* Causal Mediation Effects (CMEs) analysis has proven to be an effective methodology to understand potential mechanisms for causal analysis, particularly in the social sciences.   What are the principal limitations of this approach? Is there any advantage of taking a Bayesian approach to learning causal mediation effects?  Finally, can you imagine a scenario where CMEs could be used to address mechanism associated with anthropogenic climate change causes of increased hurricane intensity?

Comprehensive Exam Answers: Chris Wikle

What are the limitations of Causal Mediational Effects analysis?

Imai 2010b:

* Causal mediation originally was analyzed using linear structural equation models, which had serious weaknesses
  + Lack of general definition of causal mediation effects independent of a particular model
  + Inability to specify key identification assumptions
  + Difficulty of extending the framework to nonlinear models
* Imai offers instead a ‘general’ alternative approach
  + Offers definition, identification, estimation, and sensitivity analysis, without referencing a specific statistical model
  + We link these 4 elements into a single framework, accommodating linear and nonlinear relationships, parametric and nonparametric models, continuous/discrete mediators, and various types of outcome variables.

One main goal of social science is causal inference, with randomized experiments as the gold standard. However, this is only a ‘black box’ view of causality!

* Estimation of causal effects allow researchers to examine whether a treatment causally affects an outcome!
* It CANNOT tell us how and why such an effect happens
  + This is important b/c identification of these mechanisms is needed to test competing theoretical explanations of the same causal effects!
* Our solution is mediational analysis
  + Identifies intermediate variables or mediators that lie in the causal pathway

Traditionally uses Linear Structural Equation Modelling (LSEM; e.g., Baron & Kenny, 1986; Hyman, 1955; James, Mulaik, & Brett, 1982; Judd & Kenny, 1981; MacKinnon, 2008; MacKinnon & Dwyer, 1993). However, this doesn’t work for TWO reasons!

1. By definition, LSEM cannot offer a general definition of causal mediation effects that are applicable BEYONE specific statistical models
   1. This is because the key identification assumption is stated in the context of a particular model
   2. Making it hard to separate limitations of research design from those of the specific statistical model!
2. Second the methods in LSEM don’t generalize to nonlinear models
   1. Including logit and probit models, for discrete mediators and outcomes as well as non- or semiparametric models.

Again, our main assumption (but also limitation) is that of SEQUENTIAL IGNORABILITY

* Conditional on the observed pretreatment covariates, the treatment is independent of all potential values of the outcome and mediating variables.
* The observed mediator is independent of all potential outcomes given the observed treatment and pretreatment covariates.
  + This is vital b/c it’s nonparametric identification, setting a minimum base of assumptions required for mediation effects to be causal without respect to statistical models used by researchers.
* The main advantage of this assumption over other alternatives is the ease of interpretation!
* Assumption 1 is called sequential ignorability, because two ignorability assumptions are made sequentially
  + First, given the observed pretreatment confounders, the treatment assignment is assumed to be ignorable, that is, statistically independent of potential outcomes and potential mediators.
    - In the JOBS II study, this assumption is satisfied because workers were randomly assigned to the treatment and control groups
    - It would not be satisfied in any studies where the subjects may self-select into the treatment group!
    - Obtaining as many pretreatment confounders as possible helps improve the credibility of the ignorability of treatment assignment
  + The second part states that the mediator is ignorable given the observed treatment and pretreatment confounders. That is, the second part of the sequential ignorability assumption is made conditional on the observed value of the ignorable treatment and the observed pretreatment confounders.
    - The ignorability of the mediator may NOT hold even in some randomized experiments!
    - In the JOBS II study, the randomization of the treatment assignment does NOT justify this second ignorability assumption, because the post-treatment level of workers job search self-efficacy is NOT randomly assigned by the researchers.
    - In other words, the ignorability of the mediator implies that for workers with the same treatment status and pre-treatment characteristics, the mediator can be regarded AS IF it were randomized
  + This second stage is a STRONG assumption and should not be made lightly
    - It is always possible that there could be unobserved variables that confound the relationship between the outcome and the mediator variables even after conditioning on the observed treatment status and the observed covariates.
* Similar to the ignorability of treatment assignment in observational studies, it is difficult to know for certain whether the ignorability of the mediator holds EVEN AFTER the researchers collect as many pretreatment confounders as possible
  + This assumption is referred to as ‘nonrefutable’ because it CANNOT be directly tested from the observed data
* Sensitivity Analysis: SOME MORE?
  + Instead, we can test for it using sensitivity analyses that allow us to quantify the degree to which the empirical findings are ROBUST to a potential violation of the sequential ignorability assumption
  + Sensitivity analyses are appropriate because they allow us to probe whether a substantive conclusion is robust to violations of our sequential ignorability assumption!
    - In particular, the 2nd part of Assumption 1 is ‘nonrefutable’. By definition, we cannot test this assumption with our observed data.
* Sensitivity analysis addresses these nonrefutable assumptions by showing what conditions would have to exist in order for our assumptions to be proven wrong.
  + The goal is to ‘quantify the degree to which the key identification assumptions MUST be violated for the original conclusion to be reversed’
  + If an inference is sensitive, a slight violation of the assumption may lead to different conclusions!

