

Visual Representations of Meta-Analyses of Multiple Outcomes: Extensions to Forest Plots, Funnel Plots, and Caterpillar Plots

Belén Fernández-Castilla^{ab}, Lies Declercq^{ab}, Laleh Jamshidi^{ab}, Susan Natasha Beretvas^c, Patrick Onghena^a, Wim Van den Noortgate^{ab}

[a] Faculty of Psychology and Educational Sciences, KU Leuven (University of Leuven), Leuven, Belgium. [b] Imec-ITEC, KU Leuven (University of Leuven), Leuven, Belgium. [c] University of Texas at Austin, Austin, USA.

Methodology, 2020, Vol. 16(4), 299–315, <https://doi.org/10.5964/meth.4013>

Received: 2019-06-13 • Accepted: 2020-05-24 • Published (VoR): 2020-12-22

Corresponding Author: Belén Fernández-Castilla, Faculty of Psychology and Educational Science, KU Leuven, Etienne Sabbelaan 51, 8500, Kortrijk, Belgium. Tel.: +32 468 211 544, E-mail: Belen.fernandezcastilla@kuleuven.be

Abstract

Meta-analytic datasets can be large, especially when in primary studies multiple effect sizes are reported. The visualization of meta-analytic data is therefore useful to summarize data and understand information reported in primary studies. The gold standard figures in meta-analysis are forest and funnel plots. However, none of these plots can yet account for the existence of multiple effect sizes within primary studies. This manuscript describes extensions to the funnel plot, forest plot and caterpillar plot to adapt them to three-level meta-analyses. For forest plots, we propose to plot the study-specific effects and their precision, and to add additional confidence intervals that reflect the sampling variance of individual effect sizes. For caterpillar plots and funnel plots, we recommend to plot individual effect sizes and averaged study-effect sizes in two separate graphs. For the funnel plot, plotting separate graphs might improve the detection of both publication bias and/or selective outcome reporting bias.

Keywords

meta-analysis, graphical display, forest plot, funnel plot, (3-level) multilevel

The visual representation of quantitative data is valuable for any scientific discipline. The added value of using graphical displays is twofold. It can help researchers to understand better the behavior of their data (Anzures-Cabrera & Higgins, 2010) and even to detect new associations or patterns among variables that were not immediately evident before (Bulté & Onghena, 2012; Egger & Carpi, 2008; Manolov, Solanas, & Leiva, 2010). In



addition, visualizing data facilitates for non-researchers (e.g., practitioners, politicians, etc.) a better understanding of the conclusions of a study.

In the field of meta-analysis, defined as the statistical technique that combines evidence from different studies to reach more accurate conclusions (Glass, 1976), the graphical representation of data is especially important due to the nature of the meta-analytic datasets and analyses (Schild & Voracek, 2013). Meta-analytic datasets are often large, so graphical displays can help to summarize the most relevant information of the studies collected, to give a first impression of the variability among effect sizes (ESs), and to detect possible outliers. In addition, in meta-analysis it is quite common to perform additional analyses besides the main one (i.e., the derivation of an overall effect estimate and its standard error), such as moderator and sensitivity analyses. Visualizing the results of these additional analyses can further help researchers, and (especially) non-researchers, to interpret the results and even to improve the accuracy of the conclusions (Schild & Voracek, 2015).

There are two figures that are the gold standard in meta-analysis: forest plots and funnel plots. Forest plots (Lewis & Clarke, 2001) include information about the effects of the studies included in the meta-analysis and their uncertainty, as well as the combined ES (over all studies, and sometimes also per category of studies) and its confidence interval (CI). Forest plots are undoubtedly the most popular figures in meta-analysis: they constitute more than 50% of the plots reported (Schild & Voracek, 2013), and several guidelines for systematic reviews and meta-analyses explicitly recommend the inclusion of this figure in the reports (e.g., Preferred Reporting Items for Systematic Reviews and Meta-Analyses [PRISMA]; Moher, Liberati, Tetzlaff, & Altman, 2009). Caterpillar plots are a special version of forest plots where ESs are sorted by their magnitude (possibly within categories of studies) and plotted closer to each other. Another popular figure specific for meta-analysis is the funnel plot (Light & Pillemer, 1984). The funnel plot is a diagnosis plot that helps researchers to assess the variability among ESs and the possible presence of publication bias through visual inspection.

The conventions for displaying forest plots and funnel plots are widely known for the scenario where a unique ES is reported per study (see, for instance, Sterne, Becker, & Egger, 2005, for funnel plots and Lewis & Clarke, 2001, for forest plots). Nowadays, however, primary studies rarely include a single ES. For instance, it is quite common that primary studies report ESs for multiple outcome variables, or that more than two groups are compared with a single comparison group, leading to multiple ESs per study. But how should these types of meta-analytic data be represented in a forest plot or in funnel plot? Up to today, no adaptations to the forest and funnel plots have been proposed for the case where multiple ESs are reported within studies. Given the known importance of visual representation in meta-analysis, the aim of this study is to fill this gap by suggesting graphical extensions to these figures in the specific context where multiple ESs are available within primary studies.

