Mortalität als Endpunkt klinischer Studien

Peter Kleist
GlaxoSmithKline AG, Münchenbuchsee

Quintessenz

- Um valide Aussagen zur Beeinflussung der Mortalität machen zu können, müssen klinische Interventionsstudien definierten methodologischen Standards entsprechen.
- Die wichtigsten Anforderungen sind: Klare Hypothese und Fragestellung, prospektive Festlegung der Mortalität als primärer Endpunkt, exakte Festlegung der Datenauswertung, adäquate Fallzahlberechnung, Definition des Noninferioritätsbereichs bei Studien auf Nichtunterlegenheit sowie strikte Anwendung des Intention-to-treat-Prinzips.
- Dem Leser werden zudem Hilfestellungen gegeben, die Studienqualität und klinische Relevanz der Ergebnisse zu beurteilen.

Kurze Einführung

In klinischen Studien wird häufig die Frage gestellt, ob eine medizinische Intervention die Sterblichkeitsrate (Mortalität) bzw. die Überlebenszeit von Patienten günstig beeinflusst. Die Mortalität hat bei (potentiell) lebensbedrohenden oder die Lebenszeit verkürzenden Erkrankungen verständlicherweise eine unmittelbare und hohe klinische Relevanz – man spricht daher auch von einem «harten klinischen Endpunkt». Die Mortalität senken zu können ist entsprechend eines der stärksten Attribute, das einem Arzneimittel oder einem Verfahren zukommen kann.
Der nachfolgende Beitrag geht auf einige wesentliche methodologische Aspekte ein, die bei der Durchführung und der Interpretation von Studien zur Untersuchung der Mortalität zu beachten sind.

Peter Kleist

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

Spezifikation als Endpunkt im Studienprotokoll

Um überhaupt Aussagen zu dieser Fragestellung machen zu können, muss die Mortalität ein vorab definiertes Zielkriterium (Endpunkt) der klinischen Studie sein. Im Studienprotokoll (und gegebenenfalls in einem separaten statistischen Analyseplan) ist dies unmissverständlich festzuhalten. Über die Absichtserklärung hinaus gehen die Studienverantwortlichen somit auch die Verpflichtung ein, eine entsprechende Datenauswertung vorzunehmen. Aus dem Abschnitt *Methoden* in der Publikation der Studie sollte für den Leser klar hervorgehen, ob die Beeinflussung der Mortalität als Studienendpunkt im Protokoll definiert war oder nicht. Ergebnisse von ungeplanten Post-hoc-Auswertungen – insbesondere, wenn sie nicht im Einklang mit den Daten zum primären Studienendpunkt sind – können Ausdruck des Zufalls sein und haben somit bestenfalls Hypothesen-generierenden Charakter. Ein Beispiel für die mit unspezifizierten Auswertungen verbundenen Schwierigkeiten sind die frühen Studien zum Betablocker Carvedilol bei Patienten mit Herzinsuffizienz. Unerwartet wurde die Mortalität in den für eine Marktzulassung relevanten Studien statistisch hochsignifikant im Vergleich zu Plazebo gesenkt. Dennoch hat die amerikanische Gesundheitsbehörde FDA Carvedilol zunächst nicht für diese Indikation zulassen wollen, da die Ergebnisse zur Belastungstoleranz (primärer Studienendpunkt) eine Zulassung nicht stützten und die Mortalität überhaupt nicht als Endpunkt spezifiziert worden war [1, 2].
Es reicht darüber hinaus auch nicht aus, nur von «Mortalität» oder «Überleben» als Endpunkt zu sprechen. Was ausgewertet werden soll, muss von vornherein exakt festgelegt werden, um einem nachträglichen «Ausschlachten» der vorliegenden Daten mit der Suche nach statistischen Signifikanzen von Anfang an den Boden zu entziehen.
Wie wichtig die exakte Festlegung der Auswertung ist soll das folgende Beispiel zeigen: In einer Langzeittherapiestudie an COPD-Patienten, in der ein Anticholinergikum mit Plazebo verglichen wurde, war die Mortalität ein sekundärer Endpunkt [3]. Der Analyseplan sah vor, bei der späteren Auswertung alle Todesfälle bis zum Studientag 1470 zu berücksichtigen, also einschliesslich einer 30-tägigen Follow-Up-Periode nach Beendigung der Therapie [4]. Dies ist insofern von Bedeutung, als die entsprechende Datenauswertung – im Gegensatz zu einer Auswertung von Todesfällen unter Behandlung, also bis zum Tag 1440 – keinen statistisch signifikanten Unterschied zwischen aktiver Therapie und Plazebo zeigen konnte.

