International Journal of Epidemiology, 2014, 1–6
doi: 10.1093/ije/dyu107
Original article

Original article

Mediation misgivings: ambiguous clinical and public health interpretations of natural direct and indirect effects

Ashley I Naimi,[1]* Jay S Kaufman[2] and Richard F MacLehose[3]

[1]Department of Obstetrics and Gynecology and [2]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada and [3]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA

*Corresponding author. Department of Obstetrics and Gynecology, McGill University, 687 Pine Ave West, Room F 432, Montreal H3A 1A1, QC, Canada. E-mail: ashley.naimi@mcgill.ca

Abstract

Recent methodological innovation is giving rise to an increasing number of applied papers in medical and epidemiological journals in which natural direct and indirect effects are estimated. However, there is a longstanding debate on whether such effects are relevant targets of inference in population health. In light of the repeated calls for a more pragmatic and consequential epidemiology, we review three issues often raised in this debate: (i) the use of composite cross-world counterfactuals and the need for cross-world independence assumptions; (ii) interventional vs non-interventional identifiability; and (iii) the interpretational ambiguity of natural direct and indirect effect estimates. We use potential outcomes notation and directed acyclic graphs to explain 'cross-world' assumptions, illustrate implications of this assumption via regression models and discuss ensuing issues of interpretation. We argue that the debate on the relevance of natural direct and indirect effects rests on whether one takes as a target of inference the mathematical object per se, or the change in the world that the mathematical object represents. We further note that public health questions may be better served by estimating controlled direct effects.

Key words: Causal inference, epidemiological methods, effect decomposition, mediation, controlled direct effect, natural direct effect, natural indirect effect, intervention

Key Messages

- Natural direct and indirect effects are a common choice for effect measure in applied mediation analyses, but they are subject to difficult interpretational issues.
- The problems we discuss with respect to natural direct and indirect effects are intrinsic to how they are defined, and cannot be resolved using different estimation methods.
- Public health questions may be better served by estimating controlled effects instead of natural effects.

Introduction

In the past decade, the analysis of direct and indirect exposure effects has benefited from a flurry of intellectual activity and insight. Building on the social sciences literature of the early to mid 1980s,[1,2] an extensive body of work is devoted to formalizing effect decomposition under a range of conditions.[3–20] This work has given rise to an increasing number of applied papers in medical and epidemiological journals in which causal mediation analysis is used to decompose effects.[21–26] The effects estimated in these applications can be described with reference to the directed acyclic graph (DAG) in Figure 1, where the direct effect of the exposure X is the portion of the effect transmitted through the path $X \rightarrow Y$, and the indirect effect through $X \rightarrow M \rightarrow Y$.

Two broad types of effects are commonly used in mediation analysis. The first are controlled direct effects, which quantify the exposure effect under an intervention that sets the mediator to a specific value for all individuals in the population. Controlled indirect effects are notably difficult to conceptualize, and instead are defined as some contrast between the total and controlled direct effects in the absence of exposure-mediator interactions.[15] The second type of direct effects are 'natural' (or 'pure') direct effects, which are meant to estimate the exposure effect that would be observed holding the mediator fixed at the value it would have taken under some referent exposure value. Similarly, natural indirect effects seek to estimate the effect the exposure has through the mediator by changing the mediator from the value it would have taken under some referent exposure to the value it would have taken under some specific alternate exposure (all while holding the exposure fixed). In all cases (controlled direct, natural direct and natural indirect effects), the causal effect of interest is defined as a contrast of potential outcomes, or a contrast of outcomes that would be observed under different (possibly counter to fact) exposure and mediator values.

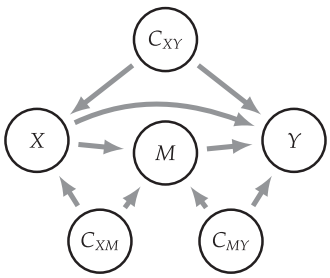

Figure 1. Directed acyclic graph representing the relations between exposure (X), mediator (M) and outcome (Y), as well as the exposure-outcome (C_{XY}), exposure-mediator (C_{XM}) and mediator-outcome (C_{MY}) confounders. Natural direct and indirect effects can be consistently estimated upon adjusting for C_{XY}, C_{XM} and C_{MY}. Note that there is no arrow from X to C_{MY} in which case natural direct and indirect effects cannot be identified.

