Epidemiologic Methods

Volume 1, *Issue* 1 2012 *Article* 7

Imputation Strategies for the Estimation of Natural Direct and Indirect Effects

Stijn Vansteelandt, *Ghent University*
Maarten Bekaert, *Ghent University*
Theis Lange, *University of Copenhagen*

Recommended Citation:

Vansteelandt, Stijn; Bekaert, Maarten; and Lange, Theis (2012) "Imputation Strategies for the Estimation of Natural Direct and Indirect Effects," *Epidemiologic Methods*: Vol. 1: Iss. 1, Article 7.
DOI: 10.1515/2161-962X.1014

Imputation Strategies for the Estimation of Natural Direct and Indirect Effects

Stijn Vansteelandt, Maarten Bekaert, and Theis Lange

Abstract

Mediation analysis is widely adopted to infer causal mechanism by disentangling indirect or mediated effects of an exposure on an outcome through given intermediaries, from the remaining direct effect. Traditional approaches build on standard regression models for the outcome and mediator, but easily result in difficult-to-interpret or difficult-to-report results when some of these models involve non-linearities. In this article, we overcome this via a general class of so-called natural effect models, which directly parameterize the (natural) direct and indirect effects of interest. We propose flexible estimation strategies for the direct and indirect effect parameters indexing these models, that are easy to perform with standard statistical software: one based on regression mean imputation and one based on doubly robust imputation. We give a theoretical discussion of the properties of these estimation strategies. We moreover assess their finite-sample performance through a simulation study, and through the analysis of the WHO-LARES study on the association between residence in a damp and moldy dwelling and the risk of depression.

KEYWORDS: causal inference, direct effect, indirect effect, mediation, imputation

Author Notes: We thank the WHO European Centre for Environment and Health, Bonn office, for providing the WHO LARES data set which was used to illustrate the methods. The paper reflects the authors' opinion and not necessarily the position of the WHO. We are also grateful to Eric Tchetgen Tchetgen and Tyler VanderWeele for helpful discussions, and to 2 reviewers for helpful comments. Stijn Vansteelandt was supported by Interuniversity Attraction Poles (IAP) research network grant P06/03 from the Belgian government (Belgian Science Policy). Maarten Bekaert acknowledges support from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

1 Introduction

Researchers in a variety of scientific fields, notably epidemiologists, have a long-standing interest in using empirical studies to disentangle causal pathways by which an exposure or treatment affects an outcome. Most popular is the use of mediation analysis (MacKinnon, 2008), whereby an exposure's effect is decomposed into an indirect effect mediated by a given intermediate variable and the remaining direct effect. Prevailing methods for mediation analysis have their roots in work by Baron and Kenny (1986), who focused on linear models for the outcome and mediator. Baron and Kenny (1986) proposed estimating the direct effect as the residual association between outcome and exposure after regression adjustment for the mediator(s). They further proposed estimating the indirect effect as the product of the exposure's effect on the mediator and the mediator's effect on the outcome. In linear models, this is equivalent to calculating the difference between the total effect and the direct effect, where the former refers to the (unadjusted) association between outcome and exposure. Both these approaches have subsequently been employed in a variety of statistical models other than the linear model and are typically referred to as 'product of coefficient methods' and 'difference in coefficient methods' for indirect effects. For instance, let A be the observed exposure of interest, M a continuous mediator, C a set of baseline covariates and Y a dichotomous outcome. Suppose moreover that the outcome and mediator obey the following models

$$\begin{aligned} \text{logit}P(Y=1|A,M,C) &= \beta_0+\beta_a A+\beta_m M+\beta_c C \quad (1)\\ E(M|A,C) &= \alpha_0+\alpha_a A+\alpha_c C. \quad (2)\end{aligned}$$

The product of coefficient method then amounts to calculating the indirect effect as $\beta_m\alpha_a$. An advantage of this approach is that when C is sufficient to control for confounding of the mediator-outcome and exposure-mediator associations, then it will yield a zero indirect effect whenever the outcome is not affected by the mediator or the mediator is not affected by the exposure, and thus it can be of some use for testing the null hypothesis of no indirect effect. However, a further theoretical basis beyond the linear model is lacking (see e.g. Pearl (2011)) and, moreover, the precise interpretation of the product $\beta_m\alpha_a$ is unclear. This difficulty of interpretation prevails in much of the traditional mediation analysis literature, where the effects are often expressed on the scale of underlying latent continuous outcomes, rather than on the scale of the observed data. The difference in coefficient method amounts to fitting an additional model for the outcome, which does not involve the mediator:

$$\text{logit}P(Y=1|A,C) = \beta_0'+\beta_a' A+\beta_c' C \quad (3)$$

and then calculating the indirect effect as the difference between β_a', which is commonly interpreted as the total effect, and β_a, which is commonly interpreted as the direct effect. A major drawback of this approach is that, as a result of non-collapsibility of the odds ratio (Greenland, Robins, and Pearl, 1999), β_a' will typically differ from β_a, even when the exposure does not affect the mediator and thus even when there is no indirect effect.

Robins and Greenland (1992) revolutionized mediation analysis by proposing a generic framework to decompose a total effect into a so-called pure or natural direct and indirect effect, without being tied to a particular statistical model. Based on this framework, use of the so-called mediation formula (Pearl, 2001, 2011) now enables combining arbitrary statistical models for the outcome and mediator to obtain valid and well-understood measures of direct and indirect effect, provided that appropriate identification conditions hold (see e.g. VanderWeele and Vansteelandt (2009), VanderWeele and Vansteelandt (2010), Imai, Keele, and Tingley (2010a), Lange and Hansen (2011), VanderWeele (2011)). The mediation formula suggests a generic approach for mediation analysis, but has a number of important limitations. First, the way to compute natural direct and indirect effects involves integration; its complexity can differ substantially depending on the model for the mediator and outcome. Second, even simple models for the mediator and outcome (e.g. a linear model for the mediator and logistic regression model for the outcome) tend to produce complex expressions of natural direct and indirect effects. This can make results difficult to report. In addition, it can make interesting hypotheses essentially impossible to test; for instance, it may not be obvious to test whether the magnitude of the direct effect varies depending on covariates, as it can be difficult to choose models for the outcome and mediator that result in a constant natural direct effect (differing from zero) at all covariate levels. Imai et al. (2010a) overcome the first concern by using Monte-Carlo integration methods, but not the second. van der Laan and Petersen (2008) accommodated both concerns by directly modeling the natural direct effect of interest. More precisely, because application of the mediation formula to even simple models for the mediator and outcome typically results in complicated models for the natural direct and indirect effect, they instead choose to a priori postulate a parsimonious model structure for the natural direct effect. Results thereby become simpler for reporting and interesting hypotheses concerning these effects become easier to test.

van der Laan and Petersen (2008) proposed so-called doubly robust estimators for the direct effect parameters indexing their model. These require correct specification of a model for the distribution of the mediator (given exposure and confounders), and for either the distribution of the exposure (given confounders) or the expected outcome (given mediator, exposure and confounders). Tchetgen Tchetgen and Shpitser (2011, 2012) focused additionally on indirect effects and pro-

posed estimators with greater robustness against model misspecification, which require an arbitrary 2 out of the 3 models for the exposure, mediator and outcome to be correctly specified. All these approaches are very elegant and appealing; arguably, their relative complexity may be a barrier to routine application.

