Maternal characteristics, childhood growth and eating disorder: a study of mediation using `gformula`

Bianca De Stavola and Rhian Daniel

London School of Hygiene and Tropical Medicine, UK

IX Convegno Italiano Utenti Stata, Bologna
20-21 September 2012

Early exposures and distal outcomes

In epidemiology we often focus on earlier exposures for distal outcomes:

Social Disadvantage in childhood → Health outcome in adulthood

Early exposures and distal outcomes

In epidemiology we often focus on earlier exposures for distal outcomes:

Social Disadvantage in childhood → Health outcome in adulthood

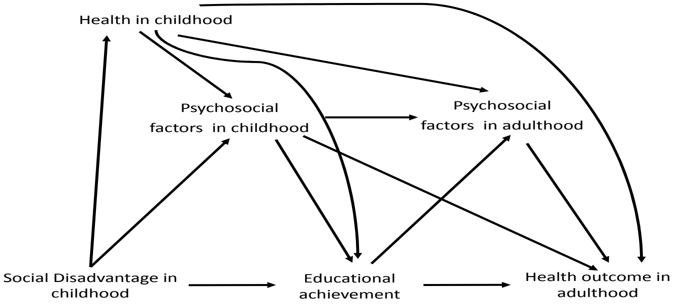
and this leads to studying the underlying processes, *e.g.*:

Early exposures and distal outcomes

In epidemiology we often focus on earlier exposures for distal outcomes:

Social Disadvantage in childhood $\longrightarrow$ Health outcome in adulthood

and this leads to studying the underlying processes, *e.g.*:

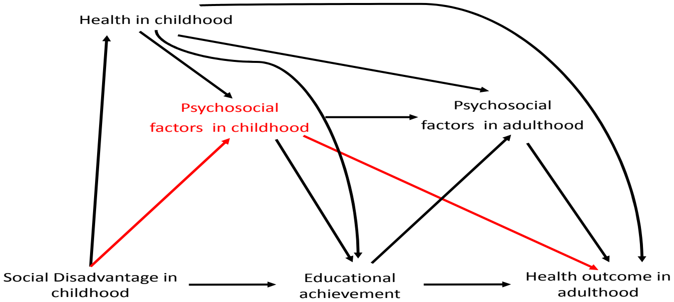


Early exposures and distal outcomes

In epidemiology we often focus on earlier exposures for distal outcomes:

Social Disadvantage in childhood ⟶ Health outcome in adulthood

and this leads to studying the underlying processes, *e.g.*:

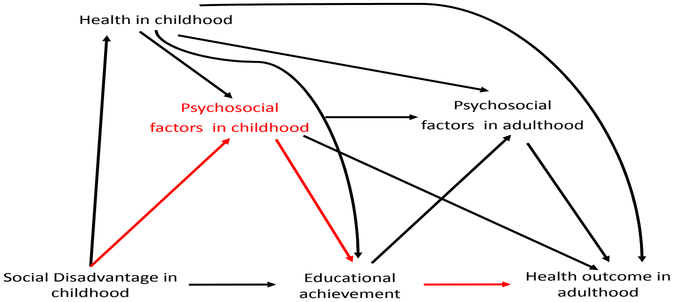


Early exposures and distal outcomes

In epidemiology we often focus on earlier exposures for distal outcomes:

Social Disadvantage in childhood → Health outcome in adulthood

and this leads to studying the underlying processes, *e.g.*:

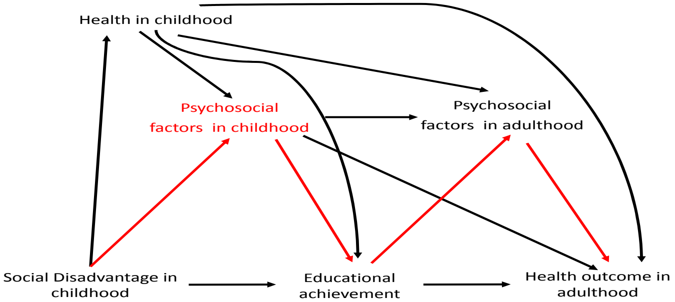


Early exposures and distal outcomes

In epidemiology we often focus on earlier exposures for distal outcomes:

Social Disadvantage in childhood ⟶ Health outcome in adulthood

and this leads to studying the underlying processes, *e.g.*:

