

Comparison of Large Versus Smaller Randomized Trials for Mental Health-Related Interventions

Despina G. Contopoulos-Ioannidis, M.D.

Simon M. Gilbody, D.Phil., M.B.Ch.B.

Thomas A. Trikalinos, M.D.

Rachel Churchill, M.Sc., Ph.D.

Kristian Wahlbeck, M.D., Ph.D.

John P.A. Ioannidis, M.D.

The EU-PSI Project

Objective: The extent of disagreement between large and smaller randomized, controlled trials on mental health issues is unknown. The authors aimed to compare the results of large versus smaller trials on mental health-related interventions.

Method: The authors screened 161 Cochrane and 254 Database of Abstracts of Reviews of Effectiveness systematic reviews on mental health-related interventions. They identified 16 meta-analyses with at least one "large" randomized trial with sample size >800 and at least one "smaller" trial. Effect sizes were calculated separately for large and smaller trials. Heterogeneity was assessed between all studies, within each group (large and smaller studies), and between large and smaller studies.

Results: Significant between-study heterogeneity was seen in five meta-analyses. By random-effects calculations, the results of large and smaller trials differed beyond

chance in four meta-analyses (25%). In three of these disagreements (effect of day care on IQ, discontinuation of antidepressants, risperidone versus typical antipsychotics for schizophrenia), the smaller trials showed greater effect sizes than the large trials. The inverse was seen in one case (olanzapine versus typical antipsychotics for schizophrenia). With fixed-effects models, disagreements beyond chance occurred in five cases (31%). In four meta-analyses, the effect size differed over twofold between large and smaller trials. Various quality and design parameters were identified as potential explanations for some disagreements.

Conclusions: Large trials are uncommon in mental health. Their results are usually comparable with the results of smaller studies, but major disagreements do occur. Both large and smaller trials should be scrutinized as they offer a continuum of randomized evidence.

(*Am J Psychiatry* 2005; 162:578–584)

Large randomized, controlled trials are considered the gold standard for the evaluation of preventive or therapeutic interventions. However, often it is practically feasible to perform only relatively small randomized controlled trials. Several investigators have addressed whether the results of large trials agree with the results of smaller trials and meta-analyses thereof on the same topic. Villar et al. (1) reported moderate agreement between meta-analyses of small trials and large trials. Cappelleri et al. (2) concluded that the two methods usually agree. Conversely, LeLorier et al. (3) concluded that meta-analyses of small trials fail to predict the results of subsequent large trials in one-third of cases. A systematic comparison of all of these empirical assessments (4, 5) concluded that disagreements may be less prominent for primary outcomes and that, overall, the frequency of significant disagreements beyond chance is 10%–23%.

Prior investigations in the comparison of large versus smaller trials have typically not targeted mental health-related interventions. The vast majority of randomized controlled trials in mental health are of small sample size. It is important to understand how reliable these results are and whether we should expect large trials to confirm or to refute results from small trials. Moreover, the conduction

of randomized controlled trials for mental health interventions may have peculiarities in terms of types of interventions, subject recruitment, and selected outcomes. Thus, we decided to evaluate the comparative results of large and smaller randomized controlled trials on mental health-related interventions across a variety of topics.

Method

Eligibility Criteria and Databases

We identified meta-analyses for mental health-related interventions that included data on at least one large and at least one small randomized controlled trial on the same topic. Data had to be available per arm for each study for the primary outcome of the meta-analysis. Large randomized controlled trials were defined a priori on the basis of sample size. Specifically, large randomized controlled trials were considered as those having more than 800 patients total in the two compared arms. We used a more conservative cutoff than the 1,000-subjects cutoff that has been proposed previously (1–3) since trials with >1,000 subjects are rare in mental health, and thus, a more lenient sample size threshold might be preferable. The cutoff definition was made a priori strictly for operational purposes and the term "large" or "small(er)" do not have any connotation for the quality of the trials.

Meta-analyses of mental health-related interventions were identified through the Mental Health Library 2002 (issue 1, last searched September 2002), a systematic compilation of all Co-

chrane and Database of Abstracts of Reviews of Effectiveness (DARE) systematic reviews that pertain to mental health (6). A total of 161 Cochrane reviews and 254 DARE reviews were scrutinized. In case of overlapping reviews, only the most updated one with complete information was retained. Whenever a meta-analysis had included also controlled trials without any random assignment, these were excluded. Both binary and continuous outcomes were considered eligible for the analysis. Meta-analyses of binary outcomes had to provide or allow the calculation of the number of events and subjects in each arm. Meta-analyses of continuous outcomes had to provide or allow the calculation of the number of subjects and means and standard deviations in each arm.

