An Update on William's Work

June 12, 2024

Preface

I have been working on three main areas: a literature review of estimation and UQ for multilevel mediation, implementing the UQ developed by B & B for the estimated total effect, and expanding their development to address direct and indirect effects. In this document, I attempt to consolidate this disparate work in one place. Unfortunately, since my implementation was written in RMarkdown, I only overview that work here and refer to an external PDF.

In the interest of minimizing the amount of re-writing necessary, I have organized this document into chapters.

Contents

Chapter 1

Literature Review

1.1 Summary and Future Work

1.1.1 May 23

I'm just finishing-up my code for the binary response, binary mediator, mixed effects case. My analysis is running, but the output doesn't look right, so I need to do some debugging.

I now better understand the uncertainty quantification method used by the Imai group. See Section 1.5.4 for details. Briefly, they do a Monte Carlo analog of the δ-method. If the target of inference is a function of some model parameters, they start by estimating the sampling distribution of our estimator for those model parameters. They then simulate from this estimated sampling distribution and evaluate the function of interest on each simulated observation. This produces a sample from the pushforward measure induced on the target of inference by our estimated sampling distribution for the model parameters. It's not completely clear to me how they generate confidence intervals from the Monte Carlo samples, but they call their method "quasi-Bayesian", so I suspect it's just percentile intervals.

Incidentally, if we could do exact (i.e. non-asymptotic) distribution theory for the model parameter (e.g. ordinary linear regression), then Imai's method is equivalent to the parametric bootstrap. They just skip the generation of datasets and work directly with the bootstrap distribution of the estimator.

A thought on how to frame the definition of mediation effects on different scales: the mediation formula of Pearl [2012] identifies causal direct/indirect effects with specific functions of conditional expectations of the outcome and mediator. This makes it natural to work on the risk-difference scale, as done by the Imai group (see Section 1.5). However, if we prefer the risk-ratio or odds-ratio, we can just express our target quantities in terms of the functions identified by Pearl [2012]. There may or may not be simplification that leads to interpretable formulas, but that's not the point. You can define mediation effects on whatever scale you like, then we will give you estimates and standard

errors, irrespective of how messy the formulas may be (they will all be hidden in the code anyway).

What to do Next

I think that at this week's meeting, we should plan out what we want to put in the paper. I propose that we do total, direct and indirect effects on risk difference, risk ratio and odds ratio scales, but stick to fixed-effects (i.e. avoid predicting group-level effects). Based on other work in the literature. I think that's enough for a paper, provided that we do a nice Monte Carlo study comparing our UQ with other methods, such as Imai's quasi-Bayesian method, the weighting method of Lange et al. [2012] and the imputation method of Vansteelandt et al. [2012]. All of these methods are implemented to varying extents in `R` packages, `mediation` for the former and `medflex` for the latter two.

Prediction of group-level effects barely comes up in the literature. The `mediation` package from Imai's group does it, but with very little discussion. They give a few sentences in the package vignette saying they can do group-level effect prediction, but don't give any further discussion or references. I had to read their source code to know what method they're using. Prediction of group-level mediation effects seems like an untapped area that we are well-positioned to develop.

1.1.2   May 16

This week, I added work by Imai and colleagues, as well as a few general references. See Section 1.5 for the Imai group. There is also a method proposed by Lange et al. [2012] based on inverse probability weighting to estimate unobserved counterfactuals. Both of these methods are included in the extensive Monte Carlo study of Samoilenko and Lefebvre [2023] and found to be comparable to inference for mediation effects based on the δ-method. However, the method of Imai et al. [2010a] is computationally intensive, and that of Lange et al. [2012] is quite different from our approach of combining estimates from two regression models.

I also found an `R` package, `merDeriv` [Wang and Merkle, 2018] which computes the full asymptotic covariance matrix for a GLMM fit using `lme4`. This will save me a lot of time in implementing the δ-method uncertainty quantification for mixed-effects models.

What to do next

I'm still not totally confident about the SEM literature, but otherwise I don't think anyone has done exact asymptotic theory for multilevel mediation analysis. I feel ready to start working on the mixed-effects version of our analysis. I will continue reading, but I think it makes sense to start spending some time writing code.

There are some extensions to the methodology outlined by B & B that I think we should consider. First, we will want to include direct and indirect effects in addition to our total effects. We can follow Samoilenko and Lefebvre [2023] for this, although we will also need to give some though to how we incorporate random effects into the definitions. It would also be nice to explicitly estimate the "mediation proportion". For continuous response, this is the indirect effect divided by the total effect. It's less clear to me what definition to use for multiplicative effects (e.g. effects on odds-ratio scale).

