NIH Public Access
Author Manuscript
Med Decis Making. Author manuscript; available in PMC 2010 January 1.

Published in final edited form as:
Med Decis Making. 2009 ; 29(1): 91–103. doi:10.1177/0272989X08323298.

Controlling for Drug Dose in Systematic Review and Meta-Analysis: A Case Study of the Effect of Antidepressant Dose

Richard A. Hansen, PhD[a,b], **Charity G. Moore, PhD**[c], **Stacie B. Dusetzina, BA**[b], **Brian I. Leinwand, MSPH**[b], **Gerald Gartlehner, MD, MPH**[d,e], and **Bradley N. Gaynes, MD, MPH**[f]

b*University of North Carolina – School of Pharmacy*

c*University of Pittsburgh – School of Medicine*

d*Danube University, Department for Clinical Epidemiology and Evidence-based Medicine*

e*Ludwig Boltzmann Institute for Health Technology Assessments, Vienna, Austria*

f*University of North Carolina – School of Medicine*

Abstract

Purpose—To describe a method for quantitatively dealing with drug dose in comparative effectiveness reviews. Second-generation antidepressants are used as an example to illustrate this method, and to determine whether dose influences conclusions on comparative effectiveness.

Methods—Studies previously identified in a systematic review of second-generation antidepressants were included if data on drug dose were available. The usual dosing range for each drug was defined, and then used to create 2- and 3-level dose categories. Placebo-controlled data were used to calculate overall effect sizes for the drug class and effect sizes stratified by drug dose. Meta-regression tested the impact of dose on effect size. Weighted mean differences and risk ratios were calculated for comparative studies, stratifying by whether compared doses were equivalent.

Results—The dose classification method was able to identify dose-response trends in the context of meta-analysis. Compared to low dose studies, medium and high dose studies had a 1 to 2 point greater differential in mean HAM-D change ($P<0.001$). Dose was not a statistically significant predictor of categorical HAM-D response. Among comparative trials with non-equivalent doses, trends favored higher dose categories, but generally were not statistically significant.

Conclusions—A structured method for quantitatively dealing with drug dose in comparative effectiveness reviews is described, with application to the second-generation antidepressants. Dose-dependent reductions in HAM-D scores were identified, although differences did not translate into better response rates for higher doses. Dose equivalency was not a significant factor among comparative studies in second-generation antidepressants.

Keywords

meta-analysis; meta-regression; dose; dose-response; comparative effectiveness

aCorresponding author and reprint address: University of North Carolina at Chapel Hill, School of Pharmacy; Division of Pharmaceutical Outcomes and Policy, CB 7360; 2205 Kerr Hall, Chapel Hill, NC 27599, Phone: (919) 966-7517, Fax: (919) 966-8486, Email: rahansen@unc.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

INTRODUCTION

An abundance of information about existing and emerging health care interventions is readily available, posing challenges for clinicians, patients and researchers to determine the best available treatment at a given time.[1] Among methods for assessing the comparative effectiveness of health care interventions, systematic reviews are an important tool for enhancing the ability of health care professionals to interpret and apply research evidence in a timely manner.[2, 3] Meta-analysis – the statistical analysis of the results from multiple independent studies – is a useful method for increasing statistical power and the precision of the estimate of treatment effect.[4] Systematic review and meta-analysis have been widely used in comparing different pharmaceutical treatments, as exemplified by the Cochrane Collaboration, the Agency for Healthcare Research and Quality, and the Drug Effectiveness Review Project.

Although systematic review and meta-analysis are valuable for comparing effects of different medications, conclusions are often limited by available evidence. One problem unique to pharmaceutical comparisons is the issue of drug dose. When comparing the benefits or harms of one treatment to another, comparable doses rarely are defined. Clinicians generally have a sense of how the dose of one drug compares to the dose of an alternative drug, although these distinctions usually are not evidence-based. This often leads to concerns as to whether differences – or lack of differences – in comparative effectiveness are related to drug dose.

Considering second-generation antidepressants, for example, a general *rule of thumb* may be that fluoxetine 20mg is roughly equivalent to sertraline 100mg or escitalopram 10mg. If this *rule of thumb* is true and we compare these drugs at equivalent doses, we assume that any differences in comparative effectiveness are attributed to the drug rather than the drug dose. However, how do you interpret the results of a study that compares non-equivalent doses such as fluoxetine 20mg, sertraline 200mg, and escitalopram 20mg? In the same regard, looking across studies or combining data from differently designed studies, can anything be said about the effect size of fluoxetine compared to the effect sizes of sertraline or escitalopram if relative drug doses differed?

This example presents several problems. First, we assume that increases in drug dose are related to an increased effectiveness, indicating a positive dose-response relationship. A clear dose-response relationship is well established for some drug classes (e.g., inhaled corticosteroids follow a logarithmic dose-response curve),[5,6] but is less clear for other drug classes such as the antidepressants. Second, there is a general lack of standardized dosing guidelines for assessing the comparative effectiveness of different drugs within the same therapeutic class. Clinicians generally have a sense of how doses of different drugs within the same class compare, but universally accepted comparative dosing evidence is not available for most drugs. Mechanisms such as the World Health Organization's Defined Daily Dose exist, but this establishes an average therapeutic dose rather than a range of comparable doses. Third, methods to handle drug dose in comparative effectiveness studies are not well developed. Most systematic reviews and meta-analyses handle the issue of drug dose qualitatively – narratively describing differences. Quantitative methods for dealing with dose differences include stratification[7,8] and meta-regression.[9] Stratification allows for comparison of the relative risk or odds ratio across dosing levels while meta-regression explores the influence of dose on the effect size across trials. Still, these methods assume that some form of comparative dose equivalence guideline exists and require that a sufficient number of eligible studies exist for analysis. Unfortunately for researchers and health care professionals, studies using such methods are the exception rather than the norm.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The purpose of this paper is to illustrate a dose classification method for quantitatively addressing if drug dose is an important factor to consider in a meta-analysis. We base our discussion on the analysis of studies previously identified by a systematic review of the comparative effectiveness of second-generation antidepressants. Our specific objectives are:

1. to describe a simple method for detecting inequalities in dose when comparing across different drugs in the same therapeutic class,
2. to describe the relationship between dose and effect size for second-generation antidepressants, and
3. to determine whether the comparative effectiveness of second-generation antidepressants is related to differences in drug dose.

