

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?

Per-protocol, As-treated oder Intention-to-treat: Was machen wir mit den „Abtrünnigen“?¹

Nehmen wir an, Sie wollen 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 zuteilen: Die Studienteilnehmer tragen dreimal täglich entweder Supersoft (Interventionsgruppe) oder ihre bisher genutzte Handcreme (Kontrollgruppe) auf. Es gelingt Ihnen ohne Probleme, eine ausreichend große Anzahl von Studienteilnehmern zu rekrutieren², aber schon wenige Wochen nach Studienbeginn läuft nicht mehr alles nach Plan:

- Herr Schmidt (Interventionsgruppe) benutzt Supersoft nach jedem Händewaschen, da sich seine Haut ansonsten trocken anfühlt.
- Frau Meyer (Interventionsgruppe) ist zurück zu ihrer bisherigen Handcreme gewechselt, weil unter Supersoft Juckreiz aufgetreten war.
- Herr Schneider (Kontrollgruppe) gibt auf Nachfrage zu, dass er im Dienst regelmäßig vergisst, sich die Hände einzucremen. Aus seiner Sicht ist die einmal tägliche Anwendung seiner Handcreme völlig ausreichend.
- Frau Müller (Kontrollgruppe) benutzt die Handcreme Supersoft, die ihr Frau Meyer freundlicherweise gegeben hat. Frau Müller freut sich, dass sie ein paar Euro sparen kann, weil sie sich keine eigene Handcreme kaufen muss.

Sie ärgern sich im Stillen über die unzuverlässigen Studienteilnehmer, die ihre sorgfältig geplante Studie ruinieren. Es erscheint naheliegend, dass Sie Herrn Schmidt, Frau Meyer, Herrn Schneider und Frau Müller nicht in die Auswertung einbeziehen, denn alle vier haben gegen die vorgesehene Anwendung der Handcreme verstoßen. Dieses Vorgehen entspräche einer **Per-protocol-Analyse (PP-Analyse)**: Hier werden nur diejenigen Teilnehmer ausgewertet, bei denen die Therapien (teilweise auch die Kontrolluntersuchungen) genau wie im Studienprotokoll geplant durchgeführt wurden. Eine PP-Analyse gefährdet jedoch die durch die Randomisierung erreichte Strukturgleichheit von Kontroll- und Interventionsgruppe (siehe „Randomisierung – der reine Zufall“), da das Abweichen mitunter nicht unabhängig von bestimmten Patientenmerkmalen ist (z. B. bei vorbestehenden Hautproblemen). Durch den Ausschluss von „Abtrünnigen“ ist somit nicht mehr gewährleistet,

Einhardt, N.

Mathes, T.

¹ Der Titel wurde durch folgenden leistungswerten Artikel inspiriert: (1). Die Bezeichnung als „abtrünnig“ soll humoristisch die Sicht des Studienleiters wiedergeben – selbstverständlich stellen wir nicht infrage, dass es den Studienteilnehmern jederzeit zusteht, von der geplanten Studienmedikation abzuweichen oder die Studie gänzlich abzubrechen.

² Bei diesem Gedankenexperiment klammern wir statistische Fragen zur Errechnung der geeigneten Stichprobengröße ebenso aus wie die notwendige Bewilligung Ihrer Studie durch die Ethikkommission.

Literatur

- 1 Schulz KF, Grimes DA. Reihe Epidemiologie Schlupflöcher in den Stichproben randomisierter Studien: Ausgeschlossene, Verlorene und Abtrünnige. Z Evid Fortbild Qual Gesundheitswes 2006; 100 (9308):467–73.