Whenever a meta-analysis used different binary or continuous outcomes, we selected only the primary outcome as stated by each meta-analysis. Whenever the primary outcome was unclear, we selected the one with the largest number of studies. If several outcomes had the same number of studies, we selected the clinically most important one, as deemed independently by two authors. When a meta-analysis had available data for several interventions, each set of interventions was considered separately.

Data Extraction

For each trial, we extracted information on author, year of publication, intervention, primary outcome, as well as events per subject in each arm for binary outcomes and number of subjects and means and standard deviations in each arm for continuous outcomes. We also extracted information on items of reported quality that may be related to the effect size in randomized controlled trials (7, 8): double blinding (yes versus no), generation of randomization sequence (adequate versus inadequate or not stated), and adequate allocation concealment (adequate versus inadequate or not stated). Generation of the randomization sequence refers to the mode of randomization; adequate modes include computer-generated numbers, random-number tables, coin or dice toss, or other methods that ensure random order. Allocation concealment means that neither the experimenter nor the subject randomly assigned has any hint beforehand to which arm the subject is most likely to be assigned. This can be achieved by use of a central facility or central pharmacy or with sealed and opaque envelopes (7, 8). Two investigators extracted data independently, and consensus was reached on all items.

Statistical Analysis

For each eligible meta-analysis, we combined data from all studies and separated data from the large studies and those from the smaller studies. Between-study heterogeneity was evaluated by using the Q statistic and was considered statistically significant for $p < 0.10$ (9). Not finding significant heterogeneity in some of these meta-analyses does not necessarily prove homogeneity since type II error cannot be excluded. We used both random-effects and fixed-effects calculations (10, 11). Studies were weighted by the inverse of their variance. Random-effects models (10) incorporate an estimate of the between-study variance, and they tend to give wider confidence intervals than fixed effects (9) when between-study heterogeneity is present; otherwise, fixed- and random-effects estimates coincide. Random effects are presented in the Results unless stated otherwise. While random-effects estimates are more appropriate, some experts suggest that in the presence of heterogeneity, no summary results should be presented and that the key issue is to try to understand the sources of heterogeneity.

For binary outcomes, we used the odds ratio as the metric of choice for the effect size. Calculations were based on the Mantel-Haenszel (11) and DerSimonian and Laird (10) models for fixed

and random effects, respectively. For continuous outcomes, we used the weighted mean difference, and data synthesis used general variance models (12).

To evaluate the concordance between the results of small and large trials, we evaluated whether the difference in the effect sizes in large versus smaller trials was more prominent than what would be anticipated by chance alone. We estimated the z score for discrepancy ($z = [\text{effect size in large studies} - \text{effect size in small studies}] / [\text{standard error of the difference of the effect sizes}]$) (2). An absolute z score > 1.96 suggests that the difference between large and smaller trials is beyond chance at the 0.05 level of statistical significance. Effect sizes are the natural logarithm of the odds ratio for binary outcomes and the absolute difference for continuous outcomes. Discrepancies were similar when we used a more conservative definition for the z score (not shown) (13).

We also assessed in how many cases the effects were in the opposite direction in large versus smaller trials and in how many cases the effect was in the same direction, but the odds ratio or absolute difference differed twofold between large and smaller studies. For each meta-analysis, we also examined whether the summary effect sizes differed beyond chance for trials with different reported quality attributes. Finally, we explored whether differences in doses and dropouts may exist whenever large trials disagreed with smaller ones.

Analyses were conducted in Meta-Analyst (Joseph Lau, Boston) and SPSS 11.0 (SPSS Inc., Chicago). All p values are two-tailed.

Results

Eligible Meta-Analyses

Sixteen eligible meta-analyses with 133 randomized controlled trials (small: $N = 101$, large: $N = 32$) were analyzed (Table 1). The number of large studies per meta-analysis ranged from one to eight (the total number of patients in all of the large studies ranged from 908 to 10,948). The number of small studies per meta-analysis ranged from one to 22 (the total number of patients in all of the small studies ranged from 147 to 3,490). In four of the 16 meta-analyses, the cumulative sample size of all of the small trials exceeded the sample size of all of the large trials. In two meta-analyses, the average number of subjects per small study was < 100 ; in nine, it was 100–250; and in five, it was 251–400. The average sample size per large study was 900–2,000 in 14 meta-analyses and over 2,000 in two meta-analyses. All large trials were multicenter, with three exceptions, whereas of the 101 small trials, 53 were multicenter, 35 were single center, and for 13, it was not specified whether they were single center or not. Fourteen meta-analyses had binary primary outcomes and two had continuous primary outcomes (Table 1).

