Published in final edited form as:

Contemp Clin Trials. 2011 March ; 32(2): 299–304. doi:10.1016/j.cct.2010.12.007.

The Use and Abuse of Multiple Outcomes in Randomized Controlled Depression Trials

Kristin M. Tyler, B.A., Sharon-Lise T. Normand, Ph.D., and Nicholas J. Horton, Sc.D.

Abstract

Objective—Multiple outcomes are commonly analyzed in randomized trials. Interpretation of the results of trials with many outcomes is not always straightforward. We characterize the prevalence and factors associated with multiple outcomes in reports of clinical trials of depression, methods used to account for these outcomes, and concordance between published analyses and original protocol specifications.

Methods—A PubMed search for randomized controlled depression trials that included multiple outcomes published between January 2007 and October 2008 in 6 medical journals. Original study protocols were reviewed where available. Parallel data collection by 2 abstractors was used to determine trial registration information, the number of outcomes, and analytical method.

Results—Of the 55 included trials, nearly half of the papers reported more than 1 primary outcome, while almost all (90.9%, $n=50$) reported more than 2 combined primary or secondary outcomes. Relatively few of the studies (5.8%, $n=3$) adjusted for multiple outcomes. While most studies had published protocols in clinical trial registries (76.4%, $n=42$), many did not specify outcomes in the protocol ($n=11$) and a number had discrepancies with the published report.

Conclusions—Multiple outcomes are prevalent in randomized controlled depression trials and appropriate statistical analyses to account for these methods are rarely used. Not all studies filed protocols, and there were discrepancies between these protocols and published reports. These issues complicate interpretability of trial results, and in some cases may lead to spurious conclusions. Promulgation of guidelines to improve analysis and reporting of multiple outcomes is warranted.

Keywords

multiplicity; clinical trial; multiple outcomes; depression studies; trial registries; joint models; global tests

INTRODUCTION

Multiple outcomes are often incorporated in randomized clinical trials (RCTs) due to interest in characterizing how a treatment influences a range of responses. Long-term mental health conditions, such as depression, are particularly reflective of this practice. Reporting more than one outcome in depression trials may be appropriate because a single measure may not sufficiently characterize the effect of a treatment on a broad set of domains. A lack

Corresponding author: Nicholas J. Horton, Department of Mathematics and Statistics/Smith College, Clark Science Center, 44 College Lane, Northampton, MA 01063-0001 USA, nhorton@smith.edu.

Table 1. Full table of articles

Note: The full table of articles, with citations, outcome counts and methodology used is available as an online Appendix (see separate attachment in submission)

of clear consensus on the most important clinical outcome, combined with the need to examine clinical effectiveness on related outcomes spanning disparate domains, encourage the use of multiple outcomes.

To adhere to statistical design issues, researchers often specify a small set of measures to serve as the primary outcomes, with another (often larger) set listed as secondary. While it is common practice to collect, analyze and report multiple measures, the efficient and appropriate analysis of multiple outcomes is not fully established. A number of approaches to accounting for multiple outcomes have been proposed [1–3], assessed [4] and reviewed [5]. The most common method for analyzing multiple outcomes is separate testing of each individual outcome, sometimes with but most often without adjustment for multiple testing. Another approach involves combining the multiple outcomes into a single (composite) outcome and performing a single test [6]. A third approach undertakes global testing using simultaneous (joint) tests [7].

The choice of an appropriate method for dealing with multiple outcomes is important because clinical interpretations can be difficult in the presence of multiple conflicting results. Simultaneous or joint models that provide an overall test, with separate reports of individual outcome, can provide useful additional information. Moreover, joint models can be more powerful if some outcomes are missing.