METHODS

Comparative Doses

In comparing doses across drugs, we first defined a usual dosing range for each drug. This range initially was based on U.S. Food and Drug Administration (FDA) labeled dosing recommendations. Because the FDA-approved range does not necessarily capture the spectrum of doses used in clinical practice or represented in clinical studies,[10, 11–12] we examined the lower and upper bounds of this range carefully. As a result, the upper bound for duloxetine was increased to 120mg/day,[13,14] and the lower bound for some drugs (e.g., immediate release bupropion, fluoxetine, and paroxetine) was decreased. Using this modified dosing range, we then defined cut-points to create a two- and three-level dose classification system (Table 1). The two-level classification system defined high doses as greater than the midpoint of the range; low doses encompassed doses at or below the midpoint of the range. The three-level classification system defined cut-points using the lower and upper 25th percentiles of the range; low doses were below the 25th percentile of the range, medium doses were greater than or equal to the 25th percentile but less than or equal to the 75th percentile of the range, and high doses were greater than the 75th percentile of the range. Because some studies defined doses outside of our pre-specified ranges, we defined doses smaller than the low dose boundary of the range as low, and doses greater than the upper boundary of the range were included in the high dose category. More finite cut-points (e.g., 4-level or 5-level) also were considered, although we realized that a dose classification method needed to capture realistic dose differences from a clinician's perspective, and sufficient evidence must be available to make realistic comparisons for each dose level.

Some clinical trials define fixed drug doses, while other studies allow clinicians to increase or decrease doses based on clinical response and side effects (i.e., flexible dosing). Although meta-analytic methods considering the effect of drug dose would ideally be limited to fixed dose trials,[9] we were concerned that focusing only on fixed dose trials would have limited application in comparative effectiveness reviews. Thus, we included flexible dose trials and defined the dose category based on the mean dose used among trial participants. We tested for potential regression to the mean among flexible dosing studies by conducting a sensitivity analysis that excluded flexible dose trials.

Study Eligibility

We analyzed randomized controlled trials that compared one second-generation antidepressant to another or to placebo. Trials had to be 1) at least 6 weeks in duration, 2) report a fixed dose or a mean dose (i.e., for flexible dose trials), 3) assess treatment effect using the Hamilton Depression Rating Scale (HAM-D), 4) have a mean population age $\leq$ 65 years, and 5) conducted primarily in outpatients. We excluded relapse or recurrence prevention studies, as well as studies that were conducted in patients with mild or subsyndromal depression (defined by a

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

mean baseline HAM-D < 18). Studies conducted in special populations (e.g., HIV, Substance Abuse Disorder, etc.) also were excluded.

Data Abstraction

For each trial we abstracted the drug, dose, duration, mean age of participants, mean and standard deviation for the baseline, endpoint, and change in the HAM-D, and the number of randomized participants characterized as a responder at endpoint. Only studies that defined response as ≥ 50% improvement in HAM-D score from baseline to endpoint were included in the responder analysis. The standard deviation for the mean HAM-D change was recorded when available; for trials that did not report a standard deviation for the mean change, we imputed a standard deviation using the mean of the standard deviations from trials that had a similar sample size and baseline HAM-D scores.

Dose and Effect Size

The goal of this analysis was to illustrate the influence of dose on effect size estimates in the context of meta-analysis. Studies were included if they compared active treatment to placebo using the mean HAM-D change as a continuous variable or a count of the number of responders as a binary variable. For studies that compared more than one active treatment to placebo, both drug-placebo comparisons were included, but the sample size of the placebo arm was reduced proportionately for each comparison so as not to double count the placebo group. We conducted stratified meta-analysis for HAM-D change and responder status to obtain pooled estimates for each dose category (2-level = low and high; 3-level = low, medium, and high). For the mean HAM-D change, we pooled the weighted mean differences between active treatment and placebo for each dose stratum. For the analysis of responders versus non-responders, the natural log risk ratios were used to estimate a pooled risk ratio within each dose stratum.

We then pooled all of the studies in a meta-regression to assess the impact of dose on the differences in HAM-D and risk ratios. For the two-level dosing scheme, high dose was compared to low dose. The 3-level dosing scheme compared medium dose to low dose and high dose to low dose. For these analyses we assumed that individual drugs do not differ substantially,[15,16] but still controlled for the effect of drug in the meta-regression. Between-study variance was compared for each model by looking at changes in tau^2.

Comparative Effectiveness for Equivalent vs. Non-Equivalent Doses

To assess whether dose influences conclusions regarding the comparative effectiveness of second-generation antidepressants, we included randomized trials that compared one second-generation antidepressant to another. Studies were stratified by whether they compared equivalent or non-equivalent doses. For non-equivalent doses in the 3-level dose classification, we treated the lower of the two doses as the reference group (i.e., medium compared to low, high compared to low, or high compared to medium). Meta-analysis of the binary response outcome and weighted mean difference in HAM-D change were conducted first for all comparative trials, and then for the equivalent and non-equivalent dose strata for the 2- and 3-level dose classification systems.

