Multiple Methods for Multiple Reporters of Child Maltreatment

Results from the Lehigh Study

Lehigh Measurement and Causal Inference Analyses Team

2025-06-12

Table of contents

1 Background

Created With `Statamarkdown`

This document was created using the amazing package by Hemken (2023), which should be cited in the final version of the paper.

The Lehigh Study presents a unique opportunity. Data are collected on experiences of abusive discipline as reported by *administrative* reports, two *parental* reports at two different time points, and two *self* reports at two different time points. However, in the absence of a gold standard measure of abusive discipline, appropriately aggregating these multiple reports across multiple time points represents an analytic challenge.

In the manuscript below, we employ multiple strategies to estimate the relationship of these multiple reports from multiple reporters at multiple time points to a mental health outcome. We compare and contrast the advantages and disadvantages of these different methods, and conclude the manuscript with suggestions on optimal methodological approaches to confront the methodological challenges that are posed by having multiple reports from multiple reporters at multiple time points.

Note 1: Key Ideas or Questions

- Many sources in the literature note that data can come from multiple reporters. It is often suggested that it is important to *triangulate* these different sources of information, yet what is meant exactly, or operationally, by such *triangulation* is often unclear.
- How do the different methods highlight, or not highlight:
  - The differences between prospective and retrospective reports, especially in predicting outcomes?
  - The level of agreement or disagreement between different reporters?
  - Issues of severity, chronicity, timing and developmental stages across reporters?

2 The Data

```
describe
```

```
Running /Users/agrogan/Desktop/GitHub/research/CAN-special-issue/profile.do

> ...



Contains data from /Users/agrogan/Library/CloudStorage/Dropbox-UniversityofMichigan/Andrew Groga
> n-Kaylor/Lehigh measurement and causal inference analyses/Data sets/Stata data/Lehigh_working_
> concept_paper.dta
 Observations:           273
    Variables:            11                  12 Jun 2025 11:31
-------------------------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------------------------
gender          double  %10.0g     gender     gender
age             double  %10.0g     a_age      Age
marital         double  %10.0g     a_lvmar    Marital status
SR2             double  %10.0g     o_ptsdchab
                                              As a child, were you ever badly beaten up by your
                                                parents or the people who rais
annual_income   double  %10.0g     y_fin      What is your total shared annual household income
```

```
                                                 before taxes from all sources?
ethrace          double  %10.0g     a_ethrace
                                               Racial background
administrative   double  %12.0g                involved with child welfare
PR2              double  %11.0g     aggwsevphys38mfo
                                               sum of mother father and other weighted severe
                                                 phys discp 38
SR1              double  %16.0g     aggsumsevw12mf_oadol
                                               sum of mother father and other weighted severe
                                                 phys discp adolescent
PR1              float   %9.0g
outcome          float   %9.0g
-----------------------------------------------------------------------------------------------
Sorted by:
```

3 Basic Conceptual Model

We begin with a basic conceptual model of the reports and time points in the data, without at this point suggesting any associational or causal relationships.

4 Variable Abbreviations

For parsimony, we use the following conventions for variable names in equations and statistical syntax.

Table 1: Variables and Variable Labels

Variable	Label
`administrative`	administrative report
`PR1`	parental report in early childhood
`PR2`	parental report in middle childhood
`SR1`	adolescent self report
`SR2`	adult self report
`covariates`	covariates (multiple variables)
	(age, gender, ethnicity or race
	marital status, annual income)
`outcome`	mental health outcome