The goal of this article is to enhance the accessibility of methods for causal mediation analysis based on models for natural direct and indirect effects. We will attempt this by widening the scope of models (and hence outcome types) which can be dealt with, and by proposing flexible estimation strategies for natural direct and indirect effects that are easy to perform with standard statistical software. In particular, we propose a simple imputation approach which relies on correct specification of a model for the outcome. In addition, we propose a doubly robust imputation approach, which relaxes the assumptions of the imputation estimator by continuing to be valid when the outcome model is misspecified, provided that models for both mediator and exposure are correctly specified. This doubly robust imputation approach can be viewed as a compromise between the simple imputation approach and an inverse probability weighting approach previously considered by Hong (2010) and Lange, Vansteelandt, and Bekaert (2012). We give a theoretical discussion of the properties of these estimators and assess their finite-sample performance through a simulation study and through the analysis of the WHO LARES study on the association between residence in a damp and moldy dwelling and the risk of depression.

2 Natural direct and indirect effects

As in Robins and Greenland (1992) and Pearl (2001) we will describe direct and indirect effects in terms of so-called nested counterfactuals, $Y(a,M(a^*))$, denoting the outcome that would have been observed if A were set to a and M to the value it would have taken if A were set to a^*. In particular, we will compare $Y(a,M(a^*))$ with $Y(a^*,M(a^*))$ to obtain a measure of the direct effect of changing the exposure from a to a^*. Such comparison can for instance be made in terms of an average difference within levels of covariates, $E[Y(a,M(a^*))-Y(a^*,M(a^*))|C]$, a risk ratio, $P[Y(a,M(a^*))=1|C]/P[Y(a^*,M(a^*))=1|C]$, etc. We will refer to such contrasts as natural direct effects. Likewise we will compare $Y(a^*,M(a))$ with $Y(a^*,M(a^*))$ to obtain a measure of the indirect effect, referred to as a natural indirect effect.

Throughout the article, we will assume that the causal assumptions underlying the causal diagram of Figure 1 hold. In particular, we will assume that the same set of covariates C is sufficient to control for confounding of the associations between exposure and outcome, exposure and mediator, and mediator and outcome. Specifically, we thus assume that there are no variables that are effects of exposure

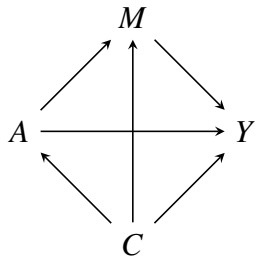

Figure 1: Causal diagram.

and that confound the mediator-outcome relationship. Formal descriptions of these assumptions in terms of counterfactuals are for instance given in VanderWeele and Vansteelandt (2009).

Under these assumptions, it follows from Pearl (2001, 2011) and Imai, Keele, and Yamamoto (2010b) that $E[Y(a,M(a^*))|C=c]$ can be calculated using the mediation formula:

$$\sum_m E(Y|A=a,M=m,C=c)P(M=m|A=a^*,C=c); \tag{4}$$

here, the sum can be substituted with an integral (integrating w.r.t. m) when M is continuous. The mediation formula advocates a form of standardization, which amounts to estimating the mean value of the outcome in each stratum defined by mediator and confounders among the individuals with exposure a, but weighting these by the likelihood of each mediator value among individuals with exposure a^*. On the basis of this, natural direct and indirect effects can be obtained. For instance, if the outcome Y and mediator M obey linear models, i.e. $E[Y|A=a,M=m,C=c]=\alpha_0+\alpha_1 a+\alpha_2 m+\alpha_3 c$ and $E[M|A=a,C=c]=\beta_0+\beta_1 a+\beta_2 c$, then equation (4) greatly simplifies and the natural direct and indirect effects are captured by $E[Y(a,M(a^*))-Y(a^*,M(a^*))|C]=\alpha_1(a-a^*)$ and $E[Y(a^*,M(a))-Y(a^*,M(a^*))|C]=\alpha_2\beta_1(a-a^*)$ (VanderWeele and Vansteelandt, 2009), which is in line with the traditional approach. However, the chief advantage of the mediation formula is that, in contrast to the traditional approach, it can also be used for non-linear models. For instance, when models (1) and (2) both hold, the mediator is normal with constant variance and the outcome mean is close to zero, VanderWeele and Vansteelandt (2010) find that the natural direct effect odds ratio

$$\frac{\text{odds}\{Y(a,M(a^*))=1|C\}}{\text{odds}\{Y(a^*,M(a^*))=1|C\}} \approx \exp(\beta_a)$$

and that the natural indirect effect odds ratio

$$\frac{\text{odds}\{Y(a^*,M(a^*))=1|C\}}{\text{odds}\{Y(a,M(a^*))=1|C\}} \approx \exp(\beta_m\alpha_a),$$

in line with the traditional (product of coefficients) approach. However, these approximations may become poor when the outcome mean is not close to zero or the mediator is non-normal. For instance, when M is dichotomous and obeys

$$\text{logit}P(M=1|A,C)=\alpha_0+\alpha_a A+\alpha_c C,$$

then application of the mediation formula shows that

$$\begin{aligned} E[Y(a,M(a^*))|C] &= \text{expit}\left(\beta_0+\beta_a a+\beta_m+\beta_c C\right)\text{expit}\left(\alpha_0+\alpha_a a^*+\alpha_c C\right) \\ &\quad +\text{expit}\left(\beta_0+\beta_a a+\beta_c C\right)\left\{1-\text{expit}\left(\alpha_0+\alpha_a a^*+\alpha_c C\right)\right\}. \end{aligned}$$

This entails natural direct and indirect effect odds ratios which, like $E[Y(a,M(a^*))|C]$, depend on exposures a and a^* and on covariates in a complicated way. This makes results difficult to report as it would require one to report separate results at each exposure value or covariate value. Moreover, it makes it essentially impossible to test whether the natural direct or indirect effect depends on covariate values, because it is essentially impossible to find plausible models for the outcome and mediator that combine into natural direct or indirect effects not depending on covariate values.

3 Simultaneous estimation of natural direct and indirect effects

3.1 Natural effect models

In this paper, as in Lange et al. (2012), we focus on conditional mean models for nested counterfactuals $Y(a,M(a^*))$:

$$g\left[E\left\{Y(a,M(a^*))|C\right\}\right]=\beta' W(a,a^*,C),$$

where $g(.)$ is a known canonical link function (e.g. the identity or logit link) and $W(a,a^*,C)$ is a known vector with components that may depend on a,a^* and C. We term such models 'natural effect models' because they encode the natural direct and indirect effects of interest. For instance, in model

$$E[Y(a,M(a^*))|C]=\beta_0+\beta_1 a+\beta_2 a^*+\beta_3 C, \tag{5}$$

$\beta_1(a-a^*)$ captures a natural direct effect and $\beta_2(a-a^*)$ captures a natural indirect effect. In model

$$\text{logit}\left[E\left\{Y(a,M(a^*))|C\right\}\right]=\beta_0+\beta_1 a+\beta_2 a^*+\beta_3 C+\beta_4 a\cdot a^*+\beta_5 a\cdot C+\beta_6 a^*\cdot C, \tag{6}$$

we find that the natural direct effect odds ratio

$$\frac{\text{odds}\{Y(a,M(a^*))=1|C\}}{\text{odds}\{Y(a^*,M(a^*))=1|C\}} = \exp\{(\beta_1+\beta_4 a^*+\beta_5 C)(a-a^*)\}$$

and that the natural indirect effect odds ratio

$$\frac{\text{odds}\{Y(a,M(a))=1|C\}}{\text{odds}\{Y(a,M(a^*))=1|C\}} = \exp\{(\beta_2+\beta_4 a+\beta_6 C)(a-a^*)\}.$$

Their product measures the total effect: $\text{odds}\{Y(a)=1|C\}/\text{odds}\{Y(a^*)=1|C\}$ (VanderWeele and Vansteelandt, 2010).

