

10-18-2011

# Depicting estimates using the intercept in meta-regression models: The moving constant technique


Blair T. Johnson Dr.

*University of Connecticut - Storrs*, [blair.t.johnson@uconn.edu](mailto:blair.t.johnson@uconn.edu)

Tania B. Huedo-Medina Dr.

*University of Connecticut - Storrs*, [tania.huedo-medina@uconn.edu](mailto:tania.huedo-medina@uconn.edu)

Follow this and additional works at: [https://opencommons.uconn.edu/chip\\_docs](https://opencommons.uconn.edu/chip_docs)

 Part of the [Applied Statistics Commons](#), [Biostatistics Commons](#), [Clinical Trials Commons](#), [Community Health and Preventive Medicine Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), [Epidemiology Commons](#), [Management Sciences and Quantitative Methods Commons](#), [Multivariate Analysis Commons](#), [Other Statistics and Probability Commons](#), [Psychological Phenomena and Processes Commons](#), [Statistical Methodology Commons](#), and the [Statistical Models Commons](#)

---

## Recommended Citation

Johnson, Blair T. Dr. and Huedo-Medina, Tania B. Dr., "Depicting estimates using the intercept in meta-regression models: The moving constant technique" (2011). *CHIP Documents*. 33.  
[https://opencommons.uconn.edu/chip\\_docs/33](https://opencommons.uconn.edu/chip_docs/33)

This pre-publication version of the article has appeared in print and should be cited as:

Johnson, B. T., & Huedo-Medina, T. B.<sup>§</sup> (2011). Depicting estimates using the intercept in meta-regression models: The moving constant technique. *Research Synthesis Methods*, 2(3), 204-220. (doi: 10.1002/jrsm.49)

## Depicting estimates using the intercept in meta-regression models: The moving constant technique

Blair T. Johnson<sup>a\*</sup> and Tania B. Huedo-Medina<sup>a</sup>

In any scientific discipline, the ability to portray research patterns graphically often aids greatly in interpreting a phenomenon. In part to depict phenomena, the statistics and capabilities of meta-analytic models have grown increasingly sophisticated. Accordingly, this article details how to move the constant in weighted meta-analysis regression models (viz. "meta-regression") to illuminate the patterns in such models across a range of complexities. Although it is commonly ignored in practice, the constant (or intercept) in such models can be indispensable when it is not relegated to its usual static role. The moving constant technique makes possible estimates and confidence intervals at moderator levels of interest as well as continuous confidence bands around the meta-regression line itself. Such estimates, in turn, can be highly informative to interpret the nature of the phenomenon being studied in the meta-analysis, especially when a comparison with an absolute or a practical criterion is the goal. Knowing the point at which effect size estimates reach statistical significance or other practical criteria of effect size magnitude can be quite important. Examples ranging from simple to complex models illustrate these principles. Limitations and extensions of the strategy are discussed. Copyright © 2011 John Wiley & Sons, Ltd.

**Keywords:** meta-regression; meta-analysis regression; weighted regression; point estimates; confidence intervals; confidence bands; prediction intervals; graphical displays

Across sciences, graphical displays of results can be an immense help in interpreting the patterns that result from statistical model tests (Light *et al.*, 1994; Tufte, 2001). Meta-analysts have an increasingly rich array of options when it comes to displaying weighted mean effect sizes and individual effect sizes (Anzures-Cabrera & Higgins, 2010; Borman and Grigg, 2009), showing how effect size indexes (*T*s) such as the standardized mean difference (SMD) or logged odds ratio performs across a literature. Variations on such displays have also been used to imply how *T*s may relate to an independent variable (Borman and Grigg, 2009; Lau *et al.*, 2006), which meta-analysts conventionally label *moderators* or *effect modifiers* and evaluate in weighted meta-analysis regression models known as *meta-regressions*, where the weights are the inverse of the variance for each *T*. Analysts often plot *T*s around the meta-regression line, which helps to capture the meaning of the meta-regression and to see how well it explains variability in *T*s (Borman and Grigg, 2009; Lau *et al.*, 2006). Relatively rare to date are meta-regression plots including confidence bands around the regression line to highlight at what levels of a moderator estimates exceed the null value or some benchmark criterion of clinical significance (Hayter *et al.*, 2007; Liu *et al.*, 2008). There also are few tools presently available to display patterns from meta-regressions with multiple moderator factors. With few exceptions, meta-analysts more often provide tabular and textual descriptions of their meta-regression results, providing standardized or unstandardized coefficients but not displaying how *T*s vary at different points along the moderator variable or variables.

Popular sources on meta-regression (e.g., Borenstein *et al.*, 2009; Cooper *et al.*, 2009; Hedges and Olkin, 1985; Higgins and Green, 2009; Lipsey and Wilson, 2001; Raudenbush, 2009) explain how this technique is applied and explicate the assumptions that underlie such analyses. They provide little detailed information on how one can extend these principles to produce confidence bands around the underlying regression line or confidence intervals at particular values of a moderator or moderators. By the same token, these reference works also do not address how to produce confidence intervals for estimates at particular values of interest along or

<sup>a</sup>Department of Psychology, 406 Babbidge Road Unit 1020, University of Connecticut, Storrs, CT 06269-1020, USA; Center for Health, Intervention, and Prevention 2006 Hillside Road Unit 1248, University of Connecticut, Storrs, CT 06269-1248, USA.

\*Correspondence to: Blair T. Johnson, Department of Psychology, 406 Babbidge Road Unit 1020, University of Connecticut, Storrs, CT 06269-1020, USA.

<sup>§</sup>E-mail: blair.t.johnson@uconn.edu.

outside the observed range of the moderator variable. Treatises that have examined graphical displays also have left these issues unexplored (Borman and Grigg, 2009; Harbord and Higgins, 2008; Light *et al.*, 1994; Viechtbauer, 2010a). Making matters worse, as we show, pre-packaged graphing software can badly mismatch conventional meta-analytic assumptions when used for plotting confidence bands.

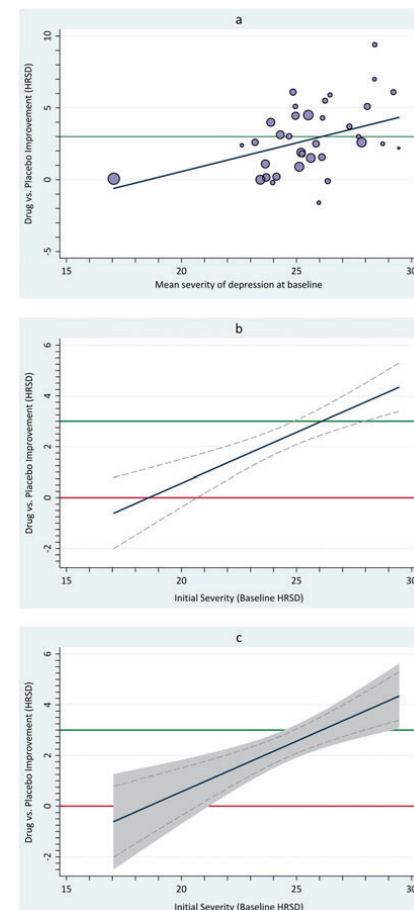
In this article, we illustrate how to move the constant in meta-regression models to plot confidence bands and to produce confidence intervals for  $T$ s at particular values of the moderator variable. Our purpose is not to develop or elucidate the underlying statistical principles underlying conventional meta-regression modeling (for these, see, e.g., Higgins and Thompson, 2002; Konstantopoulos and Hedges, 2009; Raudenbush, 2009), defined as regressions weighted by the inverse of the variance for each  $T$ , but to show how fuller use of these principles can enrich interpretation and presentation. Indeed, in many respects, the principles that guide meta-regression mirror those that guide the practice of regression with primary-level databases. We first consider estimates and confidence bands for the simplest case, bivariate meta-regressions—those with a single moderator variable. Then, we consider the moving constant technique in the case of multiple-moderator models, including nonlinear predicted values and an exploration of residuals. We conclude by discussing the limitations and potentials associated with such displays.

