

Das Intention-to-Treat-Prinzip

Peter Kleist

GlaxoSmithKline AG, Münchenbuchsee

Quintessenz

- Klinische Studien sind primär gemäss Intention-to-Treat-Prinzip (ITT) auszuwerten. Die Auswertung muss alle randomisierten Patienten unter Beibehaltung der randomisierten Gruppenzugehörigkeit und unabhängig vom Studienverlauf umfassen.
- ITT minimiert nicht den Bias, der sich durch fehlende Daten zu den Studienendpunkten ergibt. Das maximale Follow-up aller Patienten und das bestmögliche Vermeiden fehlender Daten sind wesentliche Faktoren für die Qualität von ITT-Auswertungen. Statistische Verfahren zum Umgang mit fehlenden Daten können ihrerseits mit Problemen verbunden sein.
- Trotz aller Limitationen stellt ITT das beste verfügbare Verfahren zur Vermeidung eines Auswertungsbias und zur Abschätzung von Therapieeffekten in klinischen Studien dar.

Das Problem

Da sich das Leben nicht an Lehrbücher hält, läuft auch keine klinische Studie perfekt ab. Bereits zu ihrem Beginn kann es zu einer Verletzung der vom Protokoll vorgegebenen Selektionskriterien und dadurch zu einem ungerechtfertigten Einschluss von Patienten kommen. Und da Patienten nun mal ihren eigenen Gesetzen Folge leisten, nehmen sie ihre Studienmedikation oftmals unregelmässig bzw. gar nicht ein, brechen möglicherweise die Therapie vorzeitig ab, ziehen in eine andere Stadt oder versterben, bevor die Studie zu Ende ist.

Wie geht man nun damit um? Soll man diese Patienten von der abschliessenden Auswertung ausschliessen und die Daten einfach ignorieren? Welches Vorgehen ist adäquat?

In der Wissenschaft und bei Arzneimittelbehörden hat sich das Intention-to-Treat-Prinzip (ITT) inzwischen als Goldstandard zur primären, das heisst zur massgeblichen statistischen Analyse von klinischen Studien durchgesetzt [1, 2]. Was ITT ist und welche Rolle es für randomisierte kontrollierte Studien spielt, wird nachfolgend erläutert.



Peter Kleist

Der Autor erklärt, dass er keine Interessenskonflikte im Zusammenhang mit diesem Beitrag hat.

Definition und Abgrenzung

Das Intention-to-Treat-Prinzip besagt, dass *alle* randomisierten Patienten ausgewertet werden müssen – und zwar unter Beibehaltung der randomisierten Gruppenzuteilung und unabhängig davon, was mit ihnen im Ver-

lauf der Studie geschieht. Der Begriff «Intention-to-Treat» stammt von der ursprünglichen Absicht ab, nach der man die Patienten zum Zeitpunkt der Randomisierung behandeln wollte. Falsch eingeschlossene Patienten, Therapie- oder Studienabbrecher gehen also in die Analyse ein. Unsere englischen Kollegen drücken diesen Sachverhalt sehr prägnant aus: «Once randomised, always analysed».

Eine ITT-Auswertung kann allerdings nur so gut sein, wie die Informationen zu den Patienten vollständig sind. ITT fängt daher nicht erst bei der Datenauswertung an, sondern umfasst eine gesamte Studie. Bereits die Studienprotokolle sind so zu gestalten, dass fehlende Daten die Ausnahme und nicht die Regel darstellen. Nach Möglichkeit sind alle Patienten bis zum geplanten Studienende weiter zu beobachten, selbst dann, wenn die Behandlung zum Beispiel aufgrund unerwünschter Wirkungen vorzeitig abgebrochen werden muss. Die Minimierung der Ausfallrate von Patienten (Loss-to-Follow-up) ist oberstes Gebot [3]. Nur die vollständige Datenerfassung ermöglicht letztendlich die unverzerrte Auswertung der Wirksamkeit und Sicherheit einer therapeutischen Massnahme. Da ITT also ein umfassendes Konzept ist, das die gesamte Durchführung einer Studie betrifft und nicht erst ihre Auswertung am Ende, sprechen wir von ITT als einem Prinzip.