Between-Study Heterogeneity

In five meta-analyses (identification numbers 1, 3, 4, 12, and 15 [14, 16, 17, 22, and 23]), there was statistically significant between-study heterogeneity. In two cases (numbers 1 and 12 [14, 22]), heterogeneity continued to exist between small trials. In one of the six cases where at least two large trials were available (number 12 [22]), heterogeneity was also present between the available large trials.

TABLE 1. Meta-Analyses Covering Both Large and Smaller Randomized Controlled Trials^a

Identification Number	Study	Year	Topic of Meta-Analysis	Primary Outcome	Large Trials			Small Trials		
					N	N (total)	Range (per trial)	N	N (total)	Range (per trial)
1	Zoritch et al. (14)	2002	Day care for preschool children	IQ at 36 months	1	908		3	201	35–86
2	Lopez-Arrieta and Birks (15)	2002	Nimodipine for dementia	Cognitive function change at 24 weeks	1	954		3	389	65–209
3	Roberts et al. (16)	2002	Intercessory prayer for alleviation of ill health	Intermediate or bad outcome	1	1,013		2	431	38–393
4	Duggan et al. (17)	2002	Olanzapine versus typical antipsychotics for schizophrenia	No important clinical response by 6–8 weeks	1	1,996		3	782	84–431
5	Kennedy et al. (18)	2002	Risperidone versus typical antipsychotics for schizophrenia	No change in Positive and Negative Syndrome Scale data greater than 20%	1	1,362		9	1,619	35–675
6	Barbui et al. (19)	2002	Treatment discontinuation with selective serotonin reuptake inhibitors versus clomipramine	Total dropouts	1	1,019		13	1,595	32–510
7	Ashenden et al. (20)	1997	Effect of brief smoking cessation advice	Smoking cessation	4	8,368	1,420–3,128	6	2,052	182–584
8	Ashenden et al. (20)	1997	Effect of intensive versus brief smoking cessation advice	Smoking cessation	1	2,691		4	1,268	155–629
9	Pignone et al. (21)	2002	Screening for depression in adults	Persistence of depression	1	1,156		6	1,338	41–380
10	Law and Tang (22)	1995	Advice and encouragement by physician (routine consultation)	Smoking cessation	8	10,948	843–2,138	9	3,490	163–691
11	Law and Tang (22)	1995	Advice by physician during pregnancy	Smoking cessation	1	982		1	147	
12	Law and Tang (22)	1995	Behavior therapy and self-help manuals during pregnancy	Smoking cessation	2	1,929	935–994	5	1,641	137–723
13	Law and Tang (22)	1995	Behavioral modification therapy by counselor (individual sessions)	Smoking cessation	2	2,946	1,119–1,827	3	615	144–316
14	Law and Tang (22)	1995	Behavioral modification therapy with use of self-help manuals	Smoking cessation	4	5,233	1,217–1,396	7	1,329	40–755
15	Viguera et al. (23)	1998	Discontinuation versus continuation of antidepressants in depression	Relapse rates	1	1,141		22	1,562	12–295
16	Wilk et al. (24)	1997	Brief interventions for heavy alcohol drinkers	Drinking moderation	2	2,026	907–1,119	6	736	72–197

^a All large trials were multicenter, with the exception of number 3 and number 11, as well as one of the eight large trials in number 10.

Disagreements in Large Versus Smaller Trials

In three meta-analyses of binary outcomes (numbers 4, 5, and 15 [17, 18, and 23]) and in one meta-analysis of continuous outcomes (number 1 [14]), the effect sizes of large and smaller trials differed beyond chance (Figure 1). Thus, the overall rate of disagreement was 25% (four of 16).

In trials addressing the effect of day care on the cognitive development of children (number 1 [14]), a large trial (25) found significantly more conservative results than the smaller trials. Three small trials showed that attendance of day care by preschool children was associated with better IQs at 36 months with a 24-point IQ difference. The only

available large trial showed a much smaller, 9-point IQ difference (25).

Nine small trials of risperidone versus typical antipsychotic medications for schizophrenia (number 5 [18]) showed a significant decrease in treatment failures, as measured by a reduction in positive and negative symptom scores of at least 20%. On average, treatment failures were halved in these small trials. Nevertheless, a single large trial with a sample size almost similar to that of the nine smaller trials combined showed no difference between risperidone and haloperidol (26).

The opposite scenario was seen in the case of olanzapine (number 4 [17]). Three small trials of olanzapine ver-

sus other typical antipsychotics for schizophrenia did not show any difference in clinical failure rates at 6–8 weeks. The single large trial showed a significant difference, with half the clinical failures with olanzapine compared with haloperidol (27).

