Converting between effect sizes

Version 1.1

Campbell Collaboration

Colophon

Campbell Library Methods papers

The Campbell Library Methods Series comprises three types of publications:

Methods Discussion Papers	New or innovative ideas currently in development in the field of methodology, these papers are intended for discussion and do not represent official Campbell policy or guidance
Methods Policy Notes	Current Campbell Collaboration policy on specific methods for use in Campbell systematic reviews of intervention effects
Methods Guides	Guides on how to implement specific systematic review methods
Disclaimer	Campbell Collaboration Methods Discussion Papers are published to promote discussion of new and innovative methods in systematic reviews, making these approaches available to a broad audience. Papers are published as submitted by the authors. They are not subject to review or editing by the Campbell Collaboration. The views expressed are those of the authors, and may not be attributed to the Campbell Collaboration. Campbell Collaboration Methods Discussion Papers do not represent Campbell policy.

Editorial Board

The Campbell Collaboration was founded on the principle that systematic reviews on the effects of interventions will inform and help improve policy and services. Campbell offers editorial and methodological support to review authors throughout the process of producing a systematic review. A number of Campbell's editors, librarians, methodologists and external peer-reviewers contribute.

The Campbell Collaboration
P.O. Box 7004 St. Olavs plass
0130 Oslo, Norway
www.campbellcollaboration.org

Table of contents

How to use this document

This Campbell methods policy note on converting between effect sizes sets out current Campbell Collaboration policy on calculating a common effect size from different metrics for use in meta-analyses in Campbell systematic reviews of intervention effects. It is intended to serve as a point of reference for Campbell authors and editors in both the design and production phases of Campbell reviews of intervention effects, to ensure that protocols and full reviews are prepared in line with collaboration policy. In order to provide context for the current policy, the Note is prefaced with a brief introduction to the topic and appended with a list of additional resources, tables of formulas and worked examples. It is not intended to provide comprehensive guidance or a tutorial on how to apply methods in Campbell or other systematic reviews (see below).

Campbell authors and editors should consult **Campbell Collaboration systematic reviews: Policies and Guidelines** for detailed information on the general requirements for Campbell systematic reviews, guidelines for producing them, and selected sources of further information about systematic reviews that are consistent with those requirements and guidelines.

Campbell authors and editors should also refer to **Methodological expectations of Campbell reviews of intervention effects (MEC2IR)** – standards for the conduct and reporting of Campbell Collaboration systematic reviews of intervention effects. MEC2IR conduct and reporting standards can be found at https://campbellcollaboration.org/library. These standards provide authors and users of Campbell reviews of intervention effects with clear and transparent expectations of review conduct and reporting and facilitate the editorial process. All new or updated Campbell reviews of intervention effects that proceed to the full review production stage (i.e. following approval of the protocol) after 1 October 2014 are required to comply with all mandatory MEC2IR conduct and reporting standards in order to be signed-off for publication in the Campbell systematic reviews monograph series. MEC2IR conduct standards #47 ('making maximal use of data'), #62 ('combining different scales') and #63 ('ensuring meta-analyses are meaningful') address issues related to converting between effect sizes.

Campbell also publishes Campbell Methods Guides that provide detailed guidance for authors on how to apply specific methods in the production of Campbell systematic reviews. Campbell Methods Guides supplement general guidelines for producing Campbell systematic reviews, as described in Campbell Collaboration systematic reviews: policies and guidelines. Campbell Methods Guides do not currently cover the topic of converting between effect sizes.

Introduction

Statistical meta-analysis is the preferred method for integrating studies in a systematic review of intervention effects (Campbell Collaboration, 2014). However studies identified for inclusion in a systematic review typically report a range of different measures of the same outcome construct. Without a common effect size metric across studies, the analyst has limited options to synthesize studies and compare findings across contexts. Therefore, before researcher can use statistical meta-analysis to synthesize findings from different studies estimates of effects from individual studies must be converted into a common metric. The purpose of this Methods Policy Note is to provide an introduction to transforming effect sizes into common metrics in order to synthesize studies and to outline the Campbell policy in this area.