#### A bivariate example moving the constant

Meta-regression has become an extremely popular tool to see which moderators explain discrepancies in study findings, which is especially valuable in the face of heterogeneity, when the hypothesis of homogeneity has been rejected. That is, such  $T$ s exhibit greater variability than sampling error alone would suggest. Meta-regression's popularity lies in its flexibility: It can include more than one predictor, which can help determine what moderator variables best explain unique variation in  $T$ s (Borenstein *et al.*, 2009; Hedges and Olkin, 1985). Meta-regressions can also incorporate both categorical and continuous moderator variables; the latter can be linear or nonlinear (e.g., quadratic, logarithmic, etc.). The *fixed-effects version* assumes that only sampling error is present among study findings whereas the *mixed-effects version* assumes that slopes are fixed but that the intercept is random, so between-study variance adds another source of error in the model (Hedges and Vevea, 1998). This latter approach is conventionally labeled *random-effects meta-regression* (Borenstein *et al.*, 2009; Harbord and Higgins, 2008) despite the fact that the intercept is estimated under random-effects and the slopes are estimates under fixed-effects (Berkey *et al.*, 1995; Hedges, 1992; Huedo-Medina and Johnson, 2010; Knapp and Hartung, 2003; Viechtbauer, 2010a; Viechtbauer, 2010b). Each of these forms of meta-regression invokes weights that are the inverse of the variance for each  $T$  (for calculation specifics, see Borenstein *et al.*, 2009; Harbord and Higgins, 2008; Huedo-Medina and Johnson, 2010; Lipsey and Wilson, 2001).

**The Moving Constant Technique.** Like ordinary least squares regression models, meta-regression models include a single intercept and a slope for each moderator variable or covariate. Although it is commonly ignored in practice, the constant, or intercept, in a meta-regression equation can be extremely valuable because using it permits one to estimate confidence bands and intervals. Take as an example the meta-analysis of trials evaluating the success of antidepressants at alleviating depressive symptoms presented by Kirsch *et al.* (2008). They gathered randomized controlled trials that pharmaceutical companies had submitted to the U.S. Food and Drug Administration for drug approval. Each trial evaluated change in patients' depressive symptoms after a period taking antidepressants compared with those taking placebo; patients were randomly assigned to condition. Depression was assessed in each study with the Hamilton Rating Scale of Depression (HRSD). Kirsch and colleagues evaluated the hypothesis that antidepressants should succeed better for more severely depressed samples of participants. They examined this hypothesis in two main ways: one focused on how much improvement patients in the drug or placebo groups experienced over time, and the other, on which we will focus in this example, focused on the controlled comparison, the amount of improvement, if any, in drug relative to placebo at some point after treatment commenced.  $T$  was defined in terms of the arithmetic difference in the change of depression means between the two groups, as assessed using the HRSD. The improvement from baseline is obtained for each group, and positive values imply that the drug group improved more than the placebo group. Another purpose of their work was to evaluate for what levels of severity antidepressants achieve a clinically significant level of change. For this purpose, they adopted the 2004 recommendation of the UK's National Institute for Health and Clinical Excellence (NICE), which is three HRSD points better improvement for the drug group relative to the placebo group.

Results confirmed the hypothesis presented by Kirsch and colleagues: antidepressant efficacy was indeed larger for samples with more extreme depression. Figure 1 (panel a) shows these findings; antidepressants' efficacy did not reach a conventional value of clinical significance (green line) except in samples with very severe depression. Although such a display is useful for such purposes, a version with confidence bands would enable inferences not only about whether antidepressants have a statistically significant impact on depression (relative to placebo) but also, and more importantly, for samples with what levels of depression it has this effect. Unless otherwise noted, our demonstrations use *pointwise confidence bands*, which surround the meta-regression line, and *pointwise confidence intervals*, which are made at particular values of a moderator variable. (We address two other alternatives, *simultaneous confidence bands* and *prediction bands*, in the final section.)



**Figure 1.** Improvement on depressive symptoms as a function of the samples' baseline severity of depression : (a) Kirsch *et al.*'s (2008) Figure 4, reprinted with overall trends; (b) same trends but with confidence band added; (c) same trends with confidence bands that follow differing assumptions (see text for more description). HRSD, Hamilton Rating Scale of Depression

Imagine that instead we want to know whether antidepressants have a statistically significant benefit for groups with an average level of depression of 17, the lowest value observed in any of the trials, which is in the "moderate" range of the scale. A starting point is the meta-regression equation of the line from their analysis:

$$\hat{T} = -7.416 + 0.3991(\text{Baseline depression}),$$

where  $\hat{T}$  is the estimated value of the difference between drug and placebo improvement change. By default, meta-regression statistical output includes a test of whether the coefficient for the slope differs significantly from

**Table 1.** Meta-regression equations of the line in which the constant is moved to estimate the difference in Hamilton Rating Scale of Depression values for members of antidepressant groups relative to members of placebo groups for samples with varying levels of initial depression.

Model	With mean HRSD value	Constant	Width of confidence interval	Slope (95% confidence interval)
		$\hat{T}$ (95% confidence interval)		
1	<i>Original metric (constant = 0)</i>	<i>−7.416 (−11.71, −3.12)</i>	<i>8.59</i>	<i>0.3991 (0.23, 0.57)</i>
2	Original metric − 17	−0.630 (−2.04, 0.78)	2.82	0.3991 (0.23, 0.57)
3	Original metric − 20	0.567 (−0.37, 1.51)	1.88	0.3991 (0.23, 0.57)
4	Original metric − 23	1.765 (1.21, 2.32)	1.11	0.3991 (0.23, 0.57)
5	Original metric − 26	2.962 (2.43, 3.49)	1.06	0.3991 (0.23, 0.57)
6	Original metric − 29	4.159 (3.27, 5.05)	1.78	0.3991 (0.23, 0.57)
7	Original metric − 32	5.357 (4.00, 6.71)	2.71	0.3991 (0.23, 0.57)
8	Original metric − 46	10.940 (7.23, 14.66)	7.43	0.3991 (0.23, 0.57)

Notes: The slope appears in this Table to help show that each model is estimating the same information. Rows that are rendered in *italics* are extrapolations beyond the observed data. For simplicity, models follow fixed-effects assumptions; estimates incorporating random-effects assumptions exhibited highly similar results (see Figure 3).  $\hat{T}$  = estimate of the degree of improvement in antidepressant relative to placebo group members at the specified value of Hamilton Rating Scale of Depression (HRSD).

zero, which it was in this case, confirming the prediction given by Kirsch and colleagues (Table 1, Model 1). The slope shows that drugs had increasing success relative to placebo as initial depression increased. Specifically, for every scale value worse on depression that a sample scored at baseline, the drug groups improved by 0.3991 HRSD points relative to placebo. Most meta-regression output includes estimates of the slope in two forms: (a) the unstandardized form, which we have described; and (b) the standardized form, usually characterized with the Greek symbol,  $\beta$ , which may be interpreted similarly to a correlation coefficient. In this case, the association between severity and the drug-placebo differences was  $\beta = 0.52$ .

Although in practice many analysts might be satisfied to know the magnitude of the slope and its statistical significance, potentially even more useful is the intercept, which is the value that  $\hat{T}$  takes when the moderator variable (or variables) is (are) exactly zero. By default, meta-regression statistical output includes a test of whether the intercept differs significantly from zero, which it was in this case ( $\hat{T} = -7.416$ ; 95% confidence interval =  $-11.71$ ;  $-3.12$ ; Table 1, Model 1). In other words, we are 95% confident that the difference in improvement will be in the range from  $-11.71$  to  $-3.12$  for a sample with an initial mean HRSD value of 0.<sup>1</sup> Thus, if a sample's mean baseline depression is exactly zero, this equation predicts that groups receiving placebo would average 7.416 HRSD points less depression than groups receiving antidepressants. This example helps to illustrate one reason why meta-regression intercepts are so commonly ignored in practice—because they so often yield unrealistic values. In this particular case, the intercept reflects a value that baseline HRSD means cannot plausibly take. Although it is possible for an individual to score 0 on the HRSD, a sample could not have a mean score of exactly 0 unless it also had zero variance as well, which would be a rare occurrence indeed. Moreover, when depression levels are so low, there is no need for antidepressants, let alone in predicting what their effects would be.<sup>2</sup> Much more valuable is to see what impact drugs might have on populations with a real need for them. Finally, another reason to mistrust this estimate is that there were no observations so low in the database, and therefore, the estimate is a fairly extreme extrapolation.

