Causal mediation in models with categorical latent variables: an extended approach for analysis of complex pathways in epidemiology

Leila Amorim

Federal University of Bahia, Brazil

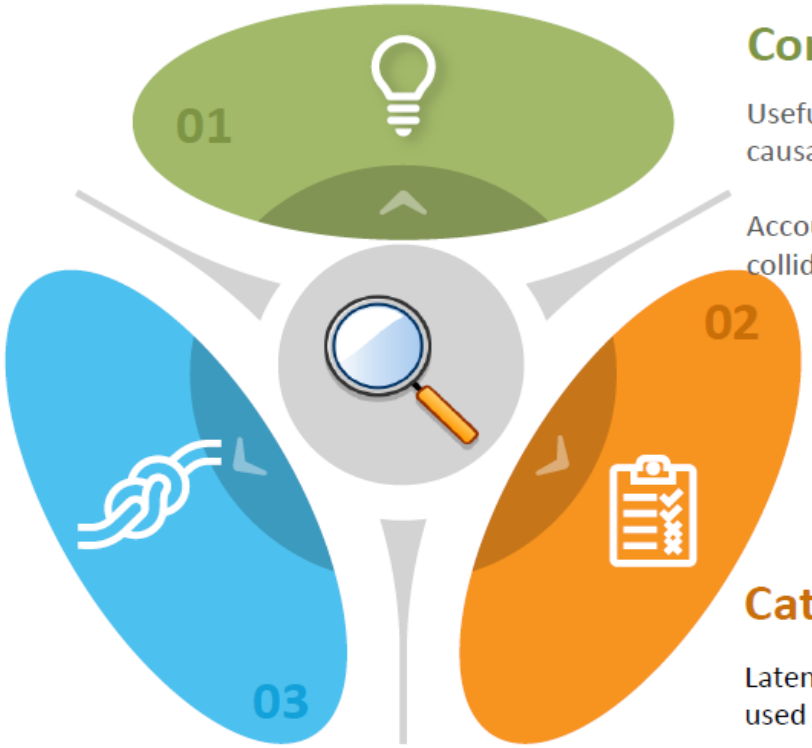
25 September 2024

Joint work with Michelle Pereira, Rosana Aquino & Marcelo Taddeo

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Background

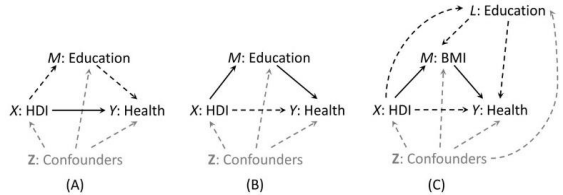
Methodological Triad



Complex Causal Pathways

Useful for improving understanding of the causal mechanisms of interventions.

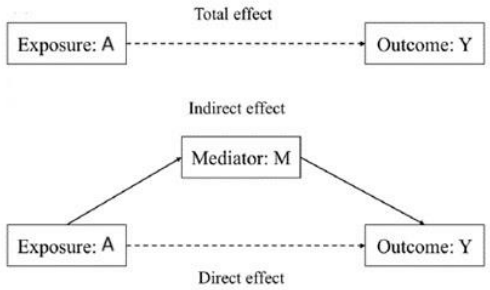
Accounting for confounding, mediators, and colliders.



Causal Mediation

Relevant to measure the individual contribution of each path between exposure and outcome.

It demands assumptions, study design, and estimation strategies to allow causal interpretation.

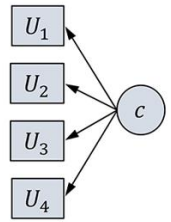


Categorical Latent Variables

Latent variables comprise abstract concepts used to explain phenomena.

They are not measured directly (without errors) – based on indicators.

LCA and extensions: categorical constructs and indicators.



WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Causal mediation

- The causal effects associated with the different paths are called **Natural Direct and Indirect Effects**:

$$NDE = E[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}] \quad \text{and}$$

$$NID = E[Y_{aM_a} - Y_{aM_{a^*}}]$$

- Using counterfactual notation

Y_a = the potential response for treatment a.
When A is binary: Y_a and Y_{a^*}, leading to $Y_{aM_{a^*}}$.

