CAMPBELL METHODS SERIES: DISCUSSION PAPER 4

October 2017

Research synthesis and meta-analysis of Monte Carlo studies: the best of two worlds

Brandon LeBeau

Version 1.0

CampbellCollaboration

Colophon

Title	Research synthesis and meta-analysis of Monte Carlo studies: the best of two worlds
Institution	The Campbell Collaboration
Authors	LeBeau, B
DOI	10.4073/cmdp.2017.3
No. of pages	9
Citation	LeBeau, B. Research synthesis and meta-analysis of Monte Carlo studies: the best of two worlds. Oslo, Norway: The Campbell Collaboration. Retrieved from: www.campbellcollaboration.org/ DOI: 10.4073/cmdp.2017.3
Main series ISSN	2535-2466
Sub-series ISSN	2535-2490
Copyright	© LeBeau This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
About this paper	Campbell discussion papers include title registration forms and protocols for new evidence synthesis products being piloted for possible inclusion in the Campbell Library. In the discussion paper the authors include some reflections on the approach, and how it differs from a Campbell systematic review.

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

ABSTRACT

Research syntheses and meta-analyses have been widely used to combine primary applied research studies. To date, there has been little work applying these methods to combine Monte Carlo studies, however narrative reviews are common to evaluate the Monte Carlo literature. This manuscript briefly describes the benefits and challenges of combining outcomes from Monte Carlo studies using research synthesis and meta-analysis methods. The particular focus is to discuss and extend ideas first described by Harwell (1992).

INTRODUCTION

Monte Carlo (MC) studies are a popular way to explore statistical properties of an estimator and are particularly useful when mathematical derivations are difficult or impossible. Monte Carlo studies of this style often explore the impact of assumption violations on estimation of parameters (e.g. Kwok, West, & Green, 2007; LeBeau, 2016; Murphy & Pituch, 2009). Common statistical assumptions explored with MC studies include, distribution of residuals, model misspecification, sample size and magnitude of effect to name a few.

Although computers have improved to speed up simulations of many models, the complexity of statistical methods has advanced at a similar pace. This makes it difficult for a single MC study to explore all possible conditions that may impact the model estimation. Instead, researchers are commonly strategic by using a purposive sampling mechanism rather than a random sampling procedure. This has the effect of being a focused, well thought out rationale for simulation conditions that may be important for a given problem, however has the drawback of poorer external validity (Skrondal, 2000).

Narrative reviews are a common way to explore the MC literature, particularly as part of a literature review for a MC study or dissertation. The primary focus of the narrative review is to motivate the current state of the research, with a focus on gaps in the coverage of the current MC studies. One of the first and most well known narrative reviews of MC literature dates back to Glass's 1972 paper exploring assumption violations when using analysis of variance or analysis of covariance (Glass, Peckham, & Sanders, 1972). Although these narrative reviews are useful, they can have an “impressionistic nature” (Harwell, 1992) in which interpretations may differ based on how the manuscript is read.

Using meta-analysis or research synthesis methods can help to improve external validity of MC studies to a wider range of simulations conditions and also attempt to overcome the “impressionistic nature” (Harwell, Rubinstein, Hayes, & Olds, 1992) of a narrative review. This manuscript will explore in more detail how meta-analysis and research synthesis methods can be used to combine MC studies and also expand on the discussion started by Harwell (1992). Since the paper by Harwell (1992), there have been studies following this framework (e.g. Guilera, Gómez-Benito, Hidalgo, & Sánchez-Meca, 2013; Harwell, 1997, 2003; Harwell et al., 1992; J. C. Keselman, Lix, & Keselman, 1996; Lix, Keselman, & Keselman, 1996; Powell & Schafer, 2001); however, this framework has been used much less frequently recently. This manuscript aims to discuss the benefits and challenges of synthesizing the Monte Carlo literature using meta-analysis methods to hopefully increase studies in this under researched area.

COMBINING MONTE CARLO STUDIES

Monte Carlo studies use quasi-random number generation to simulate many (e.g. 100, 1000, or more) replications within a single cell of the MC study. Once data generation is complete, a statistical model is fitted to each generated data set. Estimates are then compared to the population values (from the data generation) to explore potential bias or the number of rejected null hypotheses are used to compute type I error or power rates. The interested reader is directed to (Burton, Altman, Royston, & Holder, 2006) for more information on interpretation and computation of common MC outcomes. These outcomes serve as the primary outcomes of interest in the meta-analysis of MC studies. The benefits and challenges of statistically combining these outcomes are now explored in more detail.

RESEARCH QUESTIONS

Setting the research questions for a research synthesis or meta analysis of MC studies will be different than conducting a MC study. The focus is not on specific manipulated simulation conditions, rather on the aggregate impact data characteristics may have on model estimation. As a result, the research questions will naturally be broader when conducting a research synthesis or meta analysis of MC studies compared to a single MC study. Although the research question may be broader compared to a MC study, the MC literature of interest should remain narrow to ensure the studies included are exploring similar research questions in a similar framework (e.g., longitudinal linear mixed models).

