Intention-to-treat

H. Faller

Intention-to-treat

Zusammenfassung

Was bedeutet eigentlich Intention-to-treat-Auswertung? Warum soll man in kontrollierten Wirksamkeitsstudien Patienten in derjenigen Gruppe auswerten, der sie durch die Randomisierung zugewiesen wurden, unabhängig davon, ob sie die geplante Behandlung auch abgeschlossen haben? In diesem Beitrag wird die Logik des Intention-to-treat-Prinzips dargelegt. Es wird erläutert, wie es infolge einer systematischen Verzerrung der Untersuchungsstichproben dazu kommen kann, dass man Interventionseffekte erzeugt, die in der untersuchten Population gar nicht vorhanden sind, wenn man Patienten nachträglich aus der Analyse ausschließt. Dieser Gefahr kann das Intention-to-treat-Prinzip vorbeugen.

Schlüsselwörter
Intention-to-treat (ITT) · randomisierte kontrollierte Studien

Abstract

What is meant by intention-to-treat? Why should data be analyzed in controlled trials in a way that all participants are included in the group to which they were randomly assigned, regardless of whether they completed the intervention given to the group? In this Tutorial, the logic of the intention-to-treat principle is outlined. It is shown that study results may be biased by excluding patients post hoc thus producing spurious effects that do not exist in the population under study. The intention-to-treat strategy avoids this bias.

Key words
Intention-to-treat, ITT · randomized controlled trials

In den maßgeblichen Empfehlungen für die Durchführung, Auswertung und Publikation randomisierter kontrollierter Studien, wie dem CONSORT-Statement [1,2] oder dem Lehrbuch „Evidence-based medicine“ [3], wird gefordert, Daten randomisierter kontrollierter Studien nach dem „intention-to-treat“-Prinzip zu analysieren. Was ist damit gemeint? Auswertung gemäß Intention-to-treat (ITT) bedeutet, Teilnehmer an klinischen Studien in derjenigen Studiengruppe (Behandlungsgruppe oder Kontrollgruppe) auszuwerten, der sie ursprünglich zugewiesen waren, also nach der Behandlungs*intention* und nicht dem Ausmaß der tatsächlich durchgeführten Behandlung [4,5].

Dieses Vorgehen ist auf den ersten Blick kontraintuitiv: Wenn es um die Wirksamkeit einer Behandlung geht, dann erscheint es nicht sinnvoll, in der Behandlungsgruppe auch Patienten mitzuführen, die die zu prüfende Behandlung nicht in ausreichendem

Koordinatoren der Reihe „Methoden in der Rehabilitationsforschung“:
Prof. Dr. Dr. Hermann Faller, Würzburg; Prof. Dr. Thomas Kohlmann, Greifswald;
Dr. Christian Zwingmann, Frankfurt/Main
Interessenten, die einen Beitrag zur Reihe beisteuern möchten, werden gebeten, vorab Kontakt aufzunehmen,
E-mail: christian.zwingmann@vdr.de

Institutsangaben
Stiftungsprofessur Rehabilitationswissenschaften, Universität Würzburg

Korrespondenzadresse
Prof. Dr. med. Dr. phil. Hermann Faller · Institut für Psychotherapie und Medizinische Psychologie · Klinikstraße 3 · 97070 Würzburg · E-mail: h.faller@mail.uni-wuerzburg.de

Bibliografie
Rehabilitation 2004; 43: 52–55 · © Georg Thieme Verlag Stuttgart · New York · ISSN 0034-3536 · DOI 10.1055/s-2004-818553

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

Maß oder sogar überhaupt nicht erhielten; plausibel ist vielmehr, in die Behandlungsgruppe nur die tatsächlich behandelten Patienten einzubeziehen und die nicht protokollgemäß behandelten Patienten aus der Auswertung auszuschließen. Dieser naheliegende Gedankengang ist jedoch unzutreffend und kann zu einer groben Überschätzung der Wirksamkeit der geprüften Intervention führen. Was sind die Gründe dafür, und warum soll man stattdessen nach dem ITT-Prinzip auswerten?

