Point of view

The interpretation of lack of evidence of a difference in efficacy: equivalence trials and the treatment of fungal infections

The demonstration of treatment efficacy by a prospective randomised placebo controlled trial has become established in the medical literature. In situations where effective therapy already exists, the introduction of newer therapeutic agents using placebo controlled trials is controversial[1-3] and comparisons with "standard therapy" are frequently undertaken using so-called equivalence or non-inferiority trials.[4,5] The use of such trials obviously presupposes the established efficacy of a therapy, but the formulation of a "standard therapy" in the critical care setting has been difficult, as opposed to, say, the practice of cardiology. For instance, the Fibrinolytic Therapy Trialists' Collaborative Group reviewed reports of fibrinolytic and standard therapy for myocardial infarction (ST segment elevation and/or bundle-branch block with randomisation within 6 hours of symptom onset) in 58600 patients and demonstrated an overall absolute 35-day mortality reduction of -1.84% (95% CI: -2.34% to -1.35%).[6] Thus fibrinolytic therapy, in particular streptokinase, has become a standard therapy and it is "no longer ethical to withhold ...(such therapy)..from patients…".[7]

The usual (i.e. placebo) controlled trial is a superiority trial, where the aim is to rule out treatment equality by rejection of the null hypothesis that the two treatments are the same. However, the converse proposition does not hold; that the failure to reject the null hypothesis (the "negative" clinical trial) establishes equivalence.[8-11] An illustration of this was the report of a clinical trial comparing trimethoprim-sulphamethoxazole and pentamidine in the treatment of *Pneumocystis carinii* pneumonia. Forty patients were enrolled and no difference was seen in 21-day mortality rates ($p = 0.18$) or other indices of improvement or of toxicities.[12] The trialists concluded that the two study treatment arms were "probably of equal effectiveness". As Polis and Blackwelder noted, apropos the question of study sample size and β error, "With additional patients, this study may have contributed more toward the resolution of this issue… (therapy of *P. carinii* pneumonia)… . However, it does not demonstrate that trimethoprim-sulphamethoxazole and pentamidine are equally effective; failure to show a significant difference …..is not at all the same as showing equivalence".[13]

The equivalence trial reverses the logic of the superiority trial; the null hypothesis is instead that of a *specified difference* (δ) between the experimental therapy and an active control.[14-16] Thus if μ is the "true" treatment difference (experimental vs control therapy and μ is positive when the experimental is superior to standard therapy):[17,18]

i) in a superiority trial the test is $\mu = 0$ vs $\mu \neq 0$ at the 5% level (or rather, $\mu \leq 0$ vs $\mu \geq 0$ at the 2.5% level)
ii) in an equivalence trial, the purpose is to demonstrate minimal differences (i.e. experimental therapy vs standard) in either direction. Therefore, $\mu \leq -\delta$ or $\mu \geq \delta$ is tested (two-sided) against $-\delta < \mu < \delta$ Alternatively, a pair of one-sided hypotheses are tested: H_1 $\mu \leq -\delta$ vs $\mu > -\delta$ and H_2 $\mu \geq \delta$ vs $\mu < \delta$ (both one-sided hypotheses need to be rejected). True equivalence trials are usually bio-equivalence trials.[19]
iii) in a non-inferiority trial, the purpose is to demonstrate that the experimental therapy is not substantially worse than active-control. Therefore, $\mu \leq -\delta$ is tested against $\mu > -\delta$. The test is one-sided at α significance level (usually 0.05). Such testing may be subject to the known limitations of hypothesis testing in general.[20,21] Alternatively, a $100(1-2\alpha)$ percent two-sided confidence interval for the treatment difference is computed and if the lower bound of the CI is $< -\delta$, non-inferiority can be claimed. Whether this be at the 90% or 95% is a point of some dispute, although recent regulatory recommendations suggest $\alpha = 0.025$ for one-sided testing of non-inferiority.[22] It is also noted that the strategy of using a two sided 90% CI for a one-sided 5% test assumes that the 90% CI is equal-tailed (each end of the interval excludes 5%).[23] In the clinical literature equivalence is often used synonymously with non-inferiority and, unless specified otherwise, this review will conform to this practice.

In figure 1, trials 1-4 show the above as hypothetical trials with point estimates and CI (95% for trials 1 & 2 and 90% CI for trials 3 & 4). In trial 1, placebo vs standard drug, the lower 95% CI approximates, in this scenario, the value of δ which is set at 20%. Trial 2 shows a successful superiority trial with lower 95% CI above zero. Trial 3 is an equivalence trial showing upper and lower 90% CI within ±δ. In trial 4, a non-inferiority trial, the lower 90% CI is $< -\delta$ (but the upper 95% CI is $> +\delta$).