All meta-analyses used a random effects model. Heterogeneity among trial was quantified using I^2,[17,18] which measures the percentage of inconsistency across studies that is due to heterogeneity. Individual studies were not removed from the analysis based on their apparent contribution to heterogeneity. All analyses were conducted using the “metan” and “metareg” commands installed in Stata 9.1 (StataCorp, College Station, TX).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

RESULTS

Included Studies

Of the 94 head-to-head and 64 placebo-controlled trials included in the initial systematic review,[16] 74 studies were eligible for this analysis (Figure 1). Among these, 34 provided data for a head-to-head comparison and 46 provided placebo-controlled data. Six studies contributed both a head-to-head comparison and one or more placebo comparisons. All second-generation antidepressants included in our analysis were represented by at least one placebo-controlled study, although some drugs (e.g., escitalopram and mirtazapine) were represented by fewer studies than others (Table 2). Among included head-to-head trials, many possible drug-drug comparisons were not represented. All comparisons included at least one SSRI; no studies comparing an SNRI to an SNRI or an SNRI to an "other" drug were eligible for inclusion (Table 3).

Placebo-Controlled Data

Weighted mean difference in HAM-D scores—The weighted mean difference in HAM-D change between active treatment and placebo was calculated for all studies, and then separately for each stratum of the 2- and 3-level dose classification system (Figure 2). Overall, the inference drawn from all studies was similar to inferences drawn from each dose stratum. Still, we found a larger difference for higher doses, especially when using the 3-level dose classification. The 2-level classification showed a small trend favoring higher doses, although confidence intervals for the pooled estimates were overlapping for the low and high dose stratum. The meta-regression confirmed the dose-response trends illustrated in Figure 2, even after controlling for drug. For the 3-level dose classification, the weighted mean difference in HAM-D scores was 1.6 points larger for medium dose compared to low dose (95% CI: −2.2 to −0.9), and 2.3 points larger for high dose compared to low dose (95% CI: −3.4 to −1.2). The difference in HAM-D scores was 0.7 HAM-D points greater for high dose compared to low dose in the 2-level classification after controlling for drug (95% CI: −1.5 to 0.1). The estimated between-study variance (tau^2) was reduced from 0.77 to 0.39 in the 3-level model, but did not change significantly in the 2-level model. These findings were similar in the sensitivity analyses, where only fixed dose studies were included (data not shown).

Responder analysis—Risk ratios – ratios reflecting the likelihood of HAM-D response for active treatment over the likelihood for placebo – were calculated for all studies and then separately for the 2- and 3-level dose strata (Figure 3). Consistent with the analysis of mean HAM-D change, inference drawn from all studies was generally similar to inferences drawn from each dose stratum. The risk ratios were larger for higher dose categories compared to lower dose categories, suggesting better efficacy for higher doses. However, these trends were not statistically significant. In the meta-regression controlling for drug, the 3-level dose classification risk ratio was 0.08 larger for medium dose compared to low dose (95% CI: −0.2 to 0.4) and 0.13 larger for high dose compared to low dose (95% CI: −0.4 to 0.6). In the meta-regression for the 2-level dose classification, the risk ratio for high dose was 0.1 higher than low dose (95% CI: −0.2 to 0.4), again controlling for drug. The estimated between-study variance (tau^2) did not change significantly when adding the 3-level or 2-level dose variables to the model. The sensitivity analysis of only fixed dose studies revealed similar trends (data not shown).

Head-to-Head Data

Mean difference in HAM-D scores—The overall weighted mean difference in HAM-D change was approximately the same (−0.1; 95% CI: −0.4 to 0.2) between arms of comparative trials. When trials were separated by whether equal doses were compared using the 3-level dose classification (Figure 4), the weighted mean difference in HAM-D change between active

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

treatments was not statistically significantly different from zero for equivalent dose studies (−0.02; 95% CI: −0.4 to 0.4), nor for non-equivalent dose studies (0.4; 95% CI: −0.3 to 1.1). However, the point estimate favored the higher dose treatment arms, illustrating the potential over estimation of effect size when compared doses were not equivalent. The same was true using the 2-level dose classification for equivalent dose studies (0.2; 95% CI −0.2 to 0.6). Higher dose arms showed more change in the HAM-D than lower dose arms in the non-equivalent dose studies (0.7; 95% CI: 0.1 to 1.2). In other words, high dose comparators had nearly a 1-point larger reduction in HAM-D scores than lower dose comparators.

Responder analysis—The comparative trials showed similar response rates between arms (RR=0.99; 95% CI: 0.94 to 1.03). When studies were stratified by whether they compared equivalent or non-equivalent doses (Figure 5), risk ratios again did not differ significantly from 1.0. This finding was consistent for the 3-level and 2-level classification method. The point estimate favored the higher doses for the 2-level classification, again illustrating the potential over estimation of effect size when compared doses were not equivalent.

Consistency of Effects across Studies

Heterogeneity of treatment effects (i.e., I^2) generally was low to moderate[18] across studies, even after stratifying by dose. The largest heterogeneity was detected for the analysis of comparative trials with non-equivalent doses, particularly with the 3-level dose classification. The number of constituent trials was relatively small for these dose strata.

DISCUSSION

Evaluating the dose-response relationship for a single drug is relatively straightforward, but comparing across different therapeutic agents with different relative doses is a more complicated endeavor. With the exception of therapeutic classes that have physiologic data available to define comparable doses, previous meta-analyses evaluating the effect of drug dose have used ad hoc methods to define doses, or have ignored dose completely. We demonstrate a structured method for use in meta-analysis to account for inequalities in drug doses across different therapeutic agents. This method is based on the usual dosing range of constituent agents within the selected drug group, with structured cut-points to define dose levels. Our method does not treat a single drug in the therapeutic group as a reference drug, as with some previous meta-analyses (e.g., haloperidol equivalents[19] and imipramine-equivalents[20]). Our method also does not require researchers conducting systematic reviews to extensively analyze the dose-response relationship for individual agents.[21] Instead, we assume that the dosing range established by phase I-III trials and reported in product labeling provides an adequate balance among safety, tolerability, and efficacy. We also assume that deviations in these dosing recommendations will be reflected in clinical guidelines and prescribing behavior. We believe that our method allows for straightforward and generalized application in comparative effectiveness reviews, and is likely to capture gross inequalities in comparing different doses of different treatments.

