

Misclassification of the mediator matters when estimating indirect effects

Tony Blakely, Sarah McKenzie, Kristie Carter

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jech-2012-201813>).

Health Inequalities Research Programme, Department of Public Health, University of Otago Wellington, Wellington, New Zealand

Correspondence to

Dr Tony Blakely, Health Inequalities Research Programme, Department of Public Health, University of Otago Wellington, Mein Street, Wellington 6022, New Zealand; tony.blakely@otago.ac.nz

Received 7 August 2012
Revised 13 November 2012
Accepted 3 December 2012
Published Online First
5 February 2013

ABSTRACT

Background Confounding of mediator–outcome associations resulting in collider biases causes systematic error when estimating direct and indirect effects. However, until recently little attention has been given to the impact of misclassification bias.

Objective To quantify the impact of non-differential and independent misclassification of a dichotomous exposure and a dichotomous mediator on three target parameters: the total effect of exposure on outcome; the direct effect (by conditioning on the mediator); and the indirect effect (identified by the percentage reduction in the excess OR on adjusting for the mediator).

Methods Simulations were conducted for varying strength of associations between exposure, mediator and outcome, varying ratios of exposed to unexposed and mediator present to mediator absent, and varying sensitivity and specificity of exposure and mediator classification.

Results ORs before (total effect) and after adjustment (direct effect) for the mediator are both biased towards the null by non-differential misclassification of the exposure, but the percentage reduction in the excess OR is little affected by misclassification of exposure. Conversely, misclassification of the mediator rapidly biases the percentage reduction of the excess OR (indirect effect) downwards.

Conclusions If the research objective is to quantify the proportion of the total association that is due to mediation (ie, indirect effect), then minimising non-differential misclassification bias of the mediator is more important than that for the exposure. Misclassification bias is an important source of error when estimating direct and indirect effects.

INTRODUCTION

The aim of much epidemiological research is to identify cause and effect associations. Having identified a causal association, or perhaps to strengthen the case for a causal association, researchers often seek to determine what mediates the association between the exposure and outcome. For example, mediation of the association of socioeconomic position with mortality is a common focus of research.^{1–5} Such an analytical approach is known as process analysis,⁶ pathway or intermediary analysis,⁷ hierarchical analysis,⁸ and more recently as determining direct and indirect effects.^{9–11} A standard approach is to quantify the exposure–outcome (X–Y) measure of association (eg, OR) before and after adjusting for (or conditioning on) the mediating variable (Z).¹² The association before adjustment is the total effect, and the X–Y association after adjustment is the ‘direct’ effect that is *not* mediated by Z. By extension, the ‘indirect’ effect is the effect of X on Y which is

mediated though Z. There are a number of methods that are used to quantify the magnitude of the mediating pathway^{13 14}; in this paper we focus on the proportion or percentage of the X–Y association ‘explained’ by adjustment for the mediator as is common practice in social epidemiology.^{13 15 16}

Systematic error may bias the quantification of direct and indirect effects. The main methodological focus to date has been on sources of error arising from confounding (labelled C in figure 1) of the mediator (Z) to outcome (Y) association, such that when one conditions on the mediator Z, a collider bias occurs opening a backdoor pathway from X to Y through C as shown in figure 1B.^{9 17–20} While such a collider bias is plausible,^{9 10} for it to be of substantive impact *all* of the X–Z, C–Z and C–Y associations shown in figure 1 have to be strong and the confounding variable C needs to be neither rare nor ubiquitous in the source population.^{21 22} We have previously argued a decade ago that measurement error, as opposed to backdoor confounding induced by conditioning on a mediator that also happens to be a collider, is likely to be at least as important as a source of error in practice.²¹ Our argument was based on the mathematical parallel with confounder adjustment whereby non-differential measurement error of confounders rapidly attenuates adjustment, resulting in residual confounding^{23–26}; we reasoned the same logic would apply to measurement error of the mediator. However, it is only recently that methodological work has more closely looked at this issue.^{27–30} This work confirms our previous suspicions: non-differential measurement error of the mediator causes an overestimate of the direct effect (eg, the OR after adjusting for the mediator), and therefore an underestimate of the ‘percentage explained’ that is often used in social epidemiology.

In this paper, we present simulations about misclassification of a dichotomous mediator and exposure. (We use the term ‘(mis)classification’ for categorical variables, and the term ‘(mis)measurement’ more generally for both continuous and categorical variables.) While not presenting mathematical formulae for correction factors, our simulations examine the combined effects of misclassification of the exposure and mediator that (to our knowledge) has not previously been reported. We also emphasise the ‘percentage explained’ metric commonly used by applied epidemiologists such as those publishing in this journal. Recent methodological literature focuses on the impact of misclassification on the direct effect per se. Finally, we hope by presenting relatively simple simulations that our paper is accessible to a wider range of epidemiologists and public health researchers.

To cite: Blakely T, McKenzie S, Carter K. *J Epidemiol Community Health* Published Online First: [please include Day Month Year] doi:10.1136/jech-2012-201813

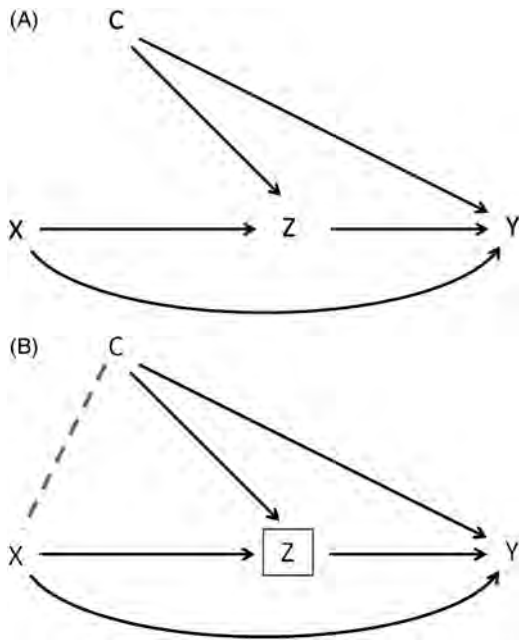


Figure 1 Directed acyclic graph of the associations between exposure (X), mediator (Z), outcome (Y) and confounder of the Z–Y association, demonstrating collider bias when conditioning on Z. (A) Before conditioning on Z. (B) After conditioning on Z.

