How to Conduct and Interpret Meta-analyses

Monica Martinussen[1,2] and Sabine Kaiser[1]

[1]Regional Centre for Child and Youth Mental Health and Child Welfare, UiT The Arctic University of Norway, Tromsø, Norway

[2]Norwegian Defence Command and Staff College, The Norwegian Defence University College, Oslo, Norway

Abstract

Research questions may be answered with empirical studies or by aggregating existing research. In contrast to single studies, meta-analyses combine the quantitative findings of multiple primary studies to estimate an overall effect size, for example the mean predictive validity of spatial tests used for pilot selection. This aggregation allows researchers to make more reliable conclusions about the size of a specific relationship, and how much it varies between studies. Moderator analyses can be applied to examine the influence of different factors on the estimated effect size, such as whether the predictive validity varies with the type of aircraft (fixed wing or rotary) or sample (ab-initio versus experienced pilots). The aim of this chapter is to present the different steps taken to conduct a meta-analysis and explain different meta-analytic approaches for estimating the mean effect sizes and variation between studies. The chapter also presents different types of software available for both screening studies and meta-analysis calculations, and uses examples from aviation and human factors explaining how the results may be interpreted and used in applied settings.

Keywords: meta-analysis, aviation, synthesising research, effect size, moderator analyses

What is a meta-analysis

Meta-analyses summarise research findings from many studies, for example validation studies in selection research or studies of the effectiveness of different training methods. By pooling and analysing findings from various articles and reports, meta-analyses enable researchers to draw more robust conclusions regarding the magnitude and variability of a specific relationship or group difference. The primary distinction between a systematic review and a meta-analysis lies in their approaches and the type of studies that may be included. While both aim to summarise research, a meta-analysis uses statistical techniques to summarise and compare the results of individual studies, whereas a review does not utilise these statistical techniques. In a meta-analysis, the inclusion of studies with quantitative findings is a prerequisite, whereas a systematic review does not have such a requirement. In some cases, articles may employ both approaches, wherein quantitative findings are subjected to meta-analytic techniques, while studies lacking quantitative data or exhibiting significant heterogeneity are qualitatively described and evaluated.

The popularity of meta-analyses has increased drastically in the last 30 years (Borenstein et al., 2021). In aviation, meta-analyses have been used in selection research to examine the predictive validity of a number of predictors of pilot performance (Hunter & Burke, 1994; Martinussen, 1996), to predict performance based on personality assessments (Campbell et al., 2010), and for predicting military pilot performance (ALMamari & Traynor, 2019) as well as for the selection of air traffic controllers (Dehn & Damitz, 2022; Martinussen et al., 2000). Meta-analysis has also been used to combine other types of studies such as effectiveness studies of Crew Resource Management (CRM) training (O'Connor et al., 2008) or studies estimating the prevalence of neck pain among fighter pilots (Riches et al., 2019). Validation studies typically report correlations between tests and criteria, and effectiveness studies typically examine mean differences between an intervention and a control group,

whereas studies examining the prevalence of a condition will report this as a percentage or proportion. These examples represent different study designs where the findings are analysed and reported differently.

When to do a meta-analysis

A meta-analysis is appropriate to conduct when there are "enough" primary studies with quantitative findings that can be summarised in the respective analyses. It is difficult to set the minimum number of studies or rather effect sizes needed to conduct meaningful analyses. Some would argue that three is the absolute minimum, while others would claim that five should be the minimum. Others, again, would argue that a much higher number is needed. Regardless of which rules are used, the number of studies will determine what kind of conclusions can be drawn based on the meta-analysis, especially with respect to what extent the findings can be generalized.

Which method to use

There are different meta-analytic approaches available to perform the meta-analyses. Traditionally a distinction has been made between the approach developed by Hunter and Schmidt (Hunter & Schmidt, 1990, 2004; Hunter et al., 1982; Schmidt & Hunter, 1977, 2015) and Glass (1976) and later Hedges and Olkin (1985). The approach pioneered by Glass (1976) was initially designed to aggregate findings from psychotherapy research, specifically focusing on differences between groups that received treatment and those that did not. In contrast, Schmidt and Hunter (1977) developed a procedure to assess the generalisability of test validity across various settings. For example, they explored whether intelligence tests consistently predicted work performance irrespective of job type or applicant characteristics. Over time, both research traditions have expanded beyond their original scope. However, within the field of work and organisational psychology, the Hunter and Schmidt method

remains widely favored, while Hedges and Olkin's approach is more commonly employed in clinical and medical research.

