KU LEUVEN

FACULTY OF SCIENCE

Transforming Effect Size Measures: A Commented Review

Meta Analysis

Daniel Gerardo GIL SANCHEZ
daniel.gilsanchez@student.kuleuven.be

Prof. Wim Van den Noortgate

Academic year 2017-2018

Contents

1 Introduction

A brief and concise definition of meta-analysis was given by Glass (1976) in his popular paper "Primary, Secondary and Meta-Analysis for research". He defined *Primary analysis* as the original data analysis in a research study, *Secondary analysis* as the re-analysis of data using better statistical techniques or with the purpose of answering new questions and *Meta-analysis* as the **analysis of analyses**. Actually, he used this term "to refer to the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings". This was a clever idea that in the beginning was not accepted by the scientific community but over the time it has gained importance as a way of summarize results in health and social sciences.

For a better understanding of how meta-analysis is performed, Lipsey and Wilson (2001) made a simile between this kind of analysis and survey analysis, they expressed that "meta-analysis can be understood as a form of survey research in which research reports, rather than people, are surveyed". Additionally, they suggested that a meta-analysis can be considered as well done, when it is conducted as a structured research technique where each step is documented and opened to scrutiny.

It is therefore important to notice that in order to get a successful result, all decisions made through the process of analysis should be documented along with their own motivation. This involves the process of expressing the research question, the selection of the documents, the process of coding effect sizes, the analysis of the data collected and the results obtained.

The process of coding effect sizes is of special focus in this document because the quality of the posterior analysis, as well as, the final conclusions depend on the quality of these measures. To *standardize* the definition of effect size, the definition given by Lipsey and Wilson (2001) is taken as a reference from now on, which is "a statistic that embody information about either the direction or the magnitude of quantitative research findings, or both". Now, the word *standardized* needs special attention because in most cases the effect sizes are a standardized version of a statistic measured in primary studies. This standardization is usually done when the same construct can be operationalized in different ways, e.g., scores on different mathematics achievement test. So in order to be comparable they need to be *standardized*.

Different effect size measures exist in the literature, but only one can be used in a specific meta-analysis. Proportions, means, difference of means, odds-ratios and correlation coefficients are the effect sizes most often used. In fact, it can be considered that the most important are the Standardized Mean Difference (SMD), the odds-ratio and correlation coefficient, because they are suggested by Campbell Collaboration, one of the biggest organizations responsible for prepare, maintain and disseminate systematic reviews and meta-analyses (see Polanin and Snilstveit, 2016).

According to Borenstein et al. (2009), there is no problem in combining results of individual studies when the effect sizes have the same meaning in all measures, e.g., studies reported the difference of blood pressure between treatment and control group.

However, when all studies selected to be part of the meta-analysis achieve the fact that are answering the same scientific question, but they are using different methods to measure the construct of interest, it is not easy to merge this information together. For instance, some studies reported a difference in means, used to compute the SMD. Other studies reported the difference in proportions, used to compute odds-ratios. And others reported a correlation. If this is the case, which usually is, it is necessary to transform these effect sizes to a common index so the posterior analysis can be performed.

The purpose of this document is to enlist different transformations used on effect sizes and explain which are the implicit assumptions made when selecting one in specific. In this way, the reader can make an informed decision in the use of particular formulas and not base its decision only on the effect size calculators available on different websites.

In the following, several formulas are presented to transform effect sizes. Then, a discussion about the use and misuse of these formulas is considered.

2 Transformations

The idea of being able to use all kind of information coming from different sources is very appealing for Meta-Analysts. After defining the articles that are going to be used in the analysis and evaluating all kinds of possible biases that can be introduced, it is time to decide which effect size to use. Borenstein et al. (2009) described a mechanism to incorporate multiple kinds of data in the same meta-analysis: "First, each study is used to compute an effect size and variance of its native index, the log odds ratio for binary data, d for continuous data, and r for correlational data. Then, we convert all of these indices to a common index, which would be either the log odds ratio, d, or r".

