

Analyse der Ergebnisse klinischer Prüfungen

Einleitung

Wenn Pharmaunternehmen klinische Prüfungen durchführen, werden medizinische Informationen der teilnehmenden Patienten (nicht jedoch ihre Identität) zusammen mit den Ergebnissen jeglicher erfolgter Messungen in einer Computerdatenbank erfasst. Danach werden statistische Analysen durchgeführt, um die Ergebnisse der Studie formal zu beurteilen.

Analysen der Ergebnisse klinischer Prüfungen decken drei Interessenbereiche ab:

- Demographische und Baseline-Informationen
- Wirksamkeit
- Sicherheit

Diese Bereiche werden nachstehend näher beschrieben. Der Typ und das Design der klinischen Prüfung spielen eine wichtige Rolle für die Interpretation der statistischen Analysen.

Demographische und Baseline-Informationen

Wer nahm an der Studie teil? Die Wirkungen eines Arzneimittels können sich in verschiedenen Gruppen von Patienten deutlich unterscheiden. Es ist daher wichtig, die Informationen zu allen Patienten der Studie zu kennen, wie etwa:

- Alter
- Geschlecht
- Ethnische Herkunft
- Schwere der Erkrankung

Je größer die Übereinstimmung zwischen einer Studiengruppe und einer Population von Interesse ist, umso relevanter sind die Befunde im Allgemeinen.

Wirksamkeit

Wie gut wirkte das Prüfpräparat? Dieser Teil der Analyse basiert auf vordefinierten „Endpunkten“. Es gibt spezifische Maße im Zusammenhang mit der betreffenden Erkrankung. Endpunkte werden im Voraus im Prüfplan (das Dokument, in dem detailliert beschrieben ist, wie die Studie durchgeführt wird) festgelegt.

Endpunkte können im Allgemeinen wie folgt kategorisiert werden:

- „Harte“ Endpunkte – jene, die sich als numerische Fakten mit spezifischer klinischer Bedeutung darstellen. Zum Beispiel, wie lange der Patient überlebte oder welcher Anteil an Patienten sich von der Infektion erholte.
- „Weiche“ Endpunkte – jene, die potenziell durch den Messvorgang beeinflusst werden oder eine fragliche Reproduzierbarkeit aufweisen. Zum Beispiel, ein Fragebogen zur Lebensqualität oder die Beschreibung der Stimmung des Patienten zu einem gegebenen Zeitpunkt. Um statistisch analysiert zu werden, müssen weiche Endpunkte in ein numerisches Format umgewandelt werden. Dieser Prozess kann kontrovers sein, da er oftmals auf subjektiven Daten beruht und potenziell Uneinheitlichkeit mit sich ziehen kann.
- „Surrogatendpunkte“ – jene, die nicht selbst Teil der Krankheitserfahrung des Patienten sind, jedoch eng damit verbunden sein können. Zum Beispiel die Ergebnisse von Laboruntersuchungen.

Im Allgemeinen, sind harte Endpunkte weichen Endpunkten und Surrogatendpunkten vorzuziehen. Weiche und Surrogatendpunkte müssen sorgfältig im Hinblick darauf, wie gut sie die

untersuchte Krankheit repräsentieren, beurteilt werden.

Die Entscheidung, welche Endpunkte verwendet werden, hängt stark vom Wesen der untersuchten Krankheit ab. Krebs bietet zum Beispiel offensichtliche harte Endpunkte in Form von Überleben, während eine Bewertung von Depression unweigerlich weichere Endpunkte beinhalten muss. Andere Erkrankungen, wie etwa Diabetes, werden mit fest etablierten Surrogatendpunkten, wie etwa Blutzuckerspiegel, in Verbindung gebracht.

Sicherheit

Welche Nebenwirkungen hatte das Arzneimittel? Immer wenn der Arzt, der eine klinische Prüfung durchführt, einen Patienten untersucht, fragt er, ob der Patient etwas Unerwünschtes bei sich festgestellt hat. Informationen zu diesen „unerwünschten Ereignissen“ werden gesammelt und später analysiert, um Aufschluss über einen möglichen Kausalzusammenhang mit dem untersuchten Arzneimittel zu geben. Wenn ein solcher Kausalzusammenhang festgestellt wird, wird das unerwünschte Ereignis zu einer „Nebenwirkung“. Besondere Aufmerksamkeit wird „schweren“ Nebenwirkungen geschenkt – jene, die lebensbedrohlich oder mit Tod, Krankenhausaufenthalt oder Geburtsfehlern verbunden sind.

Typ der klinischen Prüfung

Klinische Prüfungen variieren stark in Bezug auf Größe, Dauer und Design. Diese Faktoren spielen eine wichtige Rolle bei der Interpretation von Studienergebnissen.

