

A Primer on Systematic Reviews and Meta-Analyses: Part II

Stuart Fisher, BPsychSc, MA, BA¹, Melissa J. Pearson, PhD, AEP¹, Neil A. Smart, PhD, AEP¹

ABSTRACT

In Part I of this primer (Volume 10, Issue 4) we reviewed the basic aspects of pooling data, generating forest plots, interpreting the data, and assessment of heterogeneity. In Part II we examine how to identify publication bias, conduct subanalyses and meta-regression, and introduce the advanced concept of individual patient data meta-analysis. *J Clin Exerc Physiol.* 2022;11(1):27–30.

Keywords: publication bias, selective reporting, meta-regression, heterogeneity

PUBLICATION BIAS

Publication bias within a meta-analysis arises because some articles are never published because the results are not statistically significant and termed ‘negative data sets’, and thus this work cannot be identified, found, or included in a meta-analysis. This results in a bias favoring positive results. A funnel plot is useful for identifying unpublished data sets or publication bias. This is achieved by plotting the precision (the standard error) of a study on the Y axis and the effect size (e.g., mean difference or pre-post change) on the X axis for each study (see Figure 1). If the studies all come from a single population, an approximate bell-shaped pattern should be apparent. Larger and more precise studies will tend to cluster around the population effect size (pre-post change) and have large Y (standard error) values. The smaller studies will be scattered along the X axis and have smaller Y values. This suggests that differences in studies are largely caused by sampling error. For example, sample sizes may be too small or unbalanced between groups for critical variable(s) such as age and sex.

In Figure 1 the funnel plot is essentially symmetrical with 2 studies falling to the right of the funnel (indicated by arrows). This does not indicate significant publication bias as natural variation in only 2 out of 20 (i.e., 10%) of studies is within the expected range. However, note the area to the

bottom left of the plot does not contain any dots (highlighted by the circle) indicating no studies were included with a low or negative effect size (i.e., negative mean difference). These ‘negative’ study results are often considered less interesting and therefore, less likely to be published. If the population effect size is small this will be demonstrated in a funnel plot with a *hole* (i.e., no data) in the distribution around zero and thus only results that are significantly positive or negative are depicted. If the population effect size is large, the plot will appear skewed because of a lack of less precise studies with small effect sizes. Small effects sizes are more likely to occur in studies with small sample sizes because of low statistical power (i.e., too few subjects with data resulting in low likelihood of rejecting a false null hypothesis). This phenomenon (i.e., small study bias) is a general term describing how treatment effect sizes vary between smaller and larger studies (1).

If publication or small study bias occurs then, in turn, biases will exist in the results of meta-analyses and systematic reviews. There are several other possible reasons for small-study bias effects. One is selective reporting by authors of the included studies with only the most favorable outcomes reported, known as *selective outcome reporting bias*. Another possible cause of reduced effects size may be because of the heterogeneity observed between patients in large and small studies (e.g., patients in small sample size

¹Department of Clinical Exercise Physiology, University of New England, Armidale, NSW 2350 Australia

Address for correspondence: Neil A. Smart, School of Science and Technology, University of New England, Armidale, NSW 2350, Australia; e-mail: nsmart2@une.edu.au.

Conflicts of Interest and Source of Funding: None

Copyright © 2022 Clinical Exercise Physiology Association

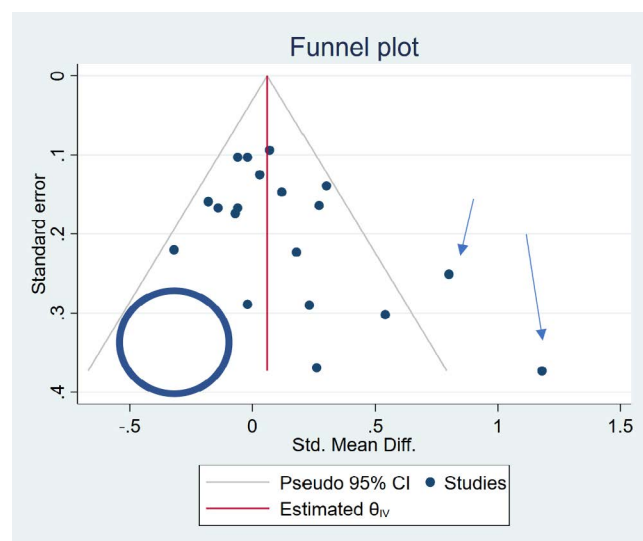


FIGURE 1. Funnel plot of effect size on the X axis and precision on the Y axis.

trials may have been preferentially selected so that a favorable outcome of the intervention is more likely).

