

Discussion

Mega-trials vs. meta-analysis: Precision vs. heterogeneity?

Ian Shrier^{a,*}, Robert W. Platt^b, Russell J. Steele^c

^a Centre for Clinical Epidemiology and Community Studies, SMBD-Jewish General Hospital, McGill University, 3755 Cote Ste-Catherine Rd, Montreal, Canada Qc H3T 1E2

^b Department of Epidemiology and Biostatistics and the Department of Pediatrics, McGill University, Canada

^c Department of Mathematics and Statistics, McGill University, Canada

Received 28 April 2006; accepted 12 November 2006

Abstract

In recent years, several authors have suggested there is a need for more very large or “mega-trials” (defined in this manuscript as a trial powered to address subgroup differences/interactions/secondary analyses) to answer important clinical questions. Because mega-trials are expensive and funding for clinical research is limited, increasing the number of mega-trials limits funding for other research. The advantages of this approach compared with funding more focused RCTs needs to be debated. Because there is no method to determine gold standard for which method gives the correct answer, we provide theoretical arguments that demonstrate that the two approaches are similar with respect to sample size requirements and the mega-trial approach provides a small advantage with respect to minimizing confounding by chance. However, the inherent heterogeneity in a series of smaller trials may represent a significant advantage over a single mega-trial.

© 2006 Published by Elsevier Inc.

Keywords: Mega trial; Meta-analysis

1. Manuscript

Empirical data suggest that the results of meta-analyses from multiple smaller trials are *usually* consistent with the results of a mega-trial (defined as powered to address subgroup differences/interactions/secondary analyses; usually considered more generalizable) [1]. Although the results may usually be consistent, several authors have suggested that the mega-trial approach is inherently superior to smaller randomized trials (even when the results of the latter are pooled together in a systematic review or “multi-trial approach”) and the results from the mega-trial should be believed when the two methods disagree [2,3]. The implications of this philosophy may affect the distribution of funds from granting agencies. If several smaller trials produce conflicting results, this philosophy leads to the conclusion that a mega-trial should be conducted to definitively answer the question across a large sample of the population. An alternative is to fund several trials that use slightly different populations and different methodologies, each with the opportunity to build on the information obtained in previous studies. Because resources are limited, it is unlikely that granting agencies will fund both approaches.

* Corresponding author. Tel.: +1 514 340 7563; fax: +1 514 340 7564.

E-mail address: ian.shrier@mcgill.ca (I. Shrier).

The arguments in favour of funding the mega-trial approach are that 1) randomization is more likely to provide equal distributions for known and unknown confounders, 2) the larger sample size leads to a more precise estimate of the main effect, and 3) the study can be powered to test important subgroup comparisons [3]. One of the better-known examples concerned the efficacy of magnesium in the early post-myocardial infarction period. Although smaller randomized trials suggested magnesium might be beneficial, some authors dismissed these results in favour of the ISIS-4 mega-trial that suggested magnesium was ineffective [3].

However, there are also advantages to continue to fund multiple smaller trials that would lead to a meta-analysis at a later date instead (Table 1). We believe that these advantages may outweigh those of the mega-trial. Because it is impossible to know whether the mega-trial or a meta-analysis (based on the multi-trial approach) provides an effect estimate closest to the population parameter of interest when the two methods disagree, one is left with only theoretical arguments. Therefore, the relative advantages and disadvantages of the two different approaches need to be debated openly and fully.

This paper is limited to a discussion that compares well-done smaller trials with well-done mega-trials; well-done trials of any size are superior to low quality trials of any size. Also, all trials should have the appropriate power to detect clinically important differences. Finally, our discussion applies to both large simple trials [4], and trials with restricted inclusion criteria. These two types of trials answer different questions and both types of questions can be answered through either approach.

2. Sample size calculations

The number of subjects required to answer research questions in the two approaches is similar. Let us assume that the 2-year mortality for a disease in untreated persons will be 15% and that treatment will reduce mortality by 50%. Setting $\alpha=0.05$ and $\text{power}=0.8$, and using an intention-to-treat analysis, we would need 304 subjects per group for the main effect. In addition, we expect a loss to follow-up of 10%, 80% compliance and we would like to examine compliance-based effects in addition to our intention-to-treat analysis. Therefore, we estimate the sample size to be 400 per group. This is true if we compare the effect in men versus women, or in those <60 versus 60+. In the multi-trial approach, one could run two different studies for a total of 1600 subjects. Alternatively, we could do one study examining both questions, and the sample size is simply 1600 subjects. If one were interested in interactions between sex and age, the sample size for both approaches would have to be increased. In fact, one could argue that the mega-trial would require slightly more subjects because it includes multiple hypotheses.

