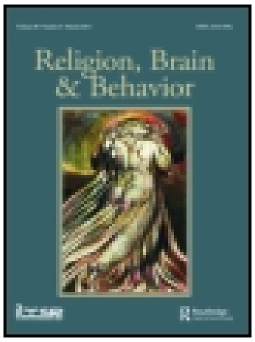

Religion, Brain & Behavior

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rrbb20

The end justifies all means: questionable conversion of different effect sizes to a common effect size measure

Marcel A.L.M. van Assen, Andrea H. Stoevenbelt & Robbie C.M. van Aert

RELIGION, BRAIN & BEHAVIOR
https://doi.org/10.1080/2153599X.2022.2070249

Routledge
Taylor & Francis Group

COMMENTARY

Check for updates

The end justifies all means: questionable conversion of different effect sizes to a common effect size measure

Marcel A.L.M. van Assen, Andrea H. Stoevenbelt, and Robbie C.M. van Aert

Meta-research center, Tilburg University, Tilburg, Netherlands

The main goal of meta-science projects (e.g., Open Science Collaboration, 2015; The MARP Team, 2022) is drawing overarching conclusions by combining the observed statistical evidence. This corresponds to the primary goal of meta-analysis where statistical results of different primary studies estimating the underlying average effect size are combined. Typically, the statistical evidence originates from analyses that give rise to different outcomes and effect size measures. For instance, outcomes can be coefficients of regression or multilevel regression models, differences between means, correlation coefficients, odds ratios, etc. The problem of meta-science projects and meta-analyses is how to deal with all these different statistical outcomes when summarizing the evidence.

The standard approach is to transform all different statistical outcomes to the same effect size measure, which is explained by classic texts on meta-analysis (e.g., Chapter 7 in Borenstein et al., 2009). However, many of these transformations make strong or even untenable assumptions, resulting in estimates of effect sizes with unknown distributional properties. For example, meta-analysts often estimate effect sizes by combining Pearson correlations between two continuous variables with (point-biserial) correlations between one dichotomous and one continuous variable. However, only after some transformations may these coefficients be integrated into the same meta-analysis (Jacobs & Viechtbauer, 2017). For many other transformed effect sizes, e.g., of odds ratios transformed into correlations, distributional properties are largely unknown. Therefore, these transformations are questionable and beg the question whether the goal to compare or summarize as much data as possible in one analysis is an end that justifies all means.

The questionable practice of combining outcomes of different statistical outcomes in one analysis also occurs in MARP. To answer MARP's two research questions on the relationship between religiosity and self-reported well-being 117 research teams were asked to analyze a large dataset containing data of 24 countries. Statistical outcomes of 100 for research question 1 (RQ1) and 99 for RQ2 of these teams were transformed into standardized beta coefficients (i.e., coefficients of models with standardized variables). However, the most prevalent analytical approach in MARP was multilevel (linear regression) analysis (35.2%, see Table 1 of MARP), an approach where standardization is discouraged because it can be done in different ways, potentially leading to different outcomes depending on how it is done (e.g., Heck et al., 2013). The reason is that the level-2 or macro-level variance affects the standardization and thereby the standardized beta coefficients, which we demonstrate with a simple example.

Consider a data set with X = 1,2,3, and half of the points lie on Y = X + 9 and the other half on Y = X + 11, with six hundred observations for each of the six different data points, resulting in a total sample size of 3,600. In this data set, the unstandardized regression coefficient equals 1 and the standardized beta coefficient equals .632, with $R^2 = .40$. We now create six equally large groups 1,2, … ,6 in the first multilevel data set, with hundred observations for each of the six data points. In the second multilevel data set a new Y_2 variable is created by $Y_2 = Y - 5 \times (\text{group} - 1)$, the third set is created by keeping Y but defining $X_2 = X + 5 \times (\text{group} - 1)$, and the fourth set uses both X_2 and Y_2. All

CONTACT Marcel A.L.M. van Assen m.a.l.m.vanassen@uvt.nl

data sets have the same within-group effect equal to 1 (see scatter plots in Figure 1), which is correctly estimated by the random-intercept multilevel analyses (using maximum likelihood estimation) on all data sets. However, the standardized beta coefficients of the multilevel analyses obtained by standardizing the variables prior to the analysis on data sets 1, 2, 3, 4 are .632, .095, 6.628, .992, respectively. SPSS and R data, syntax, and output can be found at https://osf.io/xew35/.