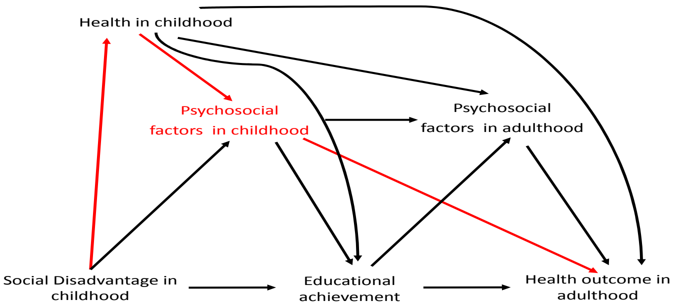


Early exposures and distal outcomes

In epidemiology we often focus on earlier exposures for distal outcomes:

Social Disadvantage in childhood ⟶ Health outcome in adulthood

and this leads to studying the underlying processes, *e.g.*:

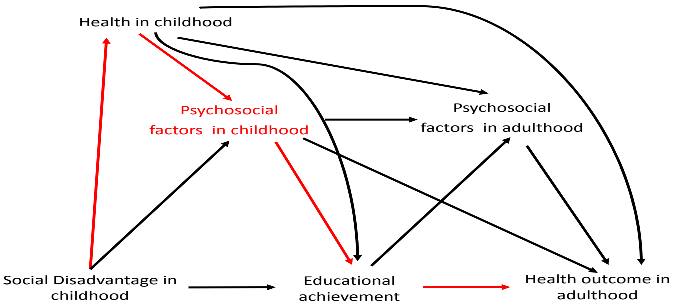


Early exposures and distal outcomes

In epidemiology we often focus on earlier exposures for distal outcomes:

Social Disadvantage in childhood → Health outcome in adulthood

and this leads to studying the underlying processes, *e.g.*:

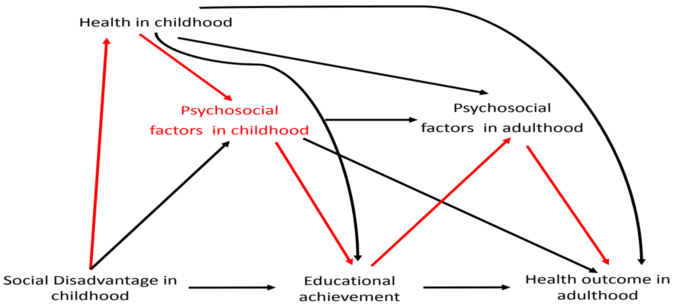


Early exposures and distal outcomes

In epidemiology we often focus on earlier exposures for distal outcomes:

Social Disadvantage in childhood ⟶ Health outcome in adulthood

and this leads to studying the underlying processes, *e.g.*:

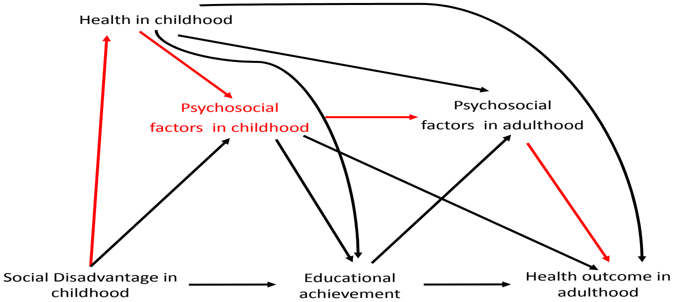


Early exposures and distal outcomes

In epidemiology we often focus on earlier exposures for distal outcomes:

Social Disadvantage in childhood → Health outcome in adulthood

and this leads to studying the underlying processes, *e.g.*:



Mediation

- Focus on how the effect of an exposure is mediated via certain pathways
- Two main strands in the literature for the study of mediation:
  - Social sciences / psychometrics (Baron and Kenny, 1986)
  - Causal inference literature (Robins and Greenland, 1992; Pearl, 2001)
- They appear to be very different, but they are linked
- The second one may seem far too complex, but in fact is the one that poses fewer restrictions.

Aims of the talk:

- review the causal inference approach to mediation
- illustrate its implementation using `gformula`

Mediation

- Focus on how the effect of an exposure is mediated via certain pathways
- Two main strands in the literature for the study of mediation:
  - Social sciences / psychometrics (Baron and Kenny, 1986)
  - Causal inference literature (Robins and Greenland, 1992; Pearl, 2001)
- They appear to be very different, but they are linked
- The second one may seem far too complex, but in fact is the one that poses fewer restrictions.

Aims of the talk:

- review the causal inference approach to mediation
- illustrate its implementation using `gformula`

Mediation

- Focus on how the effect of an exposure is mediated via certain pathways
- Two main strands in the literature for the study of mediation:
  - Social sciences / psychometrics (Baron and Kenny, 1986)
  - Causal inference literature (Robins and Greenland, 1992; Pearl, 2001)
- They appear to be very different, but they are linked
- The second one may seem far too complex, but in fact is the one that poses fewer restrictions.

