PATHWAYS

Mediation analysis for life course epidemiology

Bianca L De Stavola
LSHTM Centre for Statistical Methodology
(with Rhian Daniel)

CLOSER Seminar Series,26 March 2015

Website http://pathways.lshtm.ac.uk
Email pathways@lshtm.ac.uk
Twitter @pathwaysNCRM

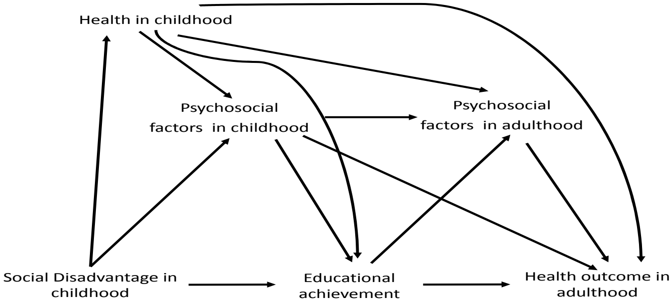
THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

NCRM National Centre for Research Methods

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

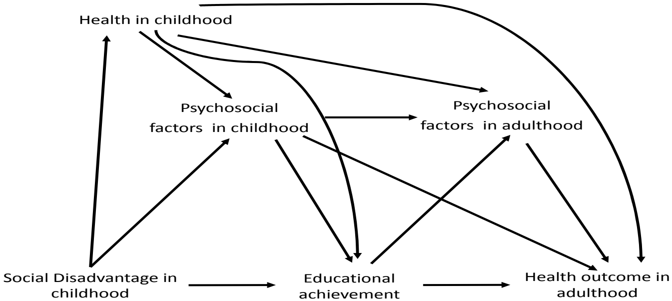
Mediation

In life course studies we focus on distal exposures (*e.g.* social disadvantage) for later life outcomes:



Mediation

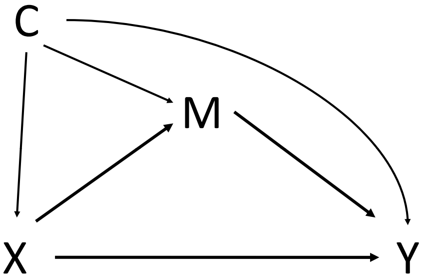
In life course studies we focus on distal exposures (*e.g.* social disadvantage) for later life outcomes:



Interest: disentangle the underlying processes.

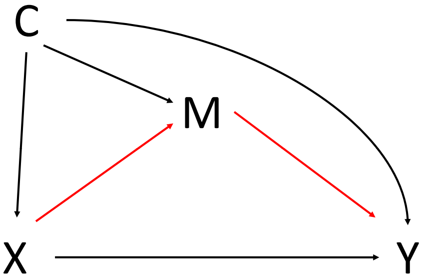
Keeping it simple

Exposure X, mediator M, outcome Y and confounders C.



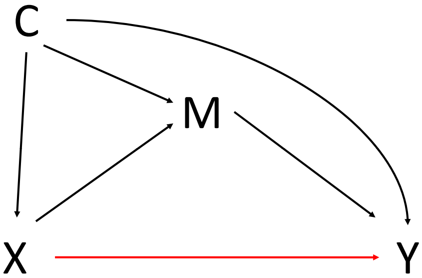
Keeping it simple

Exposure X, mediator M, outcome Y and confounders C.
Mediation leads to separate the two pathways: via M (indirect)



Keeping it simple

Exposure X, mediator M, outcome Y and confounders C.
Mediation leads to separate the two pathways: via M (indirect) and not via M (direct).

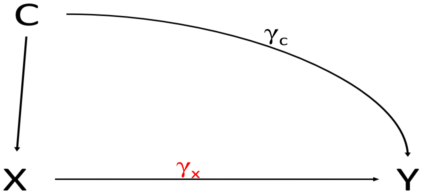


Outline

1. Standard approach
2. A more general approach
3. Example: ED in adolescent girls
4. Summary

Standard approach to mediation analysis

For Y a continuous outcome, *e.g.* birth weight:



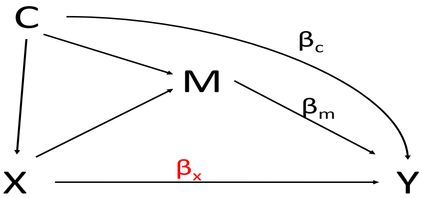
- Regress Y on the exposure X and confounding factors C:

$$E(Y|X,C) = \gamma_0 + \gamma_x X + \gamma_c C$$

γ_x this is interpreted as the total effect.