(Barron & Kenny, 1998; MacKinnon, 2008; Pearl, 2001; Imai et al.,2010; Tchetgen and Shpitser, 2012; VanderWeele, 2016)

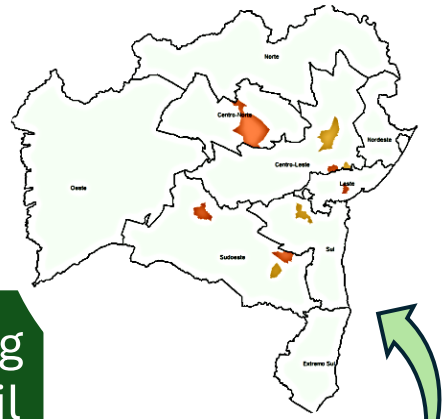
- The **estimators for NDE e NIE** for binary outcomes can be expressed on the odds ratio (OR) scale:
  - Depending on the assumption of rare events.

Methods assume that all variables are observed.

(Valeri & VanderWeele, 2013; Doretti et al.,2021)

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Motivation

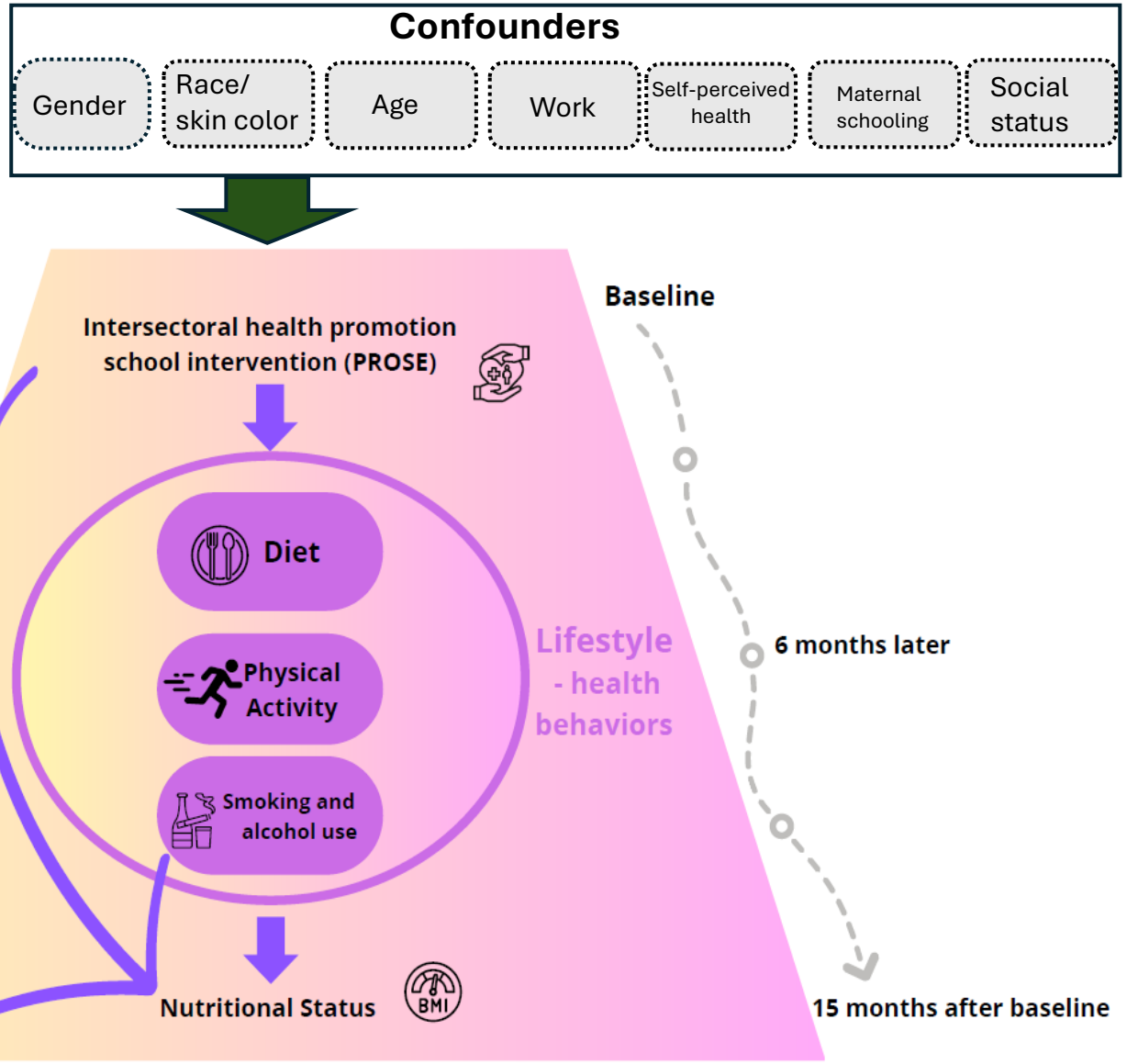


PROSE Study (Promoting Health in School) - Brazil

Randomized controlled trial conducted at the school level

10 municipalities/
20 schools

1,240 high school students follow-up for 15 months



PROSE Intervention: Set of measures to enhance the health and education of schoolchildren and their families (workshops, lectures, professional training, ...)

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Goals

Propose methodological extensions for causal mediation analysis in the presence of categorical latent variables.