How to conduct a meta-analysis

Conducting a meta-analysis takes time and includes several steps which are described in the following and depicted in Figure 1. The process resembles the ordinary research process found in a primary study, which starts with a research question, data collection and analysis, and ends with a research article.

----------Insert Figure 1 about here----------

Figure 1. The different steps in a meta-analysis.

The research question

The initial step in conducting a meta-analysis is to formulate one or more research questions. These questions can range from specific inquiries to broader topics, depending on the researcher's interests and the available studies. Sometimes the authors choose to write a protocol that describes the planned meta-analysis, and also register it in a database such as the International Prospective Register of Systematic Reviews (PROSPERO) for health sciences.

Inclusion and exclusion criteria

Inclusion and exclusion criteria define which studies should be included or excluded from the analyses. They can be outlined before the literature search is conducted. Inclusion and exclusion criteria are important in order to be systematic, transparent, and select the right studies.

Examples of inclusion criteria are English language or the year when the study was published (e.g., after 2000). More specific criteria are the type of design (e.g., an experiment),

or reporting statistics that can be used for the meta-analytic calculations (e.g., frequencies, means and standard deviations, correlations). Exclusion criteria could for example be that a certain group is excluded (e.g., commercial pilots versus general aviation pilots).

The literature search

The starting point for every meta-analysis is a comprehensive and systematic literature search that covers the most relevant databases for the respective research field and appropriate search key terms to identify relevant primary studies. A comprehensive literature search is performed in relevant databases, necessitating the careful selection of appropriate keywords and parameters. Also, gray literature (e.g., reports) is of importance and should be searched for by, for example, contacting relevant authors in the field of interest to acquire unpublished material. Accordingly, the method section should describe which databases were used (one is usually not enough), the search terms (which might have to be adapted from database to database), and how many articles that were identified in total. The literature search is very complex, and it might take several rounds before finding the right databases and search terms. It is an advantage if a librarian can assist in developing the search strategy or at least review it. After the literature search is conducted, the identified studies are reviewed for their in- or exclusion.

The review process

The review process involves the screening of the titles and abstracts of the identified studies to determine their potential relevance. Next the selected articles are thoroughly screened in full text to ascertain their eligibility for inclusion in the meta-analysis. Also, there are different tools that can help in conducting the screening of the identified articles (e.g., Rayyan, Covidence). While some meta-analyses solely rely on published articles, others include a wider range of sources, such as reports, conference presentations, and dissertations. Using only published articles ensures a certain level of quality through the peer-review and

editorial process, but incorporating grey literature can help mitigate publication bias and provide a more comprehensive overview of the research field. The review process includes a screening of the identified articles from the literature search based on the defined inclusion and exclusion criteria. The review process usually starts with the title and abstract screening to determine whether an article seems relevant for the meta-analysis or not, by applying the pre-defined inclusion and exclusion criteria. If the literature search identified a large number of articles, one can consider starting with a title screening first and continuing with an abstract screening in a second step. Those articles that are not excluded after the title and abstract screening are usually read in full-text to inform a final decision about their in- or exclusion. In this final step of the review process, it is also important to document the reason for exclusion for every article and report them in the manuscript for example in a flow diagram.

A flow diagram can be a useful way to depict the different stages of the literature search and review process, such as how many articles were identified through the original search, how many were excluded after title and abstract screening, what were the reasons for exclusion when the full-text was examined, and last but not least, what was the final number of included studies. One can either create a new flow diagram or use a template from, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).

Coding of articles

Once the final sample of studies is determined, the subsequent step entails coding all relevant information, including study characteristics, sample sizes, and findings (effect sizes). This can either be entered in a separate file or directly into the software that is going to be used to conduct the meta-analyses. The extracted information can for example include the first author (for identification purposes), publication year of the study, sample size, the relevant statistics, and possible moderators (e.g., type of aircraft [fixed wing or rotary] or sample [ab-initio versus experienced pilots]). Involving two or more individuals in both the

study screening and coding processes, can enhance the quality of the literature selection and improve coding accuracy. Furthermore, the consistency between the coders can be calculated using different types of statistics (e.g., intraclass correlation coefficient or Cohens's Kappa).