Imagine that we want to know what impact antidepressants should have on patients whose level of depression is HRSD = 17, which is the lowest value observed in this particular sample of studies and is conventionally interpreted as a moderate depression. Inserting this value into the equation of the line yields

$$\hat{T} = -7.416 + 0.3991(17) = -0.6313.$$

<sup>1</sup>Of course, if desired, other confidence intervals may be estimated other than the usual default of 95%; commonly utilized instead are 90% intervals, which will be more liberal—narrower, and 99% intervals, which will be more conservative—wider. We return to this issue in the concluding section, *General Limitations and Potential Uses of These Strategies*, where we advocate routinely using more conservative estimates than 95%, particularly under fixed-effects assumptions.

<sup>2</sup>Of course, as Thombs *et al.* (2011) discussed, it is possible that in reality, an individual who scores 0 on a standardized scale may in fact have elevated levels of depression that are not detected by the scale.

In other words, the equation predicts that members of the drug group will improve less (by 0.6313 HRSD values) than members of the placebo group. Although it is simple to determine a predicted value using the equation of the line, to create a confidence interval around it, one must use the moving constant technique.<sup>3</sup> Specifically, subtracting 17 from every observation of baseline severity will effectively move the intercept to a value of 17. Now, when we re-run the model, the results show not only the same estimated value of  $\hat{T}$  the equation implied,  $-0.630$  but also a confidence interval for it ( $-2.04, 0.78$ ), as Model 2 in Table 1 shows. Thus, for samples with moderate levels of depression, no advantage is likely to be seen for those taking antidepressants (relative to those taking placebo). Note that the confidence interval around this value is also much narrower (Figure 1, panels b & c), which makes sense because there are some data available at this point along the moderator variable on which to base an estimate (see Figure 1, panel a).<sup>4</sup>

Table 1 uses the same procedure to estimate other confidence intervals across the observed range of the moderator variable. A significant effect of drug is evident when the mean HRSD value reaches 23 (Model 4), and the drug effect continues to increase and be statistically significant at higher HRSD values. Another important benchmark is clinical significance: an effect might be statistically significant without having sufficient practical impact in people's lives. In the present case, Kirsch *et al.* (2008) noted that the three-point NICE clinical significance benchmark was achieved somewhere near the 28-point mark of the moderator variable. Indeed, at the highest observed initial depression value, 29 (Model 6), the confidence interval no longer includes the value 3. Thus, on the average for studies whose samples had mean levels of initial depression this high, the observed values exceeded the clinical significance criterion.

For the purpose of illustration, we estimated two more models that move beyond the observed range of the moderator variable. Just as the confidence interval for a sample with zero depression had a wide range, so too does the confidence interval for a sample of extremely depressed patients (46 on the HRSD), in Model 8. These extreme examples might strike some readers as controversial: conventionally, statisticians restrict estimates to observed ranges of moderator variables, but there are times when it is important to project findings beyond what was observed, such as when earth scientists project climate change over centuries. As an example, earth scientists commonly project estimates far into the future, such as how much the earth will warm by the year 2100 and even centuries farther into the future (Solomon *et al.*, 2009), or how high the seas will rise given this amount of warming. Obviously, such predictions can have profound ramifications. Having some confidence in how conditions may change can help community planners protect their territories and maximize outcomes.<sup>5</sup>

The aforementioned models imply how one can estimate *confidence bands* around the predicted values defined by  $\hat{T}$ s. Such is the logic of regression, whether in terms of regular regression with primary-level data (e.g., Myers and Well, 2003) or with meta-analytic data (Viechtbauer, 2010b). One can move the intercept across the values of a moderator and plot the estimate and its confidence interval. The problem with moving it in such a coarse fashion is that if the points estimated are very distant from each other, a graph based on this procedure may have poor smoothing between the estimates and might badly estimate certain regions along the moderator variable. Greater accuracy would result if the estimates moved in much finer increments along the moderator variable. In the current example, 1001 iterations that each move the intercept by 0.0123809 of a unit of depression creates an extremely precise estimate of the confidence bands. If we start with mean depression at its minimum observed value, 17.0631, then after 1001 iterations, the intercept would estimate the maximum observed level of mean depression in these samples, 29.444, and the resulting confidence bands will be smoother. Of course, such fine gradations might not be necessary unless a very high resolution graph is needed, such as when it will be printed in a very large size. Our Appendixes list syntax to show how one can create such graphs using two popular statistical platforms, and Figure 1 (panel b) shows such a graph. A significant difference exists when the confidence bands no longer include the red line indicating exactly no difference. A clinically significant difference exists when the confidence bands exceed the green reference line.

As we have illustrated, the moving constant strategy is also helpful for producing point estimates at points of interest, with no graphs at all. To examine estimates along a continuous moderator variable that is examined linearly, one would typically plot estimates for the lowest and highest value observed for the moderator variable. Plotting estimates at the mean or median values of the moderator might also be of interest, and other possibilities

<sup>3</sup>Those who teach meta-analysis might imbue the term *the moving constant technique* with greater intrigue for their students by instead calling it *the secret of the moving constant*. Indeed, it may effectively be a secret to those who have not yet realized its potential.

<sup>4</sup>It is worth noting, as did Kirsch and his colleagues, that statistical inferences associated with this level of depression should be taken with caution, as there was only one case in the sample with an mean depression lower than 23 on the HRSD. It is for this reason that the confidence interval is relatively wide for estimates of  $\hat{T}$  with initial mean HRSD = 17 (in Table 1, compare Model 2 with Models 4–6).

<sup>5</sup>We searched unsuccessfully for a meta-analysis within the social sciences that had explicitly projected results beyond the limits of a moderator. We believe the practice could be quite useful in some circumstances. For example, one might imagine a scenario in which a meta-analysis examines whether a given intervention improves health for a stigmatized group. Prejudice levels by those living in the communities where these trials are done might vary from medium to high. Let us imagine that the meta-analysis shows that the interventions improve health most when prejudice is lower. A meta-analysis might in turn estimate what effect such interventions would have if prejudice was low or zero. Such a pattern, if confirmed, might suggest that structural interventions to reduce stigma would multiply the impact of the intervention. Hence, using the moving constant technique might be put to fruitful use to project results into unexplored terrain, which then might stimulate more direct research on the subject.

may exist as well, such as projecting results beyond the observed range of the moderator variable. Thus, an analyst could plot  $\hat{\tau}$  at specified points along the continuum implied by the moderator variable and report the estimated weighted  $\hat{\tau}$  and the confidence interval around this estimate.

**Potential problems with bands produced by graphing functions in conventional software.** Because they are easily invoked, one temptation might be to graph such patterns with conventional, widely available software that was written for use with primary-level data and includes ordinary weighted least squares analysis functions. For example, one can use SPSS (IBM SPSS Statistics, 2010) to plot confidence bands around weighted regression lines, but because the graphing solution models the standard error of the regression coefficients (including the intercept) differently than conventional meta-analytic statistics in relation to their unstandardized coefficients,  $\hat{\beta}_j$ , the confidence bands and intervals will almost always differ from those implied by the meta-analytic regression model (see Lipsey and Wilson, 2001, pp. 138–140). Hedges and Olkin (1985) noted that “standard errors for  $\hat{\beta}_j$  printed by the program are *incorrect* by a factor of  $\sqrt{MS_E}$ , where  $MS_E$  is the error or residual mean square for the regression” (p. 174, *italics in original*) because standard software does not incorporate the known variance estimate for the meta-analytic data (Konstantopoulos and Hedges, 2004).<sup>6</sup> Although Hedges and Olkin made their statement in reference to fixed-effects meta-regression models, models that incorporate random-effects assumptions are just as susceptible to the problem.<sup>7</sup> Future research should address this issue.

To keep graphical depictions of meta-analytic results squarely in accord with the underlying inferential statistics, therefore, analysts are well advised to avoid using standard software for graphing meta-regression results, or at least to double-check that the confidence bands match the meta-analytic model. The logic we outlined in this section can be used for such a check: Estimate the confidence intervals for a value along the moderator dimension in question and compare them with the figure. In Figure 1 (panel c), the dashed lines indicate the confidence bands implied by a meta-regression following mixed-effects assumptions; gray bands resulted by invoking the weighted graphing function in primary-level statistical software (i.e., using the STATA command `twoway lfitci`). These bands (shaded area) are clearly wider; thus, statistical and clinical inference will deviate at certain points along the moderator.