Standard approach to mediation analysis

For Y a continuous outcome, *e.g.* birth weight:



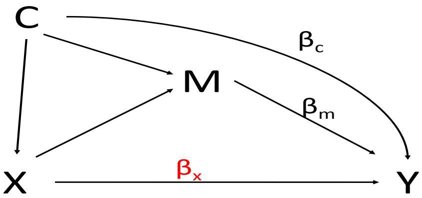
- Adding the mediator M and expanding the model for Y:

$$E\left(Y|X, M, C\right) = \beta_0 + \beta_x X + \beta_m M + \beta_c C$$

β_x is interpreted as the direct effect.

Standard approach to mediation analysis

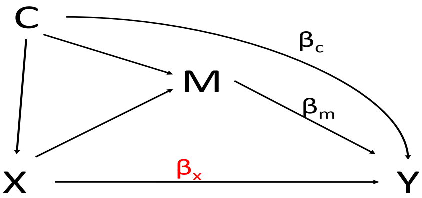
For Y a continuous outcome, *e.g.* birth weight:



- The difference between total and direct effect, $\delta_x = \gamma_x - \beta_x$, is interpreted as the indirect effect.

Standard approach to mediation analysis

For Y a continuous outcome, *e.g.* birth weight:



- The difference between total and direct effect, $\delta_x = \gamma_x - \beta_x$, is interpreted as the indirect effect.
- This can also be derived using the multiplication of effects method (as defined in SEMs).

Four limitations of the standard approach

1. If the model were:

$$E\left(Y|X,M,C\right) = \beta_0 + \beta_x X + \beta_m M + \beta_{xm} XM + \beta_c C$$

the earlier partitioning would not work.

Four limitations of the standard approach

1. If the model were:

$$E\left(Y|X,M,C\right) = \beta_0 + \beta_x X + \beta_m M + \beta_{xm} XM + \beta_c C$$

   the earlier partitioning would not work.

2. The partitioning is defined and works only for simple linear regression models.

Four limitations of the standard approach

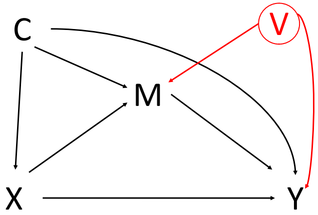
1. If the model were:

$$E\left(Y|X,M,C\right) = \beta_0 + \beta_x X + \beta_m M + \beta_{xm} XM + \beta_c C$$

   the earlier partitioning would not work.

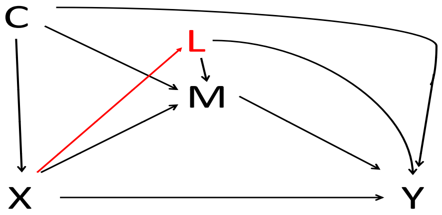
2. The partitioning is defined and works only for simple linear regression models.

3. It is not generally recognized that if there may be unaccounted confounders V of the M–Y relationship:



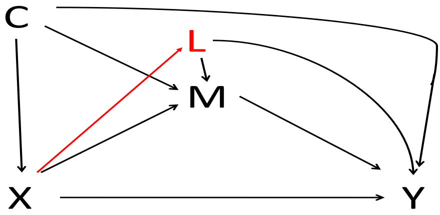
Four limitations of the standard approach (cont'd)

4. If a measured confounder is like L, *i.e.* a variable that is a consequence of X (*i.e.* intermediate confounder):



Four limitations of the standard approach (cont'd)

4. If a measured confounder is like L, *i.e.* a variable that is a consequence of X (*i.e.* intermediate confounder):



$$E(Y|X, M, C) \quad = \quad \beta_0 + \beta_x X + \beta_m M + \beta_c C + \beta_l L$$

β_x would not measure the direct effect of X:
the effect of X that is not mediated by M (the direct effect) includes $X \to L \to Y$, but controlling for L removes it!

