Comparison of five software solutions to mediation analysis

Liis Starkopf

Mikkel Porsborg Andersen

Thomas Alexander Gerds

Christian Torp-Pedersen

Theis Lange

Research Report 17/01

Department of Biostatistics
University of Copenhagen

Comparison of five software solutions to mediation analysis

Liis Starkopf, Mikkel Porsborg Andersen, Thomas Alexander Gerds, Christian Torp-Pedersen, Theis Lange

July 10, 2017

ABBREVIATIONS:

CI	Confidence interval
GPA	Grade point average
IORW	Inverse odds ratio weighting

ABSTRACT

In epidemiology and many other scientific disciplines, mediation analysis is an important tool for understanding causal mechanisms. Specifically, mediation analysis allows to disentangle the indirect effect of an exposure on outcome through a given intermediate variable, the mediator. Recent developments in causal inference have greatly extended the theoretical framework and have led to a number of distinct estimation strategies of direct and indirect effects. Still, applied researchers are faced with challenges when implementing and interpreting mediation analysis. This paper provides practical advice on how to implement mediation analysis. We compare and contrast five estimation approaches for mediation analysis including their software implementations in R, SAS, SPSS and STATA.

Keywords: causal inference, mediation analysis, software implementation

Mediation analysis is used in epidemiology, social sciences and other scientific disciplines as a tool for analysing the causal mechanisms in complex observational settings. The idea is to break down the total causal effect of an exposure on an outcome into a direct effect and an indirect effect. For illustration, throughout this paper we consider an example were interest is in understanding whether better physical fitness of students leads to increased commencement in post-compulsory education and whether effect of physical fitness is mediated through academic achievement.

Causal inference theory has led to a mathematical framework in which one can define so-called natural direct and indirect effects (1–4). In the physical exercise example, the natural direct effect describes the change in commencement in post-compulsory education that we would observe if an intervention would change the level of physical fitness of all students from low to high while keeping the academic achievement unchanged at the level it would naturally take under low level of physical fitness.

A number of distinct estimation strategies with accompanying code examples or software implementations have been developed (3, 5–10). Valeri and VanderWeele (11, 12) provided SAS and SPSS macros to obtain closed-formed expressions of natural effects as a combination of regression parameters of mediator and outcome models. Imai et al. (3, 13) suggested a general approach using repeated Monte Carlo draws of counterfactual outcomes based on the user-specified mediator and outcome model. Natural effects are functions of these potential outcomes and can be easily computed, regardless of the statistical models used (13). The approach has been made available in the R package mediation (14). Lange et al. (7) and Vansteelandt et al. (8) proposed to fit so-called natural effect models which directly parameterize natural direct and indirect effects. Two different procedures based

on natural effect models have been implemented in the R package medflex (15): a weighting procedure proposed by Lange et al. (7) and imputation strategies suggested by Vansteelandt et al. (8). Tchetgen-Tchetgen (10) suggested an inverse odds ratio weighted estimator of natural effects, which uses the estimate of the inverse of the odds ratio function relating the mediator and the exposure as a weight to estimate the natural direct effect via a weighted regression analysis. Practical guidance for conducting mediation analysis using inverse odds ratio weighted estimation approach, including STATA code examples, has been provided by Nguyen et al. (16). In addition, R code examples for implementation of inverse odds ratio estimation approach can be found in Nguyen et al. (17).

In this paper we address researchers who want to apply mediation analysis. We compare and contrast the aformentioned five estimation approaches and their software solutions. The comparison is done from both a mathematical perspective and by means of an empirical study.

The paper is organised as follows. First, we recall the mathematical definition of direct and indirect effects and necessary conditions for identifiability. Second, we review and compare the considered methods in terms of parametrization, modelling demands and software. Next, we compare small sample performance of the methods. Further discussion and practical advice for mediation analysis is postponed to the last section. Finally, coding examples and further details are provided in the eAppendix.

THEORETICAL BACKGROUND

The essential mediation analysis problem is given by the causal structure (directed acyclic graph) shown in Figure 1. Let A be an exposure, M a mediator, Y an outcome

and C a set of baseline confounders. Let $\mathscr{A}$ be the support of A, i.e the set of all the values A can take. Similarly, let $\mathscr{M}$, $\mathscr{Y}$ and $\mathscr{C}$ denote the support of M, Y and C respectively.

[Figure 1 about here.]

We describe the direct and indirect effects in terms of nested counterfactuals, $Y(a, M(a^*))$, which denotes the outcome that would have been observed if, possibly contrary to the fact, A was set to a and M was set to the value it would have taken if A was set to a^*.

The total effect of the exposure on the outcome can be decomposed into the natural direct and indirect effects marginally

$$\underbrace{g\{E[Y(a,M(a))]\} - g\{E[Y(a^*,M(a^*))]\}}_{\text{marginal total effect}} = \underbrace{g\{E[Y(a,M(a^*))]\} - g\{E[Y(a^*,M(a^*))]\}}_{\text{marginal (pure) natural direct effect}} + \underbrace{g\{E[Y(a,M(a))]\} - g\{E[Y(a,M(a^*))]\}}_{\text{marginal (total) natural indirect effect}} \tag{1}$$

$$= \underbrace{g\{E[Y(a,M(a))]\} - g\{E[Y(a^*,M(a))]\}}_{\text{marginal (total) natural direct effect}} + \underbrace{g\{E[Y(a^*,M(a))]\} - g\{E[Y(a^*,M(a^*))]\}}_{\text{marginal (pure) natural indirect effect}} \tag{2}$$

where a and a^* denote two possible exposure values corresponding to exposed and unexposed and g is some link function. The natural direct effect might depend on the natural level at which the mediator is set to ($M(a^*)$ vs. $M(a)$) and the natural indirect effect might depend on the level at which the exposure is set to (a^* vs. a).

Robins and Greenland (1), Hafeman and Schwartz (18) proposed the terms pure and total natural (in)direct effect to distinguish between the two quantities. The marginal pure natural direct effect is the change in outcome (Y) we would observe if the exposure is changed from unexposed to exposed ($a^* \rightarrow a$), but the mediator is maintained as if it would be under no exposure ($M(a^*) \rightarrow M(a*)$). Analogously, the marginal total direct effect is the change in outcome we would observe if the exposure is changed from unexposed to exposed, but the mediator is maintained as if it would be under exposure ($a^* \rightarrow a, M(a) \rightarrow M(a)$). The marginal total natural indirect effect is then the change in outcome that would occur if only the mediator is changed from what it would be under no exposure to what it would be under exposure while the exposure is kept at the exposed level ($a^* \rightarrow a^*, M(a^*) \rightarrow M(a)$). And finally, the marginal pure indirect effect is the change in outcome that would occur if the mediator is changed from what it would be under no exposure to what it would be under exposure while the exposure is kept at the unexposed level ($a^* \rightarrow a^*, M(a^*) \rightarrow M(a)$). The link function ($g$) defines the scale on which the change in the outcome is observed. If the scale is linear no link function is necessary, if the scale is odds ratio the link function is the logit function. In this paper we will consider the first representation of natural effects in equation (1) if not stated otherwise.

In addition, the natural effects can be defined conditionally on a set of baseline

covariates C

$$\underbrace{g\{E[Y(a,M(a))\,|\,C=c]\}-g\{E[Y(a^*,M(a^*))\,|\,C=c]\}}_{\text{conditional total effect (c)}}=$$

$$\underbrace{g\{E[Y(a,M(a^*))\,|\,C=c]\}-g\{E[Y(a^*,M(a^*))\,|\,C=c]\}}_{\text{conditional natural (pure) direct effect (c)}}$$

$$+\underbrace{g\{E[Y(a,M(a))\,|\,C=c]\}-g\{E[Y(a,M(a^*))\,|\,C=c]\}}_{\text{conditional natural (total) indirect effect (c)}}, \tag{3}$$

The identification and estimation of natural effects from the observed data requires the following assumptions. In particular, we assume

Consistency:

$$P(M(a)=M\,|A=a)=1 \quad \text{for all } a\in\mathscr{A}, \tag{4}$$

$$P(Y(a,m)=Y\,|A=a,M=m)=1 \quad \text{for all } a\in\mathscr{A}, m\in\mathscr{M}, \tag{5}$$

No uncontrolled confounding:

between exposure and outcome given the baseline covariates

$$Y(a,m)\perp\!\!\!\perp A\,|\,C \quad \text{for all } a\in\mathscr{A}, m\in\mathscr{M}, \tag{6}$$

between exposure and mediator given the baseline covariates

$$M(a)\perp\!\!\!\perp A\,|\,C \quad \text{for all } a\in\mathscr{A}, \tag{7}$$

between outcome and mediator given the exposure and the baseline covariates

$$Y(a,m) \perp\!\!\!\perp M \,|\, A = a, C \quad \text{for all } a \in \mathscr{A}, \tag{8}$$

where $A \perp\!\!\!\perp B \,|\, C$ states that A is independent of B given C.

No intertwined causal pathways:

No confounders of the outcome and mediator relationship given the baseline covariates that are affected by the exposure

$$Y(a,m) \perp\!\!\!\perp M(a^*) \,|\, C \quad \text{for all } a, a^* \in \mathscr{A}, m \in \mathscr{M}, \tag{9}$$

Positivity:

$$f_M(m \,|\, A, C) > 0 \text{ with probability 1 for all } m \in \mathscr{M}, \tag{10}$$

$$f_A(a \,|\, C) > 0 \text{ with probability 1 for all } a \in \mathscr{A}, \tag{11}$$

where $f_M(\cdot \,|\, A, C)$, $f_A(\cdot \,|\, C)$ are the conditional densities of M given (A, C) and A given C, respectively. In case A and/or M is not continuous, the corresponding densities in (10, 11) are replaced by probabilities.

The first three no unmeasured confounding assumptions in equations (6-8) can not be tested, but they could be guaranteed in certain designs (4). For example, assumptions in equations (6,7) are satisfied if the exposure is randomized. In contrast, assumption in equation (9) can never be guaranteed, even in an experimental design (19).

METHODS AND SOFTWARE SOLUTIONS

Parameter of interest

All estimation approaches target the natural effects. We distinguish marginal natural effects (1) from conditional natural effects (3). Only the two approaches implemented in R package medflex allow us to estimate both, marginal and conditional parameters. The R package mediation allows only estimation of marginal natural effects. The inverse odds ratio weighted estimation approach and the approach implemented in SAS and SPSS macros allow only estimation of the conditional natural effects. However, the conditional parameters targeted in the SAS and SPSS macros differ slightly from the conditional effects targeted in medflex package and by inverse odds ratio weighted estimation approach. The inverse odds ratio weighted estimation approach and approaches in the medflex package assume that the conditional natural effects are the same for any level of baseline confounders (unless in presence of exposure-confounder interactions) where as in some cases the conditional parameter estimates produced in SAS and SPSS mediation macros vary for different values of confounders even if the exposure-confounder interaction is not present (details in the Web Appendices (B.2, D.2, E.2, F.2)). By default, the SAS and SPSS macros compute the natural effects within the mean of the observed levels of the confounders.

Another distinction between the estimation approaches can be made in terms of the link function used for natural effects. Only the identtity link leading to natural effects estimates expressed as differences in expected outcome has been implemented in the R mediation package. In contrast, the approaches behind the medflex package and the SAS and SPSS macros allow to express the natural effects on a scale

corresponding to the link function in the natural effect model or outcome model respectively. For example, in presence of a binary outcome, natural effects can be expressed as odds ratios when using the medflex package where as the mediation package can provide estimates of the natural effects only as risk differences. The list of link functions implemented in R package medflex and SAS and SPSS macros is given in the Web Appendices (D.2, E.2) and (B.2) respectively.

Modelling

No matter the target parameters, all five methods require specification of two regression models (Table 1). The estimation methods implemented in the SAS and SPSS mediation macros and R mediation package are based on regression models for the mediator M given (A, C) and the outcome Y given (A, M, C). Since the analytical expressions derived from the mediation formula depend on the type of variables and models under consideration, only a limited number of scenarios have been implemented in the SAS and SPSS macros. The mediation package allows more flexible choice of the models. Furthermore, the SAS and SPSS macros allow an interaction term only between A and M and only in the outcome model where as the mediation package can accomodate all sorts of interactions in the outcome as well as the mediator model. Both approaches in the medflex package require specification of a natural effects model for the nested counterfactuals $Y(a, M(a^*))$ given C and an additional nuisance model. A model for the mediator M given (A, C) is employed as the nuisance model for weighting approach where as imputation approach relies on the regression model for the outcome Y given (A, M, C). Interactions are allowed for both, the natural effects model and the nuisance model. The inverse odds ratio weighted estimation approach requires specification of a model for the exposure A

given (M, C) and for the outcome Y given (A, C) whereby all sorts of interaction terms can be included in both models. The inverse odds ratio weighted estimation approach and approaches in the medflex package can, in principle, be used for any type of exposure, mediator and outcome. The medflex package can currently handle only a limited number of scenarios (Table 1). The inverse odds ratio weighted estimation can be implemented in any software that accomodates weighted regression analysis. However, the inverse odds ratio estimation approach requires users to compute the estimates for the weights and to carry out the weighted regression manually.

Sensitivity analysis

All the methods require rather strong conditions given in equations (4-11) for identification of natural effects. Thus, it is recommended to supplement mediation analyses with sensitivity analyses for possible violation of these assumption. The mediation package allows to conduct a sensitivity analysis to quantify the degree of potential unobserved confounding. However, the sensitivity analysis is limited to certain types of outcome and mediator models (Web Appendix C.1). The medflex package, the SPSS and SAS macros and code examples of inverse odds ratio estimation approach do not include any specific tools for sensitivity analysis. Therefore, conducting sensitivity analysis requires further coding from the user.

Standard errors

The summary output of the medflex and mediation packages provides standard errors obtained by either a bootstrap procedure or robust estimation. The default option is robust estimation in mediation package and bootstrap procedure in medflex.

The summary output of the SAS and SPSS macros displays the confidence intervals based on either robust standard errors or a bootstrap procedure. The code examples of inverse odds ratio estimation approach include a bootstrap procedure to obtain the confidence intervals. The formula for the robust estimator of the variance-covariance matrix is provided in (10). Implementation examples of robust standard errors are currently unavailable.

[Table 1 about here.]

