International Journal of Epidemiology, 2017, 1–10
doi: 10.1093/ije/dyx027
Education Corner

Education Corner

A comparison of sensitivity-specificity imputation, direct imputation and fully Bayesian analysis to adjust for exposure misclassification when validation data are unavailable

Marine Corbin,[1,2,*] Stephen Haslett[1,3,4] Neil Pearce,[1,5] Milena Maule[2] and Sander Greenland[6]

[1]Centre for Public Health Research, Massey University, Wellington, New Zealand, [2]Unit of Cancer Epidemiology, Department of Medical Sciences, University of Turin, Turin, Italy, [3]Statistical Consulting Unit, Australian National University, Camberra, Australia, [4]Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand, [5]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK and [6]Department of Epidemiology and Department of Statistics, University of California, Los Angeles, CA, USA

*Corresponding author. Centre for Public Health Research, Massey University, Wellington Campus, PO Box 756, Wellington 6140, New Zealand. E-mail: m.corbin@massey.ac.nz

Editorial decision 30 January 2017; Accepted 19 February 2017

Abstract

Purpose: Measurement error is an important source of bias in epidemiological studies. We illustrate three approaches to sensitivity analysis for the effect of measurement error: imputation of the 'true' exposure based on specifying the sensitivity and specificity of the measured exposure (SS); direct imputation (DI) using a regression model for the predictive values; and adjustment based on a fully Bayesian analysis.

Methods: We deliberately misclassify smoking status in data from a case-control study of lung cancer. We then implement the SS and DI methods using fixed-parameter (FBA) and probabilistic (PBA) bias analyses, and Bayesian analysis using the Markov-Chain Monte-Carlo program WinBUGS to show how well each recovers the original association.

Results: The 'true' smoking-lung cancer odds ratio (OR), adjusted for sex in the original dataset, was OR = 8.18 [95% confidence limits (CL): 5.86, 11.43]; after misclassification, it decreased to OR = 3.08 (nominal 95% CL: 2.40, 3.96). The adjusted point estimates from all three approaches were always closer to the 'true' OR than the OR estimated from the unadjusted misclassified smoking data, and the adjusted interval estimates were always wider than the unadjusted interval estimate. When imputed misclassification parameters departed much from the actual misclassification, the 'true' OR was often omitted in the FBA intervals whereas it was always included in the PBA and Bayesian intervals.

Conclusions: These results illustrate how PBA and Bayesian analyses can be used to better account for uncertainty and bias due to measurement error.

Key words: Misclassification, lung cancer, smoking status, sensitivity/specificity imputation, direct imputation, fully Bayesian analysis

Key messages

- We illustrate how to apply several methods for sensitivity analysis of misclassification, including imputation based on sensitivity and specificity, direct imputation based on predictive values and fully Bayesian analyses.
- Sensitivity-specificity imputation requires only values or prior distributions for sensitivity and specificity, but these values or priors should be restricted to values compatible with the data.
- Direct imputation does not require range restrictions, but does require information beyond sensitivity and specificity, such as a prior distribution for the association of interest.
- Fully Bayesian analyses require the most prior information, but can best capture the uncertainty warranted under the assumed models and priors.
- All methods should employ priors that are plausible in light of background literature.

Introduction

A major source of bias and uncertainty in epidemiological analysis is measurement error, usually termed 'misclassification' when referring to discrete variables.[1–4] Measurement error can be considered a missing-data problem[3] in that information has been recorded on a variable which is an imperfect surrogate for the missing 'true' variable of interest.

When internal validation or replication data are not available, the true values for the mismeasured variables are completely missing and no consistent point estimate can be constructed from the data without adding further, potentially arbitrary assumptions. To address this problem, simple sensitivity-analysis formulae adjust for misclassification assuming various values for fixed misclassification rates, based on background literature or on external validation data.[2,13,14] More sophisticated analyses construct and use prior distributions for these rates;[2–4,13,15–18] in that case, standard missing-data software can be used by augmenting the actual data with pseudo-validation data representing these priors.[3]. Such analyses may be repeated using different plausible priors to assess sensitivity to the assumed prior information.

In this paper we focus on the situation where exposure has been misclassified, no validation data are available and adjustment for potential confounders or matching factors is needed. We illustrate and compare methods to adjust for the misclassification of smoking status in a case-control study of smoking and lung cancer, while also adjusting for sex. Each method can be carried out with commercial software.