Decompose the total effect of an intersectoral health intervention on the nutritional status of adolescents into direct and indirect effects.

Mediator: *latent* lifestyles (diet, smoking and alcohol use, physical activities).

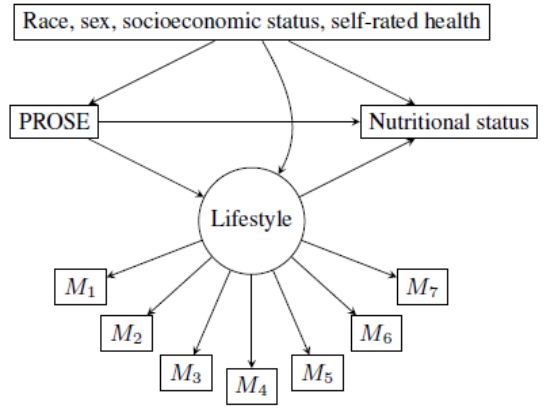


Figure 1 - Causal diagram relating the PROSE Study and nutritional status of schoolchildren mediated by the construct lifestyle.

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Methods

- The estimation of NDE and NIE is extended using methods discussed previously (Valeri & VanderWeele, 2013; Doretti et al, 2021; Hsiao et al, 2021) to:

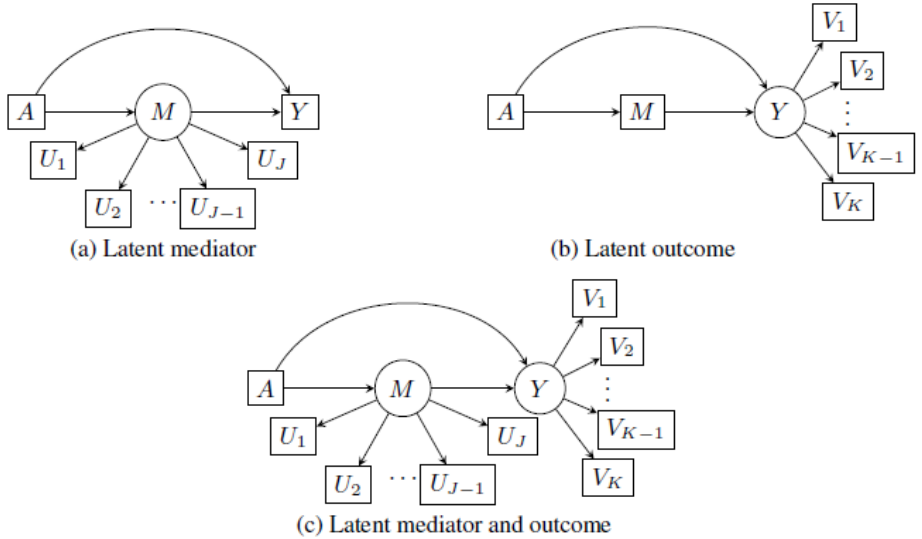
Incorporate categorical latent variables and use of estimation methods for LCA with external variables (covariates and distal outcomes)

include categorical latent mediator and/or outcome; confounders and interaction between exposure and mediator

expand the methodology to models with more than two latent classes

explore the effect of assuming (or not) rare events on the performance of the methods

study the impact of causal identification criteria on performance of the proposed estimators