DANISH EDUCATION STUDY

In this section, we show the results obtained with the estimation methods described in previous section when applied to a study of Danish school youth (20). The aim of the study was to examine how much of the effect of physical fitness on commencement in post-compulsory education is mediated through academic achievement. In Denmark last year of compulsary schooling is 9th or 10th grade, which correposnds to 10 or 11 years of schooling respectively. In the study all 8th grade students in the year 2010 from all public elementary schools with 9th grade in the Danish municipality of Aalborg were invited to participate in health examinations. Data from 1,084 students were obtained. The phyisical fitness level was measured as the relative maximal oxygen consumption during a watt-max test on a cycle ergometer which was conducted during the health examination. For the analysis, the students are divided into two groups, low ($A = 0$) or high ($A = 1$) level of physical fitness, based on gender-specific averages of physical fitness. The continuous mediator M of academic achievement was measured as a grade point average from the compulsory exams at the end of last year of compulsory education (min. 0.4, max. 12, mean 7.2,

sd. 2.2). Commencement in post-compulsory education before 2014 is considered as the outcome Y, whereby not commencing in post-compulsory education ($Y = 1$) is considered as the event of interest. The data also includes age, ethnicity, parental income and education level as baseline confounders C (Figure 2). Here we only provide a small part of the analysis for illustration, full analysis and further details about the study can be found at (20).

[Figure 2 about here.]

Since the estimation method behind SAS and SPSS mediation macros is the same, we only included the first. For the same reason, we considered the implementation of inverse odds ratio estimation approach in R and discarded the implementation in STATA. We used logistic regression to model the outcome for R mediation package, inverse odds ratio estimation approach and as the nuisance model for imputation approach in R package medflex. Since the event of interest is a rare event ($P(Y = 1) = 0.08$), we were able to use a logistic regression for the outcome also for SAS macro. Logistic regression model was also used as the natural effect model for both methods in medflex and as the exposure model for the inverse odds ratio estimation approach. A linear regression was used to relate academic achievement to physical fitness for R mediation package, SAS macro and weighting approach in medflex. Only main effects were entered in the models, interaction terms were excluded. Using these models, we estimated the natural effects as conditional odds ratios for the R package medflex, SAS macro and the inverse odds ratio estimation approach and as risk differences for the mediation package. Only the results for natural direct and indirect and total effects are reported, the controlled direct effect and the porportion mediated which are not available for all the software solutions

were not considered. The results are displayed in Table 2.

[Table 2 about here.]

The weighting method in medflex package suggests that the natural direct and indirect effect of the level of physical fitness on the odds of not commencing in post-compulsory education amount to odds ratios 0.515 (95% CI: 0.287, 0.856) and 0.749 (95% CI: 0.649, 0.846), respectively (Table 2). In particular, if the physical fitness level of all the students was changed from low to high without changing their grade point average, the odds of not commencing in post-compulsory education would be decreased $1/0.515 = 1.94$ times. Furthermore, if all the students had high level of physical fitness, then changing their grade point average to what it would be under low level of physical fitness, would decrease the odds of not commencing in post-compulsory education $1/0.749 = 1.34$ times. The corresponding odds ratio estimates for imputation method in medflex, SAS macro and inverse odds ratio estimation approach were similar (Table 2). In contrast, the estimates of the natural direct and indirect effect of the physical fitness on the risk of not commencing in post-compulsory education obtained with the R package mediation result in risk differences of -0.048 (95% CI: -0.084, -0.009) and -0.014 (95% CI: -0.024, -0.007), respectively. That is, if the physical fitness level of all students was changed from low to high without changing their grade point average, the risk of not commencing in post-compulsory education would be decreased by 0.048 and if all the students had high level of physical fitness, then changing their grade point average to what it would be under low level of physical fitness, would decrease the risk of not commencing in post-compulsory education by 0.014 (Table 2).

EMPIRICAL COMPARISON

To investigate the small sample performance of the considered estimation approaches, we performed a simulation study with 2000 runs of data sets with 200 observations. Motivated by the Danish education study, we focus on a setting with a binary outcome and continuous mediator. We consider two examples, one with a rare and one with a common outcome. Details for simulation setup can be found in Web Appendix A. To make the simulation setting close to real life, we have chosen a simulation setup, which will make all the models employed by the considered software packages slightly misspecified. This reflects reality where you can never expect your models to fit perfect and levels the playing field in the comparison as none of the methods fit perfect. The true parameter values for each causal parameter were calculated by simulating a data set with 100,000 observations and subsequently applyng each of the methods to this large data set. The procedure was repeated to confirm that the obtained estimates had in fact converged to their true values. Recall, that not all the approaches estimate the same parameters as they differ in how they handle confounders and for the R mediation package use a different scale (Table 1). To make the simulation results more comparable, we report the bias and root mean squared error (RMSE) across 2,000 simulations relative to the true parameter values estimated from a large simulated data set. The coverage probability of the 95% confidence intervals obtained with nonparametric bootstrap (1,000 bootstrap draws for each of the 2,000 simulated data set) are also reported.

Common binary outcome

In the first example, the simulation set-up results in a common binary outcome. For estimation, we have used logistic regression model as the outcome model for R mediation package, inverse odds ratio estimation approach and as the nuisance model for the imputation approach in the mdeflex package. Logistic regression model was also fitted as the exposure model for the inverse odds ratio weighted approach and as the natural effects model for both methods in medflex package. Since the outcome was common, log-linear model was used as an outcome model for SAS macro. For all methods requiring mediator model, linear regression model was fitted. All the models employed included only main effects and no interaction terms.

[Table 3 about here.]

The results from applying SAS macro suggested a large bias of the estimates we obtained from small samples with respect to the true parameter values obtained from large samples. In presence of common binary outcome, the analytic formulas implemented in SAS macro only hold for log-linear outcome models. The bias we saw might possibly be explained by a well-known issue with the non-convergence of log-linear models (21, 22). To be as fair as possible to the SAS macro, results in Table 3 are based on true values estimated from the large simulated data set by applying the imputation approach in medflex with a log-linear natural effects model instead of SAS macro. Overall, all methods besides SAS macro performed similarly. The presicion of the inverse odds ratio estimation approach for indirect effect was found slightly greater than that of the other methods (Table 3).

Rare binary outcome

In the second example, the simulation set-up results in a rare outcome. For estimation, the same models as for the common outcome were used for R mediation, both methods in medflex and inverse odds ratio estimation approach. For the SAS macro, the log-linear model was replaced by a logistic regression model for the outcome thereby making converging issues substantially smaller.

[Table 4 about here.]

For indirect effect, all the methods except inverse odds ratio estimation approach had fairly similar performance where as the inverse odds ratio estimation approach suffered a loss in precision and relative bias compared to the other methods (Table 4). For direct and total effect, the relative bias of both methods in R package medflex and inverse odds ratio estimation approach was found to be slightly greater than that of the other two methods (Table 4). However, this finding disappeared when the sample size was increased (Figure2).

[Figure 3 about here.]

DISCUSSION

In this paper we have reviewed five different estimation methods for mediation analysis. Despite of the existing well-developed theoretical mediation analysis techniques, applied researchers are still faced with challenges when implementing mediation analysis. We have showed that mediation analysis can be applied fairly easy in most of the standard software packages. The findings in our simulation study show

that all the estimation methods perform well in small samples. Therefore, choosing estimation method in practice can be partly based on the software preference.

When choosing the software, researchers should keep in mind what causal parameters are of interest. All methods described in this paper target natural direct and indirect effects. However, some methods consider marginal natural effects where as others estimate the natural effects conditionally on baseline confounders. No matter the target parameter the interpretation of a mediation analysis is often challenging. The interpretation of natural effects relies on the potential outcomes with the mediator set to a certain value. Recently Lok (23) introduced a generalization of natural effects, so-called organic effects. The interpretation of organic effects is based on organic interventions that cause the mediator to have a particular distribution rather than setting the mediator to a certain value. For instance, the measure of organic indirect effect is obtained by comparing the outcome under exposure to the outcome under exposure and organic intervention which causes the mediator to have the same distribution as under no exposure. Therefore, the interpretation holds more broadly for organic effects in comparison to natural effects. However, in any given application organic effects and natural effects are equal, only the interpretation changes.

Valid estimation for all the methods relies on the correct specification of the required models. Accordingly, differences in modelling demands can be valuable when choosing one estimation method over another, especially, when the causal parameters of interest can be estimated with many different estimation approaches (e.g. approaches in the medflex package and inverse odds ratio estimation approach). The weighting approach for natural effects models requires specification of the conditional density of the mediator. When dealing with continuous mediators, in addition

to the correct specification of the conditional expectation of the mediator, parametric assumptions about the error terms in the mediator model are therefore necessary (7, 15). Modelling the conditional density is also challenging if the mediator is multivariate (7, 15). The imputation approach of natural effects models avoids this problem by using a model for the outcome (8). However, correct specification of the outcome model might be more challenging, for example when dealing with survival outcomes. In contrast, inverse odds ratio estimation approach requires specification of regression models for the exposure and the outcome. While modelling the exposure is attractive in simple settings with binary exposure, when dealing with continuous exposure, full specification of conditional density of the exposure is needed. In the light of the foregoing discussion, the imputation approach is more attractive when the mediator is continuous or multivariate where as the weighting approach is more suitable when imputation model is complex and/or misspecification of imputation model is difficult to detect. In more complex settings where modelling the outcome and mediator is difficult, but the exposure follows a simple distribution, e.g. multivariate mediator, survival outcome and binary exposure, inverse odds ratio estimation approach might possibly be preferred over the medflex approaches. Finally, when the exposure is continuous, inverse odds ratio estimation approach and weighting method might produce unstable weights, making the imputation approach a better choice.

In summary, we have reviewed five estimation approaches for mediation analysis. We have provided a comparison of the methods, including modelling demands as well as the scope and limitations of the data types and statistical models that can be handled by the software solutions. Finally, we have provided extensive coding advice, examples and technical details.

A Simulation setup

For both examples, a binary confounder C was drawn from the binomial distribution with $P(C = 1) = 0.7$ and the binary exposure A was drawn from a binomial distribution with

$$P(A = 1 \mid C = c) = \Phi(0.2 - 0.3c),$$

where $\Phi(\cdot)$ refers to the cumulative standard normal distribution. The continuous mediator M was simulated under the following strutural equation model

$$M = 6.7 + 0.4A - 0.7C + \epsilon,$$

with the error term ϵ drawn from the t-distribution with degrees of freedom equal to 10. The binary outcome Y was drawn from the binomial distribution with

$$P(Y = 1 \mid A = a, M = m, C = c) = \Phi(\theta_0 + 0.4a + 0.2m + 0.2c)$$

with $\theta_0 = -1$ for the first example and $\theta_0 = -0.3$ for the second example.

B SAS and SPSS macros for mediation analysis

B.1 Summary

Analytic expressions of natural direct and indirect effects can be derived as a combination of regression parameters from regression models for the outcome and mediator. Valeri and VanderWeele (11, 12) provided SAS and SPSS macros to obtain these closed-formed expressions. The resulting formulas depend on the type of outcome and mediator variables and the choice of the regression models. Therefore, only a limited number of scenarios has been implemented in the macros. In particular, linear, logistic, log-linear, poisson or negative binomial regression, Cox proportional hazards or accelerated failure time models can be specified as outcome models. For the mediator, either linear or logistic regressions are allowed. Table 1 provides a summary of the available scenarios.

[Table 5 about here.]

The SAS and SPSS macros are only applicable for one mediator, but they enable to study exposure-mediator interactions. The situations with exposure-covariates or mediator-covariates interactions have not been implemented in the SAS and SPSS macros. At present, the macros do not include any specific tools for conducting sensitivity analysis. Thus, conducting sensitivity analysis requires further coding from the user. Finally, it is possible to implement mediation analysis in the SAS and SPSS macros when data arises from a case-control design, provided that the outcome in the population is rare. Further details can be found in (11, 12). Since the SPSS macro performs exactly the same tasks as the SAS macro (with some minor differences), we will further on focus only on the SAS macro. Details about SPSS macro can be found in (11, 12).

B.2 Parameter of interest

The SAS macro provides estimates of the conditional natural effects at a given level c of the covariates C. In particular, for two values of the exposure, a, a^*, and any value c of the covariates C, the conditional natural direct effect $NDE(a, a^*, c)$, natural indirect effect $NIE(a, a^*, c)$ and total effect $TE(a, a^*, c)$ are defined as

$$NDE(a,a^*,c) = g\{E[Y(a,M(a^*))\,|\,C=c]\} - g\{E[Y(a^*,M(a^*))\,|\,C=c]\} \quad \text{(B.1)}$$

$$NIE(a,a^*,c) = g\{E[Y(a,M(a))\,|\,C=c]\} - g\{E[Y(a,M(a^*))\,|\,C=c]\} \quad \text{(B.2)}$$

$$TE(a,a^*,c) = g\{E[Y(a,M(a))\,|\,C=c]\} - g\{E[Y(a^*,M(a^*))\,|\,C=c]\} \quad \text{(B.3)}$$

for some link function g. For instance, the measure of conditional natural direct effect for a binary outcome can be expressed as an odds ratio

$$\begin{aligned}
NDE(a,a^*,c) &= \text{logit}\{E[Y(a,M(a^*))\,|\,C=c]\} - \text{logit}\{E[Y(a^*,M(a^*))\,|\,C=c]\} \\
&= \text{logit}\{P(Y(a,M(a^*))=1\,|\,C=c)\} - \text{logit}\{P(Y(a^*,M(a^*))=1\,|\,C=c)\} \\
&= \log\left\{\frac{P(Y(a,M(a^*))=1\,|\,C=c)}{1-P(Y(a,M(a^*))=1\,|\,C=c)}\right\} - \\
&\quad \log\left\{\frac{P(Y(a^*,M(a^*))=1\,|\,C=c)}{1-P(Y(a^*,M(a^*))=1\,|\,C=c)}\right\} \\
&= \log\left\{\frac{odds\big(P(Y(a,M(a^*))=1\,|\,C=c)\big)}{odds\big(P(Y(a^*,M(a^*))=1\,|\,C=c)\big)}\right\} \\
&= \log\{OR_{NDE}(a,a^*,c)\}
\end{aligned}$$

by using $g = \text{logit}$ link function. In addition to the conditional natural effects, SAS macro provides an estimate of controlled direct effect $CDE(a,a^*,m,c)$ at a fixed level m of the mediator M conditional on a given level c of the covariates C given by

$$CDE(a,a^*,m,c) = g\{E[Y(a,m)\,|\,C=c]\} - g\{E[Y(a^*,m)\,|\,C=c]\}, \qquad \text{(B.4)}$$

where a, a^* are two values of exposure. In some cases, the analytic formulas for the causal parameters given in equations (B.1-B.4) can vary for different values c of C even in absence of exposure-confounder interactions. Therefore the SAS macro computes the causal parameters in equations (B.1-B.4) given the user-specified value of c. By default, the mean of the observed levels of the covariates C is chosen. Extensions of the causal effects given in equations (B.1-B.4) for failure time outcome can be found in (12).