Material and Methods

Methods

In a case-control study, let Y, T and C denote the outcome (case/control status), exposure status (exposed/unexposed) and a dichotomous covariate, respectively. In many studies the exposure T cannot be directly observed and a surrogate exposure X is measured instead. In order to retrieve information on the 'true' exposure and its association with the outcome, one has to make a priori assumptions on the relationship between T and X, i.e. on the misclassification rates. Assumptions can be made on one of the two following groups of rates:

i. In the sensitivity-specificity imputation approach (SS), assumptions are made on the proportion of subjects classified as exposed among those truly exposed, i.e. the sensitivity (*Se*) and the proportion of subjects classified as unexposed among those truly unexposed, i.e. the specificity (*Sp*).
ii. In the direct imputation approach (DI), assumptions are made on the proportion of truly exposed subjects among those classified as exposed, i.e. the positive predictive value (PPV), and the proportion of truly unexposed subjects among those classified as unexposed, i.e. the negative predictive value (NPV). The predictive

values can be expressed as functions of the sensitivity, specificity and true exposure prevalence:

$$PPV = \frac{Se \times P(T=1)}{Se \times P(T=1) + (1-Sp) \times \Big(1 - P(T=1)\Big)}$$

$$NPV = \frac{Sp \times \Big(1 - P(T=1)\Big)}{Sp \times \Big(1 - P(T=1)\Big) + (1-Se) \times P(T=1)}$$

Making assumptions about PPV and NPV is therefore equivalent to making assumptions on *Se*, *Sp* and $P(T=1)$.

Figure 1 summarizes how information from a priori assumptions and information from the data are combined to provide adjusted estimates in both methods. Detailed algorithms are included in Appendix A (available as Supplementary data at *IJE* online).

These assumptions can be expressed with more or less uncertainty. One can define a range of a priori values for the misclassification proportions (fixed-parameter-bias-sensitivity analysis or FBA) or a priori probability distributions for these proportions (probabilistic bias analysis or PBA). The most rigorous way to do PBA is via Bayesian techniques,[15] but a simple approximation is provided by Monte Carlo sensitivity analysis (MCSA) in which combinations of parameters are sampled from the prior distributions, and then an analysis is conducted for each sampled combination.[2,13,16,17] Thus, MCSA involves a sensitivity analysis using a random sample of values for adjustments, instead of fixed values. On the other hand, a fully Bayesian analysis updates the prior distributions based on the study data to yield posterior distributions for the parameters.[2,13,16,17] Procedures for MCSA have been implemented in Excel and SAS.[2,16]

We consider here both SS and DI approaches, using FBA, MCSA and a fully Bayesian analysis. The a priori values and distributions are described in Table 1.

We use updated versions of a SAS macro implementing MCSA, which allow covariates in the imputation model,[16] and the free software WinBUGS to implement fully Bayesian analysis.

We caution that the interval estimates used for our analyses do not satisfy the criterion for being valid confidence intervals (they would not have 95% coverage under all fixed parameter values); although they may provide adequate coverage when the true parameter values are very close to the parameter values used in the FBA, or close to the centres of the prior distributions in the MCSA, they can have poor coverage otherwise. Neither are the FBA and MCSA intervals valid posterior intervals (they are not a coherent integration of prior and data information) although MCSA intervals can be adequate approximations under certain simplifying assumptions.[15,17] We therefore refer to them only as FBA or MCSA intervals, as appropriate, noting that the quality of the MCSA approximation to

