Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Interpreting meta-analysis according to the adequacy of sample size. An example using isoniazid chemoprophylaxis for tuberculosis in purified protein derivative negative HIV-infected individuals

This article was published in the following Dove Press journal:
Clinical Epidemiology
8 April 2010
Number of times this article has been viewed

Kristian Thorlund[1,2]
Aranka Anema[3]
Edward Mills[4]

[1]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada; [2]The Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark; [3]British Columbia Centre for Excellence in HIV/AIDS, University of British Columbia, Vancouver, British Columbia, Canada; [4]Faculty of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada

Objective: To illustrate the utility of statistical monitoring boundaries in meta-analysis, and provide a framework in which meta-analysis can be interpreted according to the adequacy of sample size. To propose a simple method for determining how many patients need to be randomized in a future trial before a meta-analysis can be deemed conclusive.

Study design and setting: Prospective meta-analysis of randomized clinical trials (RCTs) that evaluated the effectiveness of isoniazid chemoprophylaxis versus placebo for preventing the incidence of tuberculosis disease among human immunodeficiency virus (HIV)-positive individuals testing purified protein derivative negative. Assessment of meta-analysis precision using trial sequential analysis (TSA) with LanDeMets monitoring boundaries. Sample size determination for a future trials to make the meta-analysis conclusive according to the thresholds set by the monitoring boundaries.

Results: The meta-analysis included nine trials comprising 2,911 trial participants and yielded a relative risk of 0.74 (95% CI, 0.53–1.04, $P = 0.082$, $I^2 = 0\%$). To deem the meta-analysis conclusive according to the thresholds set by the monitoring boundaries, a future RCT would need to randomize 3,800 participants.

Conclusion: Statistical monitoring boundaries provide a framework for interpreting meta-analysis according to the adequacy of sample size and project the required sample size for a future RCT to make a meta-analysis conclusive.

Keywords: meta-analysis, trial sequential analysis (TSA), randomized clinical trials (RCTs), isoniazid chemoprophylaxis, adequacy of sample size, human immunodeficiency virus (HIV), tuberculosis, purified protein derivative negative

Introduction

Meta-analyses of randomized clinical trials (RCTs) are considered the highest level of evidence for assessment of the benefits and harms of an intervention.[1,2] While comprehensively conducted meta-analyses may provide strong inferences about the quality and generalizability of the available evidence,[1–3] they can provide misleading or inadequate inferences about the reliability of the evidence, the consistency of evidence, and the definitive effectiveness of a treatment.[4–11] This is particularly the

Correspondence: Edward Mills
Faculty of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada
Email emills@cfenet.ubc.c9

case when an insufficient number of events and patients exist from trials that have been pooled.[6–10,12,13]

As with conventional RCTs, we require an estimate of the reliability of the evidence according to statistical inferences made possible through an adequate number of events and patients in the individual trials.[14] Many now claim that it is unethical to conduct a clinical trial with an insufficient sample size, as inferences of effectiveness may be insufficient to make clinical recommendations.[15] While we do not believe this is an issue of ethics,[16] a similar clinical difficulty arises in interpreting meta-analysis if the analysis includes an insufficient number of trials, patients, and ultimately, events.[6–8,10–13] While sample size calculations are almost consistently mandated for single trials,[17] there has been comparatively little discussion of the need for adequate power in a meta-analysis.

Systematic evaluations reveal that many published meta-analysis and systematic reviews are deemed inconclusive or moderately conclusive by authors.[18] Meta-analyses are frequently updated and subjected to significance testing as new trials emerge. This scenario is akin to interim analyses and data monitoring in RCTs, where data safety monitoring committees (DSMBs) meet at planned and/or unplanned times to discuss if the interim results are sufficiently convincing to recommend termination of the trial and conclude that the experimental treatment is superior, equivalent, or inferior to the control treatment.[6–8,10,12,13,19,20] In this setting, DSMBs rely on sample size calculations and use of formal monitoring boundaries to control the risk of false positive and overly positive findings caused by the play of chance (random error) and/or repeated significance testing (multiplicity).[20–22]

Some argue that the reliability of the evidence in a meta-analysis should only be established in rigorous decision-making frameworks similar to those of DSMBs for single RCTs.[6,7,19] The sample size required for a conclusive and reliable meta-analysis has typically been referred to as the required or optimal information size,[6–8,10,11] The concatenation of meta-analysis, information size calculations and formal monitoring boundaries (or stopping rules), such as the Lan-DeMets alpha-spending monitoring boundary, has been dubbed trial sequential analysis (TSA) – analogous to group sequential analysis in single RCTs.[10] A growing body of evidence from empirical studies, simulation studies, and application examples has underlined the importance of incorporating TSA to achieve reliable answers in meta-analysis and systematic reviews.[6–9,12,13,23–28] In addition, TSA may be useful for determining how many patients need to undergo further randomization before the meta-analysis can be deemed conclusive and reliable.[8,10,12,13,19] Thus, determining the sample size of future RCTs in the setting of a prospective meta-analysis using TSA may be preferable to conventional sample size calculations.[16,29]

To illustrate the utility of TSA in meta-analysis, we apply it to a pressing clinical and public health issue, the use of isoniazid chemoprophylaxis (IHZ) for tuberculosis (TB) disease in purified protein derivative negative (PPD) human immunodeficiency virus (HIV)-infected individuals.

