



AUS DEM NETZWERK EVIDENZBASIERTE MEDIZIN

Überlegenheit, Nicht-Unterlegenheit und Äquivalenz

Eine Einführung in die verschiedenen Arten
von Hypothesen klinischer Studien

**VON MAXI SCHULZ IM AUFTRAG DES NETZWERKS EVIDENZBASIERTE MEDIZIN E. V.
(WWW.EBM-NETZWERK.DE)**

R

Randomisierte kontrollierte Studien (in Englisch: Randomized controlled trials, kurz RCT) gelten als die beste Methode, um die Wirksamkeit von neuen Therapien zu überprüfen. Dabei werden die Studienteilnehmenden zufällig in zwei Gruppen aufgeteilt und verglichen: eine Gruppe erhält die neue Therapie, die andere Gruppe dient als Kontrollgruppe.

Als Kontrollgruppe kommen häufig Placebo-Gruppen zur Anwendung. Placebo-kontrollierte Studien sind jedoch nur dann ethisch vertretbar, wenn es keine bewährte Standardbehandlung gibt. In solchen Fällen werden Studien durchgeführt, bei denen die neue Therapie mit einer Standardbehandlung anstelle einer Placebo-Behandlung verglichen wird (aktiv kontrollierte Studien).

Ein wichtiger Unterschied zwischen Placebo-kontrollierten Studien und aktiv kontrollierten Studien liegt in der Fragestellung, die sie beantworten möchten. Bei Placebo-kontrollierten Studien ist es wesentlich, ob die neue Therapie einen zusätzlichen Vorteil gegenüber einer Placebo-Behandlung liefert (Überlegenheit). Im Gegensatz dazu zielen aktiv kontrollierte Studien zumeist darauf ab, zu bestimmen, ob die neue Therapie gleich gut ist wie die Kontrolltherapie (Äquivalenz) oder ob sie nicht schlechter ist als die Kontrolltherapie (Nichtunterlegenheit).

Diese Fragestellungen können in Hypothesen übersetzt werden, die im Rahmen der klinischen Studie untersucht werden. Diese Hypothesen bestimmen maßgeblich die Testentscheidung, die Interpretation der Ergebnisse und die Planung der Studie. Um Studienergebnisse bewerten zu können, ist es daher zentral, die Unterschiede zwischen den Hypothesen und Studienarten zu kennen.

In diesem Artikel werden die verschiedenen Arten von Fragestellungen und Hypothesen klinischer Studien erläutert. Dazu werden Beispiele von Therapien zur Behandlung von Typ-2-Diabetes herangezogen. Studien, die die Wirksamkeit neuer Therapien untersuchen, beschäftigen sich in diesem Zusammenhang häufig mit der Veränderung von HbA1c-Werten. Eine Reduktion dieser Werte spricht für die Wirksamkeit einer neuen Therapie, da sie auf eine verbesserte Blutzuckerkontrolle hinweist.

PRINZIP DES STATISTISCHEN TESTENS

Am Anfang jeder klinischen Studie ist es wichtig, die Annahmen an die Studiendaten in Form von Hypothesen zu formulieren. Dazu wird ein Hypothesenpaar aufgestellt, bestehend aus der Nullhypothese (H_0) und der komplementären Alternativhypothese (H_1). Die Nullhypothese stellt die Annahme dar, die es zu widerlegen gilt, während die Alternativhypothese die Annahme darstellt, die es zu beweisen gilt. Grundsätzlich gilt, dass die Alternativhypothese nicht direkt bewiesen werden kann. Stattdessen bedienen sich Statistiker:innen dem Falsifikationsprinzip: Durch die Ablehnung der Nullhypothese kann man von der (vorläufigen) Gültigkeit der Alternativhypothese ausgehen.

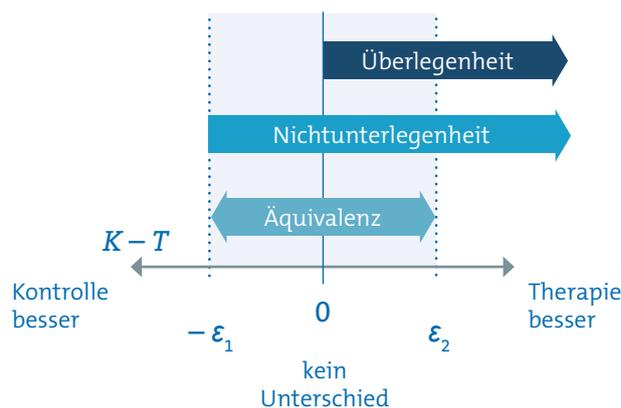
Das Ziel des statistischen Testens ist es, zu quantifizieren, inwieweit die vorliegenden Daten der Studie für oder gegen die Nullhypothese sprechen. Wenn die Daten stark von der Nullhypothese abweichen, wird die Nullhypothese abgelehnt, und es wird davon ausgegangen, dass die Alternativhypothese stimmt. Andernfalls wird die Nullhypothese beibehalten, die Studiendaten sprechen dann nicht genügend für die Alternativhypothese.