(a) Latent mediator

(b) Latent outcome

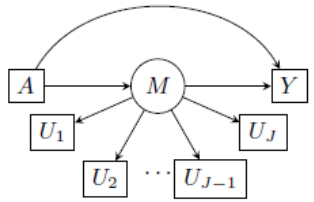
(c) Latent mediator and outcome

Software: Mplus (v 8.11) and R (v.4.2.1)

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Extended Approaches for Causal Mediation with Latent Categorical Variables

Steps for the Regression-Model Procedure:



Special Case: Latent Mediator(M) and Observed Outome (Y)
Binary Mediator and Outcome

1. Fit LCA with covariates for the Mediator (M)

$$P\left(\boldsymbol{U}=\boldsymbol{u}|\tilde{\boldsymbol{X}}\right)=\sum_{m=1}^{K}\gamma_m(\tilde{\boldsymbol{X}})\prod_{j=1}^{J}\prod_{r_j=1}^{R_j}\rho_{j,r_j|m}^{\mathbb{1}(u_j=r_j)} \qquad \tilde{\boldsymbol{X}}=(A,\boldsymbol{X})$$

$$\gamma_m(\tilde{\boldsymbol{X}})=P(M=m|\tilde{\boldsymbol{X}})=\exp\{\tilde{\boldsymbol{X}}^\top\boldsymbol{\beta}_m\}/\textstyle\sum_{m'=1}^{K}\exp\{\tilde{\boldsymbol{X}}^\top\boldsymbol{\beta}_{m'}\}$$

$$\rho_{j,r_j|m}=P(U_j=r_j|M=m)$$

Maximum likelihood estimation
Conditional independence assumption

2. Fit a distal outome model for Y

$$\operatorname{logit}\mathrm{P}(Y=1|a,m,x)=\theta_0+\theta_1a+\theta_2m+\theta_3am+\theta_4^\top x\equiv\eta_{am}^Y(x)$$

$$q_{m_2,m_1}=P(\mathcal{N}=m_1,M=m_2)$$

BCH-3 step estimation (incorporation of measurement error)

(Bolck, Croon & Hagenaars, 2004; Vermunt, 2010; Asparouhov & Muthèn, 2014)

$$\operatorname{logit}\mathrm{P}(M=1|a,x)=\beta_0+\beta_1a+\beta_2^\top x\equiv\eta_a^M(x)$$

3. Plug-in estimates in the NDE and NIE expressions (β,θ)

Assumption that the causal identifiability criteria are valid.

Method 1 :

$$\mathrm{NIE}^{\mathrm{OR}}(x)=\frac{(1+e^{\eta_{a^*}^M(x)})(1+e^{\nu_{aa}(x)})}{(1+e^{\eta_a^M(x)})(1+e^{\nu_{aa^*}(x)})}$$

$$\mathrm{NDE}^{\mathrm{OR}}(x)=\frac{e^{\theta_1a}\left(1+e^{\nu_{aa^*}(x)}\right)}{e^{\theta_1a^*}(1+e^{\nu_{a^*a^*}(x)})} \qquad \nu_{aa^*}(x)=\theta_2+\theta_3a+\beta_0+\beta_1a^*+\beta_2^\top x.$$

(Valeri and VanderWeele, 2013)

Method 2 :

$$\log\mathrm{NIE}^{\mathrm{OR}}(x)=\log\frac{A_{a,a}(x)}{A_{a,a^*}(x)} \qquad \log\mathrm{NDE}^{\mathrm{OR}}(x)=\theta_1(a-a^*)+\log\frac{A_{a,a^*}(x)}{A_{a^*,a^*}(x)}$$

$$A_{a,a^*}(x)=\frac{e^{\theta_2+\theta_3a}e_{a^*}^M(x)\left\{1+e_{a0}^Y(x)\right\}+1+e_{a1}^Y(x)}{e_{a^*}^M(x)\left\{1+e_{a0}^Y(x)\right\}+1+e_{a1}^Y(x)}$$

$$e_{am}^Y(x)=\exp\{\eta_{am}^Y(x)\} \text{ and } e_a^M(x)=\exp\{\eta_a^M(x)\}$$

Standard errors based on delta method.