BENEFITS

Many advantages of research synthesis and meta-analytic methods can be found when combining outcomes from MC studies. The benefits of adopting these methods focus on a general theme of improving external validity, increasing study coverage, and providing structured inclusion criteria.

Structured inclusion criteria

Common to both research syntheses and meta-analyses is the strict, transparent, clear, and reproducible inclusion criteria (Cooper, Hedges, & Valentine, 2009). From these inclusion criteria, the search terms for the literature search should follow naturally. In addition, having a strong set of inclusion criteria can narrow the focus of the MC literature to the area of interest. For example, a research synthesis or meta-analysis could explore the impact of missing data techniques for the type I error rate using linear regression. The inclusion criteria when performing the literature search would then be limited to studies that included different missing data techniques particularly focusing on linear regression, thus removing MC studies exploring missing data techniques within a linear mixed model. The search terms can be easily derived from these inclusion criteria and ensures that the comparison is focused and combination of MC studies are appropriate. This is not unlike well formulated inclusion criteria from more typical research syntheses or meta-analyses of empirical studies (Cooper et al., 2009).

Improved external validity

Arguably one of the largest weaknesses of MC studies stems from their lack of external validity (Paxton, Curran, Bollen, Kirby, & Chen, 2001; Skrondal, 2000) due to purposive sampling. A meta-analytic framework can greatly enhance the external validity of these studies by greatly extending the range of simulation conditions included in the analysis. Although still not representing a random sample of simulation conditions, the greater range of simulation conditions will enhance external validity to a wider range of values compared to a single MC study. Sample size is a good example, one MC study may have chosen 25, 50, and 100 for the three sample size conditions and another may have chosen 50, 200, and 500. Two important things happen here, first the sample size of 50 now has a larger sample size to estimate the effect, and second, the sample size values are now 25, 50, 100, 200, and 500 when combined using meta-analysis methods.

Expanded coverage of simulation conditions

This point is similar to the argument made for improved external validity, but another aspect is useful to note here. To keep MC studies manageable, researchers make a conscious decision about which data generating conditions to vary and which to fix. This may mean that conditions that are important were fixed to explore other aspects of model estimation that may be more problematic or more interesting. Meta-analysis can allow the combination of simulation conditions that were held fixed in one study but vary in a second study. This will improve external validity but also may allow for additional interactions between simulation conditions not originally studied by a single MC study. For example, one MC study in their data generation may fix the distribution of residuals to be normally distributed where a second study varied these to be distributed Normal, t, or gamma. These two studies together can now explore a wider range of simulation conditions and combine the study conditions that overlap.

Quasi-random number generation

The quasi-random number generation is an asset when combining MC outcomes to use in a meta-analysis. When the data generation is done well in a single MC study, each replication in the study can be treated as independent. This can greatly increase the sample size for the meta-analysis and allows for the combination of many outcomes from a single MC study in the meta-analysis.

Model based results

A recent exploration of published MC studies by Harwell, Kohli, and Peralta-Torres (2015) found that these studies tend to not have an explicitly stated research design or use inferential methods to explore the results. Instead, studies use descriptive tables which may miss complex interactions that are difficult to see. If an appropriate research design is adopted in an MC study, these are similar to a randomized design and regression models can be useful to explain variation in the MC outcomes (Paxton et al., 2001). MC studies that have already been published are likely not to be updated to fit a statistical model, however using meta-analytic methods can use the results from these studies analyzed inferentially and may find more complex interactions between simulation conditions.

CHALLENGES

There are challenges to be aware of when using research synthesis or meta-analysis methods to combine MC studies. Many of these challenges are not dissimilar from primary studies and meta-analysis studies and may include, nesting, missing data, defining appropriate outcomes, or more. Important challenges will be discussed in more detail below.

Researcher effects

There are likely nesting effects happening at the study or author level. For example, a single researcher may generate data in a particular way and may publish more than one MC study on a given topic. Researchers may also be more interested in certain simulation conditions and tend to vary these more frequently (or tend to fix other simulation conditions). These dependencies would need to be taken into account when doing the meta-analysis and could be done using a multilevel meta-analytic model with an additional level attributed to the study or author level. Deciding whether the study or author level is appropriate would depend on the specific MC literature, number of studies, or number of unique authors.

Missing data

Missing data could be prevalent when coding MC studies. Particular problems may occur when a study does not give enough information to code covariates (i.e. simulation conditions), does not explore specific simulation conditions, or only reports aggregated means of the outcomes of interest. The most difficult to overcome would be the latter problem. If a MC study does not report a table with the all possible simulation conditions, there will be missing data in other areas of the coding sheet. In addition, exploration of study effects will now be limited to aggregated effects for that study. Online supplementary materials may contain the tables of interest otherwise contacting the authors directly would be another way to attempt to overcome this issue.