ARTEN VON HYPOTHESEN IN KLINISCHEN STUDIEN

Um die verschiedenen Hypothesen zu erklären, eignet sich das Beispiel einer zweiarmigen Studie, bestehend aus einer Kontrollgruppe (K) und einer Therapiegruppe (T). Der Unterschied zwischen dem Endpunkt der Kontrollgruppe und der Therapiegruppe wird als $(K - T)$ bezeichnet.

Eine positive Differenz ($K - T > 0$) bedeutet, dass die Therapiegruppe im Mittel niedrigere Werte hat als die Kontrollgruppe, eine negative Differenz ($K - T < 0$) be-

**ABBILDUNG 1:
SCHEMATISCHE DARSTELLUNG DES
UNTERSCHIEDS ZWISCHEN ÜBERLEGENHEIT,
NICHTUNTERLEGENHEIT UND ÄQUIVALENZ**



Auf der x-Achse variiert der Unterschied zwischen Kontroll- und Therapiegruppe ($K - T$). Für blutzuckersenkende Behandlungen gilt, wenn $K - T$ größer als 0 ist, dann ist die Therapie wirksamer als die Kontrolle und vice versa. Die Pfeile zeigen an, in welchen Bereichen von Überlegenheit, Nichtunterlegenheit und Äquivalenz gesprochen wird. Überlegenheit wird angenommen, wenn der Unterschied zwischen Kontrolle und Therapie mindestens größer ist als 0 (erster Pfeil von oben). Nichtunterlegenheit wird angenommen, wenn der Unterschied mindestens größer ist als die Nichtunterlegenheitsgrenze $-\epsilon_1$ (zweiter Pfeil von oben). Von Äquivalenz spricht man, wenn der Unterschied innerhalb des Äquivalenzbereichs zwischen $-\epsilon_1$ und ϵ_2 liegt (dritter Pfeil von oben).

deutet, dass die Therapiegruppe im Mittel höhere Werte hat als die Kontrollgruppe. Je nachdem, ob eine Studie auf den Nachweis von Überlegenheit, Äquivalenz oder Nichtunterlegenheit abzielt, unterscheiden sich die Annahmen über die Differenz ($K - T$). Mithilfe von Hypothesentests können diese Annahmen überprüft werden.

Dabei ist wichtig zu beachten, dass die Interpretation der Differenz von dem jeweiligen Anwendungskontext abhängt. Zum Beispiel sprechen niedrigere HbA1c-Werte nach Gabe eines neuartigen oralen Antidiabetikums für dessen Wirksamkeit, demnach würde eine positive Differenz ($K - T > 0$) für die neue Therapie sprechen. In anderen Fällen kann eine negative Differenz erwünscht sein. *Abbildung 1* zeigt, in welchen Be-



KASTEN 1: BEISPIEL EINER ÜBERLEGENHEITSSTUDIE

Semaglutid zur Kontrolle des Blutzuckerspiegels bei Typ-2-Diabetes-Patient:innen ⁽¹⁾

Studiendesign: Rodbard et al. untersuchten in der randomisiert-kontrollierten Studie SUSTAIN 5 von 2018 den Effekt von Semaglutid, einem Antidiabetikum, auf den Blutzuckerspiegel bei Patient:innen mit Typ-2-Diabetes unter Insulintherapie (1). Die Autor:innen verglichen die Gabe von Semaglutid einmal wöchentlich in zwei Dosen (0,5 mg bzw. 1,0 mg) mit der Gabe eines Placebos.

Hypothese: Ihre Alternativhypothese lautete, dass die Gabe von Semaglutid eine bessere Wirksamkeit als Placebo bei der Kontrolle des Blutzuckerspiegels für Patient:innen unter Insulintherapie hat. In anderen Worten: Die Autor:innen vermuteten, dass Semaglutid dazu führt, dass der Blutzuckerspiegel in den Therapiegruppen deutlicher gesenkt wird als in der Placebo-Gruppe. Dementsprechend lautete

die Nullhypothese, dass sich die Semaglutid-Gruppe nicht unterscheidet von der Placebo-Gruppe hinsichtlich des Blutzuckers.

