Effect Sizes and their Variance for AB/BA Crossover Design Studies

Lech Madeyski
Faculty of Computer Science and Management,
Wroclaw University of Science and Technology
Wroclaw, Poland
Lech.Madeyski@pwr.edu.pl

Barbara Kitchenham
School of Computing and Mathematics, Keele University
Keele, Staffordshire, UK
b.a.kitchenham@keele.ac.uk

ABSTRACT

We addressed the issues related to repeated measures experimental design such as an AB/BA crossover design that have been neither discussed nor addressed in the software engineering literature.

Firstly, there are potentially two different standardized mean difference effect sizes that can be calculated, depending on whether the mean difference is standardized by the pooled within groups variance or the within-participants variance. Hence, we provided equations for non-standardized and standardized effect sizes and explained the need for two different types of standardized effect size, one for the repeated measures and one that would be equivalent to an independent groups design.

Secondly, as for any estimated parameters and also for the purposes of undertaking meta-analysis, it is necessary to calculate the variance of the standardized mean difference effect sizes (which is not the same as the variance of the study). Hence, we provided formulas for the small sample size effect size variance and the medium sample size approximation to the effect size variance, for both types of standardized effect size.

We also presented the model underlying the AB/BA crossover design and provided two examples (an empirical analysis of the real data set by Scanniello, as well as simulated data) to demonstrate how to construct the two standardized mean difference effect sizes and their variances, both from standard descriptive statistics and from the outputs provided by the linear mixed model package lme4 in R.

A conclusion is that crossover designs should be considered (instead of between groups design) only if:

- previous research has suggested that ρ is greater than zero and preferably greater than 0.25;
- there is either strong theoretical argument, or empirical evidence from a well-powered study, that the period by technique interaction is negligible.

Summarizing, our journal first paper [3]:

(1) Presents the formulas needed to calculate both non-standardized and standardized mean difference effect sizes for AB/BA crossover designs (see Section 4 and 5 of our paper [3]).
(2) Presents the formulas needed to estimate the variances of the non-standardized and standardized effect sizes which in the later cases need to be appropriate for the small to medium sample sizes commonly used in software engineering crossover designs (see Section 5 of our paper [3]).
(3) Explains how to calculate the effect sizes and their variances both from the descriptive statistics that should be reported and from the raw data (see Section 6 of our paper [3]).

It is worth mentioning that we based our formulas on our own corrections to the formulas presented earlier by Curtin et al. [1]. Our corrections for the variances of standardized weighted mean difference of an AB/BA cross-over trial were accepted by the author of the original formulas (Curtin), submitted jointly as a letter to Editor of Statistics in Medicine to assure the widespread (also beyond the software engineering domain) adoption of the corrected formulas, and accepted [2]. We proposed an alternative formulation of the standardized effect size for individual difference effects that is comparable with the standardized effect size commonly used for pretest/posttest studies. We also corrected the small sample size and moderate sample size variances reported by Curtin et al. for both the individual difference effect size and the standardized effect size comparable to independent groups trials, showing the derivation of the formulas from the variance of a t-variable. Using these results, researchers can now correctly calculate standardized effect size variances, allowing the calculation of confidence intervals for AB/BA cross-over trials, which in turn provides a direct link to null hypothesis testing and supports meta-analysis. Meta-analysts can now validly aggregate together results from independent groups, pretest/posttest and AB/BA cross-over trials. Last but not least, the presented contributions allow corrections of previously reported results.

KEYWORDS

Empirical software engineering, Meta-analysis, Effect size

REFERENCES

[1] Francois Curtin, Douglas G. Altman, and Diana Elbourne. 2002. Meta-analysis combining parallel and cross-over clinical trials. I: Continuous outcomes. *Statistics in Medicine* 21, 15 (2002), 2132–2144. https://doi.org/10.1002/sim.1205
[2] Barbara Kitchenham, Lech Madeyski, and François Curtin. 2018. Corrections to effect size variances for continuous outcomes of cross-over clinical trials. *Statistics in Medicine* 37, 2 (2018), 320–323. https://doi.org/10.1002/sim.7379 (Draft: http://madeyski.e-informatyka.pl/download/KitchenhamMadeyskiCurtinSIM.pdf).
[3] Lech Madeyski and Barbara Kitchenham. [n. d.]. Effect Sizes and their Variance for AB/BA Crossover Design Studies. *Empirical Software Engineering* ([n. d.]). https://doi.org/10.1007/s10664-017-9574-5 (Draft: http://madeyski.e-informatyka.pl/download/MadeyskiKitchenham18ESE.pdf, should appear in 2018).