

«To stop or not to stop» – das ist (und bleibt) die Frage

Peter Kleist

GlaxoSmithKline AG, Münchenbuchsee

Summary

“To stop or not to stop” – that is (still) the question

At first glance there are compelling reasons for stopping a clinical trial early on the grounds of harm, futility or treatment benefit. This article shows, however, that the decision to stop or go on with a trial on the basis of interim results may be in either case highly complex and individual. Medical, ethical, statistical and social considerations play a role, and there is occasional conflict between the obligation to minimise the risk and maximise the benefit for both trial participants and future patient populations. The ethical and scientific issue is to protect patients in a trial while also protecting society from ineffective and unsafe treatments or, in other words, to save the social value of a trial. Since a trial should remain conclusive on the true therapeutic effects, stopping it early is not always the best alternative.

Frühzeitige Studienabbrüche sind Mode geworden. Die Logik eines Studienabbruchs erscheint auf den ersten Blick zwingend, wenn sich eine Therapie vorzeitig als wirksam oder als gefährlich herausstellt. Darüber hinaus spricht auch vieles dafür, eine Studie dann nicht weiterzuführen, wenn sich die Wirksamkeit einer untersuchten Behandlung als ungenügend herausstellt (engl. *futility*). In allen genannten Fällen reduziert sich der Studienumfang (kürzere Beobachtungszeiten, eventuell auch niedrigere Patientenzahlen), die Ergebnisse sind früher verfügbar und bereits limitierte finanzielle und personelle Ressourcen stehen für andere Fragestellungen zur Verfügung. In behördlichen Richtlinien sind die formalen Anforderungen an Zwischenauswertungen und Abbruchentscheidungen innerhalb klinischer Studien festgelegt. Die Zulassungsbehörden akzeptieren heute Studienabbrüche aufgrund eines vorzeitigen Wirksamkeitsnachweises, insofern die statistischen Methoden vor Studienbeginn spezifiziert wurden und ausreichend konservativ sind. Gemäss Artikel 17 der revidierten Deklaration von Helsinki werden Studienabbrüche sogar zu einem ethischen Mandat, wenn die Risiken den potentiellen Nutzen übertreffen oder der hinreichende Nachweis positiver und nützlicher Resultate erbracht ist.

Was auf den ersten Blick so einfach und selbstverständlich erscheint, erweist sich auf den zweiten als komplex – selbst unter der Voraussetzung, dass ein von Sponsor- und Prüfarztinteressen unabhängiges Data Monitoring Committee (DMC) involviert ist. Bei der mitunter schwierigen Entscheidung, ob eine Studie vorzeitig beendet wer-

den soll oder nicht, spielen – zum Teil kontroverse – medizinische, statistische, ethische und soziale Überlegungen eine Rolle. Der folgende Beitrag soll aufzeigen, dass ein Studienabbruch nicht immer die einzige oder sogar beste Alternative darstellt.

Data Monitoring Committees und Zwischenauswertungen

Bei grossen multizentrischen Langzeitstudien ist die Implementierung eines DMC inzwischen als Standard zu betrachten [1, 2]. Wird der frühzeitige Abbruch einer Studie mit geplanten Zwischenauswertungen – am häufigsten in Studien mit einem gruppensequentiellen Design – grundsätzlich in Betracht gezogen, ist die Einrichtung eines DMC obligatorisch, um die an der Studie Beteiligten nicht zu entblinden und unabhängige Empfehlungen zum weiteren Vorgehen zu ermöglichen. Die Kriterien für einen Studienabbruch (sogenannte «stopping rules») aufgrund von mutmasslicher Therapieüberlegenheit, fehlender Wirksamkeit oder eingeschränkter Sicherheit sind im voraus und im Detail festzulegen.

Werden Zwischenauswertungen vorgenommen, ist das statistische Signifikanzniveau der Studie in bezug auf die wiederholten Tests anzupassen. Für die unterschiedlichen Abbruchentscheidungen kann das Signifikanzniveau asymmetrisch sein: Während ein möglicher Studienabbruch aufgrund inakzeptabler Nebenwirkungen einer Behandlung (z.B. eines Anstiegs der Mortalität) weniger stringente Kriterien erfordert, muss einem Studienabbruch wegen Überlegenheit eine über jeden Zweifel erhabene Beweiskraft zukommen [3]. Letzteres setzt ein sehr striktes Abbruchkriterium mit einer geringen Wahrscheinlichkeit für ein falsch positives Ergebnis voraus. Eine rasch erfolgende Marktzulassung und die potentiell grosse Aufmerksamkeit, die ein solches Ergebnis – vor allem durch Publikationen im «New England Journal of Medicine» oder in «Lancet» – in der Fachwelt erfährt, können zu einem weit verbreiteten Einsatz einer Therapie in der entsprechenden Indikation führen.