Im Gegensatz zu ITT berücksichtigt eine «Per-Protokoll-Auswertung» nur die Studienpatienten, bei denen keine

Das Intention-to-Treat-Prinzip besagt, dass *alle* randomisierten Patienten ausgewertet werden müssen – «Once randomised, always analysed»

wesentlichen Abweichungen vom Protokoll vorliegen. Protokollverletzungen, wie sie bereits genannt wurden, würden hier zum Ausschluss der Patienten von der statistischen Analyse führen. Es wird evident, dass die Nichtberücksichtigung

von Patienten und deren Daten aus einer geplanten ITT-Auswertung eine «Pseudo-ITT-Auswertung» macht, die de facto einer Per-Protokoll-Auswertung nahekommt [4].

Ein dritter Auswertungsansatz ist die «As-Treated-Analyse» (so wie tatsächlich behandelt wurde); sie spielt dann eine Rolle, wenn ein signifikanter Patientenanteil nicht die randomisiert zugewiesene Therapie, sondern die Alternativtherapie erhält. As-Treated-Analysen werden vor allem bei Vergleichen zwischen medikamentösen und operativen Verfahren angewendet. Anders als beim ITT- und Per-Protokoll-Ansatz setzt man sich dabei über die ursprünglich vorgenommene Randomisierung hinweg.

Warum ITT?

Die folgenden Gründe sprechen für das ITT-Prinzip als primäres Verfahren zur Auswertung von klinischen Studien:

- Die Vergleichbarkeit der Patientengruppen bleibt erhalten, da alle randomisierten Patienten berücksichtigt werden.
- Auch die Teststärke der Studie (Power) bleibt erhalten, da alle Patienten der statistischen Auswertung zugeführt werden.
- Das Risiko für einen Bias und für ein falsch positives Studienergebnis wird reduziert (sogenannter Typ-1-Fehler). Das wird nachfolgend näher ausgeführt: Protokollabweichungen seitens der Patienten sind oftmals nicht zufällig, sondern hängen mit der Behandlung selbst und dem Behandlungserfolg zusammen. Generell lässt sich sagen, dass Patienten, die einem Protokoll Folge leisten und eine Studie regulär abschliessen, bessere Behandlungsergebnisse aufweisen als solche, die es nicht tun. Das trifft auch auf Patienten in der Placebogruppe einer Studie zu [5]. Die Gründe für eine niedrige Compliance bzw. einen Therapieabbruch können darüber hinaus sehr unterschiedlich sein, zum Beispiel: (1.) eine deutliche Besserung der Symptome und die fehlende Einsicht für die Weiterbehandlung, (2.) die Beendigung einer anscheinend wirkungslosen Behandlung oder (3.) ein Behandlungsabbruch wegen nichttolerierter Nebenwirkungen. Da sich die genannten Gründe in aller Regel nicht identisch auf die verschiedenen Gruppen einer Studie verteilen, da sie unmittelbar von der Wirksamkeit und Verträglichkeit der jeweils untersuchten Behandlungen abhängig sind, würde eine Auswertung unter Ausschluss von Therapieabbrüchern einem Bias unterliegen – in vielen Fällen zum Vorteil der Testbehandlung.

Auch der Ausschluss von Patienten, die nach der Randomisierung die vorgesehene Behandlung nicht erhalten, ist problematisch. Dazu ein Beispiel: Eine Studie vergleicht ein operatives Verfahren mit der konservativen medikamentösen Therapie. Es wird unterstellt, dass beide Therapien absolut gleichwertig sind. Würden aber Patienten, die nach der Randomisierung, aber vor der Operation versterben, ausgeklammert, ergäbe sich ein Vorteil für die Chirurgie, und ein in Wirklichkeit nicht wirksameres Verfahren würde zu Unrecht favorisiert. Denn die scheinbar höhere Effektivität resultiert nicht aus ihrem Nutzen, sondern ist eine Folge des systematischen Ausschlusses von Patienten mit der schlechtesten Prognose aus der operierten Patientengruppe [6].