MacKinnon 2002: Many ways to test mediational analysis

As is, all of these methods are valid when considered under the perspective of an ‘intervening variable effect’, wherein “An intervening variable (mediator) transmits the effect of an independent variable to a dependent variable”

* Causal Steps Test:
  + Specifies a series of tests of links in a ‘causal chain’. Traditionally based on the work of Judd and Kenny (1981), and Baron and Kenny (1986) and is very commonly used.
  + Judd and Kenny concluded that mediation would not be sufficient unless there was complete mediation (is no effect after accounting for the mediators), Baron and Kenny found later that only partial mediation is acceptable as well, and more realistic in social sciences.
  + Weaknesses:
    - Does not provide a joint test of the three conditions.
    - Does not estimate the size of indirect effect of X on Y (vs the direct effect)
    - Does not provide standard errors for confidence intervals.
    - Has a lot of trouble w/ multiple mediation models and evaluability.
    - Cannot detect in cases of mediation where the indirect effect and direct effect ‘cancel’ each other out if the effects are in different directions.
* Difference in Coefficients:
  + Methods such like those that compare the difference b/w a regression coefficient before and after adjusting for the mediator (Freedman & Schatzkin, 1992;McGuigan & Langholtz, 1988;Olkin& Finn, 1995). Some of these methods test hypothesis about intervening variables that diverge from what psychologists are ‘used to’.
  + We can also test mediators by comparing relationship b/w independent and dependent variable before and after controlling for our mediator. We have pairs of coefficients to compare in this case, such as the regression coefficients and the correlation coefficients.
  + Each of these methods provide an estimate of the mediator and it’s standard error, however, the null hypothesis might be strangely formatted and not resemble traditional psychological sciences ones. E.g. the Clogg test assumes fixed X and I, which isn’t realistic for mediating variables.
  + Main weakness is that these methods do NOT provide a framework for generalizing the tests to estimate appropriate coefficients and test significance for multiple mediators.
* Product-of-Coefficients:
  + Uses the product of coefficients involving paths in a path model (aka the indirect effect; Alwin & Hauser, 1975;Bollen, 1987;Fox, 1980;Sobel, 1982,1988).
  + Tests the significance of the intervening variable effect by dividing the estimate of the intervening variable effect, αβ, by its standard error and comparing this value to a standard normal distribution.
  + Our model follows from path analysis where our mediator is the PRODUCT of coefficients that we hypothesize to measure causal relations. This works just fine for multiple mediator models!
    - Two main problems, sampling distribution is NOT the normal distribution, and our H0 is very complex.
* This study as a whole was concerned with balancing power and type I error, in a simulated study (where the true effects/values are known)
  + There was no main difference in methods between a binary case and a continuous case
  + Most estimates of the mediator had minimal bias, except for za x zb , as the point estimates for this were larger than point estimates for the mediator. Bias decreased as sample size and effect size increased for all methods.
  + J and K/ B and K are too underpowered. This is b/c requiring a total significant effect of X on Y leads to a lot of Type II errors. These methods are likely to miss real effects, but UNLIKELY to commit a Type I error. Good for specific use cases, but the alternative causal steps method, testing if alpha and beta are JOINTLY significant, has more power and more accurate Type I error rates.
  + Power rates for difference in coefficients methods are higher than B and K and J and K, but Type I error rates are TOO conservative except for clogg/freedman and schatzkin tests. Has the most accurate Type I error rates and greatest power for most situations. These methods underestimate standard error, but that compensates for too low critical values in the standard reference distribution! Product of coefficients method is higher power, but the Type I error rates are too low.
  + In summary – tests of mediation trade off two competing issues! Non-normal sampling distribution of our alpha x beta effect leads to tests that are associated w/ lower empirical levels of significance than stated levels, when H0 is true, and low power when H0 is false. Second, the test for the null hypothesis for alpha x beta = 0 is complex, because it is a COMPOUND form where a = 0, b = 0, a = 0 and b ≠ 0, and b = 0 and a ≠ 0. In contrast, using otherwise overly conservative critical values turns out to empirically compensate for the inflation in Type I error rate due to this compound null hypothesis!
* Requirements for Causal Inference are VERY complex and controversial!
  + Generally, using J and K/ B and K method, in the context of a single randomized experiment it provides evidence that the treatment causes the intervening variable, the treatment causes the outcome, and that the data are consistent with the proposed intervening variable model
    - Among the conditions necessary for causal inference are randomization, linear effects, and that the full effect of the treatment operates through the intervening variable (i.e., no partial intervening variable effect).