Aims of the talk:

- review the causal inference approach to mediation
- illustrate its implementation using gformula

Mediation

- Focus on how the effect of an exposure is mediated via certain pathways
- Two main strands in the literature for the study of mediation:
  - Social sciences / psychometrics (Baron and Kenny, 1986)
  - Causal inference literature (Robins and Greenland, 1992; Pearl, 2001)
- They appear to be very different, but they are linked
- The second one may seem far too complex, but in fact is the one that poses fewer restrictions.

Aims of the talk:

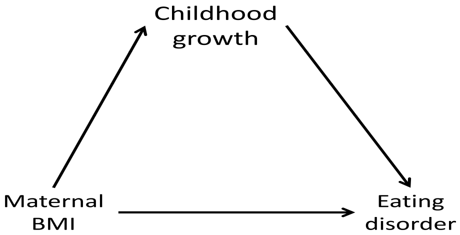
- review the causal inference approach to mediation
- illustrate its implementation using `gformula`

Outline

Outline

Motivating example
Eating disorders (ED) in adolescence

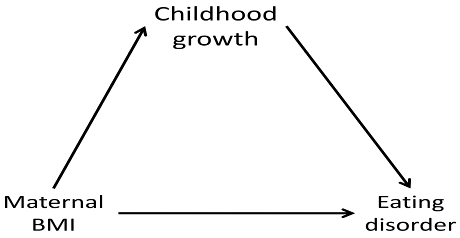
- ED comprise a variety of heterogeneous diseases
- Maternal factors possibly important (body size, education, *etc.*)
- Onset often around puberty
- Childhood growth a possible mediator



Mediation

Assuming that there is a causal effect of maternal BMI, we would like to find out, for example, how much of that effect is:

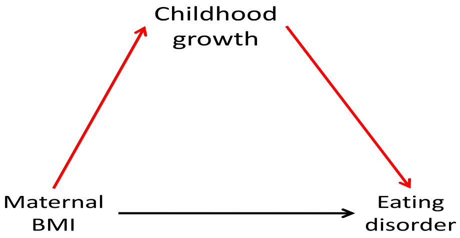
- mediated by childhood growth: the indirect effect,



Mediation

Assuming that there is a causal effect of maternal BMI, we would like to find out, for example, how much of that effect is:

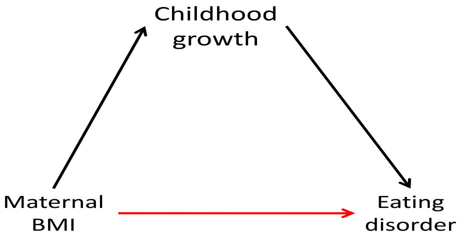
- mediated by childhood growth: the indirect effect,



Mediation

Assuming that there is a causal effect of maternal BMI, we would like to find out, for example, how much of that effect is:

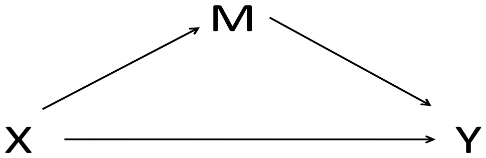
- mediated by childhood growth: the indirect effect,
- mediated via other factors: the direct effect.



Outline

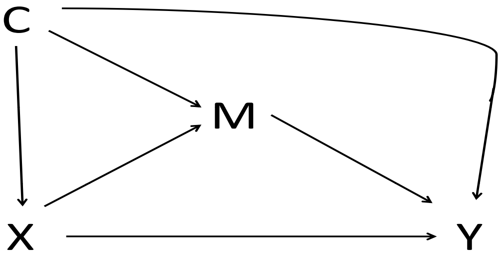
General setting

Exposure X, mediator M, outcome Y, (background) confounders C, and intermediate confounder L:



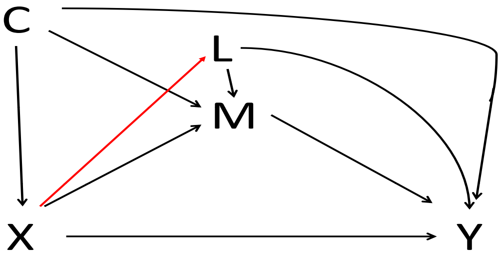
General setting

Exposure X, mediator M, outcome Y, (background) confounders C, and intermediate confounder L:



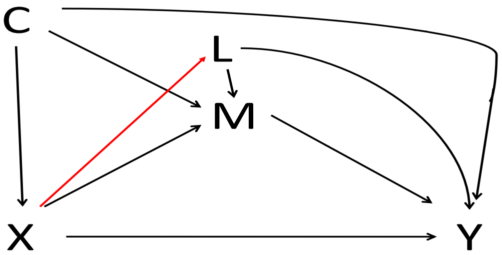
General setting

Exposure X, mediator M, outcome Y, (background) confounders C, and intermediate confounder L:



General setting

Exposure X, mediator M, outcome Y, (background) confounders C, and intermediate confounder L:



Causal inference literature:

- Many subtly different definitions of direct and indirect effect
- All involve counterfactuals (*i.e.* potential outcomes).

Potential outcomes

To make causal statements we need to compare the outcomes that would arise under different scenarios:

- $Y(x)$: 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.

Potential outcomes

To make causal statements we need to compare the outcomes that would arise under different scenarios:

- $Y(x)$: 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.

Potential outcomes

To make causal statements we need to compare the outcomes that would arise under different scenarios:

- $Y(x)$: 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.

- For simplicity consider the case where X is binary
- It also helps to start with the definition of *total causal effect*

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 [1]:

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

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.

Note that, in general: $TCE \neq E[Y|X = 1] - E[Y|X = 0]$.

[1] we are working throughout on the mean difference scale... alternatives exist

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 [1]:

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

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.

Note that, in general: $TCE \neq E[Y|X = 1] - E[Y|X = 0]$.

[1] we are working throughout on the mean difference scale... alternatives exist

Total Causal Effect (TCE): identification

To identify $TCE = E[Y(1)] - E[Y(0)]$ we need to be able to infer $E[Y(1)]$ and $E[Y(0)]$ from the observed data.

This is possible if these assumptions are satisfied:

- consistency: $Y(x)$ can be inferred from observed Y when $X = x$
- conditional exchangeability: there is no unmeasured confounding between X and Y:

Total Causal Effect (TCE): identification

To identify $TCE = E[Y(1)] - E[Y(0)]$ we need to be able to infer $E[Y(1)]$ and $E[Y(0)]$ from the observed data.

This is possible if these assumptions are satisfied:

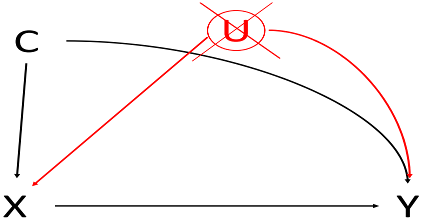
- consistency: $Y(x)$ can be inferred from observed Y when $X = x$
- conditional exchangeability: there is no unmeasured confounding between X and Y:

Total Causal Effect (TCE): identification

To identify $TCE = E[Y(1)] - E[Y(0)]$ we need to be able to infer $E[Y(1)]$ and $E[Y(0)]$ from the observed data.

This is possible if these assumptions are satisfied:

- consistency: $Y(x)$ can be inferred from observed Y when $X = x$
- conditional exchangeability: there is no unmeasured confounding between X and Y:



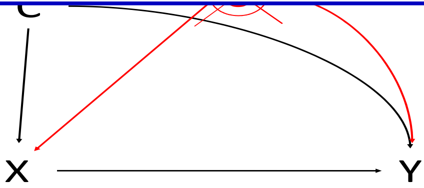
Total Causal Effect (TCE): identification

To identify $TCE = E[Y(1)] - E[Y(0)]$ we need to be able to infer $E[Y(1)]$ and $E[Y(0)]$ from the observed data.

This is possible if these assumptions are satisfied:

- consistency: $Y(x)$ can be inferred from observed Y when $X = x$
- conditional exchangeability: there is no unmeasured confounding between X and Y:

If these are satisfied, we can infer the TCE from the data



Controlled Direct Effect (CDE): definition

The average controlled direct effect of X on Y, when M is controlled at m, is:

$$CDE(m) = E[Y(1,m)] - E[Y(0,m)]$$

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.
- In both worlds, M is set to m.
- By keeping M fixed at m, we are getting at the direct effect of X, unmediated by M.
- In general $CDE(m)$ varies with m.

Controlled Direct Effect (CDE): definition

The average controlled direct effect of X on Y, when M is controlled at m, is:

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

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.
- In both worlds, M is set to m.
- By keeping M fixed at m, we are getting at the direct effect of X, unmediated by M.
- In general $CDE(m)$ varies with m.