The considered class of natural effect models is more restrictive than the classes of models considered by van der Laan and Petersen (2008) and Tchetgen Tchetgen and Shpitser (2011), who specify only the dependence of $E[Y(a,M(a^*))|C]$ on either a or a^*. However, these added restrictions bring the practical advantage of enabling simultaneous inference for natural direct and indirect effects, as well as extra flexibility by not being limited to the identity and log links. The considered models thus in particular enable the analysis of dichotomous outcomes, as in the motivating application of Section 5.

3.2 Imputation estimator

The nested counterfactual $Y(a,M(a^*))$ is only observable when a^* equals a and, in addition, a corresponds to the observed exposure level A. When a^* differs from a, then $Y(a,M(a^*))$ for a^* equaling the observed exposure level can still be predicted as $E(Y|A=a,M,C)$ under a suitable model for the outcome. This can be seen upon noting that $M(a^*)$ equals M among subjects with a^* equaling the observed value A. This gives rise to an imputation procedure whereby we will complement the observed data set with imputed data sets in which the same individuals are evaluated at different exposure levels, but corresponding to the observed mediator level. We explain the proposal first for dichotomous exposures:

1. Fit a suitable model for the outcome conditional on exposure, mediator and baseline variables using the original data set.
2. Create a new data set by repeating each observation in the original data set twice and including two additional variables: a, which is equal to the original exposure for the first replication and equal to the opposite of the actual exposure for the second replication, and a^*, which is equal to the original exposure.

3. Impute the nested counterfactual $Y(a,M(a^*))$ as Y in those lines of the database where a equals a^* and as the expected value $E(Y|A=a,M,C)$ on the remaining lines, using predict-functionality on the fitted outcome model. Thus when the outcome is binary, the imputed value will be a probability.
4. Fit the natural effect model of interest by regressing the imputed outcome on a, a^* and C on the basis of the expanded data set. Standard errors and confidence intervals can be obtained using the bootstrap (including steps 1-4).

For continuous exposures, we recommend changing the second step of the imputation algorithm as follows: create a new data set by repeating each observation in the original data set K times and including two additional variables: a, which is equal to the original exposure for the first replication and equal to a (different) random draw from the conditional exposure distribution, given C, for all $K-1$ remaining replications; and a^*, which is equal to the original exposure. While K is ideally as large as possible, computational resources will restrict the maximum number that can be chosen in any given application.

The imputation estimator has close connections to imputation-based strategies for G-computation (Snowden, Rose, and Mortimer, 2011), and shares their virtues and limitations (Vansteelandt and Keiding, 2011). It is appealing for routine use because of its simplicity and avoidance of inverse probability weights that can make alternative proposals (see e.g. Hong (2010) and Lange et al. (2012)) unstable in certain situations. However, its simplicity may be somewhat deceptive in that the difficulty in working with the imputation estimator - as with multiple imputation estimators for incomplete data analyses - is finding a coherent model specification; that is, a model for the outcome that is congenial with the natural effect model of interest. Suppose for instance that the model for the outcome were of the form

$$E(Y|M,A,C) = \alpha_0 + \alpha_1 M + \alpha_2 A + \alpha_3 C$$

and the natural effect model of interest were

$$E[Y(a,M(a^*))|C] = \beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 a a^* + \beta_4 C.$$

Then, under data generating models where M is linear in A, the outcome model would preclude the existence of interactions between a and a^* and thus bias β_3 towards zero. With concern for such interactions, the user should therefore work with outcome models which are sufficiently flexible not to preclude the existence of such interactions. As a guideline, we therefore recommend that the outcome model should at least reflect the structure of the natural effect model, but with a^* substituted by M. Thus, given the above natural effect model, the outcome model

should at least contain main effects in A, M and C, as well as the product term $A \times M$. Noteworthy is that the imputation estimator - like other imputation estimators (Tan, 2007, Vansteelandt and Keiding, 2011) - may also be more prone to model extrapolation, primarily in settings where A shows strong associations with either M or C.

3.3 Doubly robust estimator

The concern for the imputation model (i.e., the outcome model) failing to reflect the structure of the natural effect model gets exacerbated in non-linear models, where it becomes difficult or impossible to guarantee coherent model specifications. This has lead us to consider a second estimator, which is exactly obtained like the imputation estimator, but requires fitting the imputation models using weighted regression methods, with weights

$$\frac{f(A=1|C)}{f(A=0|C)}\frac{f(M|A=1,C)}{f(M|A=0,C)} \tag{7}$$

for the unexposed ($A=0$) and

$$\frac{f(A=0|C)}{f(A=1|C)}\frac{f(M|A=0,C)}{f(M|A=1,C)} \tag{8}$$

for the exposed ($A=1$). Using these weights ensures that for the unexposed, the imputation model will fit relatively better in regions of the covariate space where it is likely to find exposed subjects with the same mediator value. This makes intuitive sense as for the unexposed, the imputation model is precisely used to predict what the outcome would have been had the subject been exposed, but the mediator stayed unchanged. Likewise for the exposed, using the weight (8) ensures the imputation model will fit relatively better in regions of the covariate space where it is likely to find unexposed subjects with the same mediator value. This makes again intuitive sense as for the exposed, the imputation model is precisely used to predict what the outcome would have been had the subject been unexposed, but the mediator stayed unchanged.

We thus arrive at the following strategy for dichotomous exposure, which is identical to that of the imputation estimator, apart from the first two steps:

1. Fit suitable models for the mediator conditional on exposure and baseline variables, and for the exposure conditional on baseline variables, using the original data set.

2. Fit a suitable generalized linear model for the outcome conditional on exposure, mediator and baseline variables using the standard maximum likelihood procedure, but using the weights (7) for the unexposed ($A = 0$) and (8) for the exposed ($A = 1$).
3. Create a new data set by repeating each observation in the original data set twice and including an additional variable A^*, which is equal to the original exposure for the first replication and equal to the opposite of the actual exposure for the second replication.
4. Impute the nested counterfactual $Y(a^*, M(a))$ as Y in those lines of the database where A equals A^* and as the expected value $E(Y|A = A^*, M, C)$ on the remaining lines, using predict-functionality on the fitted models obtained in step 2.
5. Fit the natural effect model of interest by regressing the outcome on A, A^* and C on the basis of the expanded data set. Standard errors and confidence intervals can be obtained using the bootstrap (including steps 1-5).

In the Appendix, we show that the resulting estimator is a doubly robust estimator of the parameters indexing the natural effect model. More precisely, it is a consistent estimator of these parameters when the natural effect model holds and, in addition, either (1) the conditional expectation of the outcome, given the exposure and confounders, is correctly specified; or (2) the conditional distribution of the mediator, given the exposure and confounders, and of the exposure, given confounders, are correctly specified. This double robustness is theoretically appealing because it implies that the aforementioned possibility of the outcome model and the natural effect model being uncongenial, is of lesser concern when the exposure and mediator models are reasonably well specified. The doubly robust estimator is closely related to a doubly robust estimator in Tchetgen Tchetgen and Shpitser (2011) and can be viewed as a compromise between the inverse probability weighted estimator of Hong (2010) and Lange et al. (2012), and the here proposed imputation estimator. In particular, the imputation estimator is attained upon setting all weights in the doubly robust estimator to 1.

4 Simulation study

To assess the performance of the proposed natural direct and indirect effects estimators, a simulation study with 1000 runs for data sets of 500 observations was performed. Motivated by the data analysis in Section 5, our focus in Section 4.1 was on a setting with binary outcome and continuous mediator; in Section 4.2 we explore the proposed algorithm for continuous exposures.