An effect size is a measure of the *magnitude* of an observed relationship, treatment effect, or population parameter (Kelley and Preacher, 2012). Studies included in systematic reviews of effects typically provide effect sizes which estimate the magnitude of the difference in an outcome between participants in the treatment groups and participants in the control group. An effect size is different from tests of statistical significance, which indicate whether results are likely to be due to chance. Tests of statistical significance tell us nothing about the practical significance of the results in terms of the size of the effect (Ellis, 2010). To answer this question, we need an effect size.

There are many different types of effect size indices, including standardized mean-difference (d), odds ratio (OR), and correlation coefficient (r). In the case of the standardized mean-difference, an analyst divides the difference between two group's means by the standard deviation. A standard deviation represents the spread of the data, assuming that the distribution is reasonably large and normal. Using this principle, we are able to quantify the mean-difference in relation to the spread from which it derives. For example, an intervention designed to increase reading scores may result in the comparison group scoring a sample mean of 50 and a sample standard deviation of 10, and the intervention group scoring a sample mean of 80 and a sample standard deviation of 10. The principle of the standard deviation states that 68.2% of all participants' scores will fall within one standard deviation above and below the mean. Therefore, 68.2% of score will fall between $40 - 60$ in the comparison group and 68.2% of scores will fall between $90 - 110$ in the intervention group. Figure 1 illustrates this concept.

The odds ratio and correlation coefficient also represent the magnitude of an effect, but these are calculated differently because of the nature of the measures. An odds ratio can measure the size of the effect when the outcome variable is dichotomous. A correlation coefficient can also measure the effects of a treatment, but instead of measuring the difference between the two groups' means, the correlation coefficient observes the relationship between the treatment status and the outcome.
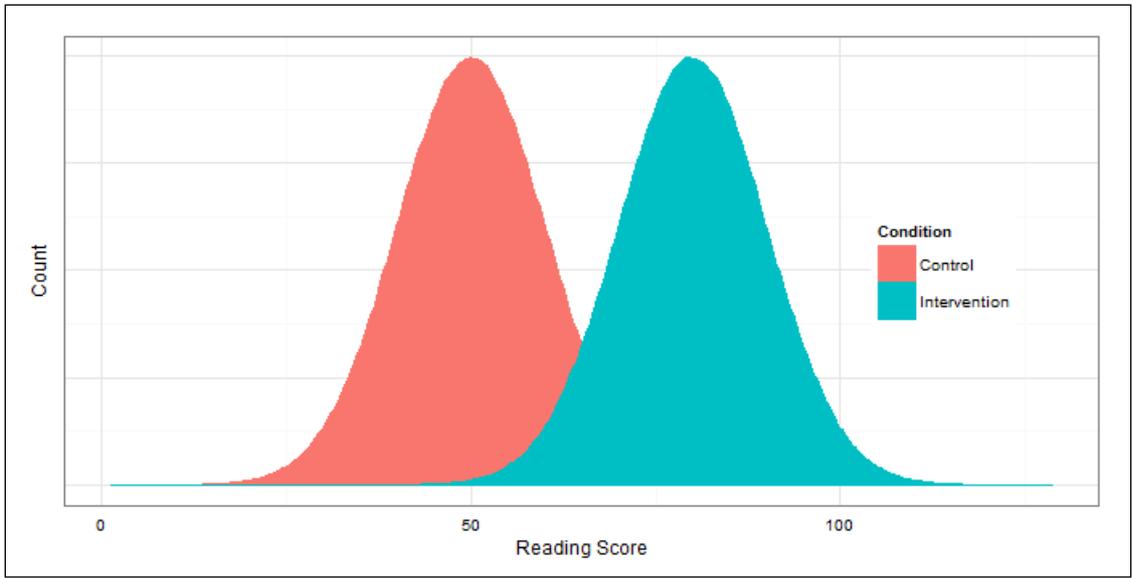
Figure 1: *Distribution of variable scores for intervention and control groups. A standardized mean-difference can be calculated using the two distributions' means and standard deviations*

Each of these concepts are related, but the results of the different effect size calculations are not directly comparable. This is because the effects are in different "metrics", or the scale of the effect size. This is easy to visualize in the Table 1. The standardized mean-difference and correlation coefficient each have "no effect" centered at 0.0, but the odds ratio's center is at 1.0. The standardized mean-difference has no theoretical lower and upper limits, while the correlation coefficient is standardized to fall between -1.0 and 1.0. The odds ratio has a lower limit of 0.0 with no theoretical upper limit.

