

Motivation

- Houtpen *et al.* [1] investigated the mediating role of DNA methylation in the association between childhood trauma and cortisol stress reactivity.
- First, they conducted association analysis between methylation loci and both childhood trauma and cortisol stress reactivity, and they identified three loci (*KITLG*: cg27512205, *C1QTNF2*: cg05608730, *JAZF1*: cg26179948) potentially involved in this relationship.
- Given the strongest association of locus cg27512205 with cortisol stress reactivity, replicated in two independent samples, the authors performed a mediation analysis targeting this locus (Figure 1, solid black arrows).

Key

The other two loci correlate with the locus cg27512205 and are associated with both childhood trauma and stress reactivity (Figure 1), potentially biasing the standard analysis of *KITLG*.

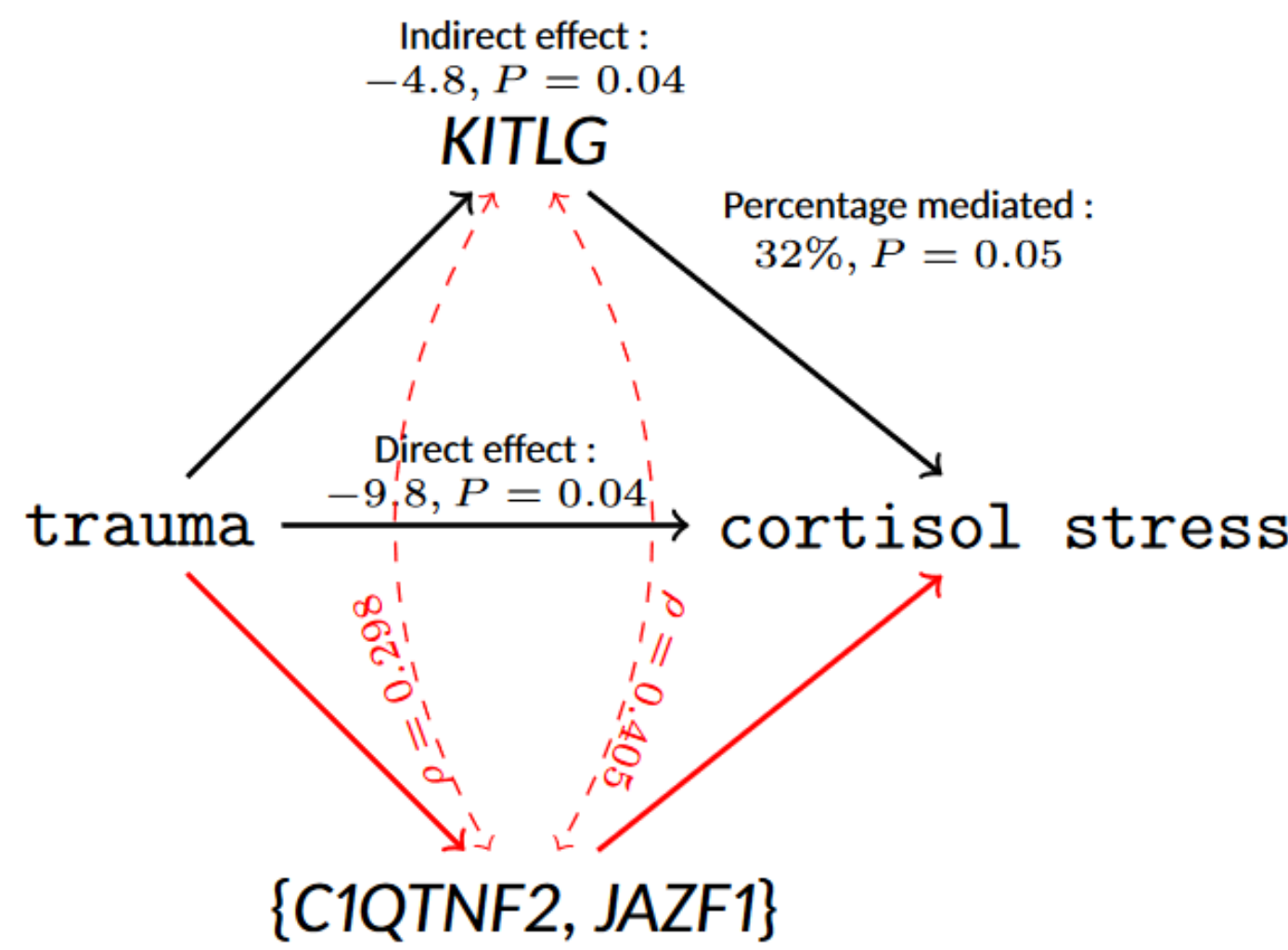


Figure 1. Causal diagram of the original study (black arrows) and its extension considering the other two loci (red arrows)

