

Wer evidenzbasiert argumentiert, bezieht sich auf den aktuellen Wissensstand aus Studien, nicht nur auf persönliche Erfahrungen oder auf die Meinung von Experten. Das bloße Zitat einer Studie ist aber noch keine evidenzbasierte Argumentation. Studien bieten keine unumstößlichen Wahrheiten, sondern Ergebnisse statistischer Analysen. Jeder Studientyp hat dabei spezifische Stärken und Schwächen. Diese Artikelreihe konzentriert sich auf klinische Studien, die experimentell Wirksamkeit und Verträglichkeit von Arzneimitteln prüfen. In kurzen Beiträgen möchten wir Sie mit dem nötigen „Werkzeug“ ausstatten, um klinische Studien zu Arzneimitteln kritisch zu lesen und sich Ihre eigene, evidenzbasierte Meinung zu bilden.

Klinische Studien zu Arzneimitteln – Wo ist der Haken?

Verloren, aber nicht unersetzlich? – Vom Umgang mit fehlenden Daten

Rückblick: Per-protokoll, As-treated oder Intention-to-treat: Was machen wir mit den „Abtrünnigen“?

In Studien weichen Teilnehmer häufig von den geplanten Therapieschemata ab. Diese Abweichungen können durch Patientenmerkmale bedingt sein, die gleichzeitig den Krankheitsverlauf oder die Wirksamkeit und Verträglichkeit des Arzneimittels beeinflussen (z. B. vorbestehende Allergien). Bei einer Intention-to-treat (ITT)-Analyse werden Studienteilnehmer auch dann entsprechend ihrer ursprünglichen Zuteilung ausgewertet, wenn sie die Studienmedikation nicht wie vorgesehen einnahmen. Dadurch bleibt die durch die Randomisierung erzeugte Strukturgleichheit von Kontroll- und Interventionsgruppe erhalten. Bei As-treated (AT)- oder Per-protocol (PP)-Analysen ist dagegen nicht gewährleistet, dass die ausgewerteten Gruppen strukturgleich sind. Die Aussagesicherheit von AT- und PP-Analysen ist deshalb geringer als die Aussagesicherheit von ITT-Analysen.

Im Beispiel unserer [Artikelserie](#) möchten Sie herausfinden, ob die medizinische Handcreme Supersoft das Risiko für Handekzeme bei Pflegepersonal stärker reduziert als in Drogerien erhältliche Handcremes. Dafür führen Sie eine kontrollierte randomisierte Studie durch, in der sie die Teilnehmer zufällig einer von zwei Behandlungen zuteilen¹: Die Studienteilnehmer tragen dreimal täglich entweder Supersoft (Interventionsgruppe) oder ihre bisher genutzte Handcreme (Kontrollgruppe) auf. Durch die Randomisierung haben Sie (tendenziell) strukturgleiche Kontroll- und Vergleichsgruppen erreicht, beispielsweise hinsichtlich der Häufigkeit vorbestehender Allergien. Damit diese Strukturgleichheit erhalten bleibt, wollen Sie eine Intention-to-treat (ITT)-Analyse durchführen: Sie möchten alle Studienteilnehmer entsprechend ihrer ursprünglichen Zuteilung auswerten („once randomized, always analyzed“ bzw. „analysiert wie randomisiert“). Sie laden deshalb alle 16 Teilnehmer ein Jahr nach Studienbeginn zur Abschlussvisite ein, auch diejenigen, die ihre zugeteilte Handcreme nicht wie vorgesehen benutzen.

Das Ergebnis der Abschlussvisite ist in Abbildung 1 dargestellt und scheint eindeutig für die Handcreme Supersoft zu sprechen: In der Abschlussvisite weist kein Patient der Interventionsgruppe, aber fast 30 % der Kontrollgruppe ein Handekzem auf. Wo könnte hier der Haken sein?

Einhart, N.

Mathes, T.

¹ Bei diesem Gedankenexperiment klammern wir statistische Fragen zur Errechnung der geeigneten Stichprobengröße ebenso aus wie die notwendige Bewilligung Ihrer Studie durch die Ethikkommission.