Before detailing these extensions, we first give an overview of the available statistical models to perform meta-analyses, because some of the elements of the forest and funnel plots depend on the model fitted. The traditional random-effects model assumes that observed differences between ESs are not only due to random sampling variation, but also due to systematic differences across studies (e.g., each study uses its specific procedures, protocols, etc.). Under this model, it is assumed that each study has its own population value and that study-specific values are normally distributed around a ‘grand’ mean population effect (γ). The traditional random-effects model can be expressed as follows:

$$d_k = \gamma + e_k + u_k, \quad (1)$$

where d_k refers to the ES reported in study k , and e_k and u_k are normally distributed residuals with variance $\sigma_{e_k}^2$ and σ_u^2 , respectively. The term $\sigma_{e_k}^2$ refers to the sampling variance and σ_u^2 refers to the systematic variability of ESs across studies. For commonly used effect size measures, this sampling variance can easily be derived per study and therefore is considered as known in the meta-analysis. If $\sigma_u^2 = 0$, this model reduces to a fixed-effect model, that assumes that there is a unique population effect and that the observed heterogeneity among ESs is only due to sampling variation.

When studies report multiple ESs, it is important to apply a statistical model that takes into account that ESs from the same study are related to each other, that is, that dependency among ESs exists. Ignoring this dependency can lead to biased estimates of the standard error and of the variance components, and hence to flawed inferences (Becker, 2000). There are three popular approaches to model dependent ESs: the multivariate approach (Raudenbush, Becker, & Kalaian, 1988), the Robust Variance Estimation method (RVE; Hedges, Tipton, & Johnson, 2010), and the application of three-level models (Cheung, 2014; Konstantopoulos, 2011; Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013, 2015). For a complete overview of under which circumstances is better to use each of these methods, we refer to López-López, Page, Lipsey, and Higgins (2018). For this study, we will focus on the three-level approach for two reasons: 1) Because the use of three-level models for treating dependent ESs has substantially increased in the last years (Fernández-Castilla et al., 2019), but researchers do not yet have options to depict data that have this three-level structure, and 2) Because by applying three-level models it is possible to get separate estimates of the variance of the observed ESs across studies and within studies.

The three-level approach consists of extending the random-effects model of Equation 1 by adding an additional random residual to model the deviation of a population ES within a study from the mean ES in that study (Cheung, 2014; Van den Noortgate et al., 2013, 2015). This extension results in a three-level model that assumes that ESs not only differ due to sampling variation and due to systematic differences between studies, but also due to systematic differences across the outcomes measured. This three-level model can be expressed as follows:

$$d_{jk} = \gamma_{00} + e_{jk} + v_{jk} + u_{0k}, \quad (2)$$

where d_{jk} refers to the observed ES of outcome j within study k , and γ_{00} is the overall population ES. Each type of residual, e_{jk} , v_{jk} , and u_{0k} is assumed to be normally distributed with mean 0 and variance $\sigma_{e_{jk}}^2$, σ_v^2 , and σ_u^2 , respectively. The sampling variance, $\sigma_{e_{jk}}^2$, is estimated in advance and therefore assumed to be known in the meta-analysis, and the terms σ_v^2 and σ_u^2 refer to the variability of ESs within studies and to the between-studies variance respectively.

Each of the following sections begins with a description of traditional plots, and continues with suggestions of extensions. In the [Supplementary materials](#), the code for constructing these augmented figures in R using package *ggplot2* (Wickham, 2011) is available.¹

Forest Plots

Forest plots give information about the precision and the weight of each study. Each ES is depicted with a dot (usually a square) on the horizontal axis, and each row in the graph represents a different study. In a meta-analysis in which just one ES per study is reported, the precision of each study is represented with the (typically) 95% CI depicted along the sides of each ES (i.e., dots). This CI is built using the sampling variance, $\hat{\sigma}_{e_k}^2$:

$$d_k \pm Z_{\alpha/2} \cdot \sqrt{\hat{\sigma}_{e_k}^2} \quad (3)$$

where $Z_{\alpha/2} = 1.96$ if $\alpha = .05$. The size of the central dot is proportional to the weight assigned to that study (W_k). In the case of a traditional random-effects model (Equation 1), the weight of each study equals:

$$W_k = \frac{1}{\hat{\sigma}_{e_k}^2 + \hat{\sigma}_u^2} \quad (4)$$

where $\hat{\sigma}_u^2$ is the estimated between-studies variance. If $\hat{\sigma}_u^2 = 0$, then the weights resulting from a random-effects model equal the ones of a fixed-effect model.

At the bottom of the graph, a diamond representing the pooled ES is typically depicted, with its horizontal diagonal representing its 95% CI. Additionally, it is recommended to plot a vertical line that indicates the value of the overall effect estimate, and also a 'no effect' vertical line that serves as an indicator of whether ESs are negative and positive, or whether studies report statistically significant ESs.