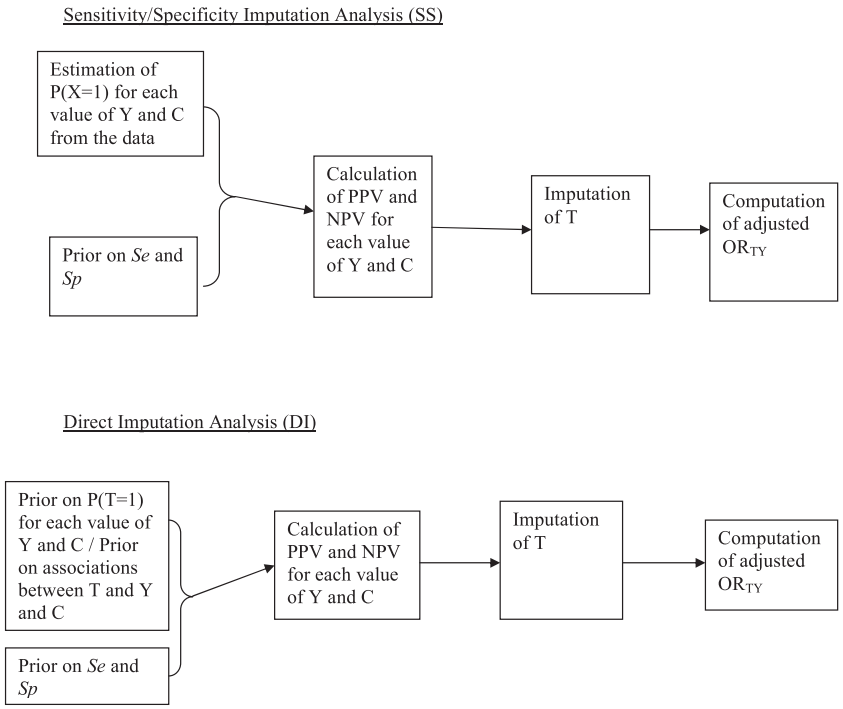

Figure 1. Steps of sensitivity/specificity imputation analysis (SS) and direct imputation analysis (DI).

Table 1. Description of priors (more details available in Appendix A, available as Supplementary data at *IJE* online)

SS FBA		
	Values for sensitivity and specificity	
Set of values	se^0	sp^0
1	0.7	0.8
2	0.7	0.9
3	0.7	1
4	0.8	0.8
5	0.8	0.9
6	0.8	1
7	0.9	0.8
8	0.9	0.9
9	0.9	1

SS PBA		
	Means [95% limits] for sensitivity and specificity	
Set of priors	se^0	sp^0
1	0.7[0.60,0.82]	0.8[0.68,0.90]
2	0.8[0.68,0.90]	0.9[0.80,0.96]
3	0.8[0.54,0.98]	0.9[0.59,0.99]
4	0.9[0.80,0.96]	0.99[0.97,1.00]

DI FBA						
	Values for sensitivity, specificity, OR_{TY}, OR_{TC} and prevalence of T = 1					
Set of values	Sensitivity	Specificity	$OR_{TY}(C=0)$	$OR_{TC}(Y=0)$	$OR_{TY}(C=1)/OR_{TY}(C=0)$	$P(T=1\|Y=0,C=0)$
1	0.7	0.8	6.93	1.11	1.59	0.40
2	0.8	0.9	6.93	1.11	1.59	0.40
3	0.9	0.99	6.93	1.11	1.59	0.40
4	0.7	0.8	3.5	1.11	1.59	0.40
5	0.8	0.9	3.5	1.11	1.59	0.40
6	0.9	0.99	3.5	1.11	1.59	0.40
7	0.7	0.8	14	1.11	1.59	0.40
8	0.8	0.9	14	1.11	1.59	0.40
9	0.9	0.99	14	1.11	1.59	0.40

(continued)

DI PBA

Set of priors	Means [95% limits] for sensitivity, specificity, OR_{TY}, OR_{TC} and prevalence of T = 1					
	Sensitivity	Specificity	$OR_{TY}(C=0)$	$OR_{TC}(Y=0)$	$OR_{TY}(C=1)/OR_{TY}(C=0)$	$P(T=1\|Y=0,C=0)$
1	0.7[0.60,0.82]	0.8[0.68,0.90]	6.93[1.76,27.44]	1.11[0.56,2.19]	1.59[0.80,3.15]	0.40[0.37,0.44]
2	0.8[0.68,0.90]	0.9[0.80,0.96]	6.93[1.76,27.44]	1.11[0.56,2.19]	1.59[0.80,3.15]	0.40[0.37,0.44]
3	0.8[0.54,0.98]	0.9[0.59,0.99]	6.93[1.76,27.44]	1.11[0.56,2.19]	1.59[0.80,3.15]	0.40[0.37,0.44]
4	0.9[0.80,0.96]	0.99[0.97,1.00]	6.93[1.76,27.44]	1.11[0.56,2.19]	1.59[0.80,3.15]	0.40[0.37,0.44]
5	0.8[0.68,0.90]	0.9[0.80,0.96]	3.5 [0.89,13.76]	1.11[0.56,2.19]	1.59[0.80,3.15]	0.40[0.37,0.44]
6	0.8[0.68,0.90]	0.9[0.80,0.96]	14[3.55,55.26]	1.11[0.56,2.19]	1.59[0.80,3.15]	0.40[0.37,0.44]