	Intervention (Supersoft Handcreme)	Kontrolle (bisherige Handcreme)
Hautbefund in der Abschlussvisite		
 Patienten mit Handekzem  Patienten ohne Handekzem		

Abbildung 1: Zur Abschlussvisite erscheinen zwölf Studienteilnehmer. Bei der Untersuchung wird festgestellt, dass in der Interventionsgruppe kein Teilnehmer ein Handekzem aufweist, während in der Kontrollgruppe bei zwei Teilnehmern ein Handekzem besteht.

Offensichtlich sind nicht alle Teilnehmer ihrer Einladung zur Abschlussvisite gefolgt. Bei 25 % der Studienteilnehmer wissen Sie deshalb nicht, ob bei ihnen zum Studienende ein Handekzem bestand. Es sind viele Gründe denkbar, warum Untersuchungsdaten fehlen. Denkbar sind beispielsweise folgende Szenarien:

1. Der Studienassistent war erkrankt und geplante Untersuchungen sind ausgefallen.
2. Studienteilnehmer haben aufgrund allergischer Reaktionen nicht nur die Studienmedikation abgesetzt, sondern ihre Teilnahme gänzlich abgebrochen.
3. Studienteilnehmer konnten nicht kontaktiert werden, weil sie den Arbeitsplatz gewechselt haben.

Fehlende Daten verzerren nur dann die Beurteilung der untersuchten Endpunkte, wenn 1. bestimmte Merkmale der Studienteilnehmer zu ihrem Fehlen führten und wenn 2. diese Merkmale durch die Wirksamkeit oder Verträglichkeit der Studienmedikation beeinflusst wurden. In Szenario 1 (Erkrankung des Studienassistenten) kann davon ausgegangen werden, dass die Eigenschaften der Studienteilnehmer keinen Einfluss darauf hatten, ob die Abschlussuntersuchung stattfand oder nicht; in Szenario 2 (Studienabbruch aufgrund unerwünschter Ereignisse) erscheint es plausibel, dass das Auftreten allergischer Reaktionen durch die Studienmedikation verursacht wurde. Komplizierter ist dagegen die Beurteilung von Szenario 3: Ein Arbeitsplatzwechsel kann durch gänzlich von der Studienmedikation unabhängige Gründe motiviert sein (beispielsweise durch Konflikte mit der Stationsleitung), aber auch auf eine unzureichende Wirksamkeit der Studienmedikation zurückzuführen sein (höheres Risiko für das Auftreten von Handekzemen). Die Beurteilung, ob Daten „zufällig“ (d. h. unabhängig von der untersuchten Intervention) fehlen, basiert somit immer auf mehr oder weniger sicheren Annahmen.

Abbildung 2 stellt ein Szenario dar, in dem mehrere Teilnehmer ihren Arbeitsplatz aufgrund eines Handekzems gewechselt haben. In beiden Behandlungsgruppen erkrankten im Studienverlauf drei von acht Teilnehmern an einem Handekzem. Unter Supersoft war das Handekzem so stark ausgeprägt, dass nach einem Jahr alle drei Pflegekräfte eine andere Tätigkeit ausübten und deshalb die Einladung zur Abschlussvisite nicht erhielten. In der Kontrollgruppe war das Handekzem deutlich leichter ausgeprägt. Nur ein Studienteilnehmer hatte aufgrund des Handekzems seinen Arbeitsplatz gewechselt, die beiden anderen Teilnehmer waren weiterhin in ihrer Station tätig und nahmen an der Abschlussuntersuchung teil. Die erfassten Daten suggerieren in diesem Szenario fälschlicherweise, dass unter Supersoft der Entwicklung von Handekzemen deutlich besser vorgebeugt wird als unter einer handelsüblichen Handcreme.

	Intervention (Supersoft Handcreme)	Kontrolle (bisherige Handcreme)
Hautbefund zu Studienende	                                      	

Forschungsschwerpunkte, konnten Sie abschätzen, welche Themen er bei Ihrer Prüfung mit einiger Wahrscheinlichkeit abfragen würde. Noch besser ließ sich eine Vorhersage treffen, wenn Sie wussten, welche Fragen er in der Vergangenheit bei Prüfungen gestellt hatte – weshalb Prüfungsprotokolle bei Medizinstudenten von Jahrgang zu Jahrgang weitergegeben werden. Aus dem gleichen Grund ist es in klinischen Studien wichtig, dass sowohl umfangreiche Baseline-Charakteristika erhoben werden als auch regelmäßige Zwischenerhebungen erfolgen.