It is well known that the formulas to transform one effect size to another are everywhere and the calculation is straightforward (see Borenstein et al., 2009; Polanin and Snilstveit, 2016; Lenhard and Lenhard, 2016). However, almost nowhere it is explained where these formulas come from, what assumptions are made, or whether there is a direct relationship between the effect size and its transformation or is just an approximation. In this section, several transformations are mentioned including the popular ones that can be found online. The notation used for each effect size is as follows:

- d for the standardized mean difference.
- OR for Odds-ration and LOR for the natural logarithm of the odds-ratio.
- r for the correlation coefficient

2.1 From binary to continuous outcomes

Generally, odds-ratio, risk ratio (RR) and risk difference (RD) are statistics used to summarize binary outcomes. According to Sanchez-Meca et al. (2003), they can be easily computed from a "2x2 contingency table, where two groups are crossed with two outcomes, giving four possible cell frequencies" (see Table 1). The RD is just the difference between success (or failure) proportions on both groups, $p_{g2} - p_{g1}$, where $p_{g1} = A/N_{g1}$ and $p_{g2} = B/N_{g2}$. The RR is just the ratio between the two proportions, p_{g2}/p_{g1}. And the OR, the relative odds that one group has more successes than the other, is computed as $p_{g2}(1 - p_{g1})/p_{g1}(1 - p_{g2})$.

Outcome	**Group 1**	**Group 2**	**Total**
Success	A	B	N_s
Failure	C	D	N_f
Total	N_{g1}	N_{g2}	N

Table 1: 2x2 Contingency table. Two groups and Binary outcome

The first transformation from OR to standardized mean difference, d, was originally proposed by Hasselblad and Hedges (1995). They defined the following transformation in terms of sensitivity ($S_n = A/(A + C)$) and specificity ($S_p = D/(D + B)$), in case of information is available in the paper, making clear that the LOR is the sum of the logits of these two measures:

$$d = \frac{\sqrt{3}}{\pi} \{log\,[S_n/(1 - S_n)] + log\,[S_p/(1 - S_p)]\} = \frac{\sqrt{3}}{\pi} LOR \tag{1}$$

Where $\pi = 3.14159$, S_n and S_p as before.

According to Hasselblad and Hedges (1995), there are different ways to compute an estimate of the variance of d. They proposed a simple relative way to compute it, derived from an approximation using the delta method:

$$Var(d) = \frac{3}{\pi^2} \left[1/A + 1/B + 1/C + 1/D\right] \tag{2}$$

These formulas were built under two key assumptions: That each group has an underlying logistic distribution (which is similar in shape to normal distribution) and that both groups have equal variances. They said that a violation of either assumption has consequences for the properties of this measure.

A similar transformation was proposed by Cox and Snell (1989), where instead of multiplying the LOR by a constant, it is divided by 1.65:

$$d = LOR/1.65 \tag{3}$$

And its sample variance is computed as:

$$Var(d) = 0.367 \left[1/A + 1/B + 1/C + 1/D\right] \tag{4}$$

They also assumed that each group follows a logistic distribution.

However, most studies rely on the assumption that their populations follow a normal distribution instead of a logistic distribution, and so the preceding transformations may yield not desirable results.

For this reason, Glass et al. (1981) proposed the probit transformation that works well under normality assumptions. They converted the success (or failure) proportion (p_{g1} and p_{g2}), via the inverse of the standard normal distribution function, to $z_{g1} = \Phi^{-1}(p_{g1})$ and $z_{g2} = \Phi^{-1}(p_{g2})$. Then, they derived the difference of these new measures:

$$d = z_{g2} - z_{g1} \tag{5}$$

The sampling variance of this measure is:

$$Var(d) = \left[\frac{2\pi p_{g1}(1-p_{g1})exp(z_{g1}^2)}{N_{g1}} + \frac{2\pi p_{g2}(1-p_{g2})exp(z_{g2}^2)}{N_{g2}}\right] \tag{6}$$

They said that assuming normal distributions in both groups makes d an unbiased estimator of the standardized mean difference.