Quality assessment of included articles

Often the included studies are assessed based on their quality regarding different criteria such as sampling, study design, method, and statistical analyses. This information can be used to describe the studies in a table or to examine differences in effect sizes based on the individual studies quality score. There are different quality assessment tools (i.e., checklists) depending on the study design you are interested in examining (e.g., RoB 2, which is Cochrane risk of bias tool in randomized trials). Possible quality indicators are for example questions like "Is the applied method appropriate to answer the research question?" or "Is the sample size large enough to conduct the respective analysis?". The quality assessment thus indicates the quality of the reporting in the primary study as well as the quality of the included studies (Guyatt et al., 2008).

The meta-analytic calculations

The next step encompasses the actual meta-analysis calculations, where the synthesized effect sizes are computed, and variation between studies is estimated and examined.

What is an effect size

Research studies investigating the same research question often exhibit variations in their design and measurement approaches. For instance, when examining pilot selection, studies may employ different types of tests and different criteria to assess pilot performance. Similarly, studies investigating the effectiveness of CRM training may adopt either a control-group design or a simple pretest-posttest design. Moreover, the outcomes measured in these studies can vary, ranging from instructor ratings to self-reported satisfaction. To synthesize

findings from studies with diverse measures and designs, it is crucial to identify commonalities that enable the integration of results across studies.

Meta-analyses rely on effect sizes derived from each study as their fundamental building blocks. Effect sizes are statistical measures that quantify the magnitude or strength of a finding. They may encompass standardised measures of co-variation between variables or differences observed between groups. Co-variation is often expressed through correlation coefficients (*r*), which indicate the strength and direction of the association between two variables. Correlation coefficients range from -1 to +1, where a value close to 0 signifies no association. A correlation coefficient of +1 or -1 indicates a perfect correlation, with a positive correlation implying that higher values on one variable align with higher values on the other, and a negative correlation indicating that higher values on one variable correspond to lower values on the other. The closer the correlation coefficient is to -1 or +1, the stronger the association. To aid the interpretation and communication of findings, Cohen (1988) coined various labels for different effect sizes in his book on statistical power calculations. Table 1 presents some guidelines for the interpretation of effect sizes and Figure 2 illustrates an example of a large correlation.

Another frequently used effect size is the standardized mean difference, typically utilised when studies report means for different groups. This effect size quantifies the difference between means (*M*) divided by a standard deviation (*SD*), often the pooled standard deviation based on both groups. One example from this family of effect sizes is denoted Hedges' *g* (= M_1-M_2/SD_{pooled}). Alternatively, the *SD* from the control group can be employed to compute the effect size. The standardised mean difference represents the difference between two groups in terms of *SD*s. In validation studies, where the focus lies on the correlation between test results and a measure of pilot performance, the common effect size is a correlation coefficient, which can be combined across studies. When evaluating CRM

training and comparing different groups based on an outcome, the most relevant effect size is the standardised mean difference.

----------Insert Table 1 about here----------

----------Insert Figure 2 about here----------

Figure 2. Scatterplot of a large correlation.

Statistical artifacts

There are several potential reasons for variations in results across studies. Firstly, sampling error contributes to random fluctuations due to the examination of samples rather than the entire population. The smaller the sample sizes the greater the sampling error. Secondly, true variation among studies may arise from specific aspects of the research design. For instance, certain interventions might prove more effective than others, or certain tests may serve as better predictors of pilot performance.

The third factor involves statistical artifacts or methodological issues that can influence the magnitude of the observed effects. In validation studies, a common problem is conducting research within a highly restricted sample (e.g., limited to accepted applicants), which reduces variation in test scores and subsequently lowers the observed correlation between the test and the criterion. Another artifact occurs when the measures used exhibit poor reliability, leading to a decrease in the observed correlation. Additionally, using a dichotomous criterion (e.g., pass/fail) instead of a more continuous measure, such as performance ratings, can introduce another artifact. These artifacts are cumulative, resulting in lower observed validation coefficients compared to a situation without range restriction, with perfectly measured variables and a continuous performance measure.

It is possible to statistically adjust the observed correlations to account for these artifacts, either addressing a specific artifact (e.g., range restriction) or considering the

combined effect of many artifacts. If the presence and magnitude of these biases differ between studies, they can contribute to the observed variation. Schmidt and Hunter (2015) describe some of these artifacts specifically in relation to validation studies, while others apply to other types of studies and effect sizes. For instance, poor reliability always diminishes the observed effect. Correcting for range restriction is not a straightforward process and depends on various factors, including the selection ratio and the number of tests used, as well as their intercorrelations. Correction for reliability relies on the reported reliability coefficient, which may or may not be reported in the primary studies. Correction for dichotomisation depends on where the split is, that is how uneven the group sizes are. The further away the groups are from a 50/50 split, the greater the correction.