MOTIVATING EXAMPLE, OBJECTIVE AND GENERAL METHOD

We have recently used cross-sectional data to analyse the association of childhood socioeconomic position (main exposure) with adult psychological distress (outcome), and attempted to determine the proportion of this association mediated by adult socioeconomic position.³¹ In this example, we had only one crude measure of childhood socioeconomic position (created from subjects’ recall of parental occupation), but multiple (potentially better) measures of adult socioeconomic position (ie, highest education, household income, labour market activity and area deprivation). The question then arises as to the relative impacts of misclassification of the exposure and mediator on the assessment of mediation in the analysis.

Figure 2 shows the causal structure with misclassification determinants U_X and U_Z , as well as true values of X and Z,

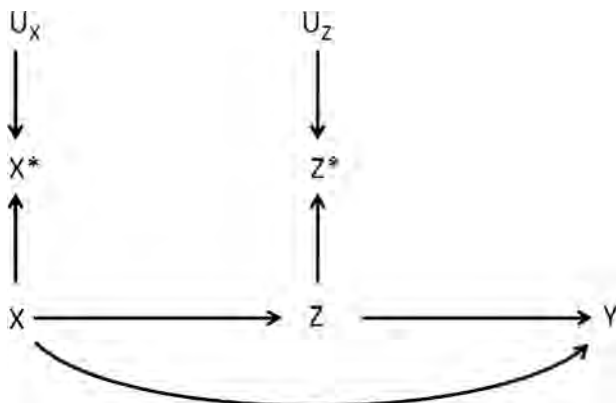


Figure 2 Directed acyclic graph of the casual associations in figure 1, additionally representing measurement error of exposure (X) and mediator (Z).

determining the observed but misclassified exposure and mediator values, X^* and Z^* , respectively (and using the framework of Hernan and Cole,³² and as subsequently elaborated³³). Using the example above, U_Z would be the process of measuring adult socioeconomic position, such as using a single census question with bin categories for total household income as opposed to an ideal set of accurate income records. In this paper we use simulated data to assess the impact of non-differential and independent misclassification of a dichotomous exposure X and a dichotomous mediator Z, separately and simultaneously, when quantifying direct and indirect effects.

We focus on three ‘target parameters’: the OR of the exposure–outcome association before adjustment for the mediator (OR_{XY} : the total effect); this OR after adjustment for the mediator ($OR_{XY|Z}$: the direct effect); and the percentage reduction in this association after adjustment for the mediator ($100\% \times (OR_{XY} - OR_{XY|Z}) / (OR_{XY} - 1)$: the indirect effect). With misclassification, these three parameters become: OR_{X^*Y} ; $OR_{X^*Y|Z^*}$; and $(100\% \times (OR_{X^*Y} - OR_{X^*Y|Z^*}) / (OR_{X^*Y} - 1))$. Space constraints in this paper mean we do not simulate misclassification of the outcome Y, differential misclassification between X and Z, nor dependent misclassification between X and Z.

We use bias analysis methods^{34–35} to estimate the observed joint distribution of exposure, mediator and outcome in the presence of varying sensitivity and specificity of exposure and mediator (mis)classification. The ratio of exposed to unexposed (ie, $X1:X0$), and mediator present to mediator absent (ie, $Z1:Z0$), also impacts on the final amount of bias. For example, for a fixed specificity of the exposure the number of false positives becomes large relative to true positives as the true prevalence of those exposed decreases, that is, the positive predictive value decreases with decreasing true prevalence. We use Mantel–Haenszel pooling across strata of Z to calculate $OR_{X^*Y|Z^*}$ (ie, the direct effect with misclassification).³⁶ Workings and calculations were undertaken in Excel (a copy of the spreadsheet is available at <http://www.uow.otago.ac.nz/sofiehealth-info.html>).

We present our simulations in the following three sections: misclassification of mediator Z only; misclassification of exposure X only; and misclassification of both exposure and mediator. All simulated misclassification is assumed to be non-differential and independent.

MISCLASSIFICATION OF MEDIATOR Z

We let $OR_{XY|Z}$, $OR_{XZ|Y}$ and $OR_{YZ|X}$ have two possible values, 2 or 5, to approximate moderate and strong associations. (Selecting even stronger OR values (eg, 10) does not alter the general findings of this paper.) For example, in the case of $OR_{XY|Z}=2$, and $OR_{XZ|Y}=OR_{YZ|X}=5$, this equates to a moderately strong association of X with Y having adjusted for mediator Z (ie, the ‘direct’ effect), and strong independent associations of the mediator Z with both X and Y. Given these parameters, the total effect of X on Y, OR_{XY} , also depends on the ratio of $Z1:Z0$. In the case of $OR_{XY|Z}=2$, and $OR_{XZ|Y}=OR_{YZ|X}=5$, the maximum value of OR_{XY} is 3.6 (using the formula $(R+1)^2/4R$ times $OR_{XY|Z}$ (1.8×2), where $R=OR_{XZ|Y}=OR_{YZ|X}$),²² and occurs when the ratio of $Z1:Z0$ among the non-cases (ie, $Y=0$) is about 1:4.

To generate data for the simulations, in addition to setting the three ORs and the $Z1:Z0$ ratio, we also set three of the cells in the two by two table of X and Y given Z=1 at 100, as shown in table 1. The remaining cells can be solved given the above specifications. Thus, the ‘true’ data has $OR_{XY}=3.08$, and both $OR_{XY|Z=0}$ and $OR_{XY|Z=1}=2.00$ (as set), meaning that

Table 1 Simulated expected data for $OR_{XY|Z}=2$, $OR_{XZ|Y}=OR_{YZ|X}=5$, when the ratio of $Z=1|Y=0$ to $Z=0|Y=0$ equals 1

a) 'True' data, where three cells (a, b and c; underlined) are set to 100 and remaining cells solved: crude $OR_{XY} = 3.08$; $OR_{XY|Z} = 2.00$; % mediation by Z = 52%.

	Z=1		Z=0		Z pooled	
	X=1	X=0	X=1	X=0	X=1	X=0
Y=1	a <u>100</u>	b <u>100</u>	e 5	f 27	i 105	j 127
Y=0	c <u>100</u>	d 200	g 27	h 273	k 127.3	l 473

b) 'Observed' data due to misclassification of Z (sensitivity = specificity = 70%): crude $OR_{XY} = 3.08$; $OR_{XY|Z}$ (Mantel Haenszel pooling) = 2.89; % mediation by Z = 9%.