Im Studienprotokoll ist exakt festzulegen, was ausgewertet werden soll

Statistische Voraussetzungen

Definitive Aussagen zum Einfluss einer medizinischen Intervention auf die Mortalität lassen sich allerdings nur machen, wenn die Mortalität *primärer* Studienendpunkt ist. Die Berechnung der notwendigen Patientenzahl sowie die Power (Teststärke) einer Studie basieren nämlich auf der primären Studienhypothese, die an-

hand des primären Endpunkts untersucht wird. Stellt die Mortalität nur einen nachgeordneten Endpunkt dar, sind die statistischen Voraussetzungen für eine konfirmative, d.h. «beweisende Studie», nicht erfüllt, denn die Fallzahl und die Teststärke sind in der Regel zu niedrig; wie bei Post-hoc-Analysen kann auch hier ein Zufallsergebnis nicht ausgeschlossen werden (Beispiel in [5]).

Eine zweite wesentliche Voraussetzung für konfirmative Studien ist, dass ein prospektiver Vergleich zu einer randomisierten Kontrollgruppe vorgenommen wird. Historische Kontrollen sind nicht verlässlich, wie zum Beispiel offene Studien zur Antiarrhythmikatherapie nach einem Myokardinfarkt in den 1970er und 1980er Jahren gezeigt haben. Die im Vergleich zu historischen Daten beobachtete höhere Überlebenswahrscheinlichkeit war – wie wir heute wissen – Ausdruck der über Jahre verbesserten Standardtherapie des Infarktes, nicht jedoch der Wirksamkeit der untersuchten Substanzen. Erst eine grosse randomisierte Vergleichsstudie Ende der 1980er Jahre (Cardiac Arrhythmia Suppression Trial; CAST [6]) war in der Lage, ihren wahren Stellenwert aufzudecken: Die Mortalität nahm unter den untersuchten Klasse-1C-Antiarrhythmika sogar um das Zwei- bis Dreifache gegenüber Plazebo zu.

Der p-Wert allein sagt nichts über die Bedeutung des Studienergebnisses aus

Wichtige Aspekte für die Fallzahlberechnung sind die mutmassliche Ereignisrate in der Kontrollgruppe und die erwartete Reduktion tödlicher Ereignisse durch die untersuchte Therapie. Obwohl in beiden Fällen die Mortalitätsabnahme relativ betrachtet 50% beträgt, macht es einen grossen Unterschied, ob die Mortalität von 20% (Kontrolle) auf 10% (untersuchte Therapie) gesenkt wird oder von 2% auf 1%. Letzteres setzt eine deutlich höhere Patientenzahl voraus, um ein valides Ergebnis zu erzielen.

Eine Überschätzung der Ereignisrate vor Beginn der klinischen Studie mindert daher per se die Teststärke. Dieses Phänomen ist im Rahmen von mehrjährigen Interventionsstudien, insbesondere bei Patienten mit Herzkreislauf-Erkrankungen, zunehmend häufiger zu beobachten. Beispielsweise erhalten immer mehr Patienten standardmässig eine begleitende Statinbehandlung, welche sich günstig auf die Gesamtprognose von Patienten mit einem kardiovaskulären Risiko auswirkt. So ging man vor Beginn der RECORD-Studie [7] von einer jährlichen Ereignisrate von 11% bei den Typ 2-Diabetikern in der Kontrollgruppe aus (primärer Endpunkt: Kombination von kardiovaskulären Todesfällen und Spitaleinweisungen); tatsächlich betrug diese in der Studie nur 3,5%.

Der antizipierte Unterschied im Behandlungsergebnis muss Ausdruck einer klinisch relevanten Veränderung sein; nur dann bildet die berechnete Fallzahl die Grundlage für eine Übereinstimmung von statistischer Signifikanz und klinischer Relevanz am Ende einer Studie. Geringe Unterschiede zwischen den Behandlungen können bei hoher Fallzahl statistisch signifikant werden, spielen aber möglicherweise klinisch keine Rolle. Umgekehrt können relevante Unterschiede die statistische Signifikanz verfehlen, wenn die Fallzahl zu gering ist.