Table 1: Effect size metrics

	No effect	Lower limit	Upper limit
Standardized Mean-Difference	0.0	∞	∞
Correlation Coefficient	0.0	-1.0	1.0
Odds Ratio	1.0	0.0	∞

To conduct a meta-analysis across all of the effect sizes, the analyst must transform all effect sizes into a common metric. The metric that all effect sizes should be transformed to depends on the measures and data provided by the included studies. A synthesis containing ten studies and ten effects may have seven standardized mean-difference effect sizes and three odds ratios. The analyst should transform the smallest number of effect sizes possible, in this case, transforming the three odds ratios into standardized mean-difference effect sizes. An example of this can be found in the Campbell review on school-based programs to decrease teen dating violence (De La Rue, Polanin, Espelage, & Pigott, 2015). The analysis of the dating violence perpetration outcome yielded three studies and a total of six effect sizes. One of the six effect sizes was originally recorded as an odds ratio and converted to the standardized mean-difference metric for the meta-analysis.

Appendix A provides a table of conversions among the common effect size metrics, standardized mean-difference (d), odds ratio (OR), and correlation coefficient (r). The conversions derive from various sources (Bonnett & Price, 2005; Borenstein et al, 2009; Lipsey & Wilson, 2001) - see these sources for derivations and additional information. We also provide the effect size variance calculations, which will be used in a traditional meta-analytic model to weight the effect sizes. Appendix B provides two worked examples for common conversions (odds ratio to standardized mean difference; and standardized mean difference to correlation coefficient) using the program R (R Core Team, 2016). Additionally, there are many adjustments and modifications that may be made for the conversion from odds ratio to a correlation coefficient. The conversion presented in Appendix A assumes that no additional information about the marginal is provided in the primary study. Should a 2 x 2 table be provided, Bonnett & Price (2005) suggested formulas that produce slightly more accurate results.

In addition to effect size conversions among common effect sizes, studies may also report limited information in which to calculate an effect size. In these cases, it is appropriate to calculate an effect size using other information provided (Lipsey & Wilson, 2001). Commonly, this requires extracting a t or F statistic, but as a last case, a p-value may also be used to estimate an effect size. As with all missing data in systematic reviews, it is important to contact the primary study authors for traditional descriptive statistics prior to using an estimation. Appendix C provides a table of formulas for computing standardized mean difference from reported study information.

Some of the suggested transformations are approximations while others are exact transformations. Therefore, reviewers should: a) carefully consider the underlying construct before transformation; b) perform sensitivity analyses, especially when using educated guesses or imputation to supply information for the transformation, and c) conduct moderator analyses that test for differences between un-transformed and transformed effect sizes (see Lipsey & Wilson, 2001 for further information).

In summary, it is often possible and necessary to transform effect sizes into a common metric or calculate effect sizes from available data before conducting meta-analysis. When data is available reviewers should always transform estimates into a common metric using appropriate formulae. The transformations provided here allow reviewers to calculate a range of effect sizes from different data.