Preacher and Hayes 2008: Showcases some extensions on multiple mediation.

* How to find mediation, in 2 parts
  + Investigating the total indirect effect (deciding which set of mediators translates the effect of X on Y
  + Testing hypothesis regarding individual mediators in the context of a multiple mediator model (the specific indirect effect associated w/ each mediator)
    - A significant total indirect effect is NOT needed to investigate specific indirect effects.
    - It is possible to find specific indirect effects to be significant in the presence of a nonsignificant total indirect effect!

Imai 2010a: Defines CME analysis as “Determining the alternative causal mechanisms by examining the roles of intermediate variables.”

* Limitations here are clearly our assumptions that are required for the method to work properly?
  + Under certain assumptions, we can prove that the average causal mediation effect can be NONPARAMETRICALLY identified! The Linear Structural Equation Model (LSEM) can be interpreted as an ACME (average causal mediation effect) estimator once we add some parametric assumptions.
  + In a lot of experimental set-ups, we can randomize our treatment assignment, but not randomize our mediator. Thus, there could be unobserved covariates that confound our proposed relationship. For example, the subject’s political ideology affects both their public order attitude and tolerance for Klan rally under BOTH treatment conditions!
* Imai notes the following as the required preconditions to identify an Average Causal Mediational Effect (ACME)
  + Random sample of size n from a population, for each unit, *i*  that we observe four traits, (Ti, Mi, Xi, and Yi). Ti is the binary treatment variable = 1 if receives treatment, 0 otherwise. Our mediator is Mi, and Y­I is our outcome variable.
  + Finally, Xi is our vector of observed pre-treatment covariates (similar to what we would use to calculate a propensity score!). *M, X,* and *Y* denote the support of the distributions.
  + Given this, what is a mediator? It MUST be a post-treatment variable that occurs before the outcome is realized! Other than this very obvious requirement, what a mediator is, is based on previous theory. For example, treatment is receiving vaccine, outcome is whether or not subject gets the flu. Scientist may say that antibodies are the mediator (Vaccine > antibodies > flu). However… the parents signing form for risks of vaccine could also be a mediator (in theory) – Hypothesis could be, getting informed of the risks will make parents LESS likely to have the child get the 2nd dose of vaccine, thereby increasing likelihood of getting flu.
  + Our causal mediation effect can be defined as “Sigma (t)”, which is the difference between the potential outcome given treatment status t, and our potential outcome if the treatment is the SAME, but with a mediator value that would result under the other treatment status.
    - We can observe the first one… but the second one is by definition unobservable!
* The total causal effect is equal to the sum of our mediator effect under one treatment condition, and the natural direct effect under our other treatment condition
  + The causal mediation effect examines whether our mediator mediates the causal relationship between our treatment and outcome, and the controlled direct effect looks at whether or not the treatment MODERATES (interacts with) the causal effect of mediator on outcome! (very interesting!)
* Sensitivity Analysis is the Next STEP!
  + ACME works under assumption 1, but these assumptions may be TOO strong.
  + For example, in situations where treatment is randomized, but the mediator is NOT randomized. In this case, part of assumption 1 is satisfied, but part of it is not, for two reasons:
    - Unmeasured pre-treatment covariates that confound the relation between mediator and outcome
    - Observed or unobserved post-treatment confounders
  + This can be addressed by assessing sensitivity of an estimated Average Causal Mediational Effect to unmeasured pre-treatment confounding. Note this does not address post-treatment confounding, which is best dealt with by experimental design.