Mean effect size

There are two primary steps involved in conducting a meta-analysis. The first step is to calculate the mean effect size, typically a weighted mean. In the Hunter and Schmidt method (2004), for instance, the effect size is weighted based on the sample size. Ideally, the effect sizes should be adjusted or corrected for any statistical artifacts before computing the mean effect size. In some meta-analyses, both corrected and uncorrected mean effect sizes are reported, and it is important to specify the type of corrections employed. Table 2 provides a meta-analysis example for correlations between test and criterion on simulated data where the data includes seven studies with individual correlations ranging from .06 to .20, as well as correlations corrected for criterion reliability. The mean sample size weighted correlation is .14, and corrected for lack of criterion reliability, slightly higher (.16) when using the software Metados (Martinussen & Bjørnstad, 1999).

Alternative meta-analytic approaches may utilize weighting methods slightly different than direct sample size (as suggested by Hedges and Olkin and implemented in the software “Comprehensive Meta-Analysis version 4”, see for example; Borenstein et al., 2022). The

approach taken in a meta-analysis depends on the specific research question. In some cases, it may be adequate to calculate an overall effect size. However, it is more common to compute mean effect sizes for different subgroups in addition to the overall effect.

How to study variation between studies

When synthesizing findings from numerous studies, the process typically involves two steps. Firstly, calculating the mean correlation provides an overall estimate of the effect. Secondly, the focus shifts to examining the variation among the studies. This entails determining whether the observed effect such as correlation coefficients exhibit greater variation than expected and assessing the extent to which the observed variation is influenced by sampling error.

In the approach proposed by Schmidt and Hunter (2015), the variation is estimated by comparing the observed variation with the variation attributed to sampling error. If the observed variation is small or predominantly caused by sampling error, then the mean effect size serves as a reliable estimate of the true effect. However, if substantial variation exists, the subsequent step involves investigating potential factors, known as moderators, that might explain the differences in variation, such as variations among sub-groups. In the example in Table 2, the estimated population variance was 0.001 and the corresponding standard deviation was 0.032. This estimated standard deviation may be used to calculate an interval (credibility interval) where the true effect is likely to be found (Whitener, 1990). Moreover, the percentage of explained variance was 64% indicating that there is some variance left to be explained. Hunter and Schmidt (2004) have suggested a 75% rule, meaning that if 75% or more of the variance is explained by sampling error, there is no need to examine moderators.

The most basic form of meta-analysis is referred to as a "bare-bones" meta-analysis (Schmidt & Hunter, 2015), where effect sizes are not adjusted for statistical artifacts such as reliability or range restriction. Alternatively, corrections for statistical artifacts can be

performed individually or by using compiled information from studies included in the meta-analysis (using artifact distributions) to improve the estimates of both the mean effect sizes as well as how the true variation between studies is estimated.

An alternative approach to estimating the variation between studies is to conduct a significance test. In this case, the null hypothesis posits that there is no significant variation among the studies. If the test yields a significant result, the null hypothesis is rejected, indicating the presence of true variation between studies. Consequently, the subsequent step involves investigating potential moderators in a similar manner as previously outlined. The examination of variation through significance testing was initially described by Hedges and Olkin, and later expanded upon by Borenstein et al. (2009, 2021). The weighting scheme employed in these calculations is contingent upon the chosen model, namely the fixed effect model or the random effects model.

In the fixed effect model, it is assumed that the included studies can be regarded as replications. However, this assumption may not hold true in most meta-analyses, as there are typically variations across studies in terms of study design, participant characteristics, and measurement methodology. On the other hand, the random effects model considers that effect sizes in the population are not constant and that factors beyond sampling error can contribute to the observed variation in effect sizes.

In the random effects model, studies are weighted by the inverse of the variance components, which encompass both random variation (sampling error) and variation between studies. Consequently, this results in more equal weights being assigned to the studies compared to the fixed effect model where the weighting is more similar to the one employed by Schmidt and Hunter (2015).

