Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2005;**34**:680–687
doi:10.1093/ije/dyi060

METHODOLOGY

Proper interpretation of non-differential misclassification effects: expectations vs observations

Anne M Jurek,[1,3]* Sander Greenland,[2] George Maldonado[1] and Timothy R Church[1]

Accepted 14 February 2005

Background Many investigators write as if non-differential exposure misclassification inevitably leads to a reduction in the strength of an estimated exposure–disease association. Unfortunately, non-differentiality alone is insufficient to guarantee bias towards the null. Furthermore, because bias refers to the average estimate across study repetitions rather than the result of a single study, bias towards the null is insufficient to guarantee that an observed estimate will be an underestimate. Thus, as noted before, exposure misclassification can spuriously increase the observed strength of an association even when the misclassification process is non-differential and the bias it produced is towards the null.

Methods We present additional results on this topic, including a simulation study of how often an observed relative risk is an overestimate of the true relative risk when the bias is towards the null.

Results The frequency of overestimation depends on many factors: the value of the true relative risk, exposure prevalence, baseline (unexposed) risk, misclassification rates, and other factors that influence bias and random error.

Conclusions Non-differentiality of exposure misclassification does not justify claims that the observed estimate must be an underestimate; further conditions must hold to get bias towards the null, and even when they do hold the observed estimate may by chance be an overestimate.

Keywords Bias, epidemiological methods, misclassification, non-differential, relative risk

Under certain conditions, non-differential exposure misclassification reduces test power and biases study estimators towards the null value.[1–6] There are several versions of this non-differential misclassification rule. One often-cited version is that non-differential misclassification of a binary exposure that is independent of other errors will bias the relative-risk estimator towards the null value of 1, i.e. towards no association. It seems underappreciated that such rules are often inapplicable. As discussed previously,[7–12] additional conditions beyond non-differentiality are required to guarantee that bias is towards the null. Less well known, and perhaps surprising to some, is that bias towards the null does not always lead to an underestimate of the relative risk.[13–17] The rules refer to expected values of estimators—which is to say, the *average* result of applying a formula (estimator) to repeated samples—not to the value estimated from a specific study. Thus, it is incorrect to claim (as authors often do) that the estimate from a study must be an underestimate because the bias is towards the null.

In this paper we briefly review these issues and then present a simulation study of the relation of observed estimates to expected and true values when bias is towards the null. The simulation is intended to illustrate how often it will be wrong to claim that an observed study result must be an underestimate when the conditions for bias towards the null are satisfied.

An overview of previous results

Non-differential misclassification rules refer to the expected (average) value of relative-risk estimates over hypothetical

[1] Division of Environmental Health Sciences, School of Public Health, University of Minnesota, Minneapolis, MN 55455-0392, USA.

[2] Department of Epidemiology and Department of Statistics, University of California, Los Angeles, CA, USA.

[3] Present address. Community Health Department, Utah Valley State College, Orem, UT 84058, USA.

* Corresponding author. Community Health Department, Utah Valley State College, Orem, UT 84058, USA. E-mail: jurekan@uvsc.edu

study repetitions (more precisely, the large-sample geometric mean of the relative-risk estimator in an infinite sequence of repetitions that vary only randomly from one another). The ratio of this *expected* estimate to the true relative risk is often used as a measure of statistical bias in relative-risk estimates.[13,14,16,17] This ratio measure for bias refers only to the average error across hypothetical repetitions. In contrast, the ratio of the observed relative-risk estimate to the true relative risk measures not only bias, but also random variation.

The rule about bias towards the null is based on the classification *process* being non-differential.[16] The process refers to the behaviour of the classification procedure over hypothetical study repetitions (i.e. the probability of misclassification). Even though misclassification may be non-differential on average, due to random variation the misclassification rates in a single study (realization) will most likely be differential. Furthermore, due to random sampling variation the correctly classified estimate may be an overestimate of the true value, and the ensuing misclassification may not counterbalance this overestimation, even if the misclassification pulls the estimate towards the null. Therefore, a non-differential misclassification process does not always lead to an underestimate of relative risk. Thus, an observed estimate can be towards the null, greater than the true, or less than the null, even when the classification process obeys conditions sufficient to produce bias towards the null.