Hierzu ist es hilfreich, sich vor Augen zu führen, aus welchen Gründen Studienteilnehmer unter Umständen nicht die geplante Behandlung erhalten: Es könnte zum Beispiel passieren, dass ein ursprünglich der Experimentalbedingung zugeteilter Patient schon kurz vor Beginn der Therapie so schwer krank war, dass er für die Behandlung nicht mehr infrage kam; das Risiko unerwünschter Nebenwirkungen könnte in einem anderen Fall zu groß geworden sein, so dass er vom behandelnden Arzt aus der Studie herausgenommen wurde. Oder der Patient selbst könnte unerwünschte Wirkungen verspürt und die Therapie von sich aus beendet haben. Ein Therapieabbruch kann seitens des Patienten auch aus wahrgenommenem mangelnden Erfolg der Behandlung resultieren; im schlimmsten Fall wechselt der Studienteilnehmer in einen anderen Studienarm, von dem er sich bessere Ergebnisse verspricht.

All diese Faktoren führen zusammengenommen dazu, dass eine hoch selektive Teilgruppe übrigbleibt, die aller Wahrscheinlichkeit nach eher gut von der Behandlung profitiert hat, weil diejenigen Studienteilnehmer, die Nebenwirkungen oder unzureichenden Erfolg aufwiesen, im Laufe der Studie ausgeschieden sind. Wenn am Ende aber nur diejenige Teilgruppe von Patienten noch in der Studie ist, die erfolgreich mit der zu prüfenden Intervention behandelt werden konnte, ist es nicht mehr überraschend, wenn der Vergleich mit der Kontrollgruppe einen Behandlungserfolg zeigt. Ein so zustande gekommenes Ergebnis ist durch das Vorgehen der Untersucher präjudiziert und deshalb zirkulär: Bei denjenigen Patienten, die erfolgreich behandelt worden sind, ist die Behandlung erfolgreich. Mit anderen Worten: Das Ausscheiden (Drop-out) von Studienteilnehmern ist selbst ein Aspekt des Ergebnisses (Outcome), und der nachträgliche Ausschluss der vermutlich weniger erfolgreich behandelten Studienteilnehmer stellt eine Ex-post-Veränderung der Stichprobe dar, die zu einer Verzerrung zugunsten der geprüften Behandlung führt.

Was hier vielleicht etwas pointiert und übertrieben wirkt, ist in publizierten Studien durchaus Realität gewesen: Behandlungen waren scheinbar erfolgreich, wenn die Stichprobe entsprechend deformiert wurde; wenn aber alle Patienten eingeschlossen blieben (ITT), löste sich der vermeintliche Benefit in Luft auf. Wenn zum Beispiel in einer Studie zur chirurgischen Schlaganfallprophylaxe (zit. n. [2]) nur diejenigen Patienten in die Auswertung einbezogen wurden, die erfolgreich gefäßoperiert werden konnten, so zeigte die Operation anscheinend einen Nutzen im Vergleich zur Kontrollgruppe. Bezog man aber nach dem ITT-Prinzip auch diejenigen Patienten ein, bei denen es Komplikationen gab oder die wegen ihrer ungünstigen Prognose gar nicht erst operiert wurden, obwohl sie eigentlich der Operation zugewiesen waren, so verschwand der Effekt. Der Unterschied kam also lediglich dadurch zustande, dass bei der ursprünglichen Auswertungsstrategie in der Behandlungsgruppe Patienten mit ungünstiger Prognose ausgeschlossen wurden, in der Kontrollgruppe aber nicht [6]. Analog war in einer Studie zum Vergleich chirurgischer und konservativer Behandlung der koronaren Herzkrankheit (zit. n. [5]) die Bypass-Operation der medikamentösen Behandlung nur dann überlegen, wenn man – in Abweichung vom ITT-Prinzip – Patienten, die schon vor der geplanten Operation verstarben, aus der Analyse ausschloss.