Wie die erwartete Differenz der Behandlungsergebnisse geht die Festlegung der Power einer Studie in die Fallzahlberechnung mit ein. Eine ausreichend hohe Power (mindestens 80%) ist ein Gütekriterium für das Studienergebnis. Bei einer Studie, die zeigen soll, dass Behandlung A im Vergleich zu Behandlung B die Mortalität senkt, bedeutet eine hohe Power, dass die Studie auch tatsächlich in der Lage ist, einen bestehenden Unterschied zwischen den Behandlungen aufzudecken. Leider weisen immer noch viele publizierte Studien, insbesondere solche mit dem Endpunkt Mortalität, eine unzureichende Teststärke auf [8]. So lassen sich nichtsignifikante Unterschiede zwischen zwei Behandlungen oftmals allein auf die zu geringe Power zurückführen – ob in Wirklichkeit nicht doch ein Unterschied besteht, lässt sich nicht ausschliessen.

Bei der Beurteilung einer Studie sind also die angenommene und tatsächliche Mortalitätsrate in der untersuchten Population, der angenommene und tatsächliche Unterschied zwischen den Behandlungen, die Frage, was eine klinisch relevante Änderung ist, und die Power der Studie einzubeziehen. Der «p-Wert» allein sagt nichts über die Bedeutung eines Studienergebnisses aus.

Überlegenheit oder Noninferiorität?

Neue Therapieoptionen zeichnen sich häufig nur durch eine bessere Verträglichkeit oder eine einfachere Anwendung aus; ein Zusatznutzen bezüglich der Sterblichkeit ist dann nur gering bzw. gar nicht vorhanden. Eine klinische Studie zum Nachweis der Überlegenheit gegenüber der Standardtherapie macht in diesen Fällen keinen Sinn. Es muss aber gezeigt werden, dass die neue Therapie einer bereits verfügbaren nicht unterlegen ist (Noninferiorität), also eine mindestens vergleichbar stark ausgeprägte Abnahme der Mortalität zur Folge hat.

Besteht die Fragestellung einer Studie im Nachweis der Noninferiorität einer Therapie zu einer Vergleichsbehandlung, so muss dies bereits bei der Studienplanung berücksichtigt werden: die Studienhypothese ist entsprechend zu formulieren, die Noninferioritätsgrenzwerte (d.h. die Grenzwerte für einen unter klinischen Gesichtspunkten gerade noch akzeptablen Unterschied) sind festzulegen und die Fallzahl ist unter Berücksichtigung der vorher definierten Noninferioritätsgrenze zu berechnen [9]. Im Vergleich zu Studien auf Überlegenheit einer Therapie gegenüber einer anderen erfordern Noninferioritätsstudien generell betrachtet deutlich höhere Patientenzahlen. Die statistischen Annahmen und das Vorgehen sind vor der Studie im Protokoll und in einer späteren Publikation (Methodenteil) exakt zu beschreiben.

Insbesondere beim Endpunkt Mortalität ist darauf zu achten, dass die tolerierte Abweichung nicht zu grosszügig definiert wird. Um ein Beispiel zu nennen: In der COMPASS-Studie (Saruplase im Vergleich zu Streptokinase bei Patienten mit akutem Myokardinfarkt; [10]) bestand der Noninferioritätsgrenzwert in einer nicht

mehr als 1,5-fach höheren Mortalität unter Saruplase. Bei erwarteten 70 Todesfällen unter 1000 mit Streptokinase behandelten Patienten wäre also eine absolute Zunahme von 35 Todesfällen im Sinne der Noninferiorität von Saruplase noch akzeptabel gewesen. Der Einwand der ethischen Fragwürdigkeit eines solchen Grenzwerts ist sicherlich berechtigt [11].

Auf der anderen Seite erfordern enge Grenzen so hohe Fallzahlen, dass viele Studien, obwohl sie berechtigte Fragestellungen angehen, nicht mehr durchführbar wären. Ein weiteres Beispiel zur Thrombolysebehandlung soll dies erläutern: In der COBALT-Studie (Vergleich zweier Alteplase-Verabreichungsschemata bei Patienten mit akutem Myokardinfarkt; [12]) wurde eine 30-Tage-Referenzmortalität von 6,3% angenommen und die Noninferioritätsgrenze auf 6,7% festgelegt – im Vergleich zur oben genannten COMPASS-Studie also eine sehr viel konservativere Grenze. Vor Studienbeginn ging man allerdings davon aus, dass das untersuchte Doppelbolus-Schema die Mortalität um 0,9% gegenüber der kontinuierlichen Infusionstherapie senken würde und berechnete eine zum Nachweis der Noninferiorität notwendige Fallzahl von etwa 8000 Patienten. Wäre man dagegen von einer identischen Wirksamkeit beider Schemata ausgegangen und hätte dann eine Abweichung von 0,4% toleriert, so hätten 50 000 Patienten in die Studie eingeschlossen werden müssen [13].