B.3 Example

In this section, we illustrate how to use SAS mediation macro for mediation analysis. We will do this by applying the SAS mediation macro to the data set from Danish education study (20). The study investigates how much of the effect of physical fitness (A) on commencement in post-compulsory education (Y) is mediated through academic achievement measured by grade point average (M).

B.3.1 The data set

The data set consists of 1084 observations and the following variables.

`fitness` Physical fitness coded with 2 levels (0-low, 1-high).

`gpa`	Grade point average of the compulsory exams at the end of compulsory education.
`postedu`	Commencement in post-compulsory education (0-commenced, 1-not commenced).
`age`	Age in years (13, 14, 15).
`ethni`	Ethnicity (0-immigrants/descendants, 1-danes).
`income`	Parental income in 4 groups (0-3) with 3 being the highest income group.
`educ`	Parental education level in 4 groups (0-3) with 3 being the highest.
`age14`	Age 14
`age15`	Age 15
`income1`	Parental income, group 1
`income2`	Parental income, group 2
`income3`	Parental income, group 3
`educ1`	Parental education, group 1
`educ2`	Parental education, group 2
`educ3`	Parental education, group 3

It is assumed that age, ethnicity, parental income and education level are sufficient to control for confounding. For more details about the data, see (20). Since SAS macro only allows baseline covariates that are coded as series of indicator variables, we have used the corresponding dummy variables for the variables age, parental income and education in the analysis.

B.3.2 Model specification

The estimation procedure requires specification of two models: one for the outcome and one for the mediator. When the outcome is binary and logistic regression is used, the formulas hold only if the outcome is rare. When the outcome is common, log-linear model has to be used instead. Since no commencement in post-compulsory education is a rare event ($P(Y=1)=0.08$), we will fit the following logistic regression model for the outcome

$$\text{logit}\{P(Y=1\,|A=a,M=m,C=c)\}=\theta_0+\theta_1 a+\theta_2 m+\theta_3^T c.$$

In addition, we use the following linear regression for the mediator

$$E[M \mid A = a, C = c] = \beta_0 + \beta_1 a + \beta_2^T c.$$

Using these models, we obtain the following (approximate) formulas for natural effects

$$\begin{aligned} \log\left\{OR_{NDE}(1,0,c)\right\} &\cong \theta_1 \\ \log\left\{OR_{NIE}(1,0,c)\right\} &\cong \theta_2\beta_1 \\ \log\left\{OR_{TE}(1,0,c)\right\} &\cong \theta_1 + \theta_2\beta_1. \end{aligned}$$

In addition, when there is no exposure-mediator interaction, controlled direct effect and natural direct effect coincide

$$\log\left\{OR_{CDE}(1,0,m,c)\right\} = \log\left\{OR_{NDE}(1,0,c)\right\}$$

for every possible mediator value m. More details about the derivation of these formulas can be found in (11).

B.3.3 Implementation

Step 1. Before conducting mediation analysis, the user must provide the data set and the mediation macro.

```
%INC "MEDIATION.sas";
PROC IMPORT DATAFILE="d1.csv" OUT=d1 DBMS=csv;
RUN;
```

Step 2. After saving the data and inserting the mediation macro we run the following statement.

```
%MEDIATION(data=d1,yvar=postedu,avar=fitness,mvar=gpa,cvar=ethni
    age14 age15 income1 income2 income3 educ1 educ2 educ3,a0=0,a1
    =1,m=0,nc=9,c=,yreg=logistic,mreg=linear,interaction=false,
    output=,boot=,casecontrol=)
RUN;
```

Here we have specified the name of the data set (`data=`) and the name of the outcome (`yvar=`), exposure (`avar=`), mediator (`mvar=`) and the baseline covariates (`cvar=`). In addition to the name of the baseline covariates, the number (`nc=`) of covariates has to be provided. The mediation macro also requires choosing the two levels of exposure, a (`a1=`) and a^* (`a0=`) to compare and the level of baseline covariates c (`c=`) at which the natural and controlled direct effect is to be estimated as well as the level of the mediator m (`m=`) at which the controlled direct effect is to

be estimated. Here, we have left the level c of the covariates C unspecified in which case the mean of the observed levels of C is used. Furthermore, the type of the outcome (`yreg=`) and mediator (`mreg=`) models need to be specified. The mediation macro employes the procedures of `PROC REG` when the variable is continuous and `PROC LOGISTIC` when the variable is binary. When the outcome model is specified as log-linear, Poisson or negative binomial regression, the procedure `PROC GENMOD` is used.
First output provided by the macro is the results from the outcome regression

The SAS System

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	1.1150	0.5612	3.9477	0.0469
fitness	1	-0.6971	0.2729	6.5252	0.0106
gpa	1	-0.4462	0.0641	48.4804	<.0001
ethni	1	-0.5950	0.4526	1.8069	0.1789
age14	1	-0.0847	0.3527	0.0576	0.8103
age15	1	0.2398	0.4881	0.2413	0.6233
income1	1	-0.00472	0.3075	0.0002	0.9878
income2	1	-0.5439	0.4419	1.5145	0.2184
income3	1	-0.1636	0.3866	0.1790	0.6722
educ1	1	-0.5493	0.4655	1.3921	0.2380
educ2	1	0.0352	0.5848	0.0036	0.9521
educ3	1	-0.4506	0.3766	1.4312	0.2316

Next, the output from the mediator regression is given.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5.56275	0.29592	18.80	<.0001
fitness	1	0.65917	0.10639	5.30	<.0001
ethni	1	-0.05600	0.23092	-0.24	0.8084
age14	1	-0.14817	0.17358	-0.85	0.3935
age15	1	-0.91395	0.28766	-3.18	0.0015
income1	1	0.13433	0.18136	0.74	0.4590
income2	1	0.72474	0.20083	3.61	0.0003
income3	1	0.62931	0.25745	2.44	0.0147
educ1	1	1.52675	0.27085	5.64	<.0001
educ2	1	2.12222	0.30248	7.02	<.0001
educ3	1	0.64497	0.18665	3.46	0.000

Finally, the estimates of the controlled direct and natural effects together with their confidence intervals, standard errors and p-values are given. The estimates and confidence intervals are given as odds ratios. For instance, natural direct effect estimate 0.545 suggests that if the physical fitness level of all the students was changed from low to high without changing their grade point average, the odds of not commencing in post-compulsory education would be decreased $1/0.545 = 1.83$ times.

The SAS System

Obs	Effect	Estimate	p_value	_95_ CI_lower	_95 _CI_upper
1	cde=nde	0.49803	0.010635	0.29171	0.85025
2	nie	0.74518	0.000024	0.65002	0.85428
3	total effect	0.37112	0.000346	0.21563	0.63875

The displayed standard errors are based on the Delta method. Bootstrap-based standard errors and confidence intervals can be obtained by setting option `boot=true` in the `MEDIATION` statement. By default, 1000 bootstrap samples will be computed. Note that when the bootstrap procedure is chosen, the estimates of the controlled direct and natural effects are no longer obtained from the observed data set but as an average of the corresponding estimates from the bootstrap samples. More details about using the mediation macro are given in Valeri and VanderWeele (11, 12).

B.4 Estimation procedure - technical details

The estimation procedure requires specification of two regression models: one for the outcome and one for the mediator. Suppose the following models are fitted to the observed data

$$g_Y\{E[Y\,|A=a,M=m,C=c]\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4^T c \tag{B.5}$$

$$g_M\{E[M\,|A=a,C=c]\} = \beta_0 + \beta_1 a + \beta_2^T c \tag{B.6}$$

where g_Y and g_M are some link functions and $\theta = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4^T)$, $\beta = (\beta_0, \beta_1, \beta_2^T)$ are vectors containing the unknown regression parameters. For a failure time outcome, the model in (B.5) is replaced by either a Cox proportional hazards model or an accelerated failure time model.

Assuming that the identifiability conditions given in equations (4)-(11) in section Theoretical background hold and the models in (B.5,B.6) are correctly specified, the analytic expressions for natural effects as well as controlled direct effect can be derived in terms of β and θ. This is done using Pearl's mediation formula (3, 24). In the presence of continuous mediator, binary (failure time) outcome and exposure-

mediator interaction in the model (B.5), further assumptions are necessary for analytic formulas to hold. Specifically, it is assumed that the residuals of the mediator model (B.6) are normally distributed with constant variance σ^2. Furthermore, in situations with binary (failure time) outcome, the exact analytic formulas are replaced by approximations and hold to the extent that the outcome is rare (in the end of follow-up). More details about the formulas and their derivations can be found in the supplemental materials of Valeri and VanderWeele (11, 12).
The estimation procedure is given by the following steps:

1. The first step of the procedure is to fit the models in equations (B.5-B.6) to the observed data to get the estimates $\hat{\theta}$ and $\hat{\beta}$ (and possibly $\hat{\sigma}^2$).

2. The second step is to plug in $\hat{\theta}$ and $\hat{\beta}$ (and possibly $\hat{\sigma}^2$) from step 1 into the analytic formulas to get the estimates for the natural direct and indirect, total and controlled direct effects. The 95% confidence intervals are obtained via a bootstrap procedure. Alternatively, robust standard errors based on the Delta method can be used.

C R package mediation

C.1 Summary

Imai et al. (3) suggested a general approach to mediation analysis that can be applied in a large number of settings. This approach is implemented by Tingley et al. (14) in the R package mediation. The estimation approach requires specification of two regression models: one for the outcome and one for the mediator.
At present, mediation package allows fitting linear regression models, generalized linear models, ordered response models, generalized additive models, quantile regression models, parametric survival models or multilevel models for both, the mediator and the outcome. For the outcome model, the censored regression model can also be used. Table 2 gives further details about possible types of variables, regression models and distribution families of the dependent variables in the regression models.
In addition, the mediation package allows estimation of the direct and indirect effects when multiple mediators or interactions are present or when the data comes from a specific research design. Imai et al. (25) described a number of research designs and assumptions required for the identification of direct and indirect effects under these designs. The mediation package can accomodate single experiment, parallel, parallel encouragement and crossover encouragement designs.
Finally, when linear regression or probit model is used for mediator and outcome models, it is possible to conduct sensitivity analysis. For further details, see (14).

[Table 6 about here.]

C.2 Parameter of interest

Irrespective of the choice of the models and the type of variables under consideration, the mediation package estimates the marginal natural direct and indirect effects on a difference scale. Furthermore, the mediation package provides both, pure and total natural effects. In particular, for two values of the exposure, a, a^*, the marginal pure natural direct effect $PNDE(a, a^*, c)$, pure natural indirect effect $PNIE(a, a^*, c)$, total natural direct effect $TNDE(a, a^*, c)$, total natural indirect effect $TNIE(a, a^*, c)$ and

total effect $TE(a, a^*, c)$ are defined as

$$PNDE(a,a^*) = E[Y(a,M(a^*))] - E[Y(a^*,M(a^*))] \quad \text{(C.1)}$$
$$PNIE(a,a^*) = E[Y(a^*,M(a))] - E[Y(a^*,M(a^*))] \quad \text{(C.2)}$$
$$TNDE(a,a^*) = E[Y(a,M(a))] - E[Y(a^*,M(a))] \quad \text{(C.3)}$$
$$TNIE(a,a^*) = E[Y(a,M(a))] - E[Y(a,M(a^*))] \quad \text{(C.4)}$$
$$TE(a,a^*) = E[Y(a,M(a))] - E[Y(a^*,M(a^*))]. \quad \text{(C.5)}$$

Besides marginal natural effects, mediation package provides estimate of the pure and total proportion mediated

$$PPM(a,a^*) = \frac{PNIE(a,a^*)}{TE(a,a^*)} \quad \text{(C.6)}$$
$$TPM(a,a^*) = \frac{TNIE(a,a^*)}{TE(a,a^*)} \quad \text{(C.7)}$$

i.e. the proportion of the indirect effect mediated through the mediator relative to the total effect of the exposure.

C.3 Example

In this section, we illustrate how to use mediation package by applying it on the data set from Danish educationd study (20). The study investigates how much of the effect of physical fitness (A) on commencement in post-compulsory education (Y) is mediated through academic achievement measured by grade point average (M).

C.3.1 The data set

The data set consists of 1084 observations and the following variables.

`fitness` Physical fitness coded with 2 levels (0-low, 1-high).

`gpa` Grade point average of the compulsory exams at the end of compulsory education.

`postedu` Commencement in post-compulsory education (0-commenced, 1-not commenced).

`age` Age in years (13, 14, 15).

`ethni` Ethnicity (0-immigrants/descendants, 1-danes).

`income` Parental income in 4 groups (0-3) with 3 being the highest income group.

`educ` Parental education level in 4 groups (0-3) with 3 being the highest.

It is assumed that age, ethnicity, parental income and education level are sufficient to control for confounding. For more details about the data, see (20).

C.3.2 Model specification

The estimation procedure requires specification of two regression models: one for the outcome and one for the mediator. We will use the following logistic regression model for the outcome

$$\text{logit}\{P(Y=1\,|A=a,M=m,C=c)\}=\theta_0+\theta_1 a+\theta_2 m+\theta_3^T c.$$

and the following linear regression model for the mediator

$$E[M\,|A=a,C=c]=\beta_0+\beta_1 a+\beta_2^T c.$$

C.3.3 Implementation

Step 1. Before conducting mediation analysis, the user must provide the data set and the mediation package. We proceed by fitting the mediator and outcome models to the observed data using the `lm` and `glm` functions, respectively.

```
library(mediation)
d1 <- read.csv("d.csv")
Mfit <- lm(formula=gpa~fitness+age+ethni+income+educ,data=d1)
Yfit <- glm(formula=postedu~fitness+gpa+age+ethni+income+educ,data=d1,
    family="binomial")
```

Step 2. The rest of the estimation procedure is done by the `mediate` function in the mediation package. The fitted mediator (`model.m` =) and outcome (`model.y`) models as well as the names of the exposure (`treat` =) and mediator (`mediator` =) variables need to be specified as the inputs to the `mediate` function.

```
med.out <- mediate(model.m=Mfit,model.y=Yfit,treat="fitness",mediator="
    gpa",robustSE=TRUE)
```

The quasi-Bayesian Monte Carlo method (details in section C.4.1) is the default estimation procedure, the nonparametric bootstrap procedure (details in section C.4.2) can be requested by setting the `boot` argument to `TRUE`. By default, the `mediate` function uses 1000 Monte Carlo draws. The number of random draws can be changed by setting the `sims` argument to a preferred number. If the argument `robustSE` is set to `TRUE`, White's heteroskedasticity-consistent estimator (`vcovHC` from R package sandwich) for the covariance matrices $Cov(\hat{\theta})$ and $Cov(\hat{\beta})$ of the

regression parametes will be used. Alternatively, this argument can be omitted if the standard uncertainty estimates are desired.