Second, it is common to include an interaction effect between the exposure and mediator in our model for the outcome[1]. This isn't conceptually any harder, it just adds another component to our formulas.

I think that we could address just the above two points and, with some nice Monte Carlo studies, call it a paper. This would resemble the work by Samoilenko and Lefebvre (e.g. Samoilenko and Lefebvre, 2023). After that, there are some more dramatic extensions we can consider.

One extension that we really should look at is predicted cluster-specific effects. The uncertainty quantification here is subtle; see, e.g., Skrondal and Rabe-Hesketh [2009], Booth and Hobert [1998]. The Imai group does give cluster-specific mediation effects, but only under their restricted framework. Otherwise, this seems to still be an open problem

Finally (for now), I think we should eventually look at sensitivity analysis. There are a bunch of assumptions required for us to give a causal interpretation to our results. My understanding is that sensitivity analysis lets us put a bound on the extent to which these assumptions can be violated without invalidating our conclusions. I'm still looking into how this works.

1.1.3 May 9

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 1.2 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 1.3 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 1.4 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

---

[1]There is no interaction effect in our trust study because a preliminary analysis found that it didn't have a significant effect on the outcome.

quantification is pretty basic. Finally, Section 1.8 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.

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].

1.2 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.

1.2.1 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.

1.2.2 Samoilenko and Lefebvre [2021]

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

1.2.3 Samoilenko and Lefebvre [2023]

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

Compares their approach to those of the Imai group (see Section 1.5), as well as Lange et al. [2012]. Found that the δ-method based approach proposed by the Samoilenko and Lefebvre group gives comparable performance with its computationally intensive alternatives.

1.2.4 Caubet et al. [2024]

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

1.3 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.

1.3.1 Maria et al. [2024]

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

1.3.2 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 is the same UQ method used by the Imai group (see Section 1.5).

1.3.3 Doretti 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.

1.4 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 effects[2].

There does seem to be some work on applying mixed-effects methodology to SEMs. See, e.g., Rabe-Hesketh et al. [2004] for early work in this direction.

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.

1.4.1 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].

1.4.2 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.

1.4.3 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].

1.5 The Imai Group

This group focuses mainly on the assumptions required to link regression-based mediation analysis with causal inference. They give definitions of mediation effects which closely resemble the "mediation formula" [Pearl, 2012], but only for differences of conditional expectation (i.e. the "risk-difference scale"; see also Samoilenko and Lefebvre, 2023). Their approach to UQ is very computational. They recommend either the bootstrap or a simulation-based approach outlined in King et al. [2000]. Briefly, simulate parameters from their fitted

---

[2]There 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.

sampling distributions, then evaluate the quantity of interest on each sample and observe the resulting distribution. I think this is connected to empirical Bayes methodology, but I haven't looked into it extensively. They use the name "quasi-Bayesian confidence intervals" in their `R` package.

Their methods are implemented in the `R` package `mediation` [Tingley et al., 2014]. They can handle mixed-effects models, including prediction of group-level effects.

1.5.1 Imai et al. [2010a]

This paper presents an UQ algorithm for estimating mediation effects defined as differences of expected values. The algorithm closely resembles the bootstrap, but is slightly different. Specifically, parameters are repeatedly sampled from the fitted sampling distributions of their estimators, then mediation effects are estimated using the sampled parameter estimates [see, e.g., King et al., 2000]. They also give an alternative method which just uses the bootstrap.

Refers to the causal inference theory from Imai et al. [2010b] to justify identification of counterfactuals with functions of conditional expectations.

1.5.2 Imai et al. [2010b]

Focuses on estimating the indirect effect for a continuous response. Defines this effect as a difference of counterfactuals and identifies this difference with a conditional expectation in the same manner as the mediation formula given by Pearl [2012].

Also gives a "non-parametric estimator" for categorical mediator based on a contingency table based approach. The idea is to partition the data based on the mediator, estimate the mediation effect within each group with a difference of averages, then re-combine by averaging.

UQ is given either by the δ-method or by the bootstrap. Some discussion of sensitivity analysis.

1.5.3 Imai et al. [2011]

Very similar to Imai et al. [2010a] and Imai et al. [2010b]. Focuses more on causal inference interpretation and applications to political science. Doesn't appear to advance the methodology.