Objective

To assess how expanding the analysis to consider the two additional mediators affects the estimation of the direct and indirect effects through the *KITLG* gene locus.

Methods

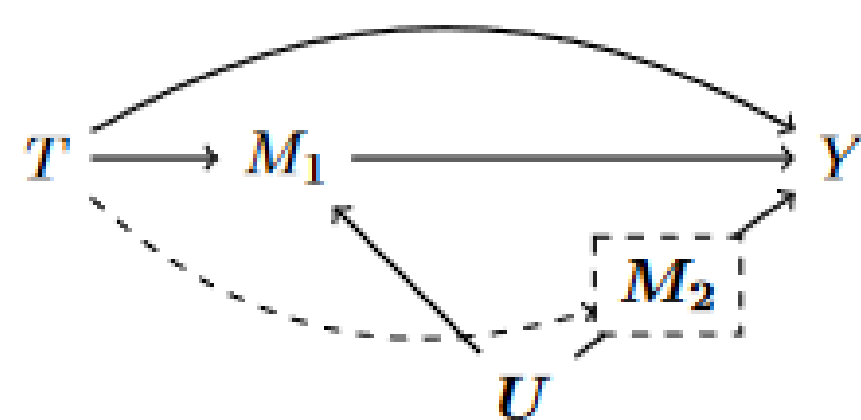


Figure 2. Multiple-mediator model. T is the exposure; M_1 is the mediator of interest; M_2 represents secondary mediators; Y is the outcome; U is a set of unobserved confounders for the relationships between mediators; the vector of pretreatment confounders X is not shown.

1. Simple effects

In the setting with only M_1 as the mediator of interest, let $Y(t, M_1(t'))$ be the value that Y would take if T were set to t and M_1 to $M_1(t')$.

The corresponding natural direct effect (DE_1^s) and natural indirect effect (IE_1^s) for a binary exposure (level 1 vs 0) are:

$$DE_1^s = \mathbb{E}\{Y(1, M_1(0)) - Y(0, M_1(0))\},$$

$$IE_1^s = \mathbb{E}\{Y(1, M_1(1)) - Y(1, M_1(0))\}.$$

Standard causal mediation analysis allows to estimate these effects under Sequential Ignorability Assumption (SIA) [2].

- Back to our study diagram, since U confounds the effect of M_1 on Y through the mediators M_2 , the assumption of no-unmeasured confounders in the mediator-outcome relationship, as part of the SIA, is violated (Figure 2).

- To account for the correlation between mediators due to the unmeasured common causes U , we consider M_1 together with the other mediators M_2 .

2. Joint effects

Let $Y(t, M_1(t'), M_2(t''))$ be the value that Y would take when T is set to t , M_1 to $M_1(t')$ and M_2 to $M_2(t'')$. We define the joint natural effects as:

$$DE = \mathbb{E}\{Y(1, M_1(0), M_2(0)) - Y(0, M_1(0), M_2(0))\},$$

$$IE = \mathbb{E}\{Y(1, M_1(1), M_2(1)) - Y(1, M_1(0), M_2(0))\}.$$

3. Composition assumption

- To link the simple and joint effects, we use the composition assumption:

$$Y(t, M_1(t')) = Y(t, M_1(t'), M_2(t)), \quad \forall t, t'.$$

- Under this assumption, we relate the natural direct and indirect effects of M_1 to the joint effects as follows:

$$DE_{1,2}^s = DE + \mathbb{E}\{Y(1, M_1(0), M_2(1)) - Y(1, M_1(0), M_2(0))\}$$

$$IE_{1,2}^s = \mathbb{E}\{Y(1, M_1(1), M_2(1)) - Y(1, M_1(0), M_2(1))\}.$$

4. Identification

- The joint natural effects as well as the natural effects for M_1 can be identified using the Sequential Ignorability for Multiple Mediators Assumption (SIMMA) [3]:

$$\{Y(t, m_1, m_2), M_1(t'), M_2(t'')\} \perp\!\!\!\perp T | \mathbf{X} = \mathbf{x},$$

$$Y(t', m_1, m_2) \perp\!\!\!\perp \{M_1(t), M_2(t)\} | T = t, \mathbf{X} = \mathbf{x},$$

$$Y(t, m_1, m_2) \perp\!\!\!\perp \{M_1(t'), M_2(t)\} | T = t, \mathbf{X} = \mathbf{x},$$

for all t, t', t'', m_1, m_2 .

- With these assumptions, U does not confound the relationship between (M_1, M_2) and Y because its influence on Y is entirely mediated through (M_1, M_2) .

5. Regression-based estimation

The natural effects $DE_{1,2}^s$ and $IE_{1,2}^s$ can be estimated using a regression-based approach.

For a continuous outcome with continuous mediators and no interactions, we fit the following regression models :

$$\mathbb{E}[Y|t, m_1, m_2, \mathbf{x}] = \beta_0 + \beta_1 t + \beta_2 m_1 + \beta_3^T m_2 + \beta_4^T \mathbf{x},$$

$$\mathbb{E}[M_j|t, \mathbf{x}] = \alpha_{0j} + \alpha_{1j} t + \alpha_{2j}^T \mathbf{x}, \quad \text{for } j = 1, \dots, K.$$

- Assuming SIMMA and the correlation matrix $Cor(M_1(t_1), \dots, M_K(t_K)|T, \mathbf{X}) = \boldsymbol{\rho}, \forall t_1, \dots, t_K \in \{0, 1\}^K$, the natural direct and indirect effects for M_1 are then given by

$$DE_{1,2}^s = \beta_1 + \sum_{j \neq 1} \beta_{3j} \alpha_{1j}, \quad IE_{1,2}^s = \beta_2 \alpha_{11}. \quad (1)$$

- The coefficients and the dependance structures $\boldsymbol{\rho}$ are estimated via maximum likelihood estimation (MLE).

Data application

1. Dataset

The dataset consists of 85 healthy individuals recruited from the general population at the University Medical Center, Utrecht, the Netherlands [1].

Table 1. Sample description ($N = 85$)

Characteristic	mean (range)
sex (% female)	50.59
age (in years)	33.8 (18; 69)
trauma (mean total score)	31.9 (24; 63)
<i>KITLG</i> (cg27512205)	0.159 (0.125; 0.188)
<i>C1QTNF2</i> (cg05608730)	0.380 (0.293; 0.469)
<i>JAZF1</i> (cg26179948)	0.123 (0.0925; 0.159)
cortisol stress (mean AUCi)	243.46 (−1029.85; 1876.28)

2. Analysis

- We applied our method on the dataset with cg27512205 as the primary mediator, while treating cg05608730 and cg26179948 as secondary mediators (Figure 1) :
 - models for the outcome and mediators, adjusted for age and sex, were fitted using MLE.
 - $DE_{1,2}^s$ and $IE_{1,2}^s$ were estimated via (1), and quasi-Bayesian approximation was used for confidence intervals.
 - The results from the initial study were obtained using the R package `mediation`.

3. Results

The mediation analysis results for the *KITLG* gene, comparing both methods, are shown in Figure 3.