Step 3. The summary output of the `mediate` function can be requested by the following command.

```
summary(med.out)

Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

                          Estimate 95% CI Lower 95% CI Upper p-value
ACME (control)           -0.02321     -0.03459     -0.01256    0.00
ACME (treated)           -0.01428     -0.02388     -0.00698    0.01
ADE (control)            -0.04797     -0.08418     -0.00890    0.01
ADE (treated)            -0.03904     -0.06872     -0.00715    0.01
Total Effect             -0.06225     -0.09616     -0.02684    0.00
Prop. Mediated (control)  0.36901      0.22062      0.76241    0.06
Prop. Mediated (treated)  0.21790      0.10151      0.70443    0.19
ACME (average)           -0.01875     -0.02805     -0.01047    0.00
ADE (average)            -0.04350     -0.07631     -0.00803    0.01
Prop. Mediated (average)  0.29345      0.16414      0.73550    0.11

Sample Size Used: 1084


Simulations: 1000
```

The estimates of `ADE` and `ACME` in the summary output of the mediate function correspond to the marginal natural direct and natural indirect effects on the difference scale respectively. In addition, the estimates of the total effect and proportion mediated are displayed. Note that all the causal parameters are estimated for `treated`, `control` and `average`. The estimates denoted as `treated` correspond to the so-called total natural effects where as the estimates denoted as `control` correspond to the so-called pure effects described in Theoretical background section. For instance, natural pure direct effect estimate (`ADE(control)`) -0.0478 suggests that if the physical fitness level of all the students was changed from low to high without changing their grade point average, the risk of not commencing in post-compulsory education would decrease by 0.0478. The estimates denoted as `average` are simply the average of the corresponding total and pure natural effects. For more details about how to conduct sensitivity analysis, how to estimate direct and indirect effects under different research designs and other functionalities of mediation package, see (14).

C.4 Estimation procedure - technical details

The estimation procedure requires two regression models: one for the outcome and one for the mediator. Consider the following models

$$g_Y\{E[Y\,|A=a,M=m,C=c]\}=\theta^T H(a,m,c) \tag{C.8}$$
$$g_M\{E[M\,|A=a,C=c]\}=\beta^T V(a,c) \tag{C.9}$$

where $H(a,m,c)$ and $V(a,c)$ are known vectors that include components that may depend on a,m,c and a,c, respectively, and θ and β are vectors of unknown parameters. In practice, $H(a,m,c)$ and $V(a,c)$ are often simply vectors $(1,a,m,c)$ and $(1,a,c)$, respectively. The estimation procedure relies on the fact that Monte Carlo draws of counterfactual outcomes $Y(a,M(a^*))$ can be obtained based on the mediator and outcome models in equations $(C.8-C.9)$. Once the draws of potential outcomes have been obtained, natural effects that are functions of these potential outcomes, can easily be computed, regardless of the statistical models used (3). The mediation package provides two different estimation procedures, one based on quasi-Bayesian Monte Carlo approximations and other based on nonparametric bootstrap techniques (14). In the next sections, we provide the description of both algorithms.

C.4.1 Parametric inference

The first algorithm is based on the quasi-Bayesian Monte Carlo approximations (described e.g. in (26)) and approximates the posterior distribution of the quantities of interest by their sampling distribution (3). The algorithm is very general since it can be applied to any parametric statistical model. The algorithm proceeds as follows:

1. The first step of the algorithm is to fit the models in equations $(C.8-C.9)$ to get the estimates $\hat{\theta}$, $Cov(\hat{\theta})$ and $\hat{\beta}$, $Cov(\hat{\beta})$, respectively. Next step is to simulate the model parameters $\tilde{\theta}$ and $\tilde{\beta}$ from their sampling distributions. The sampling distributions are approximated by the multivariate normal distributions with $\mathcal{N}(\hat{\theta}^T, Cov(\hat{\theta}))$ and $\mathcal{N}(\hat{\beta}^T, Cov(\hat{\beta}))$, respectively.

2. The next step is the following:

   (a) For each individual i, $i=1,\ldots,n$, simulate some number K copies of the counterfactual values of the mediator $M_i^k(a)$, $M_i^k(a^*)$ for each exposure value a,a^* from the distribution $F_M(\cdot\,|A,C;\tilde{\beta})$ defined by the model in

equation (C.9) and the random draw of the parameter $\tilde{\beta}$ from step 1:

$$M_i^k(a) \sim F_M(M_i \,|\, A = a, C = C_i; \tilde{\beta})$$
$$M_i^k(a^*) \sim F_M(M_i \,|\, A = a^*, C = C_i; \tilde{\beta}).$$

When the distribution of $F_M(\cdot|A, C)$ is not fully specified by the model in equation (C.9), e.g. for continuous M, additional parametric assumptions are necessary.

(b) For each individual i, $i = 1, \ldots, n$, simulate some number K copies of the counterfactual outcome values $Y_i^k(a, M_i^k(a))$, $Y_i^k(a, M_i^k(a^*))$, $Y_i^k(a^*, M_i^k(a))$, $Y_i^k(a^*, M_i^k(a^*))$ for each possible exposure level a, a^* under each simulated mediator $M_i^k(a)$ and $M_i^k(a^*)$ from the distribution $F_Y(Y \,|\, A, M, C; \tilde{\theta})$ specified by the model in equetion (C.8) and the random draw of the parameter $\tilde{\theta}$ from step 1:

$$Y_i^k(a, M_i^k(a)) \sim F_Y(Y_i \,|\, A = a, M = M_i^k(a), C = C_i; \tilde{\theta})$$
$$Y_i^k(a, M_i^k(a^*)) \sim F_Y(Y_i \,|\, A = a, M = M_i^k(a^*), C = C_i; \tilde{\theta})$$
$$Y_i^k(a^*, M_i^k(a)) \sim F_Y(Y_i \,|\, A = a^*, M = M_i^k(a), C = C_i; \tilde{\theta})$$
$$Y_i^k(a^*, M_i^k(a^*)) \sim F_Y(Y_i \,|\, A = a^*, M = M_i^k(a^*), C = C_i; \tilde{\theta}).$$

Again, if the distribution $F_Y(\cdot|A, M, C)$ is not fully specified by the model in equation (C.8), additional parametric assumptions are necessary. However, see exceptions described below.

(c) Compute the estimates of causal effects by averaging the draws of counterfactual outcomes from previous step over K copies and all individuals

in the sample:

$$\widehat{PNDE}(a,a^*) = \frac{1}{nK}\sum_{i=1}^{n}\sum_{i=1}^{K}\{Y_i^k(a,M_i^k(a^*)) - Y_i^k(a^*,M_i^k(a^*))\} \quad \text{(C.10)}$$

$$\widehat{TNDE}(a,a^*) = \frac{1}{nK}\sum_{i=1}^{n}\sum_{i=1}^{K}\{Y_i^k(a,M_i^k(a)) - Y_i^k(a^*,M_i^k(a))\} \quad \text{(C.11)}$$

$$\widehat{PNIE}(a,a^*) = \frac{1}{nK}\sum_{i=1}^{n}\sum_{i=1}^{K}\{Y_i^k(a^*,M_i^k(a)) - Y_i^k(a^*,M_i^k(a^*))\} \quad \text{(C.12)}$$

$$\widehat{TNIE}(a,a^*) = \frac{1}{nK}\sum_{i=1}^{n}\sum_{i=1}^{K}\{Y_i^k(a,M_i^k(a)) - Y_i^k(a,M_i^k(a^*))\} \quad \text{(C.13)}$$

$$\widehat{TE}(a,a^*) = \widehat{PNDE}(a,a^*) + \widehat{TNIE}(a,a^*) \quad \text{(C.14)}$$

$$= \widehat{TNDE}(a,a^*) + \widehat{PNIE}(a,a^*) \quad \text{(C.15)}$$

3. Repeat the steps 1 and 2 N times.

4. The final step is to compute the point estimates and confidence intervals of the causal effects. The point estimates are computed by taking the average over the N estimates of causal effects computed in the previous step. In addition, 2.5% and 97.5% quantiles of the N estimates of causal effects computed in the previous step are used to get the 95% confidence intervals for point estimates.

In some cases, the algorithm 1 can be simplified. For instance, if the natural effects can be analytically derived from the selected outcome model in equation (C.8), step 2(b) can be skipped and step 2(c) can be modified to compute the estimates of natural direct and indirect effects by

$$\widehat{PNDE}(a,a^*) = \frac{1}{nK}\sum_{i=1}^{n}\sum_{i=1}^{K}\big\{E[Y_i\,|A=a,M=M_i^k(a^*),C=C_i;\tilde{\theta}] - E[Y_i\,|A=a^*,M=M_i^k(a^*),C=C_i;\tilde{\theta}]\big\} \tag{C.16}$$

$$\widehat{TNDE}(a,a^*) = \frac{1}{nK}\sum_{i=1}^{n}\sum_{i=1}^{K}\big\{E[Y_i\,|A=a,M=M_i^k(a),C=C_i;\tilde{\theta}] - E[Y_i\,|A=a^*,M=M_i^k(a),C=C_i;\tilde{\theta}]\big\} \tag{C.17}$$

$$\widehat{PNIE}(a,a^*) = \frac{1}{nK}\sum_{i=1}^{n}\sum_{i=1}^{K}\big\{E[Y_i\,|A=a,M=M_i^k(a),C=C_i;\tilde{\theta}] - E[Y_i\,|A=a,M=M_i^k(a^*),C=C_i;\tilde{\theta}]\big\} \tag{C.18}$$

$$\widehat{TNIE}(a,a^*) = \frac{1}{nK}\sum_{i=1}^{n}\sum_{i=1}^{K}\big\{E[Y_i\,|A=a,M=M_i^k(a),C=C_i;\tilde{\theta}] - E[Y_i\,|A=a,M=M_i^k(a^*),C=C_i;\tilde{\theta}]\big\} \tag{C.19}$$

where the expectations are given by

$$\begin{aligned}
E[Y_i\,|A=a,M=M_i^k(a^*),C=C_i;\tilde{\theta}] &= g_Y^{-1}\{\tilde{\theta}^T H(a,M_i^k(a^*),C_i)\}\\
E[Y_i\,|A=a^*,M=M_i^k(a^*),C=C_i;\tilde{\theta}] &= g_Y^{-1}\{\tilde{\theta}^T H(a^*,M_i^k(a^*),C_i)\}\\
E[Y_i\,|A=a,M=M_i^k(a),C=C_i;\tilde{\theta}] &= g_Y^{-1}\{\tilde{\theta}^T H(a,M_i^k(a),C_i)\}.\\
E[Y_i\,|A=a^*,M=M_i^k(a),C=C_i;\tilde{\theta}] &= g_Y^{-1}\{\tilde{\theta}^T H(a^*,M_i^k(a),C_i)\}.
\end{aligned}$$

For further details about the estimation procedure, see (3).

C.4.2 Nonparametric inference

The second algorithm is based on a nonparametric bootstrap procedure and allows to use more complex models such as non- or semiparametric models and quantile regression models (3). The algorithm proceeds as follows:

1. Take a random sample with replacement of sample size n from the observed data.

2. The next step is the following:

   (a) Fit the models in equations $(C.8-C.9)$ to the random sample to get the estimates $\hat{\theta}$ and $\hat{\beta}$ respectively.

(b) For each individual i, $i = 1, \ldots, n$, simulate some number K copies of the counterfactual values of the mediator $M_i^k(a), M_i^k(a^*)$ for each possible exposure a, a^* from the distribution $F_M(\cdot \,|A, C; \hat{\beta})$ defined by the model in equation (C.9) and the estimate of the parameter $\hat{\beta}$ obtained in the previous step:

$$M_i^k(a) \sim F_M(M_i \,|A = a, C = C_i; \hat{\beta})$$
$$M_i^k(a^*) \sim F_M(M_i \,|A = a^*, C = C_i; \hat{\beta}).$$

When the distribution of $F_M(\cdot|A, C)$ is not fully specified by the model in equation (C.9), e.g. for continuous M, additional parametric assumptions are necessary.

(c) For each individual i, $i = 1, \ldots, n$, simulate some number K copies of the counterfactual outcome values $Y_i^k(a, M_i^k(a))$, $Y_i^k(a, M_i^k(a^*))$, $Y_i^k(a^*, M_i^k(a))$, $Y_i^k(a^*, M_i^k(a^*))$ for each possible exposure level a, a^* under each simulated mediator $M_i^k(a)$ and $M_i^k(a^*)$ from the distribution $F_Y(\cdot \,|A, M, C; \hat{\theta})$ specified by the model in equation (C.8) and the estimate of the parameter $\hat{\theta}$ from the step 2(a):

$$Y_i^k(a, M_i^k(a)) \sim F_Y(Y_i \,|A = a, M = M_i^k(a), C = C_i; \hat{\theta})$$
$$Y_i^k(a, M_i^k(a^*)) \sim F_Y(Y_i \,|A = a, M = M_i^k(a^*), C = C_i; \hat{\theta})$$
$$Y_i^k(a^*, M_i^k(a)) \sim F_Y(Y_i \,|A = a^*, M = M_i^k(a), C = C_i; \hat{\theta})$$
$$Y_i^k(a^*, M_i^k(a^*)) \sim F_Y(Y_i \,|A = a^*, M = M_i^k(a^*), C = C_i; \hat{\theta}).$$

Again, the distribution $F_Y(\cdot|A, M, C)$ might not be fully specified by the model in equation (C.8). Then, additional parametric assumptions are necessary. However, see exceptions described below.

(d) Compute the estimates of natural effects by averaging the draws of counterfactual outcomes from previous step over K copies and all individuals

in the random sample:

$$\widehat{PNDE}(a,a^*) = \frac{1}{nK}\sum_{i=1}^{n}\sum_{i=1}^{K}\{Y_i^k(a,M_i^k(a^*)) - Y_i^k(a^*,M_i^k(a^*))\} \quad \text{(C.20)}$$

$$\widehat{TNDE}(a,a^*) = \frac{1}{nK}\sum_{i=1}^{n}\sum_{i=1}^{K}\{Y_i^k(a,M_i^k(a)) - Y_i^k(a^*,M_i^k(a))\} \quad \text{(C.21)}$$

$$\widehat{PNIE}(a,a^*) = \frac{1}{nK}\sum_{i=1}^{n}\sum_{i=1}^{K}\{Y_i^k(a^*,M_i^k(a)) - Y_i^k(a^*,M_i^k(a^*))\} \quad \text{(C.22)}$$

$$\widehat{TNIE}(a,a^*) = \frac{1}{nK}\sum_{i=1}^{n}\sum_{i=1}^{K}\{Y_i^k(a,M_i^k(a)) - Y_i^k(a,M_i^k(a^*))\} \quad \text{(C.23)}$$

3. Repeat the steps 1 and 2 N times.

4. The final step is to compute the point estimates and confidence intervals of the natural effects. The point estimates are computed by taking the average over the N estimates of natural effects computed in the previous step. In addition, 2.5% and 97.5% quantiles of the N estimates of natural effects computed in the previous step are used to get the 95% confidence intervals for point estimates.