1) Part of the code for building forest plots has been taken from the R code supplied by Schild and Voracek (2015), that is now part of the package *metaviz* (Kossmeier, Tran, & Voracek, 2019b).

In meta-analysis, there is a direct relationship between the precision of a study and the extent to which that study contributes to the final estimation of the combined ES. This relationship is visually evident in forest plots: The width of the CI of a study (i.e., its precision) is related to the area of the dot of that study (i.e., its weight), that is, less precise studies contribute to a minor extent to the estimation of the pooled ES. It is important to mention that the study-weights depend on the model fitted. When a fixed-effect model is fitted, the weight equals the inverse of the sampling variance, whereas when a random-effects model is fitted (Equation 1), the weight assigned to a specific study also depends on the between-studies variance estimate (see Equation 4). As a consequence, small studies will get a smaller weight than large studies, especially when using a fixed-effect model.

The weight of a primary study that reports multiple ESs depends on more parameters than solely the sampling variance and the between-studies variance. This can be clearly understood by looking at the variance of the mean ES estimate obtained when we apply a random-effects model to each study k separately:

$$\text{var}(d_{.k}) = \frac{1}{\sum_j W_{jk}} = \frac{1}{\sum_j \frac{1}{\hat{\sigma}_{e_k}^2 + \hat{\sigma}_v^2}} \quad (5)$$

The precision (or inverse variance) of the effect estimate for a study k that includes j ESs is equal to the sum of the precisions of the individual ESs ($\sum_j W_{jk}$), and therefore is determined by the 1) sampling variances (which depends strongly on sample size), 2) number of ESs reported, and 3) the variability among these outcomes. If the sampling variances of ESs in study k are small, the denominator of Equation 5 will be large and hence the final $\text{var}(d_{.k})$ will be smaller, indicating a higher precision of $d_{.k}$. Second, the more ESs reported in study k , the larger the sum of precisions, and hence the smaller the estimated variance of $d_{.k}$. Third and last, if the ESs do not vary substantially within study k , that is, if $\hat{\sigma}_v^2$ is small, the denominator of Equation 5 (i.e., $\sum_j (1/(\hat{\sigma}_{e_k}^2 + \hat{\sigma}_v^2))$) will be large, which in turn will lead to a smaller value of the $\text{var}(d_{.k})$ compared to other studies where the within-ESs variance is larger.

Moving on to the weights, let us remember that in a traditional forest plot of a random-effects model, the weight of each study depends on both the sampling variance and the between-studies variance (Equation 4). Likewise, in a three-level meta-analysis, the weight of each study depends on the precision of that study (which in turn depends on the sampling variances, the number of outcomes and the between-outcome variance²) and on the between-studies variance. For obtaining the actual weight of study k in the context of multiple ESs per study, first it is necessary to calculate the total variance of

2) When performing a three-level meta-analysis, this between-outcome variance is typically assumed to be the same for all studies, and therefore is estimated across all studies.

that study, $Var(d_{.k})^*$, around the overall effect across studies. This variance (in contrast to the variance calculated using Equation 5) also incorporates the between-studies variance estimate.

$$Var(d_{.k})^* = (\mathbf{X}^T \mathbf{V}_k^{-1} \mathbf{X})^{-1} \quad (6)$$

The matrix \mathbf{V}_k^{-1} equals the inverse of the variance-covariance matrix of the J ESs reported in study k . This matrix, \mathbf{V}_k^{-1} , contains information about the between-outcomes and between-studies variance estimates (that is the same for all studies). The element \mathbf{X} refers to a $1 \times J_k$ matrix of 1. The weight of study k is therefore:

$$W_{.k} = \frac{1}{Var(d_{.k})^*} \quad (7)$$

More information about how to obtain \mathbf{V}_k^{-1} can be found in Konstantopoulos (2011), and the R code necessary to obtain the variance of a specific study (Equation 6) is inserted in the code for building the graphs, that is available in the [Supplementary materials](#).

In order to appropriately visualize the different factors that affect the precision of individual ESs and studies, we propose to draw a modified version of the ‘summary forest plots’ (Anzures-Cabrera & Higgins, 2010). In these plots, separate meta-analyses are depicted based on categories of moderator variables. Our proposal, however, is to perform separate meta-analyses for each study, so that a line in the plot represents the results of a meta-analysis of one study instead of individual ESs. To illustrate this, we will use the data of the meta-analysis of Geeraert, Van den Noortgate, Grietens, and Onghena (2004), that evaluates the effectiveness of early prevention programs on children at risk of suffering abuse. To make the forest plot not too large, we will use only a subset of 10 studies (with 208 ESs). This subset of studies is available in the [Supplementary materials](#).