Study design features should be specified prior to patient enrollment and characterized in the study protocol. However, prior work by Al-Marzouki et al [8] and Chan et al [9] found discrepancies in outcomes between published study protocols and clinical trial reports. Reasons for these discrepancies may be related to unanticipated changes while conducting the trial, such as modifications to trial inclusion and exclusion criteria to increase enrollment. Al-Marzouki and colleagues found that there were major differences in 11 out of 37 trials. Overall there was a median of one outcome in published protocols and a median of two outcomes in the published report. Chan et al concluded that reported outcomes are often incomplete and inconsistent with the registered protocols, which potentially yield bias and unreliable results. Specifically, statistically significant outcomes were more likely to be fully reported than non-significant outcomes, indicating that results may be overestimating the benefits of an intervention. Turner and colleagues [10] found a similar trend among antidepressant clinical trials. Another study by Viereck and Boudes [11] found that pharmaceutical industry practices do not generally make clinical trial protocols and results accessible to the general public. The FDA Amendment Act [12] expanded the trial registry established under the FDA Modernization Act to encompass a larger set of protocols involving treatments and devices [13]. Along with the International Committee of Medical Journal Editors [14], these new promulgations are likely to improve reporting.

We assessed the prevalence of multiple outcomes in clinical trials of depression. Major depression and related disorders were chosen because they have a profound personal, social and economic cost [15], and are the focus of a number of prevention and intervention trials. Examples of depression measures include a clinical diagnosis, measures of depressive symptoms (such as the Center for Epidemiologic Studies–Depression Scale [16], the Beck Depression Inventory [17,18], the Hamilton Depression Rating Scale [19], and the Montgomery-Asberg Depression Rating Scale [20]). The use of multiple outcomes in depression trials is particularly common because disease complexity is multifaceted [7]. No one measure encompasses all aspects of the disorder, and clinicians may be interested in the impact of a new treatment on different domains. As a result, it is naive to force a single outcome. Instead, use of a broad range of clinically relevant measures, along with a procedure to globally assess them is more realistic and useful.

We also sought to determine whether there were important associations between number of outcomes and characteristics of the depression trial and to assess the concordance between reported outcomes and those specified in published study protocols.

METHODS

Article Selection

We reviewed the use of multiple outcomes in randomized clinical trials with depression as a primary or secondary outcome that were recently published in six top-tier psychology or general medical journals (*American Journal of Psychiatry* (AJP), *Archives of General Psychiatry* (AGP), *British Medical Journal* (BMJ), *Lancet*, *Journal of the American Medical Association* (JAMA), and the *New England Journal of Medicine* (NEJM)). PubMed was used to obtain articles that matched "clinical trials", included the keywords "depression" or "depressive disorder", and were published between January 2007 and October 2008 in these six journals. These journals were selected because of their high impact and relevance to clinical researchers in psychiatry.

Abstraction from Clinical Trials Registry

We abstracted data from the appropriate clinical trials registries: ClinicalTrials.gov, Australian New Zealand Clinical Trials Registry (ANZCTR), and International Standard Randomised Controlled Trial Number (ISRCTN). We calculated the number of primary and secondary outcomes described in the protocol. In situations where it was difficult to determine the number of outcomes a consensus method was undertaken by two of the authors (KT and NH).

Data Extraction

We extracted the number of primary outcomes, secondary outcomes, and the methods (if any) used to account for multiple outcomes. An outcome was coded as primary if it was designated as such by the researchers in the abstract, methods, results or tables. Each article was reviewed independently by two of the authors (KT and NH) and a consensus process was used to address any inconsistencies in coding. Secondary outcomes include measures that were reported as randomized trial group comparisons. Outcomes listed as "additional", "tertiary" or "exploratory" were coded as secondary outcomes [e.g. 21]. Side effects and adverse events were not included in the count. If no distinction was made between primary and secondary outcomes, all were assumed to be primary outcomes.

The total number of outcomes was calculated as the sum of the number of primary and secondary outcomes. In addition, a categorical variable for the number of primary outcomes was created (1, 2–3, or 4+). Other abstracted variables included the sample size (and coded as sample size <100, 100–399, or 400+), the journal name, and the clinical trial registry code. If no registration code was reported in the paper, the registries were searched in order to abstract the appropriate protocol information.