1.5.4 Tingley et al. [2014]

This paper overviews the `mediation` package, which implements the methods developed by this group. Can handle GLMMs, but only operates on differences in the expected value of the outcome (i.e. risk-difference scale for binary outcome). Does UQ for parameters **and predicted random effects**, but doesn't go into detail on how. I'm guessing that it just follows the Monte Carlo method given by Imai et al. [2010a].

Package last updated 2019-10-08

UQ for Group-Level Effects

The Tingley et al. [2014] paper discusses inference for predicted group-level effects, but doesn't go into detail. As far as I can tell, this is also the only paper from the Imai group which discusses multilevel mediation analysis. Worse still, they give very few references.

After digging into the source code for the `mediation` package, I found what they do. First, they get predicted group-wise random effects from `lme4`. Second, it turns out that `lme4` can give a conditional covariance matrix for the random effects "conditional on the estimates of the model parameters and on the data". For now, just note that the full covariance matrix of the random effects is block diagonal, with one block per group. The `mediation` package uses these group-wise conditional covariance matrices, together with the predicted group-wise means, and simulates random effects for each group (independently, from normal distributions with the corresponding moments). The package then proceeds with its "quasi-Bayesian" inference.

Just a brief note on the conditional covariances given by `lme4`. Initially, when I was reading this portion of the documentation, I got excited that we might not need the `merDeriv` package to get standard errors for the random effects' covariance parameters. That's definitely not what I found. These are standard errors for the predicted random effects (conditional on the observed data), not standard errors for the estimators of any model parameters. In particular, the former can be thought of as K separate covariance matrices, one per group, whereas the latter is a single matrix representing estimation uncertainty based on the entire dataset.

Unfortunately, no references are given in the `lme4` documentation, but it is fairly clear what is being done. The predicted random effect within a group is the mode of the conditional distribution of the random effects given the observed data from that group (alternatively, the mean; although `lme4` uses the mode). The prediction standard error given by `lme4` is the covariance of this same conditional distribution.

1.6   The `SPSS` Macro `MLmed`

This macro appears to be the standard way to do multilevel mediation analysis in `SPSS`. As such, it is likely to be used extensively by applied researchers, and we should understand it. The macro can only handle continuous outcomes and mediators.

1.6.1   Rockwood [2019]

User manual for the `MLmed` macro. Can only handle continuous outcomes and mediators. UQ is either by the bootstrap or the "Monte Carlo method". See Preacher and Selig [2012] or Williams and MacKinnon [2008].

1.6.2 Hayes and Rockwood [2020]

A recommended reference for the methodology used in the macro.

1.6.3 Rockwood and Hayes [2017]

A conference poster presenting the macro. Popular reference for people looking to cite the macro.

1.7 Applications

It doesn't hurt for us to point to what people are doing in the applied literature. Here is some of what shows up when I search "multilevel mediation analysis" on Google Scholar.

1.7.1 Veas et al. [2019]

Does multilevel mediation analysis with continuous outcome and mediator. UQ is done with "Monte Carlo confidence intervals". Uses an `SPSS` macro called `MLmed`. See Section 1.6.

1.7.2 Arain et al. [2021]

It looks like their variables are all ordinal (1-6 or 1-8 Likert scale). I'm not sure how they fit this into the mediation analysis framework. Moving on, they use `Mplus` to do their analysis via "maximum likelihood". Not much detail given.

1.7.3 Hwang and Hopkins [2012]

Follows an earlier edition of MacKinnon [2017], as well as the very early work of Baron and Kenny [1986]. As in Arain et al. [2021], variables are ordinal/Likert. Analysis is done in `SPSS`, although few details are given. Note that this work was published before the `SPSS` macro of Rockwood [2019].

1.7.4 Rijnhart et al. [2021]

Review of applied papers using mediation analysis. Read and document this!

1.8 Others

1.8.1 Krull and MacKinnon [2001]

Throughout some parts of the literature on multilevel mediation analysis, you will see people talk about, e.g., the 2-1-1 model. This means that the exposure is measured at "level 2" (group-level), while the mediator and outcome are measured at "level 1" (individual-level). That is, in this model we would include

a random effect for the mediator, but not the exposure. Other examples of such models include 2-2-1 and 1-1-1. As far as I can tell, this notation was proposed by Krull and MacKinnon [2001].