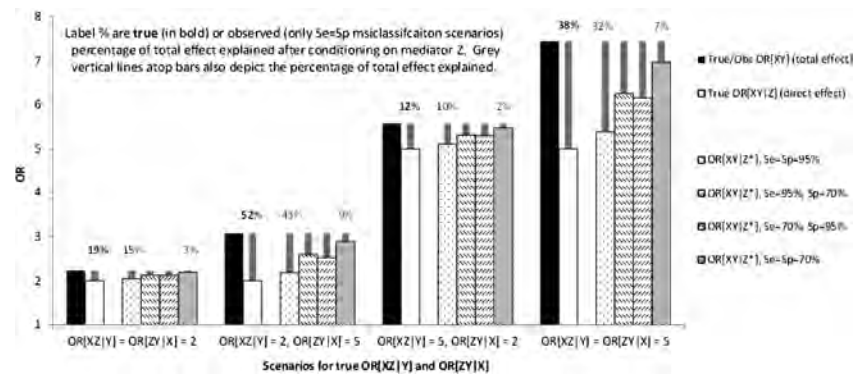
	Z*=1		Z*=0		Z pooled	
	X=1	X=0	X=1	X=0	X=1	X=0
Y=1	a 72	b 78	e 34	f 49	i 105	j 127
Y=0	c 78	d 222	g 49	h 251	k 127	l 473

Table 2 True OR_{XY} , and true and observed percentage mediation of $X \rightarrow Y$ association by Z, for selected combinations of $OR_{XY|Z}$, $OR_{XZ|Y}$, $OR_{YZ|X}$, varying ratio of Z1 to Z0 among non-cases ($Y=0$), and sensitivity (Se) and specificity (Sp) of Z measurement

	$OR_{XY Z}$	$OR_{XZ Y}$	$OR_{YZ X}$	[Z1:Z0 Y=0] = 0.1				[Z1:Z0 Y=0] = 1				[Z1:Z0 Y=0] = 10					
				True/obs OR_{XY}		True % med		True/obs OR_{XY}		True % med		True/obs OR_{XY}		True % med			
	2	2	5	5	2.12	3.26	5.33	8.37	2.23	3.08	5.57	7.44	2.05	2.17	5.12	5.35	
	2	5	2	5	11%	56%	8%	46%	19%	52%	12%	38%	5%	14%	3%	8%	
Apparent mediation by Z, given misclassification of Z †	Se[Z]=	95%	70%	95%	70%	95%	70%	95%	70%	95%	70%	95%	70%	95%	70%	95%	70%
	Sp[Z]=95%	7%	4%	39%	25%	5%	3%	32%	21%	15%	8%	43%	23%	10%	6%	32%	18%
	Sp[Z]=70%	2%	1%	13%	5%	1%	1%	11%	4%	9%	3%	26%	9%	6%	2%	20%	7%
Apparent mediation by Z, given misclassification of Z †	Se[Z]=	95%	70%	95%	70%	95%	70%	95%	70%	95%	70%	95%	70%	95%	70%	95%	70%
	Sp[Z]=95%	3%	1%	7%	2%	2%	0%	4%	1%	3%	1%	0%	3%	0%	3%	0%	0%
	Sp[Z]=70%	2%	0%	4%	1%	1%	0%	3%	0%	2%	0%	0%	1%	0%	0%	0%	0%

†The 'Apparent mediation by Z, given misclassification of Z' is $100\% \times (OR_{XY} - OR_{XY|Z}) / (OR_{XY} - 1)$, where Z* refers to misclassified value of Z. Web annex table 1 includes output for all possible combinations of $OR_{XY|Z}$, $OR_{XZ|Y}$, $OR_{YZ|X}$.

Figure 3 Effect of mediator Z misclassification scenarios on: total and direct ORs; percentage of total effect explained by conditioning on mediator Z (Z1 : Z0 ratio among non-cases=1.0).



$100\% \times (3.08 - 2.00) / (3.08 - 1) = 52\%$ of the $X \rightarrow Z$ association is mediated by Z.

Table 1 shows the data as observed (or 'expected') if the sensitivity and specificity of Z is 70%. No other misclassification bias is allowed, and thus counts are redistributed only between cells: a-e, b-f, c-g and d-h. For example, cell a in table 1 is given by the following maths using cells in table 1: $Se \times a + (1 - Sp) \times e = 70\% \times 100 + 30\% \times 5.5 = 72$. The observed crude OR_{XY} accordingly remains unchanged, but the Z-stratum specific ORs are 2.60 and 3.52, or 2.89 pooled using Mantel-Haenszel pooling. (Misclassification bias is well known to spuriously generate heterogeneity of stratum specific measures of association²³ (ie, 2.60 vs 3.52), but for this paper we focus on ORs pooled across strata.) Accordingly, with misclassification of Z the percentage mediation by Z is $100\% \times (3.08 - 2.89) / (3.08 - 1) = 9\%$, considerably less than 52% in the absence of misclassification of Z.

Table 1 shows the simulated findings for just one combination of parameter values. Table 2 shows the true OR_{XY} and true and observed percentage mediation by Z, for varying sensitivity (95% or 70%) and specificity (95% or 70%) of Z measurement, and varying ratios of Z1-Z0 among the controls (0.1, 1.0 and 10). The output for the example in table 1 is shaded grey in table 2. To simplify, we only present the results where $OR_{XY|Z}$, $OR_{XZ|Y}$, $OR_{YZ|X}$ are all 2 or 5, and when $OR_{XZ|Y}$, $OR_{YZ|X}$ are both stronger or both weaker than $OR_{XY|Z}$. (The full 96 possible simulations can be found in web annex table 1.) We also do not present $OR_{XY|Z^*}$ but rather present the observed percentage mediation; the reader can back calculate $OR_{XY|Z^*}$ from the true/observed OR_{XY} and percentage mediation (ie, $OR_{XY|Z^*} = 1 + ((OR_{XY} - 1) \times (100\% - \% \text{ med by Z}))$) or read them off figure 3 for selected scenarios (see below). Note that the observed OR_{XY} does not differ from the true OR_{XY} when only Z is misclassified, as there is as yet no conditioning on Z. The specific example outlined in table 1 of a true OR_{XY} of 3.08, true mediation of 52% and observed mediation of 9%, is highlighted in the middle panel (where Z1 : Z0=1.0) of table 2 for the combination of 70% sensitivity and 70% specificity of Z.