Statistical Analysis

Fisher's exact tests were used to test associations in cross-classification tables, the Kruskal-Wallis test was used to compare count outcomes by group, and Spearman correlation to assess associations between counts of outcomes in the initial protocol and the final report. A p-value of 0.05 was used to assess statistical significance. Because our goals were exploratory, we did not undertake any adjustment for the five tests that we undertook [22]. All p-values are two-tailed. Analyses were undertaken using Stata version 10.1.

RESULTS

Of the 105 studies initially retrieved, a total of 50 were excluded because they were not RCTs ($n=31$) (e.g. cohort study nested within a trial), were cost effectiveness studies ($n=2$), or did not have a depression measure as a primary or secondary outcome ($n=17$). After these exclusions, there were a total of 55 articles coded and analyzed (Figure 1 and Table 1). Just over half (52.7%, $n=29$) of the papers reported exactly one primary outcome. The distribution was heavily skewed to the right with 25.5% ($n=14$) of the studies reporting at least 5 primary outcomes.

Of the primary outcomes that had a depression component, the most commonly used were the Hamilton Depression Rating Scale (22.9%, $n=11$), the Montgomery-Asberg Depression Rating Scale (8.3%, $n=4$), Clinical Global Impression Scale (8.3%, $n=4$), and clinical diagnosis using the DSM-IV (8.3%, $n=4$). Table 2 displays the distribution of primary depression outcomes used in more than one study.

Secondary outcomes were also common with a median of three outcomes. This was also heavily skewed with a maximum number of secondary outcomes of 31 (25th percentile 0, 75th percentile 6 outcomes). Almost all (94.5%, $n=52$) of the articles reported more than one primary and secondary outcome. The number of secondary outcomes was significantly larger ($p=0.003$) for papers reporting only 1 primary outcome (median=5) as compared to those with more than 1 primary outcome (median=0). The median of the total number of outcomes was 7 outcomes (25th percentile 4, 75th percentile 10 outcomes). Figure 2 displays boxplots of the distribution of number of primary, secondary and total (sum of primary + secondary) outcomes.

A total of eight articles used a Bonferroni-type adjustment; however five papers reporting this approach applied it to comparisons within multiple treatments and not for multiple outcomes [23–27]. Of the 52 articles with multiple outcomes, only 5.8% ($n=3$) used a Bonferroni adjustment. While Strong et al [28] specified only one primary outcome, the seven secondary outcomes analyzed were adjusted by utilizing a modified cutoff for statistical significance at 0.01. Welton et al [29] adjusted for the 41 outcomes by using a Bonferroni corrected alpha level of 0.0001. Lesperance et al [30] clearly specified a single primary outcome and single secondary outcome in the abstract, and included 6 exploratory outcomes later in their analysis. They used a Bonferroni-like multiplicity adjustment by partitioning the experiment-wise alpha level into 0.033 for the primary outcome analysis and 0.017 the secondary outcome analysis. All additional analyses, which they designated as exploratory, used the standard alpha level of 0.05. No articles used a Hochberg [31] or similar procedure. None of the papers reported use of joint testing methods or global tests.

Only two articles used a composite measure as their primary outcome. Raskin et al [32] combined information from four cognitive tests (Verbal Learning and Recall, Symbol Digit Substitution, Two-Digit Cancellation, and Letter-Number Sequencing). The composite cognitive score was a weighted sum (in proportion to the time spent administering the test) that ranged from 0 to 51. Goldberg et al [33] used a measure of either recovery (8 weeks with partial or full remission) or recovering (4 weeks with low level of symptoms).

The median sample size of the reported RCT's was 200 with a minimum size of 28 and maximum of 7380. Studies with sample sizes less than 100 and greater than or equal to 400 participants had significantly more outcomes reported than did studies with sample sizes between 100 and 400 ($p=0.003$, see Table 3). There were more papers with total number of participants between 100 and 400 with only 1 outcome than expected.

