

Review of the Literature on Exact Formulas for Mediation Effects

May 8, 2024

1 Summary and Future Work

As far as I can tell, nobody has done exact asymptotics for mixed-effects mediation analysis. I do want to do some more searching, but the recent literature (e.g., since 2020) presents exact formulas for mediation effects with δ -method uncertainty quantification as novel.

Below, I have summarized a bunch of work on mediation analysis, roughly divided into sections. Section 2 covers work by Samoilenko, Lefebvre and others. Their work looks pretty close to what we're doing. They give exact formulas and δ -method UQ, but no random effects. Section 3 defines mediation effects in terms of partial derivatives of the mean of Y given X . They have done a bit with random effects, but only do Bayesian or bootstrap-based uncertainty quantification. Section 4 uses structural equation modelling (SEM) to do mediation analysis. I'm not very familiar with this methodology, but I tried to summarize what I could. Some of these papers do include random effects, but only seem to do continuous response and mediator. They make heavy use of the simple mediation formulas in the fully continuous setting, so uncertainty quantification is pretty basic. Finally, Section 5 is an assortment of papers. Highlights include Cheng et al. [2021], which gives definitions of direct and indirect effects which we can use, and Zheng and Cadigan [2021], which uses automatic differentiation to get the full Hessian for GLMMs. Zheng and Cadigan also do UQ for the predicted random effects, although I'm still working on understanding exactly how they do this.

1.1 What to do Next

Next week, my plan is to continue the literature search, focusing on SEMs, the UQ methodology proposed by Zheng and Cadigan [2021], and also looking more widely to see if I've missed any other lines of work. It would be helpful for me if you could look at Zheng and Cadigan [2021] to see if you think it has merit. I'm also pretty shaky on SEMs, so it would be nice if you could look at Zigler and Ye [2019] to see how it connects with our methodology.

In terms of advancing our own work, I have finished translating my analysis into R, up to the binary response and mediator. My next steps in this direction for next week are to work out as much as I can for the binary/binary setting with mixed effects. This is going to involve computing the gradient of the mediation effect, and looking for whether `lme4` gives standard errors for the random effects' covariance parameters. If not, then I will look around for other sources.

Once the mixed-effects case is done, I would like to extend our analysis to the direct and indirect effects. Formulas for these are given in Samoilenko et al. [2018] or Cheng et al. [2021]. We should also consider interaction terms between the exposure and mediator in our model for the outcome. I understand that this interaction term arises from causal inference-based concerns, but I'm still fuzzy on the details. Samoilenko and Lefebvre [2021] extend their previous work on fixed-effects models to include interactions. We also need to extend all of our methods to handle predicted group-specific effects, although this makes the uncertainty quantification more challenging, both conceptually and computationally. See, e.g., Booth and Hobert [1998], Florez-Agreda and Cantoni [2019], Zigler and Ye [2019].

2 The Samoilenko and Lefebvre group

This group does essentially what B&B are proposing with exact formulas and δ -method standard errors. In particular, they avoid any requirement for rare events. They also do direct and indirect effects. However, I haven't seen any mention of random-effects/multilevel-models.

These methods are implemented in the R package `ExactMed` [Caubet et al., 2023]. This package does not address mixed-effects models.

2.1 Samoilenko et al. [2018]

Gives exact formulas for direct and indirect mediation effects, as well as δ -method standard errors. Binary outcome, binary mediator. No analytical SE when exposure-mediator interaction is present.

2.2 Samoilenko and Lefebvre [2021]

Extends analytical results of Samoilenko et al. [2018] to handle exposure-mediator interaction term in model for outcome.

2.3 Samoilenko and Lefebvre [2023]

As Samoilenko et al. [2018], but with continuous mediator and binary response. No mixed-effects. Very nice simulation study.

2.4 Caubet et al. [2024]

Extends the work of Samoilenko and Lefebvre [2023] to case-control data.