Ein weiterer Grund, warum es in die Irre führen kann, nur diejenigen Studienteilnehmer auszuwerten, die eine hohe Compliance mit der Behandlung aufwiesen, besteht darin, dass die Tatsache der Compliance selbst ein Indikator einer günstigen Prognose ist. In einer Studie zur Wirksamkeit eines lipidsenkenden Medikaments (zit. n. [7]) zeigte sich ein signifikanter Unterschied, wenn man diejenigen Patienten, die sich an die Behandlung hielten, mit denjenigen verglich, die eine geringe Adhärenz mit der Therapie an den Tag legten. Dieser Unterschied wurde als Beweis für die Wirksamkeit des Medikaments interpretiert. Als man jedoch die gleiche Auswertung (Patienten mit Compliance versus Patienten ohne Compliance) für die Plazebo-Kontrollgruppe durchführte, fand sich ebenfalls eine signifikante Überlegenheit der Teilgruppe mit hoher Compliance im Vergleich zu derjenigen mit geringer Compliance. Compliance bedeutete in der Kontrollgruppe wohlgemerkt Compliance mit dem Plazebo! Patienten, die sich an eine wie auch immer geartete „Behandlung“ halten, haben wahrscheinlich auch generell ein günstigeres Gesundheitsverhalten, welches sich prognostisch vorteilhaft auswirkt.

Auch in onkologischen Studien besteht ein häufiger Fehler darin, nur diejenigen Patienten auszuwerten, die einen Therapieresponse zeigen; dann ist es nicht verwunderlich, dass diese Gruppe länger überlebt als die Kontrollgruppe, weil in letzterer analoge Patienten mit ungünstiger Prognose, die unter einer Behandlungsbedingung vermutlich ebenfalls seltener einen Response gezeigt hätten, nicht ausgeschlossen wurden. Noch frappierender ist ein Vorgehen, bei welchem innerhalb der Behandlungsgruppe die Patienten mit und ohne Response hinsichtlich ihrer Überlebenszeiten verglichen werden [8]. Die Aussage derartig analysierter Studie lautet letztlich wiederum: Erfolgreich behandelte Patienten haben einen Behandlungserfolg.

Diese Fallstricke können vermieden werden, wenn die Stichproben nachträglich nicht verändert werden. Deshalb gilt die Auswertung nach dem ITT-Prinzip als Qualitätsmerkmal einer Effektivitätsstudie. Das ITT-Prinzip bedeutet, wie gesagt, dass alle Patienten, die einem bestimmten Studienarm (Experimentalgruppe oder Kontrollgruppe) per Randomisierung zugewiesen wurden, in der statistischen Auswertung auch gemäß dieser Zuordnung behandelt werden. Alle Patienten, die der Experimentalgruppe zugewiesen wurden, werden auch beim Gruppenvergleich nach Abschluss der Intervention als Angehörige der Experimentalgruppe ausgewertet und mit der Kontrollgruppe verglichen, unabhängig davon, wie viel Therapie sie erhielten, ja selbst dann, wenn sie gar nicht erst mit der Behandlung begannen. Nur durch dieses konservative Vorgehen ist gewährleistet, dass das Prinzip der Randomisierung, das dafür sorgt, dass sich Behandlungs- und Kontrollgruppe lediglich zufällig (nicht aber in prognoserelevanten Merkmalen) unterscheiden, aufrechter-

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

halten bleibt. Jede willkürliche Veränderung der Gruppenzusammensetzung hingegen trägt die Gefahr einer systematischen Verzerrung (Bias) in sich, die die interne Validität, d. h. die Interpretierbarkeit von etwaigen Gruppenunterschieden als Effekte der zu prüfenden Intervention, infrage stellt. Jede Auswertung, die nicht nach ITT geschieht, zerstört die Logik der Randomisierung.

