

Studien sind einfach!

Statistik ist es nicht (immer)

Gibt es für eine neue Therapie keine etablierte Standardbehandlung als Vergleich, darf eine Plazebostudie durchgeführt werden. Gibt es Vergleichstherapien mit gut dokumentierter Wirksamkeit, kann für eine neue Therapie eine Überlegenheits-, Äquivalenz- oder Nichtunterlegenheitsstudie geplant werden. Die Charakteristika dieser methodischen Ansätze wurden an einem Symposium zuhanden der Nichteingeweihten näher erklärt.

Klinische Studien können viele Endpunkte haben, die sich in ihrer Qualität unterscheiden», sagte Stefan Anker, Professor für Innovative Klinische Studien an der Universitätsmedizin Göttingen. Die Mortalität ist ein objektiver Endpunkt, aber oft nicht der wichtigste. Zusammengesetzte Endpunkte, beispielsweise aus Mortalität und Morbidität (z.B. Hospitalisationen, Eingriffe), haben sich bei vielen

«15 bis 20 Prozent aller Ereignisse im Rahmen von Studien, insbesondere Hospitalisationen, werden nie bekannt.»

Fragestellungen durchgesetzt, bieten aber auch Probleme, da «harte» und «weiche» Endpunkte vermischt werden. Oft zählt man Ereignisse, dabei ergeben sich aber zwei Probleme: Einerseits gibt es Ereignisse, die nicht den Krankheitsprozess widerspiegeln und die wir nicht beeinflussen können, zum Beispiel das Auftreten von Krebs unter einem kardiovaskulären Medikament. Andererseits gibt es Ereignisse, die den Krankheitsprozess widerspiegeln und die wir beeinflussen können, die aber nicht gezählt werden, weil sie nicht bekannt oder in den Analysen beiseitegelassen werden.

Wiederkehrende Hospitalisationen bedürfen der Aufmerksamkeit

«15 bis 20 Prozent aller Ereignisse im Rahmen von Studien, insbesondere Hospitalisationen, werden nie bekannt», so Anker. In der CHARM-Preserved-Studie mit Candesartan beruhte die Häufigkeit der Hospitalisationen auf einigen wenigen Patienten, die sehr oft hospitalisiert wurden, was zu einer Verzerrung des Resultats führte. In der TOPCAT-Studie führte die Behandlung mit Spironolacton im Vergleich zu Plazebo zu einer 11-prozentigen Reduktion beim kombinierten Endpunkt aus kardiovaskulärer Mortalität, Reanimation wegen Herzstillstands und Hospitalisationen wegen Herzversagens. Letztere wurden um 17 Prozent reduziert. Auch hier wurde das Resultat durch wiederkehrende Hospitalisationen verzerrt.

«Tod verhindert Hospitalisationen»

Wichtig ist, dass bei einer Kombination von Zeit bis zum Tod und Zeit bis zur ersten Hospitalisation ein Studienteilnehmer schon sterben kann, bevor er hospitalisiert wird und umgekehrt ein Patient möglicherweise nicht nur einmal, sondern mehrere Male eingeliefert wird, bevor sein Tod gezählt wird. Bei einer «Time-to-first-event»-Analyse zählt keines der Folgeereignisse, damit wird die Last der durch Herzinsuffizienz bedingten Ereignisse unterschätzt. Am Beispiel der SHIFT-Studie erläuterte Anker den praktisch bedeutsamen Unterschied verschiedener Analysemethoden. In der «Time-to-first-event»-Analyse verhinderte Ivabradin 47 Hospitalisationen wegen Herzinsuffizienz pro 1000 behandelte Patienten. In der «Repeat-events»-Analyse waren es hingegen 93 Hospitalisationen. «Tod verhindert Hospitalisationen. Diese Analysen wurden nicht für diesen Effekt korrigiert», kommentierte Anker. Gerade «Repeat-events»-Analysen sind unter Kosteneffektivitätsgesichtspunkten wichtig.