Isoniazid chemoprophylaxis (IHZ)

IHZ has long been considered an effective chemoprophylaxis for preventing TB incidence among HIV-infected individuals, particularly among those testing positive for purified protein derivative (PPD+).[30,31] However, the impact of IHZ has remained unclear among those who are testing PPD–, many of whom may have cutaneous anergy related to advanced HIV progression. Existing analyses have been limited by weak inferences about the effectiveness of IHZ in this population.[32] Trials evaluating the effectiveness of IHZ among PPD– HIV-infected individuals have thus continued to use placebo controls.[33–35] Given the unacceptably high mortality rates among HIV-positive and TB co-infected individuals,[36] there is a critical need to establish definitively whether IHZ can effectively prevent TB among individuals who test PPD–. We aimed to evaluate the effectiveness of IHZ for prevention of TB disease among HIV-positive adults testing PPD.[37–40]

Methods

Eligibility criteria

We included any RCT that evaluated the effectiveness of IHZ chemoprophylaxis versus placebo for preventing the incidence of TB among HIV-positive individuals testing PPD–. We included studies from any location and of any duration. Due to difficulties associated with diagnosis of TB among HIV-positive infants and children,[41] we focused our review of studies on adult populations and excluded those evaluating the effect of IHZ versus placebo on younger age groups (<13 years). We excluded studies where participants had current or previous diagnosis of TB. We included studies where patients had negative PPD skin tests. We defined our primary outcome as active TB, probable or confirmed by microbiological, histological, or clinical methods.

Search strategy

In consultation with a medical librarian, we established a search strategy (available from corresponding author

on request). We searched independently, in duplicate, the following 10 databases (from inception to August 2009): MEDLINE; EMBASE (Exerpta Medica); Cochrane Central Register of Controlled Trials (CENTRAL); Allied and Complementary Medicine Database (AMED); Cumulative Index to Nursing and Allied Health Literature (CINAHL); TOXNET; Development and Reproductive Toxicology; Hazardous Substances Databank; Psych-info; and Web of Science, that included the full text of journals (*OVID, ScienceDirect*, and *Ingenta*, including articles in full text from approximately 1700 journals since 1993). In addition, we searched the bibliographies of published systematic reviews.[30–32] We contacted the authors of studies for clarifications, where required.

Study selection

Two investigators (AA, EM) working independently, in duplicate, scanned all abstracts and obtained the full-text reports of records, that indicated or suggested that the study met inclusion criteria for the outcomes of interest. After obtaining full reports of the candidate trials the same reviewers independently assessed eligibility from full text papers. We judged RCTs to be of adequate quality if their generation of allocation sequences was unpredictable, if methods of allocation concealment ensured patients and investigators could not foresee treatment assignment, and if patient attrition was clearly described.

Data collection

The same two reviewers conducted data extraction independently using a standardized pre-piloted form. Reviewers collected information about the: study date; location; duration; trial size; mean age of participants; treatment regimens of control and active arms; and incidence of TB. We entered the data into an electronic database such that duplicate entries existed for each study, when the two entries did not match, we resolved differences by consensus.

Data analysis

We first calculated the Phi(Φ)-statistic in order to assess inter-rater reliability on inclusion of articles. This provides a measure of inter-observer agreement independent of chance.[42] We calculated the pooled relative risk (RR) using the DerSimonian-Laird random-effects meta-analysis.[3,43] We calculated the I^2 statistic for each analysis as a measure of the proportion of the overall variation that is attributable to between-study heterogeneity.[44] Forest plots displaying individual study RR with 95% confidence intervals (CIs), and the DerSimonian-Laird pooled estimate were conducted using Review Manager version 5.[2]

TSA and meta-analysis sample size requirements

A single-trial sample size calculation is typically based on expected event rates, expected relative risks, an alpha set at 0.05, and power of 0.80.[45] In meta-analysis, however, more conservative alpha and beta-levels may be required to ensure that the evidence is sufficiently compelling to justify recommendation of the experimental treatment for widespread use, or to ensure that the evidence is so convincing that further RCTs are not required.[6,7,19] Due to heterogeneity across included trial populations, treatments and methods, the required meta-analysis information size additionally need adjustment for variation across trials.[8,10–13] Such adjustments are analogous to adjustments for variation across centers in a multi-center trial as they account for the proportion of total meta-analysis variation expected to be explained by variation across (and not within) trials.[8,10,11]

We determined the required meta-analysis information size for detecting a 25% relative risk reduction in TB, assuming a control group incidence rate of 5% (approximately the median rate across trials) and assuming that 20% of the total variation in the meta-analysis would be explained by variation across trials (heterogeneity). We calculated the information size required to yield "moderate" meta-analytic evidence based on an alpha = 5% significance level, and beta = 20% (80% power). We also calculated the information size required to yield "strong" meta-analytic evidence based on an alpha = 1% significance level, and beta = 10% (90% power).

Significance testing in RCTs and meta-analysis make use of a standardized test statistic (Z-statistic) which can be transformed to a *P*-value and subsequently used to evaluate if the observed effects in the two treatments groups differ significantly. Z-statistics that lie outside the interval −1.96 to 1.96 correspond to *P*-values smaller than 0.05, and Z-statistics that lie outside the interval −2.57 to 2.57 correspond to *P*-values smaller than 0.01. In monitoring of RCTs it is common to establish formal stopping rules for the Z-statistic rather than the *P*-value.[46] In this vein, group sequential monitoring boundaries for the cumulative Z-statistic are calculated every time a new group of randomized patients are added to the analysis up till the point where the number of patients randomized surpasses the required sample size.[21,22,46] In meta-analysis, the monitoring boundaries may be applied analogously every time one or more trials are added up till the point where the

number of patients in the meta-analysis surpasses the required meta-analysis information size.[6–8,10,12,13]

We applied TSA by utilizing the Lan-DeMets alpha-spending approach with monitoring boundaries corresponding to the O'Brien-Fleming boundaries to assess the reliability of pooled inferences from our meta-analysis on TB.[10,22] Three systematic reviews on the topic have previously been published in year 1998, 1999, and 2004.[30–32] We therefore constructed monitoring boundaries to test for statistical significance in the meta-analysis including all trials up to year 1998, up to year 1999, and up to year 2004, albeit no new trials were published in 1999.