Of the 55 eligible papers, we linked 42 (76%) to published protocols and of those, 31 (74%) reported information on outcomes. The 13 papers without protocols in the registries that we searched were published in AJP or AGP, while the 11 papers without outcomes specified in the protocol were published in AJP, AGP, JAMA, or the NEJM. There were no statistically significant differences between the distribution of the number of primary outcomes and clinical trial registry status ($p=0.92$). There were statistically significant differences between the sample size groups and the proportion with complete registry information on outcomes ($p=0.005$, 42% available for sample size <100 , 60% available for a sample size between 100 and 400, and 62% available for sample size ≥ 400). Of the 31 with reported outcomes in the protocol, 74% increased the total number of outcomes reported in the final manuscript (by an average of 4.9 outcomes). The Spearman correlation between the number of primary outcomes reported in the published manuscript and that reported in the protocol was modest ($n=31$, correlation=0.14, test of no-association yielded $p=0.45$).

DISCUSSION

The CONSORT [34,35] statement provides guidance and structure to investigators when reporting the results of clinical trials. These guidelines are intended to clarify the key outcomes of these investigations, and ensure that their description is detailed and consistent within the abstract, methods, results and tables. Furthermore, while the CONSORT statement recommends only a single primary outcome, it does not directly specify statistical methods for appropriately handling multiple outcomes. A recent study examining statistical problems found by reviewers in high-impact psychiatry journals demonstrated the need to improve reporting of multiple statistical tests [22].

The CONSORT 2010 [36] statement strengthens the discussion of multiple outcomes, and notes that while a trial may have more than one primary outcome, "having several primary outcomes, however, incurs the problems of interpretation associated with multiplicity of analyses ... and is not recommended." (p. 7).

Nearly half of depression clinical trials published between January 2007 and October 2008 in leading medical and psychiatry journals reported more than one primary outcome, while nearly all reported more than one primary or secondary outcome. The median number of total outcomes (not including our category of tertiary outcomes for side effects or similar) was seven. While depression is a multifaceted disorder that manifests itself in many ways over multiple domains, there is a need to specify what outcomes are being considered and how they will be accounted for in a clear fashion. No single primary outcome is appropriate for all depression studies.

We also found that determining the number of primary and secondary outcomes for many of the articles included in this study was not straightforward, with relatively few clearly and consistently specifying primary and secondary outcomes [e.g. 28,30,37].

Separate analyses, with no correction for multiplicity, were the most common method to analyze multiple outcomes. A familiar drawback of this approach is the risk of inflating the Type-I error rate (likelihood of obtaining significant results due to chance). While it is critically important that multiple domains of a disorder are discussed, interpretation of a large number of p-values by a clinical reader can be challenging. While we focused on randomized trials, similar issues arise in observational studies. Failure to account for the multiplicity of comparisons could lead to invalid inferences and spurious conclusions. At the very least, researchers reporting a profusion of results without adjustment should address the internal consistency of their findings [38].

The appropriate use of corrections for multiplicity is not always straightforward [38,39]. Rothman [40] notes that scientists need to explore multiple leads in the search for better interventions and treatments, and that inappropriate use of multiplicity adjustment may obscure possibly important findings. Nonetheless, inflation of Type I error is a serious concern, and in the setting of randomized trials, this must be accounted for in the trial protocol. Several papers employed a Bonferroni-type correction to address the issue of multiplicity. A particularly creative approach was undertaken by Lesperance et al [30], where the primary outcome (HAM-D) was tested at $\alpha=0.033$ while the secondary outcome (BDI) was tested at 0.017.

A common critique of the Bonferroni method is that it will tend to be conservative when the outcomes are correlated. However, the simulations of Yoon et al [4] indicated that for settings similar to that of the CATIE trial, with 5 outcomes, the Bonferroni adjustment performed adequately when correlations were moderately. For psychiatric studies, it is rare to have highly correlated endpoints.

Further use of more sophisticated approaches to account for multiplicity may be warranted. Joint testing is particularly attractive in this setting [4,7]. By capitalizing on the correlation of multiple outcomes, these methods are generally more powerful than separate analyses [7] or Bonferroni adjustment [4]. While more complicated than separate testing of multiple outcomes with multiplicity adjustment, these approaches are straightforward to fit in general purpose statistical software [4,7]. Changes in the scale of research and the use of large data banks to test hypotheses will complicate future evidence-based medicine, and will likely exacerbate these issues [41].