Ergebnisse: Dafür maßen die Autor:innen die HbA1c-Werte der Studienteilnehmenden. Die Studienergebnisse zeigen, dass die HbA1c-Werte in der Therapiegruppe um 1,4% (Dosis 1) und 1,8% (Dosis 2) reduziert wurden im Vergleich zu einer Reduktion von 0,1% in der Placebo-Gruppe.

Schlussfolgerung: Die Nullhypothese, dass kein Unterschied zwischen Semaglutid und Placebo bei Patient:innen unter Insulintherapie existiert, konnte widerlegt werden. Man kann von einer besseren Wirksamkeit von Semaglutid ausgehen.

reichen man von Überlegenheit, Nichtunterlegenheit und Äquivalenz spricht, wenn eine positive Differenz ($K - T > 0$) für eine bessere Wirksamkeit der Therapie spricht. Im folgenden Text wird angenommen, dass eine positive Differenz einer höheren Wirksamkeit der Therapie entspricht.

ÜBERLEGENHEIT

Das Ziel von Placebo-kontrollierten Studien ist es, die Überlegenheit der neuen Therapie gegenüber dem Placebo zu demonstrieren.

Für die Behandlung von Typ-2-Diabetes ist zum Beispiel dann eine Therapie überlegen, wenn sie den Blutzuckerspiegel deutlicher senkt bzw. normalisiert als eine Placebo-Behandlung (siehe Beispielstudie in Kasten 1, vgl. Abbildung 1). Überlegenheit bedeutet also eine bessere Wirksamkeit der neuen Therapie gegenüber einem Placebo.

Die Hypothesenpaare definieren sich wie folgt:

H_0^U : Es gibt keinen Unterschied zwischen der Therapie- und der Kontrollgruppe (zumeist Placebo).

Die Nullhypothese besagt, dass sich die neue Therapie im Mittel nicht unterscheidet von der Kontrolltherapie. Mit anderen Worten: Die neue Therapie hat keine zusätzliche Wirkung gegenüber der Kontrolltherapie. Eine Überlegenheitsstudie möchte diese Nullhypothese widerlegen, um zu zeigen, dass die neue Therapie tatsächlich wirksamer ist.

Die Alternativhypothese beschreibt daher das Gegenteil der Nullhypothese:

H_1^U : Es gibt einen Unterschied zwischen der Therapie- und der Kontrollgruppe.

Sie besagt, dass die neue Therapie und die Kontrolltherapie unterschiedliche Ergebnisse haben. Wenn dies der Fall ist, kann man davon ausgehen, dass die neue Therapie wirksamer ist als die Kontrolltherapie.

Die Hypothesenpaare lassen sich auch etwas mathematischer notieren als

$$H_0^U: K - T = 0$$

$$H_1^U: K - T \neq 0$$

In der Beispielstudie SUSTAIN 5 (Kasten 1) misst der mittlere Unterschied zwischen der Kontroll- und der Therapiegruppe (Dosis 1) gleich 1,3% (= (-0,1) - (-1,4)). Dieser Unterschied ist größer als 0 und somit weit weg von der Nullhypothese, dass es keinen Unterschied zwischen Therapie und Kontrolle gibt ($H_0^U: K - T = 0$). Dabei zeigt die Therapiegruppe eine wesentlich deutlichere Reduktion der HbA1c-Werte.

An dieser Stelle könnte man sich fragen, warum nicht direkt auf eine positive Differenz zwischen Therapie und Kontrolle, also $K - T > 0$, getestet wird (anstatt auf einen Unterschied ($K - T \neq 0$)), wenn doch das primäre Ziel der Studie ist, zu zeigen, dass die HbA1c-Werte deutlicher gesenkt werden (vgl. Abbildung 1).

Die Antwort ist: Ein sogenannter einseitiger Test ist nicht in der Lage, einen unerwarteten negativen Effekt der neuen Therapie zu erkennen, sprich: eine Erhöhung der HbA1c-Werte durch Semaglutid. Um einen unerwarteten negativen Effekt der neuen Therapie auszuschließen, testen klinische Studien den Unterschied in beide Richtungen.