4.1 Dichotomous exposure

In the first two experiments, a binary exposure A was drawn from a binomial distribution with $P(A=1|C) = \text{expit}(\gamma_0 + \gamma_1 C)$ with $\gamma = (0.25, -0.5)$, C standard normal and $\text{expit}(x) = \exp(x)/\{1+\exp(x)\}$. The continuous mediator M was drawn from a normal distribution with $E(M|A,C) = \alpha_0 + \alpha_1 A + \alpha_2 C$ with $\alpha_0 = 1, \alpha_2 = -5$ and variance $\sigma^2 = 1$. The binary outcome Y was drawn from a binomial distribution with $P(Y=1|A,M,C) = \Phi(\theta_0 + \theta_1 A + \theta_2 M + \theta_3 C)$, with $\theta_0 = 0.5, \theta_2 = 0.75$ and $\theta_3 = 0.5$, where $\Phi(.)$ refers to the cumulative standard normal distribution function. It can be verified upon applying the mediation formula that this choice of models entails a natural effect model of the form

$$P[Y(a, M(a^*)) = 1|C] = \Phi(\beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 C)$$

where

$$\begin{aligned} \beta_1 &= \frac{\theta_1}{\sqrt{1+\theta_2^2\sigma^2}} \\ \beta_2 &= \frac{\theta_2\alpha_1}{\sqrt{1+\theta_2^2\sigma^2}}. \end{aligned}$$

We evaluated the here proposed imputation estimator and doubly robust estimator, the inverse probability weighted (IPW) estimator of Lange et al. (2012), the more traditional estimators based on the product of coefficient method and difference in coefficient method for the indirect effect, and finally application of the mediation formula. The latter amounts to substituting θ_1, θ_2 and σ^2 in the above expressions for β_1 and β_2 by their maximum likelihood estimates. The resulting maximum likelihood estimators are efficient under the model defined by the restrictions of the outcome and mediator models, and thus set a reference to which one may compare the precision of the other methods. Standard errors were obtained using the nonparametric bootstrap (500 resamples) and the coverage of 95% standard normal approximation bootstrap confidence intervals is reported.

In the first experiment, $\alpha_1 = 3$ and $\theta_1 = 0.1$, corresponding with natural direct effect $\beta_1 = 0.08$ and natural indirect effect $\beta_2 = 1.8$. Features of the first simulation experiment are that the indirect effect is large and that the exposure-mediator and confounder-mediator associations are sufficiently strong to result in extreme inverse probability weights. In particular, the weights corresponding to the IPW estimator (min. 2.5 10^{-6}, median 1, max. 101) and the doubly robust estimator (min. 5.5 10^{-7}, median 0.79, max. 138) were highly variable. This induced large bias in the IPW estimator of the direct effect (see Table 1) and the indirect effect (see

Table 2), but not as much in the doubly robust estimator which additionally relies on the imputation model. Because of the large magnitude of the indirect effect, the product of coefficient estimator for the indirect effect was severely biased, but not the difference of coefficient estimator. Overall, the best results in terms of bias, precision and coverage of the 95% confidence intervals were attained for the maximum likelihood estimator (mediation formula) and the imputation estimator, with both having similar performance. Their precision was found to be greater or of similar magnitude than that of the traditional mediation analysis approaches, and moreover greater than that of the IPW and doubly robust estimators.

Table 1: Simulation results for natural direct effects in experiments 1 and 2: Bias, Empirical standard deviation (Emp SD), Average / Median bootstrap standard errors (Me(di)an SE) and coverage of 95% bootstrap-based confidence intervals for traditional regression adjustment (Regr), the maximum likelihood approach based on the mediation formula (Med Form), the inverse probability weighted estimator (IPW), the imputation estimator (Imp) and the doubly robust estimator (DR).

Exp		Regr	Med Form	IPW	Imp	DR
1	Bias	0.03	0.02	0.63	0.02	0.14
	Emp SD	0.44	0.34	0.70	0.35	0.84
	Mean SE	0.46	0.35	0.55	0.35	0.64
	Coverage	0.96	0.95	0.72	0.96	0.87
2	Bias	0.12	0.005	0.004	0.006	0.005
	Emp SD	0.23	0.18	0.19	0.18	0.18
	Median SE	0.23	0.18	0.80	0.18	0.19
	Coverage	0.92	0.95	1.00	0.95	0.96

In the second experiment, $\alpha_1 = 0$ and $\theta_1 = 0.5$, corresponding to natural direct effect $\beta_1 = 0.4$ and natural indirect effect 0. Since the indirect effect is now zero, the product of coefficient estimator of the indirect effect was unbiased, but not the difference of coefficient estimator as a result of non-collapsibility of the odds ratio. The IPW estimator, the imputation estimator and the doubly robust estimator now all have a fairly similar performance, but not their standard errors. In particular, the bootstrap standard errors of the IPW estimator were very ill-behaved, averaging to $3.6\ 10^{12}$ for the direct effect and $1.0\ 10^{12}$ for the indirect effect, thereby resulting in conservative inferences. The standard errors of the imputation estimator and the doubly robust estimator were much better behaved (with averages close to the medians), although some conservatism was observed for the doubly robust estimator of the indirect effect. The IPW estimator, the imputation estimator and the doubly robust estimator of the direct effect were as efficient as the maximum likelihood

estimator, but this time loss of precision was observed for the imputation estimator and the doubly robust estimator of the indirect effect. As theoretically expected, the maximum likelihood approach based on the mediation formula thus appears to be the preferred approach. However, note that the appeal of the IPW, imputation and doubly robust estimators is that they make reference to a parsimonious model for the natural direct and indirect effects. In contrast, whenever a different link function than the probit is used in the outcome model, or a different distribution than the normal in the mediator model, then application of the mediation formula would result in intractable expressions for the natural direct and indirect effects, which carry an intricate dependence on exposure (when the exposure is continuous) and covariates.

Table 2: Simulation results for natural indirect effects in experiments 1 and 2: Bias, Empirical standard deviation (Emp SD), Average / Median bootstrap standard errors (Me(di)an SE) and coverage of 95% bootstrap-based confidence intervals for the product of coefficient method (Prod), the difference of coefficient method (Diff), the maximum likelihood approach based on the mediation formula (Med Form), the inverse probability weighted estimator (IPW), the imputation estimator (Imp) and the doubly robust estimator (DR).

Exp		Prod	Diff	Med Form	IPW	Imp	DR
1	Bias	0.56	0.03	0.03	-0.33	0.04	0.00
	Emp SD	0.47	0.32	0.23	0.64	0.26	0.67
	Mean SE	0.51	0.34	0.23	0.27	0.26	0.52
	Coverage	0.87	0.96	0.93	0.48	0.94	0.90
2	Bias	0.003	-0.11	0.003	0.003	0.0004	0.001
	Emp SD	0.073	0.11	0.056	0.056	0.085	0.085
	Median SE	0.076	0.11	0.058	0.055	0.087	0.11
	Coverage	0.98	0.89	0.96	0.997	0.96	0.997

4.2 Continuous exposure

To gain insight into our proposal for handling continuous exposures, we evaluated the imputation estimator for a normally distributed exposure under 4 conditions: (a) for each individual, 3 values of A are randomly drawn from the estimated conditional distribution of A, given C; (b) for each individual, 3 values of A are chosen to be the 10%, 50% and 90% percentiles of the conditional distribution of A, given C; (c) for each individual, 5 values of A are randomly drawn from the estimated conditional distribution of A, given C; (d) for each individual, 5 values of A are chosen to

be the 10%, 25%, 50%, 75% and 90% percentiles of the conditional distribution of A, given C. Data were generated as in simulation experiments 1 and 2, but with A being normally distributed with mean $\gamma_0 + \gamma_1 C$ and standard deviation 0.5. Results were very similar for all 4 procedures. For instance, for the direct effect of 0.08 in simulation experiment 1, the bias and empirical standard deviations were 0.0157 and 0.326 for proposal (a), 0.0154 and 0.324 for proposal (b), 0.0137 and 0.325 for proposal (c), and 0.0154 and 0.324 for proposal (d). Discrepancies between the estimators were equally minor for the indirect effect, as well as under the settings of simulation experiment 2.