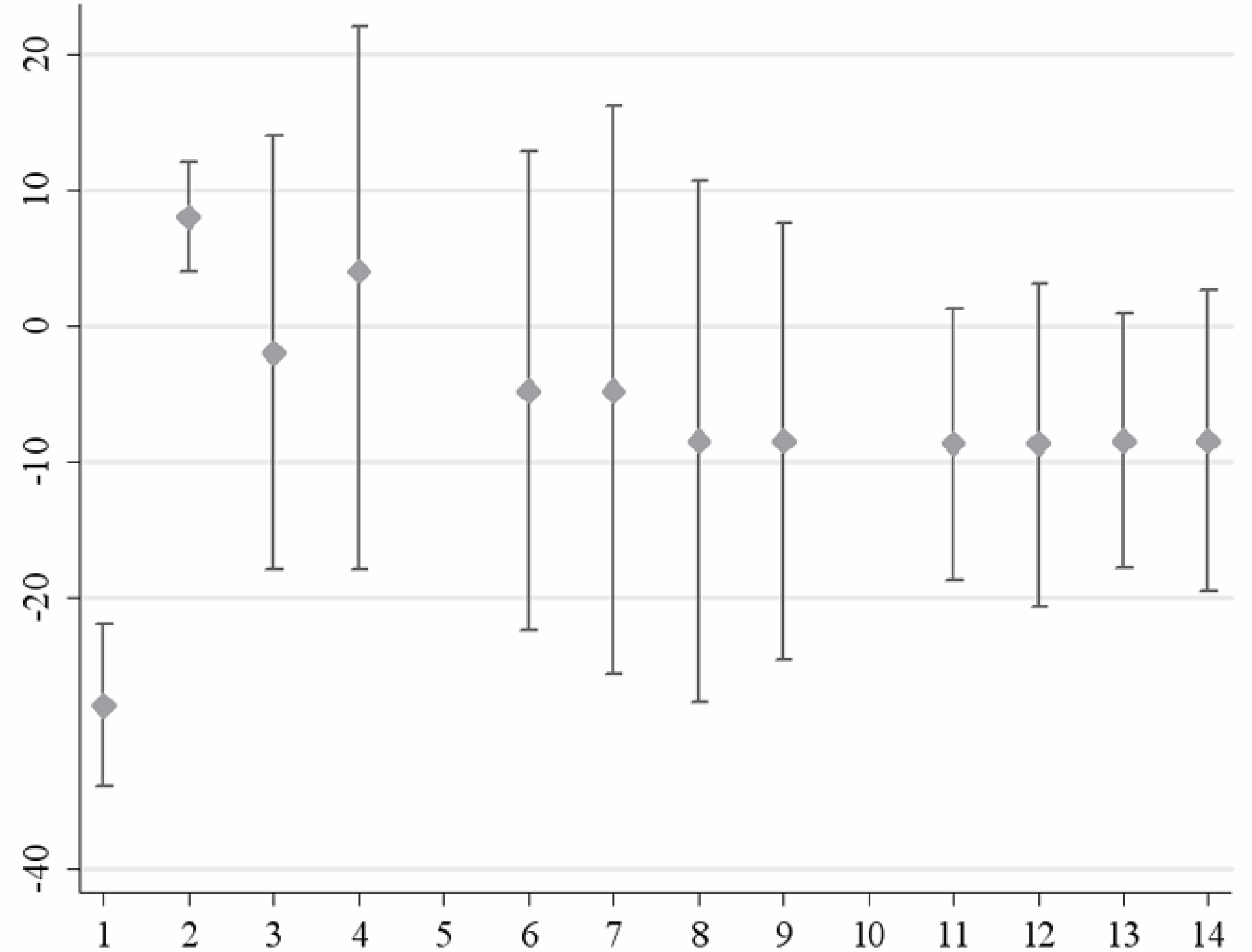

Figure 1. The vertical axis indicates the absolute risk response success (experimental – standard, as %), such that + ve values reflect efficacy of the experimental therapy and – ve values reflect efficacy of standard therapy. Diamonds represent point estimate with 90% CI, unless indicated.
The horizontal axis indicates trial type. Trials 1 - 4 are hypothetical examples of: 1. placebo/standard therapy, 2. superiority trial (successful, 95% CI, 3. equivalence trial (successful), 4. non-inferiority trial (successful). Trials 6 - 9 (Phillips *et al*, reference 69): 6. PP analysis, 7. PP analysis with 95% CI, 8. ITT analysis, 9. ITT analysis with 95% CI. Trials 11 - 14 (Rex *et al*, reference 67): 11. PP analysis, 12. PP analysis with 95% CI, 13. ITT analysis, 14. ITT analysis with 95% CI.