As practitioners of the science of public health, it has long been argued that epidemiologists should seek to estimate parameters that have a logical correspondence with some realistic intervention that might be taken to improve population health.[27–31] Similarly, in causal inference, the idea that one cannot estimate causal effects without some clearly defined (possibly hypothetical) exposure intervention is widespread.[28,32–35] This perspective has, for some time now, fuelled a debate on the usefulness of decomposing effects into their natural direct and indirect components. Some[3,5,36] have argued that because natural direct and indirect effects cannot be identified using intervention-based causal models, and cannot be estimated in a randomized trial, they cannot be interpreted as effects that have a logical correspondence with some public health action or policy. Others[4,37–39] have argued that natural direct and indirect effects are identifiable using causal inference models that make stricter assumptions about the world, that they can be used to infer mechanistic relations and that many causal effects cannot be estimated in a randomized intervention trial for logistical or ethical reasons and thus should not be ruled out as a means of providing information on the actions of various mechanisms.

Although some of these issues have been mentioned in epidemiological journals,[3,36,40,41] much of the debate on the interpretation of natural direct and indirect effects is directed to highly technical audiences in mathematical statistics and computer science. Indeed, a recent review addressed to epidemiologists on mediation analysis, which emphasized interpretation, failed to even note the controversy surrounding natural direct and indirect effects.[42] The purpose of this paper is to review and summarize the technical literature by Robins,[3,5,43] Pearl[4,38,44] and others[36,39,45,46] on the interpretation of natural direct and indirect effects. We highlight three related issues that obfuscate their interpretation: (i) the use of composite cross-world counterfactuals and the need for cross-world independence assumptions; (ii) interventional vs non-interventional identifiability; and (iii) the ambiguous policy implications of natural direct and indirect effect estimates. We start by defining relevant terms, and review the assumptions needed to identify natural direct and indirect effects that render them difficult to interpret. We consider regression models for natural direct and indirect effects to illustrate why they cannot be identified via actual interventions. We discuss ensuing issues of scientific falsifiability and interpretation, and argue that the debate on the relevance of natural direct and indirect effects rests on whether one takes as a target of inference the mathematical object per se, or the change in the world that the mathematical object is meant to represent.

Natural direct and indirect effects

Conceptually, the natural indirect effect is the expected change in the outcome when the mediator changes as though the exposure had (but when, in actuality, the exposure hadn't) changed.[44] Similarly, the natural direct effect is the expected change in the outcome if we were to 'freeze' the mediator value for each person at the level it would have taken had the person's exposure been some referent level (but when, in actuality, the person's exposure status changes).[44] To formalize these concepts, we use potential outcomes.[32] Throughout, we let capital letters denote random variables, and lower-case letters denote possible realizations. We let X denote exposure, M denote mediator, Y denote outcome and $\mathbf{C} \equiv \{C_{XY}, C_{XM}, C_{MY}\}$ denote confounders of the exposure-outcome, exposure-mediator and mediator-outcome relations, respectively. We start with a hypothetical study in which we collect information on n individuals. For each individual indexed by i, we let $Y_i[x, M_i(x)]$ denote the outcome we would have observed for individual i had she been exposed to $X_i = x$, with a mediator status defined as $M_i(x)$. Importantly, the potential outcome $Y_i[x, M_i(x)]$ is itself a function of another potential outcome $M_i(x)$: the mediator value that would have been observed under $X_i = x$. We say $Y_i[x, M_i(x)]$ is a composite counterfactual,[47] a basic element needed to define natural direct and indirect effects. For instance, the difference in means for the natural direct effect of X on Y, independent of the effect of X on M, is:

$$NDE(x^*) = \mathbb{E}\{Y[x, M(x^*)]\} - \mathbb{E}\{Y[x^*, M(x^*)]\}, \qquad (1)$$

where the expectation $\mathbb{E}\{\bullet\}$ is taken over individuals i, x and x^* represent some assigned and referent exposure values, and $Y[x, M(x^*)]$ represents the potential outcome that would have been observed under $X = x$ and under the potential mediator value that would have been observed under $X = x^*$. Such composite counterfactuals are the first problem with natural direct and indirect effects. Estimating them requires a union of two logically incompatible states: the outcome under exposure x with the mediator set to what it would have been under x^*. Because no single individual can ever exist with exposure values x and x^*, this composite counterfactual requires information that can only exist in two separate 'worlds', and has thus been referred to as a 'cross-world' counterfactual.[46]

It is worth emphasizing the unique status of the cross-world counterfactual in light of the more typical contrasts estimated in causal inference. That a single individual can never be observed in both exposed and unexposed states is known as the fundamental problem of causal inference.[32] Decades of work has been devoted to formalizing how different methods (e.g. standard regression, marginal structural models) can be used to impute summary measures (e.g. means, hazards, odds) of missing potential outcomes under an unobserved exposure state.[48–50] However, natural direct and indirect effects go beyond this by requiring that one impute information on the potential outcome that would have been observed under the union of two logically incompatible exposure states.