Other formulas to transform binary outcomes to standardized mean difference can be found in Sanchez-Meca et al. (2003). They presented seven different effect sizes measures to estimate the population standardized mean difference and compared them through Monte Carlo simulations.

2.2 From continuous to binary outcomes

In a similar fashion as the previous section, it is relatively easy to transform continuous outcomes to binary outcomes. For instance, to convert the standardized mean difference, d, to a log odds-ratio, Borenstein et al. (2009) used the definition given by Hasselblad and Hedges (1995) and presented this transformation:

$$LOR = \frac{\pi}{\sqrt{3}}d \tag{7}$$

and its sampling variance as

$$Var(LOR) = \frac{\pi^2}{3}var(d) \tag{8}$$

In the same way, the inverse of the transformation proposed by Cox and Snell (1989) can be used:

$$LOR = 1.65d \tag{9}$$

And its sample variance is:

$$Var(LOR) = 1.65\ var(d) \tag{10}$$

It is important to notice that in these transformations it is assumed that both groups follows a logistic distribution with the same variance, as a consequence of their definitions.

However, as it was mentioned before in the majority of studies is assumed normality. Whitehead (2002) made this clear in her book and propose another way to make this transformation when the outcome is continuous but it is necessary to transform it to a binary outcome, as odd-ratios.

Usually in primary studies are reported the number of people, the sample mean and the standard deviation for each group. Assuming that in each group the outcome measure (Y_{g1} and Y_{g2}) follows a normal distribution, it is possible to define a cut-point A so that $p_{g1} = P(Y_{g1} > A)$ and $p_{g2} = P(Y_{g2} > A)$, where p_{g1} and p_{g2} are the proportion of successes (or failures) in each group. Now, to estimate this proportion, she proposed that $p_{g1} = 1 - \Phi(A_{g1})$, where $A_{g1} = (A - \bar{y}_{g1}/s_{g1})$ and Φ is the standard normal distribution function. The estimate of p_{g2} is similarly defined.

So the estimate of the LOR is given by:

$$LOR = log\left[\frac{\Phi(A_{g2})(1-\Phi(A_{g1}))}{\Phi(A_{g1})(1-\Phi(A_{g2}))}\right] \tag{11}$$

And its sample variance, obtained by the delta method, is:

$$\begin{aligned} var(LOR) = & \frac{[\phi(A_{g1})]^2\left(1/N_{g1} + A_{g1}^2/2N\right)}{[\phi(A_{g1})(1-\phi(A_{g1}))]^2} + \frac{[\phi(A_{g2})]^2\left(1/N_{g2} + A_{g2}^2/2N\right)}{[\phi(A_{g2})(1-\phi(A_{g2}))]^2} \\ & - \frac{A_{g1}A_{g2}\phi(A_{g1})\phi(A_{g2})}{N\phi(A_{g1})(1-\phi(A_{g1}))\phi(A_{g2})(1-\phi(A_{g2}))} \end{aligned} \tag{12}$$

Where N_{g1} N_{g2} and N are defined in Table 1, and ϕ is the standard normal density function.

A similar result, assuming normality, can be found in Suissa (1991). In this article, the statistic is developed in a similar way but the computation of the variance is different. Da Costa et al. (2012) enlisted different methods to transform continuous outcomes into OR and made comparisons to determine the performance of each method.

2.3 From correlations to continuous outcomes

According to Borenstein et al. (2009), a simple way to convert a correlation to a standardized mean difference is by using the definition given by Friedman (1968):

$$d = \frac{2r}{\sqrt{1-r^2}} \tag{13}$$

And its sampling variance is:

$$Var(d) = \frac{4Var(r)}{(1-r^2)^3} \tag{14}$$

When this formula is used, it is assumed that the continuous data used to calculate r has a bivariate normal distribution and that the two groups are created by dichotomizing one of the two variables (Borenstein et al., 2009). It is also assumed that the sample size of each group is equal and both groups have the same variance (Friedman, 1968).