ISSUES IN EQUIVALENCE TRIALS[20,24-26]

1. *Assumptions made when conducting equivalence trials*
   a. *assay sensitivity*: that the active-control would have been superior to a placebo if such had been employed in the current trial. That is, an equivalence trial requires the consideration of "…information external to the trial".[27]
   b. *sensitivity to drug effects*: the ability of well designed trials to reliably demonstrate active-control drug effect (with respect to placebo)

   If the above two assumptions have not been met, then the interpretation of equivalence trials can be problematic. This was demonstrated by Tramer *et al*,[28] in their recent review of anti-emetics; in particular, the efficacy of ondansetron. They concluded that where no gold standard treatment existed and event rates (in this case, of emesis) varied widely "…trial designs without placebo controls are unlikely to yield sensible results".
   c. *constancy assumption*, that the historical treatment difference is preserved in the current trial. This may be difficult to sustain given changes in medical practice and the effect of different patient populations.
2. *Intention-to-treat (ITT) vs per-protocol (PP) analysis*: In superiority trials ITT is the preferred analysis compared with PP.[29] Such is not the case with equivalence trials where PP analysis is the more conservative and ITT tends to make treatment arms appear similar,[15,23,30] although this will depend upon the pattern of patient drop out and treatment assignment. Both types of analyses should be

presented, but this strategy must take into consideration the reduced patient number in a PP analysis when initial sample size calculations are made.

3. *Biocreep*: whereby a slightly inferior treatment becomes the active control for the next generation of equivalence trials and active controls become little different from placebos.[31]
4. *Conduct of the trial*: poor trial conduct in an equivalence trial will widen CI of the observed treatment effect and make the declaration of equivalence more difficult,[23,24,32] whereas in a superiority trial there will be a tendency to a null result which may be mistakenly claimed as indicating equivalence.
5. *Sample size requirements:* for equivalence studies sample size requirements are variably increased above similar superiority trials; on average about 10%.[1] Formulas for such calculations are provided in numerous articles[16,18,21,24] and specialised software is available.[33,34]
6. *The determination of δ (or non-inferiority margin)*: this may be formally defined as the largest acceptable clinical difference in treatment efficacy (experimental therapy vs standard) or, in the reverse, as a difference in patient status with an effect size ≤ δ that is non-detectable.[17] Thus δ is different (and usually smaller [35]) from the difference in proportions between two treatments ($\pi_1 - \pi_2$) used in routine sample size calculations for superiority trials.[36,16] Recommendations for the calculation of δ have been numerous and are found, not surprisingly, in the biopharmaceutical literature,[22,37-39] but the regulatory literature has been somewhat circumspect in prescribing δ a priori.[15,40] From a statistical perspective, δ has been defined as a certain fraction of:
   (a) the treatment effect of control drug vs placebo (for example, 0.2 - 0.5) or,
   (b) the lower 95% CI of this treatment effect (for example, 0.5) derived from a meta-analysis or large trial.[23,38,41,42]

In the anti-infective drug testing domain, the Food and Drug Administration (FDA) in the United States of America had informally provided a so called step-down function of δ that reflected the observed response rates in the equivalence study. For response rates (one or both arms) of at least 90%, ≥ 80% but < 90%, and ≥70% but < 80%, δ was suggested to be 10, 15 and 20% respectively.[43] The recent removal of this step-down function and the use of a more conservative δ (unofficially 10% [15]) has provoked comment regarding the unavoidable and large increment in trial size consequent upon this decision.[31]

In cardiology, where standard care in the treatment of acute myocardial infarction has been effectively established, equivalence trials have used streptokinase as the standard and δ has been set at a much lower level. In the INJECT trial,[44] which compared reteplase with streptokinase using 35 day mortality as end-point, equivalence was established if the (upper) CI of mortality difference excluded the possibility that reteplase mortality was > 1% worse than streptokinase mortality.[45] (note here and subsequently, the algebraic reversal when a positive treatment difference, i.e. μ, indicates worse outcome). The difference was in fact 0.5% with two-sided 90% CI for the difference: -1.7% to 0.71% (two-sided 95% CI: -1.96% to 0.98%).

A different perspective may be taken of the attempt to infer equivalence from a large negative superiority trial.[39] GUSTO III,[46] was designed as a superiority trial (15059 patients enrolled) to detect a 20% difference in 30-day post myocardial infarction mortality, comparing double-bolus reteplase relative to an accelerated infusion of atelpase. The mortality for the reteplase arm was 7.47% and that of alteplase 7.24, an absolute mortality difference of 0.23% (two-sided 95% CI: -0.65% to 1.11%) This would, as the trialists noted "exceed a definition of equivalence requiring a difference of less than 1%",[46,47] but they did observe, parenthetically, that the INJECT trial [44] had used 90% CI to establish equivalence. On this basis, equivalence would have been established in GUSTO III (two-sided 90% CI: -0.51% to 0.98%).