Eine prinzipiell akzeptierte Ausnahme von dieser Regel, das heisst einen nachträglichen Ausschluss von «falsch randomisierten» Patienten, stellt die modifizierte ITT-Analyse dar.¹

- Eine ITT-Auswertung entspricht am ehesten den Bedingungen des klinischen Alltags. Sie stellt einen konservativen Auswertungsansatz dar, um einer Überinterpretation des klinischen Nutzens vorzubeugen

[7]. Auch in der Praxis nehmen Patienten ihre Medikamente unregelmässig ein oder brechen eine Therapie vorzeitig ab. Die ITT-Auswertung unterstützt daher die Generalisierbarkeit von Studienergebnissen.

Vergleicht man die Ergebnisse von ITT-, Per-Protokoll- und As-Treated-Auswertungen der gleichen Studie (sofern alle drei Ansätze angebracht sind), bestätigt sich die grösste Konservativität von ITT mit den geringsten Unterschieden zwischen den Behandlungsgruppen (Tab. 1 [8]). Dagegen neigen Per-Protokoll- und As-Treated-Auswertungen dazu, Behandlungseffekte zu überschätzen.

Limitationen von ITT

Das ITT-Prinzip ist eine Strategie, um sinnvoll mit den oben genannten Aspekten umzugehen. Aber auch ITT stellt den Zustand der «perfekten Welt» nicht wieder her. Im Folgenden werden Kritikpunkte bzw. Limitationen von ITT benannt und Situationen erwähnt, bei denen alternative oder zusätzliche Auswertungsansätze angebracht sein können.

- Der «wahre» Effekt einer Behandlung wird durch ITT eher unterschätzt. Um diesen bzw. die «reine» biologische Wirkung einer Behandlung abschätzen zu können, ist eine ergänzende Per-Protokoll-Auswertung vorzunehmen.
- Bei Äquivalenzstudien oder Noninferioritätsstudien ist eine Per-Protokoll-Auswertung der konservativere Ansatz. Eine ITT-Auswertung ist hier eher in der Lage, bestehende Unterschiede zwischen den Behandlungen zu verwischen, die Anfälligkeit für einen Fehler erster Art zu erhöhen und somit falsche Schlussfolgerungen in Richtung einer Äquivalenz zu begünstigen [9]. Vieles spricht dafür, eine zusätzliche Per-Protokoll-Auswertung durchzuführen. Resultate und Interpretation einer Äquivalenz- bzw. Noninferioritätsstudie sind umso vertrauenswürdiger, wenn die Ergebnisse von ITT- und Per-Protokoll-Auswertung vergleichbar sind [10].

1 Ein nachträglicher Ausschluss von «falsch randomisierten» Patienten von der statistischen Analyse ist nur dann akzeptabel, wenn zum Zeitpunkt der Randomisierung noch nicht definitiv feststehen kann, ob tatsächlich alle Patienteneinschlusskriterien erfüllt sind. Ein Beispiel hierfür wären Antibiotika-Vergleichsstudien an Patienten mit schweren Infektionen, bei denen klinisch die Notwendigkeit einer raschen Therapieeinleitung besteht – auch ohne Kenntnis des Ergebnisses einer Blutkultur. So könnte man in einer Studie bei Patienten mit Wundinfektionen von einer grampositiven Infektion ausgehen (Einschlusskriterium) und eine Therapie mit entsprechenden Antibiotika einleiten. Nachträglich würde sich doch bei einem kleinen Teil der Patienten herausstellen, dass eine gramnegative oder Mischinfektion vorliegt. Es ist prinzipiell möglich, solche Patienten aus der Auswertung auszuklammern. Man spricht in diesem Zusammenhang von einer modifizierten ITT-Analyse, in der nur Patienten ausgewertet werden, die den vorgegebenen Einschlusskriterien tatsächlich entsprechen. Es ist dennoch sinnvoll, zusätzlich eine ITT-Auswertung mit allen Patienten vorzunehmen, da sie der klinischen Situation am nächsten kommt. Auch Arzneimittelbehörden fordern in diesen Fällen eine begleitende ITT-Analyse unter Berücksichtigung aller randomisierten Patienten.