Controlled Direct Effect (CDE): definition

The average controlled direct effect of X on Y, when M is controlled at m, is:

$$CDE(m) = E[Y(1,m)] - E[Y(0,m)]$$

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.
- In both worlds, M is set to m.
- By keeping M fixed at m, we are getting at the direct effect of X, unmediated by M.
- In general $CDE(m)$ varies with m.

Controlled Direct Effect: identification

Identification requires, as before, appropriate assumptions:

(i) consistency:
$Y = Y(x, m)$ if $X = x$ and $M = m$

(ii) sequential conditional exchangeability:

Controlled Direct Effect: identification

Identification requires, as before, appropriate assumptions:

(i) consistency:
$Y = Y(x, m)$ if $X = x$ and $M = m$

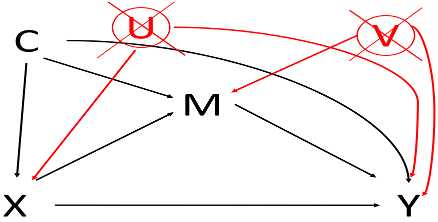
(ii) sequential conditional exchangeability:

Controlled Direct Effect: identification

Identification requires, as before, appropriate assumptions:

(i) consistency:
$Y = Y(x, m)$ if $X = x$ and $M = m$

(ii) sequential conditional exchangeability:

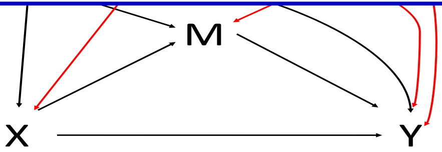


Controlled Direct Effect: identification

Identification requires, as before, appropriate assumptions:

(i) consistency:
$Y = Y(x, m)$ if $X = x$ and $M = m$

(ii) sequential conditional exchangeability:



> If these assumptions are satisfied we can infer the $CDE(m)$ from the observed data

Natural Direct Effect (NDE): definition

The average Natural Direct Effect of X on Y is:

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

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.
- 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.

Natural Direct Effect (NDE): definition

The average Natural Direct Effect of X on Y is:

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

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.
- 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.

Natural Direct Effect: identification

Identification requires, as before, appropriate assumptions:

(i) consistency
$Y = Y(x, m)$ if $X = x$ and $M = m$, $M = M(x)$ if $X = x$, and
$Y = Y\{x, M(x^*)\}$ if $X = x$ and $M = M(x^*)$.

(ii) sequential conditional exchangeability:

Natural Direct Effect: identification

Identification requires, as before, appropriate assumptions:

(i) consistency
$Y = Y(x,m)$ if $X = x$ and $M = m$, $M = M(x)$ if $X = x$, and
$Y = Y\{x, M(x^*)\}$ if $X = x$ and $M = M(x^*)$.

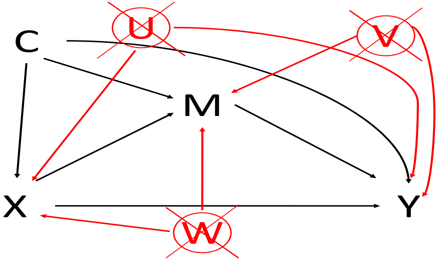
(ii) sequential conditional exchangeability:

Natural Direct Effect: identification

Identification requires, as before, appropriate assumptions:

(i) consistency
$Y = Y(x, m)$ if $X = x$ and $M = m$, $M = M(x)$ if $X = x$, and
$Y = Y\{x, M(x^*)\}$ if $X = x$ and $M = M(x^*)$.

(ii) sequential conditional exchangeability:



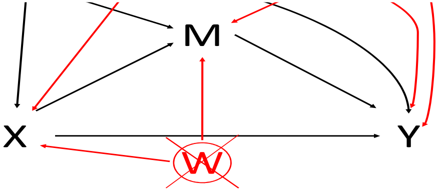
Natural Direct Effect: identification

Identification requires, as before, appropriate assumptions:

(i) consistency
$Y = Y(x, m)$ if $X = x$ and $M = m$, $M = M(x)$ if $X = x$, and
$Y = Y\{x, M(x^*)\}$ if $X = x$ and $M = M(x^*)$.

(ii) sequential conditional exchangeability:

(iii) and ... either no intermediate confounders or some restrictions on $X - M$ interactions in their effect on Y



Natural Direct Effect: identification

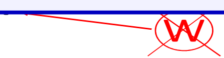
Identification requires, as before, appropriate assumptions:

(i) consistency
$Y = Y(x, m)$ if $X = x$ and $M = m$, $M = M(x)$ if $X = x$, and
$Y = Y\{x, M(x^*)\}$ if $X = x$ and $M = M(x^*)$.