What is an appropriate equivalence margin is further illustrated by a consideration of the COBALT equivalence trial,[48] where 7169 patients were enrolled to compare 30-day post myocardial infarction mortalities, weight adjusted accelerated alteplase versus double bolus alteplase. On the basis of a 0.4% lower 95% CI for the absolute difference of accelerate infusion of alteplase vs streptokinase in the GUSTO I trial,[49] equivalence was defined in the COBALT trial if the upper boundary of a one-sided 95% CI of the difference in mortality did not exceed 0.4%. The absolute mortality difference of 0.44% with two-sided 90% CI: -0.57 to 1.49% thus failed to sustain equivalence. The differences in these approaches provoked editorial comment by Ware and Antmann,[50] who also noted the consequences of the calculation of sample size when based upon the assumption of unequal mortality rates in the two arms. In the COBALT trial it was assumed that 30-day mortality rates would be 6.3% in accelerated alteplase and 5.4% with double bolus alteplase and the trialists calculated an initial equivalence sample size of 4029 per group (although, by our calculations it was 4039, using the software package PASS 2002 [34]). However, the sample size required for approximately equal mortality rates of 7.5% (the range of the two arms reported in the COBALT trial) was identified by Ware and Antman as about 50000 in each group and the power of the COBALT study as effectively 0.16 (our calculations: 53634 in each group and power 0.158). Ware and Antman further suggested a absolute difference of 1.5% as a reasonable compromise for equivalence studies of this type, which, with 80% power, would require 3832 in each arm, a not impossible task for cardiology trials.

7. *Testing for noninferiority and superiority*: within the same trial it is possible to test sequentially for non-inferiority and superiority,[51] although there are inherent problems in this strategy.[52,30] At the least, trial methodology statements must pre-specify these analyses; the direction of testing should be:
   (a) initial demonstration of non-inferiority and,
   (b) subsequent testing for superiority (the reverse is problematic in interpretation, if superiority is shown in ITT analysis, but non-inferiority is not demonstrated in PP analysis); type I error must be preserved and the potential problems of different/unequal patient populations (PP analysis for non-inferiority and ITT analysis for superiority) must be addressed, with for example, imputation of missing values.[23]

   Declaring (formal) non-inferiority when the primary trial methodology of superiority has been unsuccessful "…should be looked upon with healthy skepticism" ...[30] that is, it has the status of a post-hoc analysis.

Overviews of equivalence trials

Two recent papers have assessed the performance of trials where clinical or therapeutic equivalence had been affirmed. Greene *et al*,[53] studied 88 reports from 1992 to 1996 claiming equivalence; δ was formally set in only 23% of reports, in 67% equivalence was declared after a failed test of superiority, the sample size was calculated in advance in only 33%, and in 25% of reports n was ≤ 20 per group. Of interest, δ ranged from 0 to 76% for proportionate differences. McCalister *et al*,[27] reviewing 4 recent hypertensive trials which appeared to show equivalence between treatment arms, found a lack of fulfillment in all of these trials of the 6 additional features that were defined as distinguishing superiority from equivalence trials; that: the active control was previously shown to be effective, similarity of (current) patients and outcome variables to those of original trials, optimal application of regimens, appropriate analysis, pre-specified δ and adequate sample size.

CRITICAL CARE IMPLICATIONS

Large trials with small treatment effect margins

In this context, it is if interest to look again at the protocols of the ANZICS Clinical Trials Group SAFE trial,[54] which will compare saline and albumin solutions for resuscitation. A total of 7000 patients are to be enrolled to detect a ≥ 3% absolute mortality difference between treatment groups based upon an assumed 15% control mortality and β error 0.1. The magnitude of this mortality difference was derived from the lower 95% CI of the estimated treatment effect (albumin versus non-albumin use) from the 1998 Cochrane Injury Group Albumin Reviewers paper.[55] As the SAFE trial (a superiority trial by its methodology description) is of large size with a relatively small treatment difference targeted, the question may be posed: if the null hypothesis of no treatment effect is not rejected, are we able to conclude equivalence between the two regimens? Despite the cautions above, some support for this scenario is provided by Ng who argues that "If the sample size is such that the …β error …at some δ is sufficiently small (e.g. < 0.05) we can conclude that the two treatments are δ-equivalent".[39] However, two restrictions may apply to this proposition: the small β error and a robust δ. From the 1998 Cochrane paper, the lower 95% CI of treatment effect with a random effects estimator was 0.016 (our calculations are based upon the "metan" routine,[56] using Stata™ software[57]) . Thus possible values of δ would be 3% (lower 95% CI of fixed effects meta-analysis), 0.016 (lower 95% CI of random effects meta-analysis), 0.015 (50% of lower 95% CI of fixed effects meta-analysis). Total sample sizes under these scenarios are variably in excess of that of the SAFE trial.