Across all simulations of misclassification of Z shown in table 2, there is substantive underestimation of the percentage mediation. For the best combination of sensitivity and specificity (both 95%) where Z1 : Z0=0.1, the observed percentage mediation was underestimated by about a third across the various combinations of $OR_{XY|Z}$, $OR_{XZ|Y}$ and $OR_{YZ|X}$. For example, when all OR are 5 and Z1 : Z0=0.1 (ie, top right of table 2), the 32% observed mediation is nearly a third less than 46% even when the sensitivity and specificity are both 95%. When Z1 : Z0=1.0, the underestimate was approximately a fifth for

both sensitivity and specificity of Z equal to 95%. When Z1 : Z0=10, the OR_{XY} was very similar to the set $OR_{XY|Z}$ across all scenarios, and thus the true percentage mediation was always low. That said, the observed percentage mediation was underestimated by approximately half for the best combination of sensitivity and specificity (eg, 4% observed compared to 8% true for all OR=5, and sensitivity and specificity both 95%). When either or both the sensitivity and specificity of Z reduced to 70%, the percentage mediation was often underestimated by as much as 90%.

Figure 3 graphically shows the total and direct ORs (ie, after conditioning on Z) when Z1 : Z0=1 (ie, the middle panel of scenarios in table 2). The true total and direct effect ORs are shown as the black and white bars in each of the four groupings, with true percentage mediated being 19%, 52%, 12% and 38% from left to right. The remaining patterned bars are the direct effects ($OR_{XY|Z^*}$) for various combinations of sensitivity and specificity misclassification of Z. One can see that for worsening misclassification of Z, the $OR_{XY|Z^*}$ gets progressively larger and closer to the total effect OR_{XY} meaning that the percentage mediation is progressively underestimated.

(Simulations for weaker associations (ie, ORs of 1.2 and 1.5) are shown in web annex table 4, and demonstrate qualitatively similar patterns to those shown above.)

MISCLASSIFICATION OF EXPOSURE X

Consider the case of misclassification of X, but not Z. While the X1:X0 ratio among the non-cases does not alter the true OR_{XY} it does have a substantial impact on OR_{X^*Y} and some additional impact on $OR_{X^*Y|Z}$. Therefore, we vary ratios of X1 : X0 in our simulations. Table 3 shows an example of the calculations used in simulations, following a similar framework for the above Z misclassification analyses. Three cells in the 'true' data are set to 100, but one of the cells (e) must be among the Z=0 stratum to operationalise the X1 : X0 specification. Otherwise the calculation of solutions for other cell counts proceeds in a similar manner to table 1. For the example in table 3, $OR_{X^*Y} = 1.48$ (compared to the true $OR_{XY} = 2.89$) and $OR_{X^*Y|Z} = 1.27$ (compared to the true $OR_{XY|Z} = 2.0$), results in 45% of the $X \rightarrow Y$ association being due to mediation by Z. Despite the misclassification of X and the resulting biases in OR_{X^*Y} and $OR_{X^*Y|Z}$, there is little difference from the true percentage mediated of 47%. This is consistent with non-differential independent misclassification of X not being as important as misclassification of Z if the target parameter is the percentage of the excess OR due to mediation by Z. Table 4 presents summary output for multiple simulations for varying combinations of ORs, sensitivity and specificity (95% or 70%) of X (mis)classification, and the ratio

Table 3 Simulated expected data for $OR_{XY|Z}=2$, $OR_{XZ|Y}=OR_{YZ|X}=5$, when the ratio of $X=1|Y=0$ to $X=0|Y=0$ equals 1

a) 'True' data, where three cells (a, c and e; underlined) are set to 100 and remaining cells solved: crude $OR_{XY} = 2.89$; $OR_{XY|Z} = 2.00$; % mediation by Z = 47%.

	Z=1		Z=0		Z pooled	
	X=1	X=0	X=1	X=0	X=1	X=0
Y=1	a <u>100</u>	b 12	e <u>100</u>	f 58	i 200	j 69
Y=0	c <u>100</u>	d 23	g 500	h 577	k 600	l 600

b) 'Observed' data due to misclassification of X (sensitivity = specificity = 70%): $OR_{X*Y} = 1.48$; $OR_{X*Y|Z}$ (Mantel Haenszel pooling) = 1.27; % mediation by Z = 45%.

	Z=1		Z=0		Z pooled	
	X=1	X=0	X=1	X=0	X=1	X=0
Y=1	a 73	b 38	e 87	f 70	i 161	j 108
Y=0	c 77	d 46	g 523	h 554	k 600	l 600

of $X1:X0$ among the non-cases (1.0, 0.1 and 0.01). The output for the example in table 3 is shaded grey in table 4. Whereas it made little sense to consider very low or very high prevalence of Z above, exposures are often uncommon—hence the $X1:X0=0.01$ (ie, a prevalence of exposure among non-cases of just under 1%). A fuller tabulation of simulations is presented in web annex table 2, including a higher specificity of 99% for X, as in the context of an uncommon exposure (ie, $X1:X0=0.01$) one would probably select/set a higher specificity to prevent the true exposed being overwhelmed by the false positive exposed.

To aid interpretation, the target parameter of 'percentage mediation' in table 4 is underlined when it is more than five percentage points less than the true percentage mediation. Using this threshold, no errors of greater than five percentage points occurred for any simulations when the $X1:X0$ ratio was 1.0, and only for some simulations when the $X1:X0$ ratio was 0.1 (only when specificity of X was 70%). When the $X1:X0$ ratio dropped to 0.01 (ie, an uncommon exposure with 1% prevalence), many more simulations had greater than a five percentage point error. However, it should be noted that when more than a five percentage point error in the percentage mediated occurred, it occurred in concert with large downward bias in the OR_{X*Y} that may render the study 'unacceptable' even before undertaking an analysis of direct and indirect effects. Thus, ensuring a high specificity for the classification of an uncommon exposure is important to prevent bias in both overall and mediation analyses. That said, in most instances when OR_{X*Y} was underestimated, so too was $OR_{X*Y|Z}$ in a commensurate manner such that little or no bias resulted in the percentage mediation explained. (Simulations of exposure misclassification for weaker associations (see web annex table 5) also demonstrate little impact on the percentage mediation.)

Figure 4 graphically shows the total and direct ORs for the same exposure misclassification scenario as the top panel of

results in table 4. One can see that worsening misclassification of X biases both the total and direct effects (ie, OR_{X*Y} and $OR_{X*Y|Z}$) towards the null to a commensurate degree, such that no or little bias results in the estimated percentage mediation.

To summarise thus far, misclassification bias of the mediator alone results in substantial underestimation of the percentage mediated. Misclassification bias of the exposure alone (as is well known) causes substantial underestimation of OR_{X*Y} and in parallel a substantial underestimation of $OR_{X*Y|Z}$. However, the observed percentage mediated (in expectation at least) is only modestly underestimated when the $X1:X0$ ratio is 1.0 or 0.1, but becomes more notably biased when the $X1:X0=0.01$.