Another troubling problem, unrelated to the multiplicity issue, concerns missing data. When outcome data are only partially observed, separate analyses of the outcomes will lead to the inclusion of different subjects for the analysis of each outcome. The reader is then faced with interpreting treatment effects based on different samples of subjects, as well as assessing assumptions regarding missingness. Joint models are particularly attractive in this setting, since they incorporate partially observed data and pool information across outcomes.

The concordance between the published protocols in registries and the number of published outcomes was also discouragingly low, albeit similar to findings reported for cardiology, rheumatology and gastroenterology RCTs [42]. Although the 2007 FDA Modernization Act now requires investigators and sponsors to submit information for any applicable clinical trial to NIH/NLM, complete adherence to this act will require some time before becoming appearing in published trial results. While we anticipate that more investigators will publish their protocols in a more timely and complete fashion as part of new journal requirements, selective reporting remains a potential problem [8,9]. Investigators must not "torture their data until they speak" [38] by examining additional outcomes, undertaking unplanned subgroup analyses or similar mischief. The addition of a CONSORT checklist item to note changes in trial outcomes after the trial commences should also help with this issue.

To help improve practice in this area, we suggest that all clinical trial reports:

1. Clearly specify a single primary outcome of the study (potentially a clinically interpretable composite), or include multiple primary outcomes along with a strategy to account for multiplicity (e.g. adjustment for multiple comparisons or analysis using a joint model or global test),
2. Specify a limited number of secondary outcomes, along with a justification for their inclusion,

3. Report these analytic decisions in the published protocol in a recognized trial registry prior to the start of trial analysis,
4. Ensure that the discussion of these outcomes is consistent in the protocol, abstract, methods, results and tables, and,
5. Consider use of more principled approaches to account for multiple outcomes to help minimize the chance of spurious results due to multiplicity and help to ensure maximal gain of evidence-based knowledge accrues from these important and expensive trials.

Widespread adoption of these recommendations, all of which flow from the CONSORT guidelines and are consistent with the FDA modernization act, could be easily incorporated into common practice. If implemented, they could help improve the timely dissemination and appropriate interpretation of results from clinical trials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Partial support was provided by the National Institute of Mental Health grant R01-MH54693 and the Smith College Tomlinson Fund. Thanks to Ian White and to the anonymous reviewers for many useful comments on a previous draft.

REFERENCES

1. Pocock SJ, Geller NL, Tsiatis AA. The analysis of multiple endpoints in clinical trials. *Biometrics*. 1987; 43:487–498. [PubMed: 3663814]
2. Sankoh AJ, D'Agostino RB Sr, Huque MF. Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoints issues. *Stat Med*. 2003; 22(20):3133–3150. [PubMed: 14518019]
3. Neuhäuser M. How to deal with multiple endpoints in clinical trials. *Fundam Clin Pharmacol*. 2006; 20(6):515–523. [PubMed: 17109645]
4. Yoon F, Fitzmaurice GM, Lipsitz SR, Horton NJ, Normand SL. Alternative methods for testing treatment effects on the basis of multiple outcomes: simulation and case study. In revision.
5. Bretz, F.; Branson, M. Multiple Endpoints. In: D'Agostino; Sullivan, L.; Massaro, J., editors. *The Wiley Encyclopedia of Clinical Trials*. Vol. Volume 3. Hoboken, NJ: Wiley Interscience; 2008. p. 181-186.
6. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: Greater precision but with greater uncertainty? *JAMA*. 2003; 289(19):2554–2559. [PubMed: 12759327]
7. Teixeira-Pinto A, Siddique J, Gibbons R, Normand SL. Statistical approaches to modeling multiple outcomes in psychiatric studies. *Psychiatric Annals*. 2009; 39(7):729–735. [PubMed: 20161512]
8. Al-Marzouki S, Roberts I, Evans S, Marshall T. Selective reporting in clinical trials: analysis of trial protocols accepted by the Lancet. *Lancet*. 2008; 372(9634):201. [PubMed: 18640445]
9. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials. *JAMA*. 2004; 291(20):2457–2465. [PubMed: 15161896]
10. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008; 358:252–260. [PubMed: 18199864]
11. Viereck C, Boudes P. An Analysis of Current Pharmaceutical Industry Practices for Making Clinical Trial Results Publicly Accessible. *Contemp Clin Trials*. 2009; 30(4):293–299. [PubMed: 19348964]