Treatment of fungal infections

Therapy for mycotic infections in the intensive care unit has been recently transformed by the introduction of a group of drugs which have been marketed as "equivalent" to the enduring yardstick, amphotericin B.[58] Since its introduction in 1957, amphotericin B has been the standard (and until recently, the only) antifungal therapy,[59] despite never being compared with placebo and only two randomised trials of its action reported before the more recent comparisons with fluconazole.[60] The effect of no specific therapy for mycotic infections (e.g. candidaemia) is difficult to estimate, but the prospective observational study (n = 427) of Nguyen *et al*,[61] suggests mortality rates of 27% with and 74% without antifungal therapy.

Amphotericin B and fluconazole

There are 5 randomised trials comparing amphotericin B and fluconazole;[62-66] in 2, amphotericin B was combined with flucytosine,[62,64] and in 2 there were intention to treat and "efficacy" groups.[65,66] Study sizes ranged from 40 to 237, for a total of 619 patients, a modest number, the implications of which, for the conduct of further comparative trials in mycotic infections, have been commented upon.[60] For all of these studies, the primary evaluable end-point was responsiveness to therapy, not mortality, and in the "evaluable" groups responsiveness varied from 57% to 78% (mean, 68% for amphotericin and 61% for fluconazole) with an overall mortality ranging from 12% to 38% (mean mortality 28% and no difference between the two drugs). The Phillips *et al* 1997 study was formally conducted with an "equivalence" protocol: $\delta = 0.2$, one-sided $\alpha = 0.05$ with an assumed response rate of 70% and requiring a total of 148 patients.[65] Recruitment was limited to 106 patients: point estimates and CIs for PP ("efficacy") and ITT analyses are seen as trials 6-9 in Figure 1. Lower borders (90% and 95%) of CI for both PP and ITT analyses are > $-\delta$ and the Westlake version,[67] of the two one-sided hypothesis tests for equivalence, also fails (p = 0.08 & 0.12).[68] The largest of these trials, by Rex *et al* in 1994,[69] recruited 237 patients and tested a null (equivalence) hypothesis "…that the difference between the proportions of patients with favourable responses in the two groups would be less than 20 percent." Although surprisingly little was said in the methodology and analysis regarding "equivalence". This being said, equivalence can be established for both the PP and ITT analyses (see trials 11-14 in Figure 1), with p values for the two one-sided hypothesis tests for equivalence at 0.03 and 0.02 respectively, albeit the 95% lower border for the PP analysis (Figure 1, trial 12) was marginal. Neither trial was able to demonstrate equivalence/non-inferiority at $\delta = 0.1$. The other three trials,[62-64] used superiority methodology and inferred "equivalence" from failure to demonstrate a difference. Pooling the studies and using quantitative meta-analytic techniques,[56] the treatment effect for the PP analysis, fluconazole versus amphotericin, was – 6%, 95% CI: -13.8% to 1.9% (p = 0.14) and for the ITT analysis, -0.82%, 95% CI: -16.7% to 0.4% (p = 0.06). Although the pooled effect fail to demonstrate a definite treatment advantage for amphotericin, what can be inferred from the CI of the treatment effect is that a δ of 0.2 (20%) appears to be too large.

Four of the trials reported various toxic effects of therapy which were more common in the amphotericin group.[62,63,65,69] Surprisingly, other than in the Annaissie *et al* trial,[63] where indices of amphotericin induced renal toxicity returned to baseline level (at final review) in 73%, there was no systematic evaluation of the consequences of renal toxicity nor reporting of toxicity surrogates such hospital length of stay. This is important as it is problematic to claim "equivalence" from a randomised trial of efficacy and suggest "superiority" using arguments about toxicity (or other non-formally assessed end-points).[1] Although not germane to this paper, other areas of concern in the interpretation of these trials are patient population (e.g. neutropenic versus non-neutropenic), the number and the appropriate management of catheter associated infections (79% in the Rex *et al* trail[69]), high versus low dose amphotericin (< 500mg total vs > 500mg [61]), especially in the context of a large number of catheter associated infections, and the omission of pre-emptive management of known toxic side-effects.