Topping up a sample size evaluation

Consider the situation where a new RCT on a topic is in the planning stage and where a fully updated meta-analysis on the same topic is deemed inconclusive. Once the RCT is completed (published) it will be added to the meta-analysis. The question then remains, how large does the new RCT need to be to make the meta-analysis conclusive? In our example, we deemed the meta-analysis conclusive once the cumulative Z-statistic crossed the statistical monitoring boundaries constructed with TSA. Because we performed meta-analysis information size calculations to achieve both moderate (alpha = 5% and beta = 20%) and strong (alpha = 1% and beta = 10%) evidence and constructed the statistical monitoring boundaries separately for each degree of evidence, we made similar distinctions about the degree of conclusiveness that can be achieved from adding a new RCT. Assuming that the control group event rate in the new RCT will be 5%, and that the RCT will yield a 25% relative risk reduction in TB, we approximated the required sample size for a future RCT in order to make the meta-analysis moderately and strongly conclusive. In technical terms, we approximated the sample size of a future RCT required for the cumulative meta-analysis Z-statistic to cross the TSA monitoring boundaries demonstrating moderate evidence and the TSA monitoring boundaries demonstrating strong evidence.

Results

In our initial review, we identified 25 articles that potentially fit our study criteria. Upon in-depth review, we excluded 16 trials from our analysis. Four RCTs did not meet our eligibility criteria since they did not include a placebo group.[37–40] We excluded an additional study on the basis that it examined outcomes of isoniazid prophylaxis on HIV-positive children.[47] We excluded an RCT of IHZ with placebo control because it did not disaggregate outcomes of TB incidence by treatment group,[48] and another because it evaluated only PPD+ patients.[49] We excluded a study because it focused on the impact of IHZ on recurrent incidence of TB.[33] Two manuscripts were long-term follow-up studies of RCTs already identified by our review.[50,51] We excluded three conference abstracts on the basis that study findings have since been published in full text.[34,52,53] Finally, we excluded two additional conference posters and one manuscript since they were cohort studies.[54–56] We included nine trials in our meta-analysis, comprising a total of 2,911 trial participants ($\Phi = 0.9$, see Table 1). A total of 1,526 study participants received INH and 1,385 placebo. Figure 1 presents the meta-analysis forest plot. The meta-analysis relative risk is 0.74 (95% CI, 0.53–1.04, $P = 0.082$, $I^2 = 0\%$, $P = 0.69$).

Trial sequential analysis

We estimated that a meta-analysis information size of 10,508 patients was required to yield moderate evidence, and that a meta-analysis information size of 19,920 patients was required to yield strong evidence. From our current analysis, the cumulative Z-statistic did not cross the TSA monitoring boundaries for moderate or strong evidence (Figure 2). This suggests that the pooled meta-analytic evidence of effectiveness is neither reliable nor definitive.

Topping up a sample size evaluation

We calculated that an additional 3,800 patients would need to be randomized (1,900 randomized to IHZ, and 1,900 randomized to placebo) for the meta-analysis to yield moderate evidence of a 25% relative risk reduction. In technical terms, we calculated that a new trial showing a 5% control group incidence rate and a 25% relative risk reduction would have to include 3,800 patients for the meta-analysis to cross the monitoring boundaries for moderate evidence (ie, the monitoring boundaries based on the information size for moderate evidence, 10,508 patients). Figure 3 displays the scenario where such a trial has been added to the meta-analysis. Adding this trial to the meta-analysis would yield a meta-analysis information size of 6,711. The meta-analysis could thus be considered moderately conclusive 3,979 patients before reaching its required information size of 10,508.

We calculated that an additional 9000 patients would need to be randomized for the meta-analysis to yield strong evidence of a 25% relative risk reduction. In technical terms, we calculated that a new trial showing a 5% control group incidence rate and a 25% relative risk reduction would have to include 9000 patients for the meta-analysis to cross the

Table 1 Study characteristic and outcomes for patients testing purified protein derivative – negative (PPD–)

Author, year	Trial location	Mean age (years)	Trial duration (months)	Trial follow-up (years)	Drug regimen (active)	Drug regimen (control)	No. patients (active)	No. patients (control)	TB incidence (active)	TB incidence (control)
Mohammed, 2007	South Africa	39	12	NA	INH 900 or 800 mg twice weekly	Placebo	48	50	9	6
Rivero, 2003	Spain	32	6	2	INH 5 mg/kg to a max of 300 mg daily	Placebo	83	77	3	4
Fitzgerald, 2001	Haiti	32	12	2.5	INH 300 mg daily + Vit B_6 50 mg	Placebo + Vit B_6 50 mg	126	111	6	4
Mwinga, 1998	Zambia	30	6	1.6	INH 300 mg twice weekly	Placebo	179	166	7	8
Gordin, 1997	USA	38	6	2.8	INH 300 mg + Vit B_6 50 mg	Placebo	260	257	3	6
Hawken, 1997	Kenya	31	6	1.8	INH 300 mg	Placebo	67	69	7	10
Whalen, 1997	Uganda	30	6	1.3 (A) 1.0 (C)	INH 300 mg	Placebo	393	323	9	10
Pape, 1993	Haiti	31	12	2.5	INH 300 mg +Vit B_6 50 mg	+Vit B_6 50 mg	20	35	2	5
Wadhawan, 1993	Kenya	NA	6	1.7	INH 300 mg	Placebo	350	297	11	21