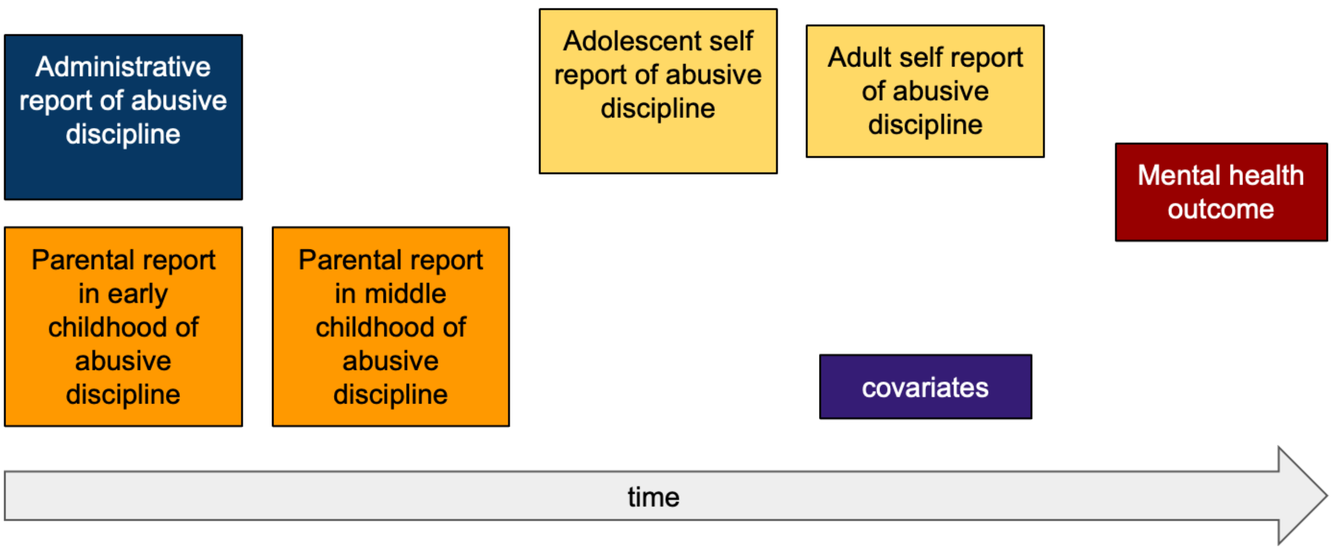

Figure 1: conceptual model

5 Methods

5.1 OLS Regression

Our outcome is continuous. Therefore we here employ *ordinary least squares regression*. Were our outcome to be dichotomous, we could as easily employ *logistic regression*.

5.1.1 Diagram

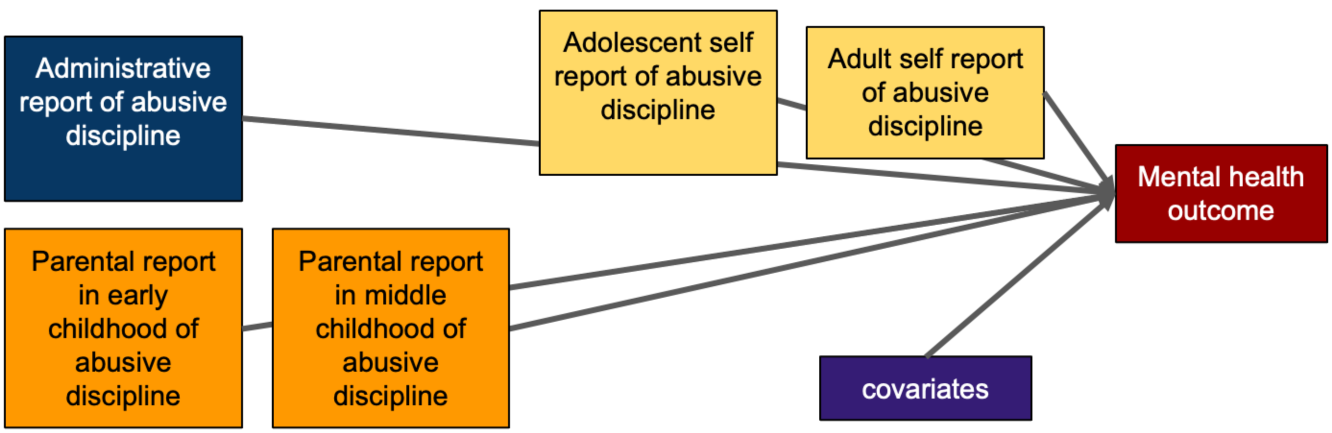

Figure 2: OLS

5.1.2 Equation

$$\text{outcome} = \beta_0 + \beta\text{P1} + \beta\text{P2} + \beta\text{SR1} + \beta\text{SR2} + \beta\text{administrative} + \Sigma\beta\text{covariates} + e_i \tag{1}$$