Assumption 1 (Sequential Ignorability): The outcome given a particular treatment prime and mediator, given a particular treatment, are independent of our treatment, GIVEN that our pretreatment covariates are x

What are the advantages of using Bayesian methods when studying Casual Mediation Effects?

Elliot 2010: Develops a Bayesian approach for estimating direct and mediated effects, when we have dichotomous mediator and dichotomous outcome, which means that many parameters cannot be fully identified.

In particular, a dissassociative effect of treatment can be estimated as the ITT effect among subjects for whom the mediator does not change under different treatment assignments, and similarly an associative effect can be estimated as the ITT effect among subjects for whom the mediator does change under different treatment assignments . A mediated effect can then be constructed by considering the value of the disassociative and associative effects when the overall treatment effect is entirely direct versus completely mediated. The Bayesian approach also allow us to incorporate constraints such as monotonicity or a relaxed stochastic monotonicity

* Bayesian approach works because likelihood theory does not address nonidentifiable parameters well.
  + Allows for the direct and mediated effects to be expressed in terms of the posterior distribution of the population parameters of interest
* The likelihoods are extremely flat, and standard asymptotic methods for point estimation and CI construction don’t really apply.
  + Thus, Bayesian methods can describe the information available about the associative and dissociative effects of interest.

McCandless 2017: Bayesian techniques work well to find potential unmeasured confounders in causal mediation analysis.

* Unique in that it assesses for confounding in all 3 potential locations
  + Mediator to outcome confounding
  + Exposure/Outcome
  + Exposure/Mediator
* This is a ‘natural’ Bayesian extension of sensitivity analysis methods that already exist in the field.

Some assumptions

* No confounding for the exposure–outcome relationship
* No confounding for the mediator–outcome relationship
* No confounding for the exposure–mediator relationship
* Finally, no mediator–outcome confounder that is itself affected by the exposure.

As an alternative to the sensitivity analysis done thus far, we can look at it as Bayesians!