Using second-generation antidepressant as an example, we found that both a 2- and 3-level dose classification system were able to identify dose-response trends in the context of meta-analysis. For active treatment compared to placebo, the weighted mean difference in HAM-D change varied across doses, with slightly larger improvements at higher doses. Dose trends also were identified in the risk ratio analysis of HAM-D response, although the risk ratio confidence intervals were generally overlapping among dose categories. Among comparative trials with non-equivalent doses, the weighted mean difference in HAM-D change favored higher doses, but was statistically significant only for the 2-level dose classification. Overall, the 3-level classification was more sensitive than the 2-level classification for detecting

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

differences in effect size. The 3-level classification also is less likely to be influenced by dose misclassification compared with the 2-level classification. Still, the 2-level classification may have utility for reviews that identify too few studies to create 3 dose strata.

Even though we observed an average 1-2 HAM-D point improvement in response for incremental increases in dose, the dose-response curve was relatively flat. This may be partially related to the strong placebo effect observed in antidepressant trials.[22] The relatively small differences in the change in HAM-D that we identified across dose strata are likely not clinically significant. This is consistent with the fact that the likelihood of actually responding to the drug as defined by the change in the HAM-D did not differ significantly across dose strata. Although this could be related to the sensitivity of our dose classification method, it is generally consistent with the small incremental improvements in efficacy previously demonstrated with antidepressants.[11–12, 23] Our analysis also is consistent with previous studies in that we found the effect size for low dose studies to be the most disparate from other evidence (i.e., using the 3-level classification in Figure 2). This parallels the findings of a meta-analysis of first-generation antidepressants that converted drugs to imipramine-equivalent doses and concluded that higher doses did not provide increased efficacy, but that the low doses showed a reduction in efficacy.[20]

In the context of the effectiveness of second-generation antidepressants, we found that drug dose does not significantly influence the evidence base for clinical decision-making. Still, this analysis does not support ignoring dose in other comparative effectiveness reviews. Instead, we advocate for testing our method in other therapeutic areas. In particular, we believe this method should be tested among drug classes that have a clear dose-response relationship. In such a case, we believe our method could help differentiate equivalent dose from non-equivalent dose trials, and help in interpretation of the evidence base. Furthermore, we did not explore the relationship between dose and adverse events. Given that clinicians need to balance evidence of effectiveness with harms, this is an important unanswered question. In fact, evidence from a recent meta-analysis of rare but harmful effects of anti-TNF drugs illustrates this point.[8] The overall results of their analysis revealed a significant risk of malignancy and infection. When stratified by dose, however, these risks were found to be significantly greater among high dose regimens; ignoring dose would have led the authors to different conclusions.

Limitations

Our dose classification method is relatively easy to use for comparing the effectiveness of diverse medications, but it requires that sufficient evidence is available to allow for stratification. Ideally, we would have limited our analysis to dose comparisons conducted within studies, rather than between studies. Data were insufficient in quantity to limit evidence to this type of design. Furthermore, in examining the relationship between dose and effect size, we assumed that individual drugs did not differ significantly,[15,16] and that unadjusted differences in effect size were related to dose rather than drug. To support this, our overall pooled risk ratio (RR=1.01; 95% CI 0.97 to 1.06) was generally similar to risk ratios from individual drug comparisons (e.g., fluoxetine vs. paroxetine RR = 1.09 (95% CI 0.99 to 1.21); fluoxetine vs. sertraline RR = 1.11 (95% CI 1.01 to 1.21).[16] Still, we controlled for drug in our meta-regression and found the dose-effect to be of comparable magnitude to the unadjusted values. In comparative assessments that pooled data across multiple different comparisons, it is feasible that the random order in which drugs were included in the analysis could affect the overall pooled estimate. To test whether this was apparent in our analysis, we randomly shuffled the order of drug comparisons multiple times and compared the pooled estimates. Although the pooled estimates were very similar, it is still possible that some remaining drug-specific effect biases our conclusions.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Because our method is rooted in dosing ranges approved by regulatory authorities, it is dependent on initial dose-ranging trials required for approval. These trials, and the ultimate dosing recommendations reported in product labeling, reflect a balance among safety, tolerability, and efficacy. It is possible that this dosing range might not fully capture the dose response curve for different drugs, thus introducing potential misclassification bias for our method. We closely evaluated the upper and lower limits of approved doses and modified the range based on other evidence, known clinical practice patterns, and our own clinical assessment. In the end, the dosing range that we used to define dose categories has some degree of subjective assessment. Finally, it is generally recommended to exclude flexible dose studies when conducting meta-regression on the effect of dose.[9] We included the mean dose from flexible-dose studies, but conducted a sensitivity analysis to determine whether this significantly impacted our conclusions. Excluding the flexible-dose data did not influence our results. Still, the potential biases that occur when doses are titrated to effect or adjusted downward based on tolerability should not be disregarded.

CONCLUSIONS

We present a structured method for use in meta-analysis to account for inequalities in doses across different therapeutic agents used to treat the same condition. Using this method, we illustrate the relationship of second-generation antidepressant dose with estimates of effect size, and demonstrate that only minimal differences exist among agents across comparative trials – even in studies that compared non-equivalent doses. We advocate for the use of this method in comparative effectiveness reviews, especially when studies compare potentially non-equivalent doses.