Importantly, ordinary weighted least squares analyses do not always overestimate the widths of confidence bands: in fact, the difference can go in either direction. As Hedges and Olkin (1985), earlier, implied, the bias is more conservative (wider confidence intervals and bands) to the extent that  $\sqrt{MS_E}$  exceeds 1, more liberal (narrower intervals and bands) to the extent that  $\sqrt{MS_E}$  is less than 1, and equivalent when  $\sqrt{MS_E} = 1$ . Consequently, using ordinary weighted least squares analysis functions to estimate confidence intervals and bands will be justified only in the very rare instance when  $\sqrt{MS_E} = 1$ .

#### The moving constant technique in more complex meta-regression models

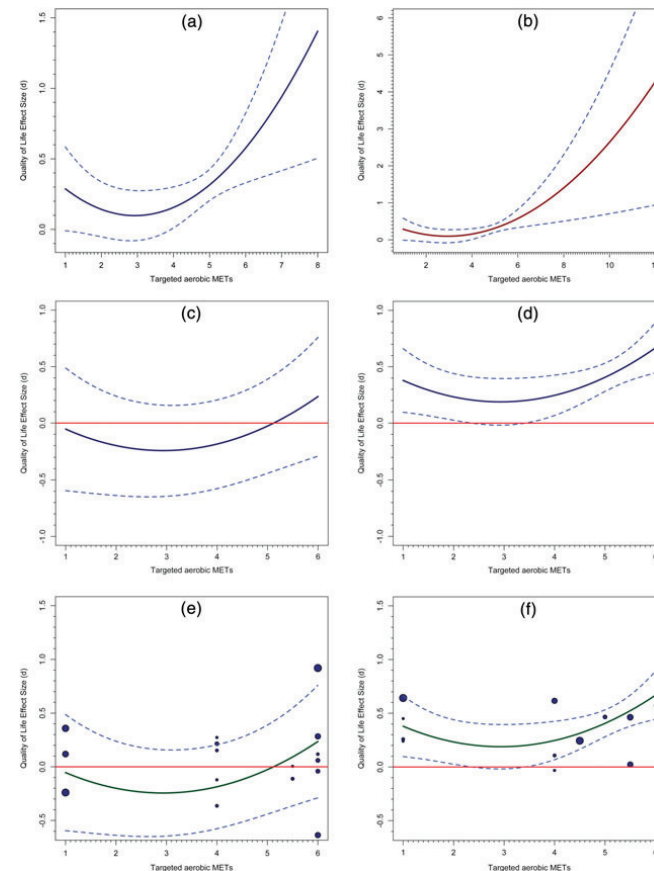
**Multiple-moderator models.** To this point, we have used a bivariate meta-regression example to make use of the intercept completely transparent. Now, we turn to a more complicated example that better maps on to the target problem. Assume we have a meta-analysis of studies using exercise to impact quality of life in cancer survivors, such as that Ferrer and colleagues (Ferrer *et al.*, 2010) recently conducted. Imagine that we want to illustrate a model with three moderator variables that we find plausible based on a reading of the literature: (a) the quantity of aerobic metabolic equivalents of task (METs); (b) its quadratic term (i.e., aerobic METs<sup>2</sup>); and (c) the percentage of the sample that is female. Aerobic METs define how active one is: At rest, one exerts one MET; at six METs, one does vigorous exercise. The studies evaluate the extent to which bouts of exercise that accrue over time relate to quality of life assessed on standard scales.

Just as in ordinary regression, a model with quadratic effects must incorporate the linear effect because otherwise, a statistically significant quadratic effect might in reality be a linear effect. Initially, let us examine trends across the range of aerobic METs, holding the impact of sample gender constant at its sample mean of 79% female; thus, gender is mean-centered at 79% in this analysis. Although all-male and all-female samples appeared in the database, more female than male samples were present. Figure 2 (panel a) shows the first figure that results from following these steps (Appendix I). We see that the impact of exercise is statistically significant across the range from medium (four MET) to very high (eight METs) targeted aerobic activity. No study had a value over

<sup>6</sup>Similarly, standard meta-regression statistical output under most if not all standard statistical packages, rarely list the  $MS_E$  associated with a model. Therefore, it is difficult for an analyst to know how differently confidence bands would be plotted if the graphing software were used. As we mentioned in the text, the moving constant technique can be used to determine whether the bands match meta-analytic assumptions.

<sup>7</sup>Of note, under heterogeneity, meta-regression models that incorporate random-effects assumptions are certain to have smaller  $MS_E$  values than would completely fixed-effects versions of these models, and indeed often exhibit  $MS_E$  values of about 1 in practice. Meanwhile, completely fixed-effects versions usually have  $MS_E$  values much larger than 1. Thus, if appropriate weights are used (see Appendix I, Step 4), conventional graphing software will often produce reasonably accurate graphs in the case of meta-regressions incorporating random-effects assumptions.

6.25 METs, so after that point, as discussed earlier, any regression line and confidence bands beyond this point amount to a prediction of what would happen at these higher activity points. Figure 2 (panel b) takes it to greater extremes by projecting estimates all the way to aerobic METs = 12, far higher than any study evaluated. Initially, note that although the confidence bands in this graph in the portion of the graph up to METs = 8 appear to be narrower, they are in fact the same; it is just that the maximum value of the y-axis is now much larger, to  $SMD = 6$ . Nonetheless, one can see that the confidence intervals widen dramatically, just as they should, because there are no observations in that region. The figure is also quite implausible from the standpoint that the largest observed effect size is less than  $SMD = 2$ , and the predicted values plotted go far beyond this value. Such a figure predicts what *might* happen if cancer survivors were given extremely strenuous aerobic activities and if they could



**Figure 2.** Complex meta-analytic models plotted under differing circumstances implied by the models. Specifically plotted is the relation between intensity of aerobic exercise on quality of life improvements for cancer survivors (see text for specific descriptions of each panel). d, standardized mean difference



undertake such a program. An analyst might use a figure such as this one if he or she was trying to encourage scholars to undertake such trials and to see whether still greater improvements in quality of life can occur.

It is important to understand that because gender was zero-centered in these models, strictly speaking, graphs such as those shown in Figure 2 (panels a and b) plot the estimated improvement of various amounts of aerobic exercise, adjusted so that they reflect samples with 79% females (the mean). Our model showed that samples with larger percentages of females had greater success improving quality of life through exercise. Thus, the plotted line is logically farther from zero for samples that are more female and closer to the line for samples that are more male. The same sequence that produced Figure 2 (panel a) was followed for these two extremes. In two new instantiations, we hold gender constant at 0% or 100%, respectively. Figure 2 (panels c and d) shows the estimates for samples of men and women, respectively. Although both lines show the curvilinearity implied by the quadratic function, the line is indeed farther from zero for female samples and closer to zero for male samples. Indeed, aerobic exercise would appear to have little impact on quality of life for samples of men unless more intense aerobic exercise is undertaken and then just barely. For female samples, in contrast, any amount of exercise appears to improve quality of life on the average but especially at higher intensities. Note too that the confidence bands are narrower for female than for male samples, consistent with the fact that there were more observations for the former group. As with any meta-analytic model, the analyst should take care to note such limitations when interpreting such findings. The authors of this particular meta-analysis used exactly these procedures to evaluate more sophisticated models of the relation, showing that the quadratic function did not appear unless the patients exercised steadily over a few months; interested readers should consult their article (Ferrer *et al.*, 2010), which lists other limitations of this particular meta-analysis.

**Complications involving categorical variables.** Up until this point, our examples of used continuous variables as moderators. Note that the moving constant technique can help an analyst avoid the problem of bifurcating continuous variables; that is, the models themselves can produce estimates for particular levels of the continuous variables, obviating the need to artificially dichotomize them. Yet, categorical variables are commonly incorporated in meta-analytic research and they present some special complications. Imagine we are still pursuing the same meta-analytic model as in the preceding paragraphs, but that now gender is represented as two categories, male and female, instead of as a continuous variable, percentage of females in the sample. Orthogonal coding gender so that male samples take  $-1$  and female samples take  $+1$  effectively controls the gender effect when estimating other moderators' effects. Now, interpretation proceeds as we have already described.