Twenty-two small trials showed that discontinuation of antidepressants in patients with depression (number 15 [23]) was associated with a sixfold increase in the rate of relapse compared with the continuation of medication. A single large trial (28) with a sample size in the same range as the total sample size of the 22 smaller trials also showed a significant increase in the relapse rate, but the increase was only about twofold rather than sixfold.

Other Potential Disagreements

The summary estimates of the large and smaller studies were always in the same direction. However, in three meta-analyses (numbers 4, 5, and 15 [17, 18, and 23]), the magnitude of the odds ratio differed over twofold between the large and smaller studies, and in another meta-analysis (number 1 [14]), with continuous outcomes, the effect size differed over twofold between large and smaller studies. These cases included the olanzapine, risperidone, antidepressant discontinuation, and day care meta-analyses described. In addition, a large difference in the effect size was also documented in the meta-analysis of intercessory prayer for the alleviation of ill health (number 3 [16]). Although the three smaller studies suggested an odds ratio of 0.31 with a potentially large benefit from intercessory prayers against bad outcomes, the largest trial, with 1,013 randomly assigned subjects (29), found absolutely no benefit from such intercessory prayers (odds ratio=0.96). Given the large confidence intervals, the results of the smaller and large trials were not statistically incompatible with each other, but the divergence in the suggested effect size was considerable.

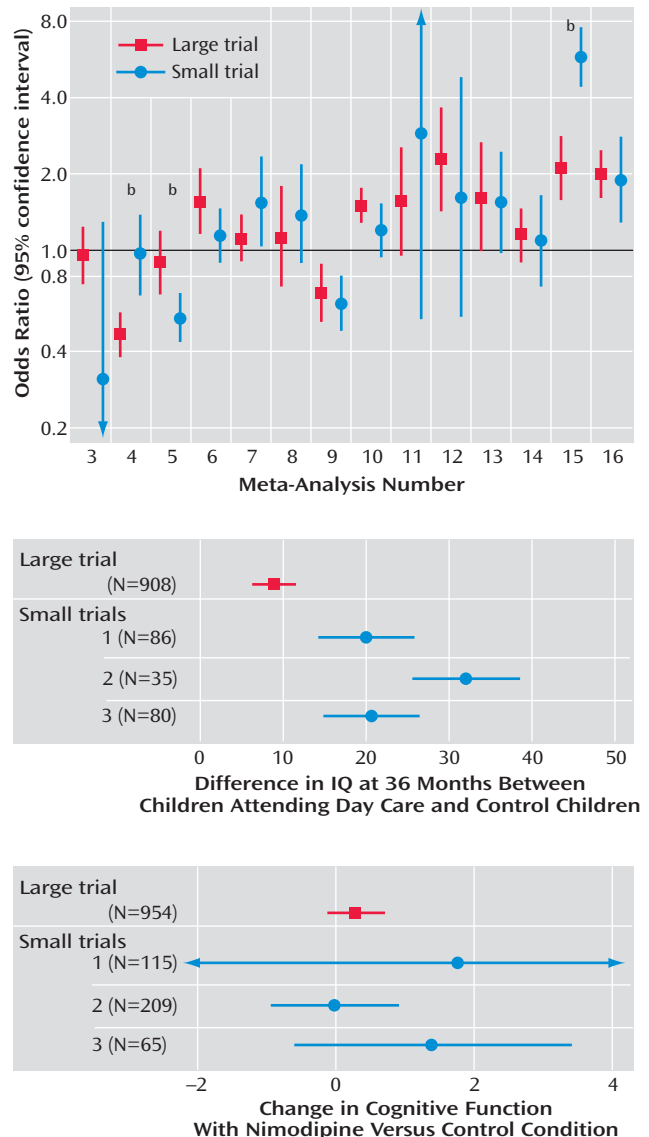
Fixed Effects

With fixed-effects models, disagreements beyond chance between large and smaller trials were observed in five cases (31.3%). Disagreements were similar to those observed with random-effects calculations, plus the intercessory prayers meta-analysis that showed significant disagreement by fixed-effects calculations. The same four meta-analyses showed over twofold differences in the effect size with fixed-effects as with random-effects calculations.