Bayesian analysis 1

Set of priors	Means [95% limits] for sensitivity, specificity, OR_{TY}, OR_{TC} and prevalence of T = 1					
	Sensitivity	Specificity	$OR_{TY}(C=0)$	$OR_{TC}(Y=0)$	$OR_{TY}(C=1)/OR_{TY}(C=0)$	$P(T=1\|Y=0,C=0)$
1	0.7[0.60,0.82]	0.8[0.68,0.90]	1[0.05,18.92]	1[0.00,2540.21]	1[0,378.4]	0.5[0.02,0.98]
2	0.8[0.68,0.90]	0.9[0.80,0.96]	1[0.05,18.92]	1[0.00,2540.21]	1[0,378.4]	0.5[0.02,0.98]
3	0.8[0.54,0.98]	0.9[0.59,0.99]	1[0.05,18.92]	1[0.00,2540.21]	1[0,378.4]	0.5[0.02,0.98]
4	0.9[0.80,0.96]	0.99[0.97,1.00]	1[0.05,18.92]	1[0.00,2540.21]	1[0,378.4]	0.5[0.02,0.98]

Bayesian analysis 2

Set of priors	Means [95% limits] for sensitivity, specificity, OR_{TY}, OR_{TC} and prevalence of T = 1					
	Sensitivity	Specificity	$OR_{TY}(C=0)$	$OR_{TC}(Y=0)$	$OR_{TY}(C=1)/OR_{TY}(C=0)$	$P(T=1\|Y=0,C=0)$
1	0.7[0.60,0.82]	0.8[0.68,0.90]	6.93[1.76,27.44]	1.11[0.56,2.19]	1.59[0.80,3.15]	0.40[0.37,0.44]
2	0.8[0.68,0.90]	0.9[0.80,0.96]	6.93[1.76,27.44]	1.11[0.56,2.19]	1.59[0.80,3.15]	0.40[0.37,0.44]
3	0.8[0.54,0.98]	0.9[0.59,0.99]	6.93[1.76,27.44]	1.11[0.56,2.19]	1.59[0.80,3.15]	0.40[0.37,0.44]
4	0.9[0.80,0.96]	0.99[0.97,1.00[	6.93[1.76,27.44]	1.11[0.56,2.19]	1.59[0.80,3.15]	0.40[0.37,0.44]
5	0.8[0.68,0.90]	0.9[0.80,0.96]	3.50[0.89,13.76]	1.11[0.56,2.19]	1.59[0.80,3.15]	0.40[0.37,0.44]
6	0.8[0.68,0.90]	0.9[0.80,0.96]	14.00[3.55,55.26]	1.11[0.56,2.19]	1.59[0.80,3.15]	0.40[0.37,0.44]

Bayesian results is adequate to the extent that the distribution of sampled parameters would be negligibly updated by a fully Bayesian analysis.

Description of the data and misclassification

The data are from a population-based lung cancer case-control study conducted in New Zealand.[19] Briefly, cases were all subjects diagnosed with incident lung cancer notified to the New Zealand Cancer Registry during 2007 and 2008 and aged 20–75 years. Controls were recruited from the New Zealand Electoral Rolls of 2003 and 2008 and were frequency matched with the cases for age and sex. For further details see Corbin *et al.*[19]

We considered the association between smoking status (ever/never) and lung cancer. The odds ratio (OR) of lung cancer for being ever-smoker vs never-smoker was estimated using unconditional logistic regression, adjusting for sex. The SAS logistic procedure (SAS V9.3) was used to estimate ORs and corresponding 95% confidence intervals (95% CI).

To provide a hypothetical reference point for evaluations, we assumed that our original dataset was correctly specified, i.e. that the 'true' smoking status indicator T was known for all subjects. We then deliberately misclassified T to X, and pretended that this was our observed measure. We attempted to use realistic misclassification rates which had been observed in previous studies. In nine studies using the cotinine validation method reported by a meta-analysis,[20] the lowest sensitivity of the self-reported smoking status was 0.82 and the lowest specificity was 0.91. We therefore took the original data, then misclassified T with a sensitivity of 0.8 and a specificity of 0.9. The misclassification was applied nondifferentially, i.e. independently of the other variables (disease status, sex). In case-control studies, the nondifferential misclassification assumption may not hold, because cases and controls may report past behaviour differently, but the methods applied here can be extended to situations where misclassification is differential.[2, 3, 16, 17]

Let Y, C and X denote the indicators for case-control status, sex (1 = Man, 0 = Woman) and misclassified smoking status, respectively, and let n_{tycx} denote the number of subjects with $T=t$, $Y=y$, $C=c$ and $X=x$. To create the misclassified smoking status X, we computed the frequencies n_{tyc+} in each of the eight combinations of the categories of T, Y and C, where a subscript '+' indicates summation over a subscript. We then calculated the frequencies of classified ever/never smokers n_{tycx} for each of these combinations as follows:

$$n_{1yc1} = n_{1yc+} \times 0.8$$
$$n_{1yc0} = n_{1yc+} \times 0.2$$
$$n_{0yc1} = n_{0yc+} \times 0.1$$
$$n_{0yc0} = n_{0yc+} \times 0.9.$$

