CAUSALLY INTERPRETABLE META-MEDIATION ANALYSIS

Marie-Félicia Beclin, Tat Thang Vo

University Paris-Est Créteil, Research Group EPIDERME, Institut Mondor de Recherche Biomédicale, Inserm U955

ABSTRACT

We propose a novel Mediation Meta-Analysis (MMA) framework that addresses these limitations. Our approach **transports study-specific natural indirect effect estimates to a well-defined target population prior** to evidence synthesis. Using semiparametric theory, we construct **flexible, data-adaptive estimators for the target parameter**. Novel random-effects MMA models are also developed to decompose **between-study heterogeneity** into distinct sources that may affect the causal interpretability of the obtained findings.

CAUSAL ESTIMANDS AND ASSUMPTIONS

Observed data: $O = (A, M, Y, C, S)$ with binary treatment A, mediator M, outcome Y, baseline covariates C, and study indicator S.

Potential outcomes:
- $Y_i(a, k, m)$: Outcome for individual i under treatment a, mediator m, in study k.
- $M_i(a, p)$: mediator under treatment a in study p.

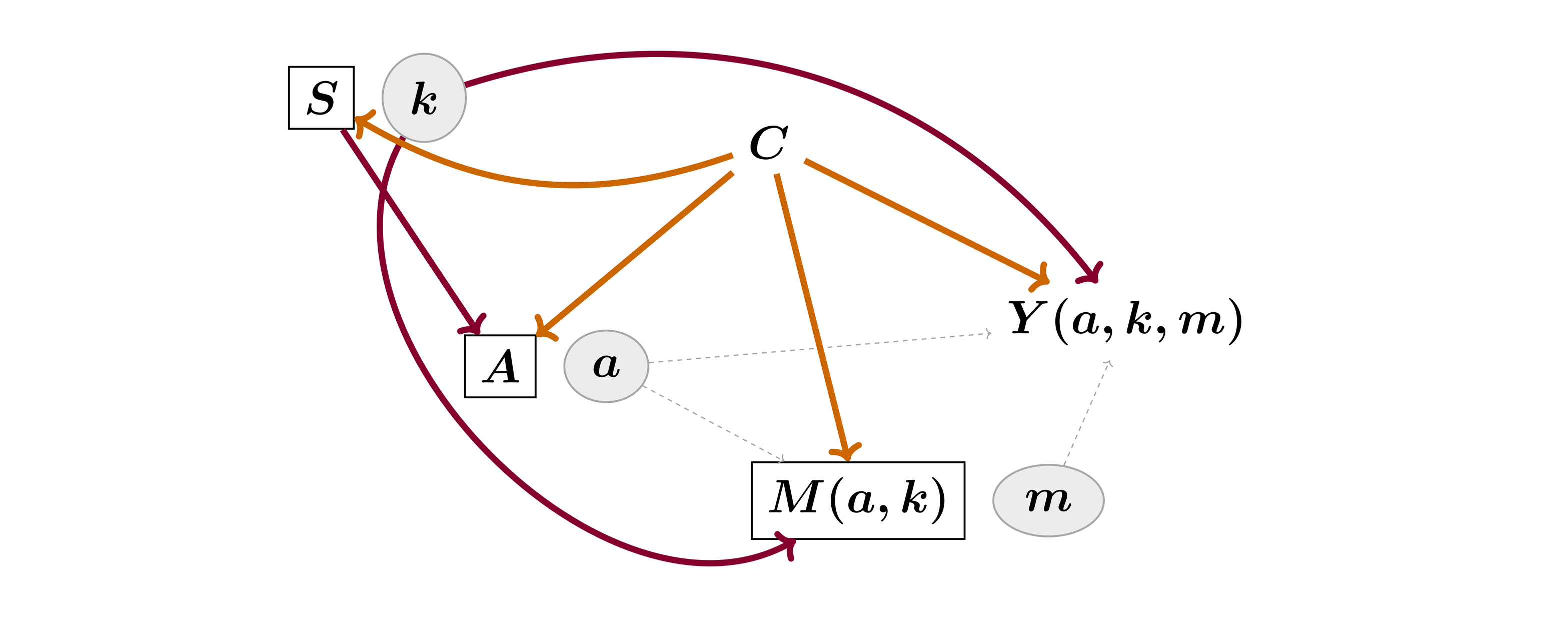
Heterogeneity across studies: $Y(a, j, m) \neq Y(a, k, m)$ and $M(a, j) \neq M(a, k)$, which captures realistic differences in treatment implementation (protocols, dosage, administration routes).

Target Causal Estimands: $\theta_{a,a^*} = \mathbb{E}\Big[Y(a, k, M(a^*, p)) \mid S = j\Big]$.