Imagine further what the results would imply if one instead dummy-coded the sample gender variable so that male samples take 0 s and female samples take 1 s, leaving aerobic METs (and its quadratic function) in the original metric. Now, the intercept in the model would reflect the point at which the moderator variables are all 0; that is, male samples with aerobic METs = 0. Once again, we find ourselves backed into a statistical corner where the intercept has little or no meaning because the MET scale begins not at value 0 but at 1, which implies being at rest. The solution of course is to commence with the intercept of our model reflecting the lowest value of the scale (1) rather than the meaningless value of 0. Finally, complexities in such models increase as the numbers of moderator variables increase, but the underlying logic to display the results is the same as we have presented here. For example, Johnson and colleagues (Johnson *et al.*, 2003) fit a meta-regression model with five moderators in their meta-analysis of studies evaluating human immunodeficiency virus prevention interventions for adolescents. Three variables were continuous and two were categorical (dummy or contrast coded). It took them 14 meta-regression runs to produce estimates for the extremes of each moderator while statistically holding the other moderators' influences constant.

*Is zero-centering or contrast-coding moderators always the best solution?*

Some readers may take the preceding examples to imply that one must zero-center or contrast code all moderators except for the one that is the focus of the moving constant technique to portray estimates and their confidence intervals or bands across that moderator. Yet, especially when interactions between moderators exist, it may be preferable to produce estimates that follow differing assumptions. Imagine that, as is often the case, the studies in a particular meta-analysis vary in methodological quality. Suppose we are examining the effects of exercise on depression levels in cancer survivors and expect to see a dose-response curve such that depression improves with greater amounts of exercise. An analyst could regress  $T_s$  on exercise dose, study quality, and the interaction of these terms. (Note that the analysis to examine the statistical significance of these terms will likely be more stable if the dose and quality terms are zero-centered before multiplying them to create the interaction term.) Suppose further that the interaction is statistically significant. In such a circumstance, an analyst might well show the dose-response pattern with study quality held constant at a high level because, logically, these are the studies whose results are the most trusted. Alternatively, one might show the dose-response function for both high-quality and low-quality studies. Brown and colleagues (Brown *et al.*, 2011, Table 4) followed just these procedures in their meta-analysis examining the effects of resistance exercise on cancer-related fatigue. A dose-response pattern was most marked for the highest quality studies.

#### *Caveats about clinical significance*

Because no standard for clinical significance yet exists for quality of life outcomes, we have focused instead on statistical significance in this section. That there are many different scales to assess quality of life makes achieving clinical significance standard more difficult but not impossible. Indeed, if one wished to parallel the NICE criterion for depression change, then one could generalize from the fact that their clinical significance criterion specifies the target of a "medium" effect size of  $SMD = 0.50$ , as per the standards given by Cohen (1988). Consequently, a medium effect size implies a change that is visible to the naked eye, with no need for statistics (Johnson and Kirsch, 2008). The graphs in Figure 2 might be interpreted as failing to meet this clinical significance standard, as none of the confidence bands exceed this value.

#### General limitations and potential uses of these strategies

It is worth discussing some other potential uses of the strategies that we have described as well as some general limitations, including assumptions that underlie the statistical model, detecting outliers, using prediction intervals, producing Bayesian estimates, strategies for alternative confidence bands, and practical limitations.

#### Assumptions underlying the model

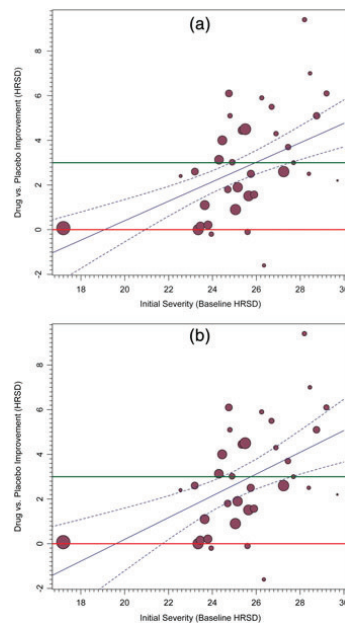
Sources on meta-analysis routinely state how, under heterogeneity, fixed-effects assumptions typically produce overly narrow confidence intervals relative to random-effects and mixed-effects assumptions. To illustrate, Figure 3 portrays the same antidepressant data that appeared in Figure 1. Figure 3 (panel a) follows fixed-effects assumptions, and Figure 3 (panel b) follows random-effects assumptions. Because heterogeneity is present, in a formal sense, the fixed-effects version is incorrectly specified; consequently, it has overly liberal, narrower confidence bands (Figure 3, panel a) and the random-effects version with its wider confidence bands might instead be favored (Figure 3, panel b). In this case, statistical inference does not dramatically differ between the two sets of estimates. It should also be noted that the variables that enter into the statistics themselves may be subject to any number of problems, including non-normal sampling distributions, invalidity, unreliability, and restriction of range, to name only a few (Hunter and Schmidt, 2004; Cohen *et al.*, 2004).

#### Outlier detection

Graphs such as we have produced can be helpful in diagnosing problems in the literature; we implied just this function at the outset, in discussing bivariate meta-regressions. The same holds for the more complex examples we presented in the preceding section. It is most valuable to use the more complex patterns implied by the model; thus, we can understand the influence of some particular relevant moderators while controlling for the effect of others. Figure 2 (panels e and f) uses the regression lines and confidence bands from panels c and d because these represent the complete model that was evaluated; superimposing observed effect sizes on the overall model, zero-centering gender, would be prone to error because it does not account for the gender effect. If we still did so, then relatively large effect sizes would be more likely to be from studies that focused on women and relatively small (or negative) effect sizes would be more likely to have focused on men. That sample gender is represented in continuous form whereas our graphs focus on the extremes of the distribution created a problem, because we had to pick which effect sizes to plot in each graph. It seems reasonable to plot cases with 50% or more men in the former plot and those with 50% or more women in the latter plot. One can see that the regression lines stay closer to studies with larger weights, which illustrates meta-analytic weighting in action. These plots also visually show residuals, which is the difference between the estimated value and the observed value for each case. Those farther from their lines have larger residuals. There are indeed far fewer cases for male cancer survivors, so statistical inferences are quite strained here. Panels e and f both show that there are few studies in the range under  $METs = 4$ , again implying that predictions here are a projection. The quadratic function appears most justifiable for the female sample studies, although there are some outliers that would be worth inspecting. Alternatively, other moderators might improve prediction and eliminate these large residuals.

#### Prediction intervals

This article focused most on estimating  $T_s$  at certain points along a moderator dimension or dimensions, which boils down to estimating where the weighted mean  $T$  lies along a moderator or set of moderators. Another alternative is to portray the *prediction interval* in which future studies are likely to be observed in relation to a moderator dimension or dimensions. As Borenstein and colleagues discuss and illustrate with a worked example, this strategy addresses the dispersion of effect sizes that one is likely to see in a particular literature. Appendix I



**Figure 3.** Improvement on depressive symptoms as a function of baseline severity of depression, with each point sized proportionally to its weight in the analysis. (panel a) Moderation pattern showing regression line and 95% confidence band under fixed-effects assumptions. (panel b) Same as panel a but under mixed-effects assumptions (i.e., with a random-effects constant and a fixed-effects slope). HRSD, Hamilton Rating Scale of Depression

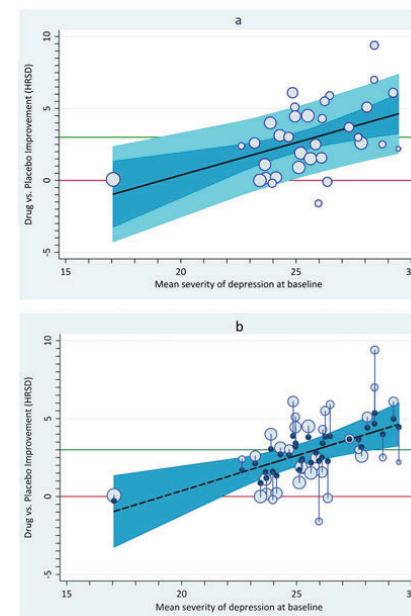
lists steps that can be used in one statistical platform to produce such estimates in tandem with meta-regression models. Figure 4 (panel a) shows a graph with such estimates for the Kirsch antidepressants data, showing that the prediction interval is much wider than the confidence bands. An analyst might display such a graph to develop expectations for how large  $T$  will be in a new study that matches a particular level of a moderator variable or variables, making it an adjunct to power analysis strategies.