Rückblick: Randomisierung – der reine Zufall: Sichere Aussagen zur Wirksamkeit und Verträglichkeit eines Arzneimittels lassen sich nur dann aus einer Studie ableiten, wenn Kontroll- und Interventionsgruppe strukturgleich sind, d. h. wenn sich die Behandlungsgruppen nicht durch Merkmale unterscheiden, die den natürlichen Krankheitsverlauf oder das Ansprechen auf die Therapie beeinflussen. Strukturgleichheit wird am besten dadurch erreicht, dass die Teilnehmer zufällig den unterschiedlichen Behandlungsgruppen zugeteilt werden („Randomisierung“). Eine Einflussnahme durch Studienpersonal oder Teilnehmer wird verhindert, indem die Zuteilungsreihenfolge zufällig bestimmt wird und die Zuteilungsreihenfolge bis zum Studieneinschluss des Teilnehmers geheim gehalten wird.

dass Kontroll- und Interventionsgruppe in allen Merkmalen übereinstimmen, die den Krankheitsverlauf oder die Wirksamkeit des Arzneimittels beeinflussen:

- Herr Schmidt (Interventionsgruppe) weicht von der vorgesehenen Intervention ab, weil diese unzureichend wirksam ist. Wahrscheinlich hat Herr Schmidt eine besonders trockene Haut, vielleicht sogar eine atopische Dermatitis. Die Nichtberücksichtigung von Studienteilnehmern wie Herrn Schmidt würde zu einer Überschätzung der Wirksamkeit von Supersoft führen, da in der Interventionsgruppe der Anteil von Teilnehmern mit einem hohen Risiko für Handekzeme sinken würde.
- Frau Meyer (Interventionsgruppe) weicht von der vorgesehenen Intervention ab, weil diese für sie schlecht verträglich ist. Eine Ursache könnte beispielsweise eine vorbestehende Allergie sein. Der Ausschluss von Studienteilnehmern wie Frau Meyer würde zu einer Überschätzung der Verträglichkeit (Sicherheit) von Supersoft führen.
- Herr Schneider (Kontrollgruppe) weicht von der vorgesehenen Intervention ab, weil eine geringere Dosierung bei ihm ausreichend wirksam war. Der Ausschluss von Studienteilnehmern wie Herrn Schneider würde dazu führen, dass in der Kontrollgruppe der Anteil von Teilnehmern mit einem geringen Risiko für Handekzeme sinkt. Die Nichtberücksichtigung von Herrn Schneider würde somit ebenfalls zu einer Überschätzung der Wirksamkeit von Supersoft führen.
- Frau Müller (Kontrollgruppe) nutzt Supersoft aufgrund finanzieller Gründe. Es ist anzunehmen, dass bei Frau Müller bislang keine Unverträglichkeiten gegenüber Hautpflegeprodukten auftraten, da sie andernfalls nicht so aufgeschlossen gegenüber einer Handcreme wäre, die bei ihrer Kollegin eine allergische Reaktion ausgelöst hat. Durch den Ausschluss von Teilnehmern wie Frau Müller würde deshalb in der Kontrollgruppe der Anteil von Studienteilnehmern mit einem geringen Risiko für unerwünschte Arzneimittelwirkungen (UAW) sinken und die Verträglichkeit (Sicherheit) von Supersoft überschätzt werden.

Alternativ könnten Sie sich entscheiden, Frau Meyer und Frau Müller in die Datenanalyse einzuschließen, jedoch nicht entsprechend ihrer ursprünglichen Zuteilung, sondern entsprechend der Therapie, die sie tatsächlich (überwiegend) erhalten haben. In diesem Fall würden Sie eine **As-treated-Analyse (AT-Analyse)** durchführen: Frau Meyer würde in der Kontrollgruppe ausgewertet werden und Frau Müller in der Interventionsgruppe. Eine AT-Analyse wird am häufigsten bei Studien angewendet, die eine chirurgische Intervention oder ein Medizinprodukt mit einer medikamentösen Therapie vergleichen, da bei diesen Studien der „cross-over“ (der Wechsel in den anderen Studienarm) eindeutig zu bestimmen ist. Auch durch eine AT-Analyse wird die durch die Randomisierung erzeugte Strukturgleichheit gefährdet. In unserem Beispiel würde die AT-Analyse die Verträglichkeit von Supersoft noch stärker überschätzen als die PP-Analyse: Durch eine AT-Analyse sinkt hier der Anteil von Patienten mit einem geringen Risiko für UAW (Frau Müller) im Kontrollarm und steigt gleichzeitig im Interventionsarm an; außerdem nimmt der Anteil von Patienten mit hohem Risiko für UAW (Frau Meyer) im Kontrollarm zu und sinkt gleichzeitig im Interventionsarm.