Dies bedeutet, dass man nicht nur testet, ob die neue Therapie besser ist als die Kontrolltherapie, sondern auch, ob sie schlechter ist. Dieses Verfahren wird als zweiseitiger Test bezeichnet (2).

ÄQUIVALENZ

Eine weitere Art von Fragestellung beschreibt die Äquivalenzhypothese, die bei aktiv kontrollierten Studien von Interesse sein kann. Hierbei ist das Ziel, zu zeigen, dass die neue Therapie im Wesentlichen gleichwertig ist wie die aktive Kontrolle, also weder besser noch schlechter. Äquivalenzstudien kommen z.B. dann zum Einsatz, wenn der Patentschutz eines bereits zugelassenen biologischen Arzneimittels ausläuft und Biosimilars, sogenannte Nachahmerpräparate, auf den Markt kommen. Um zugelassen zu werden, muss gezeigt werden, dass keine klinisch relevanten Unterschiede

zwischen dem Referenz-Arzneimittel und dem Biosimilar bestehen. Ein Beispiel hierfür sind Studien, die die Gleichwertigkeit von Biosimilar-Insulinen mit Original-Insulinen für die Insulintherapie bei Typ-2-Diabetes untersuchen (siehe Beispielstudie in Kasten 2 – nächste Seite).

Das Hypothesenpaar lässt sich wie folgt definieren:

H_0^A : Die neue Therapie ist nicht äquivalent zur Kontrolle.
 H_1^A : Die neue Therapie ist äquivalent zur Kontrolle.

Da es praktisch unmöglich ist, die exakte Gleichwertigkeit zweier Therapien zu bestimmen, beginnt eine Äquivalenzstudie mit der Frage, innerhalb welcher Spanne die Gleichwertigkeit zweier Therapien angenommen werden kann. Dieser Bereich wird auch als Äquivalenzbereich bezeichnet (vgl. Abbildung 1). Ergebnisse, die außerhalb dieses Bereichs liegen, werden nicht mehr als Gleichwertigkeit betrachtet.

Der Äquivalenzbereich kann durch die Definition einer unteren Grenze $-\epsilon_1$ und einer oberen Grenze ϵ_2 festgelegt werden. Wenn die Ergebnisse einer Studie zeigen, dass der Unterschied zwischen der Kontrollgruppe und der Therapiegruppe innerhalb dieses Bereichs $[-\epsilon_1, \epsilon_2]$ liegt, kann die Gleichwertigkeit der beiden Therapien geschlussfolgert werden (vgl. Abbildung 1).

Übertragen auf das Falsifikationsprinzip bedeutet das, dass unter der Nullhypothese geprüft wird, ob der Unterschied zwischen der Kontrollgruppe und der Therapiegruppe $K - T$ außerhalb dieses Bereichs liegt. Genauer gesagt wird geprüft, ob der Unterschied entweder niedriger ist als die untere Grenze $-\epsilon_1$ oder höher ist als die obere Grenze ϵ_2 .

Die Nullhypothese kann dann wie folgt notiert werden:

$$H_0^A: K - T \leq -\epsilon_1 \quad \text{oder} \quad K - T \geq \epsilon_2$$



KASTEN 2: BEISPIEL EINER ÄQUIVALENZSTUDIE

Vergleich von Biosimilar-Insulin mit Referenzprodukt zur Insulintherapie bei Typ-2-Diabetes ⁽³⁾

Studiendesign: Christofides et al. untersuchten in ihrer randomisierten Studie von 2024 ein Biosimilar-Insulin zur Anwendung bei Insulintherapie von Typ-2 Diabetes (3). In der Untersuchung verglichen sie das Biosimilar-Insulin mit einem bereits zugelassenen Insulin hinsichtlich Immunogenität. Anders als bei Überlegenheitsstudien kam hier keine Placebo-Gruppe als Kontrolle zur Anwendung, sondern eine sogenannte aktiv kontrollierte Gruppe, die das Referenz-Insulin erhielt. Die Therapiegruppe erhielt das Biosimilar-Insulin. Als primären Endpunkt betrachteten die Autor:innen den Anteil der Studienteilnehmenden, die behandlungsinduzierte Anti-Insulin-Antikörper über einen Behandlungszeitraum von 26 Wochen entwickelten.