### Bayesian estimates

The solutions that we described in the preceding two sections are decidedly frequentist in orientation, a practice often labeled *classical meta-analysis*. Nonetheless, those who wish to construct similar graphs following Bayesian assumptions can do so. In Figure 4 (panel b), we portray the empirical Bayesian estimates of the true effects for each study (once again in the Kirsch antidepressants database), assuming the fitted model is correct (Appendix I). In this graph, the observed SMDs appear as bubbles and the adjusted, true SMDs appear as filled bubbles. Arrows show how the observed SMDs converge on the regression line of the meta-regression model in the empirical Bayesian estimates of the true SMDs. Those cases that before appeared to be outliers as observed appear much more consistent under the assumptions of this model.

### Simultaneous confidence bands

The strategy that we have emphasized in this article is the pointwise method of constructing confidence bands and intervals. The *simultaneous confidence band strategy* attempts to control for the expanded error rate that accrues from the evaluating a statistical hypothesis across the range of a moderator variable or variables (Liu *et al.*, 2008; Macaskassy & Provost, 2004; Seber and Lee, 2003; Working and Hotelling, 1929). The pointwise strategy



**Figure 4.** Improvement on depressive symptoms as a function of baseline severity of depression, with each point sized proportionally to its weight in the analysis, and following mixed-effects assumptions. (panel a) Moderation pattern showing 95% prediction intervals depicted in light blue. (panel b) Observed (open circles) and empirical Bayesian estimates (filled circles) of true effect sizes. HRSD, Hamilton Rating Scale of Depression

we have illustrated in this article would presumably suffer from heightened Type I error rates, stemming from examining whether  $T$  is significant at different levels of a moderator variable or variables. Note that in theory, a moderator dimension can be divided into infinite increments and heightening Type I ( $\alpha$ ) error rates. The Bonferroni correction is often applied to reduce Type I error rates by dividing  $\alpha$  by the number of tests being evaluated but cannot be invoked when there is an infinity of tests. This impasse, coupled with the lack of a priori conceptualization for critical values of the moderator variables in models, suggests that analysts should routinely set  $\alpha$  to be more stringent to compensate or else interpret results more conservatively, especially under fixed-effects assumptions and in the presence of heterogeneity. Alternatively, the use of meta-regressions that incorporate random-effects assumptions is to be encouraged. As we have stated, these models are already more conservative than fully fixed-effects versions, under heterogeneity. Importantly, to date, there appears to be no easily available strategy for applying the simultaneous confidence band solution in meta-regression. Future research should address these problems directly.

### Practical limitations

It is worth discussing when graphs such as we have produced here might be most valuable. Given that most meta-analyses investigate numerous moderator variables, plotting all patterns might often prove impractical. If the goal is comprehensiveness, then tables of output are much more compact than numerous figures. Instead, the patterns that deserve the most practical attention are the ones that analysts should graph—the ones that they want to help tell the story of their meta-analytic results. Finally, note that, in practice, successful graphing is often an intensive, iterative practice in which the analyst produces successive versions of the same pattern until the right combination of informativeness and aesthetic value is achieved (Tufte, 2001).

One potential limitation of the steps we have listed here to depict meta-regression results is financial. True, some of our graphs were created using the commercial software STATA (StataCorp, 2009, College Station, TX) (Appendix I). The macros freely given by Lipsey and Wilson (2001) also are available for SPSS and SAS in addition

to Stata, but these statistical platforms are all commercial.<sup>8</sup> Those who are students rather than professionals often can get discounted prices on such software. Fortunately, all or nearly all of the solutions we suggest are also available in the open-access statistical software R (see Burns, 2006), for which Viechtbauer (Viechtbauer, 2010a; Viechtbauer, 2010b) has written a sophisticated suite of meta-analysis protocols, metafor. Appendix II shows how we used this software to create graphs such as in Figure 2 and Figure 3.

## Conclusion

Several prominent sources have given careful attention to graphing primary-level data and offer sage advice that meta-analysts might well also incorporate. Specifically, Tufte (Tufte, 2001; Tufte, 2006) argued for more careful integration of graphical and textual information, much in the way in which Leonardo da Vinci's journals richly illustrate. Research has shown that the modern practice of separating text and graphics yields poorer communication of the targeted scientific information (Sweller *et al.*, 1990). Two other sources provide very helpful advice about how best to represent more nuanced aspects of scientific databases (Cleveland, 1984; Lane and Sándor, 2009). To date, such sources emphasize primary-level investigations almost entirely to the exclusion of meta-analysis. The current article has emphasized how many of these principles of effective graphical displays generalize particularly well to effect size information.

In sum, the strategies we have described here should enable meta-analysts to create accurate graphs of meta-regression results and more effectively communicate scientific information. Because graphical depictions of statistical results are often a highly desirable way to "tell the story" related to a given phenomenon, strategies such as we have detailed should be an indispensable tool in meta-analysts' toolkit.

## Appendix I: Calculating and displaying confidence intervals and bands using meta-regression in STATA

(1) First make sure that STATA has installed the most recent version of the metareg command by using the "lookup" command:

```
. lookup metareg
```

Keyword search

Keywords: metareg

Search: (1) Official help files, FAQs, Examples, SJs, and STBs

Search of official help files, FAQs, Examples, SJs, and STBs

SJ-8-4 sbe23\_1 ..... Meta-regression in Stata  
(help metareg if installed) .... R. M. Harbord and J. P. T. Higgins  
Q4/08 SJ 8(4):493–519  
presents a revised version of the metareg command, which  
performs meta-analysis regression on study-level summary  
data

Selecting the sbe23\_1 entry (above) lets a user install or update to incorporate the metareg command presented by Harbord, which was first published in 2008 (Harbord and Higgins, 2008), updated from an earlier version that STATA published in 1998. If the metareg macro given by Lipsey and Wilson (2001) has previously been installed, then it will need to be renamed in order for the alternative to operate. For the graphs we produced for this paper, we used STATA version 11.2. After installing or updating the command, it can be used to estimate meta-analysis regression (viz. meta-regression) models. The basic form of the command follows this format:

(2) After installing or updating the command, it can be used to estimate meta-analysis regression (viz. meta-regression) models. The basic form of the command follows this format:

```
. metareg T moderator1 moderator2, wsse(se_T)
```

where  $T$  is the effect size,  $\text{moderator1}$  and  $\text{moderator2}$  are moderator dimensions, and  $\text{se}_T$  is the standard error for each  $T$ . By default, this command will invoke a random-effects constant (see discussion in the text about "mixed-" versus "random-effects" assumptions). If inverse-variance weights,  $TW$ , have been calculated for each

$T$  (see, e.g., Lipsey and Wilson, 2001, Table 3.2) but not  $\text{se}_T$ , then to use the metareg command, it is first necessary to calculate  $\text{se}_T$ :

```
. gen se_T = 1/sqrt(TW)
```

For example, in the antidepressants data described in the main text, the command

```
. metareg diffimp meanbase, wsse(se_diffimp) graph
```

where  $\text{diffimp}$  is the  $T$  (difference in improvement in depression between drug and placebo groups),  $\text{meanbase}$  is the mean severity of depression at baseline, and  $\text{se\_diffimp}$  is the standard error for each  $T$ , produces statistical output for the meta-regression model and a so-called bubble graph, where the bubbles are individual  $T$ s sized according to their weight in the analysis, and inserts a line indicating the meta-regression line (similar to Figure 3, panel b). The command can be used to generate confidence intervals at certain values of the moderator variable, such as the lowest and highest observed values. For example,

```
. gen tempvar = meanbase-17.0631
```

where 17.0631 was the lowest observed mean severity of depression at baseline, and  $\text{tempvar}$  is intended to replace  $\text{meanbase}$  in the re-instantiated command:

```
. metareg diffimp tempvar, wsse(se_diffimp) graph
```

which moves the constant or intercept to the value of 17 (the graph now literally shows the intercept at the zero point of the newly calculated variable,  $\text{tempvar}$ , on the left side of the graph). To obtain an estimate at the highest observed mean value, 29.444:

```
. replace tempvar = meanbase-29.444
```

and we re-use the  $\text{tempvar}$  variable and re-instantiate the metareg command:

```
. metareg diffimp tempvar, wsse(se_diffimp) graph
```

and the zero point now appears on the right side of the graph. (Note that the graph command will not work with more than one variable.) These models provide confidence intervals for these point estimates but not confidence bands and other additions; for these, see the next steps.