MISCLASSIFICATION OF BOTH EXPOSURE X AND MEDIATOR Z

Table 5 presents selected results for misclassification bias of both X and Z. Only options where the sensitivity and specificity of X or Z are equivalent are presented (ie, both 70% or both 95%). Scenarios about the $X1:X0$ ratio are specified, but not about Z. A fuller range of simulations are included in web annex table 3, including those with a 99% specificity of X.

All instances in table 5 where sensitivity and specificity of Z are 70% result in severe underestimation of the percentage mediation. When both the sensitivity and specificity of Z are 95%, the underestimation of the percentage mediation by Z is more severe than with just misclassification of Z (table 2) or just misclassification of X (table 4). For example, when all ORs were 5, and the ratio of $X1:X0$ among the controls was 1.0, the true percentage mediation by Z was 36%, and the observed mediation was 25% and 22% for sensitivity/specificity of X of 95%/95% and 70%/70%. Comparing these observed percentage mediation estimates with the comparable ones from table 4 for misclassification of X (36% and 32%, respectively), they are further underestimated by over a third. This underestimate by

Table 4 True OR_{XY} and observed OR_{X^*Y} , and true and observed percentage mediation of $X \rightarrow Y$ association by Z , for selected combinations of $OR_{XY|Z}$, $OR_{XZ|Y}$, $OR_{YZ|X}$, varying ratio of $X1$ to $X0$ among non-cases ($Y=0$), and sensitivity (Se) and specificity (Sp) of X measurement

True data	$OR_{XY Z}$	2				5			
		2		5		2		5	
	$OR_{XZ Y}$	2		5		2		5	
	$OR_{YZ X}$	2		5		2		5	
	$OR_{XY} \ddagger$	2.22		2.89		5.56		7.22	
	% med by Z	18%		47%		12%		36%	
Observed data									
[$X1:X0 Y=0$] = 1.0									
Observed OR_{X^*Y} , given misclassification of X	$Se[X]=$	95%	70%	95%	70%	95%	70%	95%	70%
	$Sp[X]=95\%$	2.04	1.66	2.55	1.90	4.34	2.51	5.27	2.73
	$Sp[X]=70\%$	1.78	1.36	2.16	1.48	3.42	1.77	4.05	1.87
Apparent mediat. by Z, given misclass. of X †	$Sp[X]=95\%$	18%	18%	47%	47%	12%	12%	36%	36%
	$Sp[X]=70\%$	18%	17%	46%	45%	12%	11%	34%	32%
[$X1:X0 Y=0$] = 0.1									
Observed OR_{X^*Y} , given misclassification of X	$Se[X]=$	95%	70%	95%	70%	95%	70%	95%	70%
	$Sp[X]=95\%$	1.79	1.65	2.22	1.99	3.89	3.21	4.92	3.89
	$Sp[X]=70\%$	1.28	1.17	1.44	1.26	2.03	1.57	2.39	1.73
Apparent mediat. by Z, given misclass. of X †	$Sp[X]=95\%$	17%	16%	44%	44%	10%	10%	32%	33%
	$Sp[X]=70\%$	14%	14%	40%	39%	8%	7%	26%	25%
[$X1:X0 Y=0$] = 0.01									
Observed OR_{X^*Y} , given misclassification of X	$Se[X]=$	95%	70%	95%	70%	95%	70%	95%	70%
	$Sp[X]=95\%$	1.19	1.15	1.30	1.22	1.72	1.54	1.99	1.73
	$Sp[X]=70\%$	1.04	1.02	1.06	1.03	1.14	1.08	1.19	1.11
Apparent mediat. by Z, given misclass. of X †	$Sp[X]=95\%$	13%	13%	35%	34%	5%	5%	18%	17%
	$Sp[X]=70\%$	12%	12%	32%	32%	4%	4%	14%	14%

Underlined percentage mediation results that are >5 percentage points less than true percentage mediation.
 †The 'Apparent mediat. by Z, given misclass. of X' is $100 \times (OR_{X^*Y} - OR_{XY}) / (OR_{X^*Y} - 1)$, where X^* refers to misclassified value of X.
 ‡The true OR_{XY} does not vary with variation in the ratio of $X1$ to $X0$ among controls ($Y=0$).

over a third is comparable to that occurring in table 2 for the same combination of ORs, both sensitivity and specificity of Z of 95%, and $Z1:Z0$ ratio of 0.1 (actually=0.11 for scenario in table 5), namely an observed percentage mediation of 32%

compared to a true percentage mediation of 46%. That is, the underestimation bias of the parameter 'percentage explained by Z' for both misclassification of X and Z is roughly the sum of the separate impacts of misclassification of X and Z.

Figure 4 Effect of exposure X misclassification scenarios on: total and direct ORs; percentage of total effect explained by conditioning on mediator Z ($OR_{XY|Z} = 2$, $OR_{XZ|Y} = OR_{YZ|X} = 5$, $(X1 : X0|Y = 0) = 1.0$).

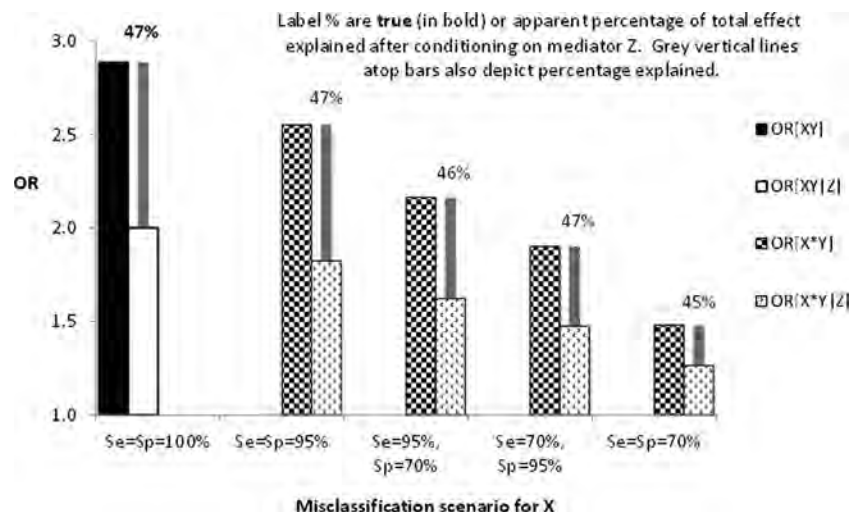


Table 5 True OR_{XY} and observed OR_{X^*Y} , and true and observed percentage mediation of $X \rightarrow Y$ association by Z , for selected combinations of $OR_{XY|Z}$, $OR_{XZ|Y}$, $OR_{YZ|X}$, varying ratio of $X1:X0$ among non-cases ($Y=0$), and sensitivity (Se) and specificity (Sp) of both X and Z