If there is a bimodal distribution (i.e., a plot with 2 peaks, rather than just 1 as depicted in Figure 1) within a funnel plot, this may suggest that there are 2 distinct populations. This should lead to an evaluation of whether the study data should have been pooled. Outliers (i.e., a group of studies that are seen isolated, far away from the rest of the funnel plot) suggest possible interaction terms in which an intervention may be very effective, or a variable that may alter the results in certain patient groups. For example, we know that having diabetes has an additional adverse effect on physical fitness in people with heart failure.

Sometimes the effects described may simply be caused by chance. To evaluate this, a useful plot is effect size on the Y axis and date of publication on the X axis (See Figure 2). It is often the case that the earliest publication will have a large effect size, which is required to achieve the first ever publication answering a particular question. Subsequent publications are often achieved with smaller effect sizes. Another explanation for differences in effects with time may be that measurements methods become more accurate over time as technological advancements occur.

PUTTING IT TOGETHER

The data presented in the forest plot below (Figure 3) are a hypothetical representation of the involvement of clinical exercise physiologists in cardiac rehabilitation over the past 20 years. Heart function is quantified by left ventricular ejection fraction (LVEF) in people with heart failure. A higher LVEF is generally considered better. Normal values for LVEF are between 50% and 60% (healthy), and systolic heart failure is diagnosed when LVEF is <40%. While there is good evidence that aerobic exercise training will increase LVEF by a few percentage points in people with systolic heart failure, suppose we desire to know if resistance exercise has the same effect?

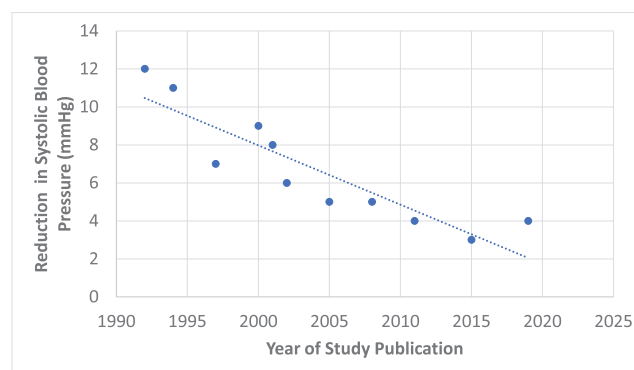


FIGURE 2. The changing effect of exercise on systolic blood pressure with time.

In Figure 3 note that resistance training produced a trend toward (but not statistically significant) an improvement in LVEF versus control (mean difference 2.05%; 95% confidence interval [CI] $-2.94, 7.04$). This is seen by observing that the green diamond touches the black line of 'no effect' informing us that this analysis was not statistically significant. Also note that the 95% CI has both negative and positive values and thus crosses zero, also indicative of a nonsignificant finding. This conclusion is also supported by the P value of 0.42 and, along with most of the diamond falling toward 'favors exercise', is why we prefaced the mean LVEF value with the term 'trend toward' as the P value was >0.05 . We also note that the heterogeneity (how different [closer to 100%] or similar [closer to 0%], the scatter of the data in the included studies are to each other) of $I^2 = 94.7\%$ is very high and suggests these data should not be pooled. So, one may conclude that resistance training does not improve LVEF, but there may be a trend toward improvement. Another point to consider is the reason(s) some studies showed better LVEF change outcomes. For example, the methods to evaluate LVEF may have been more precise in some studies; and measurement precision has changed (i.e., improved) over time. There are a few methods available to investigate these questions and presented in the next two sections.