This paper also does δ-method uncertainty quantification for the very simple mediation effects that people were using at that time (i.e. a product of coefficients). In fact, this approach for testing the presence/absence of mediation effects dates back to Sobel [1982].

1.8.2   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.

1.8.3   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.

1.8.4   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.

1.8.5   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.

1.8.6   Lange et al. [2012]

Proposes estimating mediation effects using inverse probability weighting. Duplicates the observed data and uses weighting to estimate unobserved counterfactual outcomes. Referred to as the "weighting method". Does give analytical (asymptotic) standard errors based on the sandwich variance estimator that arises from the theory of generalized estimating equations. See their "Web Appendix 3" for more details. They also suggest using the bootstrap.

Implemented in the `R` package `medflex` [Steen et al., 2017] alongside the imputation method of Vansteelandt et al. [2012].

1.8.7   Vansteelandt et al. [2012]

Proposes estimating mediation effects using regression-based imputation. Recall that the indirect and direct mediation effects depend on unobservable nested counterfactuals of the form $\phi(x, x') := Y(x, M(x'))$, where $x \neq x'$. Vansteelandt et al. [2012] proposes to estimate these counterfactuals directly by first fitting a regression model for the dependence of ϕ on x and x' using configurations of these covariates (i.e. the exposure) observed in the data. We then use this regression model to predict the value of ϕ at unobserved configurations.

Caution: This whole method is based on extrapolating from a regression model into a region where we, by definition, have no information. The usual warnings apply, both philosophical and practical (e.g. don't include high-order polynomial terms in your regression model).

Implemented in the `R` package `medflex` [Steen et al., 2017] alongside the weighting method of Lange et al. [2012].

1.8.8   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.

1.8.9   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.

1.8.10 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.

1.8.11 Wang and Merkle [2018]

Standard reference for the `merDeriv` package in `R`. This is a small package that computes the Hessian of the observed data likelihood with respect to all parameters (including RE covariance parameters) for LMMs and GLMMs fit using `lme4`. In particular, use the `vcov` function with `full=TRUE` to get the asymptotic covariance matrix of the parameters (i.e. the inverse of the negative Hessian of the log-likelihood). You can also choose your parameterization for the RE parameters from variances, SDs or terms in the Cholesky decomposition. The last one is referred to as `theta` in `lme4` and is used extensively therein.

Chapter 2

Implementation

I don't have a whole lot to say here. I implemented the δ-method based uncertainty quantification developed by B & B for the total mediation effect. Please see the RMarkdown document and accompanying PDF called "William Analysis.Rmd" and "William-Analysis.pdf" for discussion and results.

My findings for the fixed-effects models look pretty reasonable. However, the mixed-effects setting is much less nice. I'm still exploring why this is happening. I hope to have more to say here soon.

Chapter 3

Extension to Other Mediation Effects

In this chapter, we develop definitions and formulas for the direct and indirect mediation effects to complement the total effect given by B & B. We follow the work of Pearl [2012] and the Samoilenko and Lefebvre group (e.g., Samoilenko and Lefebvre, 2023).

3.1 Expected Nested Counterfactuals

To start, we define the nested counterfactual $Y(x, M(x'))$ as the value that Y would assume when X is set to x and M is set to whatever value it would have assumed if we had set X to x'. Pearl [2012] identifies the expected value of this nested counterfactual with the following expression based on conditional expectations:

$$\mathbb{E}Y(x, M(x')) = \int \mathbb{E}(Y|M = m, X = x)\mathbb{P}(M = dm|X = x') \qquad (3.1)$$

Focusing now on the case with binary response and binary mediator, 3.1 becomes

$$\mathbb{E}Y(x, M(x')) = \mathbb{P}(Y = 1|M = 1, X = x)\mathbb{P}(M = 1|X = x')+ \\ \mathbb{P}(Y = 1|M = 0, X = x)\mathbb{P}(M = 0|X = x') \quad (3.2)$$

Consider using logistic regression models for Y and M. Write $\text{logit}(\mathbb{E}(Y|M = m, X = x)) = \beta_0 + \beta_m m + \beta_x x$ and $\text{logit}(\mathbb{E}(M|X = x)) = \alpha_0 + \alpha_x x$. Then (3.2) becomes

$$\mathbb{E}Y(x, M(x')) = \frac{1}{1 + \exp(-\beta_0 - \beta_m - \beta_x x)} \cdot \frac{1}{1 + \exp(-\alpha_0 - \alpha_x x')}+ \\ \frac{1}{1 + \exp(-\beta_0 - \beta_x x)} \cdot \frac{1}{1 + \exp(\alpha_0 + \alpha_x x')} \quad (3.3)$$