Quality and Study Design

Of the 133 analyzed trials, 56 were double blind, but only 18 reported an adequate mode of random assignment, and only seven reported adequate allocation concealment. Five meta-analyses (numbers 2, 3, 4, 6, and 15 [15–17, 19, 23]) included only double-blind trials, and 10 meta-analyses (numbers 1, 7, 8, 9, 10, 11, 12, 13, 14, and 16 [14, 20–22, 24]) included no such trials. The relative benefit of risperidone was significantly larger in the two unmasked trials than in the eight double-blind trials (Figure 2). Eleven and

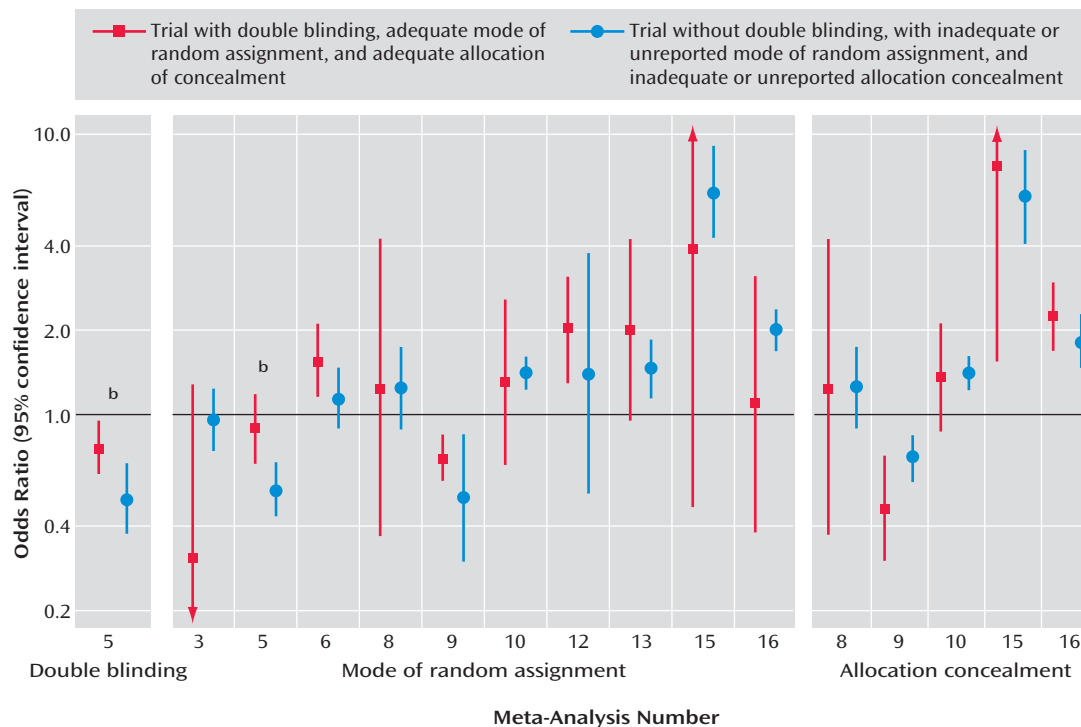
FIGURE 1. Comparison of the Effect Sizes and 95% Confidence Intervals in Large Versus Smaller Studies for 14 Topics With Binary Outcomes and Two Topics With Continuous Outcomes^a



^a The meta-analysis numbers in the top image correspond to the meta-analysis numbers in Table 1. Whenever there was more than one large or more than one small trial, their summary effect is shown and calculations were performed with random-effects models. The middle image shows effect sizes and 95% confidence intervals in the large trial and each small trial of day care attendance by preschool children. The outcome is IQ at 36 months. The image at the bottom shows effect sizes and 95% confidence intervals in the large trial and each small trial of nimodipine for dementia. The outcome is change in cognitive function.

^b Meta-analyses with significant difference beyond chance between large and smaller trials ($p < 0.05$).

six meta-analyses, respectively, had at least one trial with adequate reporting of the mode of random assignment and allocation concealment (Figure 2). In six meta-analyses, trials with adequate reporting of the mode of random assignment had less impressive results for the experimental intervention compared with trials where the mode of

FIGURE 2. Comparison of the Effect Sizes and 95% Confidence Intervals in Studies With and Without Specific Quality Attributes^a

^a The meta-analysis numbers correspond to the meta-analysis numbers in Table 1. Whenever there was more than one trial with or without a quality attribute, their summary effect is shown and calculations were performed with random-effect models. For meta-analysis number 1 (day care for preschool children), only the large trial had an adequate mode of random assignment and allocation concealment. The observed effects on IQ are shown in Figure 1. No trials had an adequate mode of random assignment in five meta-analyses, and no trials reported adequate allocation concealment in 10 meta-analyses (not shown).