SUBANALYSES

First, and probably the simplest approach, would be to conduct a subgroup analysis and separate the studies into those before and after a certain year of publication (for instance before and after 1997, when the United States Food and Drug Administration first approved a beta-blocker for treatment of heart failure). The decision to conduct a subgroup analysis can be based on a subjective view of the data or an event, such as a new therapy that has become standard (e.g., beta-blockers). The approach used should always be justified a priori in the meta-analysis methods. A visual inspection of Figure 3 suggests a cut-off before 2004 might reveal important information as all studies prior to 2004, with the exception of Saunders 1991, lie to the left of zero. From this observation 2 new forest plots can be created; one with the studies up to and including year 2003, and the second with

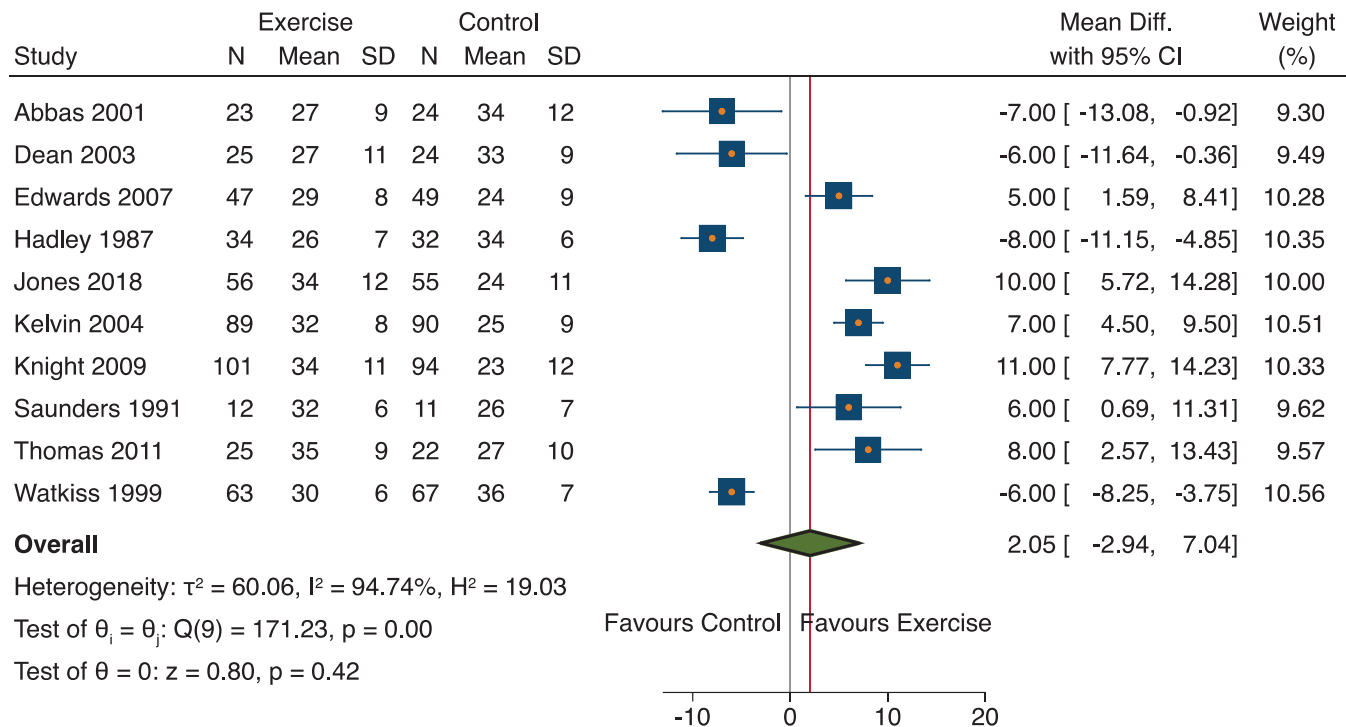


FIGURE 3. Forest plot of all resistance training studies reporting change in left ventricular ejection fraction (LVEF) in people with heart failure. Random-effects model with DerSimonian-Laird weighting method utilized.

the studies from 2004 onward. Next, assessment of the point-estimates of both plots and the 95% CIs should be made to examine if the point-estimates overlap. If there is not overlap it may indicate that the 2 analyses (before vs after and including 2004) have statistically different effect sizes.