Hypothese: Die Nullhypothese lautete, dass das Biosimilar-Insulin nicht äquivalent zum Referenzprodukt ist. Dies würde bedeuten, dass das Biosimilar-Insulin deutlich weniger oder deutlich mehr Anti-Insulin-Antikörper verursachen würde als das Referenz-Insulin. Die Alternativhypothese, die es zu beweisen galt, lautete hingegen, dass die beiden Produkte äquivalent sind (also weder deutlich besser oder noch deutlich schlechter). Die Autor:innen hatten hierfür vorab einen Äquivalenzbereich von -10,7% bis 10,7% definiert.

Eine Differenz zwischen beiden Produkten von bis zu 10,7% konnte demnach als äquivalent betrachtet werden. Kommt es zu Abweichungen innerhalb dieses Bereichs, kann die Nullhypothese abgelehnt werden. Sind die Abweichungen zwischen beiden Produkten größer als dieser Bereich, kann nicht von Äquivalenz ausgegangen werden und die Nullhypothese nicht abgelehnt werden.

Ergebnisse: Die Studie zeigt, dass bei 19,2% der Studienteilnehmenden die Entwicklung von behandlungsinduzierten Anti-Insulin-Antikörpern auftrat, wenn sie das Biosimilar-Insulin erhielten. Dies ist vergleichbar mit der Kontrollgruppe, bei der 21,3% Anti-Insulin-Antikörper entwickelten. Die Differenz zwischen den beiden Produkten betrug 2,1 Prozentpunkte (90% Konfidenzintervall der Behandlungsdifferenz: -7.6%; 3.5%).

Schlussfolgerung: Die Ergebnisse der Studie zeigen, dass die Differenz von 2,1 inkl. Konfidenzintervall zwischen den beiden Produkten innerhalb des Äquivalenzbereichs von -10,7% bis 10,7% liegt. Die Nullhypothese kann abgelehnt und es kann von Äquivalenz ausgegangen werden.

Unter der Alternativhypothese der Gleichwertigkeit befindet sich der Unterschied zwischen den Gruppen innerhalb des Bereichs $[-\epsilon_1, \epsilon_2]$:

$$H_1^A: -\epsilon_1 < K - T < \epsilon_2$$

NICHTUNTERLEGENHEIT

Die Fragestellung der Nichtunterlegenheit ist eng verwandt mit dem Äquivalenzproblem und beginnt ähnlich: Eine neue Therapie wird gegen eine etablierte Standardtherapie getestet. Im Unterschied zur Äquivalenzstudie zielt die Nichtunterlegenheitsstudie nicht

darauf ab, zu zeigen, dass die neue Therapie weder schlechter noch besser ist, sondern darauf, dass sie mindestens genauso gut ist wie die Standardtherapie.

Dies ist zum Beispiel von Interesse, wenn neue Therapien gegenüber Standardtherapien Vorteile wie geringere Nebenwirkungen, leichtere Verabreichung oder geringere Kosten in der Herstellung mit sich bringen. Ein Beispiel dafür ist ein neues orales Antidiabetikum für Typ-2-Diabetes-Patient:innen, das weniger Nebenwirkungen hat als die etablierte Tablette (*weiteres Beispiel im Kasten 3*). In diesem Fall genügt es, zu zeigen, dass das neue orale Antidiabetikum zur etablierten

KASTEN 3: BEISPIEL EINER NICHTUNTERLEGENHEITSSTUDIE**Orale Therapie gegen bewährte Standardtherapie zur Senkung des Blutzuckers bei Typ-2-Diabetes ⁽⁴⁾**

Studiendesign: Die Phase-3a-Studie PIONEER 4 von Pratley et al. untersuchte die Wirksamkeit von oral verabreichtem Semaglutid zur Senkung des Blutzuckerspiegels bei Patient:innen mit Typ-2-Diabetes (4). Diese Behandlung bietet eine Alternative zur (etablierten) subkutanen Injektion. Die Therapiegruppe, die das orale Semaglutid erhielt, wurde mit einer Kontrollgruppe, die die Standardtherapie Liraglutid per subkutaner Injektion erhielt, verglichen. Zum Vergleich wurden die HbA1C-Werte der Studienteilnehmenden gemessen.

Hypothese: Das Ziel der Studie war es, zu zeigen, dass die oral verabreichte Therapie mindestens genauso gut ist hinsichtlich der Reduzierung des Blutzuckers wie die injizierte Standardtherapie. Die Autor:innen hatten vorab eine Nichtunterlegenheitsgrenze von -0,4% definiert. Dies bedeutet, dass die Wirksamkeit von oralem Semaglutid

im Unterschied zum subkutanem Liraglutid bis zu -0,4% betragen durfte, um als nichtunterlegen betrachtet werden zu können. Die Nullhypothese lautete, dass das oral verabreichte Semaglutid nicht nichtunterlegen ist, während die Alternativhypothese lautete, dass oral Semaglutide mindestens genauso gut ist wie die Standardtherapie.