Grösstes Problem sind Berechnungen der statistischen Aussagekraft im Zusammenhang mit wiederkehrenden Ereignissen. Hier helfen nur umfangreiche Modellrechnungen. Liegen individuelle Patientendaten vor, können auch Metaanalysen von «Time-to-first-event»-Analysen und «Repeat-events»-Analysen durchgeführt werden. Sie zeigten zum Beispiel für die intravenöse Behandlung mit Ferrocarmaltose bei Herzinsuffizienz in beiden Berechnungsmethoden signifikante Ergebnisse beim kombinierten Endpunkt Mortalität und Hos-

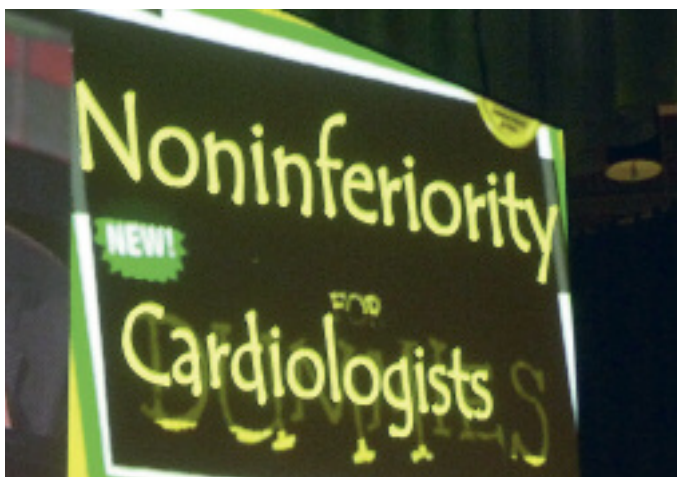


Foto: MÜ

Wer mit Studien etwas beweisen möchte, sollte sich gut auskennen – denn nicht jedes Studiendesign erlaubt auch jede Aussage.

Was Nichtunterlegenheitsstudien («Non-inferiority Trials») uns sagen

Ein nicht signifikantes Ergebnis in einer Überlegenheitsstudie («Superiority Trial») ist kein Beweis für Äquivalenz.

Die drei Studiendesigns haben unterschiedliche Ziele:

- Überlegenheitsstudie: A ist besser als B.
- Äquivalenzstudie: A ist nicht (allzu) verschieden von B.
- Nichtunterlegenheitsstudie: A ist nicht (viel) schlechter als B.
- Bei einer aktiven Kontrolle muss sichergestellt sein, dass diese in einem Vergleich mit Placebo überlegen war:
 - Dies wird oft erreicht, indem frühere plazebokontrollierte Studien beigezogen werden.
 - Konstanzvermutung: Historische Differenz zwischen aktiver Kontrolle und Placebo bleibt erhalten.
- Eine neue Therapie muss innerhalb die Nichtunterlegenheitsgrenzen der aktiven Kontrolle fallen.
- Man kann Therapie-versus-Kontrolle-Daten und Kontrolle-versus-Placebo-Daten benutzen, um das Therapie-versus-Placebo-Verhältnis abzuleiten (putatives Placebo).
- Eine schlechte Nichtunterlegenheitsstudie kann zu einem positiven Ergebnis führen.

pitalisation sowie bei kardiovaskulären Hospitalisationen und bei Hospitalisationen wegen Herzinsuffizienz.

«Randomisierte Studien haben die medizinische Praxis revolutioniert, da sie genaue Schätzungen der Therapiewirksamkeit erlauben, und dies erreichen wir, indem wir den Berechnungsfehler (Bias) der Therapiewirksamkeit reduzieren.»

Intention to treat oder On-Treatment?

«Randomisierte Studien haben die medizinische Praxis revolutioniert, da sie genaue Schätzungen der Therapiewirksamkeit erlauben», daran erinnerte Prof. Stuart J. Conolly, McMaster University, Hamilton, Ontario, «und dies erreichen wir, indem wir den Berechnungsfehler (Bias) der Therapiewirksamkeit reduzieren.» In klinischen Studien können Irrtümer entstehen einerseits durch zufällige biologische Variation, andererseits durch nicht zufällige, systematische Unterschiede zwischen den Studienarmen (Bias).

Die Randomisierung soll einen Bias bei der Zuteilung in Behandlungsarme eliminieren. Da Ärzte dazu neigen, Patienten in verschiedenen Behandlungsarmen unterschiedlich zu behandeln, kann eine doppelblinde Therapieanordnung diesen Bias minimieren.