(3) Because the metareg command can be used with post-estimation commands, a number of other possibilities for graphing become possible. Specifically, after running the metareg command, run these commands:

(4) Now the variables saved in step 3 can be used for plotting. To produce a graph similar to that in Figure 1 (panel b):

```
. twoway rarea confl confu meanbase || line fit meanbase ||
```

Command	Explanation
. predict fit	Saves predicted values from the model for each $T$ in a variable called <i>fit</i>
. predict stdp, stdp	Saves the standard error of the prediction in a variable called <i>stdp</i>
. predict stdf, stdf	Saves the standard error of the forecast in a variable called <i>stdf</i>
. predict xbu, xbu	Saves empirical Bayes estimates (predictions including random effects) in a variable called <i>xbu</i>
. local t=invttail(e(df_r)-1, 0.025)	$t$ is used in calculating 95% confidence and prediction bands; for 90%, replace 0.025 with 0.05; for 99% replace it with 0.005
. gen confl=fit - `t' * stdp	Calculate lower confidence band (using standard error of the prediction), <i>confl</i>
. gen confu=fit + `t' * stdp	Calculate upper confidence band (using standard error of the prediction), <i>confu</i>
. gen predl=fit - `t' * stdf	Calculate lower prediction band (using standard error of the forecast), <i>predl</i>
. gen predu=fit + `t' * stdf	Calculate upper prediction band (using standard error of the forecast), <i>predu</i>

where *rarea* calls an area plot to depict the confidence bands and line plots the predicted values. Note that we created more elaborate versions of these commands to enhance their appearance (e.g., inserting reference lines, tick marks). To produce a bubble graph similar to Figure 3 (panel b):

<sup>8</sup>Readers can request from the authors a version of the current strategies for constructing confidence bands using the macros by Lipsey and Wilson (2001).



```
. twoway rarea confl confu meanbase || line fit meanbase || scatter diffimp meanbase [aw = 1/
se_diffimp^2], msymbol(Oh) ||
```

This command sizes bubbles according to the fixed-effects variance estimate; the random-effects equivalent can be produced by summing the random-effects variance component,  $\tau^2$ , and the fixed-effects variance:

```
. gen TWr = 1/(fevar +  $\tau^2$ )
```

where fevar is  $1/TW$  and  $\tau^2$  is a constant value taken from the statistical output. Now, the equivalent command sizes the bubbles according to their total weight:

```
. twoway rarea confl confu meanbase || line fit meanbase || scatter diffimp meanbase [aw = TWr],
msymbol(Oh) ||
```

Note that just like the variables in step 3, TWr would need to be recalculated for each such model because  $\tau^2$  and the other variables vary from model to model.

(5) To show the prediction interval surrounding the meta-regression line, such as in Figure 4 (panel a):

```
. twoway rarea predl predu meanbase || line fit meanbase ||
```

(6) To display the empirical Bayesian estimates along with the meta-regression line and confidence bands,

```
. twoway || rarea confl confu meanbase || line fit meanbase || scatter xbu meanbase, msymbol(t) ||
```

Or, to overlay the observed  $T_s$  along with their predicted true values assuming the fitted model is correct along with the confidence bands, similar to Figure 3 (panel d):

```
. twoway || rarea confl confu meanbase || line fit meanbase || scatter diffimp meanbase [aw = TWr], msymbol
(Oh) || scatter xbu meanbase, msymbol(t) ||
```

## Appendix II: Calculating and displaying confidence bands in meta-regression using metafor with R

We used commands similar to these using the R statistical package in order to create the graphs in Figures 2 and 3.

(1) Initially, install R and the package metafor, which is available via the Comprehensive R Archive Network (CRAN) at <http://CRAN.R-project.org/package=metafor>, the author's website at <http://www.wvbauer.com/>, or through a direct command within R: `install.packages("metafor")` (one needs an internet connection and appropriate access rights on the computer). Then, open metafor and load the data set, which in this example we have named QoL, using these commands:

```
library("metafor")
data("QoL", package="metafor")
print(QoL, row.names=FALSE)
```

(2) Now, fit the meta-analytic model with the quadratic effect of targeted aerobic METs and proportion of women using the command `rma.uni()` as follows:

```
aeromet <- rma(d, vi, mods = ~aerobicmet + I(aerobicmet^2) + expwomen, subset= ((design == 1
|design == 0) & fup == 2), data = QoL, method="ME")
```

The model has three independent variables: aerobicmet, aerobicmet<sup>2</sup> (the two aerobic METs terms), and expwomen (percentage of women in the sample). If fixed-effects assumptions are desired instead of mixed-effects, then replace ME with FE. The "subset=" portion of the command can be ignored; it is used in this case to narrow the sample to the cases that we wished to model.

(3) To obtain the predicted values:

```
predsaeromet <- predict(aeromet, newmods=cbind(seq(1,8,1), seq(1,8,1)^2, 79))
wi <- 1/sqrt(QoL$vi)
size <- 0.5 + 3 * (wi - min(wi))/(max(wi) - min(wi))
```

The value 79 is the sample mean for percentage of women.

(4) Then, plot the predicted values and the confidence bands:

```
plot(QoL$aerobicmet,QoL$d,type="n",xlab = "Targeted aerobic METs",
ylab = "QoL Effect Size (d)", xlim=c(1, 8), ylim=c(-0.1, 1.5))
lines(seq(1,8,1), predsaeromet$pred, col = "navy")
lines(seq(1,8,1), predsaeromet$ci.lb, lty = "dashed", col=" blue")
lines(seq(1,8,1), predsaeromet$ci.ub, lty = "dashed", col=" blue")
```

(5) To plot the estimates beyond the observed METS range of the studies (in this case up to 12, such as in Figure 2, panel b):

```
predsaeromet <- predict(aeromet, newmods=cbind(seq(1,12,1), seq(1,12,1)^2, 79))
wi <- 1/sqrt(QoL$vi)
size <- 0.5 + 3 * (wi - min(wi))/(max(wi) - min(wi))
plot(QoL$aerobicmet,QoL$d,type="n",xlab = "Targeted aerobic METs",
ylab = "QoL Effect Size (d)", xlim=c(1, 12), ylim=c(-0.1, 6))
lines(seq(1,12,1), predsaeromet$pred, col="red")
lines(seq(1,12,1), predsaeromet$ci.lb, lty = "dashed", col=" blue")
lines(seq(1,12,1), predsaeromet$ci.ub, lty = "dashed", col=" blue")
```

(6) To plot the observed values for a particular subset of the sample, initially, create a subsample:

```
dwomen <- subset(QoL, expwomen==100 & (design == 1 |design == 0) & fup == 2)
dmen <- subset(QoL, expwomen==0)
```

(7) Then plot studies with, for example, primarily men initially (as in Figure 2, panel c):

```
plot(dmen$aerobicmet,dmen$d, pch = 21, col = "black", bg = "navy", cex = size,
xlab = "Targeted aerobic METs", ylab = "QoL", ylim=c(-0.6, 1.5), xlim=c(1, 6))
lines(seq(1,6,1), predsaeromet$pred, col="dark green")
lines(seq(1,6,1), predsaeromet$ci.lb, lty = "dashed", col=" blue")
lines(seq(1,6,1), predsaeromet$ci.ub, lty = "dashed", col=" blue")
abline(h = 0, lty = "dotted", col="red")
```

## Acknowledgements

We express our appreciation to Estrellita Ballester, Justin C. Brown, Dipek Dey, Adam R. Hafdahl, Larry V. Hedges, David Hoaglin, Jessica M. LaCroix, Mark W. Lipsey, Tom Stanley, Michelle R. Warren, and three anonymous reviewers for helpful comments on previous drafts of this article, or on a related presentation at the 5th Annual Meeting of the Society for Research Synthesis Methodology (Ottawa, Ontario, Canada), or both. We also thank Wolfgang Viechtbauer for assistance on R syntax and David A. Kenny for discussions with the senior author in 2002 that ultimately led to this article. Its preparation was facilitated by U.S. Public Health Service Grants R01-MH58563 and K18-AL094581 to Blair T. Johnson.