True data	$OR_{XY Z}$	2				5			
		2	5	2	5	2	5	2	5
	$OR_{XZ Y}$	2	5	2	5	2	5	2	5
	$OR_{ZY X}$	2	5	2	5	2	5	2	5
	OR_{XY}	2.22	2.89	5.56	7.22	2.22	2.89	5.56	7.22
	% med by Z	18%	47%	12%	36%	18%	47%	12%	36%
[X1:X0 Y=0] = 1.0		$Se/Sp[Z]=$ 95% 70% 95% 70% 95% 70% 95% 70%							
Observed OR_{X^*Y} , given misclass of X&Z	$Se/Sp[X]=95%$	2.04	2.55	4.34	5.27	2.04	2.55	4.34	5.27
	$Se/Sp[X]=70%$	1.36	1.48	1.77	1.87	1.36	1.48	1.77	1.87
Apparent mediat. by Z , given misclass of X&Z †	$Se/Sp[X]=95%$	14%	3%	33%	4%	10%	2%	25%	3%
	$Se/Sp[X]=70%$	13%	2%	31%	4%	8%	1%	22%	3%
[X1:X0 Y=0] = 0.1		$Se/Sp[Z]=$ 95% 70% 95% 70% 95% 70% 95% 70%							
Observed OR_{X^*Y} , given misclass of X&Z	$Se/Sp[X]=95%$	1.79	2.22	3.89	4.92	1.79	2.22	3.89	4.92
	$Se/Sp[X]=70%$	1.17	1.26	1.57	1.73	1.17	1.26	1.57	1.73
Apparent mediat. by Z , given misclass of X&Z †	$Se/Sp[X]=95%$	13%	2%	24%	2%	8%	1%	17%	2%
	$Se/Sp[X]=70%$	11%	2%	21%	2%	5%	1%	13%	1%
[X1:X0 Y=0] = 0.01		$Se/Sp[Z]=$ 95% 70% 95% 70% 95% 70% 95% 70%							
Observed OR_{X^*Y} , given misclass of X&Z	$Se/Sp[X]=95%$	1.19	1.30	1.72	1.99	1.19	1.30	1.72	1.99
	$Se/Sp[X]=70%$	1.02	1.03	1.08	1.11	1.02	1.03	1.08	1.11
Apparent mediat. by Z , given misclass of X&Z †	$Se/Sp[X]=95%$	10%	2%	16%	1%	4%	1%	8%	1%
	$Se/Sp[X]=70%$	9%	1%	15%	1%	3%	0%	6%	0%

†The 'Apparent mediat. by Z , given misclass of X&Z' is $100\% \times (OR_{X^*Y} - OR_{XY|Z}) / (OR_{X^*Y} - 1)$, where X^* and Z^* refers to misclassified value of X and Z .

DISCUSSION

Assuming non-differential and independent errors of the exposure and mediator, these simulations suggest that the indirect effect or estimated percentage of the total exposure–outcome association that is due to mediation by Z will be systematically and substantially underestimated by misclassification of Z (consistent with other recent methodological work²⁹), but less so by misclassification of the exposure. The magnitude of the bias depends on the ratios of exposed to unexposed, and the ratio of mediator present to mediator absent. Misclassification of both exposure and mediator causes an underestimation of the 'percentage explained' parameter that is, roughly, equivalent to the sum of the separate biases.

The majority of the methodological examination of systematic error in the estimation of direct and indirect effects within epidemiology has focused on collider bias when conditioning on Z , due to uncontrolled confounding of the $Z \rightarrow Y$ association. This is shown by 'C' in figure 1. Using formulae derived by Greenland,²² if each of the independent ORs for $X \rightarrow Z$, $C \rightarrow Z$ and $C \rightarrow Y$ were 2.0, the *maximum* (given 'worst' possible combination of prevalences of Z and C) downward bias in the direct effect (ie, $OR_{XY|Z}$ as compared to the true $OR_{XY|CZ}$) is only 2%. This maximum downward bias increases to 20% if each of the independent ORs were 5.0 (or about 17% if the prevalence

of Z and C were both 50%). If we apply this collider bias to our scenarios, assuming the ratio of $Z1:Z0$ is 1 (ie, equivalent to a 50% prevalence of Z ; middle panel of table 2) and a prevalence of C of 50%, then the apparent percentage mediation will be overestimated as shown in table 6. It is only when there is strong confounding by C of the ZY association (eg, *both* $OR_{CZ|XY} = OR_{CY|XZ} = 5$) that the collider bias becomes substantive,^{21 22} and of an equivalent magnitude to that for misclassification bias in our simulations. However: (1) collider bias may not be predictable in its direction of bias (whereas non-differential and independent misclassification bias may be more predictable in its direction); and (2) the total impact of collider bias will increase if there are many collider pathways opened by conditioning on Z (ie, C_1, C_2, C_3, \dots could be added to figure 1, and if they are all on separate back door or confounding paths then the cumulative effect of opening all these collider paths may be substantive). Thus, *both* collider bias and misclassification of the mediator are important potential sources of error in mediation analysis.