Ergebnisse: Die Studienergebnisse zeigen, dass sich die HbA1c-Werte im Mittel um 0,1 Prozentpunkte zwischen oralem Semaglutid und subkutanem Liraglutid unterschieden (95% Konfidenzintervall der Behandlungsdifferenz: -0,3 bis 0,0 Prozentpunkte).

Schlussfolgerung: Die Ergebnisse der Studie zeigen, dass die Differenz zwischen beiden betrachteten Therapien oberhalb der Grenze -0,4% liegt. Die Nullhypothese konnte abgelehnt, und es kann von Nichtunterlegenheit ausgegangen werden.

Standardtherapie höchstens eine klinisch vertretbare Effektivitätslücke aufweist und damit nicht wesentlich schlechter ist als die Standardbehandlung. Es ist also irrelevant, ob die neue Therapie möglicherweise sogar besser als die etablierte Therapie ist.

Das Hypothesenpaar lässt sich wie folgt definieren:

H_0^N : Die neue Therapie ist nicht nichtunterlegen gegenüber der Kontrolle.

H_1^N : Die neue Therapie ist nichtunterlegen gegenüber der Kontrolle.

Auch in Nichtunterlegenheitsstudien basiert die Überlegung auf einem Mindestniveau an Effektivität, das die neue Therapie gegenüber der Standardtherapie

erfüllen muss. Dieses Mindestniveau entspricht der unteren Grenzen des Äquivalenzbereichs $-\epsilon_1$. Im Gegensatz zur Äquivalenz wird jedoch keine obere Grenze definiert. Unter der Nullhypothese wird also getestet, ob der Therapieunterschied kleiner ist als der mindestvertretbare Unterschied.

$$H_0^N: K - T \leq -\epsilon_1$$

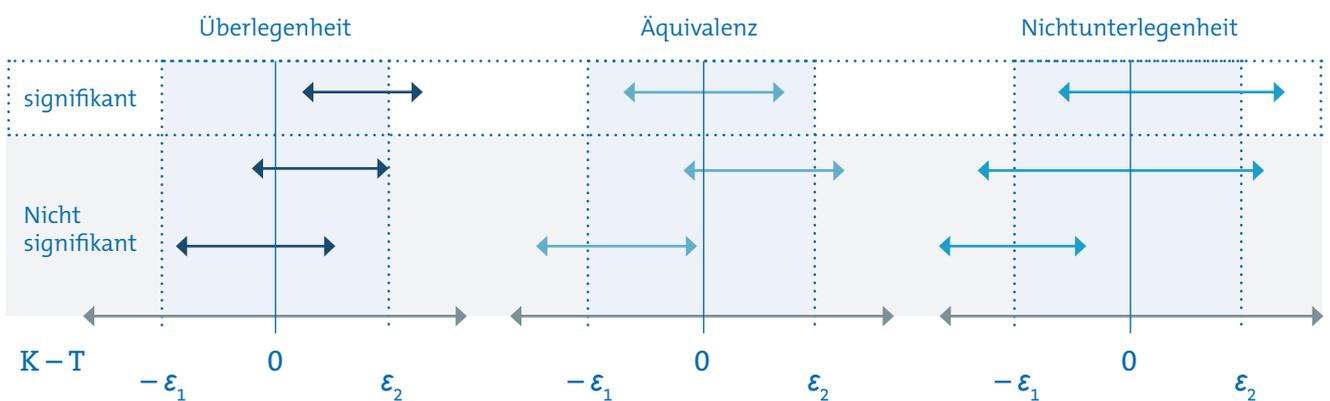
Für Nichtunterlegenheit muss der Unterschied zwischen Kontroll- und Therapiegruppe demnach mindestens größer sein als die untere Grenze $-\epsilon_1$ (vgl. Abbildung 1).

$$H_1^N: K - T > -\epsilon_1$$

Dabei stellt Nichtunterlegenheit lediglich eine Modifikation des Äquivalenzproblems dar. Denn: Die Äquivalenzhypothesen können zu Nichtunterlegenheitshypo-



ABBILDUNG 2:
TESTENTSCHEIDUNG MIT KONFIDENZINTERVALLEN, WENN EINE POSITIVE DIFFERENZ
ZWISCHEN KONTROLL- UND THERAPIEGRUPPE EINE HÖHERE WIRKSAMKEIT DER NEUEN
THERAPIE BEDEUTET



Die Pfeile bezeichnen die Länge des Konfidenzintervalls.

thesen umgewandelt werden, wenn ϵ_2 unendlich groß gewählt wird. Daher wird Nichtunterlegenheit auch häufig als einseitige Äquivalenz bezeichnet.