Ähnlich wie bei Beobachtungsstudien wird bei PP- und AT-Analysen versucht, die Gründe für Protokollverletzungen zu erfassen und bei der Auswertung zu berücksichtigen.

Alle zur Verfügung stehenden statistischen Methoden sind jedoch nicht so gut wie der reine Zufall, der bei der Randomisierung zur Strukturgleichheit der Behandlungsgruppen führte. Um Verzerrungen zu vermeiden, sollte deshalb immer eine **Intention-to-treat-Analyse (ITT-Analyse)** durchgeführt werden: Alle Teilnehmer werden bei der Auswertung in diejenige Behandlungsgruppe eingeschlossen, in die sie anfänglich randomisiert wurden, unabhängig davon, ob sie die zugeteilte Therapie in der vorgesehenen Weise erhielten, ja sogar dann, wenn sie – wie Frau Meyer und Frau Müller – die Therapie einer anderen Behandlungsgruppe erhielten (Abbildung 1).

	Intervention (Supersoft Handcreme)	Kontrolle (bisherige Handcreme)
Randomisierung		
Therapie im Studienverlauf		
Intention-to-treat-Analyse		
Per-protocol-Analyse		
As-treated-Analyse		
	Patienten mit vorbestehenden Allergien Patienten ohne vorbestehende Allergien	

Abbildung 1: Methodischer Umgang mit fehlender Adhärenz: Nach der Randomisierung sind im Interventions- und Kontrollarm ähnlich viele Patienten, die an Allergien leiden. Hier ist grafisch der Wechsel der Medikation von Frau Meyer dargestellt; der Wechsel von Frau Müller wird aus Gründen der Übersichtlichkeit nicht dargestellt. Im Studienverlauf wechselt Frau Meyer, eine Patientin mit vorbestehender Allergie, aufgrund einer Unverträglichkeit (UAW) zur Kontrolltherapie. In der ITT-Analyse wird Frau Meyer trotzdem in der Interventionsgruppe berücksichtigt. Das Verhältnis von Patienten mit und ohne Allergien bleibt deshalb bei der ITT-Analyse in Interventions- und Kontrollgruppe gleich. In der PP-Analyse wird Frau Meyer dagegen von der Analyse ausgeschlossen. Der Anteil von Patienten mit Allergien ist deshalb bei einer PP-Analyse im Kontrollarm größer als im Interventionsarm. In der AT-Analyse wird Frau Meyer der Kontrollgruppe zugeordnet. Der Anteil von Patienten mit Allergien ist hierdurch in der Kontrollgruppe deutlich größer als in der Interventionsgruppe.

In einer ITT-Analyse wird der Behandlungseffekt tendenziell geringer eingeschätzt als in PP- und AT-Analysen. In einer metaepidemiologischen Studie (2) war der durchschnittliche Unterschied zwischen PP- und ITT-Analysen mit 2 % zwar gering, nahm jedoch bei einem steigenden Anteil nicht adhärenter Patienten zu. Wenn 20 % der Teilnehmer das Studienprotokoll verletzen, ist laut dieser Auswertung mit einem 5 % größeren Behandlungseffekt in der PP-Analyse als in der ITT-Analyse zu rechnen. Bei einigen Studien ist deshalb der Behandlungseffekt nur in der PP- oder AT-Analyse statistisch signifikant bzw. klinisch relevant, nicht aber in der ITT-Analyse. Die Behandlungsentscheidung sollte sich in diesem Fall nicht auf die PP- oder AT-Analyse stützen: **Aus PP- und AT-Analysen kann kein Beleg für die Überlegenheit einer Therapie abgeleitet werden**, da durch das Brechen der Randomisierung eine kausale Rückführung des Therapieeffekts auf die Studienbehandlung unzulässig ist. Neben ihrer höheren Aussagesicherheit hat die ITT-Analyse zudem einen wei-