3 Derivative-Based

Defines mediation effects in terms of derivatives. Originally proposed by Yu et al. [2014], and later extended to multilevel models by Yu and Li [2020]. They use the name “third-variable effect analysis”. Documented extensively in a book by Yu and Li [2022]. Implemented in the R packages `mma` and `mlma` for single level and multilevel models respectively. The latter uses the bootstrap for UQ.

3.1 Maria et al. [2024]

Incorporates random-effects in the derivative-based approach. Mostly does Bayesian inference. Mentions bootstrap.

3.2 Geldhof et al. [2018]

Defines mediation effects as partial derivatives. Does UQ by bootstrap or by a very rough method called MCCI. This MCCI simulates parameter values from a normal distribution parameterized by the estimates, then evaluates the mediation effects on these simulated parameters. This gives a distribution, but it seems to me that they would be better off just using the δ -method.

3.3 Dorette et al. [2022]

Gives exact formulas for mediation effects and δ -method SEs. Decomposes total mediation effect more finely than just direct and indirect (i.e. path-specific effects). No mixed-effects.

4 Structural Equation Modelling

There are a bunch of papers which use SEM to do mediation analysis. Many of them include multilevel/mixed-effects models. I’m not very familiar with SEM methodology, but it does seem to be connected with our work. See Valeri and VanderWeele [2013] for a review and comparison with the causal-inference framework. Generally limited to linear models (i.e. continuous response and mediator), where the mediation effects’ formulas are very simple. They seem to only focus on indirect effects¹.

There does seem to be some work on applying mixed-effects methodology to SEMs. I’m not sure yet if this has made its way to the SEM-based mediation analysis literature. I’ll look into this next week.

¹There is a short paragraph in Zitzmann and Helm [2021] near the top of column 1 on page 532 which says “the direct effect is only of little interest”. They don’t give a reference for this.

4.1 Zigler and Ye [2019]

Review of multilevel mediation analysis using SEM. Does a Monte Carlo study comparing different approaches. UQ is based on Bauer et al. [2006].

4.2 Bauer et al. [2006]

Does mixed-effects mediation analysis based on SEMs. Only handles continuous outcome and mediator. Mediation effects are simple enough that variances are computed directly (in terms of asymptotic covariances of the regression parameter estimates). Fits the two regression models (for M and Y) simultaneously by concatenating the response vectors and multiplying covariates by indicators.

4.3 Preacher et al. [2010]

Does mediation analysis using structural equation models. Does UQ by bootstrap or based on the exact distribution of the product of two dependent normals [MacKinnon et al., 2007].

5 Others

5.1 Kenny et al. [2003]

Gives a simplistic definition of mediation effects for continuous models. Does UQ based on exact distribution of mediation effects obtained as a transformation of Gaussians. A nice approach, but only really works because of the simple setting.

There is a book by MacKinnon [2017] on mediation analysis, which has a whole chapter called “Multilevel Mediation Models”. However, it is mostly based on the Kenny et al. [2003] and small extensions.

5.2 Ritz and Spiegelman [2004]

Points-out that, in mixed-effects models, the regression structure for Y only holds conditional on the random effects. Thus, it’s hard to give a marginal interpretation of the regression coefficients. They give some conditions under which the conditional mean structure is preserved (in a particular sense) after marginalizing away the random effects. See, e.g., Neuhaus et al. [1991] for more discussion of the problem being addressed.

I think I’m missing something here, because it seems obvious to me that there will be a difference between the parameters in an assumed regression model for the marginal mean of Y and the conditional mean of Y given the random effects. Regardless, I don’t think this one is particularly relevant to us. We have a conditional regression model and don’t make any effort to model the marginal mean of Y using regression.

5.3 VanderWeele and Vansteelandt [2010]

Standard reference for mediation analysis with binary outcome. Focuses on the causal inference aspects of the problem. Makes heavy use of the “rare outcome” assumption. Resulting model is sufficiently simple that UQ is easy.