5 Data analysis

We re-analyze data of 5882 adult respondents from the WHO-LARES study (Shenassa, Daskalakis, Liebhaber, Braubach, and Brown, 2007), which investigated the association between residence in a damp and moldy dwelling and the risk of depression. Because a sense of compromised control over one's living environment (e.g. keeping a house clean in the face of recurrent mold) may mediate a potential link between residence in a damp and moldy dwelling and depression, interest lies in whether the effect of residence in a damp and moldy dwelling (A) on depression (Y) is mediated by perception of control over one's home, as measured on a 5-point Likert scale (reverse coded) (M). A large set of potential confounders C is available on city of residence, survey respondent characteristics (age, gender, marital status, education, employment, smoking, and environmental tobacco smoke at home) and housing characteristics (ownership, size, tenure, crowding, ventilation, natural light, and heating).

Tables 4, 5 and 6 in the Appendix show the results of a logistic regression model for the outcome and exposure, and a linear regression model for the mediator. They show differences in depression rates and in perception of control over one's home between different cities of residence, by respondent characteristics age, gender marital status, education and environmental tobacco smoke, and by housing characteristics crowding and natural light.

We first focus on the natural effect model given by

$$\text{logit}P\{Y(a,M(a^*)) = 1|C\} = \beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 C. \tag{9}$$

Using the models summarized in Table 4, 5 and 6, we estimate the natural direct and indirect odds ratios as given in Table 3. Both the IPW-estimate and the imputation-based estimate suggest that the natural direct and indirect effect of the presence of dampness or mold exposure on the odds of depression amount to odds ratios 1.32 (95% CI 1.07 to 1.63) and 1.05 (95% CI 1.02 to 1.08), respectively. That is, if

an adult was moved from not reside in a damp and moldy dwelling to doing so without changing her perception of control over one's home, the odds of depression would be increased by 32%. Furthermore, if they were all to reside in a damp and moldy dwelling, then the effect of changing their perception of control over one's home to what it would be if they were not to live in such a residence, would be to reduce the odds of depression with 5%. The weights corresponding to the IPW estimator (min. 0.48, median 1, max. 2.07, SD 0.14) were slightly more stable that those corresponding to the DR estimator (min. 0.48, median 0.97, max. 2.07, SD 0.20). There was no evidence of direct - indirect effect modification (i.e., interactions between a and a^*, P $= 0.27$).

Table 3: Estimates of the natural direct and indirect effect odds ratios under model (9).

	Estimator	OR	95% CI	P-value
Direct effect	Regression	1.36	1.10 to 1.68	0.0048
	IPW	1.32	1.07 to 1.63	0.011
	Imputation	1.32	1.07 to 1.63	0.0091
	DR	1.31	1.05 to 1.64	0.016
Indirect effect	Product	1.05	1.03 to 1.08	< 0.0001
	Difference	1.05	1.03 to 1.08	0.0001
	IPW	1.05	1.02 to 1.08	0.00019
	Imputation	1.05	1.02 to 1.08	0.00019
	DR	1.05	1.02 to 1.08	0.00019

We subsequently extended the model to allow for interactions by baseline covariates. In particular, we fitted the natural effect model given by

$$\text{logit}P\{Y(a,M(a^*))=1|C\}=\beta_0+\beta_1 a+\beta_2 a^*+\beta_3 C+\beta_4 aC+\beta_5 a^* C, \qquad (10)$$

and extended the imputation models to include the interactions of those same covariates with exposure and mediator. We found, using the imputation estimator, the magnitude of the natural direct effect to depend on the number of residents per room and the strength of the natural indirect effect to depend on the amount of natural light. In particular, the natural direct effect of the presence of dampness or mold exposure on the odds of depression amounts to odds ratios of 2.06 (95% CI 1.36 to 3.11) in less crowded homes (< 0.5 residents/room), 1.03 (95% CI 0.76 to 1.41) in medium crowded homes (0.51-1 residents/room) and 1.32 (95% CI 0.88 to 1.97) in crowded homes (> 1 residents/room). The natural indirect effect of the presence

of dampness or mold exposure on the odds of depression amounts to odds ratios of 1.10 (95% CI 1.04 to 1.16) in homes with enough light and 1.03 (95% CI 1.00 to 1.05) in homes without enough light. Such interactions would be difficult to assess upon combining standard models for the outcome and mediator via the mediation formula, because of the difficulty of choosing models for the outcome and mediator that combine to natural direct and indirect effects that do not depend on covariates.

6 Discussion

In this paper, we have focused on so-called natural effect models, whose parameters encode the natural indirect effect of an exposure on an outcome, through a given mediator, and the remaining natural direct effect. We have proposed estimators for the parameters indexing these models based on imputation. For pedagogic purposes, we have explained the approach for a class of generalized linear natural effect models, which includes linear, loglinear and logistic models. However, the imputation estimator extends to other types of models, such as parametric survival models. The proposed doubly robust estimator is essentially confined to generalized linear natural effect models, but the results in the Appendix show how such estimators may be obtained for a wider class of models. Noteworthy is that confounding adjustment in all considered approaches is based on regression adjustment, and thus that the natural effect models must include all relevant confounders, even when their association with the outcome is not of immediate interest.

A major advantage of the formalism of natural direct and indirect effects, as opposed to traditional mediation analysis approaches from the structural equations literature, is that it yields effects that can be interpreted on the scale of the observed data, and that it can validly be used for in principle any type of exposure, mediator or outcome. A major advantage of the proposed approach and related approaches (van der Laan and Petersen, 2008, Tchetgen Tchetgen and Shpitser, 2011) relative to direct application of the mediation formula (see e.g. VanderWeele and Vansteelandt (2009), VanderWeele and Vansteelandt (2010), Imai et al. (2010a), Lange and Hansen (2011), VanderWeele (2011)) is that by directly modeling the effects of interest, results become easier to report and hypotheses of interest become easier to test.

Both proposed estimators are easy to obtain using standard statistical software; the doubly robust estimator is obtained just like the imputation estimator, but relies on weighted regression techniques for fitting the imputation model. Both estimators differ in terms of their properties. Unlike the IPW estimator, the simple imputation estimator avoids inverse weighting and thereby tends to give more

precise estimators. However, it requires more care in choosing an adequate imputation model and may overstate the precision of the resulting estimators by ignoring extrapolation uncertainty. From a theoretical point of view, the doubly robust imputation estimator seems a useful compromise between the IPW estimator and the imputation estimator: it is valid when either the imputation model, or both the exposure and mediator models are correctly specified. In large samples, it may thus turn out to be less sensitive to bias resulting from the lack of a coherent model specification. However, simulations studies in Section 4 (as well as others) have demonstrated that the doubly robust estimator can sometimes be quite inefficient relative to the imputation estimator. Furthermore, the inverse probability weights on which it relies can be much less stable than those employed by the IPW estimator, and this may raise concerns about its validity in small samples. Finally, the asserted property of double robustness may be somewhat illusory (Robins and Rotnitzky, 2001) because correct specification of the exposure and mediator model may anyway be essential in the face of incoherent model specification. Relative to our proposal, the approaches of van der Laan and Petersen (2008) and Tchetgen Tchetgen and Shpitser (2011) minimize the risk of incoherent model specification and promise a more guaranteed robustness against model misspecification. However, they rely on similar inverse probability weights and will thus likely suffer the same instability as sometimes observed for the proposed doubly robust estimator. Moreover, these approaches are limited to the identity and log link, and are, to the best of our knowledge, not as readily obtainable via off-the-shelf statistical software.