Defining outcomes

The definition of outcomes of interest may vary across MC studies. For example, if a study explores the type I error rate for a statistical test, the researcher can choose which alpha value to use and may differ across MC studies. These differences should be coded and can be included in the meta-analytic model to attempt to adjust for these differences. This approach may or may not be successful and these studies may need to be excluded or analyzed separately.

Another problem that may arise is the variety of outcomes MC studies report. For instance, looking at bias of an estimator, researchers may report relative bias, simple bias, absolute bias, or others. If the primary study provides enough detail, these may be able to be transformed into the same metric, however in many instances this will likely not be possible. The combination of different outcomes would need to be justified to include in a single model as the metric can be very different.

ANALYSIS

Meta-analytic methods for analyzing outcomes from MC studies can follow directly from meta-analytic methods for analyzing empirical studies (Cooper et al., 2009). However, a few additional aspects should be thought about as well. One is the data generation. Although replications should be independent within a MC study, each researcher generates data differently. This could have implications and create dependencies in the data that should be accounted for. For example, one could use multilevel models treating author or studies as another level of nesting. A second area that should be thought about is related to covariates. The covariates may simply be the same as the simulation conditions used in each MC study. However, it may be possible to recode these into another metric that may be useful. For example, if MC studies explored different distributions when generating the data, the skewness and kurtosis for these distributions may be able to be used instead of the type of distribution used.

Lastly, the variance of the outcomes needs to be known to perform a meta-analysis. Type I error rates are in a proportion metric and theory for the proportions could be used here. However, if bias is used as an outcome, the variance of the bias would need to be known. This may be a challenge currently to overcome this aspect.

CONCLUSION

I believe that combining MC studies using research synthesis and meta-analytic methods can greatly enhance the external validity and study coverage of MC studies, by far the greatest weakness of MC studies. Narrative reviews of MC studies are still important and should also be a component of the meta-analysis of these studies. Research synthesis and meta-analysis in tandem can be important to inform which additional MC studies are warranted for a given statistical model. In addition, the meta-analysis can provice empirical evidence to support the use of a statistical model in certain data conditions, a notion very helpful to applied researchers using statistical models developed my methodologists. Challenges do remain however, particularly related to combining certain MC outcomes (e.g. bias) and knowing the variance of these statistics. Additional work in these areas may further enhance and allow for researchers to more easily combine MC studies.

REFERENCES

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, *25*(24), 4279–4292.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, *42*(3), 237–288.

Guilera, G., Gómez-Benito, J., Hidalgo, M. D., & Sánchez-Meca, J. (2013). Type I error and statistical power of the Mantel-Haenszel procedure for detecting DIF: A meta-analysis. American Psychological Association.

Harwell, M. R. (1992). Summarizing monte carlo results in methodological research. *Journal of Educational and Behavioral Statistics*, *17*(4), 297–313.

Harwell, M. R. (1997). Analyzing the results of monte carlo studies in item response theory. *Educational and Psychological Measurement*, *57*(2), 266–279. Journal Article. doi:10.1177/0013164497057002006

Harwell, M. R. (2003). Summarizing monte carlo results in methodological research: The single-factor, fixed-effects ancova case. *Journal of Educational and Behavioral Statistics*, *28*(1), 45–70.

Harwell, M. R., Kohli, N., & Peralta-Torres, Y. (2015). Monte carlo studies in methodological research in the behavioral sciences.

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing monte carlo results in methodological research: The one-and two-factor fixed effects anova cases. *Journal of Educational and Behavioral Statistics*, *17*(4), 315–339.

Keselman, J. C., Lix, L. M., & Keselman, H. (1996). The analysis of repeated measurements: A quantitative research synthesis. *British Journal of Mathematical and Statistical Psychology*, *49*(2), 275–298.

Kwok, O.-m., West, S. G., & Green, S. B. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A monte carlo study. *Multivariate Behavioral Research*, *42*(3), 557–592.

LeBeau, B. (2016). Impact of serial correlation misspecification with the linear mixed model. *Journal of Modern Applied Statistical Methods*, *15*(1), Article 21. doi: 10.22237/jmasm/1462076400

Lix, L. M., Keselman, J. C., & Keselman, H. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance f test. *Review of Educational Research*, *66*(4), 579–619.

Murphy, D. L., & Pituch, K. A. (2009). The performance of multilevel growth curve models under an autoregressive moving average process. *The Journal of Experimental Education*, *77*(3), 255–284.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte carlo experiments: Design and implementation. *Structural Equation Modeling*, *8*(2), 287–312.

Powell, D. A., & Schafer, W. D. (2001). The robustness of the likelihood ratio chi-square test for structural equation models: A meta-analysis. *Journal of Educational and Behavioral Statistics*, *26*(1), 105–132.

Skrondal, A. (2000). Design and analysis of monte carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, *35*(2), 137–167.

AUTHOR INFORMATION

Name: Brandon LeBeau
Title:
Affiliation: Psychological and Quantitative Foundations, University of Iowa
Address: Iowa City, IA 52245
Country: United States
Email: brandon-lebeau@uiowa.edu
Website: https://brandonlebeau.org