Identifiability assumptions

Every mediation analysis requires that a set of unverifiable assumptions be met.[3–6,45,51,52] These assumptions can be depicted using the directed acyclic graph (DAG) in Figure 1, which shows that, conditional on $\mathbf{C} = \{C_{XM}, C_{XY}, C_{MY}\}$, there must be:

i. no uncontrolled exposure-outcome confounding
ii. no uncontrolled mediator-outcome confounding
iii. no uncontrolled exposure-mediator confounding
iv. no measured or unmeasured effect of the exposure that confounds the mediator outcome relation.

Identifying controlled direct effects requires that assumptions (i) and (ii) be met. Identifying natural direct and indirect effects entails the addition of assumptions (iii) and (iv). In words, assumption (iv) requires that, conditional on C_{MY}, the mediator that would have been observed in a world where $X = x^*$ is independent of the outcome that would have been observed in a world where $X = x$. This assumption is known as a cross-world independence assumption.[46] Because this assumption can never be tested in an empirical setting,[44] a number of authors have raised concerns about the interpretation of natural direct and indirect effect estimates.[43,46,53] Others have developed alternate conditions under which natural direct and indirect effects can be identified.[51,54] All of them, however, require independence between some function of the potential mediator $M(x^*)$ and some function of the potential outcome $Y[x, M(x)]$.

Using DAGs, assumption (iv) leads to the idea of the 'recanting witness'. This name arose because of the fact that the exposure variable 'tries to have it both ways'.[45] Along the path in Figure 1 defined by $X \rightarrow M \rightarrow Y$, the exposure behaves as though it exists in some exposed state $X = x$. However, along the path defined by $X \rightarrow Y$, the same exposure 'changes the story' and behaves as though it exists in some unexposed referent state $X = x^*$.[45,55] The implications of this recantation are nebulous from an epidemiological perspective. How might one interpret an effect estimate in which the exposure X can behave as a realization x along one causal path, but a different realization x^* along another causal path?

Although interpreting natural direct and indirect effects may be difficult, methods have been proposed to estimate

these effects. For example, one approach[6] combines separate regression models for the outcome and mediator:

$$\mathbb{E}[Y(x,m)] = \theta_0 + \theta_1 x + \theta_2 m + \theta_3 xm + \theta'_4 \mathbf{c}$$
$$\mathbb{E}[M(x)] = \gamma_0 + \gamma_1 x + \gamma'_2 \mathbf{c}.$$

Parameters from these models are combined to obtain the natural direct effect, defined as:

$$NDE(x^*) = (\theta_1 + \theta_3\gamma_0 + \theta_3\gamma_1 x^* + \theta_3\gamma'_2 c)(x - x^*).$$

("world 1" ↓ ; ↑ "world 2")

When the above expression is factored out, it yields a term $\theta_3\gamma_1 x^* x$ that immediately reveals the cross-world nature of the natural direct effect, akin to an interaction between the exposure at its reference value in one world (x^*), and the exposure at its exposed value in another world (x). The same scenario plays out for the natural indirect effect as well. This and other methods to estimate natural direct and indirect effects ignore the philosophical problem of estimating quantities that require such cross-world interactions, and replace them with much more tractable mathematical problems of estimation.[47] Estimating these quantities raises concern as to whether they are no more than mathematical constructs, without meaningful correspondence to the world.[35] Further still is the added problem that we can only ask (and answer) such questions absent any empirical evidence in support or against our beliefs about the meaning of such quantities. Such beliefs can never be confirmed or refuted using empirical data because natural direct and indirect effects cannot be estimated via randomized intervention trials.[3,44,46]