Equation (3.3) holds for logistic regression with fixed-effects only. If we instead use mixed-effects logistic regressions for Y and M, then (3.1) and (3.2) still hold, but (3.3) must be modified. For the mixed-effects models, first write $V = (V_0, V_m, V_x) \sim \mathrm{N}(0, \Sigma_V)$ and $U = (U_0, U_x) \sim \mathrm{N}(0, \Sigma_U)$ for the random-effects in our models for Y and M respectively. Next, write $\mathrm{logit}(\mathbb{E}(Y|V, M = m, X = x)) = (\beta_0 + V_0) + (\beta_m + V_m)m + (\beta_x + V_x)x$ and $\mathrm{logit}(\mathbb{E}(M|U, X = x)) = (\alpha_0 + U_0) + (\alpha_x + U_x)x$. Returning now to identification of the expected counterfactual for Y, we get

$$\mathbb{E}Y(x, M(x')) = \left[\int \mathbb{P}(Y = 1|V = v, M = 1, X = x)\mathbb{P}(V = dv)\right] \cdot \qquad (3.4)$$

$$\left[\int \mathbb{P}(M = 1|U = u, X = x')\mathbb{P}(U = du)\right] + \qquad (3.5)$$

$$\left[\int \mathbb{P}(Y = 1|V = v, M = 0, X = x)\mathbb{P}(V = dv)\right] \cdot \qquad (3.6)$$

$$\left[\int \mathbb{P}(M = 0|U = u, X = x')\mathbb{P}(U = du)\right] \qquad (3.7)$$

and, in the logistic regression context,

$$\mathbb{E}Y(x, M(x')) = \left[\int \frac{\phi(v; 0, \Sigma_V)}{1 + \exp(-(\beta_0 + v_0) - (\beta_m + v_m) - (\beta_x + v_x)x)}dv\right] \cdot$$
$$\qquad (3.8)$$
$$\left[\int \frac{\phi(u; 0, \Sigma_U)}{1 + \exp(-(\alpha_0 + u_0) - (\alpha_x + u_x)x')}du\right] +$$
$$\left[\int \frac{\phi(v; 0, \Sigma_V)}{1 + \exp(-(\beta_0 + v_0) - (\beta_x + v_x)x)}dv\right] \cdot$$
$$\left[\int \frac{\phi(u; 0, \Sigma_U)}{1 + \exp((\alpha_0 + u_0) + (\alpha_x + u_x)x')}du\right],$$

where $\phi(.; \mu, \Sigma)$ is the multivariate normal density with mean μ and covariance matrix Σ.

Note that the four integrals in (3.8) are all multivariate, but can be transformed to univariate integrals by suitable changes of variables. To this end, write $\eta = \alpha_0 + \alpha_x x$ and $\zeta = \beta_0 + \beta_x x$ for two linear predictors (note that ζ does not contain β_m), and $\gamma_\Sigma^2(c_1, \ldots, c_r) = (c_1, \ldots, c_r)\Sigma(c_1, \ldots, c_r)^T$, where Σ is an r-by-r covariance matrix. We will generally set $\Sigma = \Sigma_V$ or $\Sigma = \Sigma_U$, in which case we write γ_V^2 or γ_U^2 respectively. We now define the function ψ as a template for the four integrals in (3.8).

$$\psi(\mu, \sigma^2) := \int \frac{\phi(z; 0, 1)}{1 + \exp(-\mu - \sigma z)}dz.$$

Note that ψ is a univariate integral, so we can expect it to be well-approximated by routine numerical quadrature techniques. We now re-write (3.8) in terms of

ψ as follows,

$$\mathbb{E}Y(x, M(x')) = \psi(\zeta + \beta_m, \gamma_V^2(1, 1, x)) \cdot \psi(\eta, \gamma_U^2(1, x)) + \tag{3.9}$$
$$\psi(\zeta, \gamma_V^2(1, 0, x)) \cdot \psi(-\eta, \gamma_U^2(1, x)).$$

3.2 Mediation Effects

Denote the expected nested counterfactual defined in (3.1) by $\mathcal{Y}(x, x') = \mathbb{E}Y(x, M(x'))$. We can define the various mediation effects in terms of expected counterfactuals. Note that mediation effects for a binary outcome are commonly defined on three different scales: risk difference, risk ratio and odds ratio. Table 3.1 gives all such definitions explicitly. In the rest of this section, we outline the processes of estimation and uncertainty quantification for the mediation effects defined in Table 3.1.