^b Meta-analyses with significant difference between trials with, versus without, the quality attribute ($p < 0.05$).

random assignment was inadequate or not stated. The opposite was seen in five meta-analyses (Figure 2). In four meta-analyses, trials with adequate allocation concealment had less impressive results for the experimental intervention compared with trials where allocation concealment was clearly inadequate or not reported; the opposite was seen in two meta-analyses (Figure 2). With two exceptions (numbers 1 and 5 [14 and 18]), the effect sizes did not differ beyond chance in trials with versus without these quality attributes in any of these meta-analyses. For day care in preschool children, the single large trial was also the only one to have adequate description of the mode of random assignment and allocation concealment. For risperidone, the single large trial was also the only one to report an adequate mode of random assignment.

In the olanzapine meta-analysis, studies differed in the doses of both olanzapine and the comparator (haloperidol). In the larger study, a high proportion of patients did not complete the 6 weeks of treatment, and the rate was higher in the haloperidol arm (53% versus 34%). This could be in part due to the high doses used in the control group and could have affected the efficacy results (30). In the risperidone and discontinuation of antidepressants meta-analyses, various doses were also employed across different trials. There was no clear pattern to separate small versus

larger trials, although for risperidone, there was a non-significant tendency toward different effect sizes, depending on whether high- or low-dose comparators were used.

Discussion

Our empirical data suggest that in the field of mental health, evidence from small trials is usually not incompatible with the results obtained from large trials on the same question. Nevertheless, in one of four cases, we observed discrepancies beyond chance in the results of large versus smaller trials. Frequently, the effect size was much different in absolute magnitude in large versus smaller trials. In most but not all discrepancies, the large trials tended to give more conservative impressions about the intervention (or the newer intervention, when interventions were compared), whereas the smaller results conveyed a more enthusiastic picture.

Our findings are in line with previous research comparing meta-analyses of large versus smaller trials in other fields, such as perinatal medicine, cardiology, infectious diseases, surgery, and oncology (1–5). In these fields, discrepancies beyond chance were noted in 10% to 23% of cases by random-effects calculations (1–5). The 25% disagreement rate that we observed in mental health is com-

patible with these figures. Fixed-effects estimates of discrepancies tend to be higher, and our current estimate of 31% is also compatible with previous investigations in other fields (1–5). We should caution that seven of the 16 meta-analyses that we addressed dealt with smoking cessation, a topic that some investigators may not consider as pertaining to the mainstream of mental health. We believe that the inclusion of these meta-analyses was warranted, given the major importance of the psychological dependence and the strong mental dimension in the establishment of the smoking epidemic. However, exclusion of these meta-analyses would increase the rate of disagreements in our series to 44% (four of nine). Therefore due caution is probably needed on interpreting the results of small trials on mental health interventions.

Empirical investigations from other medical fields suggested that variability in the treatment effects of trials may reflect quality differences. In particular, it was claimed that double-blind, adequately randomized, and adequately concealed trials may show more conservative effects for the experimental treatments, compared with trials not meeting or not stating such quality items (7, 8). However, recent work (31) has not shown any consistent dependence between these quality items and effect size. Our appraisal of these quality items in mental-health-related trials showed that the differences between “good” and “poor” reported quality trials were also not consistent. In two meta-analyses, the single large trial was also the only one to meet some of these quality aspects, so it is unknown whether differences in effect sizes were due to sample size or the quality of the trial design. Unmasked trials gave more conservative results than double-blind trials in the case of risperidone, but this was an isolated observation. Moreover, it is unknown whether the lack of reporting of specific quality items means that a trial is of poor quality or simply poorly reported (32). The advent of the Consolidated Standards of Reporting Trials guidelines (33) should help standardize the reporting of trials across medical fields. This is particularly important for mental health, where our study showed that few trials report an adequate mode of random assignment and allocation concealment. Poor reporting makes difficult the overall assessment of the randomized literature in mental health.

When small and large trials disagree, it is unknown whether the larger trials should be considered to be more reliable. There are several other considerations beyond trial quality. Small trials have large standard errors by definition. With multiple studies addressing the same research question, the most extreme effect sizes may occur with the smallest sample size. In the presence of some publication bias (34), the largest effect sizes may be seen with the smallest trials. Conversely, large multisite trials are actually summaries of many small trials (the different sites involved) with relative disregard for the potential disagreements between the results observed at different sites. Many of the smaller studies may also be conducted at sev-

eral clinical sites. In some cases, large and smaller trials may be targeting different patient populations with variable background risks (35). For various mental-health-related interventions, large trials may be less selective, whereas small trials target patients who have the highest likelihood of response. Thus, in cases of disagreement, small and large trials should be carefully scrutinized for hints to important sources of latent diversity (36). Discrepancies may be due to more than one cause in each case. For example, in the case of risperidone versus haloperidol, an updated review (18) contains a number of small trials with “negative” results that had previously remained unpublished. Moreover, these trials differed in the dose of haloperidol. Thus, the original discrepancy may be explained by a combination of publication bias and different design in small versus larger trials. Detection of such sources of heterogeneity may give us important hints about the overall integrity and validity of clinical research in mental health (37, 38).