Eine auf Basis der Ergebnisse einer Zwischenauswertung abgebrochene Studie ist auf jeden Fall abschliessend auszuwerten, da zwischen dem Stichtag der für die Zwischenauswertung berücksichtigten Daten und dem Zeitpunkt der Entscheidung, die Studie abzubrechen, weitere

Patienten eingeschlossen werden konnten und sich die Beobachtungszeit für alle Patienten verlängert hat.

Generelle Betrachtungen

Die ethische Verpflichtung eines DMC ist nicht auf den Schutz der Versuchsteilnehmer beschränkt – indem eine Studie nicht länger als notwendig durchgeführt wird –, sondern umfasst auch den Schutz zukünftiger Patienten vor ineffektiven oder unsicheren Therapien. Letzteres ist jedoch nur gewährleistet, wenn die wissenschaftliche Validität einer Studie gewahrt bleibt und eine aussagekräftige Nutzen/Risiko-Beurteilung zulässt. Man spricht in diesem Zusammenhang auch vom sozialen Wert einer klinischen Studie. Die Zielsetzungen der «Risikominimierung» und «Nutzenmaximierung» für beide Patientengruppen, d.h. diejenige *in* der Studie und diejenige *nach* der Studie, geraten jedoch mitunter in einen Konflikt [4].

Wird eine Studie beispielsweise wegen der Überlegenheit einer Behandlung frühzeitig beendet, so darf nicht darüber hinweggesehen werden, dass die damit verbundene kürzere Beobachtungsdauer und gegebenenfalls niedrigere Patientenzahl die Interpretation der Sicherheit der Behandlung grundsätzlich erschwert. Das mit der Gabe von COX-2-Hemmern assoziierte kardiovaskuläre Risiko wurde beispielsweise erst in Studien mit ausreichend langer Beobachtungszeit evident. Und – um ein weiteres, in diesem Zusammenhang konkretes Beispiel zu nennen: Eine Studie zu Vitamin E bei Frühgeborenen, die aufgrund der Reduktion von Hirnblutungen frühzeitig beendet wurde, war nicht in der Lage, ein mit der Therapie verbundenes Sepsisrisiko aufzudecken [5]. Erst weitere Studien konnten dieses Risiko identifizieren.

Ein DMC muss darüber hinaus den gegenwärtigen Status einer untersuchten Behandlung berücksichtigen. Handelt es sich um ein innovatives, noch nicht zugelassenes Arzneimittel, fällt die Entscheidung für einen Studienabbruch leichter, wenn sich ein negativer Trend abzeichnet, der einen klinisch relevanten Nutzen mit hoher Wahrscheinlichkeit ausschliesst. Befindet sich dagegen eine bekannte Therapie bereits in grosszügigem Einsatz, kommt der Unterscheidung zwischen einer neutralen oder sogar schädlichen Wirkung eine entscheidende Bedeutung zu – was eher gegen einen frühen Studienabbruch spricht [6]. Das ist insbesondere dann der Fall, wenn die Therapie andere Vorteile offeriert, zum Beispiel eine effektive Symptomkontrolle oder Lebensqualitätsverbesserung bei Patienten mit Herzinsuffizienz (siehe Beispiel im nächsten Abschnitt).

Ist der Einsatz einer Therapie bereits etabliert, muss einer Studie mit einem unerwarteten und früheren Studien widersprechenden Ergebnis eine hohe Überzeugungskraft zukommen. Diese

ist mit einer vorzeitig beendeten Studie mitunter nicht gegeben. Man denke hier nur an die Diskussion zu den ersten Untersuchungen, die eine postmenopausale Hormongabe mit einer erhöhten kardiovaskulären Ereignisrate in Verbindung brachten; einer zu früh abgebrochenen HERS- (Heart and Estrogen/Progestin Replacement Study) oder WHI-Studie (Women's Health Initiative Trial) hätten viele Ärzte aller Wahrscheinlichkeit nach keinen Glauben geschenkt. Und zu einer Zeit, in der sich die Gabe von Antiarrhythmika nach einem Herzinfarkt grosser Beliebtheit erfreute, hätte eine abgebrochene CAST-Studie (Cardiac Arrhythmia Suppression Trial), deren Ergebnis lediglich im fehlenden Nutzen der Behandlung bestanden hätte, ebenfalls nicht genügt. In diesem Fall bedurfte es des abgesicherten Nachweises der Mortalitätserhöhung durch die Antiarrhythmikatherapie.