Acknowledgements

Sources of Support

Initial funding for this research was provided to the Cecil G. Sheps Center for Health Services Research through a sub-contract with the Center for Evidence-Based Policy; Oregon Health & Science University, and through a contract from the Agency for Healthcare Research and Quality to the RTI International-University of North Carolina Evidence-based Practice Center (contract no. 290-02-0016). The analysis uses studies identified during this work, but the current analysis was unfunded. Dr. Hansen is funded by grant K12 RR023248.

References

1. Clancy, C. Statement before the U.S. House of Representatives Subcommittee on Health Committee on Ways and Means. Rockville, MD: Agency for Healthcare Research and Quality; 2007 Jun 12. AHRQ's research efforts in comparative effectiveness. Available at http://www.ahrq.gov/news/sp061207.htm
2. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Ann Intern Med 1997;126(5):376–380. [PubMed: 9054282]
3. Institute of Medicine. Learning what works best: the Nations need for evidence on comparative effectiveness in health care. 2007 Sep. Available at http://www.iom.edu/ebm-effectiveness
4. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. Ann Intern Med 1997;127(9):820–826. [PubMed: 9382404]
5. Bousquet J, Ben-Joseph R, Messonnier M, Alemao E, Gould AL. A meta-analysis of the dose-response relationship of inhaled corticosteroids in adolescents and adults with mild to moderate persistent asthma. Clin Ther 2002;24(1):1–20. [PubMed: 11833824]
6. Kelly HW. Pharmaceutical characteristics that influence the clinical efficacy of inhaled corticosteroids. Ann Allergy Asthma Immunol 2003;91(4):326–334. [PubMed: 14582810]quiz 334-5, 404
7. Zhang J, Ding EL, Song Y. Adverse effects of cyclooxygenase 2 inhibitors on renal and arrhythmia events: meta-analysis of randomized trials. Jama 2006;296(13):1619–1632. [PubMed: 16968832]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

8. Bongartz T, Sutton AJ, Sweeting MJ, Buchan I, Matteson EL, Montori V. Anti-TNF antibody therapy in rheumatoid arthritis and the risk of serious infections and malignancies: systematic review and meta-analysis of rare harmful effects in randomized controlled trials. Jama 2006;295(19):2275–2285. [PubMed: 16705109]
9. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002;21(11):1559–1573. [PubMed: 12111920]
10. Fava M. Diagnosis and definition of treatment-resistant depression. Biol Psychiatry 2003;53(8):649–659. [PubMed: 12706951]
11. Bech P, Tanghoj P, Andersen HF, Overo K. Citalopram dose-response revisited using an alternative psychometric approach to evaluate clinical effects of four fixed citalopram doses compared to placebo in patients with major depression. Psychopharmacology (Berl) 2002;163(1):20–25. [PubMed: 12185396]
12. Monteleone P, Gnocchi G. Evidence for a linear relationship between plasma trazodone levels and clinical response in depression in the elderly. Clin Neuropharmacol 1990;13:S84–S89. [PubMed: 2379183]
13. Detke MJ, Wiltse CG, Mallinckrodt CH, McNamara RK, Demitrack MA, Bitter I. Duloxetine in the acute and long-term treatment of major depressive disorder: a placebo- and paroxetine-controlled trial. Eur Neuropsychopharmacol 2004;14(6):457–470. [PubMed: 15589385]
14. Pritchett YL, Marciniak MD, Corey-Lisle PK, Berzon RA, Desaiah D, Detke MJ. Use of effect size to determine optimal dose of duloxetine in major depressive disorder. J Psychiatr Res 2007;41(3–4): 311–318. [PubMed: 16934840]
15. Hansen RA, Gartlehner G, Lohr KN, Gaynes BN, Carey TS. Efficacy and safety of second-generation antidepressants in the treatment of major depressive disorder. Ann Intern Med 2005;143(6):415–426. [PubMed: 16172440]
16. Gartlehner, G.; Hansen, RA.; Thieda, P., et al. Comparative Effectiveness Review No. 7. (Prepared by RTI International-University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016.). Rockville, MD: Agency for Healthcare Research and Quality; 2007 Jan. Comparative effectiveness of second-generation antidepressants in the pharmacologic treatment of adult depression. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm
17. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21(11): 1539–1558. [PubMed: 12111919]
18. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Bmj 2003;327(7414):557–560. [PubMed: 12958120]
19. Davis JM, Chen N, Glick ID. A meta-analysis of the efficacy of second-generation antipsychotics. Arch Gen Psychiatry 2003;60(6):553–564. [PubMed: 12796218]
20. Bollini P, Pampallona S, Tibaldi G, Kupelnick B, Munizza C. Effectiveness of antidepressants. Meta-analysis of dose-effect relationships in randomised clinical trials. Br J Psychiatry 1999;174:297–303. [PubMed: 10533547]
21. Davis JM, Chen N. Dose response and dose equivalence of antipsychotics. J Clin Psychopharmacol 2004;24(2):192–208. [PubMed: 15206667]
22. Posternak MA, Zimmerman M. Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials: meta-analysis. Br J Psychiatry 2007;190:287–292. [PubMed: 17401033]
23. Ruhe HG, Huyser J, Swinkels JA, Schene AH. Dose escalation for insufficient response to standard-dose selective serotonin reuptake inhibitors in major depressive disorder: systematic review. Br J Psychiatry 2006;189:309–316. [PubMed: 17012653]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