A more general approach to mediation analysis

The causal inference framework

- The causal inference literature on mediation provides general definitions of direct and indirect effects that:
  - Do not depend on the specification of a particular statistical model.

A more general approach to mediation analysis

The causal inference framework

- The causal inference literature on mediation provides general definitions of direct and indirect effects that:
    - Do not depend on the specification of a particular statistical model.
      Hence they can deal with any type of outcomes and mediators and do not restrict the relationships to be linear.

A more general approach to mediation analysis

The causal inference framework

- The causal inference literature on mediation provides general definitions of direct and indirect effects that:
  - Do not depend on the specification of a particular statistical model.
    Hence they can deal with any type of outcomes and mediators and do not restrict the relationships to be linear.
  - Requires quantities that are not all observable: *potential outcomes* and *potential mediators*.
    This has the advantage of forcing the user to be explicit about the underlying assumptions.

A more general approach to mediation analysis

The causal inference framework

- The causal inference literature on mediation provides general definitions of direct and indirect effects that:
  - Do not depend on the specification of a particular statistical model.
    Hence they can deal with any type of outcomes and mediators and do not restrict the relationships to be linear.
  - Requires quantities that are not all observable: *potential outcomes* and *potential mediators*.
    This has the advantage of forcing the user to be explicit about the underlying assumptions.

A more general approach to mediation analysis

The causal inference framework

- The causal inference literature on mediation provides general definitions of direct and indirect effects that:
  - Do not depend on the specification of a particular statistical model.
    Hence they can deal with any type of outcomes and mediators and do not restrict the relationships to be linear.

Potential outcomes

- $Y(x)$: the potential values of Y that would have occurred had X been set, possibly counter to fact, to the value x.
- $M(x)$: the potential values of M that would have occurred had X been set, possibly counter to fact, to the value x.
- $Y(x, m)$: the potential values of Y that would have occurred had X been set, possibly counter to fact, to the value x and M to m.

Total Causal Effect (TCE): definition

The average total causal effect of X, comparing exposure level $X = 1$ to $X = 0$, can be defined as the linear contrast:

$$TCE = E[Y(1)] - E[Y(0)]$$

This is a comparison of two hypothetical worlds: in the first, X is set to 1, and in the second X is set to 0.

We are working throughout on the mean difference scale... alternatives exist.

Mediation parameters in causal inference

(a) The Controlled direct effect (CDE(m)):

$$CDE(m) \;=\; E\left[Y\left(1, m\right)\right] - E\left[Y\left(0, m\right)\right]$$

It is a comparison of two hypothetical worlds:

- In the first, X is set to 1, and in the second X is set to 0.
- In both worlds, M is set to m.
- By keeping M fixed at m, CDE(m) is the direct effect of X, unmediated by M (in general it varies with m).

Mediation parameters in causal inference (cont'd)

(b) The Natural Direct Effect (NDE):

$$NDE = E[Y(1, M(0))] - E[Y(0, M(0))]$$

It is a comparison of two hypothetical worlds:

- In the first, X is set to 1, and in the second X is set to 0.
- In both worlds, M is set to the natural value $M(0)$, *i.e.* the value it would take if X were set to 0.
- Since M is the same (*within* individual) in both worlds, we are still getting at the direct effect of X, unmediated by M.

Mediation parameters in causal inference (cont'd)

(b) The Natural Direct Effect (NDE):

$$NDE \;=\; E[Y(1, M(0))] - E[Y(0, M(0))]$$

It is a comparison of two hypothetical worlds:

- In the first, X is set to 1, and in the second X is set to 0.
- In both worlds, M is set to the natural value $M\,(0)$, *i.e.* the value it would take if X were set to 0.
- Since M is the same (*within* individual) in both worlds, we are still getting at the direct effect of X, unmediated by M.

(c) The Natural Indirect Effect (NIE):

$$NIE = TCE - NDE$$

Mediation parameters in causal inference (cont'd)

(b) The Natural Direct Effect (NDE):

$$NDE = E[Y(1, M(0))] - E[Y(0, M(0))]$$

It is a comparison of two hypothetical worlds:

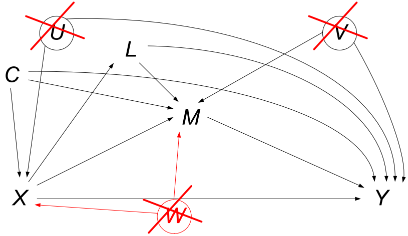
- In the first, X is set to 1, and in the second X is set to 0.
- In both worlds, M is set to the natural value $M(0)$, *i.e.* the value it would take if X were set to 0.
- Since M is the same (*within* individual) in both worlds, we are still getting at the direct effect of X, unmediated by M.