As mentioned above, non-differentiality by itself is not sufficient to guarantee that the bias is towards the null. Non-differential misclassification rules require further conditions to ensure that the bias is towards the null. First, published rules assume that the misclassification probabilities are exactly non-differential;[12] small violations of this assumption can produce substantial bias away from the null. Second, the exposure misclassification errors are assumed to be independent of errors in other variables in the analysis.[10,11,18,19] Third, further conditions are required to guarantee bias towards the null when the exposure is polytomous (>2 levels).[6,7] Fourth, the rules assume absence of interactions with other sources of systematic error, such as selection bias and confounding.

In practice it is difficult to guarantee that all these conditions are satisfied, and common practices often lead to violations of the assumptions. For example, if an exposure is continuous or polytomous with non-differential error, but it is categorized or collapsed to fewer categories in the analysis, differential misclassification can easily result.[8,9] And, if exposure is one of several measures derived from more basic data, such as one of several nutrient measures derived from a diet history, the errors in the exposure will be correlated with the errors in the other measures, thus violating one of the key assumptions needed to ensure bias towards the null when the other measures are included in the analysis.[18]

Misinterpretation of the non-differential misclassification rule dates back to 1958 at least, when Lilienfeld and Graham[20] incorrectly applied the rule to a hypothetical study result on circumcision status and cervical cancer. Although they believed differential misclassification was present in their study, they described the effect non-differential exposure misclassification could have had on one hypothetical study result, not expected study results. They stated that, were non-differential exposure misclassification present, it would have masked the true association in a study between circumcision status and cervical cancer. However, they had no way of being sure this was so.

Sorahan and Gilthorpe[15] performed simulations that show how the observed relative-risk estimate can exceed the correctly classified and true relative risk, even when all the conditions for bias towards the null are satisfied (including non-differential exposure misclassification). Thomas[13] also used simulations to demonstrate how the observed odds ratio can be greater than the correctly classified odds ratio. In a response to Thomas,[13] Weinberg *et al.*[14] suggested additional simulations of interest. Along with these suggested simulations, we present simulations to illustrate how often an observed relative risk overestimates the true value.

Simulation study

We examine the basic case of a binary exposure variable when the outcome is a correctly classified binary variable in a cohort study, using the risk ratio (incidence-proportion ratio, IPR) as the relative-risk measure. The software package used was @Risk (version 4.5).[21] The following are details of our simulation methods.

Specify simulation parameter values for simulation experiments

For each simulation experiment, the true IPR (IPR_T) was set to a value of 1, 1.5, 2, or 4. The incidence proportion (IP) in the unexposed subjects (IP_0) was set to 0.01 or 0.1. The classification probabilities (S_j) were the same for cases and non-cases, so the classification process was non-differential. Sensitivity (S_1, the probability of correctly classifying an exposed individual) and specificity (S_0, the probability of correctly classifying an unexposed individual) were examined in combinations of 1.0, 0.8, or 0.6. The true number of exposed and unexposed individuals was set to 5000 each in the first set of simulations and 1000 and 9000 in the second set, corresponding to proportions exposed of $P_E = 0.5$ and $P_E = 0.1$. This produced a total of 144 different simulation experiments (one for each combination of the four IPR_T values, two P_E values, two IP_0 values, and the three S_1 and three S_0 values).