Die Beispiele zeigen, wie schwierig der Kompromiss zwischen rationalen Annahmen und Erhaltung der Durchführbarkeit einer Studie sein kann. Wenn ein Studienergebnis einen zuvor weit gesteckten Noninferioritätsspielraum doch einmal voll ausschöpft, liegen neue Diskussionen über die klinische Relevanz und ethische Akzeptanz eines solchen Ergebnisses auf der Hand.

Mortalität als Bestandteil eines kombinierten Endpunkts

Aufgrund des medizinischen Fortschritts hat die Mortalität vieler Erkrankungen inzwischen stark abgenommen. Eine isolierte Betrachtung von wenigen fatalen Ereignissen würde dann sowohl eine lange Studiendauer als auch immens hohe Patientenzahlen erfordern und stellt somit oftmals keinen geeigneten primären Endpunkt für die Untersuchung neuer Therapieoptionen mehr dar. Als Ausweg aus diesem «Dilemma» entwickeln sich Studien mit zusammengesetzten Endpunkten, bei denen die Mortalität eine unter mehreren Endpunktkomponenten ist, zunehmend zum Standard.

Ist die Mortalität Bestandteil eines kombinierten Endpunkts, sollte man sich die Ergebnisse von jeder Endpunktkomponente genau anschauen

Was bei der Auswahl und Zusammenstellung von Endpunktkomponenten zu beachten ist, war bereits Gegenstand einer früheren Publikation [14]. An dieser Stelle sind nur ausgewählte Aspekte genannt, die zu Interpretationsschwierigkeiten bei diesem Studientyp führen.

Interpretationsschwierigkeiten ergeben sich vor allem dann, wenn Todesfälle als relativ seltenes Ereignis mit viel häufiger auftretenden, aber sogenannten «weichen» Endpunkten wie z.B. Laborvariablen kombiniert werden. Ein Beispiel hierfür ist die Kombination von Todesfällen, terminaler Niereninsuffizienz und einem Anstieg des Serumkreatinins bei Patienten mit diabetischer Nephropathie, wobei der Diagnose terminale Niereninsuffizienz wiederum Erhöhungen des Serumkreatinins zugrunde liegen. Zeigt sich in Bezug auf die Mortalität kein Unterschied zwischen zwei untersuchten Behandlungen, gründen sich allfällige positive Studienergebnisse allein (oder weit überwiegend) auf die Beeinflussung einer Laborvariable (Beispiel in [15]). Es ist daher sehr ratsam, sich ein Studienergebnis in Bezug auf jede einzelne Endpunktkomponente, insbesondere die von höherer klinischer Relevanz, genau anzuschauen; zudem ist darauf zu achten, ob gegebenenfalls ungerechtfertigte Aussagen in Bezug auf die Mortalitätsbeeinflussung gemacht werden. Ein Beispiel dafür ist die Behauptung in einem Publikationsabstract: «…compared with accelerated t-PA, primary stenting reduces death, reinfarction, stroke or repeat target vessel revascularisation» [16], zumal die Mortalität in der betreffenden Studie praktisch nicht zum Gesamtergebnis beitrug und unter der Stentbehandlung sogar leicht höher war.

Eine Beschränkung der Auswertung auf die erkrankungsspezifische Mortalität kann die Zunahme von Todesfällen aus anderen Gründen maskieren

Auswertung und Darstellung der Studienergebnisse

Gemäss dem Intention-to-treat-Prinzip sind alle randomisierten Patienten bis zu ihrem Tod bzw. bis zum geplanten Ende der Studie zu beobachten und auszuwerten. Dabei sind die Spezifikationen im Studienprotokoll und in einem separaten Analyseplan exakt zu befolgen. Ob dies der Fall ist, lässt sich häufig anhand des vorher publizierten Studiendesigns bzw. der Online-Aufschaltung des Studienprotokolls (z.B. als Appendix zur Publikation, Beispiel: Referenz [4]) überprüfen.