BESONDERHEITEN VON ÄQUIVALENZ UND NICHTUNTERLEGENHEIT

In Studien, die die Überlegenheit einer neuen Therapie gegenüber einem Placebo ermitteln, können herkömmliche zweiseitige statistische Tests zur Anwendung kommen. Eine Testentscheidung kann dann mittels p-Wert und Konfidenzintervall getroffen werden.

Eine Testentscheidung in Äquivalenz- und Nichtunterlegenheitsstudien ist wiederum etwas kniffliger. Falsch wäre es nämlich anzunehmen, dass Äquivalenz gegeben ist, wenn die Nullhypothese der zweiseiti-

gen Überlegenheitshypothese nicht abgelehnt werden kann (5). Zur Erinnerung: Die Nullhypothese des Überlegenheitsproblems lautet: $H_0^U: K - T = 0$. Wenn ein Überlegenheitsproblem also nicht signifikant ist (die Nullhypothese nicht abgelehnt wird), heißt es nicht, dass Gleichwertigkeit zwischen Therapie und Kontrolle angenommen werden kann.

Für die Testentscheidung bei Äquivalenz- und Nichtunterlegenheitsstudien sollte auf das Konfidenzintervall zurückgegriffen werden. Zwar gibt es auch geeignete Tests, diese sind aber deutlich komplizierter als herkömmliche. Das Testverfahren anhand von Konfidenzintervallen ist in *Abbildung 2* veranschaulicht.

Bei Überlegenheitsstudien ist eine Therapie überlegen, wenn das Konfidenzintervall der Differenz

zwischen Kontroll- und Therapiegruppe größer als 0 ist (Zeile 1 in *Abbildung 2*), also nicht die 0 umschließt. Umschließt das Konfidenzintervall die 0, kann die Nullhypothese nicht verworfen werden (Zeile 2 und 3 in *Abbildung 2*).

Für Äquivalenzstudien gilt: Wenn das Konfidenzintervall vollständig innerhalb des Bereichs $[-\epsilon_1, \epsilon_2]$ liegt (wie in Zeile 1 von *Abbildung 2*), dann kann die Nullhypothese verworfen und Äquivalenz angenommen werden. Liegt das Konfidenzintervall zum Teil oder ganz außerhalb des Äquivalenzbereichs (Zeile 2 und 3 in *Abbildung 2*), kann die Nullhypothese nicht abgelehnt und Äquivalenz nicht gefolgert werden.

Bei Nichtunterlegenheitsstudien kann die Nullhypothese verworfen werden, wenn das Konfidenzintervall vollständig über dem Wert $-\epsilon_1$ liegt (Zeile 1 in *Abbildung 2*). Dann spricht man von Nichtunterlegenheit. Umschließt das Konfidenzintervall $-\epsilon_1$ oder liegt darunter, dann kann die Nullhypothese nicht verworfen werden (Zeile 2 und 3 in *Abbildung 2*).

Fatal wäre es auch, die gleichen Verfahren für alle drei Arten von Studien zu verwenden, um die erforderliche Anzahl an Studienteilnehmenden zu ermitteln. Da sich die statistischen Tests unterscheiden, sind auch die Planungsverfahren unterschiedlich. Für Nichtunterlegenheits- und Äquivalenzstudien sind in der Regel höhere Stichprobenumfänge erforderlich als für Überlegenheitsstudien.

Der Bereich, in dem Äquivalenz oder Nichtunterlegenheit angenommen werden kann, ist zudem entscheidend für die richtige Interpretation von Studienergebnissen. Um Verzerrungen zu verhindern,

muss dieser Bereich unbedingt vor der Datenanalyse spezifiziert werden. Denn es können nachträglich immer Bereichsgrenzen gewählt werden, die Äquivalenz und Nichtunterlegenheit beweisen (6). Außerdem sollte die Wahl dieser Grenzen ausreichend klinisch begründet sein, damit die Studienergebnisse aussagekräftig sind.