Bei der Auswertung klinischer Studien ist die Nutzung von Imputationsverfahren nur zulässig, wenn davon ausgegangen wird, dass die nicht zufällig fehlenden Daten vollständig mit Hilfe der vorhandenen Daten erklärt werden können. Unter dieser Annahme der Ersetzbarkeit werden nicht zufällig fehlende Daten (verwirrenderweise) von Statistikern als **missing at random** (zufällig fehlend) beschrieben. Dagegen bezeichnen Statistiker Daten als **missing not at random** (nicht zufällig fehlend), wenn sie vermuten, dass unbekannte Faktoren die Datenerhebung beeinflussten, die mit den vorhandenen Daten nicht aufgeklärt werden können.³ In Szenario 3 wären fehlende Abschlussdaten somit „missing at random“, wenn der Arbeitsplatzwechsel und vorausgegangene Krankschreibungen dem Studienleiter bekannt wären (z. B. über eine Abfrage bei der Personalverwaltung) und sie wären „missing not at random“, wenn diese Informationen nicht für die Ersetzung der fehlenden Werte verwendet werden könnten.

Patienten teilen im Regelfall nicht detailliert mit, warum sie eine Studie abbrechen bzw. an einzelnen Datenerhebungen nicht teilnehmen. Studienleiter und Statistiker können somit oft nur mutmaßen, ob Daten zufällig fehlen und ob in ihre Imputation alle wesentlichen Faktoren einfließen. Zur Überprüfung der „missing completely at random“- und „missing at random“-Annahme sollten deshalb Sensitivitätsanalysen durchgeführt werden. Sensitivitätsanalysen prüfen, ob sich das Ergebnis der primären Analyse (Complete case oder Imputation) verändert, wenn bei den fehlenden Daten z. B. der schlechtestmögliche Fall („worst case scenario“), der bestmögliche Fall („best case scenario“) oder ein als plausibel eingeschätztes Szenario („base case“) angenommen wird.⁴ Abbildung 2 illustriert annähernd ein „worst case scenario“: In der Interventionsgruppe hatten alle Personen, die nicht zur Abschlussvisite erschienen sind, ein Handekzem entwickelt, während in der Kontrollgruppe die meisten Personen mit Handekzem an der Untersuchung teilnahmen.

Die Auswahl des Imputationsverfahrens kann das Studienergebnis erheblich beeinflussen (6). Die CONSORT-Leitlinie (7) gibt deshalb Empfehlungen, wie der Umgang mit fehlenden Daten bei randomisierten Studien berichtet werden soll. Leider finden sich bislang in Studienpublikationen häufig nur äußerst rudimentäre Informationen zum Umgang mit fehlenden Werten (z. B. keine Diskussion der Plausibilität der „missing at random“-Annahme). Oftmals bleibt sogar unklar, ob überhaupt eine Imputation oder Sensitivitätsanalyse angewandt wurde (8, 9). Hinweise auf ein niedriges Verzerrungspotenzial durch fehlende Daten ergeben sich dann lediglich dadurch, dass der Datenverlust in den Studienarmen ähnlich hoch ist und die Gründe für Studienabbrüche zwischen den Armen gleich verteilt sind (siehe Tabelle 1).

³ Diese Begriffe werden praxisnah in dem Artikel „What is the difference between missing completely at random and missing at random?“ (4) anhand eines fiktiven Gesprächs zwischen einem Kliniker und einem Statistiker erläutert.

⁴ Bhaskaran K, Smeeth L. What is the difference between missing completely at random and missing at random? *Int J Epidemiol* 2014; 43(4):1336–9. doi: 10.1093/ije/dyu080.

⁴ Vertiefende Informationen sowohl zu Imputationsverfahren als auch zu Sensitivitätsanalysen finden sich beispielsweise bei (5). Allgemeinverständlicher führt (2) in die Thematik ein.

⁵ Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. *Stat Med* 2020; 39(21):2815–42. doi: 10.1002/sim.8569.

⁶ Kahale LA, Khamis AM, Diab B, Chang Y, Lopes LC, Agarwal A et al. Potential impact of missing outcome data on treatment effects in systematic reviews: imputation study. *BMJ* 2020; 370:m2898. doi: 10.1136/bmj.m2898.