The summary forest plot shown in Figure 1A resembles a traditional forest plot, but there is one crucial difference: Each black line does not longer represent individual ESs, but the results of meta-analyses of outcomes within each study. Before drawing the forest plot, a separate meta-analysis has been carried out on each study (using a random-effects model), and afterwards a dot representing the meta-analytic study mean has been depicted in each line along with its 95% CI in bold black:

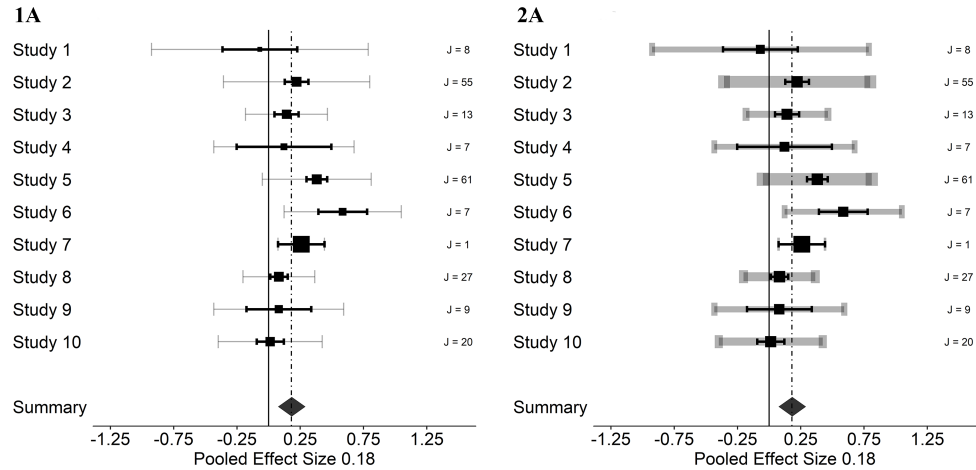
$$d_{.k} \pm 1.96 \cdot \sqrt{Var(d_{.k})} \quad (8)$$

where $d_{.k}$ designates the meta-analytic mean of observed ESs from study k , and $\sqrt{Var(d_{.k})}$ is the standard error of this estimated overall study effect, defined in Equation 5. The reason for using Equation 5 rather than Equation 6, is that we want to express how precise we can estimate the study-mean effect size, using the data from that study

only (in the same way as in a traditional forest plot, the CIs around effect sizes are calculated independently for each study).

Figure 1

Summary Forest Plots



Note. In 1A, the thickness of the grey confidence intervals is the same for all studies, and in 1B, the thickness of the grey confidence intervals is proportional to the number of effect sizes reported within studies.

As shown before, the total precision of the studies included in Figure 1A depends on the number of outcomes measured within a study, the sample size on which they are based on, and the variability among them. However, it is not possible to disentangle the influence that these three parameters have on the total study precision by looking only at the black CIs. In order to account for this information, we propose to add two more elements.

The first element is an additional grey CI, that is based on the sampling variance of individual observed ESs of the study. The formula for obtaining the upper and lower limit for these grey CIs is the following:

$$d_{.k} \pm 1.96 \cdot \text{median}(SE_{jk}), \tag{9}$$

where $\text{median}(SE_{jk})$ represents the median standard error of the observed effects for the j outcomes included in study k . The median is preferred over the mean due to the known skewed distribution of standard errors. With the inclusion of these grey CIs it is possible to know the contribution that the study sample size alone has on the total

study precision (black CI), because neither the number of outcomes nor their variance are included in Equation 9. In other words, these grey CIs reflect the median precision of one ES within that study. For instance, Study 3 and 8 show somewhat narrow grey CIs compared to other studies, meaning that the sample sizes in these studies is large, which is likely to be the reason why the precision of those studies is quite high (i.e., narrow black CI). Opposite examples can be found in Studies 1 and 4, where the grey CIs are quite long, meaning that the sample sizes are small and hence do not contribute much to the total precision of these studies, that is indeed quite low (wide black CIs). Another scenario is illustrated by Study 2 and Study 5: Although the sampling variance is rather large (i.e., wide grey CI), the black CI is narrow, which means that the total precision of those studies is quite high despite their low sample size. Therefore, the high precision exhibited by these studies has to be due to other reasons, as for instance a large number of outcomes reported in those studies and/or a low variability between them. In order to know this information, the second element we propose to incorporate is the number of outcomes reported in each study. We have depicted this information on the right side of Figure 1A. We can indeed see that a large number (55 and 61, respectively) of outcomes were reported in Study 2 and 5, which explains the high overall precision. In Study 7, only one outcome was reported ($J = 1$), and consequently the grey and black CIs coincide. It could be also the case that in a given study, the black CI was wider than the grey CI, that is, that the total precision of a study was smaller than the precision given by the sample size. This can happen when few outcomes are reported and when the variability among them is large. Although in that case the grey interval will be hidden by the bold one, the markers of the grey interval will still be visible. Another option to visually represent the number of outcomes within studies is shown in Figure 1B. In this figure, the size of the grey CIs is proportional to the number of effect sizes reported within studies. This feature is inspired by the thick forest plots proposed by Schild and Voracek (2015), where the size of the lines is proportional to the weight assigned to each study.