2.4 From continuous outcomes to correlations

The first transformation from standardized mean difference to correlation coefficient, was originally proposed by Rosenthal (1984):

$$r = \frac{d}{\sqrt{d^2 + \frac{1}{pq}}} \tag{15}$$

Where p is the proportion of the total population that is in the first of the two groups being compared and $q = (1 - p)$ is the proportion of the total population in the second group. He said that in most experimental cases it is assumed that the sample size in each group is the same, so this equation can be simplified to:

$$r = \frac{d}{\sqrt{d^2 + 4}} \tag{16}$$

However, if the groups being compared are unequal in size, a generalization provided by Aaron et al. (1998) can be used:

$$r = \frac{d}{\sqrt{d^2 + \frac{N^2}{N_{g1}N_{g2}}}} \tag{17}$$

Where N_{g1} and N_{g2} are the sample size for each group and N is total sample size ($N = N_{g1} + N_{g2}$). If the reader is interested in see the development of this formula and its relationship with the equality given by Rosenthal (1984), in the document done by Aaron et al. (1998) can be found all details.

The associated sampling variance is:

$$Var(r) = \frac{(\frac{N^2}{N_{g1}N_{g2}})^2 Var(d)}{(d^2 + \frac{N^2}{N_{g1}N_{g2}})^3} \tag{18}$$

According to Borenstein et al. (2009) when this formula is applied, it is assumed that a continuous variable was dichotomized to create the groups.

2.5 From binary outcomes to correlations

According to Bonett (2007), "the problem of estimating the correlation between two continuous variables using the information from a 2x2 table is one of the oldest problems in statistics, and involves the computation of a tetrachoric correlation". There are two kinds of approximations that can be used depending on whether there is sufficient information in the paper or not. In the case of a study that only reports an OR but no additional information to compute a 2x2 contingency table, an approximation suggested by Pearson (1900) can be used:

$$r = cos\left[\frac{\pi}{1 + OR^{1/2}}\right] \tag{19}$$

On the other hand, when more information is reported in a study, a more accurate approximation can be computed. According to Bonett (2007), when a 2x2 contingency table can be retrieved, the following formula may be more accurate that Pearson (1900) approximation:

$$r = cos\left[\frac{\pi}{1+OR^c}\right] \tag{20}$$

Where $c = \{1 - |p_s - p_{g1}|/5 - (1/2 - p_{min})^2\}/2$, $p_s = A/N_s$, $p_{g1} = A/N_{g1}$ and $p_{min} = min\{p_s, p_{g1}\}$ using the notation of Table 1.

And its sampling variance is (see Bonett and Price, 2005):

$$Var(r) = k^2\left\{(A+.5)^{-1} + (B+.5)^{-1} + (C+.5)^{-1} + (D+.5)^{-1}\right\} \tag{21}$$

Where $k = (\pi\ c\ OR^c)\ sin\{\pi/(1+OR^c)\}\ /\ (1+OR^c)^2$.

It is important to notice that the later formula is a generalization of the equation given by Pearson (1900), when $p_s \approx p_{g1}$. If the reader is interested in see these approximations with more detail, in the article of Bonett and Price (2005) can be found all details.

2.6 From correlations to binary outcomes

There is not a specific formula to make this transformation. What is usually done, is to transform correlations to standardized mean differences, using formula 13, and then transform this measure to OR (see Polanin and Snilstveit, 2016). It is crucial to be careful with the assumptions made when using this approach.