⁷ Hopewell S, Chan A-W, Collins GS, Hróbjartsson A, Moher D, Schulz KF et al. CONSORT 2025 explanation and elaboration: updated guideline for reporting randomised trials. *BMJ* 2025; 389:e081124. doi: 10.1136/bmj-2024-081124.

⁸ Ren Y, Jia Y, Huang Y, Zhang Y, Li Q, Yao M et al. Missing data were poorly reported and handled in randomized controlled trials with repeatedly measured continuous outcomes: a cross-sectional survey. *J Clin Epidemiol* 2022; 148:27–38. doi: 10.1016/j.jclinepi.2022.04.019.

⁹ Rombach I, Knight R, Peckham N, Stokes JR, Cook JA. Current practice in analysing and reporting binary outcome data—a review of randomised controlled trial reports. *BMC Med* 2020; 18(1):147. doi: 10.1186/s12916-020-01598-7.

Tabelle 1: Wo gibt es Informationen zum Umgang mit fehlenden Daten?

Methodische Frage ⁵	Fundort in der Primärpublikation
Unterscheidet sich der Anteil fehlender Daten zwischen den Armen?	Der Anteil fehlender Daten sollte pro Endpunkt und Erhebungszeitpunkt angegeben werden. Sofern publiziert, finden sich diese Informationen am ehesten in der tabellarischen Darstellung der Ergebnisse (Achtung: in der Kopfzeile der tabellarischen Endpunktdarstellung wird in der Regel die Anzahl der randomisierten, nicht die Anzahl der in der Auswertung berücksichtigten Patienten genannt.) Sofern keine endpunktbezogenen Informationen verfügbar sind, kann orientierend die Anzahl der Studienabbrecher pro Arm verglichen werden. Diese Information findet sich in der Regel in der Grafik „Flow of patients“ und/oder im Fließtext zu Beginn des Ergebnisteils („Results“). Insbesondere bei Patient-Reported Outcomes (PRO), z. B. zur Lebensqualität, kann sich die Antwortrate jedoch deutlich von der Anzahl der Studienabbrecher unterscheiden.
Unterscheiden sich die Gründe für die Studienabbrüche zwischen den Armen und könnte dies mit der Studienintervention in Zusammenhang stehen?	Die Gründe für den Studienabbruch werden zumeist in der Grafik „Flow of patients“ aufgeführt. Interessant ist insbesondere, wie viele Patienten die Studie aufgrund von Ereignissen im Zusammenhang mit der Studienintervention abgebrochen haben, z. B. aufgrund unzureichender Wirksamkeit („lack of efficacy“) oder aufgrund unerwünschter Ereignisse („adverse event“). Leider sind die genannten Gründe aber oft nicht hilfreich für die Beurteilung, ob ein Zusammenhang zur Studienintervention besteht („consent withdrawn“, „physician decision“, „other“).

⁵ Diese Fragen orientieren sich am Risk of Bias (RoB) 2 Tool von Cochrane (10). Das RoB 2 Tool wird in einem späteren Artikel dieser Reihe ausführlich vorgestellt.

10 Higgins JPT, Savovic J, Page MJ, Sterne AC, on behalf of the RoB-2-Development Group. Revised Cochrane risk-of-bias tool for randomized trials (RoB 2). 22.8.2019. Verfügbar unter: https://drive.google.com/file/d/19R9savfPdCHC8XLz2iiMvL_71IPJERWK/view.

Zusammenfassung

Fehlende Daten können das Studienergebnis verzerren, wenn ihr Fehlen durch die Verträglichkeit oder Wirksamkeit der Studienmedikation beeinflusst wurde. Mit Hilfe von Imputationsverfahren können fehlende Daten durch plausibel erscheinende Werte ersetzt werden. Die Aussagesicherheit der Imputationsverfahren sollte durch Sensitivitätsanalysen überprüft werden. Leider wird in Publikationen oftmals nur rudimentär geschildert, wie mit fehlenden Daten umgegangen wurde. Hinweise auf ein hohes Verzerrungsrisiko sind ein unterschiedlicher Anteil fehlender Daten sowie unterschiedliche Gründe für den Studienabbruch zwischen Interventions- und Kontrollgruppe.

Interessenkonflikte

Die Autorin und der Autor geben an, keine Interessenkonflikte zu haben.

Dr. med. Natascha Einhart

Prof. Dr. Tim Mathes