(ii) sequential conditional exchangeability:

(iii) and ... <u>either</u> no intermediate confounders <u>or</u> some restrictions on $X - M$ interactions in their effect on Y

If these assumptions are satisfied: we can infer the *NDE* from the observed data
Note: the Natural Indirect Effect (*NIE*) is defined as $TCE - NDE$



Estimation

Wide range of options, for most combinations of M and Y:

- G-computation:
  - suitable for estimating $CDE(m)$ and NDE
  - can deal with intermediate confounding
  - flexible and efficient but heavy on parametric modelling assumptions
  - implemented in `gformula` (Daniel et al 2011)
- Semi-parametric methods (*e.g.*g-estimation) make fewer parametric assumptions

Estimation

Wide range of options, for most combinations of M and Y:

- G-computation:
  - suitable for estimating $CDE(m)$ and NDE
  - can deal with intermediate confounding
  - flexible and efficient but heavy on parametric modelling assumptions
  - implemented in `gformula` (Daniel et al 2011)
- Semi-parametric methods (*e.g.*g-estimation) make fewer parametric assumptions

Estimation

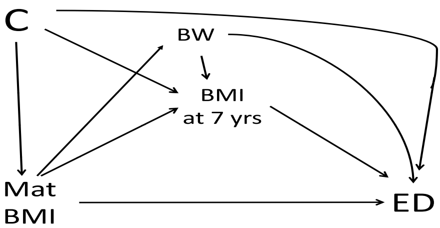
Wide range of options, for most combinations of M and Y:

- G-computation:
  - suitable for estimating $CDE(m)$ and NDE
  - can deal with intermediate confounding
  - flexible and efficient but heavy on parametric modelling assumptions
  - implemented in `gformula` (Daniel et al 2011)
- Semi-parametric methods (*e.g.*g-estimation) make fewer parametric assumptions

Outline

Back to the example

The Avon Longitudinal Study (ALSPAC)



- Birth cohort with 3,500 girls, born in 1991-2
- Outcome (Y): Binge eating at age 13yrs (ED)
- Exposure (X): maternal pre-pregnancy BMI
- Mediator (M): child BMI at age 7 yrs (bmi7)
- Intermediate confounder (L): birth weight (BW)
- Confounders (C): Maternal education and mental disorders

Using gformula (Daniel *et al.* 2011)

Binary exposure: maternal BMI> 25 kg/m^2 (overbmi)

```
gformula <varlist>, mediation outcome(ED) exposure(overbmi)
obe mediator(bmi7) post_confs(BW) base_confs(lowed anyMH)
control(bmi7:0) ...
```

	G-computation estimate	Bootstrap Std. Err.	z	P>|z|
TCE	.1473497	.0252176	5.84	0.000
NDE	.0574193	.0242394	2.37	0.018
NIE	.0899304	.0101036	8.9	0.000
CDE	.0575705	.0242823	2.37	0.018

Using gformula (Daniel *et al.* 2011)

Binary exposure: maternal BMI$>$ 25 kg/m^2 (overbmi)

```
gformula <varlist>, mediation outcome(ED) exposure(overbmi)
obe mediator(bmi7) post_confs(BW) base_confs(lowed anyMH)
control(bmi7:0) ...
```

	G-computation estimate	Bootstrap Std. Err.	z	P>|z|
TCE	.1473497	.0252176	5.84	0.000
NDE	.0574193	.0242394	2.37	0.018
NIE	.0899304	.0101036	8.9	0.000
CDE	.0575705	.0242823	2.37	0.018

Using gformula (Daniel *et al.* 2011)

Binary exposure: maternal BMI$>$ 25 kg/m^2 (overbmi)

```
gformula <varlist>, mediation outcome(ED) exposure(overbmi)
obe mediator(bmi7) post_confs(BW) base_confs(lowed anyMH)
control(bmi7:0) ...
```

```
----------------------------------------------------------
       |  G-computation   Bootstrap
       |    estimate      Std. Err.       z     P>|z|
-------+--------------------------------------------------
TCE    |    .1473497      .0252176      5.84    0.000
NDE    |    .0574193      .0242394      2.37    0.018
NIE    |    .0899304      .0101036      8.9     0.000
CDE    |    .0575705      .0242823      2.37    0.018
```

TCE: Average ED score in the world where all mothers are set to BMI$>$25 minus that where they are set to BMI$\leq$25
($TCE = NDE + NIE$)

Using gformula (Daniel *et al.* 2011)