Our recommendation is to use the simple imputation estimator for routine analysis. We make this recommendation in view of the foregoing discussion, the empirical simulation findings, and the fact that the concerns about model uncongeniality may be relatively lenient if one is careful to make the imputation model sufficiently rich. Note for instance that concerns about model uncongeniality are much more conspicuous in the widely adopted multiple imputation strategies for the analysis of incomplete data, where the imputation model typically not just refers to the outcome distribution, but to the high-dimensional distribution of outcomes and covariates; in such settings, the bias due to incoherent model specification has often been found to be relatively modest (see e.g. Van Buuren, Brand, Groothuis-Oudshoorn, and Rubin (2006)). Furthermore, if the imputation model is correct, then even if the natural effect model is misspecified and thus uncongenial with the imputation model, it can still be viewed as a (simplified) summary. We believe that the proposed doubly estimator and the multiply robust estimators of Tchetgen Tchetgen and Shpitser (2011, 2012) are also very valuable, but that further work is required to improve them via the choice of estimators of the nuisance parameters indexing the models for the exposure, mediator and outcome (Cao, Tsiatis, and Davidian, 2009, Vansteelandt, Bekaert, and Claeskens, 2012), or via the choice of

index functions (cfr. $d_{a,a^*}(C)$ in the Appendix) in the estimating equations for the parameters indexing the natural effect model. Since our goal was to concentrate on approaches that are easy enough to perform via standard software, these extensions are beyond the scope of this work.

In summary, we have proposed and evaluated procedures for mediation analysis that are easy to perform with standard software. We hope that the flexibility of the procedures for handling arbitrary mediators and outcomes, as well as their simplicity, will aid to make inference for natural direct and indirect effects accessible to a wider audience.

References

Baron, R. and D. Kenny (1986): "The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations," *J. Pers. Soc. Psychol.*, 51, 1173–1182.

Cao, W., A. Tsiatis, and M. Davidian (2009): "Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data," *Biometrika*, 96, 723–734.

Greenland, S., J. Robins, and J. Pearl (1999): "Confounding and collapsibility in causal inference," *Statistical Science*, 14, 29–46.

Hong, G. (2010): "Ratio of mediator probability weighting for estimating natural direct and indirect effects," in *Proceedings of the American Statistical Association, Biometrics Section*, Alexandria, VA: American Statistical Association, 2401–2415.

Imai, K., L. Keele, and D. Tingley (2010a): "A General Approach to Causal Mediation Analysis," *Psychological Methods*, 15, 309–334.

Imai, K., L. Keele, and T. Yamamoto (2010b): "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, 25, 51–71.

Lange, T. and J. V. Hansen (2011): "Direct and Indirect Effects in a Survival Context," *Epidemiology*, 22, 575–581.

Lange, T., S. Vansteelandt, and M. Bekaert (2012): "A simple approach for estimating natural direct and indirect effects," *American Journal of Epidemiology*, 176, 190–195.

MacKinnon, D. (2008): *An Introduction to Statistical Mediation Analysis*, New York: Lawrence Erlbaum Associates.

Newey, W. K. (1990): "Semiparametric efficiency bounds," *J. Appl. Economet.*, 5, 99–135.

Pearl, J. (2001): "Direct and indirect effects," in *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence*, San Francisco: Morgan

Kaufmann, 411–420.

Pearl, J. (2011): "The mediation formula: A guide to the assessment of causal pathways in nonlinear models," in *Causality: Statistical Perspectives and Applications*.

Robins, J. and A. Rotnitzky (2001): "Inference for semiparametric models: Some questions and an answer - Comments," *Statistica Sinica*, 11, 920–936.

Robins, J. M. and S. Greenland (1992): "Identifiability and exchangeability for direct and indirect effects." *Epidemiology*, 3, 143–155.

Shenassa, E. D., C. Daskalakis, A. Liebhaber, M. Braubach, and M. Brown (2007): "Dampness and mold in the home and depression: An examination of mold-related illness and perceived control of one's home as possible depression pathways," *American Journal of Public Health*, 97, 1893–1899.

Snowden, J. M., S. Rose, and K. M. Mortimer (2011): "Implementation of G-Computation on a Simulated Data Set: Demonstration of a Causal Inference Technique," *American Journal of Epidemiology*, 173, 731–738.

Tan, Z. (2007): "Understanding OR, PS, and DR, Comment on "Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data" by Kang and Schafer," *Statistical Science*, 22, 560–568.

Tchetgen Tchetgen, E. J. and I. Shpitser (2011): "Semiparametric estimation of models for natural direct and indirect effects," Technical report, Harvard University Biostatistics Working Paper Series. Working Paper 129.

Tchetgen Tchetgen, E. J. and I. Shpitser (2012): "Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis," *Annals of Statistics*, in press.

Van Buuren, S., J. P. L. Brand, C. G. M. Groothuis-Oudshoorn, and D. B. Rubin (2006): "Fully conditional specification in multivariate imputation," *Journal of Statistical Computation and Simulation*, 76, 1049–1064.

van der Laan, M. and J. Robins (2002): *Unified methods for censored longitudinal data and causality*, New York: Springer.

van der Laan, M. J. and M. L. Petersen (2008): "Direct effect models," *The International Journal of Biostatistics*, 4, Article 23.

VanderWeele, T. J. (2011): "Causal Mediation Analysis With Survival Data," *Epidemiology*, 22, 582–585.

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

VanderWeele, T. J. and S. Vansteelandt (2010): "Odds Ratios for Mediation Analysis for a Dichotomous Outcome," *American Journal of Epidemiology*, 172, 1339–1348.

Vansteelandt, S., M. Bekaert, and G. Claeskens (2012): "On model selection and

model misspecification in causal inference," *Statistical Methods in Medical Research*, 21, 7–30.

Vansteelandt, S. and N. Keiding (2011): "Invited Commentary: G-Computation-Lost in Translation?" *American Journal of Epidemiology*, 173, 739–742.

Appendix

Derivation of the doubly robust estimator

In this section, we will explain how to obtain doubly robust estimators of the parameters indexing the more general natural effect model

$$E\{Y(a,M(a^*))|C\} = m(a,a^*,C;\beta),$$

where $m(a,a^*,C;\beta)$ is a known function, smooth in an unknown finite-dimensional parameter β.

We start with a sketch of the derivation of the set of influence functions (Newey, 1990) for β in the above model, which is closely related to the derivation given in Tchetgen Tchetgen and Shpitser (2011). First note that

$$m(a,a^*,C;\beta) = \int E(Y|A=a,M=m,C)f(M=m|A=a^*,C)dm,$$

and let a and a^* be fixed values for now. Taking derivatives along one-dimensional parametric submodels, indexed by t, we obtain

$$\begin{aligned} 0 &= \int\int yf(Y=y|A=a,M=m,C)f(M=m|A=a^*,C) \\ &\quad \times \left\{S_y(y|A=a,M=m,C)+S_m(M=m|A=a^*,C)\right\}dydm, \end{aligned}$$

where $S_y(y|A=a,M=m,C)$ and $S_m(M=m|A=a^*,C)$ are the scores (w.r.t. t) corresponding to the components $f(Y=y|A=a,M=m,C)$ and $f(M=m|A=$

