Bayesian Causal Mediation Analysis with Stan – A g-formula approach

Belay B. Yimer, Mark Lunt, John McBeth

Center for Epidemology Versus Arthritis
University of Manchester

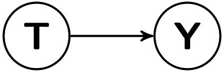
RSS International conference 2021, Manchester

September 9, 2021

The University of Manchester

Mediation analysis

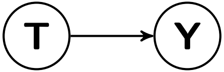
- Many scientific studies aim to infer if a given treatment or intervention influences a given outcome.



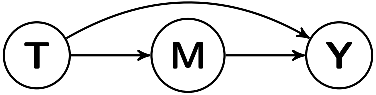
Mediation analysis

Belay B. Yimer, Mark Lunt, John McBeth

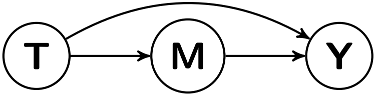
- Many scientific studies aim to infer if a given treatment or intervention influences a given outcome.



- Increasingly, many studies are interested in disentangling the pathways that link exposure to the outcome to get insight into the mechanism.
  - Why/how does the treatment/intervention affect outcome?

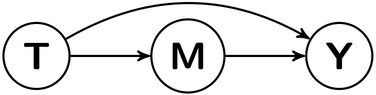


Mediation analysis



- Mediation analysis can be helpful in identifying
  - the effect of the intervention that acts through a given set of intermediate variables (indirect effect), and
  - the effect of the intervention unexplained by those same intermediate variables (direct effect).

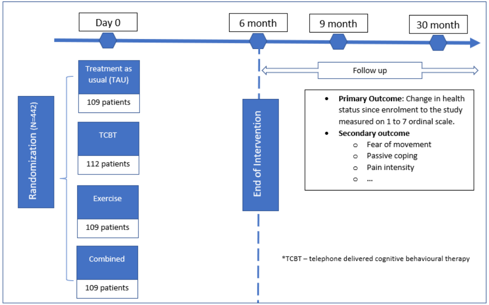
Mediation analysis



- Mediation analysis can be helpful in identifying
  - the effect of the intervention that acts through a given set of intermediate variables (indirect effect), and
  - the effect of the intervention unexplained by those same intermediate variables (direct effect).
- Question: How can we make inference about these causal effects from experimental and observational studies?

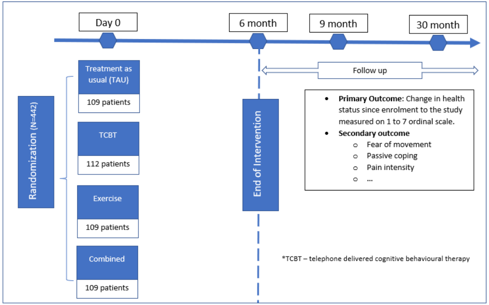
Motivating example — MUSICIAN trial

- The Managing Unexplained Symptoms (CWP) In Primary Care: Involving Traditional and Accessible New Approaches – a 2 × 2 factorial randomized controlled trial (McBeth et al., 2012)



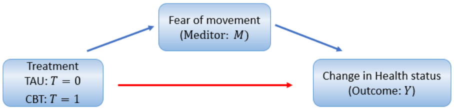
Motivating example — MUSICIAN trial

- The Managing Unexplained Symptoms (CWP) In Primary Care: Involving Traditional and Accessible New Approaches – a 2 × 2 factorial randomized controlled trial (McBeth et al., 2012)



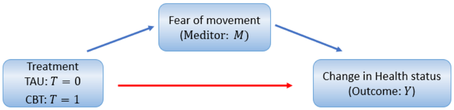
- CBT was associated with substantial improvements in patient global assessment.

Research questions



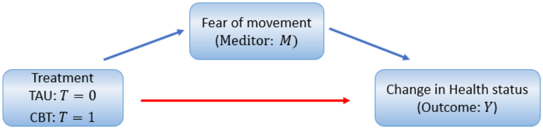
- To what extent does the treatment improve health status by inducing a change in fear of movement?
- To what extent the treatment improve health status independent of changing fear of movement?