Abbreviations: NA, not applicable; INH, isoniazid; TB, tubercolosis.

monitoring boundaries for strong evidence (ie, the monitoring boundaries based on the information size for strong evidence, 19,920 patients). Figure 4 displays the scenario where such a trial has been added to the meta-analysis. Adding this trial to the meta-analysis would yield a meta-analysis information size of 11,911. The meta-analysis could thus be considered moderately conclusive 8,009 patients before reaching its required information size of 19,920.

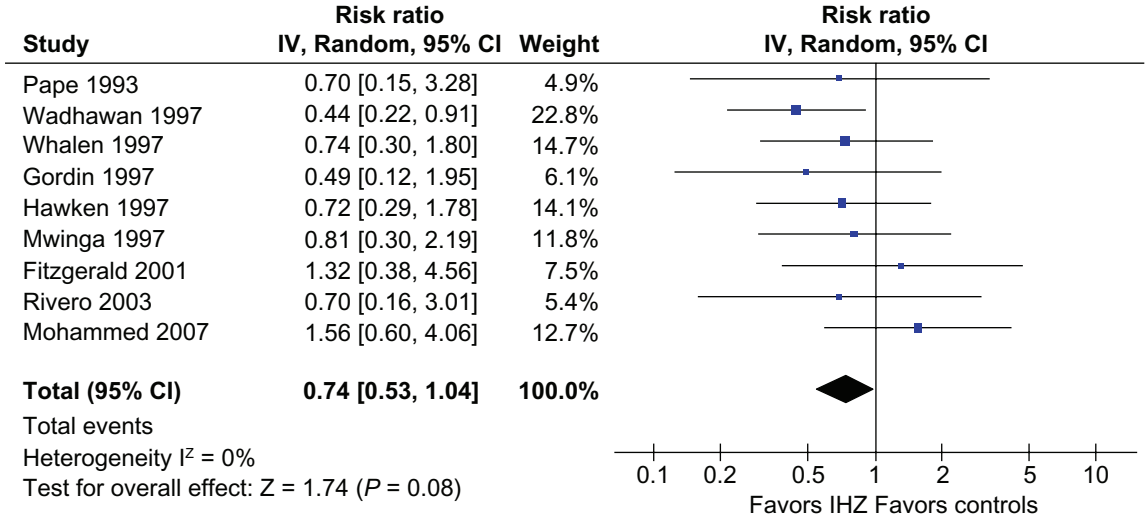

Figure 1 Forest plot of the individual trial RR estimates and 95% CIs, and the DerSimonian-Laird pooled RR estimate.
Notes: RR estimates smaller than 1.00 favor use of isoniazid chemoprophylaxis (IHZ) for prevention of tuberculosis in purified protein derivative negative HIV-infected individuals, RR estimates larger than 1.00 favor controls.