Sowohl in der Behandlungsgruppe als auch in der Kontrollgruppe ergeben sich Ereignisraten – idealerweise als Anzahl Todesfälle pro Anzahl Personenjahre ausgedrückt. Die relativen Behandlungseffekte können dann als Verhältnis von zwei Ereignisraten ausgedrückt werden; man nennt dieses Verhältnis auch «Hazard Ratio (HR)».

Ist die erkrankungsspezifische Mortalität der primäre Endpunkt oder eine Komponente eines zusammengesetzten Endpunktes, sollte man sich auf jeden Fall auch die Gesamtmortalität in der Studie anschauen. Eventuell wird durch Bezeichnungen wie «vascular death» oder «arrhythmic death» eine Zunahme von Todesfällen aus anderen Gründen maskiert [17].

Noninferioritätsstudien sind anhand von Vertrauensintervallen auszuwerten; herkömmliche statistische

Testverfahren spielen hier keine Rolle. Werden ausschliesslich p-Werte angegeben, d.h. dass die Schlussfolgerung der therapeutischen Nichtunterlegenheit auf einem statistisch nichtsignifikanten Unterschied zwischen den untersuchten Behandlungen beruht, ist die Studie in Bezug auf die Fragestellung wertlos.

In der Regel werden die Studienresultate anhand von Kaplan-Meier-Überlebenskurven graphisch dargestellt. Dieser Ansatz berücksichtigt, dass sich die Patienten unterschiedlich lange in der Studie befanden (frühe oder späte Aufnahme in die Studie; Studienabbrecher). Der Zeitpunkt Null entspricht für alle Patienten dem Zeitpunkt der Randomisierung. Die überlebenden Patienten werden am Ende der individuellen Beobachtungszeit «zensiert», oftmals durch kleine vertikale Striche auf der Überlebenskurve gekennzeichnet. Während im Verlauf der Studie aufgetretene Todesfälle eine Stufenbildung der Kurve nach sich ziehen, verändert das Zensieren die Form der Überlebenskurve nicht.

Bei der Beurteilung von Kaplan-Meier-Überlebenskurven sollte zudem auf Folgendes geachtet werden: Bildet die Graphik die gesamte Y-Achse (z.B. Überlebensrate in %) ab, oder wird nur ein Ausschnitt gezeigt? Sehr geringe Unterschiede zwischen zwei Behandlungen können durch eine entsprechende Darstellung bedeutungsvoller aussehen als sie tatsächlich sind. Weitere «Tricks» finden sich in Referenz [18].

Ist unter der X-Achse (Beobachtungsdauer) die Anzahl der Patienten angegeben, die zu einem gegebenen Studienzeitpunkt noch unter Beobachtung («at risk») standen? Kaplan-Meier-Kurven erstrecken sich häufig über viele Beobachtungsjahre, aber ab einem bestimmten Zeitpunkt sind gegebenenfalls nur noch wenige Patienten unter Beobachtung. Diese Kurvenanteile sagen unter Umständen nichts mehr aus.

Empfehlungen für den Leser von Publikationen

Trotz der grossen klinischen Bedeutung erfüllen klinische Interventionsstudien, die Aussagen zur Mortalität von Patienten machen, nicht immer die an sie zu stellenden methodologischen Anforderungen. Um den tatsächlichen Wert einer Studie zu beurteilen sind die «richtigen» Fragen zu stellen. Die 13 wichtigsten im Verlauf dieses Beitrags ausgesprochenen Empfehlungen sind deshalb in Frageform wiedergegeben. Diese mögen eine kleine Hilfestellung für Ihren «kritischen Blick» sein.

1. Ist die Mortalität der prospektiv festgelegte primäre Studienendpunkt?
2. Sind die Studienhypothese und die Fragestellung klar? Geht es um die Überlegenheit einer Therapie oder um ihre Nichtunterlegenheit?
3. Ist die Fallzahlberechnung nachvollziehbar?
4. Finden sich Angaben zur ausreichenden Power (Teststärke) der Studie?
5. Wurde bei einer Noninferioritätsstudie der Noninferioritätsbereich prospektiv festgelegt?
6. Wurde das Intention-to-treat-Prinzip strikt angewendet?
7. Entsprechen die präsentierten Ergebnisse exakt der geplanten Datenauswertung?
8. Ist das Ergebnis der Studie klinisch relevant?
9. Inwieweit trägt bei einem zusammengesetzten Endpunkt die Mortalität zum Gesamtergebnis bei? Wird der Einfluss auf die Mortalität eventuell von anderen Endpunktkomponenten maskiert?
10. Wie sieht bei Untersuchungen zur erkrankungsspezifischen Mortalität die Gesamtmortalität aus?
11. Wurde eine Noninferioritätsstudie anhand von Vertrauensintervallen ausgewertet?
12. Wird das Ergebnis bei der graphischen Darstellung eventuell «geschönt» wiedergegeben?
13. Werden unangemessene Aussagen zur Beeinflussung der Mortalität gemacht?