ITT bezieht seine Legitimation im Übrigen nicht nur aus der Forschungslogik, sondern auch aus der Nähe zur Realität klinischer Entscheidungen. Wenn ein Arzt einem Patienten eine Behandlung empfiehlt, kennt er deren voraussichtliche Wirkung bei diesem Patienten nicht mit letzter Sicherheit. Er kann ihn nicht im Vorhinein von der Behandlung ausschließen, weil er später möglicherweise Nebenwirkungen erleiden wird oder mit dem Ergebnis zu einem späteren Zeitpunkt nicht zufrieden ist; diese Ereignisse liegen noch in der Zukunft. Deshalb entspricht die ITT-Analyse mehr der klinischen Situation, in welcher ein Behandlungsverfahren einmal eingesetzt werden soll: ITT geht von der Behandlungsintention aus, ohne die späteren Komplikationen schon vorwegnehmen zu können, die der Arzt in der klinischen Situation im Einzelfall noch nicht vorauswissen kann. Ein nachträglicher Ausschluss von Patienten nutzt hingegen Informationen aus der Zukunft (ex post), die den wahrscheinlichen Behandlungserfolg voraussagen lassen. Dies widerspricht aber, wie ausgeführt, der späteren klinischen Anwendung der geprüften Intervention. Der ITT-Ansatz prüft deshalb eher das Angebot einer Intervention, wie sie später in der klinischen Realität eingesetzt wird („health policy"); er testet eher die Wirksamkeit unter Praxisbedingungen („effectiveness", „pragmatic trials") als die Effektivität unter Laborbedingungen („efficacy", „explanatory trials") [5].

Eine Studie, die die Wirksamkeit (Effektivität) einer Behandlung (Intervention) prüft, sollte also immer zunächst nach dem Intention-to-treat-Prinzip ausgewertet werden. Voraussetzung der ITT-Auswertung ist es, dass von den Studienteilnehmern, die die Behandlung abgebrochen haben, gleichwohl Daten zu denjenigen Messzeitpunkten vorliegen, die die Grundlage der Auswertung bilden. Deshalb sollten alle Anstrengungen unternommen werden, um das Ausmaß fehlender Daten (Missing Data) möglichst gering zu halten. Häufig lässt sich nicht ganz vermeiden, dass Werte fehlen; über den Umgang mit diesem Problem gibt es noch keinen Konsens. Der Ausschluss des ganzen Falls kann zu einer Verzerrung führen, wenn fehlende Werte nicht zufällig, sondern in Abhängigkeit vom Outcome verteilt sind. Mögliche weiterführende Strategien (z. B. „last response carried forward", d. h. die Verwendung der letzten vorhandenen Beobachtung; Imputationsverfahren, d. h. Einsetzung geschätzter Werte auf der Basis vorhandener Daten; Sensitivitätsanalysen, d. h. Prüfung der Stabilität der Ergebnisse unter unterschiedlichen Modellannahmen, z. B. Zählen aller fehlenden Werte als Misserfolge oder als Erfolge) können aber ebenfalls zu Fehlern führen und müssen in dieser Hinsicht diskutiert werden [5].

Es ist den Autoren einer Studie gleichwohl unbenommen, zusätzlich eine Auswertung nach der tatsächlich erhaltenen Behandlung durchzuführen („per protocol", d. h. bezogen auf diejenigen Patienten der Behandlungsgruppe, die die Behandlung wie im Protokoll geplant erhielten; „as treated", d. h. bezogen auf alle Patienten aus Behandlungs- und Kontrollgruppe, die die Behandlung erhielten). Die festgestellten Effekte der Behandlung werden in dieser sekundären Analyse wahrscheinlich deutlicher ausfallen. Warum das zu erwarten ist, dürfte nach den Ausführungen dieses Beitrags klar geworden sein. Gleichwohl sollten bei der Auswertung „wie behandelt" diejenigen Faktoren, die zu einem Abbruch der Behandlung oder einer reduzierten Dosis führten, statistisch kontrolliert werden, um zu einem möglichst wenig verzerrten Ergebnis zu gelangen. Weiterhin sollten a priori die Einschlusskriterien so gefasst werden, dass Patienten, die ein hohes Risiko für Non-Compliance tragen, nicht in die Studie gelangen, um die Gefahr von Drop-outs zu verringern. Eine engere Fassung der Einschlusskriterien schlägt sich natürlich in einer verminderten Generalisierbarkeit der Ergebnisse der Studie nieder.