Randomly generate one dataset

For a given set of simulation experiment parameters, in each simulation trial a dataset was generated (i.e. sampled) based on the usual assumption that random error follows a binomial distribution.[19] Therefore, the randomly generated number of exposed cases N_{11} was defined as a binomial ($10\,000P_E$, IP_0IPR_T) random variate, where IP_0IPR_T is the incidence proportion in the exposed individuals. Similarly, the randomly generated number of unexposed cases N_{10} was defined as a binomial ($10\,000(1 - P_E)$, IP_0) random variate, where IP_0 is the incidence proportion in the unexposed individuals. The resulting 2×2 table of counts are denoted N_{ij}, where the subscript i is 1 for cases, 0 for non-cases and the subscript j is 1 for exposed, 0 for unexposed.

Calculate the estimate without misclassification

On each simulation trial, we used the ratios of the randomly generated numbers of cases to the fixed number of individuals in each exposure group to calculate what the IPR estimate for

that trial would have been were there no exposure misclassification,

$$\widehat{\mathrm{IPR}}_C = (N_{11}/N_{+1})/(N_{10}/N_{+0}),$$

where the subscript C denotes correctly classified and the subscript + denotes summation over the subscript i (cases + non-cases).

Add non-differential exposure misclassification

We created a second, misclassified dataset by adding non-differential exposure misclassification to the first dataset. The probabilities of misclassification (false-negative and false-positive rates) were $1 - S_1$ and $1 - S_0$. Binomial (N_{ij}, $1 - S_j$) random variates were used to calculate the number of false-negative (Fn_{ij}, incorrectly classified as unexposed) and false-positive (Fp_{ij}, incorrectly classified as exposed) individuals where the subscript i is 1 for cases, 0 for non-cases, and the subscript j is 1 for exposed, 0 for unexposed. The exposed and unexposed cell counts after misclassification are thus:

$$M_{i1} = N_{i1} + \mathrm{Fp}_{i1} - \mathrm{Fn}_{i1} \quad \text{and} \quad M_{i0} = N_{i0} - \mathrm{Fp}_{i0} + \mathrm{Fn}_{i0}.$$

The denominators for the misclassified incidence proportions are the total numbers of individuals in each classified-exposure category, i.e. the sums of M_{+1} and M_{+0} of the number of cases and non-cases in each classified-exposure category j, where the subscript + denotes summation over the subscript i (cases + non-cases).

Table 1 Characteristics of $\widehat{\mathrm{IPR}}_M$ by IPR_T, incidence proportion in unexposed subjects, sensitivity and specificity where exposure prevalence = 0.5

			Incidence Proportion in Unexposed = 0.01				Incidence Proportion in Unexposed = 0.1			
IPR_T	Se	Sp	Geometric mean	% of simulations where $\widehat{\mathrm{IPR}}_M < 1$	% of simulations where $1 \leqslant \widehat{\mathrm{IPR}}_M \leqslant \mathrm{IPR}_T$	% of simulations where $\widehat{\mathrm{IPR}}_M > \mathrm{IPR}_T$	Geometric mean	% of simulations where $\widehat{\mathrm{IPR}}_M < 1$	% of simulations where $1 \leqslant \widehat{\mathrm{IPR}}_M \leqslant \mathrm{IPR}_T$	% of simulations where $\widehat{\mathrm{IPR}}_M > \mathrm{IPR}_T$
1.5	1.0	1.0	1.51[a]	1[a]	49[a]	50[a]	1.50[a]	0[a]	50[a]	50[a]
1.5	1.0	0.8	1.43	3	58	39	1.42	0	84	16
1.5	1.0	0.6	1.37	6	61	33	1.36	0	94	6
1.5	0.8	1.0	1.38	4	64	32	1.38	0	93	7
1.5	0.6	1.0	1.31	8	69	23	1.31	0	99	1
1.5	0.8	0.8	1.28	9	73	19	1.27	0	100	0
1.5	0.8	0.6	1.19	18	71	11	1.19	0	100	0
1.5	0.6	0.8	1.18	18	73	9	1.18	0	100	0
1.5	0.6	0.6	1.08	33	64	4	1.08	7	93	0
2	1.0	1.0	2.01[a]	0[a]	51[a]	49[a]	2.00[a]	0[a]	51[a]	49[a]
2	1.0	0.8	1.85	0	67	33	1.83	0	94	6
2	1.0	0.6	1.74	0	76	24	1.72	0	99	1
2	0.8	1.0	1.72	0	83	17	1.71	0	100	0
2	0.6	1.0	1.55	0	93	6	1.56	0	100	0
2	0.8	0.8	1.50	1	95	5	1.50	0	100	0
2	0.8	0.6	1.34	4	94	2	1.33	0	100	0
2	0.6	0.8	1.31	5	95	1	1.31	0	100	0
2	0.6	0.6	1.14	21	79	0	1.14	0	100	0
4	1.0	1.0	4.03[a]	0[a]	50[a]	50[a]	4.00[a]	0[a]	50[a]	50[a]
4	1.0	0.8	3.54	0	77	23	3.50	0	100	0
4	1.0	0.6	3.19	0	87	13	3.15	0	100	0
4	0.8	1.0	2.67	0	100	0	2.67	0	100	0
4	0.6	1.0	2.15	0	100	0	2.15	0	100	0
4	0.8	0.8	2.13	0	100	0	2.13	0	100	0
4	0.8	0.6	1.72	0	100	0	1.72	0	100	0
4	0.6	0.8	1.63	0	100	0	1.63	0	100	0
4	0.6	0.6	1.27	3	97	0	1.27	0	100	0