A preliminary analysis was conducted in order to check what possible values of sensitivity and specificity could have led to the misclassified odds ratio.[23] Let π_{YC} be the proportion of subjects truly ever-smokers and π_{YC}^* the proportion of subjects classified as ever-smokers in the different strata of Y and C. Then:

$$\pi_{YC}=\frac{\pi_{YC}^* + Sp - 1}{Se + Sp - 1}$$

The proportions π_{YC} must fall in the range from 0 to 1, which implies the following restrictions:

If $se + sp > 1$

$$Se > \max_{YC}(\pi_{YC}^*) \text{ and } Sp > \max_{YC}(1 - \pi_{YC}^*)$$

If $Se + Sp < 1$

$$Se < \min_{YC}(\pi_{YC}^*) \text{ and } Sp < \min_{YC}(1 - \pi_{YC}^*)$$

Table 2 shows the proportions of subjects classified as ever-smokers π_{YC}^* and never-smokers $1-\pi_{YC}^*$ in strata of Y and C. The restrictions on Se and Sp become:

If $Se + Sp > 1$

$$Se > \pi_{11}^* \text{ (i.e. } Se > 0.76) \text{ and } Sp > 1 - \pi_{00}^* \text{ (i.e. } Sp > 0.59)$$

If $Se + Sp < 1$

$$Se < \pi_{00}^* \text{ (i.e. } Se < 0.41) \text{ and } Sp < 1 - \pi_{11}^* \text{ (i.e. } Sp < 0.24)$$

As we assumed that self-reported smoking status was classified better than chance, we only considered the case where $Se + Sp > 1$.

Table 2. Prevalences of subjects classified as exposed and non-exposed in strata of Y and C

Y	C	π_{YC}^*	$1-\pi_{YC}^*$
0	0	0.41	0.59
0	1	0.52	0.48
1	0	0.69	0.31
1	1	0.76	0.24

Results

The 'true' odds ratio of lung cancer for ever-smokers vs never-smokers adjusted for sex in the original dataset was

OR = 8.18 (95% CL: 5.86, 11.43) [log odds ratio (ln OR) = 2.10, (95% CL: 1.77, 2.44)]. After misclassifying the smoking status with a sensitivity of 0.8 and a specificity of 0.9, the estimated OR was 3.08 (95% CL: 2.40, 3.96) [ln OR = 1.13 (95% CL: 0.87, 1.38)]. Tables 3 and 4 give the results obtained with the different methods using fixed-parameter and probabilistic bias analyses, respectively.

When assuming sensitivity values (Se^0) between 0.7 and 0.9 and specificity values (Sp^0) between 0.8 and 1, SS FBA produced adjusted ORs ranging from 3.96 to 15.67 and DI FBA produced adjusted ORs between 3.88 and 17.72. Lower 95% SS and DI FBA limits went down to 2.84 and 2.97, respectively, whereas upper SS and DI FBA limits went up to 44.60 and 26.30, respectively.

As expected, for larger values of Se^0 and Sp^0, the OR obtained with SS FBA became closer to the OR obtained with the misclassified smoking status. The OR estimate appeared more sensitive to changes in the sensitivity than in the specificity of the measured exposure. When Se^0 was 0.7, the sensitivity was replaced by 0.77 in step (ii) of the algorithm (see Appendix A).

Similarly, DI FBA produced adjusted ORs closer to the OR obtained with the misclassified smoking status when we assumed higher sensitivity and specificity. However, the adjusted OR was more sensitive to the value given to the OR of lung cancer in women $OR_{TY}(C = 0)$ than to the values given to the sensitivity and specificity. When the values given to the sensitivity and the specificity were equal to the actual sensitivity and specificity of the introduced misclassification ($Se = 0.8$, $Sp = 0.9$), the OR obtained with DI FBA was very close to the value given to $OR_{TY}(C = 0)$.

Both the SS FBA and DI FBA interval estimates obtained after adjustment were wider on the logarithmic scale than the intervals obtained with the 'standard' analysis using misclassified smoking status. The intervals became narrower when increasing the sensitivity and specificity and when decreasing $OR_{TY}(C = 0)$ for DI FBA. The intervals were wider when using SS FBA than when using DI FBA, as SS FBA also attempted to account for the uncertainty in estimating the prevalence of subjects classified as ever-smokers p^*.