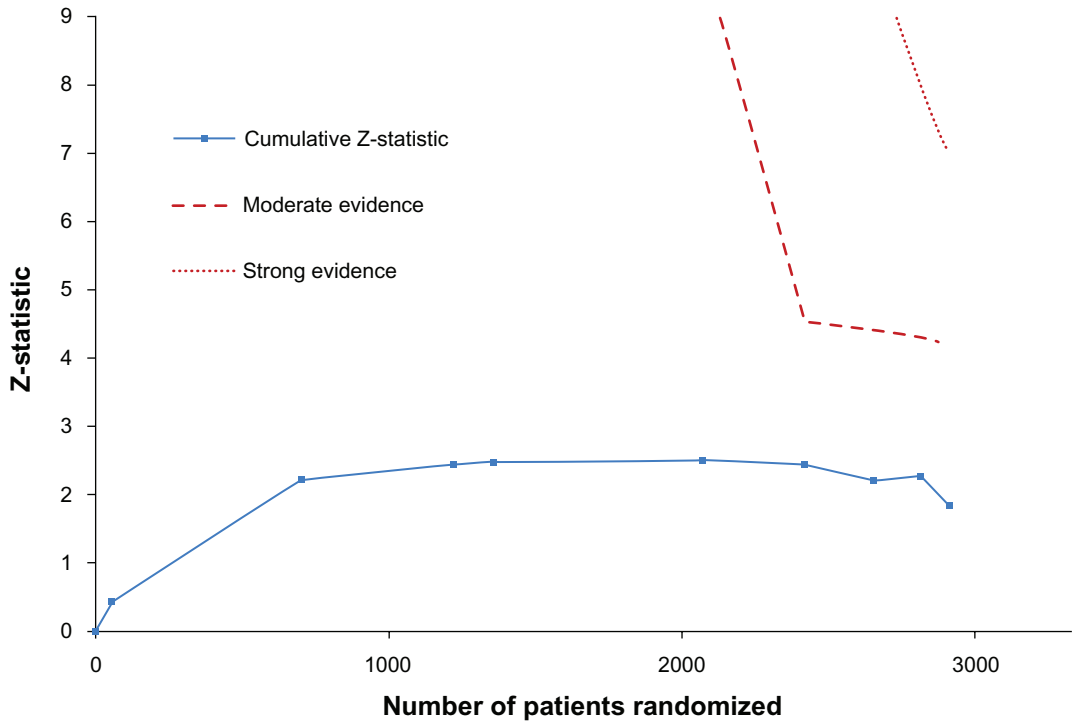

Figure 2 Trial sequential analysis (TSA) monitoring boundaries for moderate and strong evidence (the downward sloping lines) in the current meta-analysis.
Notes: The monitoring boundaries were constructed using the Lan-DeMets alpha-spending approach corresponding to the O'Brien-Fleming boundaries. The monitoring boundaries for moderate evidence (alpha = 5%, beta = 20%) were based on information size required for moderate evidence (10,508 patients). The monitoring boundaries for strong evidence (alpha = 1%, beta = 10%) were based on information size required for strong evidence (19,920 patients). Both information size calculations were based on the assumption of a 5% incidence rate in the control group, and a 25% relative risk reduction. The cumulative Z-statistic does not cross above either of the two monitoring boundaries, and thus, there is neither moderate or strong statistical evidence (significance) to support a treatment effect of at least a 25% relative risk reduction.

Discussion

Our study examined a public health and clinically important topic that was able to include data from 9 RCTs. Despite a relative consistency of treatment effect across the trials, our meta-analysis is both inconclusive and lacking information to adequately guide clinical decision-making. Based on our application of the TSA, we find that further trials including approximately 3,800 participants at a similar risk would be required to provide moderate evidence, and much more to provide strong evidence. To the extent that our example analysis applies to other meta-analyses, including Cochrane reviews, many apparently conclusive meta-analyses may in fact provide indefinite answers.[8,12,13]

Trial sequential analysis represents one of several new developments in interpreting the utility of existing meta-analyses and other forms of evidence for clinical and policy decision-making. The GRADE Working Group, a guideline development panel that has created the GRADE profiler tool for inferring the quality of evidence, has recently included precision as one of five general components which determine the quality of evidence.[3] As a recent GRADE publication points out, inferring the precision of an intervention is based on more than the CIs of a study result and we should infer the plausibility of differing treatment effects according to precision and consistency of treatment effects.[3]

In addition to trial sequential analysis of the current meta-analysis data, our analysis provides an inference on the number of patients that would be required to be further randomized to provide stronger evidence. We refer to this as 'topping-up' a sample size calculation, whereby if we plan a new clinical trial, we approximate the new number of patients required, with a similar risk profile, that must be randomized to create moderate quality meta-analytic evidence. In this example, it is 3,800 patients.

As with any analysis, there are strengths and limitations to consider. Our analysis is based on available data on an important public-health topic. We based our analysis on data from real RCTs and aimed to reduce publication bias and extraction of data bias by extensive searching and double data extraction. It is possible that other trials exist that we are unaware of. Furthermore, we distinctly chose this example of IHZ for TB prevention in HIV+ patients