5.1.3 Syntax

```
regress outcome PR1 PR2 SR1 SR2 administrative age
```

```
Running /Users/agrogan/Desktop/GitHub/research/CAN-special-issue/profile.do

> ...

      Source |       SS           df       MS      Number of obs   =       211
-------------+----------------------------------   F(6, 204)       =      4.72
       Model |   908.37924         6  151.39654   Prob > F        =    0.0002
    Residual |  6547.28901       204  32.094554   R-squared       =    0.1218
-------------+----------------------------------   Adj R-squared   =    0.0960
       Total |  7455.66825       210  35.5031821   Root MSE        =    5.6652

--------------------------------------------------------------------------------
       outcome | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
---------------+----------------------------------------------------------------
           PR1 |   .0050038   .0184412     0.27   0.786     -.031356    .0413636
           PR2 |  -.0276371    .039927    -0.69   0.490    -.1063597    .0510854
           SR1 |   .0028035   .0173291     0.16   0.872    -.0313637    .0369707
           SR2 |   3.810465   1.042226     3.66   0.000     1.755548    5.865381
administrative |   1.493418   .8286179     1.80   0.073    -.1403357    3.127171
           age |   .3125531   .2161928     1.45   0.150    -.1137058    .7388119
         _cons |  -10.42348   9.956836    -1.05   0.296    -30.05498     9.20803
--------------------------------------------------------------------------------
```

For logistic regression, the appropriate syntax would be:

```
logit outcome P1 P2 SR1 SR2 administrative covariates, or
```

5.2 Summing Across Reporters

5.2.1 Diagram

5.2.2 Equation

First, we average parental reports:

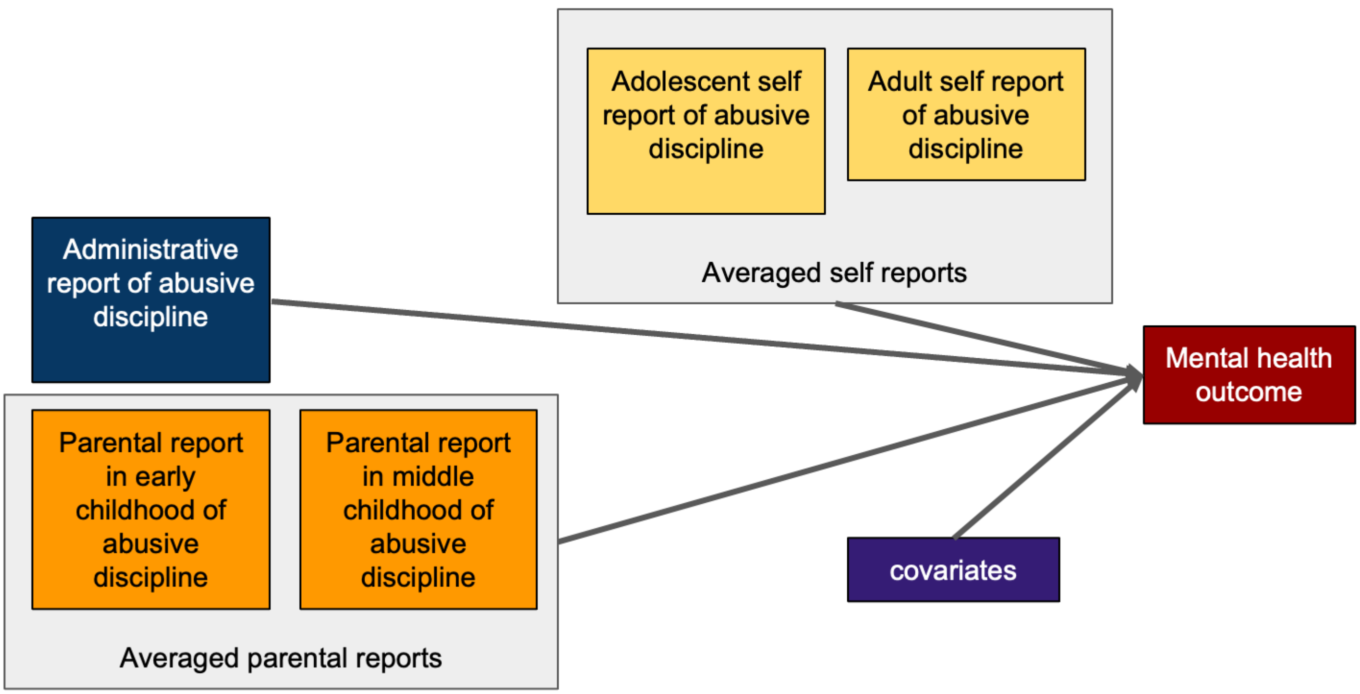

Figure 3: summing across reporters

$$P = \frac{P1 + P2}{2} \tag{2}$$

Then, we average self reports:

$$SR = \frac{SR1 + SR2}{2} \tag{3}$$

Lastly, we estimate an OLS model in which averaged parental and self reports are variables in the model.

$$\text{outcome} = \beta_0 + \beta \text{P} + \beta \text{SR} + \beta \text{administrative} + \Sigma \beta \text{covariates} + e_i \tag{4}$$

5.2.3 Syntax

```
generate P = (P1 + P2) / 2 // is averaging appropriate?

generate SR = (SR1 + SR2) / 2 // is averaging appropriate?

regress outcome P SR administrative covariates
```

5.3 Path Model

5.3.1 Diagram

5.3.2 Equation

$$\text{outcome} = \beta_0 + \beta \text{P1} + \beta \text{P2} + \beta \text{SR1} + \beta \text{SR2} + \beta \text{administrative} + \Sigma \beta \text{covariates} + e_i \tag{5}$$

$$\text{SR2} = \beta_0 + \beta \text{SR1} + e_i$$
$$\text{P2} = \beta_0 + \beta \text{P1} + e_i$$

5.3.3 Syntax

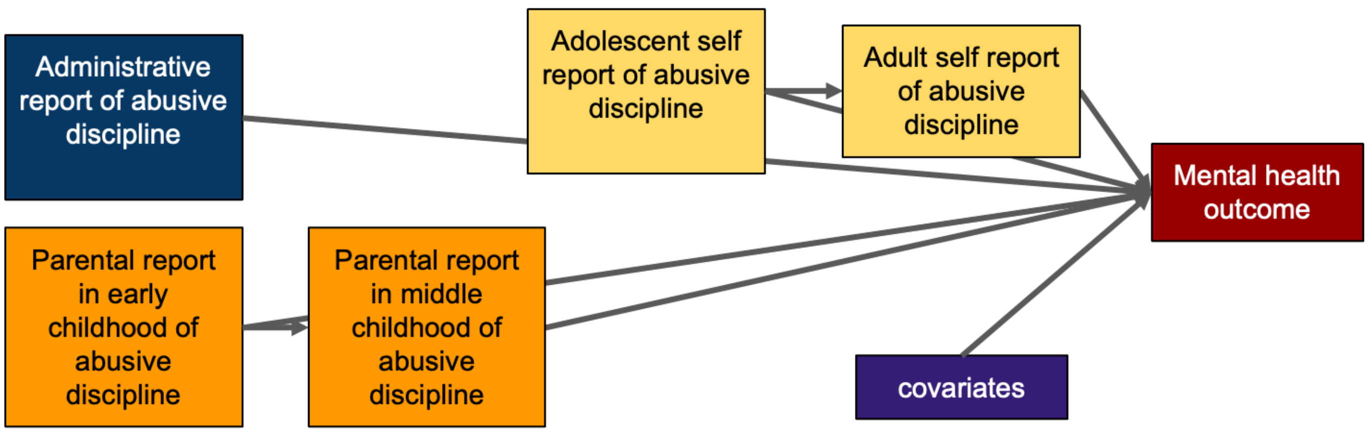

Figure 4: path model

```
sem (outcome <- covariates SR1 SR2 PR1 PR2 administrative) ///
  (SR2 <- SR1) ///
  (PR2 <- PR1) ///
  cov(e.outcome*e.SR2*e.PR2) // correlated errors
```