Table 3.1: Definitions of various mediation effects; x and x' denote different values of the exposure.

Risk Difference	Total Effect	$\mathcal{Y}(x, x) - \mathcal{Y}(x', x')$
	Direct Effect	$\mathcal{Y}(x, x') - \mathcal{Y}(x', x')$
	Indirect Effect	$\mathcal{Y}(x, x) - \mathcal{Y}(x, x')$
Risk Ratio	Total Effect	$\mathcal{Y}(x, x)/\mathcal{Y}(x', x')$
	Direct Effect	$\mathcal{Y}(x, x')/\mathcal{Y}(x', x')$
	Indirect Effect	$\mathcal{Y}(x, x)/\mathcal{Y}(x, x')$
Odds Ratio	Total Effect	$\frac{\mathcal{Y}(x,x)}{1-\mathcal{Y}(x,x)} \left/ \frac{\mathcal{Y}(x',x')}{1-\mathcal{Y}(x',x')} \right.$
	Direct Effect	$\frac{\mathcal{Y}(x,x')}{1-\mathcal{Y}(x,x')} \left/ \frac{\mathcal{Y}(x',x')}{1-\mathcal{Y}(x',x')} \right.$
	Indirect Effect	$\frac{\mathcal{Y}(x,x)}{1-\mathcal{Y}(x,x)} \left/ \frac{\mathcal{Y}(x,x')}{1-\mathcal{Y}(x,x')} \right.$

3.2.1 Estimation

Mediation effects are all defined in terms of the expected nested counterfactuals, $\mathcal{Y}(x, x')$. As such, the estimation of any mediation effect given in Table 3.1 centers around estimation of $\mathcal{Y}$. To this end, write θ for all parameters upon which $\mathcal{Y}$ depends. That is, θ contains both sets of regression coefficients, $\beta_0, \beta_m, \beta_x$ and α_0, α_x, as well as both sets of covariance parameters. For consistency with B & B, we parameterize these as τ_0, τ_m, τ_x for the standard deviations of V_0, V_m, V_x, and $\tau_{0,m}, \tau_{0,x}, \tau_{m,x}$ for the corresponding correlations[1]. Similarly, we use σ_0, σ_x for the standard deviations of U_0, U_x, and $\sigma_{0,x}$ for their correlation. We are now

---

[1]While our notation doesn't match that given by B & B, parameterizing in terms of the standard deviations and correlations does. Alternative choices include the variances and covariances, or the unique components of the Cholesky factorizations of Σ_V and Σ_U [Wang and Merkle, 2018].

equipped to write-out θ in full. The order of parameters is chosen to match my code (and to avoid refactoring thereof).

$$\theta = (\alpha_0, \alpha_x, \sigma_0, \sigma_x, \sigma_{0,x}, \beta_0, \beta_m, \beta_x, \tau_0, \tau_m, \tau_x, \tau_{0,m}, \tau_{0,x}, \tau_{m,x})$$

In order to estimate θ, we fit two mixed-effects logistic regression models. The first model predicts the mediator, M, using the exposure, X, while the second predicts the outcome, Y, using M and X. Both models also contain intercepts. We include random effects for all regression coefficients in both models. Either or both models may also contain one or more confounders, $\mathbf{C}$, although we assign these confounders only fixed effects. Following the algorithm used by the `lme4` package in R [Bates et al., 2015, Walker et al., 2023], we estimate the parameters in both regression models by maximizing an approximation to the marginal likelihood of the response (M or Y). This approximation is obtained using Laplace's Method, with a penalized iteratively re-weighted least squares algorithm maximizing to maximize the joint likelihood of the observed data and the random effects. Applying this procedure to both of our regression models yields estimates for all components of θ.

Combining our estimate of θ with Equation 3.8 from Section 3.1 and the formulas given in Table 3.1, we are now able to estimate all mediation effects on the three scales we consider. It remains however, to address uncertainty quantification.