Transported Natural Indirect Effect (NIE):
- Risk difference: $\theta_{a,1} - \theta_{a,0}$
- Risk ratio: $\theta_{a,1}/\theta_{a,0}$
- Odds ratio: $\frac{\theta_{a,1}/(1-\theta_{a,1})}{\theta_{a,0}/(1-\theta_{a,0})}$

NIE transported to a target population j_0: $\zeta_{j_0,k,p}(a)$

Single World Intervention Graph



In addition to classic causal assumptions (consistency, conditional unconfoundedness), we assume:
- **Cross-World Independence**: $Y(a, k, m) \perp\!\!\!\perp M(a^*, p) \mid C$
- **Positivity (overlap)** across studies and the target population $\forall j\ 0 < \mathbb{P}(S = j \mid C, M, a) < 1$
- **Conditional transportability**:
  - $M(a, p) \perp\!\!\!\perp S \mid C$
  - $Y(a, k, m) \perp\!\!\!\perp S \mid C$

Identification Results

G-formula:

$$\mathbb{E}_C\left[\mathbb{E}_M\left[\mathbb{E}\left[Y \mid C, M, a, k\right] \mid C, a^*, S = p\right] \mid S = j\right] \quad (1)$$

IPW 1:

$$\frac{1}{\mathbb{P}(S = j)}\mathbb{E}\left[Y\, I(a, k)\ \Omega(C, M)\right] \quad (2)$$

IPW 2:

$$\frac{1}{\mathbb{P}(S = j)}\mathbb{E}\left[I(a^*, p)\ \ \mathbb{E}\left[Y \mid C, M, a, k\right]\ W(C)\right] \quad (3)$$

ESTIMATIONS

Advantages : No need to estimate mediator density and no restriction in the mediator modality (binary, continuous, multidimensional, ect ...)

Parametric Estimations: Variance by M-Estimation
- Fit $\widehat{Q}(C, M)$ to estimate $\mathbb{E}(Y \mid C, M, a, k)$ in study k among individuals receiving treatment a.
- Predict for individuals in study p under treatment a^*, then fit $\widehat{b}(C)$ to these predictions. This yields an estimator for $\mathbb{E}\left[\mathbb{E}(Y \mid C, M, a, k) \mid C, a^*, p\right]$.

Semi-parametrics Estimations: Influence function derivation gives you the plugin bias :

$$n_j^{-1}\left\{\sum_{S_i=k;A_i=a}\widehat{\Omega}(O_i)(Y_i - \widehat{Q}(O_i)) + \sum_{S_i=p}\widehat{W}(O_i)(\widehat{Q}(O_i) - \widehat{b}(O_i))\right\}$$

One-Step AND *Target Maximum Likelihood* ESTIMATORS

One-Step Estimators $\widehat{\theta}^{naive}$ + bias

Targeted Maximum Likelihood Estimators

Algorithm 1 Two-Step TMLE Targeting Algorithm

REQUIRE: Data $\{(Y_i, C_i, M_i, A_i, S_i)\}_{i=1}^n$; initial estimators $\widehat{Q}$, $\hat{b}$; clever covariates $\widehat{\Omega}$, $\widehat{W}$.

Solving $\displaystyle\sum_{i:S_i=k,A_i=a}\widehat{\Omega}(C_i, M_i)\Big[Y_i - \widehat{Q}(C_i, M_i) - \alpha\,\widehat{\Omega}(C_i, M_i)\Big] = 0,$

Update $\widehat{Q}^{(1)} = \widehat{Q} + \hat{\alpha}\,\widehat{\Omega}.$

Solving $\displaystyle\sum_{i:S_i=p,A_i=a^*}\widehat{W}(C_i)\Big[\widehat{Q}^{(1)}(C_i, M_i) - \hat{b}(C_i) - \delta\,\widehat{W}(C_i)\Big] = 0,$

Update $\widehat{b^{(2)}} = \widehat{b} + \hat{\delta}\,\widehat{W}.$

TMLE: $\displaystyle\widehat{\theta}^{\mathrm{TMLE}} = \frac{1}{n_j}\sum_{i:S_i=j}\widehat{b}^{(2)}(C_i).$

Double Robustness: *One-Step* Estimator remains unbiased if either the weights $(\widehat{\Omega}, \widehat{W})$ or the outcome models $(\widehat{Q}, \widehat{b})$ are correctly specified.