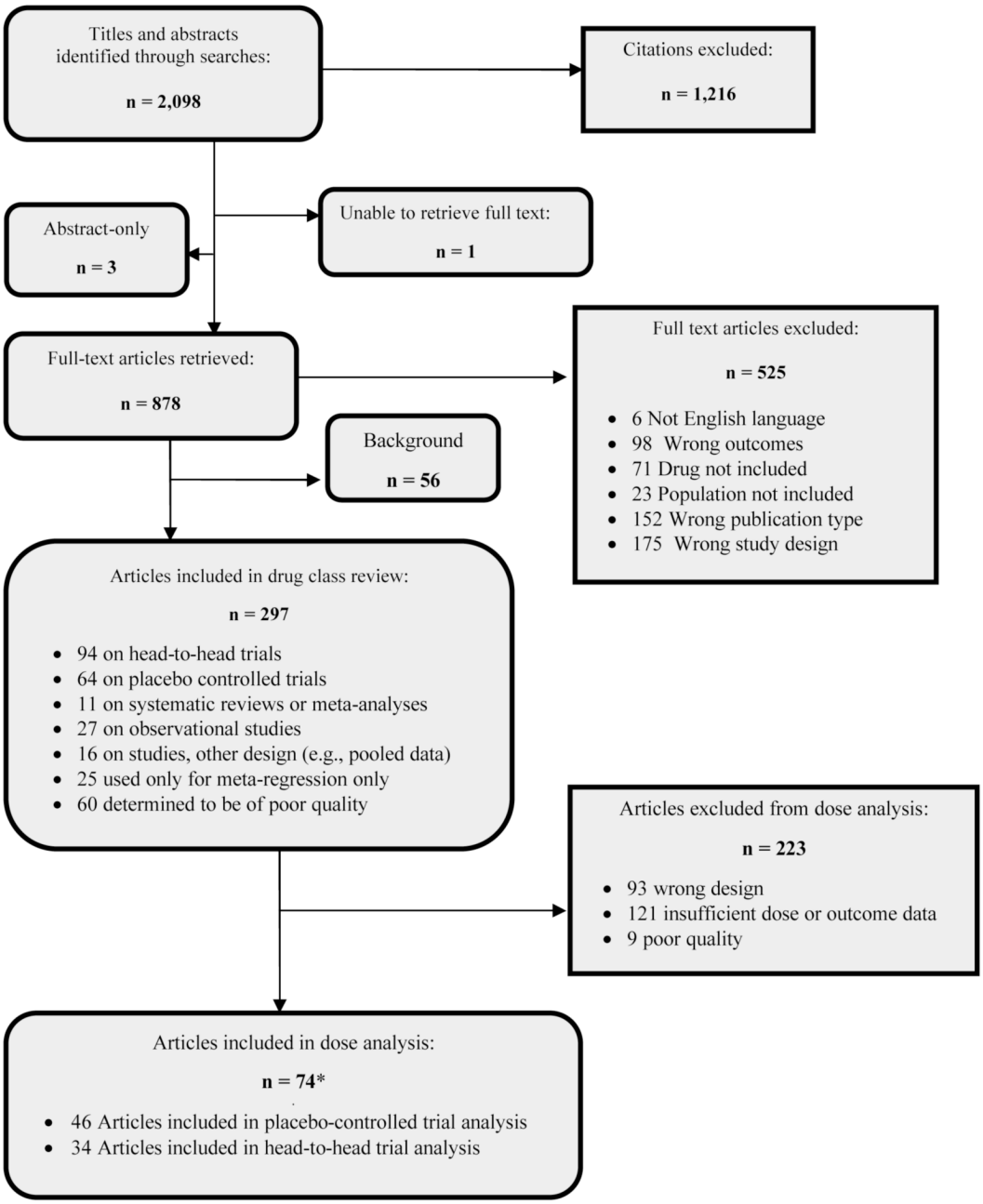

Figure 1.
QUOROM flow diagram for dose analysis
* Studies that included both head-to-head and placebo comparisons were included in both analyses (n = 6)

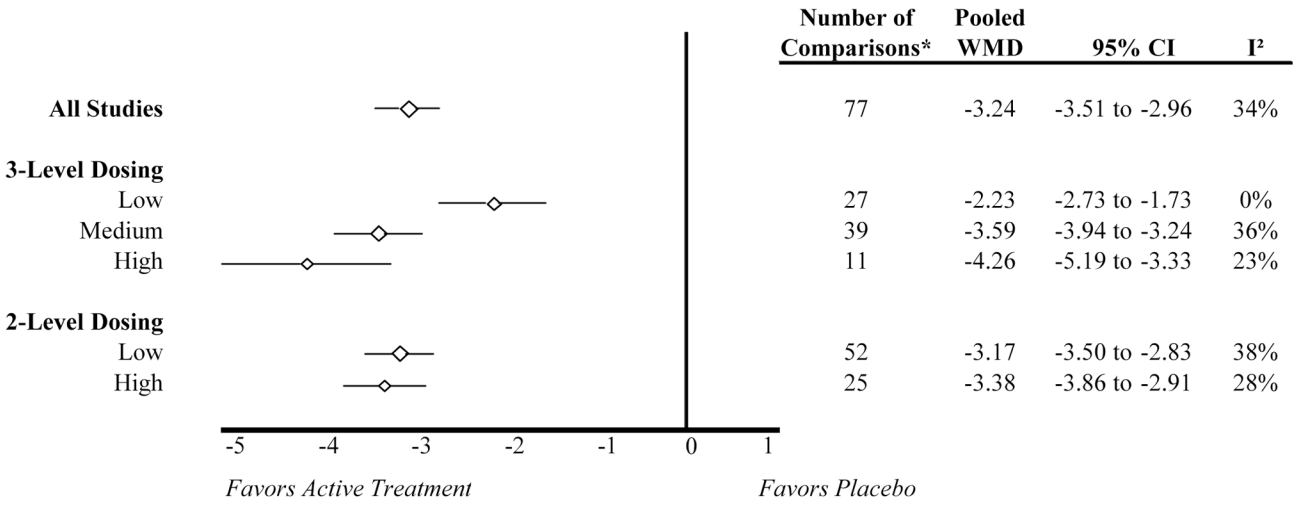

Figure 2.