Campbell policy

This policy on the conversion of effect sizes in Campbell systematic reviews of intervention effects was adopted by the Campbell Collaboration at a Steering Group meeting. Campbell methods policy on this issue is that authors using this method are expected to:

1. Define the effect size index used and explain why it was chosen. Unless a compelling rationale for an alternative is presented, that index should be one of the recognized effect size statistics, such as the standardized mean difference, odds ratio, risk ratio, or correlation coefficient.
2. Calculate each effect size using standard calculation procedures and report the formulae used for computing it, its standard error, and any other relevant statistical representations.
3. Consider and report how estimation will be done when studies do not report the most desirable statistics for calculating the selected effect size.
4. Address unit of analysis issues arising from the inclusion of cluster randomized studies (see Hedges, 2011).
5. Report and justify any adjustments to effect size calculations due to small sample sizes, missing data and outliers.
6. If multiple effect size metrics are reported within one outcome analysis, use the effect size metric most commonly reported as the primary metric and convert all of effect sizes to that metric using the conversion formulae in Appendix A.
7. Conduct moderator or sensitivity tests between the original effect sizes and the transformed effect sizes.
8. Should differences emerge between the original and transformed effects, consider synthesizing these studies separately or removing the transformed effects from the analysis.

Glossary

Correlation coefficient: A correlation coefficient represents the relationship between two continuous variables.

Effect size: An effect size represents the magnitude of a treatment effect, in the case of standardized mean-difference or odds ratio, or the relationship between two variables, as in the case of the correlation coefficient.

Odds ratio: The odds ratio is an effect size metric used when the outcome variable is dichotomous and two groups are compared.

Standardized mean-difference: The standardized mean-difference is an effect size metric used when the outcome variable is continuous and two groups are compared. It is commonly represented using Cohen's d or Hedges' g.

Transformation: A statistical transformation changes the metric of the effect size. A transformation may be exact, as in the case of transforming a t-statistic into a Cohen's d, or approximate.

Glossary

Further resources

Training videos on computing basic effect sizes and computing advanced effect sizes can be found here:

Campbell Training Video on computing basic effect sizes: https://youtu.be/Fggs7zOhw6c

Campbell Training Video on computing advanced effect sizes: https://youtu.be/MX1qdFEUodg

A range of different software packages and effect size calculators allow reviewers to calculate commonly used effect sizes from a range of different data, including:

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2005). Comprehensive meta-analysis [Version 2]. Englewood, NJ: Biostat.

Del Re, A.C. (2014). compute.es: Compute effect sizes [Software] [Version 0.2-4]. Retrieved from: http://cran.r-project.org/web/packages/compute.es/index.html

Wilson, D. B. (2015). Practical meta-analysis effect size calculator [Online software]. Retrieved from: http://www.campbellcollaboration.org/resources/effect_size_input.php

References

Bonett, D. G., & Price, R. M. (2005). Inferential methods for the tetrachoric correlation coefficient. Journal of Educational and Behavioral Statistics, 30, 213-225.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011).Introduction to meta-analysis. West Sussex, United Kingdom: John Wiley & Sons.

De La Rue, L., Polanin, J.R., Espelage, D.L., & Pigott, T.D. (2014). School-based interventions to reduce dating and sexual violence: A systematic review. Campbell Systematic Reviews, 7, 1 − 111. doi:10.4073/csr.2014.7

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. Journal of Pediatric Psychology, jsp004.

Ellis, P. D. (2010). The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results. New York, NY: Cambridge University Press.

Kelley, K., & Preacher, K. J. (2012). On effect size. Psychological methods, 17, 137-152.

Lipsey, M. W. & Wilson, D. B. (2001). Practical meta-analysis (2nd ed.). Thousand Oaks, CA: Sage Publications.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Appendices