Beschränken Sie sich nicht auf das Lesen der Zusammenfassung (Abstract) einer Studie. Um dem «Fehler 3. Art» aus dem Weg zu gehen, nämlich, «dass es keine Daten gibt, die die Schlussfolgerung der Studienautoren stützen», bedarf es einer sorgfältigen Auseinandersetzung mit der gesamten Publikation.

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

Empfohlene Literatur
– Ware JH, Antman EM. Equivalence trials. N Engl J Med. 1997; 337:1159–61.
– Gottlieb SS. Dead is dead: artificial definitions are no substitute. Lancet. 1997;349:662–3.
– Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. Lancet. 2002; 359:1686–9.

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

Mortalität als Endpunkt klinischer Studien / Mortalité comme paramètre d'études cliniques

Literatur (Online-Version) / Références (online version)

1 Fisher LD, Moye LA. Carvedilol and the Food and Drug Administration (FDA) approval process: an introduction. Control Clin Trials. 1999;20:1–15.

2 Fisher LD. Carvedilol and the Food and Drug Administration (FDA) approval process: the FDA paradigm and reflections on hypothesis testing. Control Clin Trials. 1999;20:16–39.

3 Tashkin DP, Celli B, Senn S, et al., for the UPLIFT Study Investigators. A 4-year of tiotropium in chronic obstructive pulmonary disease. N Engl J Med. 2008;359:1543–54.

4 UPLIFT trial statistical analysis plan. Appendix zu [3], N Engl J Med online. http://content.nejm.org/cgi/data/NEJMoa0805800/DC1/1.

5 Wedzicha JA, Calverley PMA, Seemungal TA, et al. The prevention of chronic obstructive pulmonary disease exacerbations by salmeterol/fluticasone propionate or tiotropium bromide. Am J Respir Crit Care Med. 2008;177:19–26.

6 Echt DS, Liebson PR, Mitchell IB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo: the cardiac arrythmia suppression trial. N Engl J Med. 1991;324:781–8.

7 Home PD, Pocock SJ, Beck-Nielsen H, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. Lancet. 2009;373:2125–35.

8 Halpern SD, Karlawish JHT, Berlin JA. The continuing unethical conduct of underpowered clinical trials. JAMA. 2002;288:358–62.

9 Kleist P. Zehn Anforderungen an therapeutische Äquivalenzstudien. Schweiz Med Forum. 2006;6(37):814–9.

10 Tebbe U, Michels R, Adgey J, et al. Randomized, double-blind study comparing saruplase with streptokinase therapy in acute myocardial infarction: the COMPASS equivalence trial. J Am Coll Cardiol. 1998;31:487–93.

11 Garattini S, Bertelé V. Non-inferiority trials are unethical because they disregard patients' interests. Lancet. 2008;370:1875–7.

12 The Continuous Infusion versus Double-Bolus Administration of Alteplase (COBALT) Investigators. A comparison of continuous infusion of alteplase with double-bolus administration for acute myocardial infarction. N Engl J Med. 1997;337:1124–30.

13 Ware JH, Antman EM. Equivalence trials. N Engl J Med. 1997;337:1159–61.

14 Kleist P. Klinische Studien mit zusammengesetzten Endpunkten (Composite Endpoints). Schweiz Med Forum. 2008;8(47):908–12.

15 Lewis EJ, Hunsicker LG, Clarke WR, et al. Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. N Engl J Med. 2001;345:851–60.

16 Le May MR, Labinaz M, Davies RF, et al. Stenting versus thrombolysis in acute myocardial infarction trial (STAT). J Am Coll Cardiol. 2001;37:985–91.

17 Gottlieb SS. Dead is dead: artificial definitions are no substitute. Lancet. 1997;349:662–3.

18 Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. Lancet. 2002;359:1686–9.