Binary exposure: maternal BMI$>25\ kg/m^2$ (overbmi)

```
gformula <varlist>, mediation outcome(ED) exposure(overbmi)
obe mediator(bmi7) post_confs(BW) base_confs(lowed anyMH)
control(bmi7:0) ...
```

	G-computation estimate	Bootstrap Std. Err.	z	P>|z|
TCE	.1473497	.0252176	5.84	0.000
NDE	.0574193	.0242394	2.37	0.018
NIE	.0899304	.0101036	8.9	0.000
CDE	.0575705	.0242823	2.37	0.018

NDE:Average ED score in the world where all mothers are set to BMI$>$25 minus that where they are set to BMI$\leq$25, with child BMI set at its natural value when the mother's BMI is set at $\leq$25

Using gformula (Daniel *et al.* 2011)

Binary exposure: maternal BMI$>25\ kg/m^2$ (overbmi)

```
gformula <varlist>, mediation outcome(ED) exposure(overbmi)
obe mediator(bmi7) post_confs(BW) base_confs(lowed anyMH)
control(bmi7:0) ...
```

	G-computation estimate	Bootstrap Std. Err.	z	P>|z|
TCE	.1473497	.0252176	5.84	0.000
NDE	.0574193	.0242394	2.37	0.018
NIE	.0899304	.0101036	8.9	0.000
CDE	.0575705	.0242823	2.37	0.018

$CDE = CDE(0)$: Average ED score in the world where all mothers are set to BMI$>$25 minus that where they are set to BMI$\leq$25, with child BMI set at 0 (bmi7 is standardized)

Using gformula (2)

Binary exposure and interactions

G-computations allows flexible specification of all association models.
⇒ Adding interactions between exposure overbmi, confounder BW and mediator bmi7:

gformula <*original varlist*> over_BW over_bmi7 , ...
derived(over_BW over_bmi7) derrules(over_BW:overbmi*BW, over_bmi7:overbmi*bmi7)

	G-computation estimate	Bootstrap Std. Err.	z	P>|z|
TCE	.1649645	.0263194	6.27	0.000
NDE	.0539651	.0251499	2.15	0.032
NIE	.1109994	.0174183	6.37	0.000
CDE	.0489107	.0235276	2.08	0.038

Using gformula (2)

Binary exposure and interactions

G-computations allows flexible specification of all association models.
⇒ Adding interactions between exposure overbmi, confounder BW and mediator bmi7:

gformula *<original varlist>* over_BW over_bmi7 , ...
derived(over_BW over_bmi7) derrules(over_BW:overbmi*BW, over_bmi7:overbmi*bmi7)

	G-computation estimate	Bootstrap Std. Err.	z	P>|z|
TCE	.1649645	.0263194	6.27	0.000
NDE	.0539651	.0251499	2.15	0.032
NIE	.1109994	.0174183	6.37	0.000
CDE	.0489107	.0235276	2.08	0.038

Using gformula (3)

Continuous exposure

Replacing binary overbmi with continuous BMIpre (standardized):

gformula <*newvarlist*> , ...linexp

	G-computation estimate	Bootstrap Std. Err.	z	P>|z|
TCE	.0199475	.0033078	6.03	0.000
NDE	.0090645	.0029689	3.05	0.002
NIE	.010883	.001978	5.5	0.000
CDE	.0077757	.0027279	2.85	0.004

Using gformula (3)

Continuous exposure

Replacing binary `overbmi` with continuous `BMIpre` (standardized):

`gformula` <*newvarlist*> `, ...linexp`

	G-computation estimate	Bootstrap Std. Err.	z	P>|z|
TCE	.0199475	.0033078	6.03	0.000
NDE	.0090645	.0029689	3.05	0.002
NIE	.010883	.001978	5.5	0.000
CDE	.0077757	.0027279	2.85	0.004

Using gformula (3)
Continuous exposure

Replacing binary `overbmi` with continuous `BMIpre` (standardized):

`gformula` <*newvarlist*> `, ...` `linexp`

```
-----------------------------------------------------------
        |  G-computation    Bootstrap
        |    estimate       Std. Err.        z      P>|z|
--------+--------------------------------------------------
TCE     |     .0199475       .0033078      6.03     0.000
NDE     |     .0090645       .0029689      3.05     0.002
NIE     |      .010883        .001978      5.5      0.000
CDE     |     .0077757       .0027279      2.85     0.004
```

This gives effects for 1 unit increase in (stand) Maternal BMI

Using gformula (4)

Categorical exposure

Replacing binary `overbmi` with categorical `bmicat` (coded: 0,1,2):

