign und eine mangelhafte Studiendurchführung verwischen potentiell vorhandene Unterschiede zwischen den Behandlungen.
- Oftmals lässt sich die Frage «Sind die Behandlungen gleich wirksam oder gleich unwirksam?» nicht beantworten.
- Der Beitrag liefert Hilfestellungen, die Qualität einer Nicht-Unterlegenheitsstudie zu beurteilen.

*«Zum Sieg gelangt man nicht, wenn man gut, sondern nur, wenn man besser spielt.»*  
Savielly Tartakower,  
russischer Schachspieler (1887–1956)

## Provokative Vorbemerkungen



Peter Kleist

Der Autor hat keine finanziellen oder persönlichen Verbindungen im Zusammenhang mit diesem Beitrag deklariert.

Eine Pub-Med-Recherche ergibt für das Stichwort *non-inferiority* über 1000 Treffer, wovon sich mehr als die Hälfte auf Veröffentlichungen aus den Jahren 2009 bis 2011 bezieht – Tendenz also stark steigend. Als Leser medizinischer Fachliteratur werden Sie daher zwangsläufig mit Nicht-Unterlegenheitsstudien (Noninferioritätsstudien) konfrontiert.

Nicht-Unterlegenheitsstudien sind in der medizinischen Wissenschaft etabliert. Zu ihrem inzwischen inflationären Gebrauch haben massgeblich zwei Aspekte beigetragen: Gibt es bereits eine wirksame Behandlung für eine Erkrankung, lässt sich eine plazebokontrollierte Studie zum Wirksamkeitsnachweis einer neuen Therapie kaum ethisch rechtfertigen. Allerdings sind viele neuen Therapieoptionen bezüglich Wirksamkeit bestenfalls gleichwertig zu den bestehenden; die Überprüfung einer allfälligen Überlegenheit der neuen Behandlung macht somit

keinen Sinn. Hinter Nicht-Unterlegenheitsstudien steht daher prinzipiell die Frage, ob sich zwei Behandlungen höchstens in einem irrelevanten Ausmass voneinander unterscheiden.

Nicht-Unterlegenheitsstudien haftet jedoch auch etwas Obskures an. Erstens ist der Begriff als solcher eine Fehlbezeichnung. Denn es geht eigentlich nicht um Nicht-Unterlegenheit, sondern um «Nicht-zuviel-Unterlegenheit». Zweitens sind sie eine Sonderform von Studien mit einer historischen Kontrolle. Dass Vergleiche mit historischen Daten problematisch sein können, wissen wir seit langem. Und drittens: Welcher andere Studientyp kann schon von sich behaupten, einem gewünschten Ergebnis durch ein schlechtes Studiendesign und eine schlechte Studiendurchführung Vorschub zu leisten?

Damit sind die drei wesentlichen Fallgruben von Nicht-Unterlegenheitsstudien bereits angesprochen. Im folgenden Beitrag soll es nur um solche Studien gehen, die vermeintlich aktive Therapien miteinander vergleichen und keine zusätzliche Plazebokontrolle vornehmen.

## 1. Fallgrube: Die Kontrollbehandlung ist nicht wirksam

In einer Nicht-Unterlegenheitsstudie setzt man die Wirkung einer Prüfbehandlung zu der einer Kontrollbehandlung in Beziehung. Die Wirkung der Kontrollbehandlung wird in der Studie allerdings nicht direkt gemessen, sondern auf der Basis von bereits vorliegenden Ergebnissen aus früheren plazebokontrollierten Studien vorausgesetzt. Oder anders ausgedrückt: Die Schlussfolgerung der Nicht-Unterlegenheit von einer Behandlung gegenüber einer anderen beruht nicht auf harten Fakten, sondern auf einer Annahme – nämlich dass sich Prüf- und Vergleichsbehandlung in einer alternativ durchgeführten plazebokontrollierten Studie als wirksam erwiesen hätten.