Publication bias

Publication bias refers to the possible overrepresentation of significant findings in published studies, and that non-significant findings are less likely to be published. If a meta-analysis is based on published studies only, it may overestimate the true effects. It is of interest to examine the extent to which the findings from unpublished studies may influence the meta-analytic results and therewith the overall conclusion (Lipsey & Wilson, 2001). This can, for example, be done by comparing the mean effect sizes between published and unpublished studies but there are also other methods that can be used (for more information see, Cooper et al., 2009).

Presenting the findings

In the final step, the results of the meta-analysis are presented to a broader audience, and hopefully provide an answer to the initial research question. This is typically done in the form of a research article or conference presentation.

Which software to use

There is different types of software that can be used to conduct meta-analyses; more general ones like R, SAS, or the newer versions of SPSS. The R software offers a package for performing meta-analysis calculations according to the Hunter and Schmidt method (1990). There is also more specialized software for conducting meta-analyses such as Comprehensive Meta-Analysis (CMA) (Borenstein et al., 2022) or Review Manager 5 (RevMan 5) (The Cochrane Collaboration, 2020). CMA uses the meta-analytic approach by Borenstein et al. (2021), while RevMan 5 is Cochrane's software and widely used.

How to interpret meta-analyses

Interpreting meta-analyses includes an assessment of the different steps that are described in this chapter and that need to be completed in order to conduct a meta-analysis. It

includes assessing the quality of the meta-analysis, possible biases, and interpreting effect sizes and other findings like from the moderator analyses. To assess a meta-analysis quality, it is important to use a systematic and transparent approach throughout the study. The literature search and the review process should be described in detail and inclusion and exclusion criteria that were applied to select the studies should be stated explicitly.

Summary and conclusion

Well conducted meta-analyses can be an important contribution to the research field. They do not just summarize statistical findings that are more reliable than the results from single studies, but also generate new knowledge beyond this. Meta-analyses provide an insight into the status quo of a research field, can identify the need for further research into a specific direction, and provide answers to research questions that cannot be answered by conducting primary studies alone. The number and quality of the primary studies will determine the overall quality of the meta-analysis, as well as the conclusions that can be drawn. As always, the more data the better, or rather the more studies the better the meta-analysis.

References

ALMamari, K., & Traynor, A. (2019). Multiple test batteries as predictors for pilot performance: A meta-analytic investigation. *International Journal of Selection and Assessment*, *27*(4), 337-356. https://doi.org/10.1111/ijsa.12258

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (Vol. 2). John Wiley & Sons.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2022). *Comprehensive Meta-Analysis* (Version 4) [Computer software]. Biostat. https://www.meta-analysis.com/

Campbell, J. S., Castaneda, M., & Pulos, S. (2010). Meta-analysis of personality assessments as predictors of military aviation training success. *The International Journal of Aviation Psychology*, *20*(1), 92-109. https://doi.org/10.1080/10508410903415872

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Vol. 2). Lawrence Erlbaum Associates.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation. http://www.jstor.org/stable/10.7758/9781610441384

Dehn, D. M., & Damitz, M. (2022). Do personality traits predict training performance? *Aviation Psychology and Applied Human Factors*, *12*(2), 72-83. https://doi.org/10.1027/2192-0923/a000222

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3-8. https://doi.org/10.2307/1174772

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., Schünemann, H. J., & GRADE Working Group (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ (Clinical research ed.), 336(7650), 924–926. https://doi.org/10.1136/bmj.39489.470347.AD

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Hunter, D. R., & Burke, E. F. (1994). Predicting aircraft pilot-training success: A meta-analysis of published research. *The International Journal of Aviation Psychology*, *4*(4), 297-313. https://doi.org/10.1207/s15327108ijap0404_1

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage Publications.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Sage Publications.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Sage Publications.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE publications, Inc.

Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. *The International Journal of Aviation Psychology*, *6*(1), 1-20. https://doi.org/10.1207/s15327108ijap0601_1

Martinussen, M., & Bjørnstad, J. F. (1999). Meta-Analysis Calculations Based on Independent and Nonindependent Cases. *Educational and Psychological Measurement*, *59*(6), 928-950. https://doi.org/10.1177/00131649921970260

Martinussen, M., Jenssen, M., & Joner, A. (2000). Selection of air traffic controllers: Some preliminary findings from a meta-analysis of validation studies. Proceedings of the

24th European Association for Aviation Psychology (EAAP) conference, Crieff, Scotland, September 4-8, 2000.