$\widehat{\mathrm{IPR}}_M$, misclassified incidence-proportion ratio estimate; IPR_T, true incidence-proportion ratio; Se, sensitivity; Sp, specificity; 10 000 simulation trials. Total number of exposed subjects = 5000, total number of unexposed subjects = 5000.

Percentages may not add to 100% due to rounding.

[a] $\widehat{\mathrm{IPR}}_M = \widehat{\mathrm{IPR}}_C$ (correctly classified incidence-proportion ratio estimate).

Table 2 Characteristics of $\widehat{IPR}_M$ by IPR_T, incidence proportion in unexposed subjects, sensitivity and specificity where exposure prevalence = 0.1

			Incidence Proportion in Unexposed = 0.01				Incidence Proportion in Unexposed = 0.1			
IPR_T	Se	Sp	Geometric mean	% of simulations where $\widehat{IPR}_M < 1$	% of simulations where $1 \leqslant \widehat{IPR}_M \leqslant IPR_T$	% of simulations where $\widehat{IPR}_M > IPR_T$	Geometric mean	% of simulations where $\widehat{IPR}_M < 1$	% of simulations where $1 \leqslant \widehat{IPR}_M \leqslant IPR_T$	% of simulations where $\widehat{IPR}_M > IPR_T$
1.5	1.0	1.0	1.46[a]	10[a]	42[a]	48[a]	1.50[a]	0[a]	50[a]	50[a]
1.5	1.0	0.8	1.17	23	66	11	1.18	1	99	0
1.5	1.0	0.6	1.11	30	64	6	1.11	4	96	0
1.5	0.8	1.0	1.43[b]	13	40	47	1.48	0	55	45
1.5	0.6	1.0	1.39	18	37	46	1.46	0	58	42
1.5	0.8	0.8	1.13	28	62	9	1.14	2	98	0
1.5	0.8	0.6	1.07	36	60	4	1.07	12	88	0
1.5	0.6	0.8	1.08	35	58	7	1.09	9	91	0
1.5	0.6	0.6	1.03	43	54	3	1.04	28	72	0
2	1.0	1.0	1.96[a]	1[a]	50[a]	49[a]	2.00[a]	0[a]	49[a]	51[a]
2	1.0	0.8	1.35	7	90	2	1.36	0	100	0
2	1.0	0.6	1.22	15	84	0	1.22	0	100	0
2	0.8	1.0	1.90	2	52	46	1.95	0	61	39
2	0.6	1.0	1.84	4	51	43	1.91	0	68	32
2	0.8	0.8	1.26	13	86	1	1.27	0	100	0
2	0.8	0.6	1.14	25	75	0	1.14	1	99	0
2	0.6	0.8	1.18	22	78	0	1.19	0	100	0
2	0.6	0.6	1.07	37	63	0	1.07	12	88	0
4	1.0	1.0	3.97[a]	0[a]	51[a]	49[a]	4.00[a]	0[a]	50[a]	50[a]
4	1.0	0.8	2.07	0	100	0	2.07	0	100	0
4	1.0	0.6	1.66	0	100	0	1.65	0	100	0
4	0.8	1.0	3.72	0	63	37	3.75	0	89	11
4	0.6	1.0	3.49	0	72	28	3.54	0	98	2
4	0.8	0.8	1.77	0	100	0	1.78	0	100	0
4	0.8	0.6	1.40	3	97	0	1.40	0	100	0
4	0.6	0.8	1.50	2	98	0	1.51	0	100	0
4	0.6	0.6	1.18	18	82	0	1.18	0	100	0