Nun stellt das kein Problem dar, wenn die Kontrollbehandlung in der Vergangenheit eindeutig und konsistent eine Überlegenheit gegenüber Plazebo aufgezeigt hat. Für viele gebräuchliche Arzneimittel, einschliesslich behördlich zugelassener, ist das jedoch nicht der Fall. Eine inkonsistente Datenlage findet man vor allem bei symptomatischen Therapien, selbst wenn die Studien mit ausreichender Teststärke versehen und ordnungsgemäss durchgeführt worden sind, und betreffen u.a. Antidepressiva, Antipsychotika, Analgetika, Antihistaminika, Antihypertensiva, Asthmaprophylaktika und Antiemetika [1]. Mögliche Ursachen für heterogene Studienergebnisse sind eine nur moderate Wirksamkeit der Therapie, variable Ansprechraten, eine hohe Plazeboresponse, un-

**Tabelle 1. Ergebnisse von sechs Studien, in denen das Antidepressivum Nomifensin jeweils mit Imipramin und Placebo verglichen wurde (adaptiert nach [3]).**

Studie	Ausgangswert (Hamilton-Depressionsskala)	Adjustierter Wert nach 4-wöchiger Therapie (Hamilton-Depressionsskala)			P-Wert (Nomifensin vs. Imipramin)
		Nomifensin	Imipramin	Placebo	
1	23,9	13,4	12,8	14,8	0,78
2	26,0	13,0	13,4	13,9	0,86
3	28,1	19,4	20,3	18,9	0,81
4	29,6	7,3	9,5	23,5	0,63*
5	37,6	21,9	21,9	22,0	1,0
6	26,1	11,2	10,8	10,5	0,85

\* P<0,001 für die Vergleiche von Nomifensin vs. Placebo und von Imipramin vs. Placebo.

#### Texttafel 1. Fragen zu Beurteilung der Validität einer Kontrollbehandlung in Nicht-Unterlegenheitsstudien.

- Ist die Kontrolle Therapiestandard, zumindest weit gebräuchlich oder die schlechteste unter mehreren Therapiealternativen?
- Hat die Kontrollbehandlung zuvor Wirksamkeit in mehreren randomisierten placebo-kontrollierten Studien mit ausreichend hohen Patientenzahlen gezeigt?
- Ist die Wirksamkeit der Kontrollbehandlung konsistent und vorhersehbar oder sehr unterschiedlich?
- Gibt es gut durchgeführte Studien mit der Kontrollbehandlung, in denen keine Wirksamkeit demonstriert werden konnte?
- Liegen Ergebnisse einer Meta-Analyse vor, die positive wie negative Studien berücksichtigt (vorzugsweise mit *Random-Effects-Modell* zur Adressierung der Heterogenität der Studien und ihrer Ergebnisse)?
- Wird die Kontrollbehandlung in der Nicht-Unterlegenheitsstudie so eingesetzt, wie sie sich zuvor als wirksam erwiesen hat (z.B. in Bezug auf die Dosis, Applikationsform und Therapiedauer)?
- Ist die Durchführung der Nicht-Unterlegenheitsstudie vergleichbar zu der von früheren Studien mit der Kontrollbehandlung (z.B. in Bezug auf Patientenauswahl, Erkrankungsschweregrad oder Studienendpunkt)?
- Stimmt das Ausmass der Wirksamkeit der Kontrollbehandlung in der Nicht-Unterlegenheitsstudie mit dem Wirkmass aus früheren Studien überein?

zureichende Compliance sowie eine starke Überlagerung durch *Regression-to-the-mean*-Effekte oder den Spontanverlauf einer Erkrankung. Gemäss einer Auswertung der amerikanischen Gesundheitsbehörde FDA konnte in 46% der Studien zu acht zugelassenen Antidepressiva und in 25% der Studien zu fünf zugelassenen Neuroleptika kein statistisch signifikanter Unterschied gegenüber Placebo demonstriert werden [2].