FAZIT

Die Wahl der Vergleichsgruppe hängt eng mit der Fragestellung zusammen, die eine klinische Studie beantworten möchte. Während Placebo-kontrollierte Studien Fragen der Überlegenheit adressieren, sind Äquivalenz und Nichtunterlegenheit häufiger bei aktiv kontrollierten Studien relevant. Äquivalenz und Nichtunterlegenheit sind jedoch häufig komplexer und schwieriger zu verstehen als Überlegenheit.

Um Studienergebnisse richtig einordnen zu können, ist es wichtig, die Unterschiede in den Hypothesen, in der statistischen Analyse und der Interpretation der Ergebnisse zwischen den verschiedenen Studienarten zu kennen. Insbesondere bei Äquivalenz- und Nichtunterlegenheitsstudien ist es wichtig zu wissen, dass Testentscheidungen vorrangig mithilfe von Konfidenzintervallen getroffen werden bzw. geeignete statistische Tests zur Anwendung kommen müssen. Darüber hinaus hängt die erforderliche Anzahl an Studienteilnehmenden davon ab, welche Hypothese untersucht wird, und eine Studie ist lediglich dann aussagekräftig, wenn die Grenzen der Nichtunterlegenheit und Äquivalenz vor der Studiendurchführung spezifiziert und ausreichend begründet sind.

→ Fortsetzung nächste Seite



→ Fortsetzung

Erklärung über die Nutzung von KI

Während der Vorbereitung dieses Artikels kamen eine lokal gehostete Version von Llama 3.3 70B Instruct und Llama 3.1 SauerkrautLM 70B Instruct zum Korrekturlesen und zur Textoptimierung zur Anwendung. Nach der Nutzung dieses Tools überprüfte die Autorin den Inhalt und redigierte ihn nach Bedarf. Sie übernimmt die volle Verantwortung für den Inhalt des veröffentlichten Artikels.



MAXI SCHULZ

Institut für Medizinische Statistik,
Universitätsmedizin Göttingen,
Humboldtallee 32, 37073 Göttingen
maxi.schulz@med.uni-goettingen.de

Literatur

1. Rodbard HW, Lingvay I, Reed J, La Rosa R de, Rose L, Sugimoto D et al. Semaglutide Added to Basal Insulin in Type 2 Diabetes (SUSTAIN 5): A Randomized, Controlled Trial. *J Clin Endocrinol Metab* 2018; 103(6):2291–301.
2. Dunnett CW, Gent M. An alternative to the use of two-sided tests in clinical trials. *Stat Med* 1996; 15(16):1729–38.
3. Christofides EA, Puente O, Norwood P, Denham D, Maheshwari H, Lillestol M et al. Immunogenicity, efficacy, and safety of biosimilar insulin glargine (Gan & Lee glargine) compared with originator insulin glargine (Lantus®) in patients with type 2 diabetes after 26 weeks' treatment: A randomized open label study. *Diabetes Obes Metab* 2024; 26(6):2412–21.
4. Pratley R, Amod A, Hoff ST, Kadowaki T, Lingvay I, Nauck M et al. Oral semaglutide versus subcutaneous liraglutide and placebo in type 2 diabetes (PIONEER 4): a randomised, double-blind, phase 3a trial. *Lancet* 2019; 394(10192):39–50.
5. Wellek S, Blettner M. Establishing equivalence or non-inferiority in clinical trials: part 20 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2012; 109(41):674–9.
6. Committee for Proprietary Medicinal Products. Points to consider on switching between superiority and non-inferiority. *Br J Clin Pharmacol* 2001; 52(3):223–8.

Weitere Quellen

- Kishore K, Mahajan R. Understanding Superiority, Noninferiority, and Equivalence for Clinical Trials. *Indian Dermatol Online J* 2020; 11(6):890–4.
- Hilgers R-D, Bauer P, Scheiber V. Einführung in die medizinische Statistik. Berlin, Heidelberg, New York, Hongkong, London, Mailand, Paris, Tokio: Springer; 2003. (Statistik und ihre Anwendungen).
 - Wiens BL. Choosing an equivalence limit for noninferiority or equivalence studies. *Control Clin Trials* 2002; 23(1):2–14.
 - Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJW. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006; 295(10):1152–60.
 - Wellek S. Testing statistical hypotheses of equivalence. Boca Raton, Fla.: Chapman & Hall/CRC; 2003. Available from: URL: <http://www.loc.gov/catdir/enhancements/fy0646/2002031307-d.html>.
- D'Agostino RB, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Stat Med* 2003; 22(2):169–86.