META-ANALYSIS STEPS AND HETEROGENEITIES

We consider the NIE, we have $\zeta_{j_0,k,p}(a)$

$$\hat{\zeta}_{j_0}(a) = \zeta_{j_0}(a) + \epsilon \qquad \epsilon \sim \mathcal{N}(0, \Sigma)$$

$$\zeta_{j_0,k,p}(a) = Z_{j_0}(a) + \gamma_k + \alpha_p \qquad Var(\gamma_k) = \xi^2 \text{ and } Var(\alpha_p) = \eta^2$$

with $Z_{j_0}(a) = (z, \dots, z)^T$

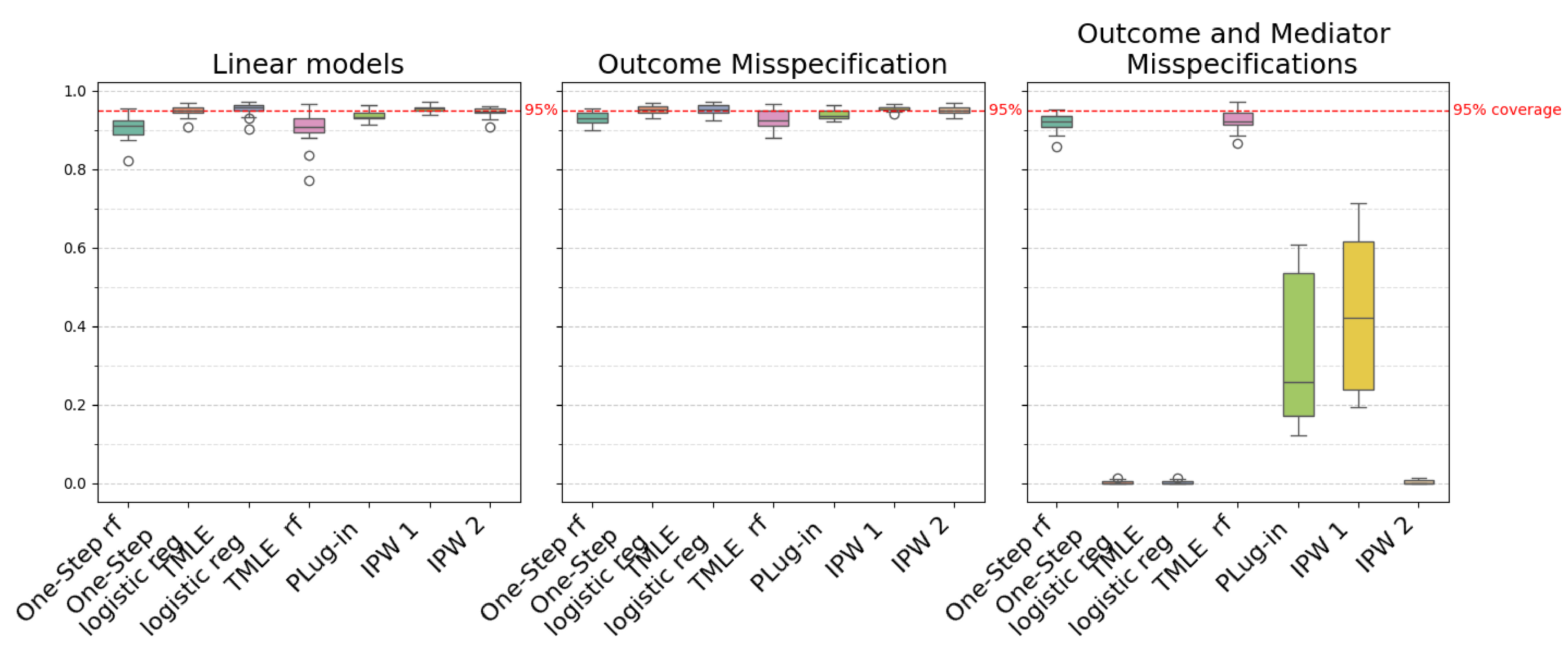
Assumptions : $\forall k', k\ Var\big[\zeta_{j_0,k,p} \mid k'\big] = Var\big[\zeta_{j_0,k,p} \mid k\big] = \eta^2$ and symetrically for condition variance on p

ξ^2 treatment-outcome heterogeneity

η^2 treatment-mediator heterogeneity

We use nonparametric estimations based on law of total variance and under assumptions we have : $\mathbb{E}\Big[Var\big[\zeta_{j_0,k,p} \mid k\big]\Big] = Var\Big[\mathbb{E}\big[\zeta_{j_0,k,p} \mid p\big]\Big]$.

SIMULATIONS



Coverage of 95% Confidence Intervals, $n_{\text{sample}} = 20000$, $n_{\text{simulations}} = 500$

COMBINING VARIETY OF NATURE OF STUDIES

In plug-in estimators and IPW, outcome data are not required for populations in P. **Studies that only investigate the causal path $A \rightarrow M$ can be included.**

For populations in K, in plug-in and IPW2, an outcome model is needed for individuals in the population receiving treatment a.

Studies that investigate the causal path $M \rightarrow Y$ under treatment a can also be included.

Survival Data and binary data can also been mixed in the same analysis.

PERSPECTIVES

- Simulations with survival and Real-life data application
- Construct consistent estimators to include aggregated data