Medflex: flexible mediation analysis using natural effect models in R

Johan Steen[1], Tom Loeys[1], Beatrijs Moerkerke[1], Theis Lange[2], Stijn Vansteelandt[1]

[1]Ghent University, [2]University of Copenhagen

UNIVERSITEIT GENT

UNIVERSITY 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.

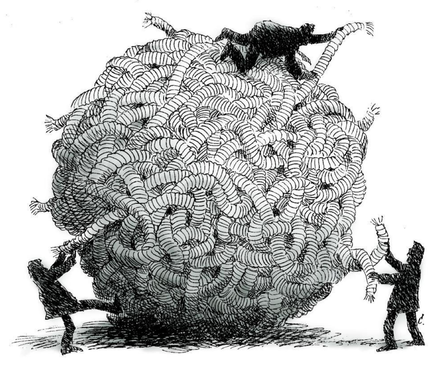
working models

$$\text{logit}P\{Y=1|X,M,C\} = \theta_0+\theta_1X+\theta_2M+\theta_3XM+\theta_4C$$

$$E(M|X,C) = \gamma_0+\gamma_1X+\gamma_2C$$

mediation formula

$$\int E(Y|X=x_0,M=m,C) \times dF(M=m|X=x_1,C)$$



- 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 still 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.

Natural effect models

Alternatively, *natural effect models* focus on *direct parameterization* of the natural direct and indirect effects of interest (Lange, Vansteelandt & Bekaert, 2012; Vansteelandt, Bekaert & Lange, 2012).

Fitting natural effect models entails the use of well-established missing data methods.

natural effect model

$$\text{logit}P\{Y(x_0,M(x_1))=1|C\} = \beta_0+\beta_1x_0+\beta_2x_1+\beta_3x_0x_1+\beta_4C$$

natural direct effect odds ratio

$$\frac{\text{odds}\{Y(\mathbf{1},M(x_1))|C\}}{\text{odds}\{Y(\mathbf{0},M(x_1))|C\}} = \exp(\beta_1+\beta_3x_1)$$

natural indirect effect odds ratio

$$\frac{\text{odds}\{Y(x_0,M(\mathbf{1}))|C\}}{\text{odds}\{Y(x_0,M(\mathbf{0}))|C\}} = \exp(\beta_2+\beta_3x_0)$$

Fitting natural effect models and making statistical inferences using R package *medflex*[1] in three simple steps

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

i	X_i	x_0	x_1	$Y_i(x_0,M_i(x_1))$
1	1	1	1	Y_1
1	1	1	**0**	**?**
1	1	**0**	1	**?**
1	1	**0**	**0**	**?**
2	0	0	0	Y_2
2	0	0	**1**	**?**
2	0	**1**	0	**?**
2	0	**1**	**1**	**?**
⋮	⋮	⋮	⋮	⋮

WEIGHTING-BASED APPROACH[2]

fit a model for the **mediator** distribution and calculate regression **weights**

$$w_i = p_i(x_1)/p_i(x_0) = \frac{\hat{P}(M_i|X_i=x_1,C_i)}{\hat{P}(M_i|X_i=x_0,C_i)}.$$

in a single R command:

```
expData <- neWeight(M~X+C,
 family=gaussian, data=data)
```

i	X_i	x_0	x_1	$Y_i(x_0,M_i(x_1))$	w_i
1	1	1	1	Y_1	1
1	1	1	**0**	Y_1	$p_1(\mathbf{0})/p_1(1)$
2	0	0	0	Y_2	1
2	0	0	**1**	Y_2	$p_2(\mathbf{1})/p_2(0)$
⋮	⋮	⋮	⋮	⋮	⋮

IMPUTATION-BASED APPROACH[3,4]

or fit a model for the **outcome** mean and **impute** unobserved $Y_i(x_0,M_i(x_1))$ with

$$\hat{Y}_i(x_0,M_i) = \hat{E}(Y_i|X_i=x_0,M_i,C_i).$$

in a single R command:

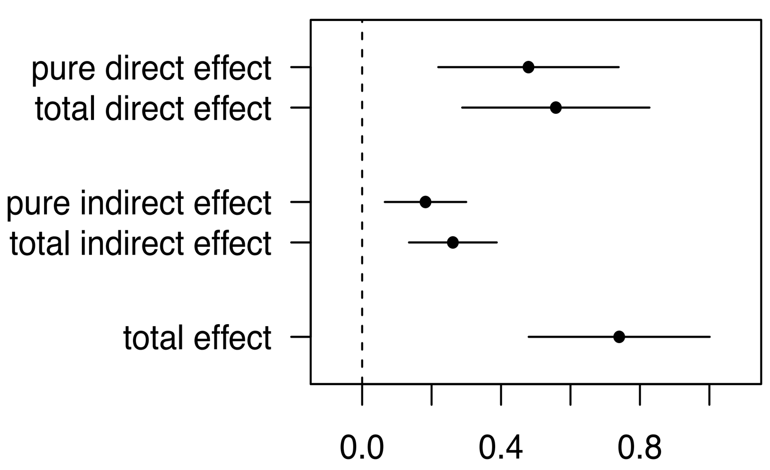
```
expData <- neImpute(Y~X*M+C,
 family=binomial, data=data)
```

i	X_i	x_0	x_1	$Y_i(x_0,M_i(x_1))$
1	1	1	1	Y_1
1	1	**0**	1	$\hat{Y}_1(\mathbf{0},M_1)$
2	0	0	0	Y_2
2	0	**1**	0	$\hat{Y}_2(\mathbf{1},M_2)$
⋮	⋮	⋮	⋮	⋮

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

```
fit <- neModel(Y~X0*X1+C,
 family=binomial, expData=expData)
```

```
                      Estimate
pure direct effect      0.4790
total direct effect     0.5578
pure indirect effect    0.1824
total indirect effect   0.2613
total effect            0.7403
```



3 Utility functions for **effect decomposition**,

```
neEffdecomp(fit)
```

or general linear hypotheses

```
neLht(fit)
```

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) targeted model parameters
- ✓ provides *robust standard errors* (for `glm` working models): *less computer-intensive* than bootstrap or Monte Carlo integration

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

References

[1]Steen, J., Loeys, T., Moerkerke, B., & Vansteelandt, S. (2015). Medflex: An R Package for Flexible Mediation Analysis Using Natural Effect Models. *Submitted Manuscript.*

[2]Lange, T., Vansteelandt, S., & Bekaert, M. (2012). A Simple Unified Approach for Estimating Natural Direct and Indirect Effects. *American Journal of Epidemiology, 176*(3), 190–195.

[3]Vansteelandt, S., Bekaert, M., & Lange, T. (2012). Imputation Strategies for the Estimation of Natural Direct and Indirect Effects. *Epidemiologic Methods, 1*(1), Article 7.

[4]Loeys, T., Moerkerke, B., De Smet, O., Buysse, A., Steen, J., & Vansteelandt, S. (2013). Flexible Mediation Analysis in the Presence of Nonlinear Relations: Beyond the Mediation Formula. *Multivariate Behavioral Research, 48*(6), 871–894.

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

fwo Research Foundation Flanders Opening new horizons