$\widehat{IPR}_M$, misclassified incidence-proportion ratio estimate; IPR_T, true incidence-proportion ratio; Se, sensitivity, Sp, specificity; 10 000 simulation trials. Total number of exposed subjects = 1000, total number of unexposed subjects = 9000.

Percentages may not add to 100% due to rounding.

[a] $\widehat{IPR}_M = \widehat{IPR}_C$ (correctly classified incidence-proportion ratio estimate).

[b] Unable to calculate $\ln(\widehat{IPR}_M)$ for one simulation trial since there were zero exposed cases in the misclassified dataset.

Calculate the misclassified estimate

An IPR estimate with individuals misclassified on exposure status, $\widehat{IPR}_M$, was calculated on each iteration from the misclassified counts in the second dataset,

$$\widehat{IPR}_M = (M_{11}/M_{+1})/(M_{10}/M_{+0}).$$

Analyse simulation data

Comparisons were made among the true value (IPR_T), correctly classified estimates ($\widehat{IPR}_C$), and misclassified estimates ($\widehat{IPR}_M$) on each simulation trial. The frequency of four conditions ($\widehat{IPR}_C > IPR_T$, $\widehat{IPR}_M > IPR_T$, $1 \leqslant \widehat{IPR}_M \leqslant IPR_T$, and $\widehat{IPR}_M < 1$) were all computed for each different combination of IPR_T, P_E, IP_0, and S_j values. For each simulation experiment (which consisted of 10 000 simulation trials), the geometric means of $\widehat{IPR}_C$ and $\widehat{IPR}_M$, i.e. the antilogs of the average values of $\ln(\widehat{IPR}_C)$ and $\ln(\widehat{IPR}_M)$, were calculated. The distributions of $\widehat{IPR}_C$ and $\widehat{IPR}_M$ were graphed.

The selected number of 10 000 simulation trials for each simulation experiment ensured that the widths of the 95% confidence interval for the percentages and geometric means shown in Tables 1 and 2 are <2%, e.g. a percentage shown as '50%' in the tables has 95% confidence limits within 49 and 51%.

Results

Our simulation results are shown in Figures 1 and 2 and Tables 1 and 2. We plotted the number of times each IPR