The conclusions we can draw from this simple example are that standardized beta coefficients of multilevel analysis are different from those of regression analysis, and that variation across groups affects the standardized beta coefficients, making a comparison of these standardized beta coefficients very difficult or even nonsensical. This was just one example, but likely similar cases can be made for standardized beta coefficients of many of the other statistical approaches used to answer MARP's RQs. We therefore argue not to compare these standardized regression coefficients between statistical approaches, or even within the set of multilevel analyses. Instead, we corroborate others (e.g., Jacobs & Viechtbauer, 2017; Lipsey & Wilson, 2001; Sánchez-Meca et al., 2003) recommending researchers to consider the statistical approach, design (e.g., within- and between subjects, Reeves et al., 2021), and effect size measures, when synthesizing statistical evidence by including dummy variables in the analysis to assess systematic differences.

For meta-science projects, we recommend summarizing the results by focusing on outcomes that do not require transforming data. The interpretation of the results of MARP can be focused on (i) the sign of the estimated effect size, (ii) evidence against the null-hypothesis (p-values), and (iii) evidence in favor of a hypothesis (e.g., using Bayes factors). In any case, more rigorous

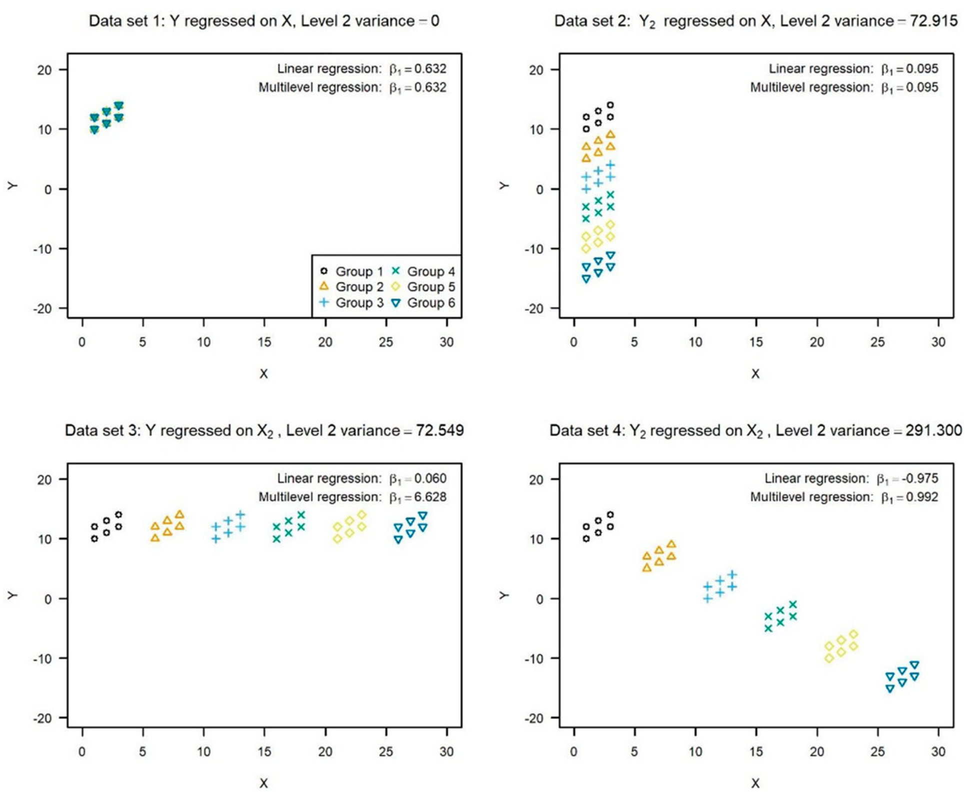

Figure 1. Scatter plots of multilevel data sets of the example. The within-group effect equals 1 for all data sets but the level-2 variances differ. Standardized beta coefficients of both linear and multilevel regression are also shown. Note that the participants of all groups have the same scores in data set 1 such that the symbols are plotted on top of each other.

meta-science and meta-analysis is needed, where the end does not justify all means but where the differences in statistical approach, design, and effect size measure are respected.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Marcel A.L.M. van Assen http://orcid.org/0000-0002-7517-6081

References

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

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2013). *Multilevel and longitudinal modeling with IBM SPSS*. Routledge.

Jacobs, P., & Viechtbauer, W. (2017). Estimation of the biserial correlation and its sampling variance for use in meta-analysis. *Research Synthesis Methods*, *8*(2), 161–180. https://doi.org/10.1002/jrsm.1218

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

The MARP Team. (2022). A Many-Analysts Approach to the Relation Between Religiosity and Well-being. *Manuscript in preparation*.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), https://doi.org/10.1126/science.aac4716

Reeves, B. C., Deeks, J. J., Higgins, J. P. T., Shea, B., Tugwell, P., & Wells, G. A. (2021). Including non-randomized studies on intervention effects. In J. P. T. Higgins, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions (version 6.2 ed.)*. Cochrane. www.training.cochrane.org/handbook.

Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, *8*(4), 448–467. https://doi.org/10.1037/1082-989X.8.4.448