(c) The Natural Indirect Effect (NIE):

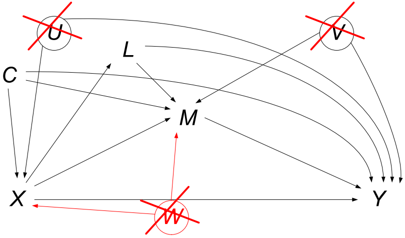
$$NIE = TCE - NDE$$

Will focus on these natural effects

Assumptions for estimation of natural effects

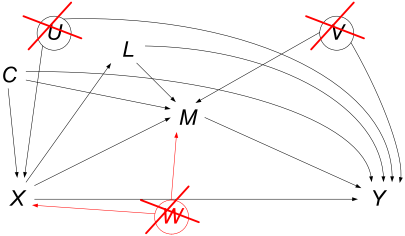


Assumptions for estimation of natural effects



- In addition, either:
    - no intermediate confounding,
    - or some model restrictions.

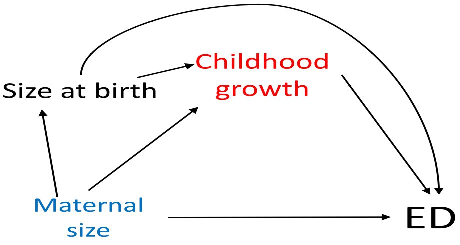
Assumptions for estimation of natural effects



- In addition, either:
  - no intermediate confounding,
  - or some model restrictions.
- Estimation: choice of fully parametric or semi-parametric approaches.

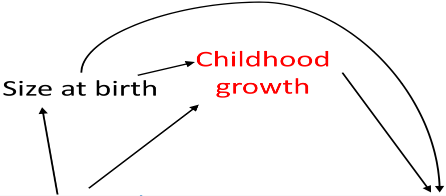
Eating disorders (ED) in adolescent girls

- ED comprise a variety of heterogeneous diseases
- Maternal body size is a possible risk factor
- Childhood growth may act as mediator (with size at birth an intermediate confounder).



Eating disorders (ED) in adolescent girls

- ED comprise a variety of heterogeneous diseases
- Maternal body size is a possible risk factor
- Childhood growth may act as mediator (with size at birth an intermediate confounder).



What effect does intervening on maternal BMI have on the child's ED symptoms in a world where maternal BMI has no effect on her child growth trajectory?

The ALSPAC Study

- Cohort of children born in 1991-92 in SW England, followed from birth at set intervals; 5,000 girls.

The ALSPAC Study

- Cohort of children born in 1991-92 in SW England, followed from birth at set intervals; 5,000 girls.
- Outcomes: 3 types of ED symptoms scores, derived from parental reports collected when child was 13.5y (Micali *et al.* 2014):
  - *“Binge eating”*,
  - *“Fear of weight gain”*,
  - *“Food Restriction”*

The ALSPAC Study

- Cohort of children born in 1991-92 in SW England, followed from birth at set intervals; 5,000 girls.
- Outcomes: 3 types of ED symptoms scores, derived from parental reports collected when child was 13.5y (Micali *et al.* 2014):
  - *"Binge eating"*,
  - *"Fear of weight gain"*,
  - *"Food Restriction"*
- Exposure: pre-pregnancy maternal BMI ($< 18.5, 18.5 - 25.0, > 25.0\mathrm{kg/m}^2$).

The ALSPAC Study

- Cohort of children born in 1991-92 in SW England, followed from birth at set intervals; 5,000 girls.
- Outcomes: 3 types of ED symptoms scores, derived from parental reports collected when child was 13.5y (Micali *et al.* 2014):
  - *"Binge eating"*,
  - *"Fear of weight gain"*,
  - *"Food Restriction"*
- Exposure: pre-pregnancy maternal BMI $(< 18.5, 18.5 - 25.0, > 25.0\mathrm{kg/m}^2)$.
- Bivariate Mediator: BMI at 7y and BMI velocity at 7-12y.