$a^*, C)$, respectively. Note that

$$\begin{aligned}
&\int\int yf(Y=y|A=a,M=m,C)f(M=m|A=a^*,C)S_y(y|A=a,M=m,C)dydm\\
&=\int\int\int y\frac{I(A=a)}{f(A=a|C)}\frac{f(M=m|A=a^*,C)}{f(M=m|A,C)}S_y(y|A,M=m,C)\\
&\qquad\times f(Y=y,M=m,A|C)dydmdA\\
&=\int\int\int\{y-E(Y|A=a,M=m,C)\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M=m|A=a^*,C)}{f(M=m|A=a,C)}\\
&\qquad\times S_y(y|A,M=m,C)f(Y=y,M=m,A|C)dydmdA\\
&=\int\int\int\{y-E(Y|A=a,M=m,C)\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M=m|A=a^*,C)}{f(M=m|A=a,C)}\\
&\qquad\times\left\{S_y(y|A,M=m,C)+S_m(M=m|A,C)+S_a(A|C)\right\}\\
&\qquad\times f(Y=y,M=m,A|C)dydmdA,
\end{aligned}$$

and

$$\begin{aligned}
&\int E(Y|A=a,M=m,C)f(M=m|A=a^*,C)S_m(M=m|A=a^*,C)dm\\
&=\int\int E(Y|A=a,M=m,C)\frac{I(A=a^*)}{f(A=a^*|C)}S_m(M=m|A,C)\\
&\qquad\times f(M=m,A|C)dmdA\\
&=\int\int\{E(Y|A=a,M=m,C)-m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}\\
&\qquad\times S_m(M=m|A,C)f(M=m,A|C)dmdA\\
&=\int\int\int\{E(Y|A=a,M=m,C)-m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}\\
&\qquad\times\left\{S_y(y|A,M=m,C)+S_m(M=m|A,C)+S_a(A|C)\right\}\\
&\qquad\times f(Y=y,M=m,A|C)dydmdA.
\end{aligned}$$

We conclude from both these expressions that for each one-dimensional parametric submodel, the score $U=S_y(y|A,M,C)+S_m(M|A,C)+S_a(A|C)$ satisfies

$$\begin{aligned}
0 \;=\; & E\left[U\left\{\{Y-E(Y|A=a,M,C)\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M|A=a^*,C)}{f(M|A=a,C)}\right.\right.\\
&\left.\left.+\{E(Y|A=a,M,C)-m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}\right\}|C\right],
\end{aligned}$$

for each a, a^*. It is now immediate from Theorem 2.2 in Newey (1990) that all influence functions under the natural effect model for fixed a and a^* are given by

$$d_{a,a^*}(C)\left[\{Y-E(Y|A=a,M,C)\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M|A=a^*,C)}{f(M|A=a,C)}\right.$$
$$\left.+\{E(Y|A=a,M,C)-m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}\right],$$

where $d_{a,a^*}(C)$ is an arbitrary index function. Because the natural effect model is a union model corresponding to all possible choices a and a^*, all influence functions under the natural effect model for arbitrary a and a^* are given by (van der Laan and Robins, 2002)

$$\sum_a\sum_{a^*} d_{a,a^*}(C)\left[\{Y-E(Y|A=a,M,C)\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M|A=a^*,C)}{f(M|A=a,C)}\right.$$
$$\left.+\{E(Y|A=a,M,C)-m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}\right].$$

It follows from the above that for binary A taking the values 0 and 1, all influence functions for β are of the form

$$\begin{aligned}
&d_{0,0}(C)\frac{I(A=0)}{f(A=0|C)}\{Y-m(0,0,C;\beta)\}\\
&+d_{0,1}(C)\left[\frac{I(A=0)}{f(A=0|C)}\frac{f(M|A=1,C)}{f(M|A=0,C)}\{Y-E(Y|A=0,M,C)\}\right.\\
&\left.+\frac{I(A=1)}{f(A=1|C)}\{E(Y|A=0,M,C)-m(0,1,C;\beta)\}\right]\\
&+d_{1,0}(C)\left[\frac{I(A=1)}{f(A=1|C)}\frac{f(M|A=0,C)}{f(M|A=1,C)}\{Y-E(Y|A=1,M,C)\}\right.\\
&\left.+\frac{I(A=0)}{f(A=0|C)}\{E(Y|A=1,M,C)-m(1,0,C;\beta)\}\right]\\
&+d_{1,1}(C)\frac{I(A=1)}{f(A=1|C)}\{Y-m(1,1,C;\beta)\}.
\end{aligned}$$

Writing

$$d_{a,a^*}(C)=f(A=a^*|C)d^*_{a,a^*}(C),$$

we obtain,

$$
\begin{aligned}
&I(A=0)d^*_{0,0}(C)\{Y-m(0,0,C;\beta)\}\\
&+I(A=0)d^*_{1,0}(C)\{E(Y|A=1,M,C)-m(1,0,C;\beta)\}\\
&+I(A=1)d^*_{0,1}(C)\{E(Y|A=0,M,C)-m(0,1,C;\beta)\}\\
&+I(A=1)d^*_{1,1}(C)\{Y-m(1,1,C;\beta)\}\\
&+\frac{I(A=0)f(A=1|C)}{f(A=0|C)}\frac{f(M|A=1,C)}{f(M|A=0,C)}d^*_{1,0}(C)\{Y-E(Y|A=0,M,C)\}\\
&+\frac{I(A=1)f(A=0|C)}{f(A=1|C)}\frac{f(M|A=0,C)}{f(M|A=1,C)}d^*_{0,1}(C)\{Y-E(Y|A=1,M,C)\}.
\end{aligned}
$$

In the proposed procedure, $d^*_{a,a^*}(C)$ is taken to be the vector of covariates appearing in the outcome model, including 1 for the intercept. The impact of the third step of the proposed algorithm is then to set the sample average of the last 2 lines in the above expression equal to zero. The parameters of interest can now be estimated by setting the sample average of the remaining components in this expression to zero and solving for β. It is easily seen that this amounts to the proposed imputation procedure in lines 4-6 of the proposed algorithm.

Demonstration of double robustness

In this section, we confirm that the proposed estimation strategy yields a doubly robust estimator. The development is again similar to that in Tchetgen and Schpitser (2011). Suppose first that the outcome model for Y is correctly specified. Then we have that

$$
\begin{aligned}
&E\left[d^*_{a,a^*}(C)\{Y-E(Y|A=a,M,C)\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M|A=a^*,C)}{f(M|A=a,C)}\right.\\
&\left.+d^*_{a,a^*}(C)\{E(Y|A=a,M,C)-m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}|C\right]\\
&=d^*_{a,a^*}(C)E\left[\{E(Y|A=a,M,C)-m(a,a^*,C;\beta)\}\frac{I(A=a^*)}{f(A=a^*|C)}|C\right],
\end{aligned}
$$

where, by the mediation formula,

$$
\begin{aligned}
&E\left[E(Y|A=a,M,C)\frac{I(A=a^*)}{f(A=a^*|C)}|C\right]\\
&=E\left[\int E(Y|A=a,M=m,C)f(M=m|A=a^*,C)dm\frac{I(A=a^*)}{f(A=a^*|C)}|C\right]\\
&=E\left[E(Y(a,M(a^*))|C)\frac{I(A=a^*)}{f(A=a^*|C)}|C\right],
\end{aligned}
$$

which confirms the unbiasedness of the estimating functions. Suppose next that the models for M and A are correctly specified, then with $\tilde{E}(Y|A=a,M,C)$ denoting the possibly misspecified model,

$$
\begin{aligned}
&E\left[\left\{Y-\tilde{E}(Y|A=a,M,C)\right\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M|A=a^*,C)}{f(M|A=a,C)}\right.\\
&\left.+\left\{\tilde{E}(Y|A=a,M,C)-m(a,a^*,C;\beta)\right\}\frac{I(A=a^*)}{f(A=a^*|C)}|C\right]\\
&=E\left[\left\{E(Y|A=a,M,C)-\tilde{E}(Y|A=a,M,C)\right\}\frac{I(A=a)}{f(A=a|C)}\frac{f(M|A=a^*,C)}{f(M|A=a,C)}\right.\\
&\left.+\left\{\tilde{E}(Y|A=a,M,C)-m(a,a^*,C;\beta)\right\}\frac{I(A=a^*)}{f(A=a^*|C)}|C\right]\\
&=\int\left\{E(Y(a,M(a^*))|C)-\int\tilde{E}(Y|A=a,M,C)f(M|A=a^*,C)dM\right\}\frac{I(A=a)}{f(A=a|C)}dA\\
&+E\left[\left\{\tilde{E}(Y|A=a,M,C)-m(a,a^*,C;\beta)\right\}\frac{I(A=a^*)}{f(A=a^*|C)}|C\right]\\
&=E(Y(a,M(a^*))|C)-\int\tilde{E}(Y|A=a,M,C)f(M|A=a^*,C)dM\\
&+\int\tilde{E}(Y|A=a,M=m,C)f(M=m|A=a^*,C)dm-m(a,a^*,C;\beta)=0,
\end{aligned}
$$

which shows that the estimating functions maintain their unbiasedness.