The weight assigned to each study is represented by the area of the central dot, and in a three-level meta-analysis the study weight can be calculated with Equation 6. The weights of a traditional random-effects model are normally more similar (compared to the ones from fixed-effect model) because the between-studies variance estimate is the same for all studies. In a three-level meta-analysis, this effect is even more pronounced because all studies have in common the between-studies variance and the between-outcomes variance estimates.

The final conclusion of the forest plot in Figure 1 is that Study 7 is the one that contributes more to the estimation of the pooled effect, apparently due to the large sample size of this study. Study 2, 5 and 8 also get a relatively high weight, reflecting their high precision because of the large number of outcomes reported. On the other hand, the precision of Study 1 and Study 4 is quite low, possibly due to their small sample

sizes (i.e., wide grey CIs). Therefore, these studies are the ones contributing the least to the summary pooled effect and to its precision.

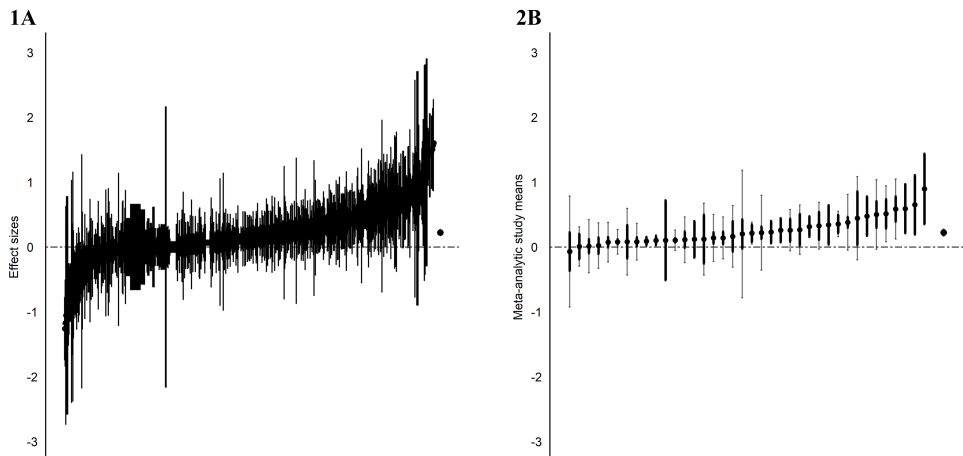
Caterpillar Plots

Caterpillar plots are nothing but rotated forest plots. The main differences are that ESs are sorted and plotted by their magnitude (from negative to positive) and that ESs are depicted closer to each other, making it possible to get a general view of the distribution of all ESs. An advantage of plotting ESs closer to each other is that many of them can be plotted at the same time, whereas including too many ESs in the forest plot would make it difficult to interpret.

Because caterpillar plots allow the inclusion of many ESs, we have used the complete dataset of Geeraert et al. (2004) for plotting Figure 2 (38 studies and 583 ESs). In Figure 2A we have plotted each ES and its 95% CI, ordered by their magnitude, and in Figure 2B we have plotted the study effects, using the same two black and grey CIs as in the forest plots of Figure 1. At the end of both graphs, the overall effect and its CI have been plotted. Our recommendation is to include both plots together, so it is possible to see the amount of (negative or positive) ESs that are statistically different from zero, and at the same time it is possible to know which studies report (positive or negative) significant pooled study-effects.

Figure 2

(A) Caterpillar Plot With All Effect Sizes and Their Confidence Intervals and (B) Study-Caterpillar Plot



Funnel Plots

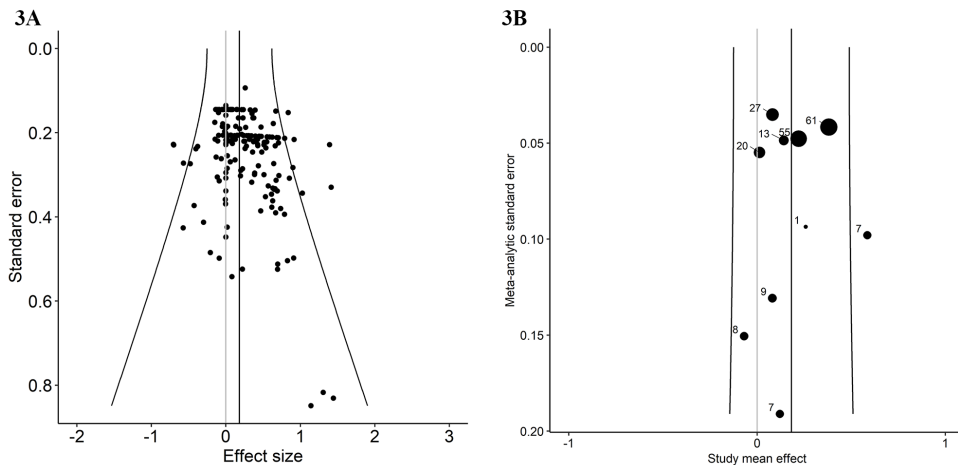
This figure is commonly used in meta-analysis to investigate the presence of selection bias. The funnel plot is a scatter plot in which ESs (horizontal axis) are plotted against a measure of their precision (vertical axis). This measure of precision can be the sampling variance, the standard error, or their inverses. In this paper, we will use the standard error for the vertical axis, as recommended by [Anzures-Cabrera and Higgins \(2010\)](#). The scale of this vertical axis is inverted, so ESs with small standard errors are drawn at the top of the graph. Ninety-five percent pseudo-CIs are depicted around the pooled effect to mark the region where 95% of the ESs are expected in the absence of between-studies heterogeneity. If selection bias does not exist, we would expect studies based on large samples to be closer to the pooled effect in comparison to more imprecise studies, that are expected to be more widely spread along both sides of the overall effect at the base of the graph, leading to a symmetrical inverted funnel shape. If a part of the funnel is empty (normally in the lower part), publication bias could exist. However, funnel plots can be asymmetrical for other reasons, such as for instance the existence of large between-studies heterogeneity ([Terrin, Schmid, Lau, & Olkin, 2003](#)).