3 Comments

According to Lipsey and Wilson (2001), one of the main disadvantages of Meta-Analysis is the amount of effort and expertise it requires. He is referring to the fact that every step done, needs to be documented along with the main motivation to do so. For instance, if a Meta-analyst decided to use the standardized mean difference as the unique effect size, it should be explicitly explained why this is the final choice. It might be because of the majority of studies used this effect size or just because of convenience. The more clear the process is, the more transparent and useful the analysis can be considered.

The quality of the results obtained after analyzing the data is based on the quality of the data. If the data is not consistent, the analysis will not be consistent. If the data are biased in the coding process, the results will be biased as well. This is the main reason to carefully select a transformation in the process of combining studies. It is not just matter of finding an online effect size converter, it is about the importance of making a correct decision and being aware of the implications of using a specific formula in the posterior analysis.

Now, even though this is an important step in the process of creating a good analysis, Meta-Analysts do not usually declare in their documents what methodology was used to transform effect sizes or what assumptions were made. They do not even mention if a sensitivity analysis was performed to validate the final conclusions (see

Borenstein et al., 2009).

Hopefully, this document makes Meta-Analysts realize the importance of using correct effect size transformations and the impact that any choice can have in their results. This document was made with the purpose of encouraging the people to really understand the formulas that can be easily found online and take an informed decision about the final effect size. As it can be seen in the reference list, in the literature can be found a diverse and broad discussion about this topic.

References

Aaron, B., Kromrey, J. D., and Ferron, J. (1998). Equating ”r”-based and ”d”-based effect size indices: Problems with a commonly recommended formula. Paper presented at the Annual Meeting of the Florida Educational Research Association, Orlando, FL.

Bonett, D. G. (2007). Transforming odds ratios into correlations for meta-analytic research. *American Psychologist*, pages 254–255.

Bonett, D. G. and Price, R. M. (2005). Inferential methods for the tetrachoric correlation coefficient. *Journal of Educational and Behavioral Statistics*, pages 213–225.

Borenstein, M., Hedges, L., Higgins, J., and Rothstein, H. (2009). *Introduction to Meta-Analysis*. Statistics in practice. Wiley.

Cox, D. and Snell, E. (1989). *The Analysis of Binary Data*. Chapman and Hall/CRC, London, 2 edition.

Da Costa, B. R. d., Rutjes, A. W., Johnston, B. C., Reichenbach, S., Nüesch, E., Tonia, T., Gemperli, A., Guyatt, G. H., and Jüni, P. (2012). Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. *International Journal of Epidemiology*, 41(5):1445–1459.

Friedman, H. (1968). Magnitude of experimental effect and a table for its rapid estimation. *Psychological Bulletin*, 70(4):245–251.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10):3–8.

Glass, G. V., Smith, M. L., and McGaw, B. (1981). *Meta-analysis in social research*. Sage Publications, Beverly Hills.

Hasselblad, V. and Hedges, L. (1995). Meta-analysis of screening and diagnostic tests. *Psychological bulletin*, 117:167–78.

Lenhard, W. and Lenhard, A. (2016). *Calculation of Effect Sizes*. `https://www.psychometrica.de/effect_size.html`.

Lipsey, M. and Wilson, D. (2001). *Practical Meta-Analysis*. Applied Social Research Methods. SAGE Publications.

Pearson, K. (1900). Mathematical contributions to the theory of evolutionvii: On the correlation of characteristics not quantitatively measurable. *Philosophical Transactions of the Royal Society of London*, 195:1–47.

Polanin, J. R. and Snilstveit, B. (2016). *Campbell Methods Policy Note on Converting Between Effect Sizes*. The Campbell Collaboration, Oslo.

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Sage Publications, 1 edition.

Sanchez-Meca, J., Marin-Martinez, F., and Chacon-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, 8(4):448–467.

Suissa, S. (1991). Binary methods for continuous outcomes: A parametric alternative. *Journal of Clinical Epidemiology*, 44(3).

Whitehead, A. (2002). *Meta-Analysis of Controlled Clinical Trials*. John Wiley and Sons.