(Doretti et al, 2021)

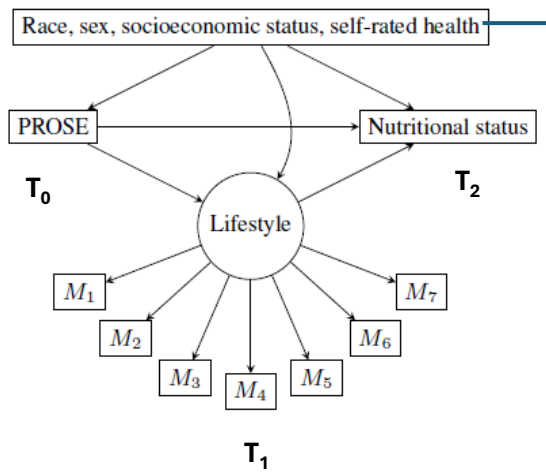
WCE WORLD CONGRESS OF EPIDEMIOLOGY 2024

Data Collection and Descriptives

Instruments

Questionnaires

1. Socioeconomic and demographic data
2. International Physical Activity Questionnaire - IPAQ
3. Smoking and Alcohol consumption
4. Food Frequency Questionnaire - FFQ
5. Measurement of nutritional status (weight, height)



30.2% females, age 13-19 y-old (mean =16), 70.9% black or brown, 13.5% work **(T_0)**

M1: Moderate or vigorous physical activity (insufficient: < 300 min/week)

M2: Leisure physical activity (low: < 60 min/week)

M3: Screen Time (high: >= 4h/day)

M4: Smoking and alcohol consumption (yes: use in the last month)

M5: Binge Drinking (yes: 6 or more drinks on one occasion in the last year)

M6: Consumption of health food based on daily consumption of fruits, vegetables, milk, ... (low: < 2082gr/day - 75%ile)

M7: Energy (Kcal) consumed daily in ultra-processed foods (high: >50% of the diet)

Undesired behaviors

M1: 97.4% **(T_1)**
M2: 56.7%
M3: 69.8%
M4: 12.7%
M5: 12.6%
M6: 21.0%
M7: 75.7%

Body Mass Index (BMI) calculated based on age and sex (WHO, 2007).

Classification: overweight (85th-97th %ile) and obese (> 97th %ile)