O'Connor, P., Campbell, J., Newon, J., Melton, J., Salas, E., & Wilson, K. A. (2008). Crew resource management training effectiveness: A meta-analysis and some critical needs. *The International Journal of Aviation Psychology*, *18*(4), 353-368. https://doi.org/10.1080/10508410802347044

Riches, A., Spratford, W. A., Witchalls, J., & Newman, P. (2019). A systematic review and meta-analysis about the prevalence of neck pain in fast jet pilots. *Aerospace medicine and human performance*, *90*(10), 882-890. https://doi.org/10.3357/AMHP.5360.2019

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, *62*(5), 529. https://doi.org/10.1037/0021-9010.62.5.529

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Sage Publications. https://doi.org/10.4135/9781483398105

Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology*, *75*(3), 315–321. https://doi.org/10.1037/0021-9010.75.3.315

The Cochrane Collaboration. (2020). *Review Manager* (Version 5.4) [Computer software]. Cochrane. https://training.cochrane.org/online-learning/core-software/revman

Table 1

Interpretation of effect sizes

	Effect size	
Effect size labels	Correlation	Hedges' *g*
Small effect	*r* = .10	*g* = 0.20
Medium effect	*r* = .30	*g* = 0.50
Large effect	*r* = .50	*g* = 0.80

Note. Based on Cohen (1988).

Table 2

Meta-analysis example for correlations (r between test and criterion)

Study	*n*	*r*	r_{yy}	r_c
1	200	.20	.60	.26
2	150	.20	.70	.24
3	1500	.10	.80	.11
4	200	.15	.60	.19
5	100	.13	.70	.16
6	1200	.20	.80	.22
7	400	.06	.70	.07

Note. *N* = sample size; *r* = correlation; r_{yy} = criterion reliability; r_c = correlation corrected for criterion reliability.

Formulas

The mean sample-size weighted correlation: $\bar{r} = \frac{\sum[N_i r_i]}{\sum N_i}$

The population variance: $\sigma_\rho^2 = \sigma_o^2 - \sigma_e^2$

Results

Meta-analysis calculations (bare-bones) using the Hunter and Schmidt method:

The number of studies = 7

The total number of participants = 3750 (mean number per study = 535)