5.4 Latent Construct(s)

5.4.1 Diagram

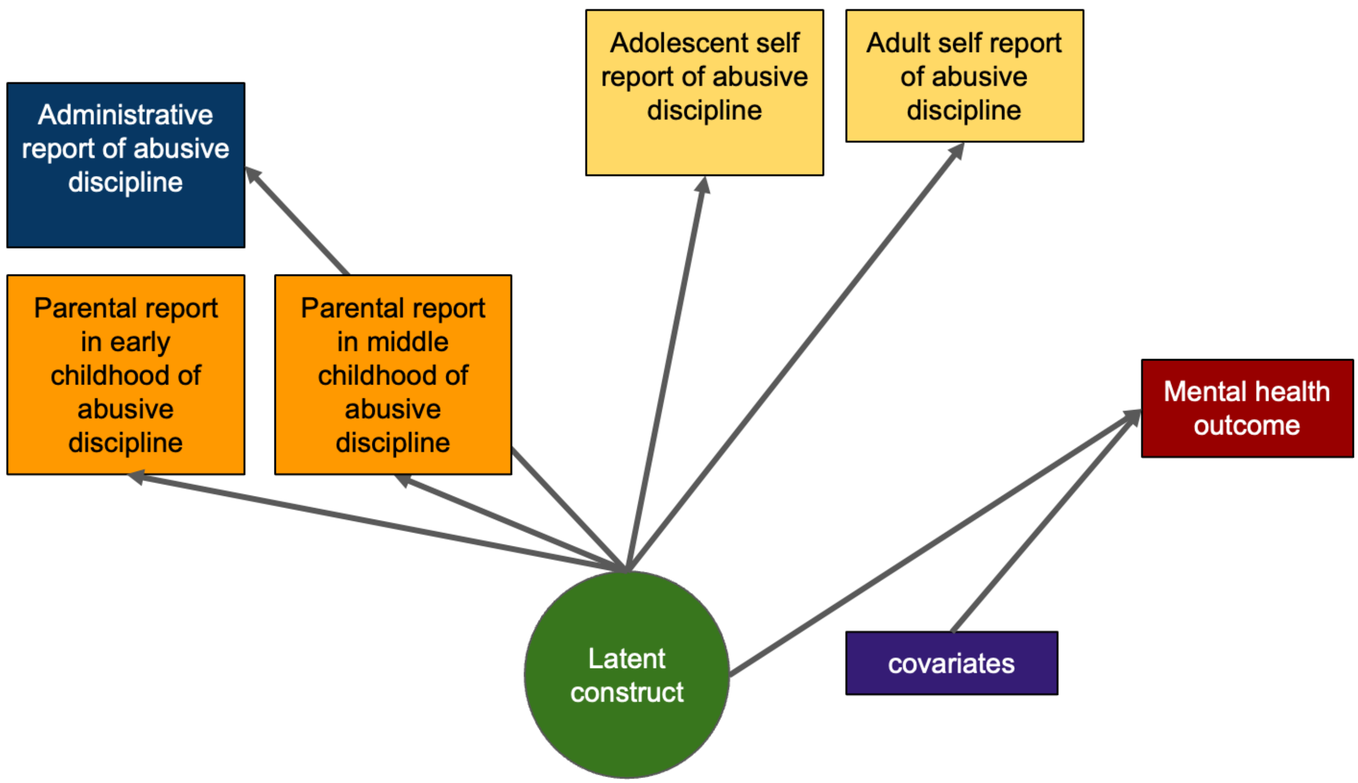

Figure 5: latent construct

5.4.2 Equation

5.4.3 Syntax

```
sem ///
  (P1 P2 SR1 SR2 administrative <- X) /// measurement
  (outcome <- covariates X) // structural
```