The existing body of work on measurement error of confounders and its impact on observed exposure–outcome associations^{24 25 37 38} is directly applicable from a mathematical viewpoint—one just shifts from conceptualising Z as a confounder to conceptualising Z as a mediator. The methods presented

Table 6 Observed direct effect and percentage mediation in the presence of collider bias by C (as shown in figure 1) but no misclassification bias, for ratio of Z1 to Z0 among controls (Y=0) equal to 1 and prevalence of U=50%

OR _{XY CZ} (i.e. true direct effect)	2	2	5	5
OR _{XZ YC} = OR _{ZY XC} = OR _{CZ XY} = OR _{CY ZX}	2	5	2	5
Obs OR _{XY} (i.e. true total effect)	2.23	3.08	5.57	7.44
True % med	19%	52%	12%	38%
Obs OR _{XY Z} (i.e. collider-biased estimate of direct effect)	(1/1.02) × 2 = 1.96	(1/1.21) × 2 = 1.65	(1/1.02) × 5 = 4.90	(1/1.21) × 5 = 4.13
Observed % mediation of XY association by Z	(2.23-1.96)/ (2.23-1) = 22%	(3.08-1.65)/ (3.08-1) = 69%	(5.57-4.90)/ (5.57-1) = 15%	(7.44-4.13)/ (7.44-1) = 51%

The collider bias ratios (1/1.02 and 1/1.21) are as per calculations in Greenland²² assuming the prevalence of Z and U are both 50%. These bias ratios are for the observed direct effect (OR_{XY|Z}) compared to the true direct effect (OR_{XY|CZ}) due to inability to block the backdoor path through C 'opened' by conditioning on Z (ie, collider bias), and apply to moderate (ie, OR_{CZ|XY}=OR_{CY|ZX}=2) and strong (ie, OR_{CZ|XY}=OR_{CY|ZX}=5) confounding of the Z→Y association by C, as shown in figure 1.

here facilitate an intuitive shift in interpretation from confounding to mediation of direct and indirect effects: residual (or unmeasured) confounding now equates to residual (or unmeasured) mediation; the crude (confounded) measure of association now equates to the total effect; the adjusted measure of association now equates to the direct effect; and the percentage reduction from the crude to adjusted measure of association now equates to the percentage mediation.

An example of the applicability of the methodological work on measurement error of confounders to measurement error of mediators is one finding of Fewell *et al.*³⁷ They found that 'residual confounding' by multiple mismeasured or unmeasured confounders is less problematic when the confounders are all correlated—adjusting for two of four such correlated confounders 'captured' most of the confounding. Going back to our motivating example (ie, adult socioeconomic position as a potential mediator of the association of childhood socioeconomic position with adult psychological distress³¹), we know that varying measures of adult socioeconomic position (such as income, education and asset wealth) are all correlated. Therefore, including a couple of such measures of socioeconomic position that are well measured, or a handful of such socioeconomic factors that are moderately mismeasured, should 'capture' most of the percentage mediation (again, assuming independent and non-differential measurement errors). Including multiple measures of the mediating variable or construct, *ceteris paribus*, will be a better strategy for fully capturing percentage mediation than including only one mismeasured proxy of the mediator.

In this paper we focus on the percentage of the excess OR explained by adjusting for the mediator, a common practice in social epidemiological studies where the goal is to explain causal pathways from social position to health outcomes. Hafeman and Schwartz¹¹ have connected this 'explanation' agenda with recent methodological work on direct and indirect effects. Briefly, a key distinction in the estimation of direct and indirect effects is between a focus on controlled effects (by setting the value of the mediator) and natural effects (by allowing the mediator to vary 'naturally' with exposure).³⁹ Hafeman and Schwartz¹¹ argue that natural effects are most appropriate for an 'explanation' agenda that is implicit in our paper. More specifically, Hafeman and Schwartz's second motivation for undertaking mediation analysis, to 'test a pathway specific hypothesis',¹¹ matches our motivating

example. Thus, using their framework, it is the 'total indirect effect' (TIE; the effect of exposure that would be prevented if the exposure did not cause the mediator) that is of interest to us. Accordingly, we also calculated TIEs and compared them to the total risk difference (RD) for the above simulations (ie, P(Y=1 | X=1)–P(Y=1 | X=0); workings available in spreadsheets). Given that our simulations forced OR homogeneity, there was (mathematically necessary) heterogeneity of the RD for the XY association across strata of Z. (It is mathematically impossible to specify RD homogeneity for all three of RD_{XY|Z}, RD_{XZ|Y} and RD_{ZY|X} using a framework similar to that used above.) Nevertheless, examining the TIEs using RDs within the framework of Schwartz and Hafeman gave a qualitatively similar conclusion to our OR simulations: for mediator misclassification the TIE was underestimated (often grossly), and therefore the TIE as a percentage of the total RD was underestimated; for exposure misclassification, both the TIE and total RD were underestimated, but the TIE as a percentage was (usually) little affected.

A potential concern with our simulations is non-collapsibility of the OR.³⁶ For strong effects and large differences in the proportion of exposed across strata of a third variable, one may see a lower crude OR than by stratum of the third variable in the *absence* of any confounding. Such a concern may extend to when there is an absence of mediation. While noting scale dependency (homogeneity of the three ORs in our simulations), we do note that before modelling any misclassification, if we set either OR_{XZ|Y} or OR_{ZY|X} to 1 then there was perfect collapsibility of OR_{XY} (ie, OR_{XY|Z=0}=OR_{XY|Z=1}=OR_{XY}). Also, the above alternative examination using the TIEs provides further reassurance that our key findings are not distorted by non-collapsibility. Nevertheless, this might be a worthwhile area of further methodological research.

CONCLUSION

Accurate classification of the mediator is essential and more important than accurate classification of the exposure when the goal is to estimate the percentage of the total X–Y association mediated by Z. However, accurate classification of the exposure is also still highly desirable for the following reasons. It gives an observed total effect (OR_{X→Y}) that is both more accurate, and is far enough away from the null to allow statistically stable estimates of the percentage mediation to be calculated. Accurate

classification of the exposure also guards against substantial differential or dependent misclassification of the exposure, which will alter the conclusions that apply above for non-differential and independent errors. Finally, accurate exposure assessment combined with accurate mediator assessment will usually result in more accurate estimation of the direct effect (ie, $OR_{XY|Z}$).

The accurate estimation of percentage mediation will require attention to study design to minimise misclassification of the mediator. Additionally, it is possible to use quantitative bias analyses methods,^{28 29 34 35} whereby one posits plausible ranges of misclassification of the mediator and exposure and then calculates corrected total, direct and indirect effects. Such quantitative bias analyses could also simultaneously model both collider and misclassification bias.

What this study adds

- ▶ Determining direct and indirect effects, or the proportion of and exposure–outcome association explained by a mediating variable, is a common analysis in epidemiology.
- ▶ Confounding and collider bias as sources of error in determining direct and indirect effects have been understood for at least a decade, but only recently has there been serious methodological scrutiny of misclassification bias.
- ▶ For the target parameter of ‘proportion of the exposure–mediator association explained by the mediator’, we find using simulations about non-differential and independent misclassification of both a dichotomous exposure and mediator that misclassification of the mediator is a far more important source of error than misclassification of the exposure.

Acknowledgements Comments were received on drafts of this paper from John Lynch, Diana Sarfati, Neil Pearce, Jennie Connor and Ana Diez-Roux.