As before, in some cases the algorithm 2 can be simplified to skip the step 2(c) and modify the step 2(d) such that the estimates of natural effects are computed as in equations $(C.16 - C.19)$.

D Weighting method implemented in R medflex package

D.1 Summary

Mediation approaches that build on simple regression models for outcome and mediator often result in difficult expressions for direct and indirect effects (7, 8, 10). To overcome these difficulties, Lange et al. (7), Vansteelandt et al. (8) suggested using so-called natural effects models (natural effect models), which directly parameterize the natural effects of interest. Two different procedures for estimating natural effects by fitting natural effect models have been implemented in the R package medflex (15). In this section we will discuss the approach based on weighting strategies suggested by Lange et al. (7). This approach requires specification of two models: a regression model for the mediator and a natural effects model for the counterfactual outcome. Even though the approach can, in principle, be applied to almost any combination of variable types and statistical models, a limited number has been implemented in the medflex package (7, 15). When using the weighting-based approach, currently models for the mediator can be fitted using the `glm` function or the `vglm` function from the VGAM package. Models for the outcome can be fitted only using the `glm` function. An overview of the type of variables and models available in medflex package when using the weighting approach, is given in Table 3.

[Table 7 about here.]

In the presence of multiple mediators, applying weighting approach entails fitting a model for each of the mediators separately which is a taunting task (15). Hence, estimation of joint mediated effect of several mediators is not possible with the weighting approach in medflex package. The weighting approach can be applied in the presence of interactions. Currently the medflex package does not include any specific tools for conducting sensitivity analysis. Thus, conducting sensitivity analysis requires further coding from the user.

D.2 Parameter of interest

By default, the weighting-based estimation procedure implemented in the medflex package estimates the conditional natural effects within the levels of the covariates C. In particular, for two values of the exposure, a, a^*, and any value c of the covariates C, the conditional natural direct effect $NDE(a, a^*, c)$, natural indirect effect

$NIE(a,a^*,c)$ and total effect $TE(a,a^*,c)$ are defined as

$$NDE(a,a^*,c) = g\{E[Y(a,M(a^*))\,|\,C=c]\} - g\{E[Y(a^*,M(a^*))\,|\,C=c]\} \quad \text{(D.1)}$$
$$NIE(a,a^*,c) = g\{E[Y(a,M(a))\,|\,C=c]\} - g\{E[Y(a,M(a^*))\,|\,C=c]\} \quad \text{(D.2)}$$
$$TE(a,a^*,c) = g\{E[Y(a,M(a))\,|\,C=c]\} - g\{E[Y(a^*,M(a^*))\,|\,C=c]\} \quad \text{(D.3)}$$

where g is some link function. For example, the measure of natural direct effect for a binary outcome can be expressed as an odds ratio

$$\begin{aligned}
NDE(a,a^*,c) &= \text{logit}\{E[Y(a,M(a^*))\,|\,C=c]\} - \text{logit}\{E[Y(a^*,M(a^*))\,|\,C=c]\} \\
&= \text{logit}\{P(Y(a,M(a^*))=1\,|\,C=c)\} - \text{logit}\{P(Y(a^*,M(a^*))=1\,|\,C=c)\} \\
&= \log\left\{\frac{P(Y(a,M(a^*))=1\,|\,C=c)}{1-P(Y(a,M(a^*))=1\,|\,C=c)}\right\} - \\
&\quad \log\left\{\frac{P(Y(a^*,M(a^*))=1\,|\,C=c)}{1-P(Y(a^*,M(a^*))=1\,|\,C=c)}\right\} \\
&= \log\left\{\frac{odds\big(P(Y(a,M(a^*))=1\,|\,C=c)\big)}{odds\big(P(Y(a^*,M(a^*))=1\,|\,C=c)\big)}\right\} \\
&= \log\{OR_{NDE}(a,a^*,c)\}
\end{aligned}$$

by using $g = \text{logit}$ link function. Natural effects given in equations (D.1-D.3) are assumed to be the same for any chosen value of c (unless exposure-covariate interactions are present). In addition to conditional natural effects, the medflex package enables to estimate marginal natural effects. The marginal natural effects are defined similarly to conditional natural effects by replacing the conditional expectations $E[Y(a,M(a^*))\,|\,C=c]$ in equations (D.1-D.3) with $E[Y(a,M(a^*))]$.

D.3 Example

In this section, we illustrate how to use medflex package by applying it on the data set from Danish education study (20). The study investigates how much of the effect of physical fitness (A) on commencement in post-compulsory education (Y) is mediated through academic achievement measured by grade point average (M).

D.3.1 The data set

The data set consists of 1084 observations and the following variables.

`fitness` Physical fitness coded with 2 levels (0-low, 1-high).

`gpa` Grade point average of the compulsory exams at the end of compulsory education.

`postedu` Commencement in post-compulsory education (0-commenced, 1-not commenced).

`age` Age in years (13, 14, 15).

`ethni` Ethnicity (0-immigrants/descendants, 1-danes).

`income` Parental income in 4 groups (0-3) with 3 being the highest income group.

`educ` Parental education level in 4 groups (0-3) with 3 being the highest.

It is assumed that age, ethnicity, parental income and education level are sufficient to control for confounding. For more details about the data, see (20).

D.3.2 Model specification

The estimation procedure requires specification of two models: one for the mediator and one for the counterfactual outcome. We will consider the following linear regression model for the mediator

$$E[M \mid A = a, C = c] = \beta_0 + \beta_1 a + \beta_2^T c. \tag{D.4}$$

Here we implicitly assume that the error terms of model in equation (D.4) are normally distributed with fixed unknown variance σ^2. We will consider the following natural effects model for the counterfactual outcome

$$\text{logit}\{P(Y(a, M(a^*)) = 1 \mid C = c)\} = \theta_0 + \theta_1 a + \theta_2 a^* + \theta_3^T c. \tag{D.5}$$

The vector θ in the logistic regression model in equation (D.5) includes parameters that capture the conditional causal effects of interest. In particular, the conditional natural effects log-odds ratos are given by

$$\begin{aligned} log\{OR_{NDE}(1,0,c)\} &= \theta_1, \\ log\{OR_{NIE}(1,0,c)\} &= \theta_2, \\ log\{OR_{TE}(c)\} &= \theta_1 + \theta_2. \end{aligned}$$

D.3.3 Implementation

Step 1. Before conducting mediation analysis, the user must provide the data set and the medflex package. We proceed by fitting the mediator model for the observed data using the `glm` function. Note that, when the exposure is not continous, it needs to be specified as a factor, either in the formula for the mediator model or in the observed data set.

```
library(medflex)
d1 <- read.csv("d.csv")
Mfit <- glm(formula=gpa~fitness+age+ethni+income+educ,data=d1,
    family="gaussian")
```

Step 2. Next, using the `neWeight` function of the medflex package we prepare the data for fitting natural effects model. The fitted mediator model object needs to be specified as the first argument in the `neWeight` function. As a result, we obtain a new data set with two new variables corresponding to the exposure values a and a^* (details in Appendix D.4). The `neWeight` function uses the name of the exposure in the original data set with indeces "0" and "1" to denote the new variables.

```
expData <- neWeight(Mfit)
```

Step 3. The natural effects model in equation (D.5) is fitted to the new data set using the `neModel` function. Note that the formula for the natural effects model must include the new variables instead of the exposure and mediator.

```
Yout <- neModel(formula=postedu~fitness0+fitness1+age+ethni+income+educ
    ,expData=expData,family="binomial")
summary(Yout)

Natural effect model
with standard errors based on the non-parametric bootstrap
---
Exposure: fitness
Mediator(s): gpa
---
Parameter estimates:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.1006     0.5124   -2.15    0.032 *
fitness01    -0.6640     0.2807   -2.37    0.018 *
fitness11    -0.2896     0.0689   -4.20 0.000027 ***
age14        -0.0242     0.3623   -0.07    0.947
age15         0.4747     0.5388    0.88    0.378
ethni        -0.6644     0.5389   -1.23    0.218
income1      -0.0308     0.2947   -0.10    0.917
income2      -0.7467     0.3880   -1.92    0.054 .
income3      -0.8438     0.4754   -1.77    0.076 .
educ1        -0.4053     0.4095   -0.99    0.322
educ2        -1.1993     0.5265   -2.28    0.023 *
educ3        -0.8416     0.9669   -0.87    0.384
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The estimates of `fitness01` and `fitness11` in the summary output of the fitted natural effects model correspond to the natural direct and natural indirect effects on the log-odds scale respectively. For instance, the natural direct effect estimate (`fitness01`) -0.664 suggests that if the physical fitness level of all the students was

changed from low to high without changing their grade point average, the odds of not commencing in post-compulsory education would be decreased $1/\exp(-0.664) = 1/0.515 = 1.94$ times. By default, the bootstrap standard errors are calculated using 1000 replications. The number of replications can be set in the `nBoot` argument. Alternatively, robust standard errors can be computed by setting argument `se = "robust"`.

Step 4. The summary output of natural effect model obtained in Step 3 does not include the estimate of the total effect. The `neEffdecomp` function can be used for effect decomposition. The summary output of the effect decomposition includes estimates, standard errors and p-values for the natural effects as well as for the total effect.

```
effdecomp <- neEffdecomp(Yout)
summary(effdecomp)

