

Statistik in klinischen Studien: Schlüsselkonzepte

Einleitung

Statistische Methoden ermöglichen eine formale Beurteilung der Quellen für Schwankungen in der Art und Weise, in der Patienten auf eine Behandlung ansprechen. Die Anwendung statistischer Methoden ermöglicht es dem klinischen Forscher, aus erfassten Daten vernünftige und präzise Schlussfolgerungen zu ziehen und trotz Unsicherheit zu fundierten Entscheidungen zu gelangen. Statistische Methoden sind für die Vermeidung von Fehlern und Verzerrungen in der medizinischen Forschung unverzichtbar. Dieser Artikel behandelt bestimmte Grundkonzepte statistischer Methoden und ihrer Anwendung im Rahmen klinischer Studien.

Hypothesentestung

Eine Hypothese ist eine Annahme oder eine Gruppe von Annahmen, die entweder a) auf vorläufiger Basis eine Behauptung aufstellen, um die weitere wissenschaftliche Untersuchung zu leiten, oder b) einen bestimmten Sachverhalt angesichts anerkannter Fakten als mit größter Wahrscheinlichkeit zutreffend bestätigen.

Im Rahmen dieser Erörterung interessieren uns Hypothesen, die eine Behauptung aufstellen – beispielsweise, dass eine neue Behandlung für eine Erkrankung besser als die bestehende Standardbehandlung ist. Nennen wir die neue Behandlung „B“ und die Standardbehandlung „A“, so besagt die Hypothese, dass „B“ besser ist als „A“.

Man könnte nun annehmen, dass Wissenschaftler sich daran machen, diese Hypothese zu beweisen, jedoch ist dies

mitnichten der Fall. Stattdessen wird dieses Ziel indirekt angegangen. Anstatt zu versuchen, die Hypothese B zu beweisen, gehen wissenschaftliche Methoden von der Annahme aus, dass in Wirklichkeit A wahr ist, dass also kein Unterschied zwischen der Standardbehandlung und der neuen Behandlung besteht. Dies wird als „Nullhypothese“ bezeichnet. Die Wissenschaftler versuchen dann, A zu widerlegen. Dies bezeichnet man als Falsifikation der Nullhypothese. Gelingt ihnen dies – also zu beweisen, dass die Hypothese A falsch und die Standardbehandlung nicht besser als die neue Behandlung ist –, folgt daraus, dass B wahr und die neue Behandlung besser als die Standardbehandlung ist.

Warum wird auf diese Weise vorgegangen?

Auf diese Frage gibt es keine einfache Antwort; es handelt sich um die allgemein akzeptierte Vorgehensweise, die sich in der modernen Wissenschaft entwickelt hat. Vielleicht hilft eine Analogie aus dem Rechtswesen. Die Nullhypothese gibt unsere aktuelle Situation oder unser aktuelles Wissen wieder (vor Gericht also: „Der Angeklagte ist unschuldig“), dem wir vertrauen müssen, sofern wir keine ausreichenden, diesem Wissen widersprechenden Beweise oder Indizien besitzen. Wenn wir uns jedoch bemüht haben, die „Alternativhypothese“ (die Bezeichnung für das Gegenteil der „Nullhypothese“) zu beweisen, gilt tatsächlich: „Der Angeklagte ist schuldig“).

Ein anderer, möglicherweise einfacherer Weg, zu diesem Punkt zu gelangen, ist ein Zitat von Albert Einstein:

„Noch so viele Experimente können niemals beweisen, dass ich Recht habe; ein einziges Experiment kann beweisen, dass ich falsch liege.“

Das sagt uns, dass der Versuch, die Nullhypothese zu widerlegen, ein schlüssigeres und eher zu erreichendes Ziel ist, als der Versuch, die Gültigkeit der Alternativhypothese zu beweisen. Beachten Sie bitte, dass diese Ausführungen NICHT

exakt erklären, warum in der Wissenschaft dieser Ansatz verfolgt wird. Sie ermöglichen es uns jedoch möglicherweise, ein kniffliges Konzept leichter zu verstehen und zu akzeptieren.

Fehler 1. und 2. Art

Ein Blick auf die nachstehende Tabelle macht den Unterschied zwischen Fehlern 1. Art (falsch positiv) und Fehlern 2. Art (falsch negativ) deutlich.

	Die Nullhypothese ist wahr	Die Nullhypothese ist falsch
Die Nullhypothese zurückweisen	Fehler 1. Art „Falsch Positiv“	Korrektes Ergebnis „Richtig Positiv“
Die Nullhypothese kann nicht zurückgewiesen werden	Korrektes Ergebnis „Richtig Negativ“	Fehler 2. Art „Falsch Negativ“

Dies ist weiterhin sehr verwirrend. Ein – vergleichsweise krasses – Beispiel soll dies in einfacheren Worten ausdrücken:

- Fehler 1. Art könnten einen Patienten töten – Stellen Sie sich eine Studie vor, die fälschlicherweise zu dem Ergebnis gelangt, dass die Standardversorgung nicht besser ist als die neue Behandlung, wodurch Patienten die neue Behandlung erhalten – mit katastrophalen Folgen. Auftretende Fehler 1. Art stellen fälschlicherweise eine Wirkung fest, die es nicht gibt.
- Fehler 2. Art führen dazu, dass potenziell wertvolle Forschungsergebnisse verworfen werden. Diese Forschungsergebnisse wären möglicherweise wirklich nützlich gewesen, da jedoch keine weiteren diesbezüglichen Untersuchungen stattfinden, werden keine Patienten geschädigt. Auftretende Fehler 2. Art führen dazu, dass eine vorhandene Wirkung nicht festgestellt

wird.