Policy relevance

An additional issue is the disconnect that exists between natural direct and indirect effects and policy. Suppose we are interested in estimating the mediating role of screening mammography in the relation between hormone replacement therapy and breast cancer. As pointed out in Joffe *et al.*,[56] women taking hormone replacement therapy are under more medical supervision, and thus subject to more mammographic screening. As a result, more breast cancer is observed among women taking hormone replacement therapy, and some portion of the excess cases is due to this increased detection. But how much? At face value, natural direct and indirect effects may seem worthwhile to pursue in this setting. A natural indirect effect risk ratio of 0.2 would tell us that the breast cancer risk under the mammographic screening that a woman would have had under postmenopausal hormone therapy is one-fifth the risk of breast cancer under the mammographic screening that a woman would have had under no postmenopausal hormone therapy. Whereas one can implement clinical or public health policy decisions to perform mammography for some well-defined set of women, there is no way to define an intervention that results in unexposed women undergoinging the mammography screening that they would have undergone had they been exposed to postmenopausal hormone therapy (and vice versa). Thus, natural effects cannot be used as a basis for policy, whether individual, clinical or population-level. Such an approach to epidemiological inference stands at odds with the consistently emphasized *raison d'etre* of our field: to intervene, to change elements of the world and to improve the population's health.[27–31]

Discussion

In this paper, we have highlighted the salient issues in the longstanding debate on the relevance of natural direct and indirect effects. We reviewed a number of characteristics of natural direct and indirect effects that often receive little or no attention in the epidemiological literature. Natural direct and indirect effects: (i) require a union of states that can never occur together in the world (states that are logically incompatible); (ii) require independence assumptions to hold between these logically incompatible states; and (iii) can never be interpreted so as to inform individual, clinical or public health interventions. This last condition we believe markedly curtails the usefulness of natural direct and indirect effects, especially in light of the extensive history of calls for pragmatic epidemiological research.

The debate on the relevance of natural direct and indirect effects in epidemiology has been going on for some time. Much of it boils down to whether one takes as a target of inference the information in the mathematical object defining such effects, or the change in the world that might be brought about by acting on this information. The former objective is often encountered as the claim that the primary purpose of mediation analysis is to understand epidemiological mechanisms.[37,47] We question this claim. An analysis with ambiguous implications for clinical or public health action appears to have limited use from an epidemiological perspective. Rather than estimating natural effects, epidemiologists may be better off estimating controlled direct effects.[4] These effects correspond to the exposure effect that would remain after an intervention that sets the mediator to a specific level. Such effects are well suited to epidemiology conducted in the context of public health research, in that they seek to understand the change in the outcome that might be brought about by intervening on the exposure and mediator under study.

Interestingly, various attempts at clearing up the ambiguities inherent in natural direct and indirect effects have only reinforced the importance of controlled direct effects. Natural effects are often informally described by invoking metaphors of 'freezing' the mediator,[44] or 'disabling' or 'de-activating'[4] the arrow from the exposure to the mediator. Invariably, more detail on how one might go about freezing a mediator, or disabling or de-activating an arrow, leads to the introduction of a separate variable along the exposure-mediator path, whose value is set to a certain level. Whereas it might be argued that this conceptual innovation preserves the 'natural' relation between the exposure and the mediator, it does not yield a natural direct effect, but yields a controlled direct effect in which the newly introduced variable is intervened upon and set to a given level. Indeed, general conditions have been provided proving one can interpret a natural direct effect on a given DAG as a controlled direct effect in an expanded DAG in which nodes are added along the path between the exposure and mediator.[5,43]

We have discussed many issues involved in estimating natural direct and indirect effects. We have ignored, however, a number of related and important issues. In particular, the notion that causation cannot be inferred without manipulation is not without its critics.[57] Although non-manipulable exposures (such as race or sex) undoubtedly play an important role in influencing health outcomes, counterfactual theory has yet to accommodate such exposures. Moreover, as mentioned in the introduction, even though controlled direct effects are more relevant to epidemiological research, and indirect effects are of general interest in the field, there is no simple and general way to define controlled indirect effects. These limitations are currently the subject of active research.[58] Finally, our criticism of natural direct and indirect effects is possible because they are rigorously defined mathematical objects—a strength not shared by all targets of epidemiological inference.[35] Our intent was to highlight that this strength does not necessarily imbue natural direct and indirect effects with policy-relevant qualities.

In epidemiology, natural direct and indirect effects are becoming effect estimates of choice in applied mediation analyses. The merits of this choice have not been reconciled with the repeated calls for a more consequential epidemiology.

Funding

Canada Research Chairs Program grant to J.S.K.

Acknowledgements

We thank Miguel Hernán for helpful comments on previous drafts.

Conflict of interest: None declared.

