Medflex: flexible mediation analysis using natural effect models in R

Johan Steen¹, Tom Loeys¹, Beatrijs Moerkerke¹, Theis Lange², Stijn Vansteelandt¹
¹Ghent University, ²University of Copenhagen

Beyond the mediation formula: in search of flexibility and parsimony

Mediation analysis is routinely adopted in a wide range of applied disciplines as a statistical tool to disentangle the causal pathways by which an exposure X affects an outcome Y. Within the counterfactual framework, the mediation formula, can be considered the predominant vehicle for effect decomposition.

\[
\begin{align*}
\text{mediation formula} & \quad = \int E(Y|X=x_0, M=m, C) \times dF(M=m|X=x_1, C) \\
\logit(Y=1|X,M,C) & \quad = \theta_0 + \theta_1 X + \theta_2 M + \theta_3 XM + \theta_4 C \\
E(M|X,C) & \quad = \gamma_0 + \gamma_1 X + \gamma_2 C
\end{align*}
\]

- Despite widespread application, the mediation formula often produces complex expressions for natural direct and indirect effects.
- For instance, even if no modification by X and/or covariate C levels are allowed for in the working models (for the outcome Y and mediator M), the resulting expressions may depend on X and/or C in a complicated way.
- This makes results difficult to report and hypotheses infeasible (or even impossible) to test and may hence pose an impediment to routine application of the mediation formula.

Fitting natural effect models and making statistical inferences using R package medflex¹ in three simple steps

1. Create a hypothetical dataset by expanding the original data along unobserved \((x_0, x_1)\) combinations and...

What’s in it for practitioners?

- handles a larger class of parametric working models than software applications that rely on closed-form expressions
- embedded within framework of existing model-fitting functions in R (mainly glm), allowing estimation on most natural (mostly multiplicative) effect scale (e.g., odds ratios)
- simplifies testing, especially when dealing with continuous exposures or covariates, as hypotheses of interest can be captured by (a linear combination of) target model parameters
- provides robust standard errors (for glm working models): less computer-intensive than bootstrap or Monte Carlo integration

2. Fit a natural effect model to the expanded data:

![Estimates for different effect components](image)

Many thanks to Patrick Corrigan for granting permission to reproduce his cartoon

References


correspondence: johan.steen@ugent.be  medflex is freely available on CRAN: http://cran.r-project.org/web/packages/medflex/index.html