R-code for the data analysis

In the first step of the analysis we fit models for the outcome, mediator and exposure. For continuous mediator and dichotomous outcome and exposure, for instance, we define logistic regression models for the outcome and exposure:

```
my <- glm(Y ~ A + M + C , family = binomial)
ma <- glm(A ~ M + C , family = binomial)
```

and a linear regression model for the mediator:

```
mm <- lm(M ~ A + C)
```

We construct a dataset `newDAT` by repeating each observations in the original data set `DAT` twice with an additional variable `Astar`, which is equal to the original exposure. We redefine `A` to be the original exposure for the first replication and equal to the opposite of the actual exposure for the second replication.

```
newDAT=rbind(DAT,DAT)
newDAT$Astar=newDAT$A
newDAT$A=c(rep(0,length(dat$Y)),rep(1,length(dat$Y)))
```

The imputation estimator is now obtained by constructing a new outcome variable `Ynew`, which is equal to the observed outcome if $A = A^*$ and equal to $E(Y|A = A,M,C)$ if $A \neq A^*$:

```
Ynew=ifelse(newDAT$A==newDAT$Astar,newDAT$Y,
       predict.glm(my,newDAT,type=c("response")))
```

This outcome is then used in next model

```
glm(Ynew ~ A + Astar + C, family = binomial, data = newDAT)
```

Because `Ynew` may contain non-integer values, fitting this model may produce warnings, which can be safely ignored. The doubly robust estimator is obtained in the same way, but requires substituting `my` with:

```
my <- glm(Y ~ A + M + C , family = binomial, weights = W)
```

with corresponding weights `WDR` equal to:

```
W = as.numeric((DAT$A*pA0*f0)/(pA1*f1)
        + ((1-DAT$A)*pA1*f1)/(pA0*f0))
```

where

```
pA1 = fitted.values(ma)
pA0 = 1-pA1
DAT1 = cbind(1,1,DAT$C)
```

```
M1 = predict(mm,DAT1)
DAT0 = cbind(1,0,DAT$C)
M0 = predict(mm,DAT0)
f1 = as.numeric(dnorm(DAT$M,M1,summary(mm)$sigma))
f0 = as.numeric(dnorm(dat$M,M0,summary(mm)$sigma))
f = dnorm(DAT$M,as.numeric(fitted.values(mm)),summary(mm)$sigma)
```

The standard errors of the estimators can be obtained via the bootstrap.

Tables with standard regression model results for outcome, exposure and mediator

Table 4: Results from logistic outcome regression model.

Estimator	Level	OR	95% CI	P-value
Dampness and mold		1.36	1.10 to 1.67	0.0040
Perception control		1.31	1.19 to 1.43	< 0.0001
Age		1.02	1.01 to 1.03	< 0.0001
Sex	female	1.63	1.34 to 1.99	< 0.0001
Crowding	> 1 per room	1		
	< 0.5	0.95	0.74 to 1.22	0.68
	0.51-1	1.42	1.02 to 1.99	0.040
Light	enough	1.29	1.04 to 1.59	0.017
City	Angers	1		
	Bonn	0.66	0.34 to 1.25	0.21
	Bratislava	1.92	1.17 to 3.21	0.012
	Budapest	2.41	1.50 to 3.98	0.0004
	Ferreira	4.08	2.62 to 6.58	< 0.0001
	Forli	1.00	0.61 to 1.69	0.99
	Geneva	1.01	0.54 to 1.85	0.97
	Vilnius	1.64	1.03 to 2.67	0.041
Marital status	married	1		
	separated	1.34	1.02 to 1.74	0.032
	single	0.91	0.67 to 1.23	0.55
Education	primary	1		
	secondary	0.86	0.67 to 1.11	0.25
	higher	0.49	0.35 to 0.68	< 0.0001
Env. tobacco smoke		1.54	1.26 to 1.88	< 0.0001

Table 5: Results from logistic exposure regression model.

Estimator	Level	OR	95% CI	P-value
Age		0.99	0.99 to 1.00	0.0002
Crowding	>1 per room	1		
	<0.5	1.66	1.43 to 1.93	<0.0001
	0.51-1	3.18	2.58 to 3.94	<0.0001
Light	enough	1.28	1.12 to 1.47	0.0003
City	Angers	1		
	Bonn	0.54	0.42 to 0.70	<0.0001
	Bratislava	0.26	0.20 to 0.35	<0.0001
	Budapest	0.32	0.25 to 0.43	<0.0001
	Ferreira	2.93	2.11 to 4.07	<0.0001
	Forli	0.96	0.76 to 1.23	0.76
	Geneva	0.32	0.24 to 0.42	<0.0001
	Vilnius	0.41	0.31 to 0.52	<0.0001
Marital status	married	1		
	separated	1.06	0.88 to 1.27	0.56
	single	0.78	0.66 to 0.93	0.0059
Home type		1.82	1.53 to 2.17	<0.0001
Home size	150m^2	1		
	0m^2-49m^2	0.82	0.69 to 0.98	0.029
	50m^2 - 99m^2	0.61	0.49 to 0.76	<0.0001
	100m^2-149m^2	0.53	0.38 to 0.72	<0.0001
Ventilation	free	1		
	no	0.71	0.60 to 0.84	0.0001
	forced	0.88	0.76 to 1.01	0.062
Heating		1.71	1.40 to 2.08	<0.0001

Table 6: Results from linear mediator regression model.

Estimator	Level	Estimate	95% CI	P-value
Dampness and mold		0.20	0.14 to 0.25	< 0.0001
Age		-0.0062	-0.0081 to -0.0042	< 0.0001
Crowding	> 1 per room	0		
	< 0.5	0.28	0.22 to 0.35	< 0.0001
	0.51-1	0.53	0.44 to 0.63	< 0.0001
Light	enough	0.10	0.042 to 0.16	0.0007
City	Angers	0		
	Bonn	-0.32	-0.43 to -0.21	< 0.0001
	Bratislava	0.047	-0.066 to 0.16	0.42
	Budapest	-0.52	-0.63 to -0.40	< 0.0001
	Ferreira	0.097	-0.020 to -0.40	0.11
	Forli	-0.076	-0.18 to 0.031	0.17
	Geneva	-0.029	-0.15 to0.091	0.64
	Vilnius	0.13	0.021 to 0.23	0.019
Marital status	married	0		
	separated	0.077	-0.0016 to 0.16	0.055
	single	0.31	0.24 to 0.38	< 0.0001
Education	primary	0		
	secondary	0.092	0.017 to 0.17	0.016
	higher	0.10	0.019 to 0.19	0.016
Smoke	> 15 cigarettes	0		
	no	0.019	-0.062 to 0.10	0.65
	occasionally	-0.0039	-0.082 to 0.075	0.92
	5-15 cigarettes	0.12	0.039 to 0.21	0.0043
Home type		0.12	0.051 to 0.20	0.0008