APPENDIX A: EFFECT SIZE CONVERSION FORMULAE

	Standardized Mean-Difference	Odds Ratio	Correlation Coefficient
Standardized Mean-Difference	-	$d = Log(OR) * \left(\frac{\sqrt{3}}{\pi}\right)$ $V_d = V_{Log(OR)} * \left(\frac{3}{\pi^2}\right)$	$d = \frac{2 * r}{\sqrt{1 - r^2}}$ $V_d = \frac{4 * (V_r)}{(1 - r^2)^3}$
Odds Ratio	$Log(OR) = d * \left(\frac{\pi}{\sqrt{3}}\right)$ $V_{Log(OR)} = V_d * \left(\frac{\pi^2}{3}\right)$	-	$Log(OR) = \left(\frac{2 * r}{\sqrt{1 - r^2}}\right) * \left(\frac{\pi}{\sqrt{3}}\right)$ $V_{Log(OR)} = \frac{4 * (V_r)}{(1 - r^2)^3} * \left(\frac{\pi^2}{3}\right)$
Correlation Coefficient	$r = \frac{d}{\sqrt{d^2 + a}}$ $a = \frac{(n_1 + n_2)^2}{n_1 * n_2}$ $V_r = \frac{a^2 * V_d}{(d^2 + a)^3}$	$r = \frac{Log(OR) * \left(\frac{\sqrt{3}}{\pi}\right)}{\sqrt{\left(Log(OR) * \left(\frac{\sqrt{3}}{\pi}\right)\right)^2 + a}}$ $a = \frac{(n_1 + n_2)^2}{n_1 * n_2}$ $V_r = \frac{a^2 * V_{Log(OR)} * \left(\frac{3}{\pi^2}\right)}{\left(\left(Log(OR) * \left(\frac{\sqrt{3}}{\pi}\right)\right)^2 + a\right)^3}$	-

APPENDIX B: WORKED EXAMPLES

Example 1: Odds ratios to standardized mean difference

One common conversion is the conversion from an odds ratio to the standardized mean-difference. A likely mistake is to use the untransformed odds ratio instead of the log-odds ratio. Below is an example, starting first with the calculated odds ratio and moving to the conversion. Assume a two-group, intervention vs. control design where the intervention is assumed to increase an outcome. The odds ratio is equal to 1.50. The example formulas are presented using R and can be copied and pasted directly into an R console.

```
logOR <- log(1.50)
d_logOR <- logOR * (sqrt(3)/pi)
d_logOR

## [1] 0.2235


V_logOR <- .08
V_d_logOR <- V_logOR * (3/(pi^2))
V_d_logOR

## [1] 0.02432
```

Example 2: Standardized mean-difference to correlation coefficient

Another common conversion is the standardized mean-difference to the correlation coefficient. Here, we need to provide a measure of the sample size (represented as a in the table). Assume, again, a two-group, intervention vs. control design. The estimated standardized mean-difference is .35; 75 individuals are included in each group.

```
d <- .35
a <- ((75 + 75)^2)/(75*75)
r_d <- d/sqrt(((d^2)+a))
r_d

## [1] 0.1724


V_d <- ((75 + 75)/(75 * 75)) + ((d^2)/(2*(75 + 75)))
V_r_d <- ((a^2)*V_d)/(((d^2)+a)^2)
V_r_d

## [1] 0.02549
```

APPENDIX C: FORMULAE FOR COMPUTING STANDARDIZED MEAN DIFFERENCE FROM REPORTED STUDY INFORMATION

Data Provided	Conversion
T-Statistic from two groups; equal sample sizes	$$d = \frac{2 * t}{\sqrt{N}}$$
T-Statistic from two groups; unequal sample sizes	$$d = t * \sqrt{\frac{n_1 + n_2}{n_1 * n_2}}$$
F-Statistic from two groups (i.e., 1 df); equal sample sizes	$$d = 2 * \sqrt{\frac{F}{N}}$$
F-Statistic from two groups (i.e., 1 df); unequal sample sizes	$$d = \sqrt{\frac{F(n_1 + n_2)}{n_1 * n_2}}$$
F-Statistic from two-group ANCOVA (i.e., 1 df; unequal sample sizes	$$d = \sqrt{\frac{(n_1 + n_2) * (1 - r^2) * F}{n_1 * n_2}}$$ where r is the correlation between the covariate and the outcome.
Estimating means from medians and range	$$Mean = \frac{a + (2 * Median) + b}{4}$$ where a is the minimum and b is the maximum.
Estimating standard deviation from the median and range	$$SD = \sqrt{\frac{1}{12} * \left(\frac{(a - (2 * Median) + b)^2}{4}\right) + (b - a)^2}$$ where a is the minimum and b is the maximum.
Unstandardized beta coefficient (i.e., 1 df)	$$d = \frac{b}{s_{pooled}}$$ where s_{pooled} is the pooled standard deviation.