Table 1

Summary of advantages and disadvantages to conducting a future mega-trial versus future multiple smaller trials when current evidence from trials is conflicting

Sample size	Mega-trial may have slightly higher total sample size required because of multiple hypothesis testing
Confounding by chance	Mega-trial is more likely to have equal distributions of covariates for the entire population compared to any one smaller trial, but not compared to a meta-analysis based on many smaller trials. Further, subgroups in the mega-trial are just as likely to have unequal distributions if the sample size of the subgroup is the same as the sample size of a smaller trial examining only the effect of the subgroup.
Precision	Mega-trial and meta-analysis of multiple smaller trials will have the same precision if total sample size is the same and patient-level data is available. Even if no patient-level data is provided, the precision for any subgroup comparison is the same as that of a smaller trial examining only the effect of the subgroup. It may or may not be easier to obtain patient-level data from one mega-trial compared to several smaller trials.
Heterogeneity	
Time-independent differences	Mega-trials are normally designed so that each centre uses the exact same methodology. However, there will be regional differences in quality of personnel, demographics of subjects recruited, prognosis, etc. These subgroup differences are normally hidden within a mega-trial cluster analysis, or ignored altogether if no cluster analysis is performed. The differences are transparent in the multi-trial approach.
Time-dependent differences	By definition, a mega-trial is conducted at one point in time. The interaction between the intervention and changes to standard medical therapy (or population demographics, etc) cannot be investigated. Therefore, it is not possible to have one definitive trial for any intervention.

3. Confounding by chance

The mega-trial approach has some advantage over the multi-trial approach with respect to adequacy of randomization. As the number of subjects increases, the probability that an important confounder will be unequally distributed between the two comparison groups decreases. Although often taken for granted, it is possible to calculate the expected distributions. In brief, if the prevalence of the only important potential confounder for the study is 20% and we randomize 400 subjects into 2 groups, there is approximately 95% probability that the prevalence of the confounder in one group will lie between $1.96 \times \text{SE}$ of 20% [5], i.e. the “sampling interval” is between 15.6% and 24.4% (Note: if the prevalence is 16% in one group, the prevalence must be 24% in the other group if the two groups are of equal size). Therefore, there is a 5% probability of at least a 10% difference between groups for the important confounder. For a trial of 1600 subjects there would be 800 subjects in each group and the corresponding interval is 17.8% and 22.2%. As the number of important confounders increases, the probability that there is an unequal distribution of any one confounder increases if we assume the confounders are relatively independent of one another. For example, if there are 5 important and independent confounders each with 20% prevalence, there is a 23% probability that the distribution for at least one of the five confounders will lie outside these marginal 95% sampling intervals (i.e. $1 - 0.95^5$). To obtain a 95% sampling region for all 5 confounders, one can use the 99% sampling interval for a single confounder (i.e. $0.99^5 = 0.95$); the sampling interval for the trial with 200 subjects per group is 14.2% to 25.8%. In the mega-trial, the sampling interval is between 17.1% and 22.9%. This simply confirms that confounders are much more likely to be equally distributed in larger trials.

Although smaller trials are more likely to have unequal distributions, the multi-trial approach uses many trials and therefore one expects the unequal distribution of confounders in one study to be offset by the unequal distribution of confounders from a different study; the distribution of confounders should average out over the collection of trials. Therefore, a meta-analysis based on the unadjusted data is expected to provide an unbiased result.

4. Precision vs. heterogeneity

Because variance is related to the size of the sample under study, mega-trials provide a very precise estimate of the overall effect (e.g. across sex and age). The multi-trial approach provides the same precision if the studies used homogeneous methodology [3]. In both cases, one should include intra- and inter-centre variability in the analysis. In essence, a multi-centre mega-trial is a collection of many smaller studies performed within different institutions at the same time with the same protocol. Even if one is interested in subgroup comparisons, the multi-trial approach provides the same level of analysis if one has access to the individual patient data (although this is often difficult at the current time, we believe that authors of publicly funded studies should be required to make their data publicly available once their primary analyses are published), or if one or more studies directly tested the subgroup comparison. If the subgroup comparison hypothesis is developed after the study has been conducted, one needs access to the raw data in both types of approaches.