When assuming 95% prior limits of 0.68 and 0.90 for the sensitivity, of 0.80 and 0.96 for the specificity and an average for $OR_{TY}(C = 0)$ of 6.93, 95% of the odds ratios from SS PBA were between 2.99 and 23.17; whereas 95% from Bayesian analysis 1 (Table 1) were between 4.44 and 48.51. Using DI PBA, 95% of the odds ratios were between 3.06 and 26.07; whereas from Bayesian analysis 2, 95% were between 4.23 and 21.78. As expected, prior means for the sensitivity and the specificity equal to the actual misclassification sensitivity and specificity gave the closest median ORs to the 'true' OR.

Table 3. Smoking-lung cancer odds ratios from SS FBA and DI FBA; 95% interval estimates in brackets

Sensitivity	0.7	0.7	0.7	0.8	0.8	0.8	0.9	0.9	0.9
Specificity	0.8	0.9	1	0.8	0.9	1	0.8	0.9	1
SS FBA	15.67 [5.51,44.60]	13.00 [4.77,35.40]	11.59 [4.35,30.89]	9.54 [4.75,19.16]	8.00 [4.15,15.43]	7.18 [3.80,13.54]	5.15 [3.47,7.63]	4.38 [3.07,6.26]	3.96 [2.83,5.54]
DI FBA ($OR_{TY}(C = 0) = 3.5$)	5.40 [4.10,7.11]				4.86 [3.69,6.39]				3.88 [2.97,5.07]
DI FBA ($OR_{TY}(C = 0) = 6.93$)	9.56 [6.94,13.17]				7.60 [5.59,10.33]				5.06 [3.81,6.72]
DI FBA ($OR_{TY}(C = 0) = 14$)	17.72 [11.94,26.30]				12.89 [8.96,18.54]				7.40 [5.41,10.13]

Table 4. Smoking-lung cancer odds ratios from SS PBA, DI PBA and Bayesian analyses 1 and 2; 95% interval estimates in brackets

Sensitivity	0.70[0.60,0.82]	0.80[0.68,0.90]	0.80[0.54,0.98]	0.90[0.80,0.96]
Specificity	0.80[0.68,0.90]	0.90[0.80,0.96]	0.90[0.59,0.99]	0.99[0.97,1.00]
SS PBA	15.31[4.78,48.20]	8.18[2.99,23.17]	9.05[2.15,63.86]	4.12 [2.64,8.09]
DI PBA (mean $OR_{TY}(C=0)=3.5$)		4.83[2.11,14.09]		
DI PBA (mean $OR_{TY}(C=0)=6.93$)	9.33[2.89,36.90]	7.45[3.06,26.07]	6.81[2.93,28.26]	5.07[2.92,13.94]
DI PBA (mean $OR_{TY}(C=0)=14$)		12.59[4.45,52.00]		
Bayesian analysis 1	18.5[6.72,86.24]	11.63[4.44,48.51]	12.08[3.60,64.51]	4.75[2.93,18.48]
Bayesian analysis 2 (mean $OR_{TY}(C=0)=3.5$)		6.70[3.85,15.40]		
Bayesian analysis 2 (mean $OR_{TY}(C=0)=6.93$)	12.10[5.67,35.38]	8.03[4.23,21.78]	7.09[3.28,23.50]	4.47[3.14,8.12]
Bayesian analysis 2 (mean $OR_{TY}(C=0)=14$)		10.63[4.83,37.09]		

In SS PBA, out of 10 000 draws of initial sensitivity Se^0, 8,799 (88%), 3128 (31%), 4283 (43%) and 76 (0.76%) values for prior distributions 1, 2, 3 and 4, respectively, were lower or equal to 0.76 and were adjusted to 0.77. In draws of initial specificity Sp^0 from prior distribution 3, 415 (4%) were lower or equal to 0.59 and were adjusted to 0.60. An increase of the prior means for the sensitivity and specificity resulted in a decrease of the median ORs. When expanding the 95% limits for the sensitivity and specificity, the median ORs increased slightly, moving away from the 'true' OR, and the 95% simulation intervals (95% SI) were much wider. For DI PBA, as for DI FBA, an increase in the prior means for the sensitivity and the specificity still resulted in a decrease of the median ORs; whereas increasing the prior mean for $OR_{TY}(C=0)$ considerably increased the median ORs. Expanding the 95% limits for the sensitivity and the specificity slightly increased the median ORs and the 95% SI. Both SS and DI MCSA intervals were much wider than the interval estimates obtained with the original and the misclassified smoking status.