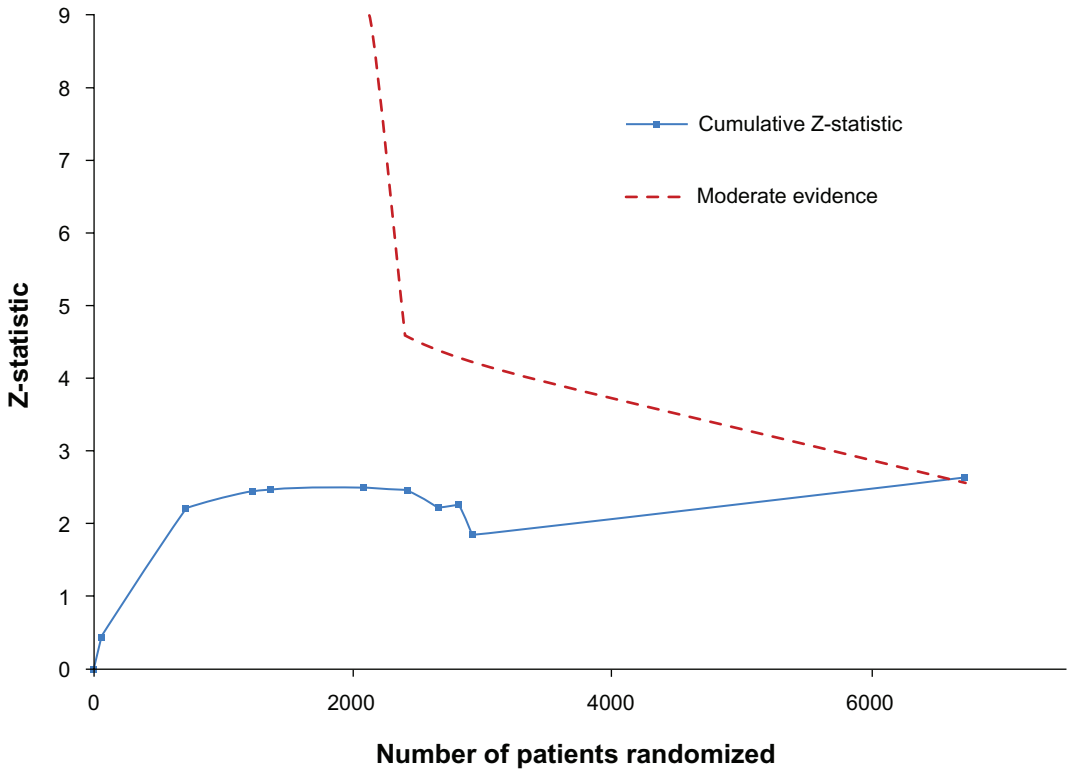

Figure 3 Trial sequential analysis (TSA) monitoring boundaries for moderate evidence (the downward sloping line) after adding a 'fake trial' with 3800 patients.
Notes: The monitoring boundaries for moderate evidence (alpha = 5%, beta = 20%) were constructed using the Lan-DeMets alpha-spending approach corresponding to the O'Brien-Fleming boundaries, and were based on the information size required for moderate evidence (10,508 patients). The information size calculation was based on the assumption of a 5% incidence rate in the control group, and a 25% relative risk reduction. The cumulative Z-statistic crossed the monitoring boundaries for moderate evidence after adding a 'fake trial' with an intervention effect estimate of a 25% relative risk reduction, a 5% control group incidence rate and 3800 patients (1900 in each group). This result demonstrates that under the event rate assumptions made for the information size calculations, a new trial would need to include 3800 patients (1900 in each group) to make the meta-analysis conclusive with moderate statistical support.

because we were aware that event rates were small and that this important public-health topic remains unanswered. It is possible that if we had chosen many other topics, we would not be able to demonstrate insufficiency, and hence this study is used as an example.

Furthermore, as with conventional sample size calculations for RCTs, our analyses are only reliable to the extent that the projected control group incidence rate and intervention effect is a good approximation of the "truth". Our projected control group incidence rate and intervention effect, reflect what we might realistically expect, given the current data. However, some future RCTs may include design features (high or low methodological quality, high or low risk population groups, variants of the intervention, etc.) which will render a higher or lower control group incidence rate and intervention effect. Deviations from the projected event rates may also occur due to the play of chance. As a result, a future RCT including the projected 3,800 patients will, in some situations, not be enough to make the meta-analysis conclusive (ie, cross the monitoring boundaries for moderate evidence). We are, however, not too concerned with this issue. Firstly, in clinical research we never know the "truth", and the uncertainty associated with any sample size estimation is simply an inherent and unavoidable part of the cumulative nature of science. Secondly, even in the situation where some future RCT does not make the meta-analysis conclusive according to the thresholds set by the monitoring boundaries, it will still provide valuable information to the existing body of evidence, and thus, allow for better informed policy making and planning of future research.

The role of meta-analysis in informing clinical and policy decision-making has received a tremendous amount of attention. Much less so, has been the adequate statistical interpretation of these analyses. Meta-analysis, particularly when including a small number of trials may overestimate the actual treatment effects of an intervention or yield false positive results.[8,9] We have known for a considerable time that small trials may yield exaggerated treatment effects and spuriously small *P*-values compared to larger trials or