12. Food and Drug Administration. Food and Drug Administration Amendment Act 2007 (FDAAA-2007). Washington, DC: Government Printing Office; 2007.
13. Dartmouth-Hitchcock Medical Center [Online]. Clinical Trials Registration. c2010 [cited 2010 Nov 29]. Available from: http://www.dhmc.org/webpage.cfm?site_id=2&org_id=102&morg_id=0&sec_id=44393&gsec_id=36824&item_id=55844.
14. International Committee of Medical Journal Editors [Online]. Uniform Requirements for Manuscripts Submitted to Biomedical Journals. c2009 [cited 2010 Nov 29]. Available from: <http://www.icmje.org/>.
15. Simon GE, VonKorff M, Barlow W. Health care costs of primary care patients with recognized depression. *Arch Gen Psychiatry*. 1995; 52(10):850–856. [PubMed: 7575105]
16. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas*. 1977; 1(3):385–401.
17. Beck AT, Ward CH, Mendelson M, Mock JE, Erbaugh JK. An inventory for measuring depression. *Arch Gen Psychiatry*. 1961; 4:561–571. [PubMed: 13688369]
18. Beck AT, Steer RA, Ball R, Ranieri W. Comparison of Beck Depression Inventories IA and II in psychiatric outpatients. *J Pers Assess*. 1996; 67(3):588–597. [PubMed: 8991972]
19. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 2009; 23:56–62. [PubMed: 14399272]
20. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979; 134:382–389. [PubMed: 444788]
21. Lautrette A, Darmon M, Megarbane B, Joly LM, Chevret S, Adrie C, et al. A communication strategy and brochure for relatives of patients dying in the ICU. *N Engl J Med*. 2007; 356(5):469–478. [PubMed: 17267907]
22. Harris AHS, Reeder R, Hyun JK. Common statistical and research design problems in manuscripts submitted to high-impact psychiatry journals: what editors and reviewers want authors to know. *J Psychiatr Res*. 2009; 43(15):1231–1234. [PubMed: 19435635]
23. Bolton P, Bass J, Betancourt T, Speelman L, Onyango G, Clougherty K, et al. Interventions for depression symptoms among adolescent survivors of war and displacement in northern Uganda: a randomized controlled trial. *JAMA*. 2007; 298(5):519–527. [PubMed: 17666672]
24. Bryant RA, Mastrodomenico J, Felmingham K, Hopwood S, Kenny L, Kandris E, et al. Treatment of acute stress disorder: a randomized controlled trial. *Arch Gen Psychiatry*. 2008; 65(5):659–667. [PubMed: 18519824]
25. Eranti S, Mogg A, Pluck G, Landau S, Purvis R, Brown R, et al. A randomized, controlled trial with 6-month follow-up of repetitive transcranial magnetic stimulation and electroconvulsive therapy for severe depression. *Am J Psychiatry*. 2007; 164(1):73–81. [PubMed: 17202547]
26. Nurnberg H, Hensley P, Heiman J, Croft H, Debattista C, Paine S. Sildenafil treatment of women with antidepressant-associated sexual dysfunction: A randomized controlled trial. *JAMA*. 2008; 300(4):395–404. [PubMed: 18647982]
27. Sultzer DL, Davis SM, Tariot P, Dagerman K, Lebowitz B, Lyketsos C, et al. Clinical symptom responses to atypical anti-psychotic medications in Alzheimer's disease: phase 1 outcomes from the CATIE-AD effectiveness trial. *Am J Psychiatry*. 2008; 165(7):844–854. [PubMed: 18519523]
28. Strong V, Waters R, Hibberd C, Murray G, Wall L, Walker J, et al. Management of depression for people with cancer (SMaRT oncology 1): a randomized trial. *Lancet*. 2008; 372:40–48. [PubMed: 18603157]
29. Welton AJ, Vickers MR, Kim J, Ford D, Lawton BA, MacLennan AH, et al. Health related quality of life after combined hormone replacement therapy: a randomized controlled trial. *BMJ*. 2008; 337:a1190. [PubMed: 18719013]
30. Lespérance F, Frasur-Smith N, Koszycki D, Laliberté M, van Zyl L, Baker B, et al. Effects of citalopram and interpersonal psychotherapy on depression in patients with coronary artery disease: the Canadian Cardiac Randomized Evaluation of Antidepressant and Psychotherapy Efficacy (CREATE) trial. *JAMA*. 2007; 297(4):367–379. [PubMed: 17244833]
31. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*. 1988; 75(4):800–802.