Effect definition



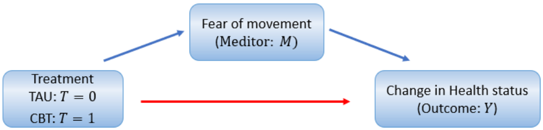
	CBT ($T_i = 1$)	TAU ($T_i = 0$)
Fear of movement if CBT ($M_i(1)$)		
Fear of movement if TAU ($M_i(0)$)		

Effect definition



	CBT ($T_i = 1$)	TAU ($T_i = 0$)
Fear of movement if CBT ($M_i(1)$)	$Y_i(1, M_i(1))$	
Fear of movement if TAU ($M_i(0)$)	$Y_i(1, M_i(0))$	$Y_i(0, M_i(0))$

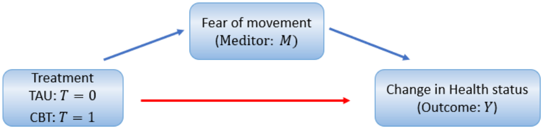
Effect definition



	CBT ($T_i = 1$)	TAU ($T_i = 0$)
Fear of movement if CBT ($M_i(1)$)	$Y_i(1, M_i(1))$	
Fear of movement if TAU ($M_i(0)$)	$Y_i(1, M_i(0))$	$Y_i(0, M_i(0))$

Population Average Natural Indirect effect: $E[Y(1, M(1))] - E[Y(1, M(0))]$

Effect definition



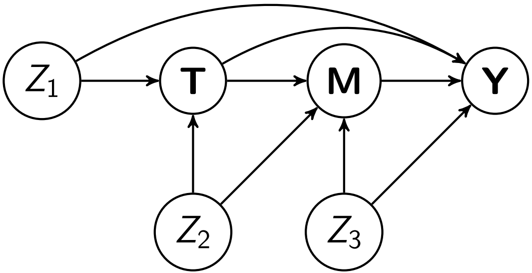
	CBT ($T_i = 1$)	TAU ($T_i = 0$)
Fear of movement if CBT ($M_i(1)$)	$Y_i(1, M_i(1))$	
Fear of movement if TAU ($M_i(0)$)	$Y_i(1, M_i(0))$	$Y_i(0, M_i(0))$

Population Average Natural direct effect: $E[Y(1, M(0))] - E[Y(0, M(0))]$

Identification of causal effects

To estimate natural direct and indirect effect we need:

- There are no unmeasured exposure-outcome confounder given $\boldsymbol{Z}$
- There are no unmeasured mediator-outcome confounder given $(T, \boldsymbol{Z})$
- There are no unmeasured exposure-mediator confounder given T
- There are no mediator-outcome confounder affected by exposure



Identification of causal effects

Under the four identification assumptions, natural direct and indirect effects are given by

$$E\left[Y(1, M(0)) - Y(0, M(0))\right] = \int\int \{E\left[Y \mid t = 1, m, z\right] - E\left[Y \mid t = 0, m, z\right]\} dF_{M|t=1,z}(m) dF_Z(z)$$

$$E\left[Y(1, M(1)) - Y(1, M(0))\right] = \int\int E\left[Y \mid t = 1, m, z\right] \{dF_{M|t=1,z}(m) - dF_{M|t=0,z}(m)\} dF_Z(z)$$

Estimation

- Analytical integration (Valeri, L., and VanderWeele, T. J. (2013))
  - Parametric regression model for Y and M and computing the integration analytically
  - `SAS` and `SPSS macros` available
  - Frequentest approach
- Sampling (Imai et al., 2010)
  - Proposed to use a broad class of parametric or semiparametric model for Y and M
  - Use simulations to calculate NIE and NDE and the standard errors for this using bootstrap
  - Popular `R` - package `mediation`
  - Quasi-Bayesian approach

Estimation

- We propose to use Bayesian framework for estimation of NIE and NDE.
- Why bayes?

Estimation

- We propose to use Bayesian framework for estimation of NIE and NDE.
- Why bayes?
  - Full posterior inference for any function of model parameters, hence, point and interval estimates can be easily constructed for causal risk ratios, odds ratios, and risk differences
  - Priors can help us compute causal effects under sparsity - avoid *ad hoc* approaches
  - Probabilistic sensitivity analysis

Estimation — How it works?