References

1. Judd CM, Kenny DA. Process analysis: Estimating mediation in treatment evaluations. *Eval Rev* 1981;**5**:602–19.
2. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986;**51**:1173–82.
3. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992;**3**:143–55.
4. Pearl J. Direct and indirect effects. In: *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann, 2001.
5. Robins JM. Semantics of causal DAG models and the identification of direct and indirect effects. In: Green PJ, Hjort NL, Richardson S (eds). *Highly Structured Stochastic Systems*. Oxford, UK: Oxford University Press, 2003.
6. VanderWeele TJ. Marginal structural models for direct and indirect effects. *Epidemiology* 2009;**20**:18–26. See also erratum in *Epidemiology* 2009;**20**:629.
7. Ten Have TR, Joffe MM, Lynch KG, Brown GK, Maisto SA, Beck AT. Causal mediation analyses with rank preserving models. *Biometrics* 2007;**63**:926–34.
8. Hafeman DM, Schwartz S. Opening the black box: a motivation for the assessment of mediation. *Int J Epidemiol* 2009;**38**:838–45.
9. Hafeman DM. 'Proportion explained': a causal interpretation for standard measures of indirect effect? *Am J Epidemiol* 2009;**170**:1443–48.
10. Hafeman DM. A sufficient cause based approach to the assessment of mediation. *Eur J Epidemiol* 2008;**23**:711–21.
11. Ten Have TR. Invited commentary: pushing the mediation envelope. *Am J Epidemiol* 2010;**172**:1352–54.
12. Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. *Am J Epidemiol* 2012;**176**:190–95.
13. Daniels MJ, Roy JA, Kim C, Hogan JW, Perri MG. Bayesian inference for the causal effect of mediation. *Biometrics* 2012;**68**:1028–36.
14. Albert JM, Nelson S. Generalized causal mediation analysis. *Biometrics* 2011;**67**:1028–38.
15. Kaufman JS, Maclehose RF, Kaufman S. A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. *Epidemiol Perspect Innov* 2004;**1**:4.
16. VanderWeele TJ. A three-way decomposition of a total effect into direct, indirect, and interactive effects. *Epidemiology* 2013;**24**:224–32.
17. Albert JM. Distribution-free mediation analysis for nonlinear models with confounding. *Epidemiology* 2012;**23**:879–88.
18. Lange T, Hansen JV. Direct and indirect effects in a survival context. *Epidemiology* 2011;**22**:575–81.
19. Albert JM. Mediation analysis via potential outcomes models. *Stat Med* 2008;**27**:1282–304.
20. Ten Have TR, Joffe MM. A review of causal estimation of effects in mediation analyses. *Stat Methods Med Res* 2012;**21**:77–107.
21. Ananth CV, VanderWeele TJ. Placental abruption and perinatal mortality with preterm delivery as a mediator: disentangling direct and indirect effects. *Am J Epidemiol* 2011;**174**:99–108.
22. Nandi A, Glymour MM, Kawachi I, VanderWeele TJ. Using marginal structural models to estimate the direct effect of