5.5 Latent Profile Analysis (Person Centered Approach)

5.5.1 Diagram

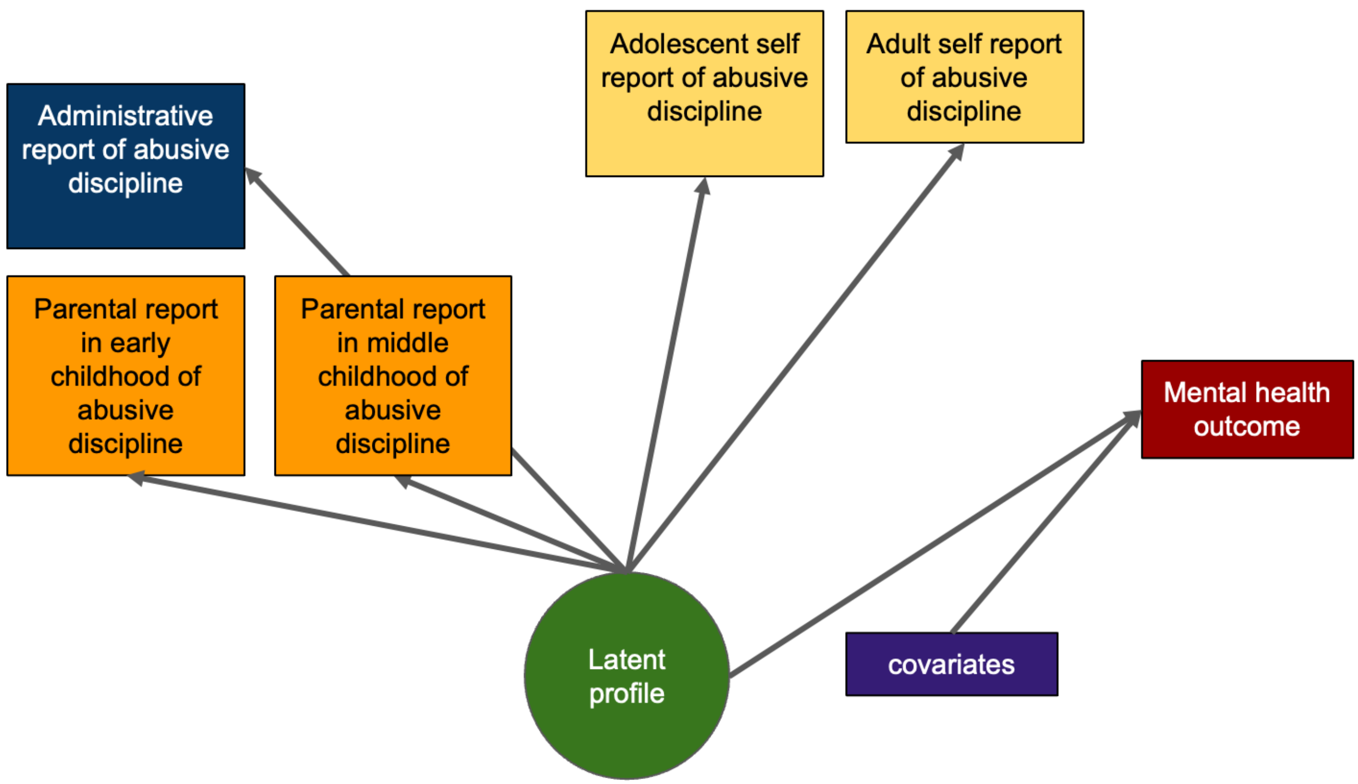

Figure 6: latent profile

5.5.2 Equation

5.5.3 Syntax

We first run a latent class analysis to generate latent underlying classes based upon the reports of discipline from the different reporterers

In the syntax below, we estimate three latent classes. The actual number of latent classes is determined by running models with different numbers of latent classes, and comparing those models using *fit statistics*, and *likelihood ratio tests*.

```
gsem (P1 P2 SR1 SR2 administrative <-, gaussian), (lclass(C 3))
```

We then use class membership to predict the outcome.

```
regress outcome i.class covariates
```

5.6 Network Analysis

5.6.1 Diagram

5.6.2 Equation

$$\begin{array}{c|cccccc} & \text{P1} & \text{P2} & \text{SR1} & \text{SR2} & \text{administrative} & \text{outcome} \\ \text{P1} & 1 & r_{\text{P1, P2}} & r_{\text{P1, SR1}} & r_{\text{P1, SR2}} & r_{\text{P1, administrative}} & r_{\text{P1, outcome}} \\ \text{P2} & & 1 & r_{\text{P2, SR1}} & r_{\text{P2, SR2}} & r_{\text{P2, administrative}} & r_{\text{P2, outcome}} \\ \text{SR1} & & & 1 & r_{\text{SR1, SR2}} & r_{\text{SR1, administrative}} & r_{\text{SR1, outcome}} \\ \text{SR2} & & & & 1 & r_{\text{SR2, administrative}} & r_{\text{SR2, outcome}} \\ \text{administrative} & & & & & 1 & r_{\text{administrative, outcome}} \\ \text{outcome} & & & & & & 1 \end{array} \tag{6}$$