Wenn in einer Nicht-Unterlegenheitsstudie ein Vergleich zu einer Kontrollbehandlung vorgenommen wird, bei der man nicht sicher sein kann, ob sie auch in dieser Studie wirksam ist, lassen sich keine tauglichen Schlussfolgerungen ziehen. Denn eine Vergleichbarkeit der Wirkungen kann sowohl Ausdruck der Wirksamkeit als auch der Unwirksamkeit beider Therapien sein. Ein Beispiel soll das verdeutlichen (Tab. 1 ↩, nach [3]): Das aufgrund toxischer Effekte letztendlich nicht zugelassene Antidepressivum Nomifensin wurde in sechs Studien sowohl mit Imipramin als auch mit Placebo verglichen. Die unter Nomifensin ausgeprägte Reduktion der Werte auf der Hamilton-Depressionsskala zeugte von klinischer

Relevanz. Vor allem: In keiner der sechs Studien gab es einen signifikanten Unterschied zu Imipramin, dessen Überlegenheit gegenüber Placebo sich zuvor in vielen Studien erwiesen hatte. Nur aufgrund der gleichzeitig in den Studien vorgenommenen Placebokontrolle wurde deutlich, dass beide Antidepressiva faktisch nur in einer Studie (Tab. 1, Studie Nr. 4) wirksam waren.

Gemäss den bisherigen Ausführungen ist das Problem mit invaliden Kontrollbehandlungen vor allem bei Nicht-Unterlegenheitsstudien mit psychiatrischen Fragestellungen vorhanden. Es kann aber in jedem Indikationsbereich auftreten, wie ein abschliessendes Beispiel aus der Oto-Rhino-Laryngologie demonstrieren soll. In einer vor Kurzem publizierten Studie wurde gezeigt, dass bei Patienten mit Hörsturz die Wirksamkeit einer intratympanalen Kortikosteroidgabe der einer oralen Kortikosteroidtherapie nicht unterlegen ist [4]. Die Evidenz für die Wirksamkeit oraler Kortikosteroide – der Kontrollbehandlung – ist jedoch insuffizient und beruht im Wesentlichen auf einer 30 Jahre alten Untersuchung. Zudem sind die in der aktuellen Studie beobachteten Remissionsraten vergleichbar mit der Häufigkeit von Spontanremissionen unbehandelter Patienten [5]. Die Frage «Gleich wirksam oder gleich unwirksam?» bleibt im Raum stehen.

Die Texttafel 1 ↩ gibt Ihnen Hilfestellungen, die Eignung einer Kontrollbehandlung in einer Nicht-Unterlegenheitsstudie zu hinterfragen.

## 2. Fallgrube: fehlerhafte Annahmen für die Nicht-Unterlegenheit

Der eindeutigste Ansatz, die Gleichwertigkeit von zwei Behandlungen aufzuzeigen, würde im Nachweis ihrer absoluten Äquivalenz bestehen. Im klinischen Bereich ist das nicht möglich – ausser man weist die Überlegenheit der Prüfbehandlung nach. Doch genau das ist ja im Kontext von Nicht-Unterlegenheitsstudien auch nicht möglich. Daher wird in diesen Studien untersucht, ob die Prüfbehandlung höchstens in einem irrelevanten Ausmass schlechter ist als die Kontrollbehandlung (weshalb der Ausdruck *Nicht-Unterlegenheit* nicht ganz zutreffend ist).

Hierzu muss *vor* Beginn der Studie der Irrelevanzbereich festgelegt werden, konkret: Der sogenannte Nicht-Unterlegenheitsgrenzwert ist zu definieren. Wird dies versäumt oder erst nach Beendigung der Studie getan, ist die Studie unbrauchbar. Denn diesem Grenzwert kommt sowohl für die untersuchte Hypothese als auch für die Ermittlung der notwendigen Patientenzahl eine massgebliche Bedeutung zu. Die subjektive Post-hoc-Festlegung des Grenzwertes unter Kenntnis der Daten ist zudem anfällig für einen Bias und würde ausserdem eine Manipulation der Studie ermöglichen. Und was nicht oft genug wiederholt werden kann: Eine fehlgeschlagene Überlegenheitsstudie, die keinen statistisch signifikanten Unterschied zwischen zwei untersuchten Behandlungen aufzeigen konnte, ist nicht geeignet, Aussagen zur Nicht-Unterlegenheit zu machen. Denn dass die untersuchte Behandlung der Vergleichsbehandlung eventuell doch unterlegen ist, lässt sich aufgrund des Studiendesigns