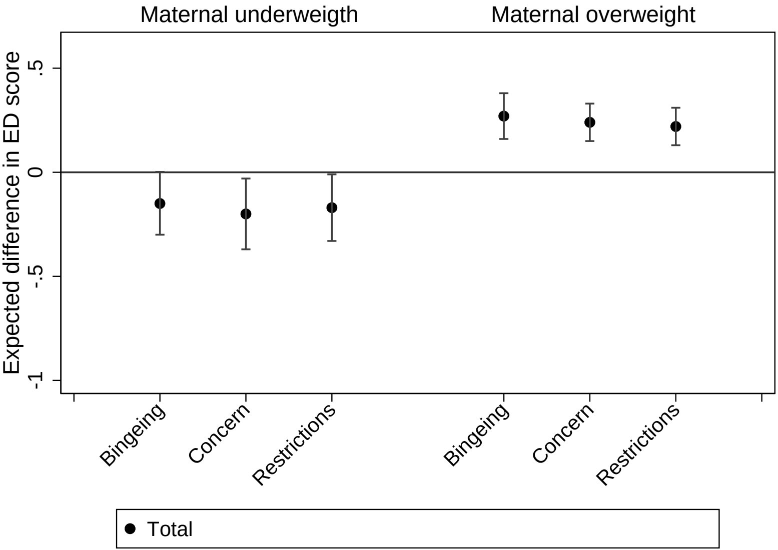
The ALSPAC Study

- Cohort of children born in 1991-92 in SW England, followed from birth at set intervals; 5,000 girls.
- Outcomes: 3 types of ED symptoms scores, derived from parental reports collected when child was 13.5y (Micali *et al.* 2014):
  - *"Binge eating"*,
  - *"Fear of weight gain"*,
  - *"Food Restriction"*
- Exposure: pre-pregnancy maternal BMI ($< 18.5, 18.5 - 25.0, > 25.0\text{kg/m}^2$).
- Bivariate Mediator: BMI at 7y and BMI velocity at 7-12y.

- Background confounders: pre-pregnancy maternal psychopathology, maternal age, education, social class.
- Assumptions: No unmeasured confounding of the X–Y, X–M, M–Y relations; no X–M interactions.
- Estimation: Fully-parametric via Monte Carlo simulation (with imputation and bootstrapped SEs).