5.6.3 Syntax

```
corr P1 P2 SR1 SR2 administrative outcome
```

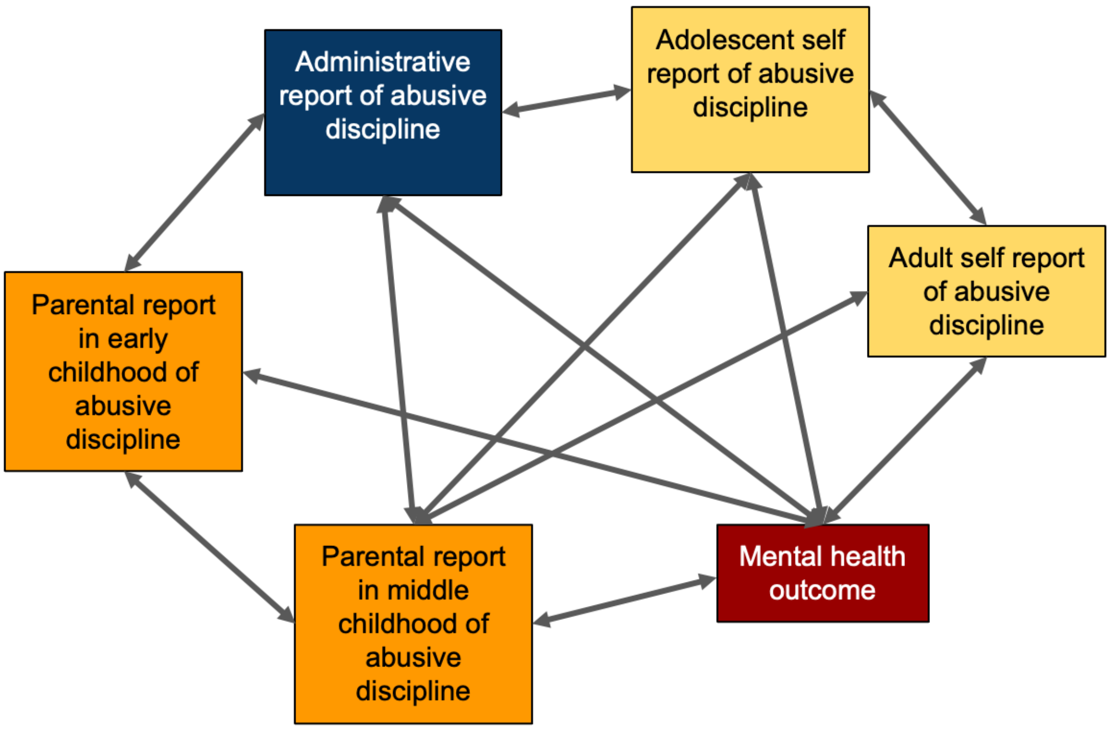

Figure 7: network model

5.7 ~~Item Response Theory (IRT)~~

5.8 ~~Multilevel Modeling~~

5.9 ~~Classification and Regression Tree (CART) (Machine Learning)~~

5.10 ~~Random Forest (Machine Learning)~~

6 Notes

Note 1

References

Hemken, D. (2023). *Statamarkdown: 'Stata' markdown*. https://CRAN.R-project.org/package=Statamarkdown

Machlin, L., Sheridan, M. A., Tsai, A. P.-T., & McLaughlin, K. A. (2025). Research review: Assessment of early-life adversity and trauma – cumulative risk and dimensional approaches. *Journal of Child Psychology and Psychiatry*, *n/a*. https://doi.org/https://doi.org/10.1111/jcpp.14170

McLaughlin, K. A., Sheridan, M., Humphreys, K. L., Belsky, J., & Ellis, B. J. (2020). *The value of dimensional models of early experience: Thinking clearly about concepts and categories*. PsyArXiv. https://doi.org/10.31234/osf.io/29fmt