2 Mostazir M, Taylor G, Henley WE, Watkins ER, Taylor RS. Per-Protocol analyses produced larger treatment effect sizes than intention to treat: a meta-epidemiological study. *J Clin Epidemiol* 2021; 138:12–21. doi: 10.1016/j.jclinepi.2021.06.010.

teren Vorteil: Die Übertragbarkeit des per ITT-Analyse geschätzten Effekts auf die Versorgungspraxis ist besser als bei einer PP- oder AT-Analyse, denn in der Routineversorgung ist der perfekt adhärenste Patient vermutlich noch seltener anzutreffen als unter Studienbedingungen. Die PP- und AT-Analysen haben deshalb ihren Stellenwert vor allem als ergänzende Analysen bei häufigen Protokollverletzungen bzw. hohem Cross-over (siehe unten: “Pro und Kontra: Methodischer Umgang mit Abweichungen vom Studienprotokoll”). Insbesondere bezüglich der Verträglichkeit kann hier eine AT-Analyse Zusatzinformationen bieten.

Die ITT-Analyse ist schwierig umzusetzen, wenn relevante Ergebnisdaten fehlen. Wie können Sie eine ITT-Analyse durchführen, wenn Herr Schneider und Frau Meyer die Studie gänzlich abbrechen und nicht zur Abschlussvisite erscheinen? Der Umgang mit fehlenden Werten ist Thema des nächsten Artikels.

Pro und Kontra: Methodischer Umgang mit Abweichungen vom Studienprotokoll

Angewandte Methode	Pro	Kontra
Intention-to-treat (ITT)	Die Strukturgleichheit der Behandlungsgruppen bleibt erhalten, so dass eine hohe Aussagesicherheit besteht. Die Situation der Behandlungsentscheidung – mit unsicherer Adhärenz des Patienten – wird abgebildet.	Bei häufigen Protokollverletzungen sind die Studienergebnisse nicht auf adhärenste Patienten übertragbar.
Per-protocol (PP)	Bei häufigen Verletzungen des Studienprotokolls als ergänzende Analyse sinnvoll.	Die Strukturgleichheit der Behandlungsgruppen ist nicht gesichert. Die Aussagesicherheit ist deshalb reduziert.
As-treated (AT)	Bei hoher Cross-over-Rate als ergänzende Analyse sinnvoll, insbesondere bei Sicherheitsanalysen.	

Modifiziert = „gemauschelt“?

Als modifizierte Intention-to-treat (ITT)-Analyse wird häufig eine Analyse bezeichnet, in der Studienteilnehmer aufgrund bestimmter, erst nach der Randomisierung festgestellter Merkmale ausgeschlossen werden. Ist die Anzahl dieser Patienten sehr gering, beeinflusst ihr Ausschluss die Aussagesicherheit der Studie nicht. In einzelnen Studien wird jedoch ein beträchtlicher Anteil der randomisierten Teilnehmer aus der primären Analyse ausgeschlossen. Ein Beispiel hierfür ist die Studie CAPSTONE-2 (3). In dieser Studie wurden 2184 Patienten mit influenzatypischen Symptomen, deren Beginn maximal 48 Stunden zurück lag, entweder zu Baloxavir, Oseltamivir oder Placebo randomisiert. Analysiert wurden allerdings nur 1163 Patienten, nämlich diejenigen, deren Influenza-Infektion per PCR-Test bestätigt wurde – das heißt nur etwas mehr als die Hälfte der randomisierten Patienten. Warum ist dies problematisch? Das IQWiG erklärte in seiner Dossierbewertung (4), dass eine Auswertung der Gesamtpopulation die Verhältnisse in der Versorgung deutlich besser widerspiegelt als die Beschränkung auf Patienten mit labor diagnostischem Influenznanachweis: In der Versorgung wird üblicherweise die Entscheidung zur antiviralen Therapie in Abhängigkeit von der klinischen Symptomatik getroffen, ohne dass zuvor das Ergebnis des PCR-Tests abgewartet wird, da ein früher Behandlungsbeginn mit einer höheren Wirksamkeit der antiviralen Therapie assoziiert ist. Bei der Therapieentscheidung für oder gegen Baloxavir ist deshalb nicht die Wirksamkeit bei labor diagnostisch bestätigter Influenza relevant, sondern die Wirksamkeit bei klinischem Verdacht auf Influenza. Da Baloxavir bei einer nicht durch Influenza bedingten Symptomatik wirkungslos ist, überschätzt die publizierte Analyse³ die Wirksamkeit von Baloxavir im Versorgungsalltag deutlich (siehe auch (5)).