Practically speaking, a meta-analysis of multiple trials has slightly reduced power compared to a mega-trial due to increased heterogeneity between trials; it is unreasonable to assume that all studies within the multi-trial approach will use the exact same methodology for inclusion criteria (e.g. age, co-morbidity, disease severity), intervention (e.g. dose or timing of a drug; frequency or intensity or duration or type of exercise, etc), outcome measure (e.g. quality of life measured with the SF-36 vs. Euro-QOL), timing of the outcome (e.g. 1-month mortality, 1-year mortality), etc. Although this results in a decrease in statistical power, we believe the methodological heterogeneity may represent an overall strength of the multi-trial approach. If the results of the individual small trials are all consistent, there is limited heterogeneity in the results and a meta-analysis should yield the same answer as a mega-trial. The important situation is when at least one of the multiple smaller trials has results that represent clinically important differences from the other studies. In this situation, confusion arises [3] and there are two possibilities: the heterogeneity is due to chance because of fewer subjects, or it is due to methodological differences between studies. With respect to chance, this possibility also exists in the mega-trial because as above, the mega-trial is conceptually identical to multiple smaller trials performed in different institutions at the same time with the same protocol. Whereas truly random variation in results between studies in the multi-trial approach is presented in the results of a meta-analysis, the random variation between centres in a mega-trial is hidden if an overall effect estimate is presented from a cluster analysis, or is ignored altogether if a cluster analysis is not performed. Therefore, the mega-trial approach is different because it is less

transparent (a potential weakness of the reporting of mega-trials), and because it eliminates methodological heterogeneity as a factor. Within this context, the preference for a mega-trial or multi-trial approach must be related to whether methodological heterogeneity is considered beneficial or detrimental.

We believe methodological heterogeneity is beneficial in most situations. First, we agree that caution is advised for subgroup analyses in systematic reviews [6,7]. Second, subgroup analyses based on a priori protocols can be tested statistically, but this is inappropriate when the subgroup analyses are developed after data exploration. That said, an exploratory analysis of heterogeneity is important [7,8]. Often, patient care cannot follow the protocols from research studies and it is essential to understand how and why differences in methodology might affect the results when the treatment is used in routine clinical practice. For example, methodological differences may be related to population differences, dose of drug, patient compliance, etc. Further, the practice of medicine is continually changing. Therefore, the results of any trial are time-dependent; we do not know how the effect of the treatment will change under other conditions. The following example illustrates the usefulness of embracing heterogeneity.

5. Time-dependent/population differences

The first smaller trials of magnesium in the immediate post-myocardial infarction (post-MI) period suggested it was beneficial with respect to both mortality and arrhythmias [9–13]. Later studies questioned the benefit and the mega ISIS-4 trial eventually concluded that there was no effect [14]. This mega-trial has been cited as evidence for the need of mega-trials [3]. However, the sample size of the mega-trial was only one difference between the trials. The original trials were conducted in the mid-1980s and both patient behaviours and treatments have changed in the interim, resulting in different populations being studied. For example, both acetyl-salicylic acid (ASA) and thrombolytic agents were used in the ISIS trial. Given that basic science evidence suggests the effect of magnesium might be mediated through actions on platelets and/or thrombolysis [15–17], treating with magnesium when ASA and/or thrombolytic agents were already given would not be expected to have a major effect. Further, the mechanisms of action suggest magnesium would only be beneficial if given early in the post-MI period. Although the ISIS-4 trial provided some sub-analyses showing no effect in the absence of thrombolytic agents, these patients only received magnesium an average of 7 h post-MI. Finally, the trial also investigated the effects of captopril and mononitrates so that the large numbers of magnesium patients reflected those with and without these other medications. The actual number of magnesium only patients was 7000 per group, and there are no analyses of only this group with and without thrombolytic agents or based on time of administration. Although post-hoc suggestions, these hypotheses are based on the known biological mechanisms for magnesium. A full systematic review of the effects of magnesium in the post-MI period is beyond the scope of this article but the conclusion that the ISIS-4 mega-trial provided a definitive and correct answer reflects a pre-conceived notion that mega-trials are inherently superior; other equally plausible interpretations exist to explain the heterogeneity of results.

6. Implications for systematic reviews and meta-analyses

On theoretical reasoning, a random-effects model is more rational than a fixed-effects model for most meta-analyses because individual studies with methodological differences do not come from one overall population of studies but from a series of populations; one should account for the methodological variability between studies [18]. In the random-effects model, the inter-study variance is not sample-size dependent, and therefore a new study of 10,000 subjects contributes much less weight to inter-study variability than 10 studies of 1000 subjects. Further, repeatability is an important feature of all science; if something is done only once and there is an error, even a small error, one is stuck with the results. Whether the error is related to content (e.g. omitting a confounder, timing of drug administration), process, or is unknown/known at the time of the study, the effect is the same. In the multi-trial approach, any error in one trial is not likely to be duplicated in all other trials and can be factored out in a sensitivity analysis within a systematic review; there are no alternatives with a mega-trial approach.