Spezifische Aspekte

Die plazebokontrollierte ISIS-1-Studie (First International Study of Infarct Survival; [7]), die den Nachweis eines verbesserten Herzinfarktüberlebens durch den Betablocker Atenolol erbrachte, liefert ein Beispiel für die bereits mit dem fortlaufenden Sicherheitsmonitoring in einer Studie verbundenen Schwierigkeiten. In der Frühphase der Studie lag ein signifikanter Trend für eine Mortalitätserhöhung durch den Betablocker vor; das DMC entschied sich jedoch aufgrund der Instabilität dieses Trends und der bisher geringen Anzahl eingeschlossener Patienten für eine Fortführung der Studie. Am Ende der Studie, nach Auswertung von 16 000 Patienten, zeigte sich eine klinisch relevante Reduktion der Infarktsterblichkeit durch Atenolol um 15%. Auch wenn solche Beispiele selten sind, zeigen sie doch, dass auf der Basis negativer Zwischenresultate nicht überreagiert werden darf. Zu frühe Abbruchentscheidungen können dazu führen, dass verheissungsvolle Therapien nicht weiter untersucht werden.

Steht im Rahmen einer Studie die Abwägung bekannter bzw. möglicher weiterer Vorteile einer Therapie gegenüber einem eventuell erhöhten Mortalitätsrisiko im Vordergrund, ist eine Abbruchentscheidung aufgrund einer fehlenden Überlebenszeitverlängerung (Abbruch wegen «futility») wenig sinnvoll. Ein Beispiel soll dies erläutern: In der PROMISE-Studie (Prospective Randomized Milrinone Survival Evaluation trial [8]) wurde der Effekt von Milrinone auf die Mortalität bei Patienten mit fortgeschrittener Herzinsuffizienz untersucht. Die Substanz verbessert die Leistungsfähigkeit herzinsuffizienter Patienten markant, und Patienten würden eventuell sogar ein leicht erhöhtes Sterberisiko in Kauf nehmen. Die zentrale Frage der Studie bestand also darin, ob sich die Substanz in bezug auf die Mortalität

neutral verhält oder einen negativen Einfluss ausübt. Daher wurde die Studie erst zu einem Zeitpunkt beendet, als sich das Mortalitätsrisiko eindeutig und statistisch signifikant abzeichnete. Grundsätzlich gilt: Eine frühe Studienbeendigung aufgrund des fehlenden Nutzens der untersuchten Therapie ist aus ethischen Erwägungen nicht zwingend, solange mit der Studie keine Gefährdung der Patienten verbunden ist. Im Gegenteil – ein Studienabbruch kann sogar Sicherheitsprobleme verdecken und öffnet daher Türen zu einem potentiellen Missbrauch – indem eine Studie abgebrochen wird, bevor ein Sicherheitsrisiko evident ist [9].

Das hauptsächliche Problem von Studienabbrüchen aufgrund der Überlegenheit einer Behandlung besteht in der Überschätzung von Therapieeffekten. Vor zwei Jahren konnte eine systematische Untersuchung zeigen, dass ein relevanter Anteil von initialen «enthusiastischen» und anschließend vielzitierten Studien von nachfolgenden Studien mit nur moderaten oder zum Teil entgegengesetzten Resultaten relativiert wurde [10]. Das Risiko einer Wirksamkeitsüberschätzung wird auf Basis der Ergebnisse von Zwischenauswertungen nochmals erhöht, da die Wahrscheinlichkeit von Zufallsergebnissen zunimmt. Zwar ist die Präzision für die Abschätzung der «wahren» Behandlungseffekte generell ziemlich hoch, insofern einem Studienabbruch geeignete, konservative statistische Methoden zu Grunde liegen; dennoch unterliegt die Abschätzung des Therapieeffektes potentiell einem Bias: Würde die gleiche Studie nämlich vielfach wiederholt, dann führten zufällige Trends in Richtung eines grösseren Therapie Nutzens eher zu einer vorzeitigen Studienbeendigung («at random high») als Fluktuationen in Richtung geringerer Effekte [11]. Man muss sich daher der prinzipiellen Gefahr einer Überschätzung der Behandlungseffekte bewusst sein.