³ In der Publikation der Studie CAPSTONE-2 wird die vorgenommene ITT-Analyse nicht als „modifiziert“ bezeichnet, sondern als „ITTI“ (Intention-to-treat infected).

- Ison MG, Portsmouth S, Yoshida Y, Shishido T, Mitchener M, Tsuchiya K et al. Early treatment with baloxavir marboxil in high-risk adolescent and adult outpatients with uncomplicated influenza (CAPSTONE-2): a randomised, placebo-controlled, phase 3 trial. *Lancet Infect Dis* 2020; 20(10):1204–14. doi: 10.1016/S1473-3099(20)30004-9.
- Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Baloxavir marboxil (Influenza) – Nutzenbewertung gemäß § 35a SGB V – Dossierbewertung. IQWiG-Berichte – Nr. 1107; Auftrag A21-21; Version 1.0; 12.5.2021. Verfügbar unter: https://www.iqwig.de/download/a21-21_baloxavir-marboxil_nutzenbewertung-35a-sgb-v-v1-0.pdf.
- Arzneimittelkommission der deutschen Ärzteschaft. Baloxavir marboxil (Xofluzar®). *Arzneiverordnung in der Praxis (AVP)* 2021; 48(3–4):100–7.

Wo finde ich Angaben zur Intention-to-treat-, Per-protocol- oder As-treated-Analyse?

Die gewählte Analyse wird im Methodenteil unter "Statistical Analysis" beschrieben. In manchen Studien erfolgt eine „modifizierte“ ITT-Analyse, bei der bestimmte Studienteilnehmer nicht berücksichtigt werden. Dies birgt das Risiko für Verzerrungen. Die Behauptung einer „ITT“-Analyse sollte deshalb überprüft werden:

- anhand der graphischen Darstellung des Patientenflusses (häufig im Anhang, bezeichnet als „Flow of patients“ oder „Trial Consort Diagram“)
- oder durch einen Vergleich zwischen der Anzahl randomisierter Teilnehmer (angegeben im Abstract sowie im Ergebnisteil unter „Results“) und der Anzahl der in die Analyse eingeschlossenen Teilnehmer (angegeben in der Tabelle zu Wirksamkeits-/Sicherheitsendpunkten).

Fazit

Bei einer Intention-to-treat(ITT)-Analyse werden Studienteilnehmer entsprechend ihrer ursprünglichen Zuteilung ausgewertet, auch dann, wenn sie die Studienmedikation nicht wie vorgesehen einnahmen („once randomized, always analyzed“ bzw. „analysiert wie randomisiert“). Dadurch ist gewährleistet, dass Kontroll- und Interventionsgruppe in allen Merkmalen übereinstimmen, die den Krankheitsverlauf oder die Wirksamkeit des Arzneimittels beeinflussen. Per-protocol(PP)- und As-treated(AT)-Analysen sollten aufgrund ihrer geringeren Aussagesicherheit immer nur ergänzend herangezogen werden. Insbesondere ist Skepsis angebracht, wenn ein Behandlungseffekt sich nur in einer PP- oder AT-Analyse zeigt, nicht aber in der ITT-Analyse.

Interessenkonflikte

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

Dr. med. Natascha Einhart, Berlin
natascha.einhart@baek.de

Prof. Dr. Tim Mathes, Köln