Amphotericin and new-generation antifungals

The "high" incidence of toxicity, especially renal,[70] associated with amphotericin B has prompted re-assessments of the first-line position of this drug in mycotic infections.[71,72] Two recent large equivalence trials of Caspofungin, an echinocandin (total n = 687, $\delta = 0.1$)[73] and liposomal amphotericin B (total n = 239, $\delta = 0.2$)[74] versus amphotericin B have suggested comparable efficacy of these two drugs with respect to amphotericin B and a reduction in toxicities. A similarly large superiority trial of Amphotericin B colloidal-dispersion versus the parent compound concluded "comparable efficacy",[75] but noted increased non-renal infusion toxicities with colloidal-dispersion amphotericin, a point re-iterated by Drew *et al* in correspondence over the liposomal amphotericin B trial.[76] What is pertinent in this series of trials is the high cost of the group of amphotericin comparators; that is, the trade-off of the extra costs of toxicities (especially renal) versus drug costs. Cagnoni *et al*,[77] undertook a pharmaco-economic analysis of the liposomal amphotericin B trial above and found that the hospital acquisition cost of the drug was critical in determining the break-even point, but it is noted that only 60% of patients were evaluated and "costs" were inferred from billing data, a strategy which has been previously criticized.[78] O'Connell *et al* undertook a similar cost-effectiveness study of the use of liposomal amphotericin B in 2002 and estimated the cost per additional life saved to be £ 23,819 (≡ $AUS 57, 574).

Conclusions

Claims in the literature as to the demonstration of "equivalence" must be subjected to careful scrutiny. The particular methodology of equivalence trials is of critical importance with respect to the conclusions that may be inferred, especially as these trials require the concurrent assessment of appropriate "external information". Extension of conclusions beyond the "equivalence" hypothesis must also be formally assessed.

J. L. MORAN
Department of Intensive Care Medicine, Queen Elizabeth Hospital, Woodville, SOUTH AUSTRALIA

P. J. SOLOMON
School of Applied Mathematics, University of Adelaide, Adelaide, SOUTH AUSTRALIA

REFERENCES

1. Djulbegovic B, Clarke, M. Scientific and ethical issues in equivalence trials. JAMA 2001;285:1206-1208.
2. Rothman KJ, Michels, K B. The continuing unethical use of placebo controls. N Engl J Med 1994;331:394-398.
3. Temple R. Problems in interpreting active control equivalence trials. Accountability in Research 1996;4:267-275.
4. Ellenberg SSP, Temple, R M. Placebo-Controlled Trials and Active-Control Trials in the Evaluation of New Treatments: Part 2: Practical Issues and Specific Cases. Ann Intern Med 2000;133:464-470.
5. Temple RM, Ellenberg, S S P. Placebo-Controlled Trials and Active-Control Trials in the Evaluation of New Treatments: Part 1: Ethical and Scientific Issues. Ann Intern Med 2000;133:455-463.
6. Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. Fibrinolytic Therapy Trialists' (FTT) Collaborative Group. Lancet 1994;343:311-322.
7. White HD. Thrombolytic therapy and equivalence trials. J Am Coll Cardiol 1998;31:494-496.
8. Altman DG, Bland, J M. Statistics Notes: Absence of evidence is not evidence of absence. BMJ 1995;311:485.
9. Freiman JA, Chalmers, T C, Smith, H, Jr., Kuebler, R R. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. N Engl J Med 1978;299:690-694.