Eine ITT-Auswertung entspricht am ehesten den Bedingungen des klinischen Alltags

- Zur Auswertung der Häufigkeit und des Schweregrads von Nebenwirkungen ist das As-Treated-Verfahren zu bevorzugen.
- Bei hoher Cross-over-Rate der Behandlungen ist eine zusätzliche As-Treated-Analyse sinnvoll.
- Das grösste Problem in Verbindung mit ITT-Auswertungen ist jedoch folgende Tatsache: ITT kann grundsätzlich den Bias nicht minimieren, der sich durch fehlende Daten zu den Studienendpunkten ergibt. Ist die Loss-to-Follow-up-Rate substantiell, das heisst, es liegen bei vielen Patienten keine Informationen zum Studienausgang vor, ist der Wert einer Studie erheblich eingeschränkt [4]. Unvollständigen Daten lässt sich nur mit Methoden begegnen, die ihrerseits ebenfalls limitiert sind und prinzipiell einem Bias unterliegen (siehe nachfolgenden Abschnitt). Dieser Aspekt unterstreicht erneut die Forderung, Daten so vollständig wie möglich zu erheben.

Der Vorteil aller Verfahren zum Umgang mit fehlenden Werten besteht darin, dass sie besser sind als der Ausschluss einzelner Patienten von der statistischen Analyse

zugeteilt wurden, hauptsächlich zu Beginn wegen starker Nebenwirkungen aus; es sind gleichzeitig diejenigen Patienten mit dem stärksten Ansprechen auf die Behandlung. Dagegen sind Studienabbrecher in der Placebogruppe vornehmlich solche Patienten, bei denen sich der klinische Zustand verschlechtert. Es liegt auf der Hand, dass die Wirksamkeit der untersuchten Behandlung durch die Anwendung der LOCF-Methode deutlich überschätzt wird. Erstens sind die Drop-outs in den Gruppen nicht vergleichbar; zweitens ist es unrealistisch, dass die guten Werte von den Patienten in der Interventionsgruppe auch nach Therapieabbruch konstant bleiben. Rangbasierte Verfahren kommen im Rahmen von Studien zum Einsatz, bei denen Informationen zu sogenannten di-

chotomen Variablen fehlen (z.B. überlebt oder gestorben). Häufig werden Best-and-Worst-Case-Analysen (Sensitivitätsanalysen) vorgenommen, das heisst Effektabschätzungen auf der Basis von theoretisch möglichen extremen Resultaten. Je höher die Rate an Drop-outs mit ungewissem Studienausgang ist, umso grösser ist jedoch die Unsicherheit eines auf diese Weise abgeschätzten Ergebnisses. Bei Worst-Case-Betrachtungen wird der tatsächliche Effekt in einem unbekanntem Ausmass unterschätzt, bei Best-Case-Betrachtungen ist es umgekehrt. Werden Studienergebnisse durch Worst-Case-Betrachtungen neutral oder verkehren sich unter Umständen sogar ins Gegenteil, weiss man letztlich nicht, ob die Therapie ineffektiv ist oder das Ergebnis nur eine Konsequenz von Studienabbrechern. Das Beispiel in Tabelle 2 (11) illustriert die mit diesem Verfahren verbundenen Schwierigkeiten.

Der Vorteil aller angewendeten Verfahren zum Umgang mit fehlenden Werten besteht eigentlich nur darin, dass sie immer noch besser sind als der Ausschluss einzelner Patienten von der statistischen Analyse. Nachteile sind das Risiko für einen Bias und die Gefahr von falsch positiven bzw. negativen Ergebnissen, ein hohes Mass von Unsicherheit bezüglich des realen Effekts und der Verlust an Power [3, 12].