Das informativste Design für eine klinische Prüfung ist der „doppelblinde, randomisierte Vergleich“, bei dem einigen Patienten das neue Arzneimittel verabreicht wird, während andere eine alternative Behandlung erhalten. Die alternative Behandlung, die manchmal als „Kontrollbehandlung“ bezeichnet wird, kann Folgendes sein:

- Ein **Placebo** – eine unwirksame „Scheinbehandlung“
- Ein **wirksames Vergleichspräparat** – im Allgemeinen eine gut etablierte Behandlung für die untersuchte Erkrankung.

Die Teilnehmer werden per Zufallsprinzip den einzelnen Studiengruppen zugewiesen. Die Studie ist so aufgebaut, dass während der Laufzeit der Studie weder der Arzt noch der Patient weiß, wer welche Behandlung erhält. Eine Studie, die so aufgebaut ist, wird als „doppelblind“ bezeichnet. Eine Doppelverblindung senkt das Potenzial für Verzerrungen der Ergebnisse.

Bei solchen Studien werden die Ergebnisse im Hinblick auf den Unterschied zwischen der Gruppe, die das neue Arzneimittel erhalten hat, und der Gruppe, die die Kontrollbehandlung erhalten hat, präsentiert:

- Bei einem Vergleich mit Placebo ist dieser Unterschied ein Maß für die reale Wirkung des neuen Arzneimittels.
- Bei einem Vergleich mit einem wirksamen Vergleichspräparat, gibt der Unterschied Aufschluss darüber, wie sich das neue Arzneimittel im Vergleich zur aktuellen medizinischen Praxis verhält.

In beiden Fällen wird wahrscheinlich über zwei Aspekte des Unterschieds berichtet:

- **Größe:** Diese wird oft als der tatsächliche Unterschied in einer bestimmten Studie im Zusammenhang mit einem „95-%-Konfidenzintervall“ berichtet. Dies ist der Bereich, innerhalb dessen wir zu 95 % sicher sein können, dass dort der wahre Unterschied für die Population liegen würde. Obwohl man eine statistische Signifikanz feststellen kann, muss sie nicht klinisch relevant sein. Allgemein gesprochen: je größer dieser Unterschied ist, desto wahrscheinlicher ist es, dass er klinisch relevant ist (eine Verlängerung des Überlebens

um Jahr ist klinisch relevanter als eine Verlängerung um einen Tag).

- **Statistische Signifikanz:** Da einige Einzelpersonen besser auf eine Behandlung ansprechen als andere, besteht immer das Risiko, dass sich der beobachtete Unterschied zwischen den Gruppen in einer klinischen Prüfung durch Zufall vergrößert hat. Wenn zum Beispiel alle von sich aus guten Responder in eine Gruppe randomisiert wurden und die schlechten Responder in die andere. Statistiker können berechnen, wie wahrscheinlich es ist, dass dieses Szenario in einer bestimmten klinischen Prüfung eingetreten ist, und bezeichnen ihr Ergebnis als „p-Wert“.

Ein p-Wert von 0,05 bedeutet, dass eine Wahrscheinlichkeit von 5 % bzw. 1 zu 20 besteht, dass der Unterschied durch Zufall entstanden ist. Er wird konventionell als Grenze genommen, bis zu der Ergebnisse als „statistisch signifikant“ akzeptiert werden. Es ist wichtig zu verstehen, dass der Begriff „signifikant“ in diesem Sinne nichts über die medizinische Bedeutung der Ergebnisse aussagt – er bietet lediglich eine Rückversicherung, dass das Ergebnis mit sehr geringer Wahrscheinlichkeit zufällig ist. Zum Beispiel könnte sich eine Verlängerung eines sechsminütigen Fußmarsches um einen Meter in einer ausreichend großen Studie als statistisch signifikant erweisen (d. h., dass es unwahrscheinlich ist, dass sie zufällig entstanden ist). Sie würde für einen Patienten mit Herzversagen oder seinen Arzt jedoch nicht als von klinischem Wert betrachtet werden.

Eine zweite wichtige Gruppe klinischer Prüfungen, die oft zur Untersuchung der Langzeit-Sicherheit durchgeführt werden, bilden die „offenen“ Studien. In diesen Studien gibt es keine Kontrollgruppe – alle werden mit dem neuen Arzneimittel behandelt und ihre Erfahrungen werden erfasst. Es können keine Unterschiede zwischen den Gruppen entstehen (entweder zufällig oder durch die tatsächlichen therapeutischen Wirkungen) und

somit ist eine Signifikanzberechnung überflüssig. Als Ausgleich zu diesen Mängeln schließen offene Studien oft eine große Anzahl an Patienten (bis zu mehrere Tausend) ein, die über lange Zeiträume (in einigen Fällen mehrere Jahre) untersucht werden. Diese Studien erleichtern es daher, seltene Nebenwirkungen sowie Nebenwirkungen, die sich erst nach langer Zeit entwickeln, aufzudecken.

Die Ergebnisse solcher Studien werden als überschaubare Tabellen präsentiert, in denen verschiedene unerwünschte Ereignisse und deren beobachtete Häufigkeit aufgelistet werden.