12.5% overweight/obese **(T_2)**

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Results

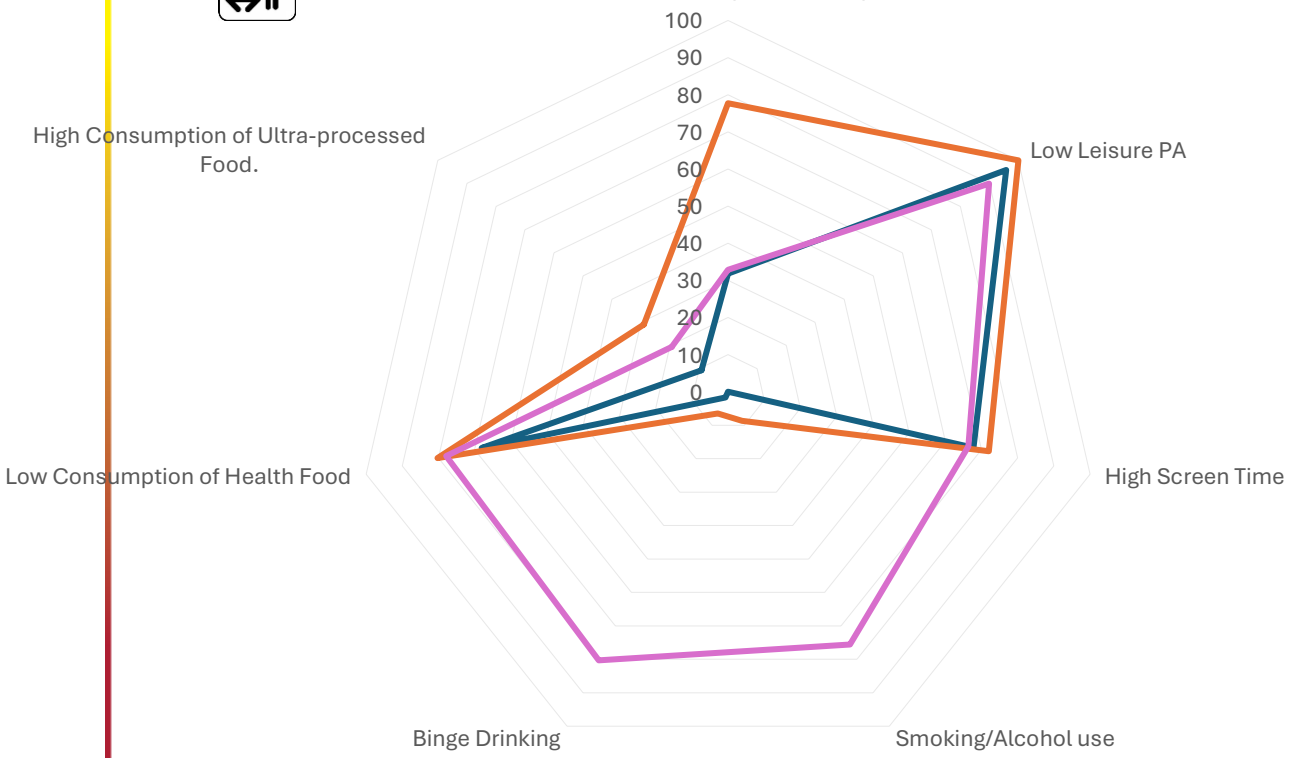
Latent Classes Analysis

Lifestyle (Mediator)

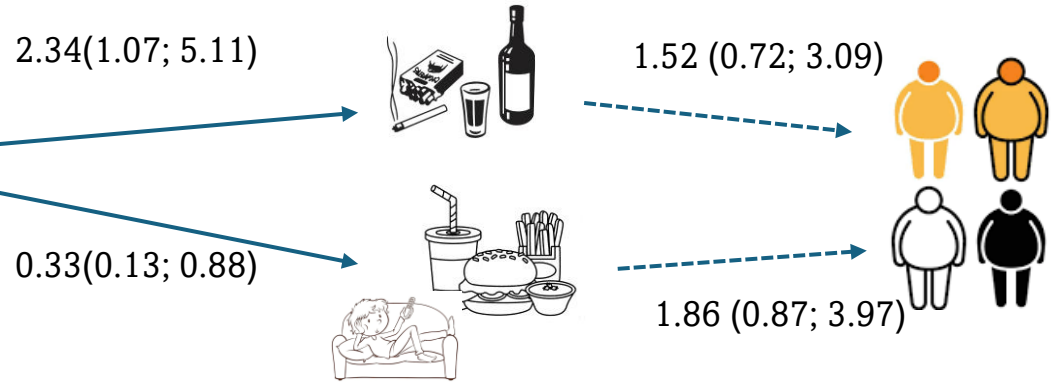
Best health behaviors (33.5%)

Poor diet and insufficiently active (56.5%)

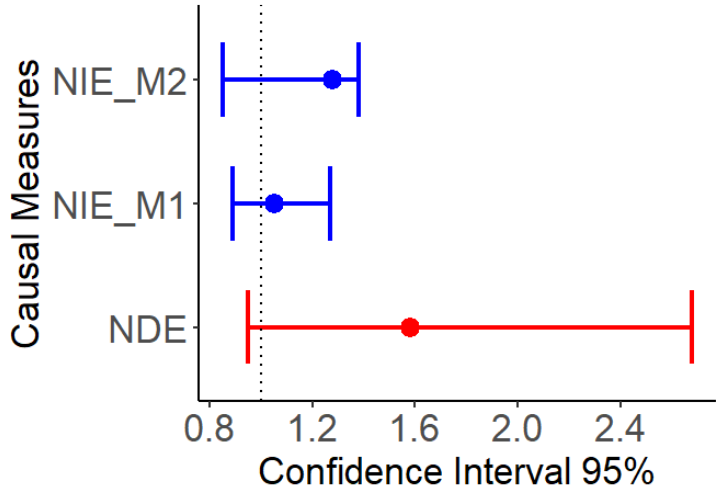
Alcohol and tobacco users (10.0%)



Entropy: 0.69



Causal Mediation Analysis



WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Final Remarks

- We have extended previous approaches for causal mediation analysis, allowing for the inclusion of latent categorical variables:
  - The new methodologies are helpful in evaluating the causal effects (direct and indirect) of interventions in health,
  - Their performance was evaluated through simulation studies (not presented here),
  - Performance may depend on the outcome prevalence, measurement error (entropy), and sample size.

- Several applications in Epidemiology might benefit from these methodological developments.

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Thank You!

Acknowledgments

fapesb CNPq

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024