3.2.2 Uncertainty Quantification

As mentioned above, all mediation effects are defined in terms of expected nested counterfactuals $\mathcal{Y}$, which themselves depend on the parameter vector θ. Quantification of uncertainty for our estimator of a mediation effect thus proceeds in three stages. First, we obtain a covariance matrix for our estimator of θ. Second, we apply the δ-method to get a joint covariance matrix for all values of $\mathcal{Y}$ required to evaluate the mediation effect of interest. Finally, we use the δ-method again to translate uncertainty in the $\mathcal{Y}$s to uncertainty in the mediation effect. In fact, we can obtain the covariance matrix for all three mediation effects defined on a particular scale (e.g., risk ratio), or, indeed, between all nine mediation effects given in Table 3.1. Note that the variances and covariances discussed above are asymptotic.

We now proceed with the three steps discussed above, starting with an asymptotic covariance matrix for θ. For now, we treat the two regression models as independent. That is, we assume that the covariance between all parameter estimators in our M model and all parameter estimators in our Y model are zero. This is likely not true; call it a working assumption. See Bauer et al. [2006] for a method to model this inter-model dependence. This step is fairly straightforward, since both regression models are fit using (approximate) maximum likelihood. We simply compute the negative Hessian of the marginal log-likelihood for the observed data in both models, and stack the results into a block-diagonal matrix. Call the resulting covariance matrix $\hat{\Sigma}_\theta$. This computation is performed by the `merDeriv` package in R [Wang and Merkle, 2018].

To quantify uncertainty in $\mathcal{Y}$ based on our estimator for θ, we use the δ-method [see, e.g., Chapter 3 of van der Vaart, 1998]. Briefly, given the asymptotic covariance of our estimator for θ, $\sqrt{n}(\hat{\theta} - \theta) \rightsquigarrow N(0, \Sigma)$, the asymptotic covariance of some function of our estimator can be obtained by differentiating that function and multiplying by the limiting distribution. That is, $\sqrt{n}[f(\hat{\theta}) - f(\theta)] \rightsquigarrow \nabla f(\theta) N(0, \Sigma) \stackrel{d}{=} N(0, \nabla f(\theta) \Sigma \nabla f(\theta)^T)$. We estimate this limiting covariance by $\nabla f(\hat{\theta}) \hat{\Sigma} \nabla f(\hat{\theta})^T$, where typically $\Sigma = \Sigma(\theta)$ and $\hat{\Sigma} = \Sigma(\hat{\theta})$.

Returning now to our problem, in order to apply the δ-method we need the gradient of $\mathcal{Y}$ with respect to θ. While tedious, this calculation is easily performed with the help of symbolic differentiation software like Maple [Maplesoft, a division of Waterloo Maple Inc., 2020]. Next, we evaluate our gradient formula at $\hat{\theta}$, $\nabla \mathcal{Y}|_{\theta = \hat{\theta}}$, then pre- and post-multiply $\hat{\Sigma}_{\theta}$ by this estimated gradient. If we require the covariance matrix for multiple values of $\mathcal{Y}$ (e.g., $\mathcal{Y}(x, x)$ and $\mathcal{Y}(x, x')$ for some $x \neq x'$), we frame the problem as a vector-valued transformation, and compute the Jacobian by stacking the gradient for each individual $\mathcal{Y}$. More precisely, to get a covariance matrix for $\mathcal{Y}_1, \ldots, \mathcal{Y}_r$, we first construct the r-by-$|\theta|$ Jacobian matrix $\mathcal{J} := [\nabla \mathcal{Y}_1, \ldots, \nabla \mathcal{Y}_r]$. We then evaluate $\mathcal{J}$ at $\theta = \hat{\theta}$, call the result $\hat{\mathcal{J}}$, and pre- and post-multiply $\hat{\Sigma}_{\theta}$ by $\hat{\mathcal{J}}$, giving $\hat{\mathcal{J}} \hat{\Sigma}_{\theta} \hat{\mathcal{J}}^T$.

Finally, to get the asymptotic variance of one or more estimated mediation effects, we just apply the δ-method to the corresponding formulas from Table 3.1. Since the mediation effects are all simple functions of the $\mathcal{Y}$, any gradients required for this step are easily obtained.

3.3   To Do

- Incorporate an interaction term between X and M in the model for Y.

- Explore dependence between the models for M and Y using the 'stacking' technique described in Bauer et al. [2006]. This will give non-zero covariance between parameter estimators from the two models.

- Incorporate covariates/confounders

Bibliography

Ghulam Ali Arain, Imran Hameed, Waheed Ali Umrani, Abdul Karim Khan, and Abdullah Zafar Sheikh. Consequences of supervisor knowledge hiding in organizations: a multilevel mediation analysis. *Applied Psychology: An International Review*, 70(3), 2021.