If our proposal is adopted, situations may occur where trials are very different from each other and it will not be logical to combine the studies in one meta-analysis. As before, we believe this apparent weakness may also be a strength. First, although it is possible that different studies will measure different outcomes, the granting agencies can certainly require that all proposals for a particular competition include the granting agency's primary outcome. Second, if the different methodologies do not affect the efficacy of the intervention, there will not be any heterogeneity in the

results and the results could be combined. If the different methodologies do affect the efficacy, this is important information for the clinician to know.

The other implication is cost. At this point, it is not clear which approach is more expensive and it may depend on the particular nature of the trial(s). A mega-trial would save on fixed costs due to economies of scale (e.g. randomization, database management). However, there may be additional variable costs associated with running a study over the large geographical areas necessary to obtain the required sample sizes (e.g. travel and accommodations for investigator meetings, training of research associates, quality assurance).

7. Conclusion

Both the multi-trial and the mega-trial approaches allow for precise estimates of effect and subgroup analyses. The multi-trial approach provides additional information about heterogeneity between study methodologies and populations that might not be evident at the time of one mega-trial. If heterogeneity exists between smaller studies, a close examination of the methodologies may lead to new hypotheses — and these may not be possible with a mega-trial approach. Finally, if funding agencies and/or journals required the release of patient-level data for future meta-analyses as a pre-requisite for financial support (i.e. similar to the registration of trials that was recently adopted), then the multi-trial and mega-trial approach become equivalent with respect to the ability to conduct subgroup analyses or adjust for baseline imbalances of important covariates, but the multi-approach has the added advantage of methodological heterogeneity.

References

- [1] Cappelleri JC, Ioannidis JPA, Schmid CH, et al. Large trials vs meta-analysis of smaller trials. *JAMA* 1996;276:1332–8.
- [2] LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;337:536–42.
- [3] Egger M, Ibrahim S, Smith GD. Where now for meta-analysis? *Int J Epidemiol* 2002;31:1–5.
- [4] Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984;3:409–22.
- [5] Agresti A. *Categorical data analysis*. New York: John Wiley and Sons; 2002.
- [6] Smith GD, Egger M. Going beyond the grand mean: subgroup analysis in meta-analysis of randomised trials. In: Egger M, Smith GD, Altman DG, editors. *Systematic reviews in health care. Meta-analysis in context*. London: BMJ Publishing Group; 2001. p. 143–56.
- [7] Thompson SG. Why and how sources of heterogeneity should be investigated. In: Egger M, Smith GD, Altman DG, editors. *Systematic reviews in health care. Meta-analysis in context*. London: BMJ Publishing Group; 2001. p. 157–75.
- [8] Anello C, Fleiss JL. Exploratory or analytic meta-analysis: should we distinguish between them? *J Clin Epidemiol* 1995;48:109–16.
- [9] Morton BC, Nair RC, Smith FM, McKibbin TG, Poznanski WJ. Magnesium therapy in acute myocardial infarction—a double-blind study. *Magnesium* 1984;3:346–52.
- [10] Rasmussen HS, McNair P, Norregard P, Backer V, Lindeneg O, Balslev S. Intravenous magnesium in acute myocardial infarction. *Lancet* 1986;327:234–6.
- [11] Smith LF, Heagerty AM, Bing RF, Barnett DB. Intravenous infusion of magnesium sulphate after acute myocardial infarction: effects on arrhythmias and mortality. *Int J Cardiol* 1986;12:175–83.
- [12] Abraham AS, Rosenmann D, Kramer M, et al. Magnesium in the prevention of lethal arrhythmias in acute myocardial infarction. *Arch Intern Med* 1987;147:753–5.
- [13] Ceremuzynski L, Jurgiel R, Kulakowski P, Gebalska J. Threatening arrhythmias in acute myocardial infarction are prevented by intravenous magnesium sulfate. *Am Heart J* 1989;118:1333–4.
- [14] ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. *Lancet* 1995;345:669–85.
- [15] Adams JH, Mitchell JR. The effect of agents which modify platelet behaviour and of magnesium ions on thrombus formation in vivo. *Thromb Haemost* 1979;42:603–10.
- [16] Swartz WM, Brink RR, Buncke Jr HJ. Prevention of thrombosis in arterial and venous microanastomoses by using topical agents. *Plast Reconstr Surg* 1976;58:478–81.
- [17] Hughes A, Tonks RS. Magnesium, adenosine diphosphate and blood platelets. *Nature* 1966;210:106–7.
- [18] Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Smith GD, Altman DG, editors. *Systematic reviews in health care. Meta-analysis in context*. London: BMJ Publishing Group; 2001. p. 285–312.