Diese Überschätzungen sind nicht nur Theorie, sondern treten in der Praxis tatsächlich auf. Das wird durch solche Studien evident, die man wegen statistisch signifikant überlegener Wirksamkeit der untersuchten Therapie hätte abbrechen können, die dann aber dennoch bis zum vorgesehenen Ende weitergeführt wurden. Nach planmässigem Abschluss dieser Studien waren die Resultate neutral oder verkehrten sich sogar ins Gegenteil. Hierzu ein Beispiel: In der plazebokontrollierten OPTIMIST-Studie (Optimized Phase 3 Tifacogin in Multicenter International Sepsis Trial [12]) bei Patienten mit schwerer Sepsis betrug die absolute Verbesserung der 4-Wochen-Überlebensrate durch den untersuchten Gewebefaktor-Hemmstoff Tifacogin zwischenzeitig 10%. Aufgrund des in diesem Ausmass unerwarteten Effekts empfahl das DMC die Studie fortzusetzen. In der Endauswertung zeigte sich bezüglich der Mortalität kein Unterschied, dafür wiesen die Patienten unter aktiver Therapie ein höheres Blutungsrisiko auf. Weitere eindruckliche Beispiele finden sich

in einer früheren Übersicht [13]. Generell sind Zweifel an positiven Zwischenergebnissen angebracht, wenn diese zu stark ausgeprägt sind, um plausibel zu sein, die Studie insgesamt (zu) klein und die beobachtete Ereignisrate generell sehr niedrig ist; alle genannten Faktoren leisten einem zufälligen, falsch positiven Ergebnis Vorschub [14, 15].

Darüber hinaus kann ein Bias dadurch zustande kommen, dass klinische Ereignisse nicht konstant während des Beobachtungszeitraums auftreten. Das CHARM-Studienprogramm (Candesartan in Heart failure: Assessment of Reduction in Mortality and Morbidity), in welchem der Angiotensin-II-Blocker Candesartan bei herzinsuffizienten Patienten in drei separaten randomisierten Studien mit Plazebo verglichen wurde, liefert hierfür ein Beispiel [16]. Das Abbruchkriterium für das gesamte Programm wurde mit der vierten Zwischenauswertung erreicht, als sich unter Candesartan eine 24prozentige Mortalitätsreduktion gegenüber Plazebo zeigte. Da die Mortalitätsabnahme in zwei der drei Studien nicht signifikant und die durchschnittliche Beobachtungszeit relativ kurz war, empfahl das DMC, die Studien fortzusetzen. In der finalen Auswertung des Studienprogramms schrumpfte diese Differenz auf nicht mehr signifikante 9%. Nachfolgende Auswertungen zur Erklärung dieser Entwicklung konnten demonstrieren, dass die Reduktion kardiovaskulärer Todesfälle vornehmlich auf das erste Behandlungsjahr beschränkt war.

Die geschilderten Beispiele zeigen, dass Studien nur dann wegen (anscheinender) Wirksamkeit bzw. Überlegenheit der untersuchten Behandlung abgebrochen werden sollten, wenn eine «überwältigende» Evidenz für einen Behandlungseffekt vorliegt. In den genannten Fällen war die jeweilige Entscheidung des DMC, die Studien weiterzuführen, absolut richtig. Im Fall von Tifacogin wurden zukünftige Patienten vor der Einnahme einer wirkungslosen Substanz bewahrt und im Fall von Candesartan wurde ein ungerechtfertigter Anspruch auf Verlängerung der Überlebenszeit verhindert.