nämlich nicht ausschliessen [6]. In der Veröffentlichung von Nicht-Unterlegenheitsstudien, die diesen Namen verdienen, sind daher zwingend die folgenden Angaben zu finden: a) eine klare Hypothese bezüglich Nicht-Unterlegenheit, b) die Festlegung des entsprechenden Grenzwerts und c) eine davon abgeleitete Fallzahlberechnung. Für das Vorgehen bei der Bestimmung des Nicht-Unterlegenheitsgrenzwerts gibt es keine goldene Regel. Üblicherweise geht man vom unteren Wert des Vertrauensbereichs (bzw. des Konfidenzintervalls) der Kontrollbehandlung aus früheren plazebokontrollierten Studien aus, wie das folgende hypothetische Beispiel zeigen soll: Hat die Kontrollbehandlung die Ereignisrate (z.B. Myokardinfarkte) im Vergleich zu Plazebo um 2,5% gesenkt (Vertrauensintervall von 2,0 bis 3,1%), spiegelt der untere Wert des Vertrauensintervalls (2,0%) – vereinfacht ausgedrückt – den Mindesteffekt der Behandlung wider. Würde man jetzt für eine neue Therapie in einer Nicht-Unterlegenheitsstudie einen Grenzwert von 2% akzeptieren (ebenfalls unterer Wert des Vertrauensbereichs, in diesem Fall für den Unterschied zwischen Prüf- und Kontrollbehandlung), wäre unter Ausnutzung des Spielraums der komplette Effekt der Vergleichsbehandlung aufgebraucht und keine Wirksamkeit gegenüber Plazebo mehr vorhanden. Aus diesem Grund nimmt man für die Grenzwertbestimmung meistens nur eine Fraktion der unteren Vertrauenslimite für die Kontrollbehandlung gegenüber Plazebo – z.B. 50%, entsprechend der Hälfte des gezeigten Effektes. In unserem Beispiel hätte das einen Nicht-Unterlegenheitsgrenzwert von  $0,5 \times 2\% = 1\%$  zur Folge. Die Festlegung des Grenzwerts ist letztlich keine statistische Entscheidung, sondern muss einer Überlegung auf der Basis von klinischen Gesichtspunkten folgen. Zwei Voraussetzungen müssen dabei auf jeden Fall erfüllt sein: a) die Prüfbehandlung ist immer noch besser als Plazebo und weicht b) maximal um einen klinisch nicht relevanten Unterschied von der Vergleichsbehandlung ab. Welchem Unterschied klinische Relevanz zukommt, ist mitunter schwierig zu bestimmen, denn oftmals gibt es keine offiziellen Expertenempfehlungen. Als Faustregel gilt: Je ernsthafter die Erkrankung, desto konservativer muss die Grenzwertfestsetzung sein. So kostet jeder «Kompromiss» in Bezug auf den Endpunkt Mortalität Menschenleben.


Dass in einigen Fällen ein zu grosszügiger Grenzwert gewählt wird, hat zwei Gründe: a) Einer vermeintlich schwächer wirksamen Therapie soll dennoch das Siegel der Nicht-Unterlegenheit zukommen, und b) dieses Vorgehen soll die benötigte Fallzahl reduzieren, um die Machbarkeit der Studie zu verbessern. Das Problem ist, dass ein inadäquater Grenzwert die Schlussfolgerung von Nicht-Unterlegenheit begünstigen kann, obwohl die Prüfbehandlung klinisch relevant schlechter wirkt als die Kontrollbehandlung und sich ihre Wirkung kaum oder gar nicht mehr von Plazebo unterscheidet. Ein konkretes Beispiel aus der Literatur soll dies verdeutlichen.

Der Thrombin-Inhibitor Ximelagatran sollte in einer Nicht-Unterlegenheitsstudie bei Patienten mit Vorhofflimmern mit Warfarin verglichen werden. Als Grenzwert für die Nicht-Unterlegenheit wurde eine 2%-ige jährliche Zunahme von Schlaganfällen und thromboembolischen Ereignissen gegenüber Warfarin festgelegt [7].