10. Kim MY, Buyon, J P, Petri, M, Skovron, M L, Shore, R E. Equivalence trials in SLE research: issues to consider. Lupus 1999;8:620-626.
11. Spriet A, Beiler, D. When can 'non significantly different' treatments be considered as 'equivalent'? Brit J Clin Pharmacol 1979;7:623-624.
12. Wharton JM, Coleman, D L, Wofsy, C B, et al. Trimethoprim-sulfamethoxazole or pentamidine for Pneumocystis carinii pneumonia in the acquired immunodeficiency syndrome. A prospective randomized trial. Ann Intern Med 1986;105:37-44.
13. Polis MA, Blackwelder, W C. Trimethoprim-sulphamethoxazole or pentamidine for Pneumocystis carinii pneumonia. Ann Intern Med 1987;106:475.
14. Blackwelder WC. "Proving the null hypothesis" in clinical trials. Control Clin Trials 1982;3:345-353.
15. D'Agostino RB, Massaro, J M, Sullivan, L M. Non-inferiority trials: design concepts and issues- the encounter of academic consultants in statistics. Stat Med 2003;22:169-186.
16. Makuch RW, Johnson, M F. Some issues in the design and interpretation of 'negative' clinical studies. Arch Int Med 1986;146:986-989.
17. Aras G. Superiority, noninferiority, equivalence, and bioequivalence - Revisited. Drug Inf J 2001;35:1157-1164.
18. Hwang IK, Morikawa, T. Design issues in noninferiority/equivalence trials. Drug Inf J 1999;33:1205-1218.
19. Williams RL, Chen, M-L, Hauck, W W. Equivalence approaches. Clin Pharmacol Ther 2002;72:229-237.
20. Makuch R, Johnson, M. Issues in planning and interpreting active control equivalence studies. J Clin Epidemiol 1989;42:503-511.
21. Makuch RW, Simon, R M. Sample size considerations for non-randomized comparative studies. J Chronic Dis 1980;33:175-181.
22. Hauschke D. Choice of delta: A special case. Drug Inf J 2001;35:875-879.
23. Hauck WW, Anderson, S. Some issues in the design and analysis of equivalence trials. Drug Inf J 1999;33:109-118.
24. Jones B, Jarvis, P, Lewis, J A, Ebbutt, A F. Trials to assess equivalence: the importance of rigorous methods. BMJ 1996;313:36-39.
25. Kirshner B. Methodological standards for assessing therapeutic equivalence. J Clin Epidemiol 1991;44:839-849.
26. Windeler J, Trampisch, H-J. Recommendations concerning studies on therapeutic equivalence. Drug Inf J 1996;30:195-200.
27. McAlister FA, Sackett, D L. Active-control equivalence trials and antihypertensive agents. Am J Med 2001;111:553-558.
28. Tramer MR, Reynolds, D J, Moore, R A, McQuay, H J. When placebo controlled trials are essential and equivalence trials are inadequate. BMJ 1998;317:875-880.
29. Lewis JA, Machin, D. Intention to treat--who should use ITT? Br J Cancer 1993;68:647-650.
30. Wiens BL. Something for nothing in noninferiority/superiority testing: A caution. Drug Inf J 2001;35:241-245.
31. Shlaes DM, Moellering, R C, Jr. The United States Food and Drug Administration and the end of antibiotics. Clin Infect Dis 2002;34:420-422.
32. Chuang-Stein C. Clinical equivalence - A clarification. Drug Inf J 1999;33:1189-1194.
33. Elashoff, J. D. nQuery Advisor V 4.0. http://www.statsol.ie/nquery/nquery.htm
34. Hintze, J. L. PASS 2002. http://www.ncss.com/
35. Makuch R, Simon, R. Sample size requirements for evaluating a conservative therapy. Cancer Treat Rep 1978;62:1037-1040.
36. Hatala R, Holbrook, A, Goldsmith, C H. Therapeutic equivalence: all studies are not created equal. Can J Clin Pharmacol 1999;6:9-11.
37. Holmgren EB. Establishing equivalence by showing that a specified percentage of the effect of the active control over placebo is maintained. J Biopharm Stat 1999;9:651-659.