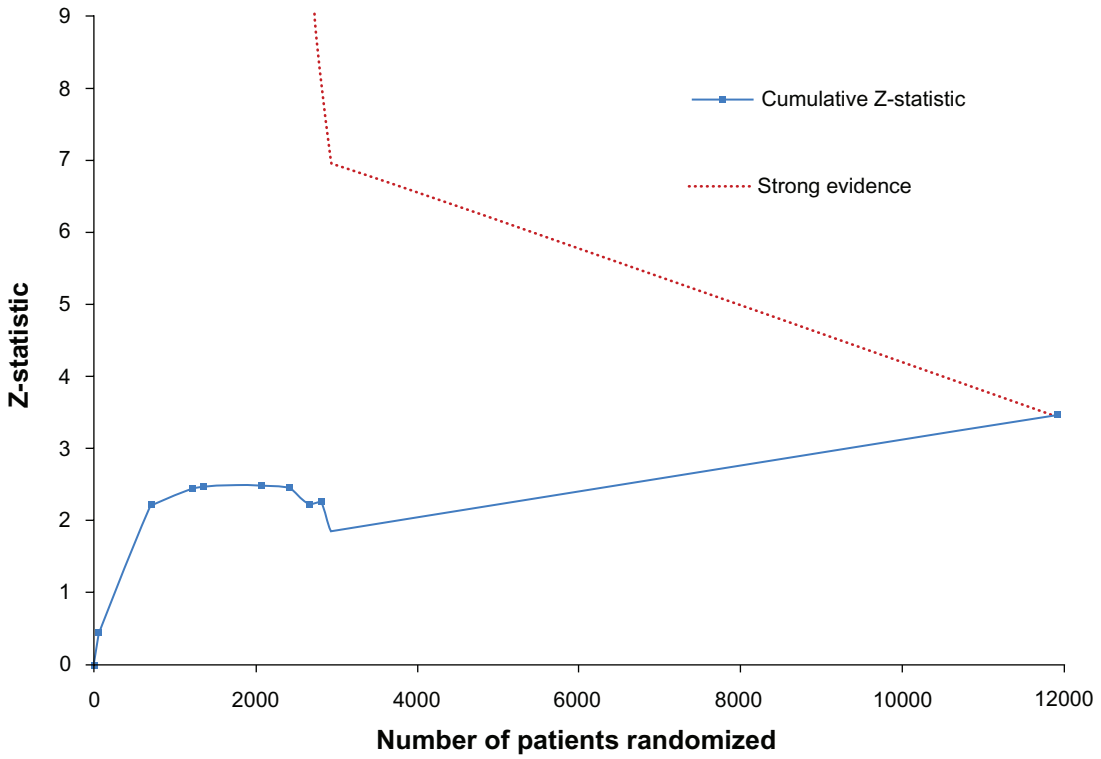

Figure 4 Trial sequential analysis (TSA) monitoring boundaries for strong evidence (the downward sloping line) after adding a 'fake trial' with 9000 patients.
Notes: The monitoring boundaries for moderate evidence (alpha = 1%, beta = 10%) were constructed using the Lan-DeMets alpha-spending approach corresponding to the O'Brien-Fleming boundaries, and were based on the information size required for moderate evidence (19,920 patients). The information size calculation was based on the assumption of a 5% incidence rate in the control group, and a 25% relative risk reduction. The cumulative Z-statistic crossed the monitoring boundaries for moderate evidence after adding a 'fake trial' with an intervention effect estimate of a 25% relative risk reduction, a 5% control group incidence rate and 9000 patients (4500 in each group). This result demonstrates that under the event rate assumptions made for the information size calculations, a new trial would need to include 9000 patients (4500 in each group) to make the meta-analysis conclusive with strong statistical support.

multiple trials pooled in a meta-analysis. Given the status that meta-analysis may receive in decision-making, falsely large treatment effects and spuriously small *P*-values may inhibit subsequent trials on the specific questions. This viewpoint suggests skepticism in interpreting small meta-analyses. Examples of meta-analyses generating falsely conclusive findings that are then refuted by subsequent larger trials are common.[4,8,9,24]

This is now an important time to refine our tools for inferring the quality and necessary conduct of these analyses. Recent reporting guidelines for meta-analysis, such as the PRISMA guidelines, make no reference to the adequacy of power and precision; and organizations such as The Cochrane Collaboration have had no written guidance on this issue.[57] Given the small number of included trials within many Cochrane reviews this represents an important rallying opportunity for the conduct of future RCTs. Given the past emphasis on sample size calculations of individual trials that many organizations place on investigators, the number of patients and events should be given much greater consideration in meta-analysis.

Disclosures

The authors report no conflicts of interest relevant to this research.

References

1. Guyatt G, Haynes RB, Jaeschke RZ, et al. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. *JAMA*. 2002;284:1290–1296.
2. Higgins JP, Green S. *Cochrane Handbook for systematic reviews of interventions, version 5.0.0*. Oxford: John Wiley & Sons; 2009.
3. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336:924–926.
4. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *NEJM*. 1997;337:536–542.
5. Moses LE, Mosteller F, Buehler JH. Comparing results of large clinical trials to those of meta-analysis. *Stat Med*. 2002;21:793–800.
6. Pogue J, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Contr Clin Trials*. 1997;18:580–593.
7. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. *Lancet*. 1998;351:47–52.
8. Thorlund K, Devereaux PJ, Wetterslev J, et al. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? *Int J Epidemiol*. 2009;38:276–286.