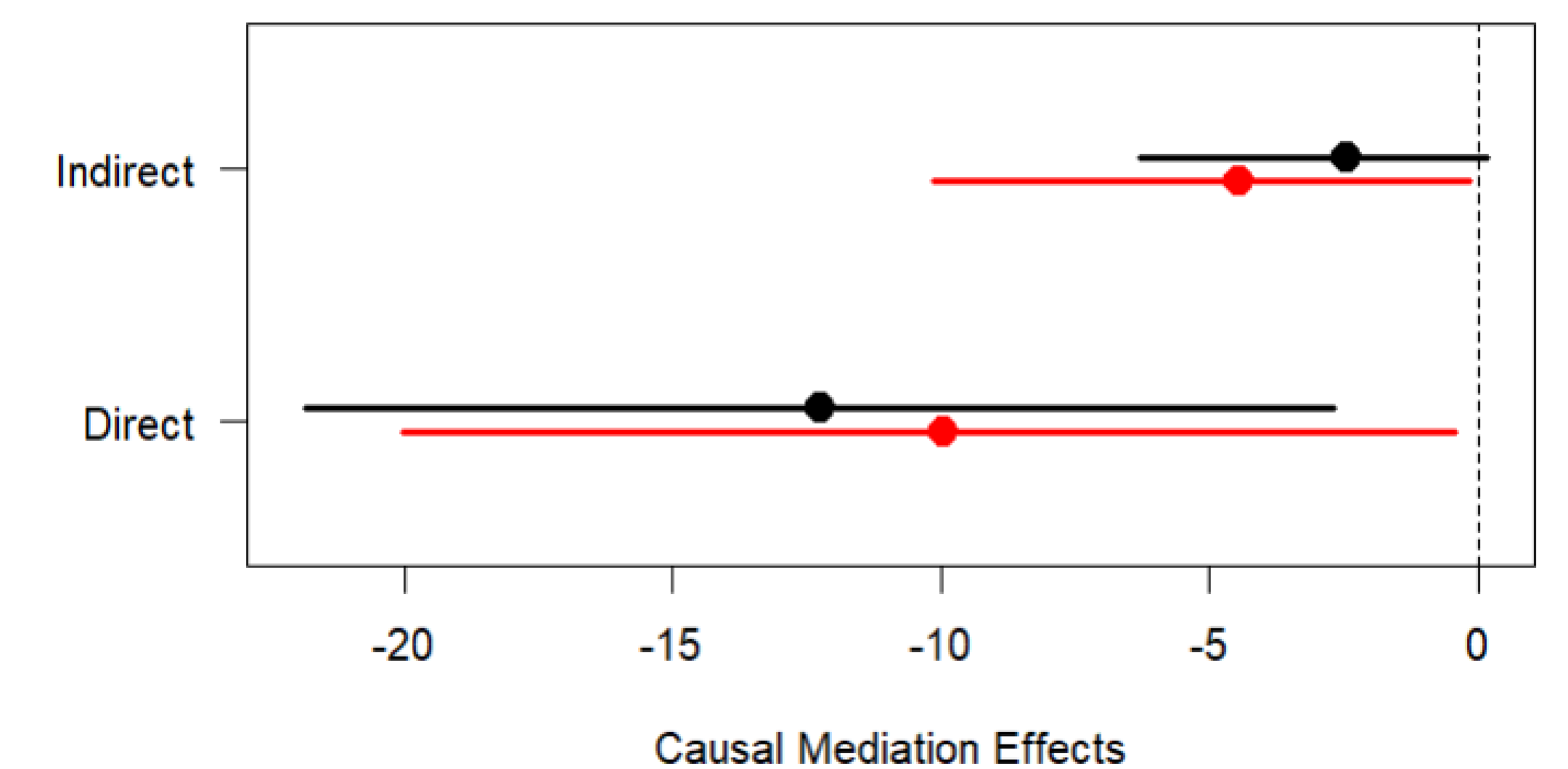


Figure 3. Causal mediation effects: comparison between our method (black) and initial study method (red)

- We found a proportion mediated by the *KITLG* gene of 16%, substantially lower than the 32% proportion mediated reported in the initial study.

Conclusion

- Standard mediation analyses involving the *KITLG* gene are likely problematic due to potential violations of the no-unmeasured confounders of the mediator-outcome relationship in the SIA.
- Simple mediation analysis of the *KITLG* gene, including the two additional mediators is thought to have reduced the confounding effects.

References

- L. C. Houtepen, C. H. Vinkers, T. Carrillo-Roa, M. Hiemstra, P. A. Van Lier, W. Meeus, S. Branje, C. M. Heim, C. B. Nemeroff, J. Mill, et al. Genome-wide dna methylation levels and altered cortisol stress reactivity following childhood trauma in humans. *Nature communications*, 7(1):10967, 2016.
- K. Imai, L. Keele, and T. Yamamoto. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1):51 – 71, 2010.
- A. Jérolon, L. Baglietto, E. Birmel , F. Alarcon, and V. Perduca. Causal mediation analysis in presence of multiple mediators uncausally related. *The International Journal of Biostatistics*, 17(2):191–221, 2021.