As with SS PBA, median ORs obtained from fully Bayesian analysis 1 decreased when increasing the sensitivity and the specificity. However, median ORs obtained from Bayesian analysis 1 were higher than median ORs obtained with SS PBA; 95% credibility intervals (95% CI) obtained from Bayesian analysis 1 were also wider than the 95% SI obtained with SS PBA, suggesting that SS PBA underestimates the uncertainty in the prevalence of true smokers in strata of *T* and *Y*.

In comparison with median ORs obtained from DI PBA, median ORs obtained from Bayesian analysis 2 were more sensitive to the prior means assigned to the sensitivity and specificity and less sensitive to the prior mean assigned to $OR_{TY}(C=0)$. Credibility intervals (95% CI) obtained from Bayesian analysis 2 were slightly narrower than DI PBA 95% SI.

When the means assigned to sensitivity and specificity equalled the actual misclassification sensitivity and specificity, the informative prior distributions used in Bayesian analysis 2 yielded median ORs closer to the 'true' OR than in Bayesian analysis 1. Credibility intervals were narrower after Bayesian analysis 2 than after Bayesian analysis 1.

Discussion

We have illustrated the use of several currently available methods for bias analysis which can be implemented using standard statistical software. Sensitivity/specificity (SS) imputation analysis has the advantage of requiring only the specification of a priori values for sensitivities and specificities. When one wishes to account for uncertainty about these values, one can specify prior distributions for the values and then sample from those.[2,13,17]

Nonetheless, the apparent simplicity of the SS approach has its own difficulties, since seemingly intuitive guesses for sensitivity and specificity may turn out to be highly implausible when compared with what one might deduce by considering the actual classification mechanism and background literature, particularly when covariates are also taken into account. Furthermore, prior distributions for sensitivity and specificity in PBA require restriction to the range of values compatible with the data (because some values may be impossible given the observed data) whereas fully Bayesian methods automatically accommodate such restrictions.[15,17]

Direction imputation (DI) analysis directly models predictive values, thus eliminating the need for constraints on sensitivity and specificity.[3] Its main limitation is that the user needs to specify values or prior distributions for coefficients about which there may be poor prior information, such as the association of interest (here, the odds ratio of lung cancer for being ever-smoker), and the resulting adjusted estimate can be very sensitive to that distribution.

Both SS and DI methods have been applied using fixed-parameter bias-sensitivity analysis (FBA) and probabilistic bias-sensitivity analysis (PBA). FBA is simpler and faster to run, since one only needs to specify fixed values. It is also very useful to check which values are compatible with the data in the SS method. Nonetheless, it does not account for uncertainty in the specification of the bias parameters. PBA takes this uncertainty into account and as a result produces wider interval estimates, thus producing inferences less sensitive to misspecification of the bias parameters.

Rough allowance for uncertainty due to random error in PBA can be made via the addition of a random number to estimates during simulation. This shortcut thus leads to fast run times, but should be used with caution as it may seriously underestimate the actual contribution of random error to uncertainty about the *TY* association; this underestimation will be a problem if uncertainty due to random error is not minor compared with uncertainty about the classification parameters. Bootstrap or jackknife methods for adding random error are preferable, but can lead to long run times; bootstrapping in particular can also encounter technical problems in small samples.[21]

The choice between SS and DI depends on what information is available. In particular, one needs to evaluate the amount and the quality of prior information to decide between setting priors on sensitivity and specificity or on regression coefficients for predictive values. When both validation data and prior information are available, all the information can be combined using data augmentation[3,24–26] in which prior distributions are translated into new data records and added to the validation data. Such an approach enables analysis with standard methods for missing data.

Bayesian procedures may be preferable to PBA, especially when one feels comfortable assigning priors to parameters beyond the classification model.[15] Our Bayesian analyses indicated that the uncertainty in the prevalence of exposure might be underestimated when using SS PBA. In addition, unlike SS PBA, Bayesian analyses do not require truncation of the prior distributions when the sensitivity or specificity prior extends below the range compatible with the data. For further analysis and contrast of SS PBA and Bayesian analyses, see Maclehose and Gustafson.[15]

It has been remarked that most epidemiologists write their methods and results sections as frequentists and their introduction and discussion sections as Bayesians.[3,27] In their methods and results sections, they analyse their data as if those are the only data that exist, and as if there is no bias left uncontrolled by the study design or by covariate adjustment (i.e. they implicitly use point-null priors on hidden bias parameters[3]). In the discussion, they then assess their results relative to background information, examining consistency with previous studies, biological plausibility and the possibility of various biases. It has been lamented, however, that in the latter discussions they severely overweight their own results, and tend to understate biases in these results, displaying especially poor intuitions about potential misclassification and measurement-error effects.[2,13,17,28]

These problems can be mitigated by including bias analyses.[2,3,13,29] FBA is particularly simple and may be useful for initial bias analyses, but we recommend PBA or Bayesian analyses when doing a risk assessment that must account for all sources of uncertainty. We have reviewed and illustrated several methods feasible using standard statistical software. We hope that sensitivity and bias analyses will become options in standard statistical packages to supplement existing methods, facilitating their conduct and presentation before inferences are offered. This will enable readers to better quantitatively assess the uncertainty warranted in the face of methodological problems.[29]

Supplementary Data

Supplementary data are available at *IJE* online.