Umgang mit fehlenden Daten

Es gibt verschiedene statistische Verfahren zum Umgang mit fehlenden Werten. Grundsätzlich lassen sich numerische Imputationsstrategien (z.B. die «Last-Observation-Carried-Forward»-[LOCF-]Methode) und rangbasierte Verfahren (z.B. extreme Fallbetrachtungen wie die «Best-and-Worst-Case-Analyse») unterscheiden. Nachfolgend soll nur auf die genannten Beispiele kurz eingegangen werden.

Bei der LOCF-Methode wird der zuletzt erhobene Post-Randomisierungswert bis zum Ende der Studie fortgeschrieben. Diese Methode bezieht also die zuvor im Verlauf der Studie gewonnene Information mit ein. Sie geht jedoch von der unrealistischen Annahme aus, dass die Werte bei einem Patienten konstant bleiben. Ein Beispiel soll zeigen, dass das LOCF-Verfahren zusätzlich einem erheblichen Bias unterliegen kann und das Risiko für falsch positive Befunde erhöht: In einer hypothetischen Studie scheiden Patienten, die der Interventionsgruppe

Tabelle 1.

Gemäss Randomisierung	Zugewiesene bzw. tatsächlich vorgenommene Intervention				Unterschied in der Mortalität (medikamentös vs. operativ)
	Medikamentös	Chirurgisch			
Tatsächlich durchgeführt	medikamentös (a)	chirurgisch (b)	chirurgisch (c)	medikamentös (d)	
Anzahl Überlebende	296	48	353	20	
Anzahl Verstorbene	27	2	15	6	
Mortalität	8,4%	4,0%	4,1%	23,1%	
ITT (a+b) vs. (c+d)	7,8% (29/373)		5,3% (21/394)		2,4%
Per Protokoll (a) vs. (c)	8,4% (27/323)		4,1% (15/368)		4,3%
As Treated (a+d) vs. (b+c)	9,5% (33/349)		4,1% (17/418)		5,4%

Vergleich zwischen Bypass-Chirurgie und medikamentöser Therapie bei Männern mit stabiler Angina pectoris; überlebende bzw. verstorbene Patienten 2 Jahre nach Randomisierung (nach [8]). Am höchsten ist die Mortalität von Patienten, die gemäss Randomisierung der chirurgischen Therapie zugewiesen wurden, letztendlich aber nicht operiert werden konnten und eine medikamentöse Behandlung erhielten (23,1%). Die ITT-Analyse, die diese Patienten der Chirurgiegruppe zuordnet, zeigt den geringsten Vorteil der Chirurgie gegenüber dem konservativen (medikamentösen) Vorgehen. In der Per-Protokoll-Analyse werden diese Patienten wegen Protokollverletzung nicht berücksichtigt. Die As-Treated-Analyse zeigt den deutlichsten Unterschied zugunsten der Chirurgie, da sie die erwähnten Patienten der medikamentös behandelten Gruppe zuordnet.

Tabelle 2. Vergleich zwischen Ballon-Angioplastie und Stent-Behandlung bei 110 Patienten mit verschlossenen koronaren Bypass-Gefässen (nach [11]).

	Angioplastie (n = 110)	Stent (n = 110)	Unterschied in der Restenoserate (aus Sicht der Stenttherapie) und 95%-Konfidenz-Intervalle
Restenose nach 6 Monaten			
Ja	37	32	
Nein	43	54	
Unbekannt	30	24	
Restenoserate in %			
Patienten mit bekanntem Status	46% (37/80)	37% (32/86)	-9% (-24 bis +6)
Best-Case-Analyse für Stents	61% (67/110)	29% (32/110)	-32% (-44 bis -19)
Worst-Case-Analyse für Stents	34% (37/110)	51% (56/110)	+17% (+4 bis +30)

Bei 30 Patienten, d.h. fast einem Drittel der Patienten, ist der Gefässstatus nach 6 Monaten nicht bekannt. Die Aussagekraft einer Analyse von Patienten mit bekanntem Status, aber auch die von Best- and Worst-Case-Analysen sind entsprechend limitiert. Bei Patienten mit bekanntem Gefässstatus zeigt die Stent-Behandlung eine 9-prozentige Restenosereduktion gegenüber der alleinigen Gefässaufdehnung. Nach Vornahme einer Worst-Case-Analyse für Stents verkehrt sich dieser Unterschied ins Gegenteil. Der «wahre» Effekt kann praktisch nicht abgeschätzt werden.