adverse childhood social conditions on onset of heart disease, diabetes, and stroke. *Epidemiology* 2012;**23**:223–32.
23. Bennett AC, Rankin KM, Rosenberg D. Does a medical home mediate racial disparities in unmet healthcare needs among children with special healthcare needs? *Matern Child Health J* 2012;**16**(**Suppl 2**):330–38.
24. VanderWeele TJ, Lauderdale DS, Lantos JD. Medically induced preterm birth and the associations between prenatal care and infant mortality. *Ann Epidemiol* 2013;**23**:435–40.
25. Subbaraman MS, Lendle S, van der Laan M, Kaskutas LA, Ahern J. Cravings as a mediator and moderator of drinking outcomes in the COMBINE study. *Addiction* 2013; **108**:1737–44.
26. Smith PM, Smith BT, Mustard CA, Lu H, Glazier RH. Estimating the direct and indirect pathways between education and diabetes incidence among Canadian men and women: a mediation analysis. *Ann Epidemiol* 2013;**23**:143–49.
27. Galea S. An argument for a consequentialist epidemiology. *Am J Epidemiol* 2013;**178**:1185–91.
28. Glass TA, Goodman SN, Hernán MA, Samet JM. Causal inference in public health. *Annu Rev Public Health* 2013;**34**: 61–75.
29. Ahern J, Hubbard A, Galea S. Estimating the effects of potential public health interventions on population disease burden: a step-by-step illustration of causal inference methods. *Am J Epidemiol* 2009;**169**:1140–47.
30. Rockhill B. Theorizing about causes at the individual level while estimating effects at the population level: implications for prevention. *Epidemiology* 2005;**16**:124–29.
31. Kaufman JS, Cooper RS. Seeking causal explanations in social epidemiology. *Am J Epidemiol* 1999;**150**:113–20.
32. Holland PW. Statistics and causal inference. *J Am Stat Assoc* 1986;**81**:945–60.
33. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period–application to control of the healthy worker survivor effect. *Mathematical Modelling* 1986;**7**:1393–512.
34. Robins JM, Greenland S. Causal inference without counterfactuals: Comment. *J Am Stat Assoc* 2000;**95**:431–35.
35. Hernán MA. Invited commentary: Hypothetical interventions to define causal effects—afterthought or prerequisite? *Am J Epidemiol* 2005;**162**:618–20.
36. Kaufman JS. Commentary: Gilding the black box. *Int J Epidemiol* 2009;**38**:845–47.
37. Vansteelandt S. Estimation of controlled direct effects on a dichotomous outcome using logistic structural direct effect models. *Biometrika* 2010;**97**:921–34.
38. Pearl J. *Causality: Models, Reasoning and Inference*. 2nd edn. Cambridge, UK: Cambridge University Press, 2009.
39. Schwartz S, Hafeman D, Campbell U, Gatto N. Authors' response to: Commentary: Gilding the black box. *Int J Epidemiol* 2010;**39**:1399–401.
40. VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology* 2014;**25**:300–06.
41. Tchetgen Tchetgen EJ, VanderWeele TJ. Identification of natural direct effects when a confounder of the mediator is directly affected by exposure. *Epidemiology* 2014;**25**:282–91.
42. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol* 2013;**42**:1511–19.
43. Robins J, Richardson T. Alternative graphical causal models and the identification of direct effects. In: Keyes KM, Ornstein K, Shrout PE (eds). *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*. Oxford, UK: Oxford University Press, 2011.
44. Pearl J. Interpretation and identification in causal mediation analysis. *Psychol Methods* 2014. In press.
45. Avin C, Shpitser I, Pearl J. Identifiability of path specific effects. In: *Proceedings of the International Joint Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press, 2005.
46. Richardson TS, Robins JM. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. Tech. Rep. Number 128. Center for Statistics and the Social Sciences, University of Washington. http://www.csss.washington.edu/Papers/wp128.pdf 26 August 2013, date last accessed).
47. Vansteelandt S. Estimation of direct and indirect effects. In: Carlo Berzuini LB, Philip Dawid, Bernardinelli L *et al*. (eds). *Causality: Statistical Perspectives and Applications*. Chichester, UK: Wiley, 2012.
48. Rosenbaum P. *Observational Studies*. New York, NY: Springer, 1995.
49. Morgan SL, Winship C. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York, NY: Cambridge University Press, 2007.
50. Hernán MA, Robins J. *Causal Inference*. New York, NY: Chapman/Hall/CRC. In press. http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book (12 March 2014, date last accessed).
51. Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. *Epidemiology* 2006;**17**:276–84.
52. Robins JM. Testing and estimation of direct effects by reparameterizing directed acyclic graphs with structural nested models. In: Glymour C, Cooper G (eds). *Computation, Causation, and Discovery*. Menlo Park, CA / Cambridge, MA: AAAI Press/The MIT Press., 1999.
53. Kaufman JS. Invited commentary: Decomposing with a lot of supposing. *Am J Epidemiol* 2010;**172**:1349–51.
54. Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. *Stat Sci* 2012;**25**:51–71.
55. Shpitser I. Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding. *Cogn Sci* 2013;**37**:1011–35.
56. Joffe MM, Byrne C, Colditz GA. Postmenopausal hormone use, screening, and breast cancer: characterization and control of a bias. *Epidemiology* 2001;**12**:429–38.
57. Bollen KA, Pearl J. Eight myths about causality and structural equation models. In: *Handbook of Causal Analysis for Social Research*. New York, NY: Springer, 2013.
58. VanderWeele TJ, Hernán MA. Causal effects and natural laws: Towards a conceptualization of causal counterfactuals for nonmanipulable exposures, with applications to the effects of race and sex. In: Berzuini C, Dawid AP, Bernardinelli L (eds). *Causality: Statistical Perspectives and Applications*. Chichester, UK: Wiley, 2012.