Effect decomposition on the scale of the linear predictor
with standard errors based on the non-parametric bootstrap
---
conditional on: age, ethni, income, educ
with x* = 0, x = 1
---
                        Estimate Std. Error z value Pr(>|z|)
natural direct effect    -0.6640     0.2807  -2.37  0.01799 *
natural indirect effect  -0.2896     0.0689  -4.20 0.000027 ***
total effect             -0.9536     0.2835  -3.36  0.00077 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Univariate p-values reported)
```

For more details about medlfex package, see (15).

D.4 Estimation procedure - technical details

The estimation procedure requires specification of two models: a regression model for the mediator and a natural effects model for the nested counterfactuals, $Y(a, M(a^*))$. Consider

$$g_M\{E[M \mid A = a, C = c]\} = \beta^T V(a, c) \tag{D.6}$$

as a mediator model, where g_M is some link function, $V(a, c)$ is a known vector with components that may depend on a and c and β is a vector of unknown regression parameters. Let

$$g_Y\{E[Y(a, M(a^*)) \mid C = c]\} = \theta^T H(a, a^*, c), \tag{D.7}$$

with a, a^* two values of exposure to compare, g_Y some link function, $H(a, a^*, c)$ a known vector with components that may depend on a, a^* and c and θ a vector of unknown regression parameters. In practice, $H(a, a^*, c)$ and $V(a, c)$ often corresponds to simple vectors, such as $(1, a, a^*, c)$ and $(1, a, c)$, respectively.
Under identifiability conditions given in equations (4−11) given in Theoretical background section, it has been showed (e.g. in (15)) that

$$E[Y(a, M(a^*)) \,|\, C = c] = E[W \cdot Y \,|\, A = a, C = c], \tag{D.8}$$

with

$$W = \frac{f_M(m \,|\, A = a^*, C)}{f_M(m \,|\, A = a, C)}, \tag{D.9}$$

where $f_M(\cdot \,|\, A, C)$ denotes the conditional density of the mediator given the exposure and the covariates. When the mediator is categorical, the density is replaced by the probability. Property in equation (D.8) gives arise to a weighting procedure whereby the observed data set is expanded to construct a pseudo-population where the same individuals are evaluated at different mediator levels, but the observed exposure level, by using the weights W in equation (D.9) (7). For simplicity, we will describe the estimation procedure for a binary exposure A.

1. Fit a regression model for the mediator given in equation (D.6) to the observed data set to obtain the estimates $\hat{\beta}$.

2. Expand the observed data by repeating each observation in the original data set twice and including two additional variables: A_0, which is equal to the observed exposure A, and A_1, which is equal to the observed exposure for the first replication and equal to the opposite of the observed exposure, $1 - A$, for the second replication. The medflex package denotes the new variables A_0 and A_1 by using the name of the exposure from the original data set with indeces 0 and 1, respectively.

3. Compute the weights

$$W = \frac{f_M(m \,|\, A = A_0, C; \hat{\beta})}{f_M(m \,|\, A = A_1, C; \hat{\beta})}$$

   by applying the fitted mediator model in equation (D.6) to the expanded data set twice, once setting the exposure to A_1 defined in the previous step and once setting the exposure to A_0 defined in the previous step. Here, it is implicitly assumed that M is generated from a particular distribution $F_M(\cdot|A, C)$

with density $f_M(\cdot|A, C)$, i.e not only the mean but also the variance of M is determined by the model in equation (D.6).

4. The last step requires fitting the model in equation (D.7) to the expanded data set from the previous steps by using $a = A_0$ and $a^* = A_1$ and weights W computed in the previous step. The estimates for natural direct and indirect effects are given by the certain components of the vector of estimated regression parameters $\hat{\theta}$. The standard errors are obtained from a bootstrap procedure. Alternatively, robust standard errors can be used.

The estimation procedure is similar for continuous or categorical exposures with more than two possible levels. Furthermore, it can easily be extended to obtain estimates of the marginal natural effects. For futher details, see (7, 15).

E Imputation method implemented in R medflex package

E.1 Summary

Mediation approaches that build on simple regression models for outcome and mediator often result in difficult expressions for direct and indirect effects (7, 8, 10). To overcome these difficulties, Lange et al. (7), Vansteelandt et al. (8) suggested using so-called natural effects models (natural effect models), which directly parameterize the natural effects of interest. Two different procedures for estimating natural effects by fitting natural effect models have been implemented in the R package medflex (15). In this section, we will discuss the approach based on imputation strategies suggested by Vansteelandt et al. (8). This approach requires specification of two models: an imputation model for the outcome and the natural effects model for the counterfactual outcome. Even though the approach can, in principle, be applied to almost any combination of variable types, a limited number has been implemented in the medflex package (8, 15). When using the imputation-based approach, imputation and natural effects models can only be fitted using the `glm` function. An overview of the types of variables and models that can be used with the imputation approach, is given in Table 4.

[Table 8 about here.]

The imputation approach in the medflex package allows to estimate the direct and indirect effects in the presence of multiple mediators and interactions. At present, the medflex package does not include any specific tools for conducting sensitivity analysis. Thus, conducting sensitivity analysis requires further coding from the user. For further details, see (15).

E.2 Parameter of interest

By default, the imputation-based estimation procedure implemented in the medflex package estimates the conditional natural effects within the levels of the covariates C. In particular, for two values of the exposure, a, a^*, and any value c of covariates C, the conditional natural direct effect $NDE(a, a^*, c)$, natural indirect effect $NIE(a, a^*, c)$ and total effect $TE(a, a^*, c)$ are defined as

$$NDE(a,a^*,c) = g\{E[Y(a,M(a^*))\,|\,C=c]\} - g\{E[Y(a^*,M(a^*))\,|\,C=c]\} \quad \text{(E.1)}$$

$$NIE(a,a^*,c) = g\{E[Y(a,M(a))\,|\,C=c]\} - g\{E[Y(a,M(a^*))\,|\,C=c]\} \quad \text{(E.2)}$$

$$TE(a,a^*,c) = g\{E[Y(a,M(a))\,|\,C=c]\} - g\{E[Y(a^*,M(a^*))\,|\,C=c]\} \quad \text{(E.3)}$$

where g is some link function. For example, the measure of natural direct effect for a binary outcome can be expressed as an odds ratio

$$\begin{aligned}
NDE(a,a^*,c) &= \text{logit}\{E[Y(a,M(a^*))\,|\,C=c]\} - \text{logit}\{E[Y(a^*,M(a^*))\,|\,C=c]\} \\
&= \text{logit}\{P(Y(a,M(a^*))=1\,|\,C=c)\} - \text{logit}\{P(Y(a^*,M(a^*))=1\,|\,C=c)\} \\
&= \log\left\{\frac{P(Y(a,M(a^*))=1\,|\,C=c)}{1-P(Y(a,M(a^*))=1\,|\,C=c)}\right\} - \\
&\quad \log\left\{\frac{P(Y(a^*,M(a^*))=1\,|\,C=c)}{1-P(Y(a^*,M(a^*))=1\,|\,C=c)}\right\} \\
&= \log\left\{\frac{odds\big(P(Y(a,M(a^*))=1\,|\,C=c)\big)}{odds\big(P(Y(a^*,M(a^*))=1\,|\,C=c)\big)}\right\} \\
&= \log\{OR_{NDE}(a,a^*,c)\}
\end{aligned}$$

by using g = logit link function. Natural effects given in equations (E.1-E.3) are assumed to be the same for any chosen value of c (unless exposure-covariate interactions are present). In addition to conditional natural effects, the medflex package enables to estimate marginal natural effects. The marginal natural effects are defined similarly to conditional natural effects by replacing the conditional expectations $E[Y(a,M(a^*))\,|\,C=c]$ in equations (E.1-E.3) with $E[Y(a,M(a^*))]$.

E.3 Example

In this section, we illustrate how to use medflex package by applying it on a data set from a study by (20). The study investigates how much of the effect of physical fitness (A) on commencement in post-compulsory education (Y) is mediated through academic achievement measured by grade point average (M).

E.3.1 The data set

The data set consists of 1084 observations and the following variables.

`fitness` Physical fitness coded with 2 levels (0-low, 1-high).

`gpa` Grade point average of the compulsory exams at the end of compulsory education.

`postedu` Commencement in post-compulsory education (0-commenced, 1-not commenced).

`age` Age in years (13, 14, 15).

`ethni` Ethnicity (0-immigrants/descendants, 1-danes).

`income` Parental income in 4 groups (0-3) with 3 being the highest income group.

`educ` Parental education level in 4 groups (0-3) with 3 being the highest.

It is assumed that age, ethnicity, parental income and education level are sufficient to control for confounding. In the analysis we will consider not being commenced in post-compulsory education ($Y = 0$) as the event of interest. For more details about the data, see (20).

E.3.2 Model specification

The estimation procedure requires specification of two models: an imputation model for the outcome and so-called natural effects model for the nested counterfactual outcome. We will consider the following logistic regression model as the imputation model

$$\text{logit}\{P(Y = 1 \,|\, A = a, M = m, C = c)\} = \beta_0 + \beta_1 a + \beta_2 m + \beta_3^T c.$$

and the following natural effects model for the counterfactual outcome

$$\text{logit}\{P(Y(a, M(a^*)) = 1 \,|\, C = c)\} = \theta_0 + \theta_1 a + \theta_2 a^* + \theta_3^T c. \tag{E.4}$$

The vector θ in the logistic regression model in equation (E.4) includes parameters that capture the conditional causal effects of interest. In particular, the conditional natural effects log-odds ratos are given by

$$\begin{aligned} log\big\{OR_{NDE}(1,0,c)\big\} &= \theta_1, \\ log\big\{OR_{NIE}(1,0,c)\big\} &= \theta_2, \\ log\big\{OR_{TE}(c)\big\} &= \theta_1 + \theta_2. \end{aligned}$$

E.3.3 Implementation

Step 1. Before conducting mediation analysis, the user must provide the data set and the medflex package. We proceed by fitting the imputation model to the observed data using the `glm` function. Note that when the exposure is not continous, it needs to be specified as a factor either in the formula for imputation model or in the observed data set.

```
library(medflex)
d1 <- read.csv("d.csv")
Yfit <- glm(formula=postedu~fitness+gpa+age+ethni+income+educ,data=d1,
    family="binomial")
```

Step 2. Next, using the `neImpute` function of the medflex package we prepare the observed data set for fitting natural effects model (details in section E.4). The fitted imputation model object needs to be specified as the first argument in the `neImpute` function. As a result, we have a new data set with two new variables corresponding to two exposure values a and a^* (details in section E.4). The `neImpute` function uses the name of the exposure in the original data set with indeces "0" and "1" to denote the new variables.

```
impData <- neImpute(Yfit)
```

Step 3. The natural effects model in equation (E.4) is fitted to the new data set using the `neModel` function. Note that the formula for the natural effects model must include the two new variables from the new data set instead of exposure and mediator.

```
Yout <- neModel(formula=postedu~fitness0+fitness1+age+ethni+income+educ
    ,expData=impData,family="binomial")
summary(Yout)

Natural effect model
with standard errors based on the non-parametric bootstrap
---
Exposure: fitness
Mediator(s): gpa
---
Parameter estimates:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.1401    0.5039   -2.36   0.02367 *
fitness01   -0.6488    0.2775   -2.34   0.01940 *
fitness11   -0.2969    0.0823   -3.61   0.00031 ***
age14       -0.0075    0.3626   -0.02   0.98350
age15        0.5458    0.5366    1.02   0.30910 *
ethni       -0.5766    0.5263   -1.10   0.27329
income1     -0.0446    0.2923   -0.15   0.87874
income2     -0.6902    0.3774   -1.83   0.06742 .
income3     -0.8546    0.4561   -1.87   0.06097 .
educ1       -0.3899    0.4018   -0.97   0.33186
educ2       -1.1841    0.5112   -2.32   0.02055 *
educ3       -0.8615    0.9395   -0.92   0.35916
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The estimates of `fitness01` and `fitness1`‘ in the summary output of the fitted natural effects model correspond to the natural direct and natural indirect effects on the log-odds scale, respectively. For instance, the natural direct effect estimate (`fitness01`) -0.649 suggests that if the physical fitness level of all the students was changed from low to high without changing their grade point average, the odds of not commencing in post-compulsory education would be decreased $1/\exp(-0.649) =$

$1/0.523 = 1.91$ times. By default, the bootstrap standard errors are calculated using 1000 replications. The number of replications can be set in the `nBoot` argument. Alternatively, robust standard errors can be computed by setting argument `se = "robust"`.

Step 4. The summary output of the natural effect model obtained in Step 3 does not include the estiomate of the total effect. The `neEffdecomp` function can be used for effect decomposition. The summary output of the effect decomposition includes estimates, standard errors and p-values for the natural effects as well as for the total effect.

```
effdecomp <- neEffdecomp(Yout)
summary(effdecomp)


Effect decomposition on the scale of the linear predictor
with standard errors based on the non-parametric bootstrap
---
conditional on: age, ethni, income, educ
with x* = 0, x = 1
---
                        Estimate Std. Error z value Pr(>|z|)
natural direct effect    -0.6488     0.2775   -2.34  0.01940 *
natural indirect effect  -0.2969     0.0823   -3.61  0.00031 ***
total effect             -0.9458     0.2802   -3.38  0.00074 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Univariate p-values reported)
```

For more details about medflex package, see (15).

E.4 Estimation procedure - technical details

The estimation procedure requires specification of two models: an imputation model for the outcome and a natural effects model for the nested counterfactuals, $Y(a, M(a^*))$. Let

$$g_Y\{E[Y \mid A = a, M = m, C = c]\} = \beta^T V(a, m, c), \tag{E.5}$$

where g_Y is a link function, $V(a, m, c)$ is a known vector with components that may depend on a, m and c and β is a vector of unknown regression parameters. In addition, let

$$g\{E[Y(a, M(a^*)) \mid C = c]\} = \theta^T H(a, a^*, c), \tag{E.6}$$

with a, a^* two values of exposure to compare, g_Y some link function, $H(a, a^*, c)$ a known vector with components that may depend on a, a^* and c and θ a vector of

unknown regression parameters. In practice, $H(a, a^*, c)$ and $V(a, m, c)$ often corresponds to simple vector such as $(1, a, a^*, c)$ and $(1, a, m, c)$, respectively. Moreover, the vector $V(a, m, c)$ should at least contain all the elements of the vector $H(a, a^*, c)$ in with a^* replaced by m
Under identifiability conditions given in equations (4−11) given in Theoretical background section, it has been showed (e.g. in (15)) that

$$E[Y(a, M(a^*)) \mid C] = E\Big[E[Y \mid A = a, M, C] \Big| A = a^*, C\Big].$$

This gives arise to an imputation procedure whereby the observed data set is complemented with imputed data sets in which the same individuals are evaluated at different exposure levels, but corresponding to the observed mediator level (8). For simplicity, we will describe the estimation procedure for a binary exposure A.

1. Fit the imputation model in equation (E.5) to the observed data to obtain estimated $\hat{\beta}$.

2. Expand the observed data by repeating each observation in the original data set twice and including two additional variables: A_1, which is equal to the observed exposure A, and A_0, which is equal to the observed exposure for the first replication and equal to the opposite of the observed exposure, $1 - A$, for the second replication. The medflex package denotes the new variables A_0 and A_1 by using the name of the exposure from the original data set with indeces 0 and 1, respectively.

3. Impute the counterfactual outcomes, $Y(a, M(a^*))$, in the expanded data set as the expected values

$$E[Y \mid A = A_0, M = m, C = c; \hat{\beta}] = g_Y^{-1}\{\hat{\beta}^T V(A_0, m, c)\},$$

   using the the fitted outcome model in equation (E.5) with the exposure set to A_0 defined in the previous step and the mediator M and the covariates C set to their observed values.

4. The last step requires fitting the model in equation (E.6) to the imputed data set from the previous step by using $a = A_0$ and $a^* = A_1$. The estimates for natural direct and indirect effects are given by the certain components of the vector of estimated regression parameters $\hat{\theta}$. The standard errors can be obtained from a bootstrap procedure. Alternatively, robust standard errors can be used.

The estimation procedure is similar for continuous or categorical exposures with more than two possible levels. Furthermore, it can easily be extended to obtain estimates of the marginal natural effects. For futher details, see (8, 15).

F Inverse odds ratio weighted estimation

F.1 Summary

Tchetgen Tchetgen (10) suggested a simple yet general inverse odds ratio weighted estimation approach for making inferences about conditional natural effects. This approach involves inverse odds ratio weights that relate the exposure and mediator and requires specification of two regression models: a regression model for the exposure given the mediator and baseline covariates and a regression model for the outcome given the exposure and baseline covariates. In principle, inverse odds ratio estimation approach can be implemented with almost any type of variables and regression models, including generalized linear models (even those with nonlinear link functions), quantile regression or survival models, and can be implemented in any standard software that accomodates weighted regression (10). However, the available software solutions only include STATA and R code examples for limited settings (16, 17). Therefore, implementing inverse odds ratio estimation approach requires further programming from the user. IORW accomodates multiple mediators and interactions without actually having to specify the interactions (10, 16). Conducting sensitivity analysis requires further programming from the user.

[Table 9 about here.]

F.2 Parameter of interest

The inverse odds ratio weighted estimation estimates the conditional natural effects within the levels of the covariates C. In particular, for any value c of the covariates C, the conditional natural direct effect $NDE(a,a^*,c)$, natural indirect effect $NIE(a,a^*,c)$ and total effect $TE(a,a^*,c)$ are defined as

$$NDE(a,a^*,c) = g\{E[Y(a,M(a^*))\,|\,C=c]\} - g\{E[Y(a^*,M(a^*))\,|\,C=c]\} \quad \text{(F.1)}$$

$$NIE(a,a^*,c) = g\{E[Y(a,M(a^*))\,|\,C=c]\} - g\{E[Y(a,M(a^*))\,|\,C=c]\} \quad \text{(F.2)}$$

$$TE(a,a^*,c) = g\{E[Y(a,M(a))\,|\,C=c]\} - g\{E[Y(a^*,M(a^*))\,|\,C=c]\} \quad \text{(F.3)}$$

where g is some link function and a, a^* two exposure levels corresponding to being exposed and not exposed respectively. For example, the measure of natural direct

effect for a binary outcome can be expressed as an odds ratio

$$\begin{aligned}
NDE(a,a^*,c) &= \text{logit}\{E[Y(a,M(a^*))\,|\,C=c]\} - \text{logit}\{E[Y(a^*,M(a^*))\,|\,C=c]\} \\
&= \text{logit}\{P(Y(a,M(a^*))=1\,|\,C=c)\} - \text{logit}\{P(Y(a^*,M(a^*))=1\,|\,C=c)\} \\
&= \log\left\{\frac{P(Y(a,M(a^*))=1\,|\,C=c)}{1-P(Y(a,M(a^*))=1\,|\,C=c)}\right\} - \\
&\quad \log\left\{\frac{P(Y(a^*,M(a^*))=1\,|\,C=c)}{1-P(Y(a^*,M(a^*))=1\,|\,C=c)}\right\} \\
&= \log\left\{\frac{odds\big(P(Y(a,M(a^*))=1\,|\,C=c)\big)}{odds\big(P(Y(a^*,M(a^*))=1\,|\,C=c)\big)}\right\} \\
&= \log\{OR_{NDE}(a,a^*,c)\}
\end{aligned}$$

by using g = logit link function. Natural effects given in equations (F.1-F.3) are assumed to be the same for any chosen value of c (unless exposure-covariate interactions are present).