Es ist klar, dass – soweit es um Patienten geht – Fehler 1. Art schwerer wiegen als Fehler 2. Art.

Signifikanzniveau

Das Signifikanzniveau bezeichnet die Wahrscheinlichkeit des Auftretens eines Fehlers 1. Art. Diese hängt u. a. vom Umfang der Stichprobe und der „Teststärke“ des statistischen Verfahrens ab.

Teststärke

Die „Stärke“ eines statistischen Verfahrens bezeichnet die Wahrscheinlichkeit, dass es zu einer korrekten Zurückweisung einer Nullhypothese führt – oder anders ausgedrückt: das Vermögen des Verfahrens, eine Wirkung zu erkennen, wenn eine solche tatsächlich vorliegt. Eine andere Beschreibung wäre, dass die Stärke eines statistischen Verfahrens die Wahrscheinlichkeit bezeichnet, dass es NICHT zu einem Fehler 2. Art kommt.

p-Werte

Der p-Wert (von Engl. „probability“, Wahrscheinlichkeit) wichtet die Stärke der Evidenz gegen die Nullhypothese auf einer Skala von 0 bis 1. Ein kleiner p-Wert (typischerweise unter 0,05 bzw. 5 %) besagt, dass überzeugende Erkenntnisse *gegen* die Nullhypothese sprechen, was Sie dazu bringen könnte, die Nullhypothese zu verwerfen, wohingegen ein großer p-Wert (größer als 0,05) das Gegenteil besagt.

Korrelation und Kausalität

Bei der Analyse der Ergebnisse einer Studie muss man sich unbedingt vergegenwärtigen, dass Korrelation und Kausalität

nicht dasselbe sind. Eine Korrelation besteht, wenn zwei Variablen auf die eine oder andere Weise miteinander verbunden sind; dies bedeutet jedoch nicht, dass die eine die andere herbeiführt (es besteht ein Zusammenhang zwischen den beiden Variablen). Hierzu ein Beispiel: Es wurde beobachtet, dass Frauen, die eine Hormon-Substitutions-Therapie (HST) erhielten, ein geringeres Risiko für die Koronare Herzkrankheit (KHK) aufwiesen. Ursächlich hierfür war jedoch nicht die eigentliche HST-Behandlung, sondern der Umstand, dass dieser Personenkreis tendenziell einen höheren sozioökonomischen Status mit einer gesünderen Ernährungsweise und ausgeprägterer sportlicher Betätigung aufwies.

Eine Kausalität kann beobachtet werden, wenn ein Faktor ein Ergebnis verursacht. Oftmals ist ein Kausalfaktor nur eine der Ursachen für ein Ergebnis. Um zwischen Korrelation und Kausalität unterscheiden zu können, müssen so viele Informationen wie möglich über die an einer klinischen Studie teilnehmenden Personen erfasst werden. Zudem müssen klinische Studien unbedingt unter sorgfältiger Berücksichtigung wissenschaftlicher Methodiken ausgelegt und mögliche Verzerrungen bewertet werden.

Datenmanipulation

Unter Datenmanipulation versteht man die selektive Falschdarstellung von Daten oder das Herbeiführen falscher Ergebnisse. Dies wäre beispielsweise dann der Fall, wenn Daten, die dem erwarteten Ergebnis widersprechen, absichtlich verworfen werden, um den Anteil der Ergebnisse zu erhöhen, die die formulierte Hypothese bestätigen. Wenn ein Forscher die Ausreißer (Werte, die sehr viel größer oder kleiner als der jeweils benachbarte Wert sind) aus den Ergebnissen eliminieren, müssen sie unbedingt verifizieren, dass es sich tatsächlich um Ausreißer handelt und nicht etwa nur um Ergebnisse, die von dem erwarteten oder erwünschten Ergebnis abweichen. Datenmanipulation liegt beispielsweise auch dann

vor, wenn ein Datenerfasser aus den von einem einzigen Patienten erfassten Messwerten eine ganze Datenserie generiert.

Datentransformation

Datentransformation bezeichnet die Anwendung einer mathematischen Formel auf bestimmte im Rahmen einer Studie erfasste Daten. Dies geschieht oftmals, um die Darstellung der Daten klarer oder verständlicher zu machen. Beispielsweise liegt es bei der Bestimmung des Kraftstoffverbrauchs eines Autos nahe, die Wirtschaftlichkeit in „Kilometer je Liter“ anzugeben. Lautet die Fragestellung jedoch, wieviel Kraftstoff erforderlich ist, um eine gewisse Strecke zu fahren, ist die Darstellung in „Liter je Kilometer“ günstiger. Würde in einem solchen Fall eine falsche Formel verwendet, um die neue Darstellung der Daten zu generieren, hätte dies Auswirkungen auf das Gesamtergebnis der Studie.

Zusammenführung von Daten

Unter dem Zusammenführen von Daten versteht man das Kombinieren von Daten aus mehreren Studien mit der Zielsetzung, ein besseres Verständnis der Situation zu erlangen. Eines der gebräuchlichsten Verfahren der Zusammenführung von Daten ist die Metaanalyse, in der Ergebnisse aus verschiedenen Studien zusammengelegt werden, um sie zu vergleichen und zu einem Gesamtergebnis zu gelangen. Bei der Durchführung einer Metaanalyse muss sorgfältig darauf geachtet werden, dass die Methoden der Studien identisch oder vergleichbar sind. Etwaige Unterschiede im Studiendesign müssen berücksichtigt werden, damit keine Unterschiede in den zugrundeliegenden Variablen (Störvariablen) bestehen. Ein Beispiel für eine unzulässige Zusammenführung von Daten wäre die Aggregation der Daten aus Studien an unterschiedlichen Mäusespezies zu einer Tiermodellstudie.

A2-4.33.3-V1.1