Weighted mean difference in HAM-D change for placebo-controlled trials stratified by dose

* The number of *comparisons* differs from the number of *articles* included in the systematic review because some studies contributed more than one placebo-controlled comparison, and some studies contributed data to the responder analysis but not the HAM-D change analysis

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

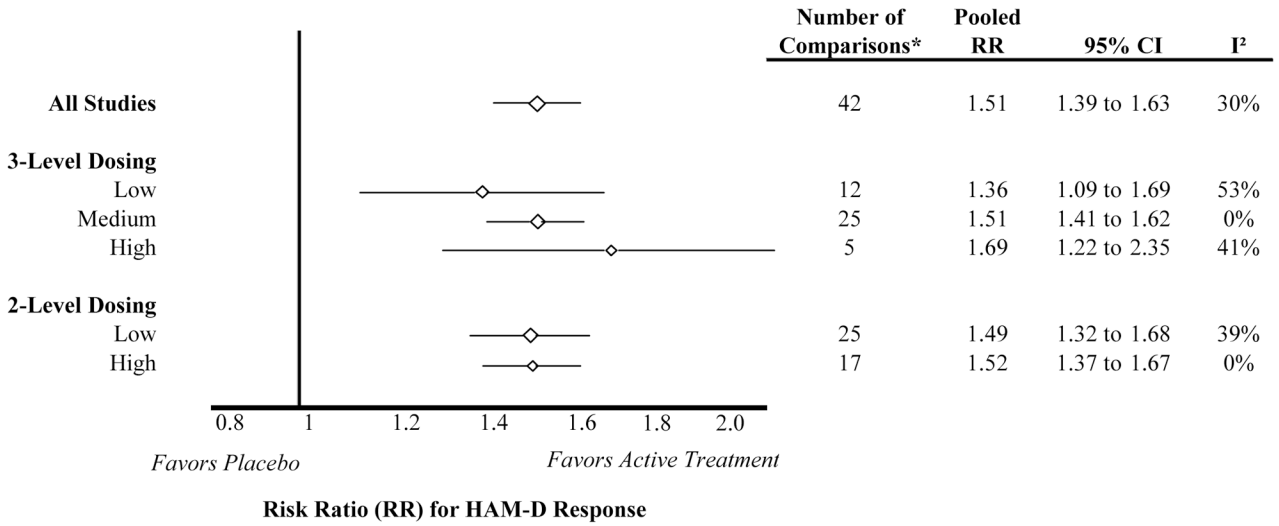

Figure 3.

Risk ratio for placebo-controlled trials stratified by dose

* The number of *comparisons* differs from the number of *articles* included in the systematic review because some studies contributed more than one placebo-controlled comparison, and some studies contributed data to the HAM-D change analysis but not the responder analysis

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

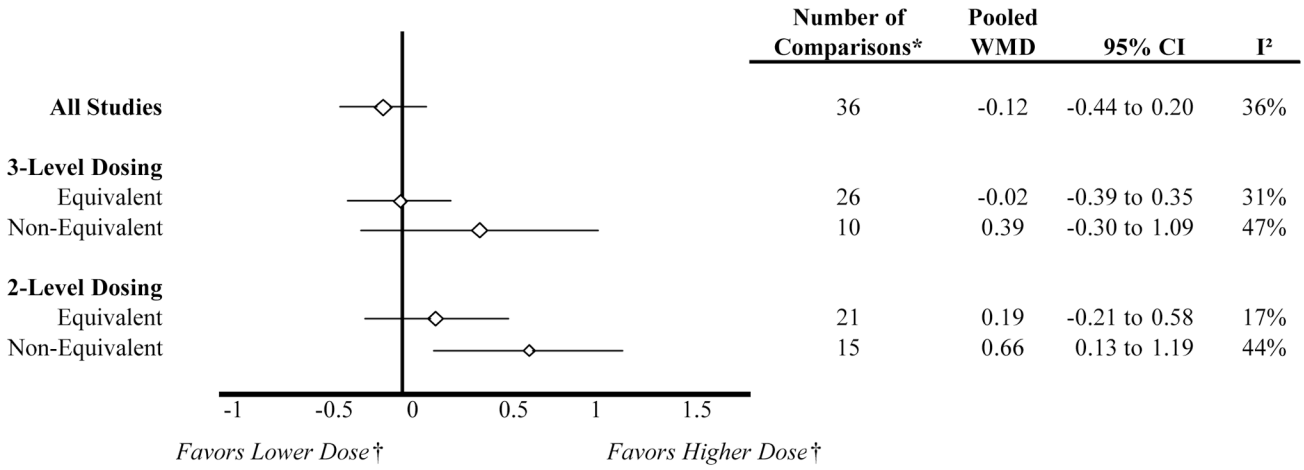

Figure 4.