5.4 Pearl [2012]

Presents the so-called “mediation formula”. A set of identities for the total, direct and indirect mediation effects based on counterfactuals, but expressed in terms of conditional expectations. Effects are on risk-difference scale.

5.5 Nevo et al. [2017]

Does inference for mediation effects defined using the “difference method”. Briefly, fit models to predict Y using X with or without M , then check how much the coefficient on X changes. They apply their analysis to continuous, binary (log- and logit-link), and survival data (Cox model).

Inference is done with the δ -method. They need the covariance between coefficient estimates from different models (i.e. with and without M). Since this doesn’t come directly from a standard analysis, they propose a “data duplication algorithm”.

No mixed effects. Not obviously relevant.

5.6 Cheng et al. [2021]

Defines mediation effects (indirect and direct) as differences between transformed conditional expectations. This facilitates use with GLM link functions. Does continuous and binary outcome and mediator using GEEs. Does UQ with δ -method.

Includes a Monte Carlo study comparing δ -method with bootstrap. Finds that pretty large samples are required for δ -method UQ to be accurate, although bootstrap works with somewhat smaller samples.

No mixed-effects.

Does δ -method UQ for GLMs (more precisely, GEEs) with exact formulas for mediation effects. These mediation effects (indirect and direct effects) are defined as differences between transformed conditional expectations.

5.7 Zheng and Cadigan [2021]

Does joint uncertainty quantification for the parameters and random effects. Based on maximizing the Laplace Approximation to the marginal likelihood. Gives δ -method SEs with gradients from automatic differentiation.

Based on the R package `TMB`, which does automatic differentiation [Kristensen et al., 2016]. There is a related package, `glmmTMB`, which is specifically designed to apply automatic differentiation methods to GLMMs [Brooks et al.,

2017] (their paper focuses on zero-inflated count GLMMs, but their methodology applies more generally). Zheng and Cadigan [2021] raises some concerns about the two TMB packages, and presents a solution. See Skaug and Fournier [2006] for more discussion of automatic differentiation in GLMM fitting.

References

- Daniel J. Bauer, Kristopher J. Preacher, and Karen M. Gil. Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. *Psychological Methods*, 11(2), 2006.
- James G. Booth and James P. Hobert. Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association*, 93(441), 1998.
- Mollie E. Brooks, Kasper Kristensen, Koen J. van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Mächler, and Benjamin M. Bolker. `glmmTMB` balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9(2), 2017.
- Miguel Caubet, Mariia Samoilenko, Jesse Gervais, and Geneviève Lefebvre. *ExactMed: Exact Mediation Analysis for Binary Outcomes*, 2023. URL <https://caubm.github.io/ExactMed/>. R package version 0.2.0.9000.
- Miguel Caubet, Kevin L’Espérance, Anita Koushik, and Geneviève Lefebvre. An empirical evaluation of approximate and exact regression-based causal mediation approaches for a binary outcome and a continuous or a binary mediator for case-control study designs. *BMC Medical Research Methodology*, 24(1), 2024.
- Chao Cheng, Donna Spiegelman, and Fan Li. Estimating the natural indirect effect and the mediation proportion via the product method. *BMC Medical Research Methodology*, 21(1), 2021.
- Marco Doretti, Martina Raggi, and Elena Stanghellini. Exact parametric causal mediation analysis for a binary outcome with a binary mediator. *Statistical Methods & Applications*, 31(1), 2022.
- Daniel Florez-Agreda and Eva Cantoni. Bootstrap estimation of uncertainty in prediction for generalized linear mixed models. *Computational Statistics and Data Analysis*, 130, 2019.
- G. John Geldhof, Katherine P. Anthony, James P. Selig, and Carolyn A. Mendez-Luck. Accommodating binary and count variables in mediation: A case for conditional indirect effects. *International Journal of Behavioral Development*, 42(2), 2018.