`gformula` <*varlist*> `, ...oce baseline(1)`

	G-computation estimate	Bootstrap Std. Err.	z	P>|z|
TCE(0)	-.1041429	.0152158	-6.84	0.000
TCE(2)	.1055137	.015159	6.96	0.000
NDE(0)	.0160151	.0163919	.98	0.329
NDE(2)	.0429508	.0143828	2.99	0.003
NIE(0)	-.1201579	.0125716	-9.56	0.000
NIE(2)	.0625629	.0067193	9.31	0.000
CDE(0)	-.0431978	.0143389	-3.01	0.003
CDE(2)	.0437326	.0143348	3.05	0.002

Using gformula (4)

Categorical exposure

Replacing binary `overbmi` with categorical `bmicat` (coded: 0,1,2):

`gformula` <*varlist*> `, ...oce baseline(1)`

	G-computation estimate	Bootstrap Std. Err.	z	P>|z|
TCE(0)	-.1041429	.0152158	-6.84	0.000
TCE(2)	.1055137	.015159	6.96	0.000
NDE(0)	.0160151	.0163919	.98	0.329
NDE(2)	.0429508	.0143828	2.99	0.003
NIE(0)	-.1201579	.0125716	-9.56	0.000
NIE(2)	.0625629	.0067193	9.31	0.000
CDE(0)	-.0431978	.0143389	-3.01	0.003
CDE(2)	.0437326	.0143348	3.05	0.002

Using gformula (4)

Categorical exposure

Replacing binary `overbmi` with categorical `bmicat` (coded: 0,1,2):

`gformula <varlist> , ...oce baseline(1)`

	G-computation estimate	Bootstrap Std. Err.	z	P>|z|
TCE(0)	-.1041429	.0152158	-6.84	0.000
TCE(2)	.1055137	.015159	6.96	0.000
NDE(0)	.0160151	.0163919	.98	0.329
NDE(2)	.0429508	.0143828	2.99	0.003

Results with "(0)" refer to BMI<18 (low) relatively to 18≤BMI<25 (normal);
those with "(2)" to BMI>25 relatively to normal BMI.

Outline

Summary

- The study of mediation implies causal questions
- Causal inference literature offers general definitions
- There is a choice of estimation methods, each asking a different causal question
- `gformula` offers a very flexible tool to estimate these estimands
- Their identification requires stringent, unverifiable, assumptions:
  "To claim that effects are causal, it is not sufficient to use causally defined effects." (Muthèn, 2011)
- Need for sensitivity analyses (Imai et al, 2010)

Thank you!

Acknowledgements: Nadia Micali, UCL and George Ploubidis, LSHTM

Outline

References

1. 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.
2. Cole SR, Hernán MA. Fallibility in estimating direct effects. *Int J Epidemiol* 2002; 31: 163165.
3. 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.
4. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychological Methods* 2010; 15, 309-334.
5. Muthén B. Applications of Causally Defined Direct and Indirect Effects in Mediation Analysis using SEM in Mplus. 2011.
6. 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;32
7. Pearl J. Direct and indirect effects. *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence* 2001; San Francisco: Morgan Kaufmann.
8. Robins, J. M. and Greenland, S. Identi?ability and exchangeability for direct and indirect e?ects *Epidemiology* 1992, 3:143-155.
9. VanderWeele T, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface* 2009; 2, 457468
10. 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.

Additional slide

Alternative $CDE(m)$

- The inclusion of interaction terms implies that exposure effect is allowed to vary with the mediator
- Hence the CDE(m) will vary with the value assigned to M

Estimand	G-computation estimate	Bootstrap Std. Err.
CDE(0)	0.049	0.024
CDE (-1)	0.021	0.034
CDE (1)	0.077	0.032

G-computation formula for the CDE

Robins, 1986

- Lets look at how the CDE is estimated, when there is also an intermediate confounder L:

$$
\begin{aligned}
CDE(m,c) &= E\{Y(1,m)\,|\,C=c\} - E\{Y(0,m)\,|\,C=c\} \\
&= \int E\left(Y|C=c, X=1, L=l, M=m\right) f_{L|C,X}\left(l|c,1\right) dl \\
&\quad - \int E\left(Y|C=c, X=0, L=l, M=m\right) f_{L|C,X}\left(l|c,0\right) dl
\end{aligned}
$$

- This is the g-computation formula.
- It requires correct specification of these parametric associational models for $Y|C,X,L,M$ and $L|C,X$.
- Both models can be completely flexible: they can include non-linearities and interactions.
- By marginalising over $L|C,X$, intermediate confounding is appropriately dealt with.