Die Intention-to-Treat-(ITT)-Analyse berücksichtigt alle randomisierten Patienten – unabhängig davon, ob sie die ihnen zugewiesene Behandlung erhalten haben oder nicht. Die ITT-Analyse gibt das wieder, was in der Praxis passiert – dass Patienten die Therapie nicht befolgen, sie vorzeitig abbrechen oder zu Kontrollen nicht erscheinen.

Demgegenüber schließt die On-Treatment-(OT)-Analyse typischerweise Patienten aus, welche die Therapie gar nicht erhalten haben. Hat ein Patient die Behandlung gestoppt, er-

folgt keine weitere Nachbeobachtung. Die OT-Analyse ist erklärend, sie schätzt ab, was passiert, wenn der Patient die Therapie tatsächlich anwendet.

Beide Analysemethoden haben Vor- und Nachteile. Die ITT-Analyse eliminiert den Zuteilungsbias und ist der Goldstandard für Zulassungsstudien. Der Nutzen der OT-Analyse liegt darin, dass sie mechanistische Informationen gibt, die auch Hypothesen für zukünftige Studien generieren. Sie wird oft in Sicherheitsanalysen eingesetzt, wenn es darum geht, die beste Schätzung für einen tatsächlichen Effekt eines Medikaments zu erstellen. Die ITT-Analyse unterschätzt den «wahren» Therapienutzen, da Patienten, welche die Therapie gar nicht bekommen haben, das Ergebnis «verwässern». Die OT-Analyse birgt immer die Gefahr eines potenziellen Bias, da sie die Randomisierung durchbricht.

Überlegenheit, Äquivalenz, Nichtunterlegenheit

In den späten Neunzigerjahren war etabliert, dass ACE-Hemmer bei Herzinsuffizienz einen therapeutischen Nutzen bieten. Dann kamen die Angiotensinrezeptorblocker (ARB) mit dem Anspruch, eine theoretisch komplettere Blockade des Renin-Angiotensin-Aldosteron-Systems (RAAS) zu bewirken. In einer sehr kleinen Studie (ELITE) erwies sich der ARB Losartan hinsichtlich der kardiovaskulären Endpunkte dem damals aktuellen ACE-Hemmer Captopril bei älteren Patienten mit Herzinsuffizienz als nicht überlegen. In einer späteren Studie (ELITE II) sollte der in ELITE überraschend beobachtete Mortalitätsvorteil für Losartan nachgewiesen werden. Dies gelang in dieser Überlegenheitsstudie an viel grösseren Patientenzahlen jedoch nicht (immerhin wurde die bessere Verträglichkeit dokumentiert). «Man kann natürlich fragen, weshalb in ELITE Losartan nicht gegen Placebo verglichen wurde», sagte Prof. Scott D. Solomon, Harvard University, Brigham and Women's Hospital, Boston. Dies wäre jedoch in Konflikt mit der Helsinki-Deklaration gestanden, die verlangt, dass eine neue Methode mit der besten verfügbaren Methode verglichen wird. Dies war in den Augen der ELITE-II-Autoren ein ACE-Hemmer, aber sie gingen von der Hypothese aus, dass der ARB besser sei als der ACE-Hemmer. Da sie keine Nichtunterlegenheitsstudie durchgeführt hatten, konnten sie auch nicht beweisen, dass Losartan nicht schlechter war als Captopril.

Die meisten klinischen Studien werden so ausgestaltet, um zu zeigen, dass eine neue Therapie einer alten überlegen ist. Es handelt sich also um Überlegenheitsstudien. Manchmal besitzen neue Therapien aber keine oder nur eine sehr geringfügige Überlegenheit gegenüber etablierten Therapien, zeichnen sich aber durch weniger Nebenwirkungen, geringere Invasivität, geringere Kosten, einfachere Dosierungsschemata oder eine längere Patentlebensdauer aus. Fallen solche Gesichtspunkte ins Gewicht, kann sogar eine bescheidene Wirksamkeitseinbusse akzeptabel erscheinen, wie Solomon meinte.