Besides publication bias, other types of bias might also be present, as for instance selective outcome reporting bias ([Ioannidis, Munafo, Fusar-Poli, Nosek, & David, 2014](#)). Authors might measure a construct using several measurements, and then only report the outcome(s) in which they found significant results. Selective outcome reporting bias can at the same time propitiate publication bias, because studies with (many) positive outcomes are more likely to be published than studies with negative or non-significant outcomes.

In a three-level meta-analysis, we recommend to first plot all ESs in the funnel plot to check for both selective outcome reporting bias and publication bias. In [Figure 3A](#), there are three ESs at the lower-right part of the graph, whereas there are no study results with a similar imprecision around the mean ES or at the negative side of the plot. This could be an indication of selective outcome reporting bias or publication bias; studies might have reported only the more salient outcomes and/or small studies with small or unexpected study effects might have been censored.

Figure 3

(A) Funnel Plot of All Effect Sizes and (B) Study-Funnel Plot that Includes the Number of Outcomes Reported in Each Study



Note. In addition, the size of the dots is proportional to the number of effect sizes included in those studies. Notice that the y- and x-axis of 3A differ in their dimensions from the ones of 3B.

In order to further clarify which of these kinds of bias took place, we propose to additionally plot study effects against their meta-analytic standard errors in another study-funnel plot (Figure 3B). These study effects have been estimated by performing separate random-effects meta-analyses to each study. It is important to remember that now the precision of each study (reflected in the vertical axis) not only depends on the sample size, but also on the number of outcomes reported and on their variability (Equation 5). Therefore, ESs of studies with a small sample size, reporting few outcomes, and/or with a large between-outcomes variance will be plotted at the bottom of the graph, while studies with large sample sizes, reporting many outcomes, and with a small within-study variability will be plotted at the top of the graph, because they are expected to vary less around the pooled effect. In Figure 3B we can see that the distribution of study-specific effects is actually quite symmetrical, so there is no indication of the existence of publication bias, although ESs in Figure 3A were not that symmetrically distributed around the mean ES. However, it should be kept in mind that we cannot discard the possibility that the asymmetry of any funnel plot is due to some other reasons, as mentioned before. Moreover, the number of studies in our example is probably too small to assess asymmetry across studies. There are some strategies, such as plotting enhanced-contour funnel plots (Peters, Sutton, Jones, Abrams, & Rushton, 2008)

or comparing real-data funnel plots with null-hypothesis funnel plots (Kossmeier, Tran, & Voracek, 2019a) that can help identifying whether publication bias actually exists.

An additional suggestion for the study-funnel plot is to depict the number of outcomes reported in each study next to the dots representing the study effects and/or to plot the size of the study-effect proportionally to the number of effects included in that study (as in Figure 3B)³. The shape of the study-funnel plot might be symmetric, but it could be the case that studies with large positive effects reported fewer outcomes in comparison with studies that reported on average smaller or even negative effects, which would again be an indication of reporting bias (Van den Noortgate et al., 2013).

Another special characteristic of the funnel plots in Figure 3 are the 95% confidence bands. In a traditional, two-level - random-effects model, it is expected that highly precise studies also vary around the overall mean (i.e., black vertical line) due to systematic differences between studies. Therefore, under a traditional random-effects model 95% pseudo-CIs should leave some space on the sides of the overall effect at the top of the graph, so the between-studies variability among highly precise ESs is taken into account. The formula for obtaining these 95% pseudo-CIs for a traditional random-effects model is the following:

$$d_{..} \pm 1.96 \cdot \sqrt{\hat{\sigma}_{e_{jk}}^2 + \hat{\sigma}_u^2} \quad (10)$$

where $\hat{\sigma}_{e_{jk}}^2$ corresponds to each of the standard error values in the y -axis of the plot, and $d_{..}$ represents the estimate of the overall ES. When several ESs are reported within studies, it is necessary to adapt this formula by adding the between-outcomes variance estimate ($\hat{\sigma}_v^2$) in the equation. Hence, the formula used for the contours depicted in Figure 3A is the following:

$$d_{..} \pm 1.96 \cdot \sqrt{\hat{\sigma}_{e_{jk}}^2 + \hat{\sigma}_u^2 + \hat{\sigma}_v^2} \quad (11)$$