Zum Zeitpunkt der Studienplanung lagen bereits die Ergebnisse aus zwei Meta-Analysen vor, die für Warfarin eine jährliche Schlaganfall-Reduktion von im Mittel 3,1 bzw. 2,8% im Vergleich zu Plazebo zeigten [8, 9]. In der *Random-Effects*-Meta-Analyse zu Warfarin [9] betrug der Vertrauensbereich 1,4–4,2%. Im Hinblick auf das untere Ende dieses Vertrauensbereichs (1,4%) wäre es also bereits vom Ansatz her möglich gewesen, für Ximelagatran eine Nicht-Unterlegenheit gegenüber Warfarin nachzuweisen, ohne ausschliessen zu können, dass die Wirkung schlechter als die von Plazebo ist. Faktisch betrogen nun die jährlichen Ereignisraten am Studienende 1,6% unter Ximelagatran und 1,2% unter Warfarin, mit einem Vertrauensbereich für die Behandlungsdifferenz von 0,13 bis 1,03% [10]. Basierend auf dem zuvor festgelegten Grenzwert von 2% konstatierten die Autoren der Studie die Nicht-Unterlegenheit von Ximelagatran gegenüber Warfarin – aber bei einer adäquaten Berücksichtigung der Datenlage zu Warfarin hätte selbst ein Nicht-Unterlegenheitsgrenzwert von 1% (anstatt 2%) das Fehlschlagen der Studie nicht verhindern können.

### 3. Fallgrube: Das Design und die Durchführung der Studie sind unzureichend

Im Gegensatz zu einer Studie zum Nachweis der Überlegenheit weist bei einer Nicht-Unterlegenheitsstudie die Verblindung der Behandlungen keinen Schutz gegenüber einem Bias auf. Theoretisch könnten alle von den Patienten entnommenen Proben gemischt oder immer der gleiche Messwert in die Datenbögen eingetragen werden, um alle Unterschiede aufzuheben. Weniger extrem, dafür schon realitätsnäher sind die folgenden Beispiele: Durch die Auswahl von Patienten mit einem niedrigen Erkrankungsschweregrad gibt man der schwächer wirksamen Behandlung eine gute Chance, ein zur Kontrollbehandlung vergleichbares Wirkungsausmass zu erreichen. Oder durch die Überbetonung von potentiellen Nebenwirkungen bei der Aufklärung der Patienten bzw. eine lange Studiendauer wird dafür gesorgt, dass nur eine Minderheit der Patienten den Therapieempfehlungen Folge leistet; dann spiegelt das Studienergebnis weitgehend den Spontanverlauf der Erkrankung wider [11].

Ob beabsichtigt oder unbeabsichtigt: Schwächen im Design der Studie und vor allem in ihrer Durchführung haben den gleichen Effekt: Eventuell vorhandene Unterschiede zwischen den Behandlungen werden verwischt und einer Nicht-Unterlegenheit als Endresultat der Studie Vorschub geleistet. Dementsprechend kommen qualitätssichernden Massnahmen und der objektiven Überprüfung der Studienendpunkte eine hohe Bedeutung zu. In der Texttafel 2  sind verschiedene Faktoren aufgelistet, die ein Studienergebnis in Richtung Nicht-Unterlegenheit beeinflussen. Sie sollen Ihnen eine Hilfestellung bieten, die richtigen Fragen zu stellen und die Qualität einer Nicht-Unterlegenheitsstudie beurteilen zu können.

Dass die in der Texttafel 2 genannten Aspekte nicht nur Theorie, sondern Realität sind, sollen die folgenden konkreten Beispiele aus der Literatur belegen: Eine Studie bei Patienten mit Schizophrenie kam zu dem Ergebnis, dass eine Behandlung mit Olanzapin der einer mit