32. Raskin J, Wiltse CG, Siegal A, Sheikh J, Xu J, Dinkel JJ, et al. Efficacy of duloxetine on cognitive, depression, and pain in elderly patients with major depressive disorder: an 8-week, double-blind, placebo-controlled trial. *Am J Psychiatry*. 2007; 164(6):900–909. [PubMed: 17541049]
33. Goldberg JF, Perlis RH, Ghaemi SN, Calabrese JR, Bowden CL, Wisniewski S, et al. Adjunctive antidepressant use and symptomatic recovery among bipolar depressed patients with concomitant manic symptoms: Findings from the STEP-BD. *Am J Psychiatry*. 2007; 164:1348–1355. [PubMed: 17728419]
34. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001; 134(8):663–694. [PubMed: 11304107]
35. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*. 2001; 285(15):1987–1991. [PubMed: 11308435]
36. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomized trials. *J Clin Epidemiol*. 2010; 63(3):834–840. [PubMed: 20346629]
37. Robinson RG, Jorge RE, Moser DJ, Acion L, Solodkin A, Small SL, et al. Escitalopram and problem-solving therapy for prevention of post stroke depression: A randomized controlled trial. *JAMA*. 2008; 299(20):2391–2400. [PubMed: 18505948]
38. Schulz KF, Grimes DA. Multiplicity in randomized trials 1: endpoints and treatments. *Lancet*. 2005; 365:1348–1353. [PubMed: 15823387]
39. Millis SR. Statistical Practices: The Seven Deadly Sins. *Child Neuropsychology*. 2003; 9(3):221–233. [PubMed: 13680411]
40. Rothman, KJ.; Greenland, S. *Modern Epidemiology*. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 1998.
41. Enhancing the Vitality of the National Institutes of Health: Organizational Change to Meet New Challenges. National Research Council (US) and Institute of Medicine (US) Committee on the Organizational Structure of the National Institutes of Health. Washington (DC): National Academies Press (US); 2003.
42. Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA*. 2009; 302(9):977–984. [PubMed: 19724045]

Figure 1.
Flowchart of article inclusion and exclusion

Figure 2.
Boxplot of primary, secondary, and total (primary + secondary) number of outcomes excluding extreme outliers (total number of outcomes > 18)

Table 2

Frequency of use of primary depression outcomes (for those reported by more than one study)

Outcome	Frequency reported
Hamilton Depression Rating Scale	11
Clinical Global Impression	4
DSM-IV diagnosis of depression	4
Montgomery-Asberg Depression Rating Scale	4
Beck Depression Inventory	3
Children's Depression Rating Scale	3
Hospital Anxiety and Depression Score	3
Quick Inventory of Depressive Symptoms	2

Table 3

Cross-classification of grouping of number of primary outcomes by sample size grouping

Sample Size Group	Number of Primary Outcomes (Row Percentage)			
	<i>1</i>	<i>2-3</i>	<i>4+</i>	<i>Total</i>
<i>n<100</i>	4 (33%)	3 (25%)	5 (42%)	12 (100%)
<i>100<=n<400</i>	22 (73%)	5 (17%)	3 (10%)	30 (100%)
<i>n>=400</i>	3 (23%)	2 (15%)	8 (62%)	13 (100%)
Total	29 (53%)	10 (18%)	16 (29%)	55 (100%)