In Figure 3B, study means are plotted instead of individual ESs. The x -axis refers to the estimated study effects, and the y -axis represents the standard error of these estimated study effects. Therefore, the formula for the contours should be adapted by substituting the between-outcomes variance for the mean between-outcomes variance:

$$d_{..} \pm 1.96 \cdot \sqrt{\frac{\hat{\sigma}_{e_{.k}}^2 + \hat{\sigma}_v^2}{\text{median}(J_k)} + \hat{\sigma}_u^2} \quad (12)$$

3) In the functions included in the R code in the Supplementary material, the user can select whether or not to plot the number of effect sizes within studies next to each dot, or whether or not to plot the size of the dots proportional to the number of effect sizes within studies.

In Equation 12, each of the meta-analytic standard error values of the y -axis has less influence in the calculation of the upper and lower bound of the 95% CIs because this value is always divided by a constant value [$\text{median}(J_k)$]. As a consequence, the contours or bands will look more vertical than in a standard funnel plot.

Discussion

Visualization of data is very important to understand, interpret, and summarize results, regardless of the research area of interest (Egger & Carpi, 2008). In meta-analysis, two types of plots are commonly reported: the forest plot (or its ordered variant, the caterpillar plot) and the funnel plot (Schild & Voracek, 2013). In this study, we have suggested adaptations to these graphs to the scenario in which multiple ESs are reported within primary studies. These extensions can help researchers to summarize more precisely the information obtained when a three-level meta-analysis is performed, to understand better the information reported in primary studies, and to make more accurate visual inferences about the possible existence of selection bias, selective outcome reporting bias, or both.

An advantage of the proposed forest plot is that a lot of information is included in just one image: the study-effects and their uncertainty, the uncertainty due to sampling variation, the number of outcomes reported in each study, the importance of each study in the estimation of the pooled ES, the extent to which each study contributes to the between-outcomes variance estimate, the value of the overall effect, and its precision. Regarding caterpillar plots, by plotting separate caterpillar plots for the observed ESs and for the study-effects, it is possible to get information of the pattern of the effects at the lowest level and at the upper-study level. Other existing graphs, such as Galbraith plots (Galbraith, 1994), L'Abbé plots (L'Abbé, Detsky, & O'Rourke, 1987) or rainforest plots (Zhang, Kossmeier, Tran, Voracek, & Zhang, 2017) can be adapted in similar ways. In the same way, by plotting separate funnel plots for individual ESs and for study-effects it is possible to check for indications of the presence of selective outcome reporting bias, publication bias, or both, keeping in mind that asymmetry of the funnel plot does not necessarily indicate the presence of any bias. Finally, we have also proposed some modifications in the formula for constructing the 95% pseudo-CI so that they account for the between-outcomes variability.

In sum, we believe that the incorporation of these extensions to forest plots and funnel plots in three-level meta-analysis can greatly help to accurately summarize and understand these (often large) datasets, as well as to improve initial inferences about the presence of selection bias.

Funding: The authors have no funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Acknowledgments: The authors have no support to report.

Author Note: This article was written within the scope of the Doctoral dissertation of the first author (Fernández-Castilla, 2019).

Supplementary Materials

For this article the following supplementary materials are available (Fernández-Castilla, Declercq, Jamshidi, Beretvas, Onghena, & Van den Noortgate, 2020):

- Database used to generate the figures of this manuscript.
- R Code to generate the figures described in this manuscript.

Index of Supplementary Materials

Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2020). *Supplementary materials to: Visual representations of meta-analyses of multiple outcomes: Extensions to forest plots, funnel plots, and caterpillar plots* [Code]. *PsychOpen*. <https://doi.org/10.23668/psycharchives.4420>

References

- Anzures-Cabrera, J., & Higgins, J. (2010). Graphical displays for meta-analysis: An overview with suggestions for practice. *Research Synthesis Methods*, 1(1), 66-80. <https://doi.org/10.1002/jrsm.6>
- Becker, B. J. (2000). Multivariate meta-analysis. In H. E. A. Tinsley & E. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 499-525). Orlando, FL, USA: Academic Press. <https://doi.org/10.1016/B978-012691360-6/50018-5>
- Bulté, I., & Onghena, P. (2012). When the truth hits you between the eyes. *Methodology*, 8(3), 104-114. <https://doi.org/10.1027/1614-2241/a000042>
- Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, 19(2), 211-229. <https://doi.org/10.1037/a0032968>
- Egger, A. E., & Carpi, A. (2008). *Using graphs and visual data in science*. Retrieved from <https://www.visionlearning.com/en/library/Process-of-Science/49/Using-Graphs-and-Visual-Data-in-Science/156/reading>