The mean correlation (sample size weighted) = .14

The mean correlation unweighted = .15

The estimated population variance = 0.001 and standard deviation = 0.032

Lower credibility value (90%): .10

Percentage of observed variance accounted for by sampling error = 64%

Corrected and *N*-weighted mean correlation (for criterion reliability) = .16

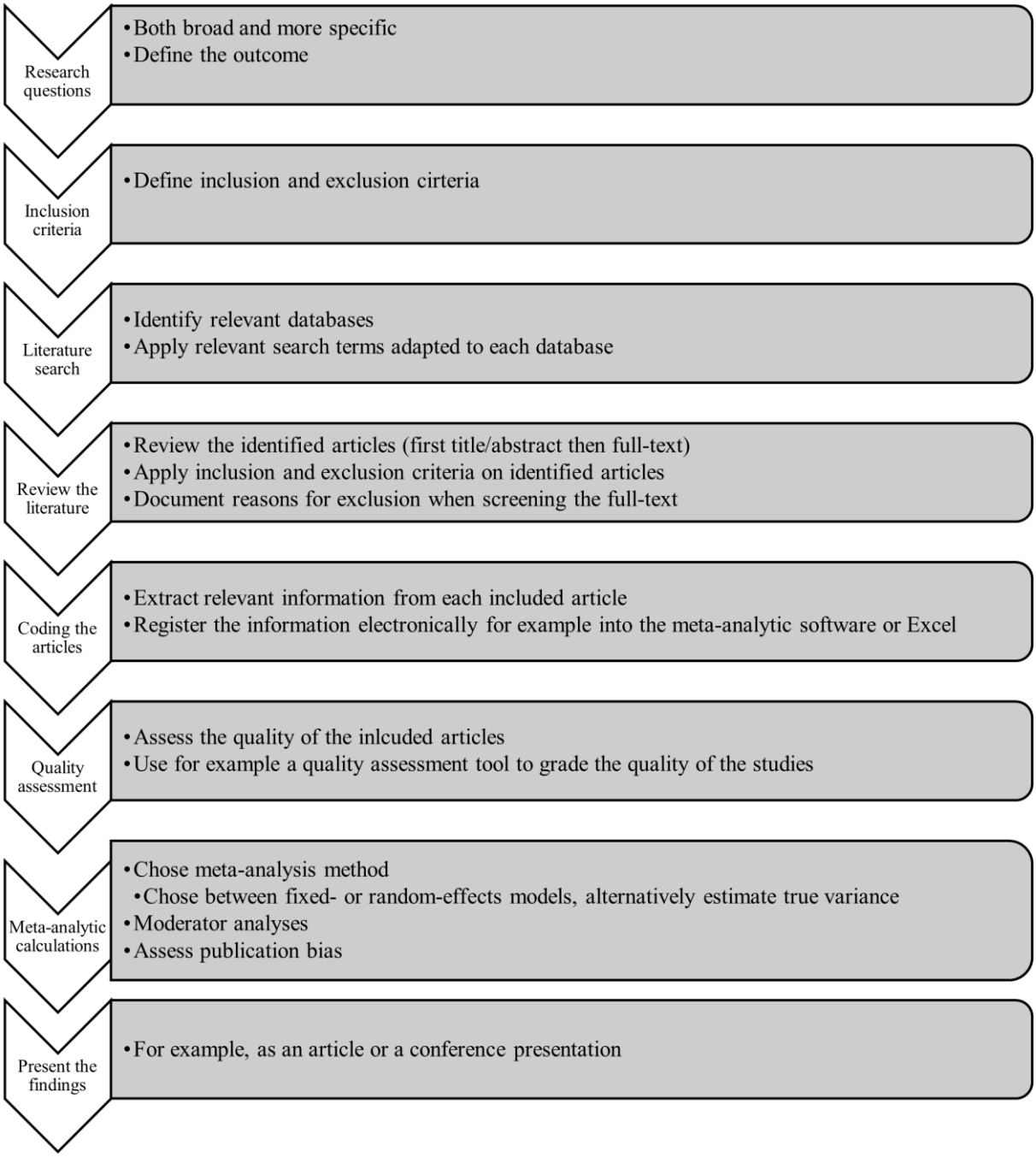

Figure 1. The different steps in a meta-analysis.

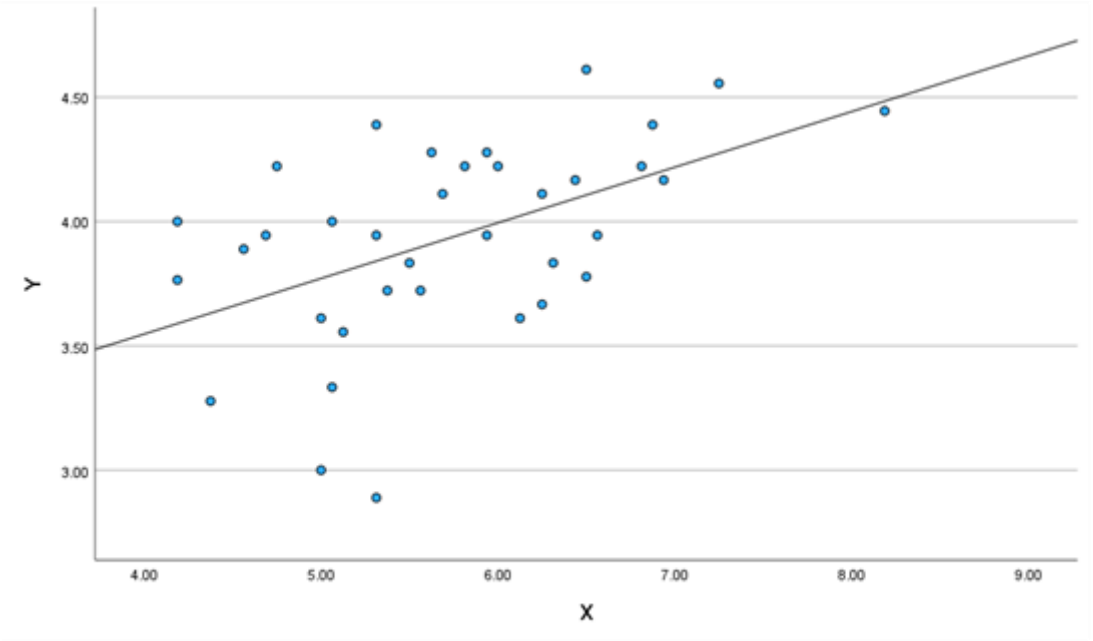

Figure 2. Scatterplot of a large correlation.