Conflicts of interest: None declared.

Acknowledgements

This work was supported by: the Health Research Council of New Zealand; the New Zealand Department of Labour; Lottery Health Research; the Cancer Society of New Zealand; and the Accident Compensation Corporation (ACC). The Centre for Public Health Research is supported by a Programme Grant from the Health Research Council. We wish to thank Jonathan Bartlett for his comments on earlier analyses, which led to the production of this paper.

References

1. Rothman KJ, Greenland S, Lash TL. Validity in epidemiologic studies. In: *Modern Epidemiology*. 3rd edn.Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.
2. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York, NY: Springer, 2009.
3. Greenland S. Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods. *Int J Epidemiol* 2009;**38:**1662–73.
4. Gustafson P. *Measurement Error and Misclassification in Statistics and Epidemiology*. Boca Raton, FL: Chapman and Hall, 2003.
5. Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction. *Int J Epidemiol* 2006;**35:**1074–81.
6. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement Error in Nonlinear Models: A Modern Perspective* 2nd edn. Boca Raton, FL: Chapman and Hall/CRC, 2006.
7. Little RJA, Rubin DB. In: Sons JW (ed). *Statistical Analysis with Missing Data*. New York, NY: 2002.
8. Allison PD. *Missing Data*. London: SAGE Publications, 2001.

9. Daniels MJ, Hogan JW. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Abingdon, UK: Taylor & Francis, 2008.
10. Tsiatis A. *Semiparametric Theory and Missing Data*. New York, NY: Springer, 2007.
11. Schafer JL. *Analysis of Incomplete Multivariate Data*. Abingdon, UK: Taylor & Francis, 1997.
12. Carpenter J, Kenward M. Missing data in randomised controlled trials - a practical guide. Birmingham, UK: National Institute for Health Research, 2007.
13. Greenland S, Lash TL. Bias analysis. In: Rothman KJ, Greenland S, Lash TL (eds). *Modern Epidemiology*. 3rd edn. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.
14. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods*. Hoboken, NJ: Wiley, 1982.
15. MacLehose RF, Gustafson P. Is probabilistic bias analysis approximately Bayesian?. *Epidemiology* 2012;**23:**151–58.
16. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol* 2005;**34:**1370–76.
17. Greenland S. Multiple-bias modelling for analysis of observational data. *J R Stat Soc* 2005;**168:**267–91.
18. Chu H, Wang Z, Cole SR, Greenland S. Sensitivity analysis of misclassification: a graphical and a Bayesian approach. *Ann Epidemiol* 2006;**16:**834–41.
19. Corbin M, McLean D, Mannetje A *et al*. Lung cancer and occupation: A New Zealand cancer registry-based case-control study. *Am J Ind Med* 2011;**54:**89–101.
20. Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. The validity of self-reported smoking: a review and meta-analysis. *Am J Public Health* 1994;**84:**1086–93.
21. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall, 1993.
22. Hinkley D. Jackknife methods. In:*Encyclopedia of Statistical Sciences*. Hoboken, NJ: John Wiley, 2004.
23. Blettner M, Wahrendorf J. What does an observed relative risk convey about possible misclassification?. *Methods Inf Med* 1984;**23:**37–40.
24. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *Int J Epidemiol* 2006;**35:** 765–75.
25. Greenland S. Bayesian perspectives for epidemiological research. II. Regression analysis. *Int J Epidemiol* 2007;**36**(1): 195–202.
26. Greenland S. Prior data for non-normal priors. *Stat Med* 2007;**26**(19):3578–90.
27. Pearce N, Corbin M. Why we should be Bayesians (and often already are without realising it). . In: Venables K, editor. *Current topics in occupational epidemiology*. Oxford, UK: Oxford University press; 2013.
28. Fischer HJ, Greenland S, Kheifets L. Methods to explore uncertainty and bias introduced by job exposure matrices. Risk Anal. To appear;35.
29. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol* 2014;**43**(6):1969–85.