Nichtunterlegenheitsstudien haben auch eine dunkle Seite

Worum handelt es sich bei einer Nichtunterlegenheitsstudie? Dies ist eine Studie, die belegt, dass eine neue Therapie nicht schlechter – oder nicht sehr viel schlechter – ist als eine bestehende. Hier kommt die Nullhypothese ins Spiel: In einer Überlegenheitsstudie lautet die Nullhypothese, dass die zwei verglichenen Therapien gleich effektiv sind. Der Nachweis der Überlegenheit gelingt, indem die Nullhypothese verworfen

wird. Das Gegenteil trifft jedoch nicht zu: Ungenügende Evidenz, um die Nullhypothese zu verwerfen, bietet nicht genügend Evidenz, um sie zu akzeptieren. In einer Nichtunterlegenheitsstudie verhält sich die Sache mit der Nullhypothese im Grunde genau umgekehrt. Die Studie ist darauf ausgelegt, die Nullhypothese zu widerlegen, dass zwischen den verglichenen Therapien mindestens ein minimaler nachweisbarer Unterschied besteht. Wichtig sind in diesem Zusammenhang die Konfidenzintervalle und die Festlegung der Grenzen, von ihnen hängt ab, «wie viel unterlegen» die untersuchte Therapie ist. Äquivalenzstudien erlauben die Aussage: «Mit 95 Prozent Sicherheit kann ich sagen, dass die Mortalitätsrate (oder irgendein anderer Parameter) für beide Vergleichsmedikamente innerhalb von 1 Prozent liegt.» Demgegenüber erlaubt eine Nichtunterlegenheitsstudie die Aussage: «Mit 95 Prozent Sicherheit kann ich sagen, dass Medikament A nicht mehr als 1 Prozent schlechter ist als Medikament B.»

Welches sind die Ansprüche an eine Nichtunterlegenheitsstudie? Die aktive Kontrolle muss ein etablierter Standard für die untersuchte Indikation und darf keinesfalls anderen bekannten Therapien unterlegen sein. Für die aktive Kontrolle muss ein Effektivitätsnachweis vorliegen, und zwar in der gewählten Formulierung und Dosis und aus neuester Zeit. Zudem müssen die Forscher erklären, was sie unter Nichtunterlegenheit beziehungsweise Äquivalenz verstehen. Eine sehr wichtige und anspruchsvolle Arbeit für Statistiker ist die Festlegung der Grenzen.

«Es gibt auch eine dunkle Seite der Nichtunterlegenheit», betonte Solomon. «Sie können Nichtunterlegenheit «beweisen», indem Sie eine schlechte Studie machen.»

- Schlechte Patienten auswählen: Zum Beispiel müssen Patienten ausgewählt werden, für welche die Wirksamkeit der aktiven Kontrolle nachgewiesen wurde;
- Kontrolle der Noncompliance: Wenn niemand die Studienmedikamente einnimmt, werden die Vergleichstherapien identisch sein;
- nicht aussagekräftige Verlaufsparameter: Outcomes sollen einen Nutzen für den Patienten widerspiegeln und schon früher als Placebo überlegen dokumentiert sein;
- fehlende Daten;
- zu kurze Beobachtungszeit: Das Follow-up muss lang genug sein, damit sich irgendwelche potenziellen Differenzen zeigen können.

Am besten eine Kombination aus ITT- und OT-Analyse

Die meisten Nichtunterlegenheitsstudien benützen eine OT-Analyse (auch als Per-Protocol-[PP-]Analyse bezeichnet), welche die Compliance berücksichtigt. Am besten ist es, wenn eine Kombination von ITT- und OT-Analyse berichtet wird. Zu den Fallstricken bei Nichtunterlegenheitsstudien gehört es, wenn die Annahme der Effektstärken zwischen aktiver Kontrolle und Placebo sich geändert hat oder wenn sich Behandlungsstandards oder Ereignisraten mit der Zeit verändert haben. «Und wenn man die Nichtunterlegenheitsgrenze zu weit setzt, läuft man Gefahr, eine neue Therapie zu akzeptieren, die nicht viel besser ist als Placebo», mahnte Solomon.

Halid Bas

Quelle: Symposium «Trials are easy! A guide for the clinical cardiologist», 51. Jahreskongress der European Society of Cardiology (ESC), 1. September 2015 in London.