- Fernández-Castilla, B. (2019). Models and techniques for the meta-analysis of dependent effect sizes (Doctoral dissertation, University of Leuven). Retrieved from <https://lirias.kuleuven.be/2831603?limo=0>
- Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2019). *The application of meta-analytic models with multiple random effects: A systematic review*. PsyArXiv. <https://doi.org/10.31234/osf.io/ugwn8>
- Galbraith, R. F. (1994). Some applications of radial plots. *Journal of the American Statistical Association*, 89(428), 1232-1242. <https://doi.org/10.1080/01621459.1994.10476864>
- Geeraert, L., Van den Noortgate, W., Grietens, H., & Onghena, P. (2004). The effects of early prevention programs for families with young children at risk for physical child abuse and neglect. A meta-analysis. *Child Maltreatment*, 9(3), 277-291. <https://doi.org/10.1177/1077559504264265>
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3-8. <https://doi.org/10.3102/0013189X005010003>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39-65. <https://doi.org/10.1002/jrsm.5>
- Ioannidis, J. P., Munafo, M. R., Fusar-Poli, P., Nosek, B. A., & David, S. P. (2014). Publication and other reporting biases in cognitive sciences: Detection, prevalence, and prevention. *Trends in Cognitive Sciences*, 18(5), 235-241. <https://doi.org/10.1016/j.tics.2014.02.010>
- Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods*, 2(1), 61-76. <https://doi.org/10.1002/jrsm.35>
- Kossmeier, M., Tran, U. S., & Voracek, M. (2019a). Visual inference for the funnel plot in meta-analysis. *Zeitschrift fur Psychologie mit Zeitschrift fur Angewandte Psychologie*, 227(1), 83-89. <https://doi.org/10.1027/2151-2604/a000358>
- Kossmeier, M., Tran, U. S., & Voracek, M. (2019b). *Metaviz: Forest plots, funnel plots, and visual funnel plot inference for meta-analysis* (R package version 0.3.0). Retrieved from <https://CRAN.R-project.org/package=metaviz>
- L'Abbé, K. A., Detsky, A. S., & O'Rourke, K. (1987). Meta-analysis in clinical research. *Annals of Internal Medicine*, 107(2), 224-233. <https://doi.org/10.7326/0003-4819-107-2-224>
- Lewis, S., & Clarke, M. (2001). Forest plots: Trying to see the wood and the trees. *British Medical Journal*, 322, 1479-1480. <https://doi.org/10.1136/bmj.322.7300.1479>
- Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA, USA: Harvard University Press.
- López-López, J. A., Page, M. J., Lipsey, M. W., & Higgins, J. P. T. (2018). Dealing with effect size multiplicity in systematic reviews and meta-analyses. *Research Synthesis Methods*, 9(3), 336-351. <https://doi.org/10.1002/jrsm.1310>
- Manolov, R., Solanas, A., & Leiva, D. (2010). Comparing “visual” effect size indices for single-case designs. *Methodology*, 6(2), 49-58. <https://doi.org/10.1027/1614-2241/a000006>

- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS MEDICINE*, *6*(7), Article e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*(10), 991-996. <https://doi.org/10.1016/j.jclinepi.2007.11.010>
- Raudenbush, S. W., Becker, B. J., & Kalaian, H. (1988). Modeling multivariate effect sizes. *Psychological Bulletin*, *103*(1), 111-120. <https://doi.org/10.1037/0033-2909.103.1.111>
- Schild, A. H., & Voracek, M. (2013). Less is less: A systematic review of graph use in meta-analyses. *Research Synthesis Methods*, *4*(3), 209-219. <https://doi.org/10.1002/jrsm.1076>
- Schild, A. H., & Voracek, M. (2015). Finding your way out of the forest without a trail of bread crumbs: Development and evaluation of two novel displays of forest plots. *Research Synthesis Methods*, *6*(1), 74-86. <https://doi.org/10.1002/jrsm.1125>
- Sterne, J. A., Becker, B. J., & Egger, M. (2005). The funnel plot. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 75-98). Hoboken, NJ, USA: Wiley. <https://doi.org/10.1002/0470870168.ch5>
- Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, *22*(13), 2113-2126. <https://doi.org/10.1002/sim.1461>
- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*, 576-594. <https://doi.org/10.3758/s13428-012-0261-6>
- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods*, *47*, 1274-1294. <https://doi.org/10.3758/s13428-014-0527-2>
- Wickham, H. (2011). ggplot2. *Interdisciplinary Reviews*, *3*(2), 180-185. <https://doi.org/10.1002/wics.147>
- Zhang, Z., Kossmeier, M., Tran, U. S., Voracek, M., & Zhang, H. (2017). Rainforest plots for the presentation of patient-subgroup analysis in clinical trials. *Annals of Translational Medicine*, *5*(24), 485. <https://doi.org/10.21037/atm.2017.10.07>



Methodology is the official journal of the European Association of Methodology (EAM).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.