Schlussfolgerungen

1. Für Ärztinnen und Ärzte, die in der Planung und Durchführung von Studien involviert sind:

Patienten sind vor der Randomisierung sorgfältig zu screenen. Die Studie ist so anzulegen, dass eine hohe Compliance mit dem Protokoll gewährleistet werden kann und fehlende Daten vermieden werden – beispielsweise durch Vermeidung zu häufiger Visiten oder einer unnötigen Diagnostik. In Bezug auf den Studienendpunkt ist ein maximales Follow-up der Patienten bis zum geplanten Studienende oder einem vorzeitigen fatalen Ereignis anzustreben – auch dann, wenn eine Weiterbehandlung der Patienten, z.B. aufgrund von Unverträglichkeiten, nicht möglich ist. Das Studienprotokoll sollte bereits festhalten, in welchem Umfang mit fehlenden Daten gerechnet wird und welche statistischen Verfahren in diesem Zusammenhang zur Anwendung kommen [13].

Empfohlene Literatur

- Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med.* 2001;134:663-94.
- Fergusson D, Aaron SD, Gyatt G, Hébert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ.* 2002;325:652-4.

2. Für die Leser von Publikationen:

Es sind die richtigen Fragen zu stellen, um den Wert einer Studie beurteilen zu können. Wurde eine ITT-Auswertung an allen randomisierten Patienten vorgenommen? Wie hoch ist die Drop-out-Rate? Wie wurde mit Studienabbruchern und fehlenden Daten im Rahmen der statistischen Auswertung umgegangen?

Fehlen für einen wesentlichen Teil der Patienten Angaben zum Studienendpunkt, ist der Aussage, dass eine ITT-Analyse durchgeführt wurde, und den Schlussfolgerungen der Studie mit Skepsis zu begegnen.

Korrespondenz:

Dr. med. Peter Kleist
Glaxo Smith Kline AG
Talstrasse 3-5
CH-3053 Münchenbuchsee
peter.m.kleist@gsk.ch

- Gotzsche PC. Lessons from and cautions about noninferiority and equivalence randomised trials. *JAMA.* 2006;295:1172-4.

Die vollständige nummerierte Literaturliste finden Sie unter www.medicalforum.ch.

Weiterführende Literatur (Online Version)

- 1 Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med.* 2001;134:663–94.
- 2 International Conference on Harmonisation (ICH). Guideline E9: Statistical Principles for Clinical Trials. www.ich.org/LOB/media/MEDIA485.pdf.
- 3 Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ.* 1999;319:670–4.
- 4 Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ.* 2001;165:1339–41.
- 5 Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med.* 1980;303:1038–41.
- 6 Bhandari M, Gyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature: how to use an article about a surgical therapy. *J Bone Joint Surg Am.* 2001;83:916–26.
- 7 Fergusson D, Aaron SD, Gyatt G, Hébert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ.* 2002;325:652–4.
- 8 European Coronary Surgery Study Group. Coronary-artery bypass surgery in stable angina pectoris: Survival at two years. *Lancet.* 1979;i:889–93.
- 9 Gotzsche PC. Lessons from and cautions about noninferiority and equivalence randomised trials. *JAMA.* 2006;295:1172–4.
- 10 Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ.* 1996;313:36–9.
- 11 Savage MP, Douglas JS Jr, Fischman DL, et al. Stent placement compared with balloon angioplasty for obstructed coronary bypass grafts. *N Engl J Med.* 1997;337:740–7.
- 12 Streiner D, Geddes J. Intention to treat analysis in clinical trials when there are missing data. *Evid based Ment Health.* 2001;4:70–1.
- 13 Committee for Proprietary Medicinal Products (CPMP). Points to consider on missing data. London, 15 November 2001. www.emea.europa.eu/pdfs/human/ewp/177699EN.pdf.