Contributors TB initiated the paper, led all simulations, and led drafting of the paper. SM assisted with simulations and contributed to drafting. KC contributed to drafting.

Funding This work was supported by Health Research Council of New Zealand grant number 08/048. All authors receive funding for the Health Research Council of New Zealand (10/248, 08/048, 06/256).

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Simulations (in Excel spreadsheets) are available from the authors on request.

REFERENCES

- 1 Marmot MG, Shipley M, Rose G. Inequalities in death—specific explanations of a general pattern? *Lancet* 1984;1:1003–6.
- 2 Kivimäki M, Shipley MJ, Ferrie JE, *et al.* Best-practice interventions to reduce socioeconomic inequalities of coronary heart disease mortality in UK: a prospective occupational cohort study. *Lancet* 2008;372:1648–54.
- 3 Lynch J, Kaplan G, Cohen R, *et al.* Do cardiovascular risk factors explain the relation between socioeconomic status, risk of all-cause mortality, cardiovascular mortality, and acute myocardial infarction? *Am J Epidemiol* 1996;144:934–42.
- 4 Blakely T, Fawcett J, Hunt D, *et al.* What is the contribution of smoking and socioeconomic position to ethnic inequalities in mortality in New Zealand? *Lancet* 2006;368:44–52.
- 5 Stringhini S, Dugravot A, Shipley M, *et al.* Health behaviours, socioeconomic status, and mortality: further analyses of the British Whitehall II and the French GAZEL prospective cohorts. *PLoS Med Pub Libr Sci* 2011;8:e1000419.
- 6 Judd CM, Kenny DA. Process analysis. Estimating mediation in treatment evaluations. *Eval Rev* 1981;5:602–19.
- 7 Singh-Manoux A. Commentary: modelling multiple pathways to explain social inequalities in health and mortality. *Int J Epidemiol* 2005;34:638–9.
- 8 Victora CG, Huttly SR, Fuchs SC, *et al.* The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. *Int J Epidemiol* 1997;26:224–7.
- 9 Cole SR, Hernan MA. Fallibility in estimating direct effects. *Int J Epidemiol* 2002;31:163–5.
- 10 Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992;3:143–55.
- 11 Hafeman DM, Schwartz S. Opening the black box: a motivation for the assessment of mediation. *Int J Epidemiol* 2009;38:838–45.
- 12 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.
- 13 Hafeman DM. “Proportion explained”: a causal interpretation for standard measures of indirect effect? *Am J Epidemiol* 2009;170:1443–8.
- 14 Mackinnon DP, Lockwood CM, Hoffman JM, *et al.* A comparison of methods to test mediation and other intervening variable effects. *Psychol Methods* 2002;7:83–104.
- 15 Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Philadelphia: Lippincott Williams & Wilkins, 2008.
- 16 Fritz MS, Mackinnon DP. Required sample size to detect the mediated effect. *Psychol Sci (Wiley-Blackwell)* 2007;18:233–9.
- 17 Glymour M, Greenland S. Causal diagrams. In: Rothman K, Greenland S, Lash T. *Modern epidemiology*. 3rd edn. Philadelphia: Lippincott Williams & Wilkins, 2008:183–212.
- 18 Greenland S, Brumback B. An overview of relations among causal modelling methods. *Int J Epidemiol* 2002;31:1030–37.
- 19 Hafeman DM, Vanderweele TJ. Alternative assumptions for the identification of direct and indirect effects. *Epidemiology* 2011;22(6):753–764.
- 20 Hafeman DM. Confounding of indirect effects: a sensitivity analysis exploring the range of bias due to a cause common to both the mediator and the outcome. *Am J Epidemiol* 2011;174(6):710–717.
- 21 Blakely T. Estimating direct and indirect effects: fallible in theory, but in the real world? *Int J Epidemiol* 2002;31:166–7.
- 22 Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 2003;14:300–6.
- 23 Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980;112:564–9.
- 24 Marshall J, Hastrup J. Mismeasurement and the resonance of strong confounders: uncorrelated errors. *Am J Epidemiol* 1996;143:1069–78.
- 25 Marshall J, Hastrup J, Ross J. Mismeasurement and the resonance of strong confounders: correlated errors. *Am J Epidemiol* 1999;150:88–96.
- 26 Phillips A, Davey Smith G. Bias in relative odds estimation owing to imprecise measurement of correlated exposures. *Stat Med* 1992;11:953–61.
- 27 Vanderweele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology* 2010;21:540–51.
- 28 Vanderweele TJ, Valeri L, Ogburn EL. The role of measurement error and misclassification in mediation analysis: mediation and measurement error. *Epidemiology* 2012;23:561–4.
- 29 Le Cessie S, Debeij J, Rosendaal FR, *et al.* Quantification of bias in direct effects estimates due to different types of measurement error in the mediator. *Epidemiology* 2012;23:551–60.
- 30 Ogburn E, Vanderweele T. Analytic results on the bias due to nondifferential misclassification of a binary mediator. *Am J Epidemiol* 2012;176:555–61.
- 31 McKenzie S, Carter K, Blakely T, *et al.* The association of childhood socioeconomic position and psychological distress in adulthood: is it mediated by adult socioeconomic position? *Int J Longitudinal Life Course Stud* 2010;1:359–8.
- 32 Hernan MA, Cole SR. Invited commentary: causal diagrams and measurement bias. *Am J Epidemiol* 2009;170:959–62.
- 33 Vanderweele TJ, Hernan MA. Results on differential and dependent measurement error of the exposure and the outcome using signed directed acyclic graphs. *Am J Epidemiol* 2012;175:1303–10.
- 34 Lash T, Fox M, Fink A. *Applying quantitative bias analysis to epidemiological data*. New York: Springer, 2009.
- 35 Greenland S. Basic methods for sensitivity analysis and external adjustment. In: Greenland S, Rothman K. *Modern epidemiology*. 2nd edn. Philadelphia: Lippincott-Raven, 1998:343–58.
- 36 Rothman K, Greenland S, Lash T. *Modern epidemiology*. 3rd edn. Philadelphia: Lippincott Williams & Wilkins, 2008.
- 37 Fewell Z, Davey Smith G, Sterne JaC. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol* 2007;166:646–55.
- 38 Phillips A, Davey Smith G. Bias in relative odds estimation owing to imprecise measurement of correlated exposures. *Soc Sci Med* 1992;11:953–61.
- 39 Pearl J. Direct and Indirect Effects. Proceedings of the American Statistical Association Joint Statistical Meetings 2005:1572–81.