We should acknowledge that the definition of “large” trial is unavoidably arbitrary. Moreover, we could identify only 16 meta-analyses in which both large and smaller trials were available, despite the fact that we searched more than 400 systematic reviews. The dearth of meta-analyses with large trials reflects mostly the fact that few such large trials are conducted in the field of mental health. In a recent evaluation of schizophrenia trials, only 2% of 2,000 studies had a sample size exceeding 400 patients (39). The performance of large trials should be encouraged in the broader mental health field (40). Nevertheless, large trials will never be performed for most clinical questions because of limitations in resources. For behavioral and psychosocial interventions, standardization of interventions across multiple therapists is difficult. Moreover, even small trials may provide evidence conclusive enough for practical purposes. Additional large trials may be deemed unnecessary or even unethical if small studies have shown a large treatment effect. Therefore, evidence-based medicine will often have to depend on the results of small trials and meta-analyses thereof. Attention to the design, conduct, analysis, and reporting of all trials, regardless of sample size, is thus important, and all trials should be examined as forming a continuum of evidence.

Received Dec. 30, 2003; revision received March 31, 2004; accepted May 14, 2004. From the Department of Hygiene and Epidemiology and the Department of Pediatrics, University of Ioannina School of Medicine; the Department of Pediatrics, George Washington University School of Medicine and Health Sciences, Washington, D.C.; the Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds, Leeds, U.K.; the Health Services Research Department, Institute of Psychiatry, London; the STAKES National Research and Development Centre for Welfare and Health, Helsinki, Finland; and the Institute for Clinical Research and Health Policy Studies, Tufts-New England Medical Center, Tufts University School of Medicine, Boston. Address correspondence and reprint requests to Dr. Ioannidis, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece; jioannid@cc.uoi.gr (e-mail).

The following people have participated in the EU-PSI Project: A. Tuunainen, C. Almeida, M. Starr, D. Hermans, H. Lee, S. Leucht, R. Stancliffe, M. Fenton, P. Pörtfors, S. Cardo, M. Ferri, M. Davoli, P. Papanicolaou, S. Pirkola, S. Vecchi, J. Birks, G. McDonald, and J.A. Dennis.

The authors thank Athina Tatsioni for helping them retrieve some of the pertinent articles.