Reuben M. Baron and David A. Kenny. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1986.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using `lme4`. *Journal of Statistical Software*, 67(1), 2015.

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.

Andrew F. Hayes and Nicholas J. Rockwood. Conditional process analysis: concepts, computation and advances in the modeling of the contingencies of mechanisms. *American Behavioral Scientist*, 64(1), 2020.

Jeongha Hwang and Karen Hopkins. Organizational inclusion, commitment, and turnover among child welfare workers: a multilevel mediation analysis. *Administration in Social Work*, 36, 2012.

Kosuke Imai, Luke Keele, and Dustin Tingley. A general approach to causal mediation analysis. *Psychological Methods*, 15(4), 2010a.

Kosuke Imai, Luke Keele, and Teppei Yamamoto. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1), 2010b.

Kosuke Imai, Luke Keele, Dustin Tingley, and Teppei Yamamoto. Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, 105(4), 2011.

David A. Kenny, Josephine D. Korchmaros, and Niall Bolger. Lower level mediation in multilevel models. *Psychological Methods*, 8(2), 2003.

Gary King, Michael Tomz, and Jason Wittenberg. Making the most of statistical analyses: improving interpretation and presentation. *American Journal of Political Science*, 44(2), 2000.

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.

Jennifer L. Krull and David P. MacKinnon. Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, 36(2), 2001.

Theis Lange, Stijn Vansteelandt, and Maarten Bekaert. A simple unified approach for estimating natural direct and indirect effects. *American Journal of Epidemiology*, 176(3), 2012.

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.

Maplesoft, a division of Waterloo Maple Inc. Maple, 2020.

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 and James P. Selig. Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, 6 (2), 2012.

Kristopher J. Preacher, Michael J. Zyphur, and Zhen Zhang. A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, 15(3), 2010.

Sophia Rabe-Hesketh, Anders Skrondal, and Andrew Pickles. Generalized multilevel structural equation modeling. *Psychometrika*, 69(2), 2004.

Judith J. M. Rijnhart, Sophia J. Lamp, Matthew J. Valente, David P. MacKinnon, Jos W. R. Twisk, and Martijn W. Heymans. Mediation analysis methods used in observational research: a scoping review and recommendations. *BMC Medical Research Methodology*, 21(226), 2021.

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.

Nicholas J. Rockwood. MLMED user guide, 2019.

Nicholas J Rockwood and Andrew F Hayes. `MLmed`: An `SPSS` macro for multi-level mediation and conditional process analysis. In *Poster presented at the annual meeting of the Association of Psychological Science (APS), Boston, MA*, 2017.

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.

Anders Skrondal and Sophia Rabe-Hesketh. Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society: Series A*, 172(3), 2009.

Michael E. Sobel. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13, 1982.

Johan Steen, Tom Loeys, Beatrijs Moerkerke, and Stijn Vansteelandt. `medflex`: an `R` package for flexible mediation analysis using natural effect models. *Journal of Statistical Software*, 76(11), 2017.

Dustin Tingley, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. `mediation`: `R` package for causal mediation analysis. *Journal of Statistical Software*, 59(5), 2014.

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.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

Tyler J. VanderWeele and Stijn Vansteelandt. Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology*, 172(12), 2010.

Stijn Vansteelandt, Maarten Bekaert, and Theis Lange. Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiologic Methods*, 1(7), 2012.

Alejandro Veas, Juan-Luis Castejón, Pablo Miñano, and Raquel Gilar-Corbí. Relationship between parent involvement and academic achievement through metacognitive strategies: a multiple multilevel mediation analysis. *British Journal of Educational Psychology*, (89), 2019.

Steven Walker, Rune Haubo Bojesen Christensen, Douglas Bates, Benjamin M. Bolker, and Martin Mächler. Fitting generalized linear mixed-effects models using `lme4`. *Unpublished*, 2023. URL `https://github.com/lme4/lme4/blob/master/misc/glmer_JSS/glmer.pdf`.

Ting Wang and Edgar C. Merkle. `merDeriv`: derivative computations for linear mixed effects models with applications to robust standard errors. *Journal of Statistical Software*, 87(1), 2018.

Jason Williams and David P. MacKinnon. Resampling and distribution of the product method for testing indirect effects in complex models. *Structural Equation Modeling*, 15, 2008.

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.