F.3 Estimation procedure - technical details

The inverse odds ratio weighted estimation relies on the fact that

$$E\left[\frac{f_M(m_0|A=a,C=c)}{f_M(m_0|A=a^*,C=c)}Y(a,M(a^*))\,\middle|\,C=c\right] = E[W\cdot Y\,|A=a,C=c], \qquad \text{(F.4)}$$

with

$$\begin{aligned}
W &= \frac{f_M(m_0\,|A=a,C=c)f_M(M\,|A=a^*,C=c)}{f_M(m_0\,|A=a^*,C=c)f_M(M\,|A=a,C=c)} \\
&= \frac{f_A(a^*\,|M,C=c)f_A(a\,|M=m_0,C=c)}{f_A(a\,|M,C=c)f_A(a^*\,|M=m_0,C=c)}
\end{aligned} \qquad \text{(F.5)}$$

where $f_M(\cdot|A,C)$ denotes the conditional density of the mediator given the exposure and the covariates, $f_A(\cdot|M,C)$ denotes the conditional density of the exposure given the mediator and the covariates and m_0 is some reference value of the mediator. When the mediator or exposure is categorical, the corresponding density is replaced by the probability. The property in equation (F.4) gives arise to the estimation procedure whereby the natural direct effect is estimated via a weighted regression model for the outcome given the exposure and covariates and using the inverse exposure-mediator odds ratio (given in equation (F.5)) as a weight. The total effect is estimated by fitting a regression model akin to the direct effect model, but omitting the weights. Finally, the difference between the total and direct effects on

a scale corresponding to the linear predictor is taken to compute the estimate of indirect effect. Because of the second representation of the IORW weights in equation (F.5), an exposure model rather than a mediator model can be fitted to estimate the IORW weights (10).
For simplicity we will describe the estimation procedure for a binary exposure. Consider the following regression models

$$g_A\{P(A=1\,|\,M=m,C=c;\beta)\}=\beta^T V(m,c) \tag{F.6}$$

and

$$g_Y\{E[Y\,|A=a,C=c]\}=\theta^T H(a,c) \tag{F.7}$$

where g_A, g_Y are some link functions, $V(m,c)$ and $H(a,c)$ are known vectors with components that may depend on m,c and a,c, respectively, and β,θ are unknown parameter vectors. In practice, $V(m,c)$ and $H(a,c)$ are often specified such that

$$\begin{aligned}\beta^T V(m,c)&=\beta_0+\beta_1 m+\beta_2^T c\\ \theta^T H(a,c)&=\theta_0+\theta_1 a+\theta_2^T c.\end{aligned}$$

In case of binary exposure, the IORW weights are given by

$$W=\frac{P(A=0\,|\,M=M_i,C=C_i)P(A=1\,|\,M=m_0,C=C_i)}{P(A=1\,|\,M=M_i,C=C_i)P(A=0\,|\,M=m_0,C=C_i)},$$

for each subject i in the exposed group $A_i=1$. For subjects in the unexposed group, $A_i=0$, the weights are equal to 1. The estimation procedure proceeds as follows.

1. Fit the exposure model in equation (F.6) to the observed data to get the estimates $\hat{\beta}$.

2. For each sucject i in the exposed group, compute the IORW weights

$$\begin{aligned}W_i&=\frac{P(A=0\,|\,M=M_i,C=C_i;\hat{\beta})P(A=1\,|\,M=m_0,C=C_i;\hat{\beta})}{P(A=1\,|\,M=M_i,C=C_i;\hat{\beta})P(A=0\,|\,M=m_0,C=C_i;\hat{\beta})}\\&=\frac{(1-g_A^{-1}\{\hat{\beta}^T V(M_i,C_i)\})}{g_A^{-1}\{\hat{\beta}^T V(M_i,C_i)\})}\times\frac{g_A^{-1}\{\hat{\beta}^T V(m_0,C_i)\}}{(1-g_A^{-1}\{\hat{\beta}^T V(m_0,C_i)\})}\end{aligned} \tag{F.8}$$

   by applying the fitted model from the first step to the observed data set twice, once using the observed mediator M_i and once using the reference value, say $m_0=0$. In most software packages, this can be done by using "predict-functionality". For subjects in the unexposed group, set the weight equal to

1.

3. Fit a suitable outcome model in equation (F.7) to the observed data set twice: once by weighting it with the weights from the previous step and once without the weights to obtain the estimates of natural direct and total effect, respectively.

4. Compute the estimate of the natural indirect effect by substracting the direct effect from the total effect.

5. The 95% confidence intervals for natural effects can be obtained by a bootstrap procedure.

The estimation procedure can be extended for continuous and categorical exposures (with more than two levels) as long as the reference level in the inverse odds ratio function remains a^*. In addition, more efficient estimation may be obtained by stabilizing the weights. For more details, see Tchetgen-Tchetgen (10) and Nguyen et al. (16).

F.4 Example

Although full software implementation of inverse odds ratio weighted estimation approach is not available, code examples of how to implement the inverse odds ratio estimation approach in software packages STATA (16) and R (17) are available. In this section, we illustrate how to implement the inverse odds ratio estimation approach in R by applying it to the Danish education study (20). The study investigates how much of the effect of physical fitness (A) on commencement in post-compulsory education (Y) is mediated through academic achievement measured by grade point average (M).

F.4.1 The data set

The data set consists of 1084 observations and the following variables.

`fitness` Physical fitness coded with 2 levels (0-low, 1-high).

`gpa` Grade point average of the compulsory exams at the end of compulsory education.

`postedu` Commencement in post-compulsory education (0-commenced, 1-not commenced).

`age` Age in years (13, 14, 15).

`ethni` Ethnicity (0-immigrants/descendants, 1-danes).

`income` Parental income in 4 groups (0-3) with 3 being the highest income group.

`educ` Parental education level in 4 groups (0-3) with 3 being the highest.

It is assumed that age, ethnicity, parental income and education level are sufficient to control for confounding. For more details about the data, see (20).

F.4.2 Model specification

The estimation procedure requires specification of two models. We will consider the following logistic regression model for the outcome

$$\text{logit}\{P(Y=1\,|A=a,C=c)\}=\theta_0+\theta_1 a+\theta_2^T c. \tag{F.9}$$

and another logistic regression model for the exposure

$$\text{logit}\{P(A=1\,|M=m,C=c)\}=\beta_0+\beta_1 m+\beta_2^T c. \tag{F.10}$$

Using a logistic regression for the outcome results in estimating the conditional natural effects as log odds ratios.

F.4.3 Implementation

Step 1. Before conducting mediation analysis, the user must provide the data set. Since grade point average takes values between 0.4 and 12, we use a reference value $m_0=7$. The first step of the estimation algorithm, fitting the exposure model for the observed data, is done using the `glm` function.

```
d1 <- read.csv("d.csv")
Afit <- glm(fitness~gpa+age+ethni+income+educ,data=d1,
    family="binomial")
```

Step 2. Next, we construct a new data set by including everything in the observed data set and changing the mediator variable to the reference value $m_0=7$ for all subjects. The inverse odds ratio weights for the subjects in the exposed group are computed by applying the `predict` function twice, once to the observed data set and once to the new data set. The inverse odds ratio weights are set to 1 for subjects in the unexposed group.

```
newdata <- d1
newData$gpa <- 7
p1 <- predict(Afit,newdata=d1,type="response")
p2 <- predict(Afit,newdata=newData,type="response")
W <-((1-p1)*p2)/(p1*(1-p2))
W[d1$fitness==0] <- 1
```

Note that more efficient estimation can be obtained by using stabilized weights. Weights can be stabilized by multiplying each individual's inverse odds ratio weight by the inverse of the predicted odds of the exposure, where the mediator is evaluated at its refrence value. The resulting inverse odds weights can then be used insted of inverse odds ratio weights.

```
W2 <- (1-p1)/p1
W2[d1$fitness==0] <- 1
```

Furthermore, if the mediator is centered around it's mean (e.g. using `scale` function) and the refrence value is $m_0 = 0$, then the inverse odds ratio weights can be computed using the alternative code:

```
d1$gpa2 <- scale(d1$gpa)
Afit <- glm(fitness~gpa2+age+ethni+income+educ,data=d1,
    family="binomial")
W <-1/(exp(coef(Afit)[2]*d1$gpa2))
W[d1$fitness==0] <- 1
```

Step 3. To get the estimates of the natural direct effect we use the `glm` function and set the `weights` argument equal to the inverse odds ratio weights computed in the previous step. To estimate the total effect we use the same regression model, but omit the weights.

```
wgtYfit <- glm(postedu~fitness+age+ethni+income+educ,data=d1,weights=W,
       family="binomial")
Yfit <- glm(postedu~fitness+age+ethni+income+educ,data=d1,
       family="binomial")
```

Step 4. The estimated natural direct effect and total effect given as log odds ratios are equal to the regression coefficient for the exposure in the weighted regresson model and standard regression model, respectively. The estimated natural indirect effect (on the log-odds scale) is equal to the difference between the estimated total and estimated natural direct effect (on the log-odds scale).

```
Est <- c(Total=coef(Yfit)[2],Direct=coef(wgtYfit)[2],Indirect=coef(Yfit
    )[2]-coef(wgtYfit)[2])
Est
```

```
        Total           Direct         Indirect
     -0.95125         -0.67865          -0.2726
```

For instance, the natural direct effect estimate -0.679 suggests that if the physical fitness level of all the students was changed from low to high without changing their grade point average, the odds of not commencing in post-compulsory education would be decreased $1/\exp(-0.679) = 1/0.507 = 1.97$ times.

Step 5. To obtain the 95% confidence intervals we use a bootstrap procedure. This is done by repeating the estimation procedure (function `getIORWEst`) on a large num-

ber of random samples with replacement from the observed data. The confidence intervals are obtained by taking the suitable quantiles from the bootstrap estimates.

```
getIORWEst <- function(data){
    Afit <- glm(fitness~gpa+ethni+age+income+educ,data=data,
        family="binomial")
    newData <- data
    newData$gpa<- 7
    p1 <- predict(Afit,newdata=data,type="response")
    p2 <- predict(Afit,newdata=newData,type="response")
    W <- ((1-p1)*p2)/(p1*(1-p2))
    W[bootData$fitness==0] <- 1
    wgtYfit <- glm(postedu~fitness+ethni+age+income+educ,weights=W,data
      =data
        ,family="binomial")
    Yfit <- glm(postedu~fitness+ethni+age+income+educ,data=data,
        family="binomial")
    Est <- c(coef(Yfit)[2],coef(wgtYfit)[2],
        coef(Yfit)[2]-coef(wgtYfit)[2])
    return(Est)
}
N <- 1000
bootSample <- data.frame(matrix(NA,ncol=3,nrow=N))
for (i in (1:N)){
    inx <- sample(n,replace=TRUE)
    bootData <- d1[inx, ]
    bootSample[i,] <- getIORWEst(bootData)
}
CI <- apply(bootSample,2,function(k)quantile(k,c(0.025,0.975)))
Estimates <- t(rbind(Est,CI))
rownames(Estimates) <- c("Total effect","Direct effect","Indirect
      effect")
colnames(Estimates) <- c("Estimate","2.5% CI","97.5% CI")
Estimates
```

This gives the following output

```
                 Estimate    2.5% CI   97.5% CI
Total effect     -0.95125   -1.53547    -0.46503
Direct effect    -0.67865   -1.28391    -0.19098
Indirect effect  -0.2726    -0.47037    -0.11482
```

References

1. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992. 3(2):143–155.

2. Pearl J. Direct and indirect effects. In *Proceedings of the seventeenth conference on uncertainty in artificial intelligence*, pages 411–420. Morgan Kaufmann Publishers Inc., 2001.

3. Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*. 2010. pages 51–71.

4. VanderWeele T, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface*. 2009. 2:457–468.

5. VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *American journal of epidemiology*. 2010. 172(12):1339–1348.

6. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*. 2009. 20(1):18–26.

7. Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. *American journal of epidemiology*. 2012. 176(3):190–195.

8. Vansteelandt S, Bekaert M, Lange T. Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiologic Methods*. 2012. 1(1):131–158.

9. Tchetgen Tchetgen EJ, Shpitser I. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Annals of statistics*. 2012. 40(3):1816.

10. Tchetgen Tchetgen EJ. Inverse odds ratio-weighted estimation for causal mediation analysis. *Statistics in medicine*. 2013. 32(26):4567–4580.

11. Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure–mediator interactions and causal interpretation: Theoretical assumptions and implementation with sas and spss macros. *Psychological methods*. 2013. 18(2):137.

12. Valeri L, VanderWeele TJ. Sas macro for causal mediation analysis with survival data. *Epidemiology*. 2015. 26(2):e23–e24.

13. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychological methods*. 2010. 15(4):309.

14. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R package for causal mediation analysis. *Journal of Statistical Software*. 2014. 59(5):1–38.

15. Steen J, Loeys T, Moerkerke B, Vansteelandt S. *medflex: Flexible mediation analysis using natural effect models*, 2015. R package version 0.5-0.

16. Nguyen QC, Osypuk TL, Schmidt NM, Glymour MM, Tchetgen EJT. Practical guidance for conducting mediation analysis with multiple mediators using inverse odds ratio weighting. *American journal of epidemiology*. 2015. page kwu278.

17. Nguyen TT, Tchetgen E, Kawachi I, Gilman SE, Walter S, Glymour M. Comparing alternative effect decomposition methods: the role of literacy in mediating educational effects on mortality. *Epidemiology (Cambridge, Mass.)*. 2016.

18. Hafeman DM, Schwartz S. Opening the black box: a motivation for the assessment of mediation. *International Journal of Epidemiology*. 2009. page dyn372.

19. Robins JM, Richardson TS. Alternative graphical causal models and the identification of direct effects. *Causality and psychopathology: Finding the determinants of disorders and their cures*. 2010. pages 103–158.

20. Porsborg Andersen M, Starkopf L, Nørmark Mortensen R, Lange T, Torp-Pedersen C. The indirect and direct pathways between physical fitness and academic achievements on post-compulsory education commencement in a retrospective cohort of danish school youth. *Unpublished*.

21. Wacholder S. Binomial regression in glim: estimating risk ratios and risk differences. *American journal of epidemiology*. 1986. 123(1):174.

22. Deddens JA, Petersen MR, Lei X. Estimation of prevalence ratios when proc genmod does not converge. In *Proceedings of the 28th annual SAS users group international conference*, volume 30, pages 270–28, 2003.

23. Lok JJ. Defining and estimating causal direct and indirect effects when setting the mediator to specific values is not feasible. *Statistics in medicine*. 2016.

24. Pearl J. The causal mediation formula—a guide to the assessment of pathways and mechanisms. *Prevention Science*. 2012. 13(4):426–436.