- David A. Kenny, Josephine D. Korchmaros, and Niall Bolger. Lower level mediation in multilevel models. *Psychological Methods*, 8(2), 2003.
- Kasper Kristensen, Anders Nielsen, Casper W. Berg, Hans Skaug, and Bradley M. Bell. TMB: automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70(5), 2016.
- David P. MacKinnon. *Introduction to statistical mediation analysis*. Routledge, 2017.
- David P. MacKinnon, Matthew S. Fritz, Jason Williams, and Chondra M. Lockwood. Distribution of the product confidence limits for the indirect effect: program PRODCLIN. *Behaviour Research Methods*, 39(3), 2007.
- Chiara Di Maria, Claudio Rubino, and Alessandro Albano. The derivative-based approach to nonlinear mediation models: insights and applications. *Quality and Quantity*, 2024.
- J. M. Neuhaus, J. D. Kalbfleisch, and W. W. Hauck. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review*, 59(1), 1991.
- Daniel Nevo, Xiaomei Liao, and Donna Spiegelman. Estimation and inference for the mediation proportion. *International Journal of Biostatistics*, 13(2), 2017.
- Judea Pearl. The causal mediation formula - a guide to the assessment of pathways and mechanisms. *Prevention Science*, 13(4), 2012.
- Kristopher J. Preacher, Michael J. Zyphur, and Zhen Zhang. A general multi-level SEM framework for assessing multilevel mediation. *Psychological Methods*, 15(3), 2010.
- John Ritz and Donna Spiegelman. Equivalence of conditional and marginal regression models for clustered and longitudinal data. *Statistical Methods in Medical Research*, 13(4), 2004.
- Mariia Samoilenko and Geneviève Lefebvre. Parametric-regression-based causal mediation analysis of binary outcomes and binary mediators: moving beyond the rareness or commonness of the outcome. *American Journal of Epidemiology*, 190(9), 2021.
- Mariia Samoilenko and Geneviève Lefebvre. An exact regression-based approach for the estimation of natural direct and indirect effects with a binary outcome and a continuous mediator. *Statistics in Medicine*, 42(3), 2023.
- Mariia Samoilenko, Lucie Blais, and Geneviève Lefebvre. Comparing logistic and log-binomial models for causal mediation analyses of binary mediators and rare binary outcomes: evidence to support cross-checking of mediation results in practice. *Observational Studies*, 4(1), 2018.

- Hans J. Skaug and David A. Fournier. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. *Computational Statistics and Data Analysis*, 51, 2006.
- Linda Valeri and Tyler J. VanderWeele. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions with SAS and SPSS macros. *Psychological Methods*, 18(2), 2013.
- Tyler J. VanderWeele and Stijn Vansteelandt. Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology*, 172(12), 2010.
- Qingzhao Yu and Bin Li. Third-variable effect analysis with multilevel additive models. *PLoS ONE*, 15(10), 2020.
- Qingzhao Yu and Bin Li. *Statistical methods for mediation, confounding and moderation analysis using R and SAS*. CRC Press, 2022.
- Qingzhao Yu, Ying Fan, and Xiaocheng Wu. General multiple mediation analysis with an application to explore racial disparities in breast cancer survival. *Journal of Biometrics & Biostatistics*, 5(2), 2014.
- Nan Zheng and Noel Cadigan. Frequentist delta-variance approximations with mixed-effects models and TMB. *Computational Statistics and Data Analysis*, 160, 2021.
- Christina K. Zigler and Feifei Ye. A comparison of multilevel mediation modeling methods: recommendations for applied researchers. *Multivariate Behavioral Research*, 54(3), 2019.
- Steffen Zitzmann and Christoph Helm. Multilevel analysis of mediation, moderation, and nonlinear effects in small samples, using expected a posteriori estimates of factor scores. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(4), 2021.