Abschliessende Bemerkungen

Die Entscheidung, eine klinische Studie auf Basis einer Zwischenauswertung abzubrechen oder fortzuführen, ist eine Einzelfallentscheidung, die eine Vielzahl von Faktoren berücksichtigen muss. Klar formulierte Abbruchkriterien sind notwendig, aber die Entscheidung in die eine oder andere Richtung ist letztlich keine statistische, sondern eine wissenschaftliche und ethische. Das Dilemma besteht darin, einerseits nicht zu rasch zu reagieren und andererseits die Grenze der ethischen Akzeptanz nicht zu überschreiten. Ohne Zweifel müssen Studienteilnehmer einem besonderen Schutz unterliegen; aber der Schutz zukünftiger Patienten vor wirkungslosen oder sogar

gefährlichen Therapien ist ein genauso wichtiger ethischer Anspruch. Eine Studie vorzeitig zu beenden muss daher nicht die beste Alternative darstellen. Data Monitoring Committees werden sich weiterhin einer schwierigen Herausforderung stellen müssen. Denn die gesamtgesellschaftlichen Fragen zur angemessenen Abwägung des Individual- und Kollektivinteresses sowie der akzeptablen Geschwindigkeit für medizinischen Fortschritt sind zur Zeit unbeantwortet.

Anmerkung

Ein Beitrag des Autors zu dieser Thematik erschien bereits in der Mitgliederzeitschrift der Deutschen Gesellschaft für Pharmazeutische Medizin e.V. DGPharMed (DGPharMedNews Oktober 2007; «Ein kritischer Balanceakt – Der Abbruch klinischer Studien aus medizinischen Gründen»).

Ein ergänzender Kommentar zu vorzeitigem Studienabbrüchen aus nichtmedizinischen Gründen wird in Heft 18 erscheinen.

Literatur

- 1 European Medicines Agency. CHMP Guideline on data monitoring committees. London, 27 July 2005. www.emea.europa.eu/pdfs/human/ewp/587203en.pdf
- 2 Damocles Study Group. A proposed charter for clinical trial data monitoring committees: helping them to do their job well. *Lancet*. 2005;365:711–22.
- 3 Pocock SJ. Current controversies in data monitoring for clinical trials. *Clin Trials*. 2006;3:513–21.
- 4 Slutsky AS, Lavery JV. Drug safety and monitoring boards. *N Engl J Med*. 2004;350:1143–7.
- 5 Fish WH, Cohen M, Franzek D, et al. Effect of intramuscular vitamin E on mortality and intracranial hemorrhage in neonates of 1000 grams or less. *Pediatrics*. 1990;85:578–84.
- 6 DeMets DL, Pocock SJ. The agonising negative trend in monitoring of clinical trials. *Lancet*. 1999;354:1983–8.
- 7 First International Study of Infarct Survival (ISIS) Collaborative Study Group. Randomised trial of intravenous atenolol among 16 027 cases of suspected acute myocardial infarction: ISIS-1. *Lancet*. 1986;ii:57–66.
- 8 Packer M, Carver JR, Rodehoffer RJ, et al., for the PROMISE Study Research Group. Effect of oral milrinone on mortality in severe chronic heart failure. *N Engl J Med*. 1991;325:1468–75.
- 9 Schoenfeld DA, O Meade M. Pro / con clinical debate: it is acceptable to stop large multicentre randomised controlled trials at interim analysis for futility. *Crit Care*. 2005;9:34–6.
- 10 Ioannidis JPA. Contradicted and initial stronger effects in highly cited clinical research. *JAMA* 2005;294:218–28.
- 11 Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. *Lancet*. 2005;365:1591–5.
- 12 Abraham E, Reinhart K, Opal S, et al., for the OPTIMIST Trial Study Group. Efficacy and safety of tifacogin (recombinant tissue factor pathway inhibitor) in severe sepsis. A randomised controlled trial. *JAMA*. 2003;290:238–47.
- 13 Kleist P. Stopping a trial early for treatment benefit: pros & cons. *Applied Clinical Trials* 2007;January:34–7.
- 14 Montori VM, Devereaux PJ, Adhikari NKJ, et al. Randomized trials stopped early for benefit. A systematic review. *JAMA*. 2005;294:2203–9.
- 15 Mueller PS, Montori VM, Bassler D, et al. Ethical issues in stopping randomised trials early because of apparent benefit. *Ann Intern Med*. 2007;146:878–81.
- 16 Pocock S, Wang D, Wilhelmssen L, Hennekens CH. The data monitoring experience in the Candesartan in Heart Failure Assessment of Reduction in Mortality and Morbidity (CHARM) program. *Am Heart J*. 2005;149:939–43.

Korrespondenz:
 Dr. med. Peter Kleist
 GlaxoSmithKline AG
 Talstrasse 3–5
 CH-3053 Münchenbuchsee
peter.p.kleist@gsk.com