References

- Villar J, Carroli G, Belizan JM: Predictive ability of meta-analyses of randomized controlled trials. *Lancet* 1995; 345:772-776
- Cappelleri JC, Ioannidis JP, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC, Lau J: Large trials vs meta-analyses of small trials: how do their results compare? *JAMA* 1996; 276:1332-1338
- 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; 337:536-542
- Ioannidis JP, Cappelleri JC, Lau J: Meta-analyses and large randomized controlled trials (letter). *N Engl J Med* 1998; 338:59
- Ioannidis JP, Cappelleri JC, Lau J: Issues in comparisons between meta-analyses and large trials. *JAMA* 1998; 279:1089-1093
- EU-PSI Project: Evidence-Based Treatment in Mental Health and Optimised Use of Databases: The EU-PSI Project, PsiTri, and the Mental Health Library. www.psitri.helsinki.fi
- Schulz KF, Chalmers I, Hayes RJ, Altman DG: Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408-412
- Juni P, Altman DG, Egger M: Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001; 323:42-46
- Lau J, Ioannidis JP, Schmid CH: Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997; 127:820-826
- DerSimonian R, Laird N: Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7:177-188
- Mantel N, Haenszel WH: Statistical aspects of the analysis of data from retrospective studies of diseases. *J Natl Cancer Inst* 1959; 22:719-748
- Petitti DB: Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis, 2nd ed. New York, Oxford University Press, 2000
- Moses LE, Mosteller F, Buehler JH: Comparing results of large clinical trials to those of meta-analyses. *Stat Med* 2002; 21:793-800
- Zoritch B, Roberts I, Oakley A: Day care for pre-school children. *Cochrane Database Syst Rev* 2002; (3):CD000564
- Lopez-Arrieta JM, Birks J: Nimodipine for primary degenerative, mixed and vascular dementia. *Cochrane Database Syst Rev* 2002; (3):CD000147
- Roberts L, Ahmed I, Hall S: Intercessory prayer for the alleviation of ill health. *Cochrane Database Syst Rev* 2002; (3):CD000368
- Duggan L, Fenton M, Dardennes RM, El-Dosoky A, Indran S: Olanzapine for schizophrenia. *Cochrane Database Syst Rev* 2002; (3):CD001359
- Kennedy E, Song F, Hunter R, Clarke A, Gilbody S: Risperidone versus typical antipsychotic medication for schizophrenia. *Cochrane Database Syst Rev* 2002; (3):CD000440
- Barbui C, Hotopf M, Freemantle N, Boynton J, Churchill R, Eccles MP, Geddes JR, Hardy R, Lewis G, Mason JM: Selective serotonin reuptake inhibitors versus tricyclic and heterocyclic antidepressants: comparison of drug adherence. *Cochrane Database Syst Rev* 2002; (3):CD002791
- Ashenden R, Silagy C, Weller D: A systematic review of the effectiveness of promoting lifestyle change in general practice. *Fam Pract* 1997; 14:160-176
- Pignone MP, Gaynes BN, Rushton JL, Burchell CM, Orleans CT, Mulrow CD, Lohr KN: Screening for depression in adults: a summary of the evidence for the US Preventive Services Task Force. *Ann Intern Med* 2002; 136:765-776
- Law M, Tang JL: An analysis of the effectiveness of interventions intended to help people stop smoking. *Arch Intern Med* 1995; 155:1933-1941
- Viguera AC, Baldessarini RJ, Friedberg J: Discontinuing antidepressant treatment in major depression. *Harv Rev Psychiatry* 1998; 5:293-306
- Wilk AI, Jensen NM, Havighurst TC: Meta-analysis of randomized control trials addressing brief interventions in heavy alcohol drinkers. *J Gen Intern Med* 1997; 12:274-283
- Brooks-Gunn J, McCarton CM, Casey PH, McCormick MC, Bauer CR, Bernbaum JC, Tyson J, Swanson M, Bennett FC, Scott DT, et al: Early intervention in low-birth-weight premature infants: results through age 5 years from the Infant Health and Development Program. *JAMA* 1994; 272:1257-1262
- Peuskens J (Risperidone Study Group): Risperidone in the treatment of patients with chronic schizophrenia: a multi-national, multi-centre, double-blind, parallel-group study versus haloperidol. *Br J Psychiatry* 1995; 166:712-726
- Tollefson GD, Beasley CM Jr, Tran PV, Street JS, Krueger JA, Tamura RN, Graffeo KA, Thieme ME: Olanzapine versus haloperidol in the treatment of schizophrenia and schizoaffective and schizophreniform disorders: results of an international collaborative trial. *Am J Psychiatry* 1997; 154:457-465
- Rouillon F, Serrurier D, Miller HD, Gerard MJ: Prophylactic efficacy of maprotiline on unipolar depression relapse. *J Clin Psychiatry* 1991; 52:423-431
- Harris WS, Gowda M, Kolb JW, Strychacz CP, Vacek JL, Jones PG, Forker A, O'Keefe JH, McCallister BD: A randomized, controlled trial of the effects of remote, intercessory prayer on outcomes in patients admitted to the coronary care unit. *Arch Intern Med* 1999; 159:2273-2278
- Leucht S, Pitschel-Walz G, Abraham D, Kissling W: Efficacy and extrapyramidal side-effects of the new antipsychotics olanzapine, quetiapine, risperidone, and sertindole compared to conventional antipsychotics and placebo: a meta-analysis of randomized controlled trials. *Schizophr Res* 1999; 35:51-68
- Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, Lau J: Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002; 287:2973-2982
- Ioannidis JP, Lau J: Can the quality of randomized trials and meta-analyses be quantified? (letter). *Lancet* 1998; 352:590-591
- Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T (CONSORT [Consolidated Standards of Reporting Trials] group): The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; 134:663-694
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR: Publication bias in clinical research. *Lancet* 1991; 337:867-872
- Ioannidis JP, Lau J: The impact of high-risk patients on the results of clinical trials. *J Clin Epidemiol* 1997; 50:1089-1098
- Lau J, Ioannidis JP, Schmid CH: Summing up evidence: one answer is not always enough. *Lancet* 1998; 351:123-127
- Gilbody SM, Song F: Publication bias and the integrity of psychiatry research. *Psychol Med* 2000; 30:253-258
- Gilbody SM, Song F, Eastwood AJ, Sutton AJ: The causes, consequences and detection of publication bias in psychiatry. *Acta Psychiatr Scand* 2000; 102:241-249
- Thornley B, Adams C: Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998; 317:1181-1184
- Yusuf S, Collins R, Peto R: Why do we need some large, simple randomized trials? *Stat Med* 1984; 3:409-422