- $\boldsymbol{D} = \{Y_i, T_i, M_i, \boldsymbol{Z}_i\}$
- T_i is a binary treatment assignment $t \in \{0, 1\}$
- Y is binary and M is continuous
- Assume IA (1) – IA (4) hold, the following regression models are correctly specified
  - $\text{logit}(P(Y_i = 1 \mid T_i, M_i, \boldsymbol{Z}_i)) = \alpha_0 + \boldsymbol{\alpha}_Z \boldsymbol{Z}_i + \alpha_T T_i + \alpha_M M_i,$
  - $E\left[M_i \mid T_i, \boldsymbol{Z}_i\right] = \beta_0 + \boldsymbol{\beta}_Z \boldsymbol{Z}_i + \beta_T T_i$
- $\boldsymbol{\theta} = (\alpha_0, \boldsymbol{\alpha}_Z, \alpha_T, \alpha_M, \beta_0, \boldsymbol{\beta}_Z, \beta_T)$
- Appropriate prior is assumed for elements of $\boldsymbol{\theta}$.
- The model is fitted in `stan`.

Estimation — How it works?

- Obtain b^{th} set of posterior draws $\boldsymbol{\theta}^{(b)},\ b = 1, ..., B$

Estimation — How it works?

- Obtain b^{th} set of posterior draws $\boldsymbol{\theta}^{(b)},\ b = 1, ..., B$
- Bootstrap sampling to integrate out the confounder distribution:
  - Sample n new values of $\boldsymbol{Z}$ with replacement from the observed Z distribution and denote these resampled values as $\boldsymbol{Z}^{(1,b)}, \ldots, \boldsymbol{Z}^{(n,b)}$

Estimation — How it works?

- Obtain b^{th} set of posterior draws $\boldsymbol{\theta}^{(b)}$, $b = 1, ..., B$
- Bootstrap sampling to integrate out the confounder distribution:
  - Sample n new values of $\boldsymbol{Z}$ with replacement from the observed Z distribution and denote these resampled values as $\boldsymbol{Z}^{(1,b)}, \ldots, \boldsymbol{Z}^{(n,b)}$
- Draw the potential outcome values
  - Potential values of M: $M(t)^{(i,b)} \sim \text{Normal}(\beta_0^{(b)} + \beta_Z^{(b)} \boldsymbol{Z}^{(i,b)} + \beta_T^{(b)} t, \sigma^{(b)})$

Estimation — How it works?

- Obtain b^{th} set of posterior draws $\boldsymbol{\theta}^{(b)}$, $b = 1, ..., B$
- Bootstrap sampling to integrate out the confounder distribution:
  - Sample n new values of $\boldsymbol{Z}$ with replacement from the observed Z distribution and denote these resampled values as $\boldsymbol{Z}^{(1,b)}, \ldots, \boldsymbol{Z}^{(n,b)}$
- Draw the potential outcome values
  - Potential values of M: $M(t)^{(i,b)} \sim \text{Normal}(\beta_0^{(b)} + \beta_Z^{(b)}\boldsymbol{Z}^{(i,b)} + \beta_T^{(b)}t, \sigma^{(b)})$
  - Given, the potential values of M, draw potential values of Y: $Y(t, M(t)^{(i,b)})^{(i,b)} \sim$ $\text{Bernoulli}(\text{logit}^{-1}(\alpha_0^{(b)} + \alpha_Z^{(b)}\boldsymbol{Z}^{(i,b)} + \alpha_T^{(b)}t + \alpha_M^{(b)}M(t)^{(i,b)}))$

Estimation — How it works?

- Obtain b^{th} set of posterior draws $\boldsymbol{\theta}^{(b)}$, $b = 1, ..., B$
- Bootstrap sampling to integrate out the confounder distribution:
  - Sample n new values of $\boldsymbol{Z}$ with replacement from the observed Z distribution and denote these resampled values as $\boldsymbol{Z}^{(1,b)}, \ldots, \boldsymbol{Z}^{(n,b)}$
- Draw the potential outcome values
  - Potential values of M: $M(t)^{(i,b)} \sim \text{Normal}(\beta_0^{(b)} + \beta_Z^{(b)}\boldsymbol{Z}^{(i,b)} + \beta_T^{(b)}t, \sigma^{(b)})$
  - Given, the potential values of M, draw potential values of Y: $Y(t, M(t)^{(i,b)})^{(i,b)} \sim$ $\text{Bernoulli}(\text{logit}^{-1}(\alpha_0^{(b)} + \alpha_Z^{(b)}\boldsymbol{Z}^{(i,b)} + \alpha_T^{(b)}t + \alpha_M^{(b)}M(t)^{(i,b)}))$
- Compute NIE and NDE
  - $NDE^{(b)} = \frac{1}{n}\sum_{i=1}^{n}\{Y(1, M(0)^{(i,b)})^{(i,b)} - Y(0, M(0)^{(i,b)})^{(i,b)}\}$
  - $NIE^{(b)} = \frac{1}{n}\sum_{i=1}^{n}\{Y(1, M(1)^{(i,b)})^{(i,b)} - Y(1, M(0)^{(i,b)})^{(i,b)}\}$
- Compute average and quartiles of NDE and NIE