**Texttafel 2. Faktoren, die ein Studienergebnis in Richtung Nicht-Unterlegenheit beeinflussen.**

- Inadäquates Titrationsschema/inadäquate Dosierung der Kontrollbehandlung
- Inhomogene Patientenpopulation
- Niedriger Erkrankungsschweregrad → geringer Bedarf für eine ausgeprägte Wirkung
- Inadäquate Behandlungs- bzw. Beobachtungsdauer
- Keine standardisierten Messungen
- Hohe Endpunktvariabilität
- Erlaubte Begleitmedikation, die Einfluss auf den Studienendpunkt nimmt
- Viele Non-Responder
- Hohe Anzahl von Therapie- bzw. Studienabbrechern
- Hohe Anzahl von Protokollverletzern
- Niedrige Compliance
- Viele fehlende Messwerte
- *Intention-to-treat*-Auswertung (eher als Per-Protokoll-Auswertung)

Clozapin nicht unterlegen ist; allerdings sah das Protokoll zu wenig Titrationsschritte vor, so dass die Mehrzahl der mit Olanzapin behandelten Patienten bis zum Studienende keine Dosierung im therapeutischen Zielbereich erhielt [12]. Eine weitere Studie bei Patienten mit Anämie konstatierte die Gleichwertigkeit zweier Behandlungsschemata mit Erythropoietin, obwohl nur 63% der Patienten eine protokollgemässe Behandlung erhielten [13]. Und am Ende von grossen Antihypertensiva-Vergleichsstudien nahmen nur noch zwei Drittel der Patienten die ursprünglich vorgesehene Studienmedikation ein [14].

**Ernüchternde Schlussbemerkungen**

Ob eine Nicht-Unterlegenheitsstudie valide ist oder nicht, hängt eng mit ihrer sogenannten Assay-Sensitivität zusammen. Diese bezeichnet die Fähigkeit einer Studie,

**CME [www.smf-cme.ch](http://www.smf-cme.ch)**

1. Ein Studienergebnis belegt die Nicht-Unterlegenheit der Prüfbehandlung zur Vergleichsbehandlung. Welcher der folgenden Aspekte lässt die Schlussfolgerung der Nicht-Unterlegenheit trotzdem *fraglich* erscheinen?

- Es wurde ein Parallelgruppen-Design für die Studie verwendet.
- Die Rate an Patienten, die die Therapie nur unregelmässig einnahmen, betrug 2% unter der Prüfbehandlung und 4% unter der Vergleichsbehandlung.
- Die Ergebnisse der Intention-to-treat- und der Per-Protokoll-Auswertung wichen geringfügig voneinander ab.
- In der 6-monatigen Therapiestudie lagen stabile Messwerte des Endpunkts bereits nach 4 Monaten, unter der Vergleichsbehandlung erst nach 5 Monaten unter der Prüfbehandlung vor.
- In beiden Studiengruppen brach mehr als die Hälfte der Patienten die Behandlung vorzeitig ab.

eine Differenzierung zwischen verschiedenen Behandlungen vornehmen zu können, oder anders gesagt: eine wirksame von einer weniger oder gar nicht wirksamen Behandlung zu unterscheiden. Alle in diesem Beitrag geschilderten Fallgruben betreffen Situationen, in denen die Assay-Sensitivität nicht (oder nicht ausreichend) gegeben ist. Bei einer inadäquaten Auswahl der Kontrollbehandlung wissen wir nicht, ob diese überhaupt wirksam ist und als Referenz dienen kann; bei einem zu grosszügig gewählten Nicht-Unterlegenheitsgrenzwert oder einer schlechten Durchführung der Studie werden möglicherweise bestehende Unterschiede zwischen Prüf- und Kontrollbehandlung verwischt. Letztlich wird in diesen Fällen die Schlussfolgerung von Nicht-Unterlegenheit zu Unrecht gezogen. In der Realität bleibt somit die Frage, ob zwei Behandlungen gleich wirksam oder gleich unwirksam sind, in vielen Fällen offen.

Handelt es sich hier um Einzelfälle? Sicher nicht, denn noch vor wenigen Jahren kam eine systematische Untersuchung zu dem Ergebnis, dass nicht einmal jede fünfte Studie den an Nicht-Unterlegenheitsstudien gestellten Qualitätsanforderungen genügt [15]. Publizierten Nicht-Unterlegenheitsstudien sollten Sie also nur die Hälfte glauben. Hoffentlich hilft Ihnen dieser Beitrag, eine Antwort auf die Frage «Aber welcher Hälfte?» zu finden.