Obwohl das Bewusstsein für die methodische Qualität von Studien in den letzten Jahren zugenommen hat, wurde nur bei knapp der Hälfte der im Jahr 1997 in hochkarätigen wissenschaftlichen Zeitschriften publizierten randomisierten kontrollierten Studien ein ITT-Ansatz realisiert, und dessen konkrete Durchführung wurde häufig unzureichend oder gar falsch beschrieben [5]. Autoren sollten explizit angeben, ob es Abweichungen vom Studienprotokoll gab und wie sie mit denjenigen Studienteilnehmern umgingen, die nicht wie im Protokoll vorgesehen behandelt wurden oder bei denen die Daten hinsichtlich des hauptsächlichen Outcome-Kriteriums fehlen; die bloße Nennung des Begriffs Intention-to-treat reicht nicht aus.

Der ITT-Ansatz ist sicherlich kein Allheilmittel. Ferguson et al. [9] diskutieren Situationen, in denen es gerechtfertigt sein kann, Patienten nachträglich aus der Analyse auszuschließen, z. B. wenn sich herausstellt, dass sie die Einschlusskriterien nicht erfüllten und fälschlicherweise randomisiert wurden. In diesem Fall ist nicht davon auszugehen, dass ein nachträglicher Ausschluss das Ergebnis der Studie beeinflusst. Auch der ITT-Ansatz hat seine Grenzen, insofern er eher zu einer Unterschätzung als zu einer Überschätzung der Wirksamkeit einer Behandlung führen kann. Dadurch ist jedoch andererseits die Gefahr verringert, Effekte zu finden, die in Wirklichkeit gar nicht vorhanden sind, und fälschlicherweise Konsequenzen zu ziehen, die weitreichende klinische und ökonomische Implikationen haben können. Zwar kann auch die Datenanalyse gemäß ITT zu Irrtümern führen – aber wenn man schon irrt, dann besser auf der konservativen Seite der Entscheidung.

Die Empfehlung lautet deshalb zusammenfassend, die Auswertung hinsichtlich der Hauptzielgröße primär nach dem ITT-Prinzip durchzuführen und zusätzlich sekundär eine Per-protocol- oder As-treated-Auswertung vorzunehmen, die Art und Ausmaß der tatsächlich durchgeführten Therapie berücksichtigt.

Literatur

[1] Moher D, Schulz KF, Altman DG, for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001; 357: 1191 – 1194

[2] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T, for the CONSORT Group. The revised CONSORT state-

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

ment for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001; 134: 663 – 694
[3] Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine. 2. Aufl. Edinburgh: Churchill Livingstone, 2000
[4] Fisher LD, Dixon DO, Herson J, Frankowski RK, Hearron MS, Peace KE. Intention-to-treat in clinical trials. In: Peace KE (Hrsg). Statistical issues in drug research and development. New York: Marcel Dekker, 1990
[5] Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomized controlled trials. BMJ 1999; 319: 670 – 674
[6] Guyatt GH, Sackett DL, Cock DJ, for the Evidence-Based Medicine Working Group. Users' Guides to the medical literature; II: How to use an article about therapy or prevention; A: Are the results of the study valid? JAMA 1993; 270: 2598 – 2601
[7] Piantadosi S. Clinical trials. A methodological perspective. New York: Wiley, 1997
[8] Altman DG, de Stavola BL, Love SB, Stepniewska KA. Review of survival analyses published in cancer journals. British Journal of Cancer 1995; 72: 511 – 518
[9] Ferguson D, Aaron SD, Guyatt G, Hébert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. BMJ 2002; 325: 652 – 654

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.