9. Trikalinos TA, Churchill R, Ferri M, et al. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *J Clin Epidemiol*. 2004;57:1124–1130.
10. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol*. 2008;61:64–75.
11. Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. *BMC Med Res Methodol*. 2009;9:86.
12. Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. *J Clin Epidemiol*. 2008;61:763–769.
13. Brok J, Thorlund K, Wetterslev J, Gluud C. Apparently conclusive meta-analyses may be inconclusive – Trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. *Int J Epidemiol*. 2009;38:287–298.
14. Keech AC, Gebski V. Managing the resource demands of a large sample size in clinical trials: can you succeed with fewer subjects? *MJA*. 2002;177:445–447.
15. Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA*. 2002;288: 358–362.
16. Guyatt G, Mills E. In the era of systematic reviews, does the size of an individual trial still matter. *PLoS Med*. 2008;5(1):e4.
17. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001;357:1191–1194.
18. Tricco AC, Tetzlaff J, Pham B, Brehaut J, Moher D. Non-Cochrane vs Cochrane reviews were twice as likely to have positive conclusion statements: cross-sectional study. *Journal of Clinical Epidemiology*. 2009;62:380–386.
19. Devereaux PJ, Beattie WS, Choi PT, et al. How strong is the evidence for the use of perioperative beta blockers in non-cardiac surgery? Systematic review and meta-analysis of randomised controlled trials. *BMJ*. 2005;331:313–321.
20. Pocock SJ. When to stop a clinical trial. *BMJ*. 1992;305:235–240.
21. DeMets D, Lan KK. Interim analysis: the alpha spending function approach. *Stat Med*. 1994;12:1341–1352.
22. Lan KK, DeMets D. Discrete sequential monitoring boundaries for clinical trials. *Biometrika*. 1983;659–663.
23. Afshari A, Wetterslev J, Brok J, Moller A. Antithrombin III in critically ill patients: systematic review with meta-analysis and trial sequential analysis. *BMJ*. 2007;335:1248–1251.
24. Bangalore S, Wetterslev J, Pranesh S, Sawhney S, Gluud C, Messerli FH. Perioperative beta blockers in patients having non-cardiac surgery: a meta-analysis. *Lancet*. 2008;372:1962–1976.
25. Berkey C, Mosteller F, Lau J. Uncertainty of the time of first sifnificance in random-effects cumulative meta-analysis. *Contr Clin Trials*. 1996;17:357–371.
26. Gehr B, Weiss C, Porzsolt F. The fading of reported effectiveness. A meta-analysis of randomised controlled trials. *BMC Med Res Methodol*. 2006;6:25.
27. Hu M, Cappeleri J, Lan KK. Applying the law of the iterated logarithm to control type I error in cumulative meta-analysis of binary outcomes. *Clin Trials*. 2007;4:329–340.
28. Lan KK, Hu M, Cappelieri J. Applying the law of the iterated logarithm to cumulative meta-analysis of a continuous endpoint. *Stat Sinica*. 2003;13:1135–1145.
29. Sutton AJ, Cooper NJ, Jones DR. Evidence synthesis as the key to more coherent and efficient research. *BMC Med Res Methodol*. 2009;9.
30. Bucher HC GLGGS. Isoniazid prophylaxis for tuberculosis in HIV infection: a meta-analysis of randomized controlled trials. *AIDS*. 1999;13:501–507.
31. Wilkinson D. Effect of preventive treatment for tuberculosis in adults infected with HIV: systematic review of randomised placebo controlled trials. *BMJ*. 1998;317:625–629.
32. Woldehanna S V. Treatment of latent tuberculosis infection in HIV infected persons (Review). *Cochrane Database Syst Rev*. 2004;CD000171.
33. Fitzgerald DW DM. Effect of post-treatment isoniazid on prevention of recurrent tuberculosis in HIV-1-infected individuals: a randomised trial. *Lancet*. 2000;356:1470–1474.
34. Mohammed AA ML. Randomized controlled trial of isoniazid preventive therapy administered by patient-nominated supervisors, in advanced HIV disease. 2006 – XVI International AIDS Conference: Abstract no. MOPE0185. 2006.
35. Rivero A. Randomized trial of three regimens to prevent tuberculosis in HIV-infected patients with anergy. *Enferm Infecc Microbiol Clin*. 2003;21:287–292.
36. World Health Organization Global Tuberculosis Control. WHO Report. Available at: http//www.who/int/tb/publications/global_report/2008/pdf. Accessed on: June 22, 2008.
37. Grant AD CSFKDJCECRDCKHRCG. Effect of routine isoniazid preventive therapy on tuberculosis incidence among HIV-infected men in South Africa: a novel randomized incremental recruitment study. *JAMA*. 2005,293:2719–2725.
38. Halsey NA CJD. Randomised trial of isoniazid versus rifampicin and pyrazinamide for prevention of tuberculosis in HIV-1 infection. *Lancet*. 1998;351:786–792.
39. Matteelli A. Multinational study for preventive therapy of latent tuberculosis in HIV positive individuals: isoniazid vs rifabutin. *12th AIDS Conference*. 1998;Abstract: BO7–B11.
40. Rivero A. Randomized clinical trial investigating three chemoprophylaxis regimens for latent tuberculosis infection in HIV-infected patients. *Enferm Infecc Microbiol Clin*. 2007;25:305–310.
41. Chintu C. Tuberculosis in children with human immunodeficiency virus infection. *Int J Tuberc Lung Dis*. 2005;9:477–484.
42. Meade MO, Guyatt GH, Cook RJ, et al. Agreement between alternative classifications of acute respiratory distress syndrome. *AJRCCM*. 2001;163:490–493.
43. DerSimonian L, Laird N. Meta-analysis in clinical trials. *Contr Clin Trials*. 1986;7:177–188.
44. Higgins JP, Thompson S. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21:1539–1558.
45. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet*. 2002;359:519.
46. Jennison C, Turnbull B. *Group sequential methods with applications to clinical trials*. Boca Raton, USA: Chapman&Hall/CRC Press; 2000.
47. Zar HJ CMSSK. Effect of isoniazid prophylaxis on mortality and incidence of tuberculosis in children with HIV: randomised controlled trial. *BMJ*. 2007;334:136.
48. Lim HJ OA. Effect of tuberculosis preventive therapy on HIV disease progression and survival in HIV-infected adults. *HIV Clin Trials*. 2006;7:172–183.
49. Valdespino JL GMDEZ. Outcomes of the pilot study of TB chemoprophylaxis trials. *Int Conf AIDS.* 1993; Abstract no PO-C35-3387.
50. Johnson JL OA. Duration of efficacy of treatment of latent tuberculosis infection in HIV-infected adults. *AIDS*. 2001;15:2137–2147.
51. Quigley MA. Long-term effect of preventive therapy for tuberculosis in a cohort of HIV-infected Zambian adults. *AIDS*. 2001;15:215–222.
52. Rivero A. A randomized trial of 3 regimens to prevent tuberculosis in anergic patients infected with HIV. *9th Conf of Retroviruses and Opportunistic Infection (CROI)*. 2002;Session 82:(Poster):625W.
53. Whalen C. Preventive Therapy for Tuberculosis in HIV-infected Ugandans. 1997 4th Conference on Retroviruses and Opporutnistic Infections. Abstract no. 676.
54. Gourevitch MN, Hartel D, Selwyn PA, Schoenbaum EE, Klein RS. Effectiveness of isoniazid chemoprophylaxis for HIV-infected drug users at high risk for active tuberculosis. *AIDS*. 1999;13:2069–2074.
55. Golub JE SV. Tuberculosis (TB) incidence by HAART and isoniazid prophylactic therapy (IPT) in HIV-infected patients in Rio de Janeiro, Brazil. 2006 AIDS – XVI International AIDS Conference" Abstract no. MOPE0395.

56. Golub J. The effectiveness of HAART and isoniazid preventive treatment (IPT) in preventing TB in HIV-infected adults – a prospective cohort from two South Africa sites. 2007 Poster exhibition: 4th IAS Conference on HIV Pathegenesis, Treatment and Prevention: Abstract no. MOPEB025.

57. Liberatti A, Altman D, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339: DOI: 10.1136/bmj.b2700.