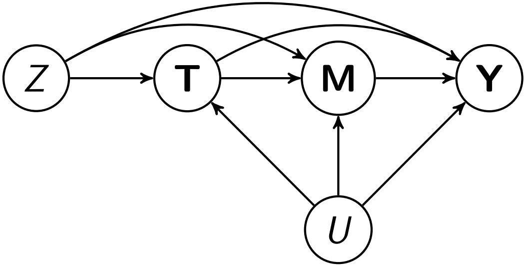
Application to MUSICIAN trial

- The Bayesian g-formula approach and the approach by Imai et al, 2010 (R-package `mediation`) leads to a comparable result.
- The effect of CBT on change in health status is mainly through mechanisms other than fear of movement.

Estimand	Imai et al. 2010	Bayesian g-formula
Direct effect	0.217 (0.107, 0.350)	0.224 (0.078, 0.369)
Indirect Effect	0.040 (-0.02, 0.100)	0.037 (-0.071, 0.142)

Bayesian sensitivity analysis

- The identification assumptions are often too strong
- Need to assess the robustness of findings via sensitivity analysis
- Question: How large a departure from the key assumption must occur for the conclusions to no longer hold?



Bayesian sensitivity analysis

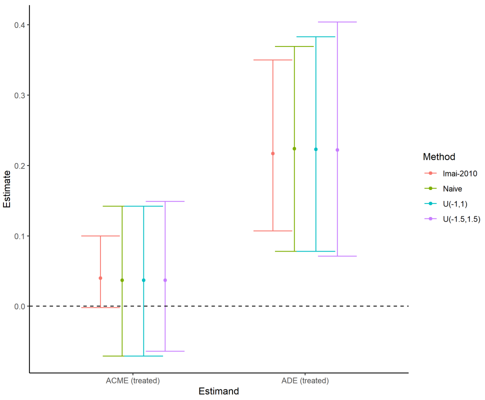
- We follow the approach by McCandless et al, 2007

$$Y|T, M, \boldsymbol{Z} \sim \text{Bernoulli}(\text{expit}(\alpha_0 + \boldsymbol{\alpha}_Z \boldsymbol{Z} + \alpha_T T + \alpha_M M + \alpha_U U))$$
$$M|T, \boldsymbol{Z}, \sigma \sim \text{Normal}(\beta_0 + \boldsymbol{\beta}_Z \boldsymbol{Z} + \beta_T T + \beta_U U, \sigma^2)$$
$$U|\boldsymbol{Z} \sim \text{Bernoulli}(\text{expit}(\gamma_0 + \gamma_Z \boldsymbol{Z}))$$

- We assign Uniform mean-zero bounded priors for the sensitivity parameters.

$$\alpha_U \sim U(-\delta, \delta)$$
$$\beta_U \sim U(-\delta, \delta)$$
$$\gamma_Z \sim U(-\delta, \delta)$$
$$\gamma_0 \sim U(-\delta, \delta)$$

Bayesian sensitivity analysis



Summary

- We have demonstrated the application of g-formula to Bayesian models for conducting mediation analysis.
- We show a flexible Bayesian model to explore sensitivity to unmeasured confounding in causal mediation analysis.
- Our goal is to make the methodology accessible to practitioners.
  - The development version of the R-package, `BayesGmed`, are available at https://github.com/belayb/BayesGmed.

Selected references

- McBeth, J., Prescott, G., Scotland, G., Lovell, K., Keeley, P., Hannaford, P., ... & Macfarlane, G. J. (2012). Cognitive behavior therapy, exercise, or both for treating chronic widespread pain. Archives of internal medicine, 172(1), 48-57.
- Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. Statistical science, 25(1), 51-71.
- Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychological methods, 18(2), 137.
- McCandless, L. C., Gustafson, P., & Levy, A. (2007). Bayesian sensitivity analysis for unmeasured confounding in observational studies. Statistics in medicine, 26(11), 2331-2347.