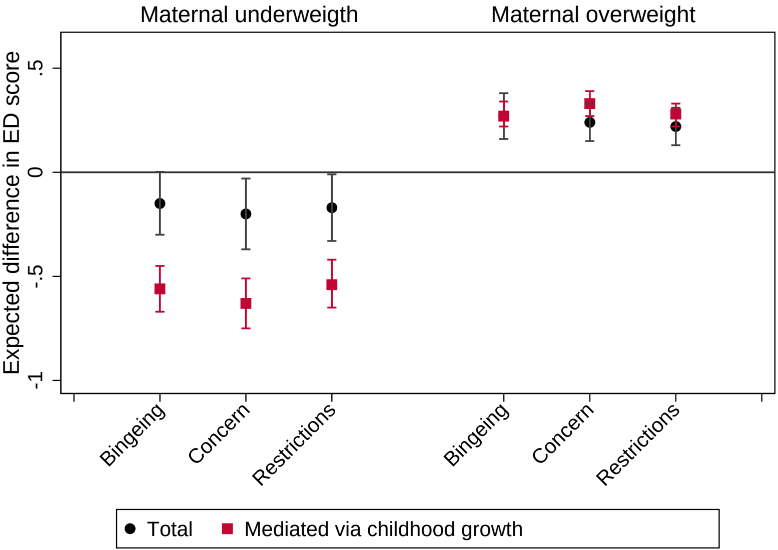
Results

N=3,526



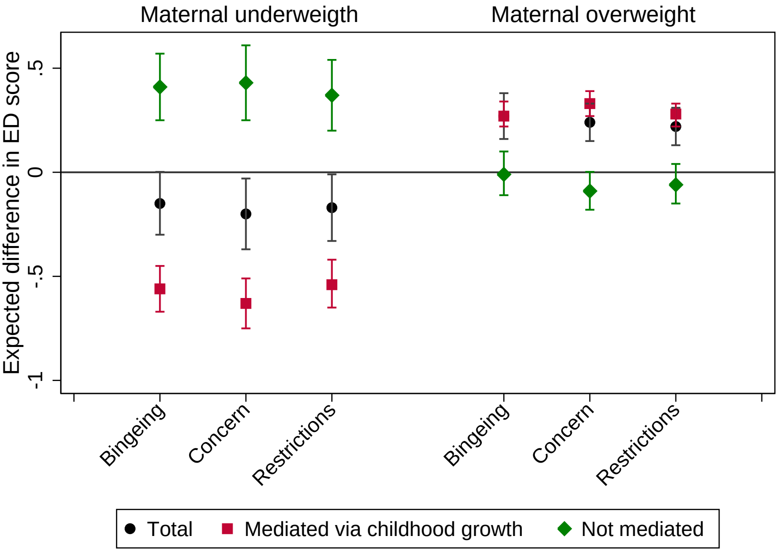
Results

N=3,526



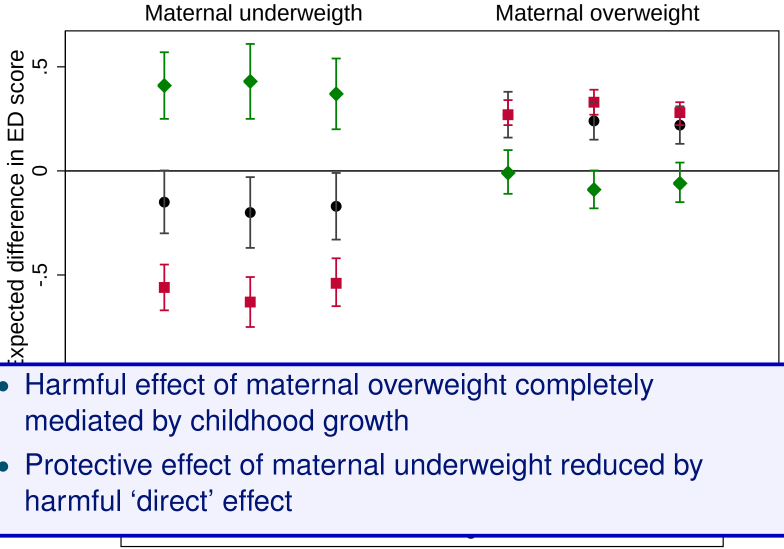
Results

N=3,526



Results

N=3,526



- Harmful effect of maternal overweight completely mediated by childhood growth
- Protective effect of maternal underweight reduced by harmful ‘direct’ effect

Summary

- Life course epidemiology involves studying processes, hence understanding issues in the field of mediation analysis is essential.

Summary

- Life course epidemiology involves studying processes, hence understanding issues in the field of mediation analysis is essential.
- Two main approaches for the study of mediation:
  - the traditional approach suffers from several limitations
  - modern causal inference is very general, but also more formal and hence harder to grasp.

Summary

- Life course epidemiology involves studying processes, hence understanding issues in the field of mediation analysis is essential.
- Two main approaches for the study of mediation:
  - the traditional approach suffers from several limitations
  - modern causal inference is very general, but also more formal and hence harder to grasp.
- Worth learning about the causal approach if research aims at studying pathways: pretending it is not difficult is not a safe solution!

Summary

- Life course epidemiology involves studying processes, hence understanding issues in the field of mediation analysis is essential.
- Two main approaches for the study of mediation:
  - the traditional approach suffers from several limitations
  - modern causal inference is very general, but also more formal and hence harder to grasp.
- Worth learning about the causal approach if research aims at studying pathways: pretending it is not difficult is not a safe solution!

Thank you!

References

- Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 1986; 51, 1173-1182.
- Daniel RM, De Stavola BL, and Cousens SN. gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. *Stata Journal* 2011; 11: 479–517.
- Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychological Methods* 2010; 15, 309-334.
- Pearl J. Direct and indirect effects. *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence* 2001; San Francisco: Morgan Kaufmann.
- Vansteelandt S. Estimation of direct and indirect effects (chapter 4.2). In *Causality: Statistical Perspectives and Applications*, Berzuini C, Dawid AP, Bernardinelli L (eds). Wiley, 2011.