## References

- Anzures-Cabrera, J., Higgins JPT. 2010. Graphical displays for meta-analysis: An overview with suggestions for practice. *Research Synthesis Methods* 1: 68–80.
- Berkey, C., Hoaglin DC, Mosteller F, Colditz GA. 1995. A random-effects regression model for Meta-analysis. *Statistics in Medicine* 14: 395–411.
- Borenstein, M., Hedges LV, Higgins JPT, Rothstein HR. 2009. Introduction to Meta-Analysis. John Wiley & Sons, West Sussex.
- Borman, G., Grigg JA. 2009. Visual and narrative interpretation, in The Handbook of Research Synthesis and Meta-analysis. 2nd Ed. (Eds: Cooper H, Hedges LV, Valentine JC) Russell Sage Foundation, New York, NY 497–519.
- Brown, J., Huedo-Medina TB, Pescatello LS, Pescatello SM, Ferrer RA, Johnson BT. 2011. Efficacy of exercise interventions in modulating cancer-related fatigue among adult cancer survivors: a meta-analysis. *Cancer Epidemiology Biomarkers and Prevention* 20: 123–133.
- Burns, P. 2006. R relative to statistical packages: a look at STATA, SAS and SPSS. Accessed 10 June 2010 ([http://lib.stat.cmu.edu/S/Spoetry/Tutor/R\\_relative\\_statpack.pdf](http://lib.stat.cmu.edu/S/Spoetry/Tutor/R_relative_statpack.pdf)).

- Cleveland, W. 1984. Graphs in scientific publications. *The American Statistician* **38**: 261–269.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd Ed. Lawrence Erlbaum Associates: Mahwah Hillsdale, NJ.
- Cohen, J, Cohen P, West SG, Aiken LS. 2003. *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd Ed. Lawrence Erlbaum Associates Publishers: Mahwah, NJ.
- Cooper, H, Hedges LV, Valentine JC. 2009. *The Handbook of Research Synthesis and Meta-analysis*. 2nd Ed. Russell Sage Foundation, New York, NY.
- Ferrer, R, Huedo-Medina TB, Johnson BT, Stacey R, Pescatello L. 2010. Exercise interventions for cancer survivors: a meta-analysis of quality of life outcomes. *Annals of Behavioral Medicine* **41**: 32–47.
- Harbord, R, Higgins JPT. 2008. Meta-regression in STATA. *The Stata Journal* **8**: 493–519.
- Hayter, A, Liu W, Wynn HP. 2007. Easy-to-construct confidence bands for comparing two simple linear regression lines. *Journal of Statistical Planning and Inference* **137**: 1213–1225.
- Hedges, L. 1992. Meta-analysis. *Journal of Educational Statistics* **17**: 279–296.
- Hedges, L, Olkin I. 1985. *Statistical Methods for Meta-analysis*. Academic Press, Orlando, FL.
- Hedges, L, Vevea JL. 1998. Fixed- and random-effects models in meta-analysis. *Psychological Methods* **3**: 486–504.
- Higgins, J, Green S. 2009. *Cochrane handbook for systematic reviews of interventions*, version 5.0.2. Accessed 18 February 2011 ([www.cochrane-handbook.org](http://www.cochrane-handbook.org)).
- Higgins, J, Thompson SG. 2002. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**: 1539–58.
- Huedo-Medina, T, Johnson BT. 2010. *Modelos Estadísticos en Meta-Análisis [Statistical Models in Meta-Analysis]*. Netbiblio, La Coruña.
- Hunter, J, Schmidt FL. 2004. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, 2nd Ed. Sage Publications, Thousand Oaks, CA.
- IBM. SPSS Statistics. 2010. Release 19.0, SPSS Inc., Chicago, IL.
- Johnson, BT, Carey MP, Marsh KL, Levin KD, Scott-Sheldon LAJ. 2003. Interventions to reduce sexual risk for the human immunodeficiency virus in adolescents, 1985–2000. *Archives of Pediatrics and Adolescent Medicine* **157**: 381–388.
- Johnson, BT, Kirsch I. 2008. Do antidepressants work? Statistical significance versus clinical benefits. *Significance* **5**: 54–58.
- Kirsch, I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. 2008. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Medicine* **5**(e45): 0260–0268.
- Knapp, G, Hartun J. 2003. Improved tests for a random-effects meta-regression with a single covariate. *Statistics in Medicine* **22**: 2693–2710.
- Konstantopoulos S, Hedges LV. 2004. Meta-analysis, in *The Handbook of Quantitative Methods in the Social Sciences*, (Ed: Kaplan M), Sage Publications, Thousand Oaks, CA 281–297.
- Konstantopoulos, S, Hedges LV. 2009. Analyzing effect sizes: fixed-effects models, in *The Handbook of Research Synthesis and Meta-Analysis*. 2nd Ed., (Eds: Cooper H, Hedges LV, Valentine JC). Russell Sage Foundation, New York, NY 279–294.
- Lane, D, Sándor A. 2009. Designing better graphs by including distributional information and integrating words, numbers, and images. *Psychological Methods* **14**: 239–257.
- Lau, J, Ioannidis JPA, Terrin N, Schmid CH, Olkin I. 2006. The case of the misleading funnel plot. *British Medical Journal* **333**: 597–600.
- Light, E, Singer JD, Willett JB. 1994. The visual presentation and interpretation of meta-analyses, in *The Handbook of Research Synthesis and Meta-analysis*, 2nd Ed., (Eds: Cooper H, Hedges LV, Valentine JC) Russell Sage Foundation, New York, NY 439–453.
- Lipsey, M, Wilson DB. 2001. *Practical Meta-Analysis*. Sage Publications, Thousand Oaks, CA.
- Liu, W, Lin S, Piegorsch WW. 2008. Construction of exact simultaneous confidence bands for a simple linear regression model. *International Statistical Review* **76**: 39–57.
- MacKassay, S, Provost, F. 2004. Confidence Bands for ROC Curves: Methods and an Empirical Study. Proceedings of the First Workshop on ROC Analysis in AI.
- Myers, J, Well AD. 2003. *Research Design and Statistical Analysis*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Raudenbush, S. 2009. Analyzing effect sizes: random-effects models, in *The Handbook of Research Synthesis and Meta-Analysis*, 2nd Ed., (Eds: Cooper H, Hedges LV, Valentine JC), 295–315. Russell Sage Foundation, New York, NY.
- Seber, G, Lee AJ. 2003. *Linear Regression Analysis*. 2nd Ed. John Wiley & Sons, Hoboken, NJ.
- Solomon, S, Plattner G, Knutti R, Friedlingstein P. 2009. PNAS: irreversible climate change due to carbon dioxide emissions. *Proceedings of the National Academy of Science of the United States of America* **106**: 1704–1709.
- StataCorp. 2009. STATA statistical software: release 11, StataCorp LP, College Station.
- Sweller, J, Chandler P, Tierney P, Cooper M. 1990. Cognitive load as a factor in the structure of technical material. *Journal of Experimental Psychology* **119**: 176–192.
- Thombs, B, Coyne JC, Cuijpers P, de Jonge P, Gilbody S, Ioannidis JP, Johnson BT, Patten SB, Turner EH, Ziegelstein RC. 2011. Rethinking recommendations for screening for depression in primary care. *Canadian Medical Association Journal*. DOI: [cmaj.111035](https://doi.org/10.1503/cmaj.111035).

- Tufte E. 2006. *Beautiful Evidence*. Graphic Press, Cheshire, CT.
- Tufte, E. 2001. *The Visual Display of Quantitative Information*. 2nd Ed. Graphic Press, Cheshire.
- Viechtbauer, W. 2010a. The metafor package: a meta-analysis package for R. Accessed 10 June 2010 (<http://www.viechtbauer.com/downloads.html>).
- Viechtbauer, W. 2010b. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* **36**: 1–48.
- Working, H, Hotelling H. 1929. Applications of the theory of error to the interpretation of trends. *Journal of the American Statistical Association* **24**: 73–85.