**Korrespondenz:**

Dr. med. Peter Kleist  
GlaxoSmithKline AG  
Talstrasse 3-5  
CH-3053 Münchenbuchsee  
[peter.m.kleist@gsk.com](mailto:peter.m.kleist@gsk.com)

**Empfohlene Literatur**

- Piaggio G, Elbourne DR, Altman, DG, et al. Reporting of noninferiority and equivalence randomized trials. An extension of the CONSORT statement. *JAMA*. 2006;295:1152–60.  
Die vollständige nummerierte Literaturliste finden Sie unter [www.medicalforum.ch](http://www.medicalforum.ch).

2. In einer Studie wird als Grenzwert für die Nicht-Unterlegenheit der Prüfbehandlung gegenüber der Standardtherapie eine relativ um 20% höhere Ereignisrate festgesetzt. Nach Abschluss der Studie betragen die absoluten Ereignisraten in beiden Gruppen genau 1%. Welches Vertrauensintervall für die Behandlungsdifferenz spricht für die Nicht-Unterlegenheit der Prüfbehandlung?

- 0,5–1,5%.
- 0,5–0,75%.
- 0,25–2%.
- 0,9–1,1%.
- Die Ereignisrate ist in beiden Gruppen gleich, somit lässt sich das Vertrauensintervall nicht berechnen.

## **Drei Fallgruben bei Nicht-Unterlegenheitsstudien /**

### **Trois pièges dans les études de non-infériorité**

#### **Literatur (Online-Version) / Références (online version)**

- 1 Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. *Ann Intern Med* 2000; 133: 455-63.
- 2 Laughren TP. The scientific and ethical basis for placebo-controlled trials in depression and schizophrenia: an FDA perspective. *Eur Psychiatry* 2001; 16: 418-23.
- 3 Leber P. Hazards of inference: the active control investigation. *Epilepsia* 1989; 30 (Suppl 1): S57-63.
- 4 Rauch SD, Halpin CF, Antonelli PJ, et al. Oral vs intratympanic corticosteroid therapy for idiopathic sudden sensorineural hearing loss. A randomized trial. *JAMA* 2011; 305: 2071-9.
- 5 Piccirillo JF. Steroids for idiopathic sudden sensorineural hearing loss. Some questions answered, others remain (editorial). *JAMA* 2011; 305: 2114-5.
- 6 Anderson P. Absence of evidence is not evidence of absence. *BMJ* 2004; 328: 476-7.
- 7 Halperin JL, and the Executive Steering Committee, on behalf of the SPORTIF III and V Study Investigators. Ximelagatran compared with warfarin for prevention of thromboembolism in patients with nonvalvular atrial fibrillation: Rationale, objectives, and design of a pair of clinical studies and baseline characteristics (SPORTIF III and V). *Am Heart J* 2003; 146:431-8.
- 8 Hart RG, Benavente O, McBride R, et al. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: A meta-analysis. *Ann Internal Med* 1999; 131:492-501.
- 9 [No authors listed]. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomized controlled trials. *Arch Intern Med* 1994; 154: 1449-57.
- 10 SPORTIF Executive Steering Committee for the SPORTIF V Investigators. Ximelagatran vs Warfarin for stroke prevention in patients with nonvalvular atrial fibrillation. A randomized trial. *JAMA* 2005; 293: 690-8.
- 11 Lüdtke R. Nicht unterlegen, oder doch? Editorial. *Forsch Komplementärmed* 2006; 13: 332-3.
- 12 Tollefson GD, Birkett MA, Kiesler GM, et al. Double-blind comparison of olanzapine versus clozapine in schizophrenic patients clinically eligible for treatment with clozapine. *Biol Psychiatry* 2001; 49: 52-63.
- 13 Steensma DP, Molina R, Sloan JA, et al. Phase III study of two different schedules of erythropoietin in anemic patients with cancer. *J Clin Oncol* 2006; 24: 1079-89.
- 14 McAlister FA, Sackett DL. Active-control equivalence trials and antihypertensive agents. *Am J Med* 2001; 111: 553-8.
- 15 Le Henaff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA* 2006; 295: 1147-51.