38. Ng T-H. A specification of treatment difference in the design of clinical trials with active controls. Drug Inf J 1993;27:705-719.
39. Ng T-H. Conventional null hypothesis testing in active control equivalence studies. Control Clin Trials 1995;16:356-358.
40. Wiens BL. Choosing an equivalence limit for noninferiority or equivalence studies. Control Clin Trials 2002;23:2-14.
41. Blackwelder WC. Showing a treatment is good because it is not bad: when does "noninferiority" imply effectiveness? Control Clin Trials 2002;23:52-54.
42. Ng T-H. Choice of delta in equivalence testing. Drug Inf J 2001;35:1517-1527.
43. Wiens BL, Iglewicz, B. Testing noninferiority of response rates for regulatory filings using transformations. Drug Inf J 2001;35:1165-1171.
44. Randomised, double-blind comparison of reteplase double-bolus administration with streptokinase in acute myocardial infarction (INJECT): trial to investigate equivalence. International Joint Efficacy Comparison of Thrombolytics. Lancet 1995;346:329-336.
45. Hampton JR. Mega-trials and equivalence trials: experience from the INJECT study. European Heart Journal 1996;17:Suppl E 28-34.
46. A comparison of reteplase with alteplase for acute myocardial infarction. The Global Use of Strategies to Open Occluded Coronary Arteries (GUSTO III) Investigators. N Engl J Med 1997;337:1118-1123.
47. Fleming TR. Design and interpretation of equivalence trials. Am Heart J 2000;139:S171-S176.
48. A comparison of continuous infusion of alteplase with double-bolus administration for acute myocardial infarction. The Continuous Infusion versus Double-Bolus Administration of Alteplase (COBALT) Investigators.N Engl J Med 1997;337:1124-1130.
49. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. The GUSTO investigators. N Engl J Med 1993;329:673-682.
50. Ware JH, Antman, E M. Equivalence trials. N Engl J Med 1997;337:1159-1161.
51. Dunnett CW, Gent, M. An alternative to the use of two-sided tests in clinical trials. Stat Med 1996;15:1729-1738.
52. Chuang-Stein C. Testing for superiority or inferiority after concluding equivalence? Drug Inf J 2001;35:141-143.
53. Greene WL, Concato, J, Feinstein, A R. Claims of equivalence in medical research: are they supported by the evidence? Ann Int Med 2000;132:715-722.
54. Finfer S, Bellomo, R, Myburgh, J, Norton, R. Efficacy of albumin in critically ill patients. BMJ 2003;326:559-560.
55. Cochrane Injuries Group Albumin Reviewers. Human albumin administration in citically ill patients: systematic review of randomised controlled trials. BMJ 1998;317:235-240.
56. Bradburn MJ, Deeks, J, Altman, D G. metan-sbe24 an alternative meta-analysis command. Stata Technical Bulletin Reprints 1998;8:100.
57. Stata Statistical Software, Version 8.0, 2003. Stata Corporation, College Station, TX
58. Kam LW, Lin, J D. Management of systemic candidal infections in the intensive care unit. American Journal of Health-System Pharmacy 2002;59:33-41.
59. Gallis HA, Drew, R H, Pickard, W W. Amphotericin B: 30 years of clinical experience. Rev Infect Dis 1990;12:308-329.
60. Rex JH, Walsh, T J, Nettleman, M, et al. Need for alternative trial designs and evaluation strategies for therapeutic studies of invasive mycoses. Clin Infect Dis 2001;33:95-106.
61. Nguyen MHM, Peacock, J E J M, Tanner, D C M, et al. Therapeutic Approaches in Patients With Candidemia: Evaluation in a Multicenter, Prospective, Observational Study. Arch Int Med 1995;155:2429-2435.
62. Abele-Horn M, Kopp, A, Sternberg, U, et al. A randomized study comparing fluconazole with amphotericin B/5-flucytosine for the treatment of systemic Candida infections in intensive care patients. Infection 1996;24:426-432.
63. Anaissie EJ, Darouiche, R O, Abi-Said, D, et al. Management of invasive candidal infections: results of a prospective, randomized, multicenter study of fluconazole versus amphotericin B and review of the literature. Clin Infect Dis 1996;23:964-972.
64. Kujath P, Lerch, K, Kochendorfer, P, Boos, C. Comparative study of the efficacy of fluconazole versus amphotericin B/flucytosine in surgical patients with systemic mycoses. Infection 1993;21:376-382.
65. Phillips P, Shafran, S, Garber, G, et al. Multicenter randomized trial of fluconazole versus amphotericin B for treatment of candidemia in non-neutropenic patients. Canadian Candidemia Study Group. Eur J Clin Microbiol Infect Dis 1997;16:337-345.
66. Rex JH, Pappas, P G, Karchmer, A W, et al. A randomized and blinded multicenter trial of high-dose fluconazole plus placebo versus fluconazole plus amphotericin B as therapy for candidemia and its consequences in nonneutropenic subjects. Clin Infect Dis 2003;36:1221-1228.
67. Westlake WJ. Symmetrical confidence intervals for bioequivalence trials. Biometrics 1976;32:741-744.
68. Goldstein R. Equivalence testing sg21. Stata Technical Bulletin Reprints 1994;3:107-112.
69. Rex JH, Bennett, J E, Sugar, A M, et al. A randomized trial comparing fluconazole with amphotericin B for the treatment of candidemia in patients without neutropenia. Candidemia Study Group and the National Institute. N Engl J Med 1994;331:1325-1330.
70. Wingard JR, Kubilis, P, Lee, L, et al. Clinical significance of nephrotoxicity in patients treated with amphotericin B for suspected or proven aspergillosis. Clin Infect Dis 1999;29:1402-1407.
71. Ostrosky-Zeichner L, Marr, K A, Rex, J H, Cohen, S H. Amphotericin B: time for a new 'gold standard'. Clin Infect Dis 37:415-25, 2003.
72. Rex JH, Walsh, T J. Estimating the true cost of amphotericin B. Clin Infect Dis 1999;29:1408-1410.
73. Mora-Duarte J, Betts, R, Rotstein, C, et al. Comparison of caspofungin and amphotericin B for invasive candidiasis. N Engl J Med 2002;347:2020-2029.

74. Walsh TJ, Finberg, R W, Arndt, C, et al. Liposomal amphotericin B for empirical therapy in patients with persistent fever and neutropenia. National Institute of Allergy and Infectious Diseases Mycoses Study Group. N Engl J Med 1999;340:764-771.
75. White MH, Bowden, R A, Sandler, E S, et al. Randomized, double-blind clinical trial of amphotericin B colloidal dispersion vs. amphotericin B in the empirical treatment of fever and neutropenia. Clin Infect Dis 1998;27:296-302.
76. Winston DJ, Schiller, G J, Territo, M C. Liposomal amphotericin B for fever and neutropenia. N Engl J Med 1999;341:1154-1155.
77. Cagnoni PJ, Walsh, T J, Prendergast, M M, et al. Pharmacoeconomic analysis of liposomal amphotericin B versus conventional amphotericin B in the empirical treatment of persistently febrile neutropenic patients. J Clin Oncol 2000;18:2476-2483.
78. Gyldmark M. A review of cost studies of intensive care units: problems with the cost concept. Crit Care Med 1995;23:964-972.