Difference in weighted mean HAM-D change for comparative trials stratified by dose equivalence

* The number of *comparisons* differs from the number of *articles* included in the systematic review because some studies contributed multiple dosing arms for an active comparator, and some studies contributed data to the HAM-D change analysis but not the responder analysis

† Only applicable to non-equivalent dose comparisons

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

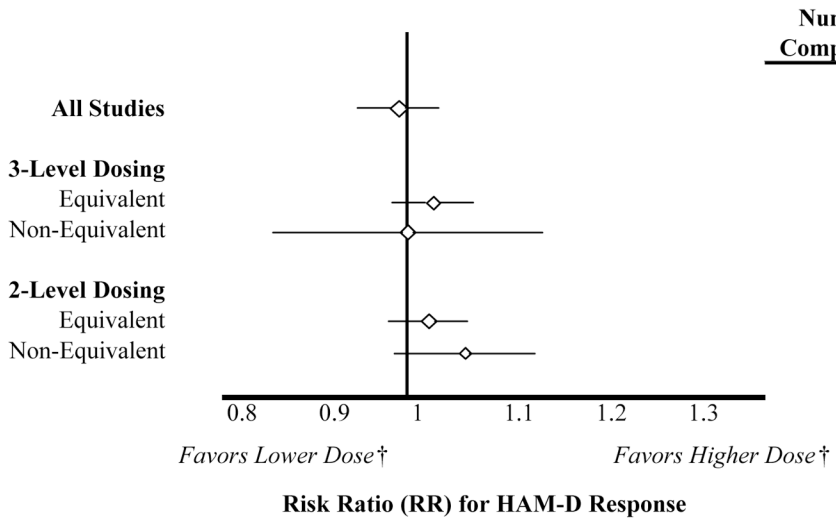

Figure 5.
Risk ratio for comparative trials stratified by dose equivalence
* The number of *comparisons* differs from the number of *articles* included in the systematic review because some studies contributed multiple dosing arms for an active comparator, and some studies contributed data to the responder analysis but not the HAM-D change analysis
† Only applicable to non-equivalent dose comparisons

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Dose Classification of Second-Generation Antidepressants

Generic Name	US Trade Name	Usual Daily Dosing Range*	2-Level Dose Classification		3- Level Dose Classification		
			Low	High	Low	Medium	High
Bupropion	Wellbutrin®	100–450 mg	≤275	>275	< 187.5	187.5–362.5	> 362.5
	Wellbutrin SR®	150–400 mg	≤275	>275	< 212.5	212.5–337.5	> 337.5
	Wellbutrin XL®	150–450 mg	≤300	>300	< 225	225–375	> 375
Citalopram	Celexa®	20–60 mg	≤40	>40	< 30	30–50	> 50
Duloxetine	Cymbalta®	40–120 mg	≤80	>80	< 60	60–100	> 100
Escitalopram	Lexapro®	10–20 mg	≤15	>15	< 12.5	13–17.5	> 17.5
Fluoxetine	Prozac®	10–80 mg	≤45	>45	< 27.5	28–62.5	> 62.5
Fluvoxamine	Luvox®	50–300 mg	≤175	>175	< 112.5	113–237.5	> 237.5
Mirtazapine	Remeron®	15–45 mg	≤30	>30	< 22.5	22.5–37.5	> 37.5
Nefazodone	Serzone®	200–600 mg	≤400	>400	< 300	300–500	> 500
Paroxetine	Paxil®	10–50 mg	≤30	>30	< 20	20–40	> 40
	Paxil CR®	25–62.5 mg	≤43.75	>43.75	< 34.4	34.4–53.1	> 53.1
Sertraline	Zoloft®	50–200 mg	≤125	>125	< 87.5	87.5–162.5	> 162.5
Venlafaxine	Effexor®	75–375 mg	≤225	>225	< 150	150–300	> 300
	Effexor XR®	75–225 mg	≤150	>150	< 112.5	112.5–187.5	> 187.5

*The usual dosing range represents doses recommended by FDA-approved product labeling. The upper bound for duloxetine was increased to 120mg/day based on practice patterns and other clinical evidence.[12,13] The lower bound for some drugs was decreased (e.g., immediate release bupropion, fluoxetine, and paroxetine) when it was deemed that the FDA-approved range did not capture the full range of doses used in clinical practice and shown to be efficacious in clinical studies.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Summary of placebo-controlled studies included in the dose analyses

Drug	Number of Unique Placebo Comparisons by Drug*	Total sample size (drug)	Total sample size (placebo)	Number of unique dosing arms				
				2-level		3-level		
				Low	High	Low	Medium	High
Bupropion	3	580	344	1	3	1	3	0
Citalopram	2	646	168	4	1	2	2	1
Duloxetine	2	439	323	3	1	0	3	1
Escitalopram	1	241	80	1	1	1	1	0
Fluoxetine	7	757	387	8	1	5	4	0
Fluvoxamine	4	133	130	2	2	1	2	1
Mirtazapine	2	96	71	2	0	2	0	0
Nefazodone	5	446	322	5	2	2	4	1
Paroxetine	10	1035	825	7	4	3	6	2
Sertraline	6	921	628	3	5	2	5	1
Venlafaxine	10	1456	668	16	5	8	9	4
Total	52	6750	3946	52	25	27	39	11

* The total number of unique placebo comparisons by drug includes the 46 separate placebo controlled studies noted in the Quorum Tree (Figure 1). Six of these studies compared more than one active drug to placebo, thus they appear twice. All arms of the study were included as placebo controlled comparisons.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 3

Summary of comparative studies included in the dose analyses

Drug	Number of Active Drug Comparisons*	Total sample size	Dose comparison and equivalency			
			2-level		3-level	
			Equivalent	Non-equivalent	Equivalent	Non-equivalent
SSRI vs. SSRI	13	2273	9	4	9	4
SSRI vs. SNRI	9	2085	6	3	8	1
SSRI vs. Other	14	2971	6	8	9	5

SSRI: citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, sertraline

SNRI: venlafaxine

Other: bupropion, duloxetine, mirtazapine, nefazodone (Note – duloxetine and mirtazapine are sometimes classified as SNRIs, although it is argued that mechanisms of action are distinct)

* The total number of active drug comparisons includes the 34 separate head-to-head studies noted in the Quorum Tree (Figure 1). Two of these studies included more than one head-to-head comparison, thus they appear twice.