* Bayesian Sensitivity Analysis for unmeasured confounding in observational studies.
  + Incorporates uncertainty about unmeasured confounding by using prior distributions for sensitivity parameters.
  + The posterior distribution for the causal effect estimate incorporates uncertainty from bias (systematic error) in addition to uncertainty from random sampling (random error).
  + An advantage of BSA is that it gives dimension reduction when there are multiple sensitivity parameters inputs because it averages over uncertainty in the prior.
  + Furthermore, posterior credible intervals will often have better frequentist coverage probability compared to interval estimates that ignore unmeasured confounding
* In our case however… we can use BSA-like techniques to explore sensitivity to unmeasured confounding in CMA
  + It has the unique advantage that it is able to simultaneously assess unmeasured mediator–outcome, exposure–outcome and exposure–mediator confounding.
* Conceptually, introduces a latent binary variable U that takes values 1 or 0 to indicate the presence or absence of an unmeasured confounder.
  + This latent variable U is linked to each of our measured variables, T, M, and X, in our ‘causal diagram’ for the hypothesized relationship b/w our variables.
  + It is a UNIQUE feature of this Bayesian method that U is simultaneously a confounder for mediator/outcome, exposure/outcome, and exposure/mediator relationships.
* We have a model thus, as a series of equations contingent on U, or with U contingent on other variables.
  + In these equations, we have bias parameters that work as the basis for our sensitivity analysis.
    - Betau is the log hazard ratio of the association between U and survival time T, conditional on X, M, and C
    - Alphau governs the association between U and the mediator M, given X and C
    - Lambda0 and Lambdax is the log odds ratio for the association between the unmeasured confounder and X given C.
  + We then assign uniform prior distributions to our four bias parameters.
    - This is a uniform grid of values, similar to those in sensitivity analysis w/ high dimensional tables.
  + Each of our priors are parameterized by a mean and width parameter, specified by the investigator to represent the size and direction of unmeasured confounding that they suspect.
    - Otherwise, all other parameters are assigned default non-informative priors.
* Samples from this posterior distribution for model parameters by using marginal likelihood function that integrates over U.
  + How do we sample from our posterior distribution, if our model for unmeasured confounding is non-identifiable?
  + We use a novel Monte Carlo sensitivity Analysis (MCSA)
    - Repeatedly sample bias parameters directly from our priori distributions
    - Use the sampled values to obtain point estimates for our remaining parameters that are corrected for unmeasured confounding
    - At each monte-carlo iteration, random sampling error is incorporated using standard asymptotic approximations to the standard error of the point estimates
  + Resulting in a sample of bias-corrected estimates from which we can calculate summary statistics like mean, percentiles, and interval estimates!
* MCSA is a perfect sampler in the sense that it samples from the desired distribution from the start, and there is no Markov chain simulation involved.
  + Consequently, it is not necessary to assess sampler convergence.
* MCSA has the advantage that it is much more computationally efficient than direct MCMC simulation.

Kim 2019: Bayesian methods in multiple mediator analysis, for power plant emissions

* Bayesian methods here are used to enable a nonparametric approach to modeling the observed distribution of emissions and pollution outcomes
  + Multivariate Gaussian copula model, linking flexibly-modeled marginal distributions of observed outcomes to a joint distribution of potential outcomes.
* Posterior distributions cannot be computed directly from observed data because potential outcomes are never jointly observed in both the presence and absence of a scrubber and a priori counterfactuals are never observed.
  + Nonparametric models were specified for the observed data
    - The marginal distribution for EACH observed mediator observed for power plants (e.g. ones who installed scrubbers and those who didn’t) is specified separately and linked to a coherent joint distribution using a gaussian copula model.
  + The models for potential outcomes are specified conditional on covariates and all potential mediators that are never observed simultaneously!
    - Thus, the conditional outcome models are ESTIMATED via the data augmentation for unobserved potential mediators.
* In general, value is found because “Bayesian nonparametric modeling approaches provided flexible models for the observed data (marginal distribution for each mediator and conditional distribution for the outcome under each intervention z = 0, 1), and linked observed data distributions to joint distributions of potential mediators using explicit and transparent assumptions about both observable and a priori counterfactuals.”

What is a scenario where CMEs could be used to address mechanisms associated with anthropogenic climate change causes of [Increased hurricane intensity]?

Kim 2019: Bayesian CME is considered super useful “To accommodate the setting of multiple pollutants that are emitted contemporaneously and possibly interact with one another, we have developed methods to accommodate multiple contemporaneous and non-independent mediators.”

* This falls in line with what I perceive to be the biggest problems/challenges with studying the effects of climate change, as this is literally research on a form of climate change.