25. Imai K, Yamamoto T. Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. *Political Analysis*. 2013. 21(2):141–171.

26. King G, Tomz M, Wittenberg J. Making the most of statistical analyses: Improving interpretation and presentation. *American journal of political science*. 2000. pages 347–361.

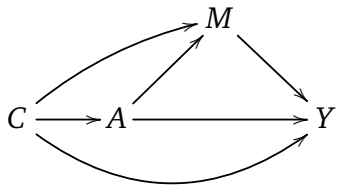

Figure 1: Generic causal structure in mediation analysis, where A denotes the exposure, M the mediator, Y the oucome and C a set of baseline confounders.

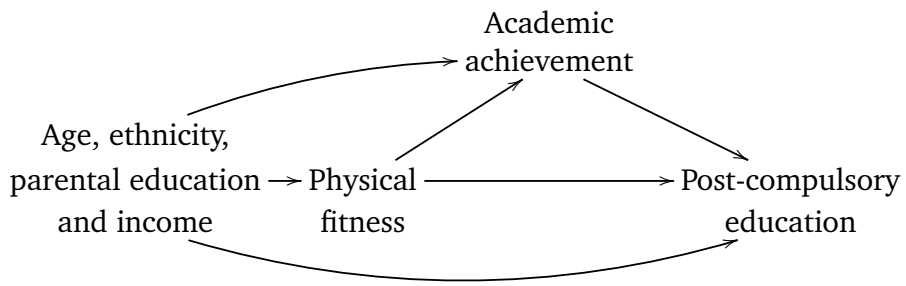

Figure 2: Causal structure of Danish education data.

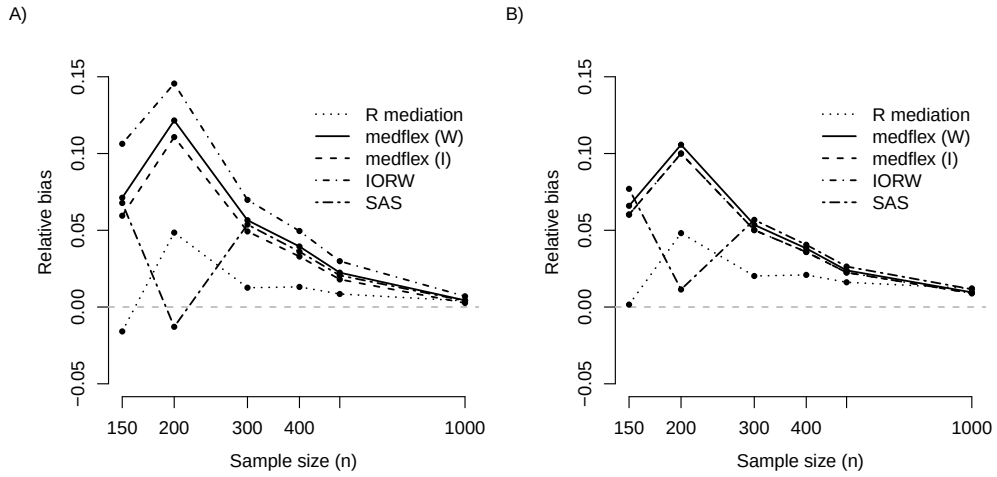

Figure 3: Relative bias for natural direct (panel A)) and total (panel B)) effect from 2000 simulation runs as a function of sample size.

Table 1: Comparison of the Five Estimation Methods and Their Software Solutions.

Variable	SAS/SPSS	medflex (W)	medflex (I)	mediation	inverse odds ratio estimation approach
Type of variables					
Exposure	Continuous Binary Polytomous	Continuous Binary Polytomous	Continuous Binary Polytomous	Continuous Binary Polytomous	Continuous Binary Polytomous
Mediator	Continuous Binary	Continuous Binary Count Polytomous	Continuous Binary Count Polytomous Multidimensional	Continuous Binary Count Polytomous Failure time Multidimensional	Continuous Binary Count Polytomous Failure time Multidimensional
Outcome	Continuous Binary Count	Continuous Binary Count	Continuous Binary Count	Continuous Binary Count Polytomous Failure time	Continuous Binary Count Polytomous Failure time
Parameters of interest					
Marginal or conditional	Conditional natural effects[a]	Conditional natural effects[b] Marginal natural effects	Conditional natural effects[b] Marginal natural effects	Marginal natural effects	Conditional natural effects[b]
Scale	Corresponds to g	Corresponds to g	Corresponds to g	Always difference, i.e. $g = identity$	Corresponds to g
Modelling					
Required models	$M \mid A, C$ $Y \mid A, M, C$	$M \mid A, C$ $Y(a, M(a^*)) \mid C$	$Y \mid A, M, C$ $Y(a, M(a^*)) \mid C$	$M \mid A, C$ $Y \mid A, M, C$	$A \mid M, C$ $Y \mid A, C$
Model flexibility	Main effects $A \times M$ interaction	All regression models[c]	All regression models[c]	All regression models[c]	All regression models[c]
Sensitivity analysis	Requires further coding	Requires further coding	Requires further coding	For confounders between M and Y	Requires further coding
Standard errors	Robust Bootstrap	Robust Bootstrap	Robust Bootstrap	Robust Bootstrap	Robust (not implemented) Bootstrap

Abbreviations: I, imputation method in medflex; IORW, inverse odds ratio weighted; W, weighting method in medflex package.

[a] Parameters can vary across different levels of the confounders.

[b] Parameters are assumed to be the same for any level of the confounders (unless exposure-confounders interactions are present) .

[c] Models including main effects, all sorts of interaction terms, splines etc. Further details in eAppendices.

Table 2: Estimates of (pure) natural direct, (total) natural indirect and total effect from the Danish education data. The estimates of natural effects are given as odds ratios, except for R mediation package for which risk differences are given instead.

Effect	Method	Estimate	95% CI
Direct effect	R mediation	-0.048	-0.084, -0.009
	R medflex (W)	0.515	0.287, 0.856
	R medflex (I)	0.523	0.296, 0.878
	IORW estimation	0.507	0.277, 0.826
	SAS mediation	0.498	0.292, 0.850
Indirect effect	R mediation	-0.014	-0.024, -0.007
	R medflex (W)	0.749	0.649, 0.846
	R medflex (I)	0.743	0.621, 0.862
	IORW estimation	0.761	0.625, 0.892
	SAS mediation	0.745	0.650, 0.854
Total effect	R mediation	-0.062	-0.096, -0.027
	R medflex (W)	0.385	0.214, 0.654
	R medflex (I)	0.388	0.217, 0.653
	IORW estimation	0.386	0.215, 0.628
	SAS mediation	0.371	0.216, 0.639

Abbreviations: CI, confidence interval, I, imputation method in medflex; IORW, inverse odds ratio weighted; W, weighting method in medflex package.

Table 3: Simulation results for natural effects in the setting with common binary outcome: relative bias, relative RMSE and coverage of 95% bootstrap-based confidence intervals from 2000 simulations.

Effect	Method	Rel.Bias	Rel. RMSE	Cov.P of 95%CI
Direct effect	R mediation	0.0179	0.491	0.955
	medflex (W)	0.0510	0.529	0.947
	medflex (I)	0.0451	0.523	0.951
	IORW estimation	0.0510	0.557	0.948
	SAS macro	-0.2232	0.486	0.909
Indirect effect	R mediation	-0.0399	0.589	0.941
	medflex (W)	-0.0412	0.616	0.944
	medflex (I)	-0.0181	0.601	0.956
	IORW estimation	-0.0400	0.910	0.923
	SAS macro	-0.4023	0.530	0.817
Total effect	R mediation	0.0079	0.403	0.959
	medflex (W)	0.0348	0.432	0.954
	medflex (I)	0.0342	0.430	0.955
	IORW estimation	0.0342	0.430	0.956
	SAS macro	-0.2536	0.432	0.889

Abbreviations: CI, confidence interval; Cov.P, coverage probability; I, imputation method in medflex; IORW, inverse odds ratio weighted; Rel.Bias, relative bias; Rel.RMSE, relative root mean squared error; W, weighting method in medflex package.

Table 4: Simulation results for natural effects in the setting with rare binary outcome: relative bias, relative RMSE and coverage of 95% bootstrap-based confidence intervals from 2000 simulations.

Effect	Method	Rel.Bias	Rel. RMSE	Cov.P of 95%CI
Direct effect	R mediation	0.0477	0.609	0.961
	medflex (W)	0.1208	0.728	0.944
	medflex (I)	0.1100	0.705	0.950
	IORW estimation	0.1447	0.804	0.931
	SAS macro	-0.0147	0.731	0.946
Indirect effect	R mediation	0.0478	0.750	0.954
	medflex (W)	0.0270	0.744	0.955
	medflex (I)	0.0444	0.775	0.958
	IORW estimation	-0.1046	1.374	0.923
	SAS macro	0.0688	0.799	0.957
Total effect	R mediation	0.0477	0.515	0.959
	medflex (W)	0.1054	0.607	0.945
	medflex (I)	0.0997	0.594	0.945
	IORW estimation	0.0996	0.594	0.945
	SAS macro	0.0109	0.613	0.946

Abbreviations: CI, confidence interval; Cov.P, coverage probability; I, imputation method in medflex; IORW, inverse odds ratio weighted; Rel.Bias, relative bias; Rel.RMSE, relative root mean squared error; W, weighting method in medflex package.

Table 5: Scenarios implemented in SAS and SPSS macros.

Variable	Type of variable	Regression model
Exposure	Continuous	Not necessary
	Binary	
	Polytomous	
Mediator	Continuous	Linear
	Binary	Logistic
Outcome	Continuous	Linear
	Binary	Logistic
		Log-linear
	Count	Poisson
		Negative binomial
	Failure time	Cox proportional hazard
		Accelerated failure time

Table 6: Types of variables, models and distribution families that can be used with mediation package.

Variable	Type of variable	Regression model	Family
Exposure	Continuous Binary Polytomous	Not necessary	
Mediator	Continuous	Linear (`lm`, `lmer`)	-

		GLM (`glm`, `bayesglm`)	`gaussian`
			`Gamma`
			`inverse.gaussian`

		Quantile (`rq`)	-

		GAM (`gam`)	`gaussian`
			`Gamma`
			`inverse.gaussian`

	Binary/Count	GLM (`glm`, `glmer`, `bayesglm`)	`binomial/poisson`
	Polytomous	Ordered logistic (`polr`) Ordered probit (`polr`)	-
	Failure time	Parametric survival (`survreg`)	All available distributions for `survreg`
Outcome	Continuous	Linear (`lm`, `lmer`)	-
		GLM (`glm`, `bayesglm`)	All families available for `glm`, `bayesglm`
		Quantile (`rq`)	-
		GAM (`gam`)	All families available for `gam`
		Censored (`vglm`)	`tobit`
	Binary/Count	GLM (`glm`, `glmer`, `bayesglm`)	`binomial/poisson`
	Polytomous	Ordered logistic (`polr`) Ordered probit (`polr`)	-
	Failure time	Parametric survival (`survreg`)	All available distributions for `survreg`

Table 7: Combination of variables and models available in medflex package for weighting approach.

Variable	Type of Variable	Regression model	Family
Exposure	Continuous Binary Polytomous	Not necessary	

Mediator	Continuous
	Binary
	Count
		GLM (`glm` or `vglm`)	`gaussian`
			`binomial`
			`poisson`

	Polytomous	Multinomial logit (`vglm`)	`multinomial`
Outcome	Continuous Binary Count	GLM (`glm`)	All suitable options available for `glm`

Table 8: Combination of variables and models available in medflex package for imputation approach.

Variable	Type of Variable	Regression model
Exposure	Continuous Binary Polytomous	Not necessary
Mediator	Continuous Binary Count Polytomous Failure time	Not necessary
Outcome	Continuous Binary Count	GLM (`glm`)

Table 9: Combination of variables and models that can be used for inverse odds ratio approach. Implementation requires additional coding.

Variable	Type of Variable	Regression model
Exposure	Continuous Binary Polytomous	Any kind of model
Mediator	Continuous Binary Count Polytomous Failure time	Not necessary
Outcome	Continuous Binary Count Polytomous Failure time	Any kind of regression model that accomodates weights

Research Reports available from Department of Biostatistics

http://www.pubhealth.ku.dk/bs/publikationer

Department of Biostatistics
University of Copenhagen
Øster Farimagsgade 5
P.O. Box 2099
1014 Copenhagen K
Denmark

14/01 Ambrogi, F. & Andersen, P.K. Predicting Smooth Survival Curves through Pseudo-Values.

14/02 Olsbjerg, M. & Christensen, K.B. Modeling local dependence in longitudinal IRT models.

14/03 Olsbjerg, M. & Christensen, K.B. LIRT: SAS macros for longitudinal IRT models.

14/04 Jacobsen, M. & Martinussen, T. A note on the large sample properties of estimators based on generalized linear models for correlated pseudo-observations.

14/05 De Neve, J. & Gerds, T. A note on the interpretation of the Cox regression model.

14/06 Gerds, TA. The Kaplan-Meier theater.

14/07 Martinussen, T., Holst, K.K. & Scheike, T. Cox regression with missing covariate data using a modified partial likelihood method.

15/01 Mansourvar, Z., Martinussen, T. & Scheike, T. Semiparametric regression for restricted mean residual life under right censoring.

15/02 Mansourvar, Z., Martinussen, T. & Scheike, T. An additive-multiplicative restricted mean residual life model.

15/03 Mansourvar, Z. & Martinussen, T. Estimation of average causal effect using the restricted mean residual lifetime as effect measure.

15/04 Ekstrøm, C.E., Gerds, T.A., Jensen, A.K. & Brink-Jensen, K. Sequential rank agreement methods for comparison of ranked lists.

15/05 Christensen, K.B., Makransky, G. & Horton, M. Critical Values for Yen's *Q3*: Identification of Local Dependence in the Rasch model using Residual Correlations.

15/06 Müller, M. & Kreiner, S. Item Fit Statistics in Common Software for Rasch Analysis.

16/01 Andersen, P.K. Life years lost among patients with a given disease.

16/02 Strohmaier, S., Haase, N., Wetterslev, J. & Lange, T. A simple to implement algorithm for natural direct and indirect effects in survival studies with a repeatedly measured mediator.

16/03 Andersen, P.K., Syriopoulou, E. & Parner, E. Causal inference in survival analysis using pseudo-observations.

16/04 Paul Blanche, Michael W. Kattan & Thomas A. Gerds. The c-index is not proper for the evaluation of *t*-year predicted risks.

17/01 Liis Starkopf, Mikkel Porsborg Andersen, Thomas Alexander Gerds, Christian Torp-Pedersen, Theis Lange. Comparison of five software solutions to mediation analysis.