META-REGRESSION

Figure 4 was generated from the data in Figure 3. However, instead of generating a forest plot, a meta-regression analysis was conducted using the publication year for each study as a moderator (i.e., explanatory) variable. Not all software packages have meta-regression capability. The regression analysis produced a regression equation: Change in LVEF (%) = $-1432 + 0.71 \times \text{Year of publication}$. Figure 4 shows the plot of this regression equation, and this data illustrate a clear linear relationship between publication year and change in LVEF%. We may consider that the results may not be predicted by publication year and are more likely to be affected by improvements in technology and measurement methods over time.

INDIVIDUAL PATIENT DATA META-ANALYSES

This brings us to what is considered the gold standard approach to meta-analyses: the individual patient data (IPD) meta-analysis. The main difference between a group-level meta-analysis and an IPD is that for the former most data can be obtained from the original manuscripts of the included studies. However, for an IPD the authors of the original studies must provide the original datasets of the deidentified individual patient-level data. Ethical collection of data must

be considered. Some researchers may request data usage agreements before providing their data. Furthermore, institutional/privacy review boards may be a complication if data is acquired from several countries.

It may be an oversimplification to state that an IPD is the *sewing together* of numerous datasets following a systematic literature search that has identified a group of studies that meet predetermined inclusion and exclusion criteria. An extensive summary of how to conduct an IPD can be found in the article by Riley et al., but in essence such an analysis is a 2-stage process (2). An IPD requires researchers to

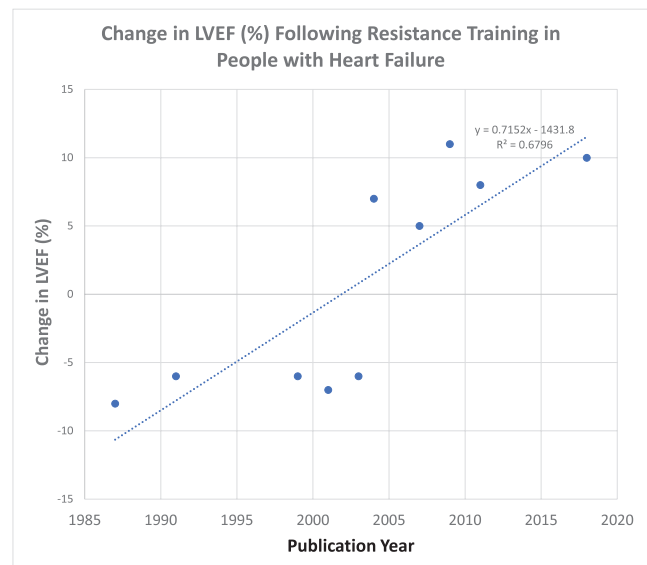


FIGURE 4. Meta-regression plot of change in left ventricular ejection fraction (LVEF, %) vs year of publication.

conduct both a traditional group-level meta-analysis, as described in Part I of this primer series, but also to then conduct a regression analysis similar to that presented in Figure 4. This linear regression process requires the pooling together all of the data as if it came from a single study, while also retaining a study label variable that identifies which of the original studies the data has come from. The group-level and patient-level analyses are then compared for similarity of results.

The primary advantage of conducting an IPD analysis, as opposed to a group-level (or aggregate-level) analysis is that direct relationships between patient characteristics (e.g., age, gender, body mass, etc.) can be established. If these relationships are inferred with group-level data, it can result in unsubstantiated assumptions, known as *ecological fallacy*.

REFERENCES

1. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol*. 2000;53(11):1119–29
2. Jones AP, Riley RD, Williamson PR, Whitehead A. Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clin Trials*. 2009;6(1):16–27

It is possible with group-level analyses to conduct meta-regression on study characteristics; however, these assumptions cannot be extended to patient characteristics. The IPD process usually concludes with a series of subanalyses aimed at determining the effect size for select variables (e.g., age, gender, body mass, number of medications, etc.).

SUMMARY

Meta-analysis is a key tool for medical and health practitioners because it can clarify whether a treatment, or an approach to delivering a treatment, is effective in the presence of conflicting data from different publications. Part II of this primer has shown that additional tools are available such as funnel plots, subanalyses, meta-regression, and IPD analyses to optimize pooled data analyses.