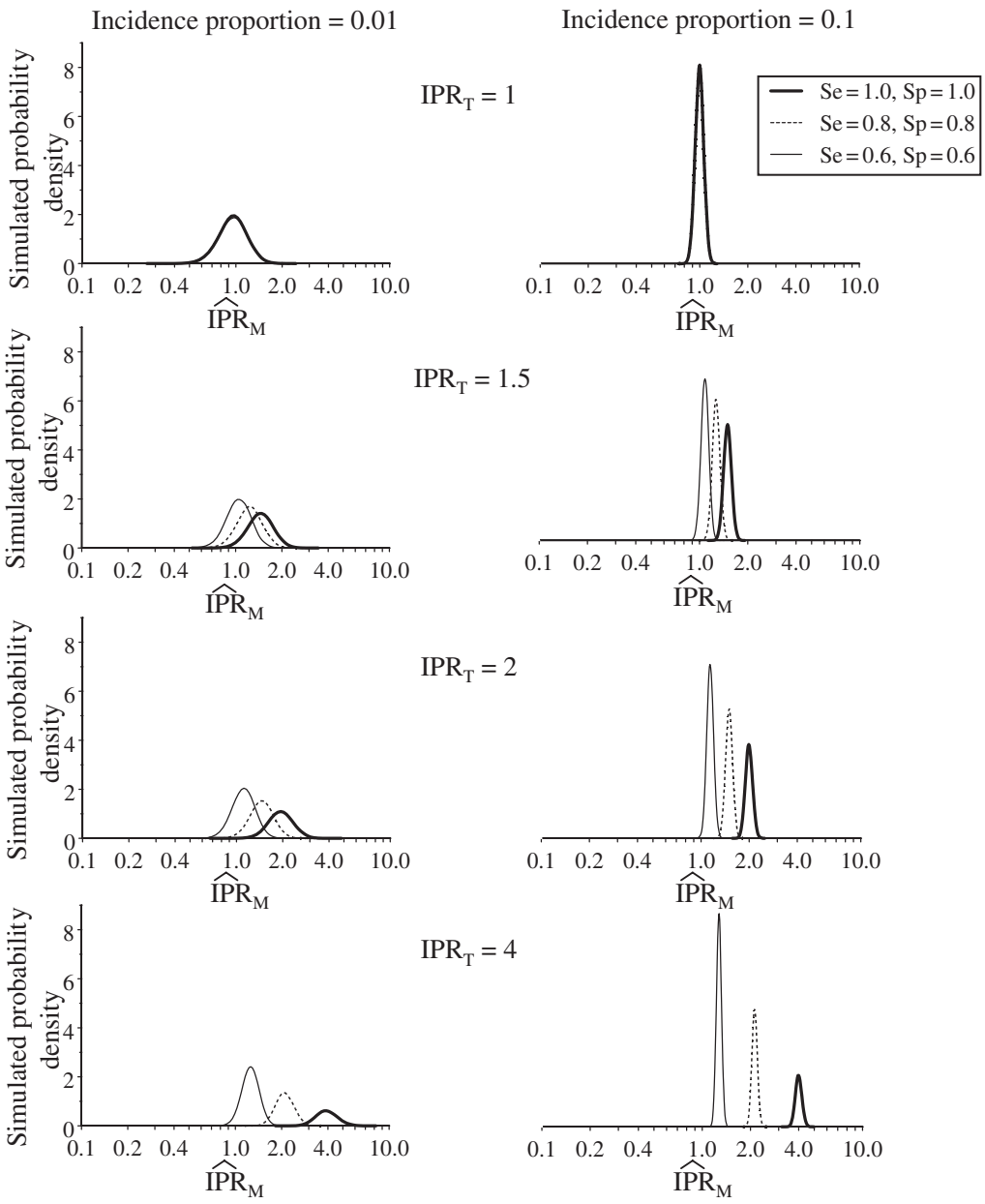

Figure 1 Distribution of $\widehat{IPR}_M$, by IPR_T, incidence proportion in unexposed subjects, sensitivity and specificity where exposure prevalence = 0.5. $\widehat{IPR}_M$, misclassified incidence-proportion ratio; IPR_T, true incidence-proportion ratio; Se, sensitivity; Sp, specificity; 10 000 simulation trials. Total number of exposed subjects = 5000, total number of unexposed subjects = 5000

estimate occurred during the 10 000 simulation trials. Each graph shows distributions of $\widehat{IPR}_M$ for three different exposure classification scenarios. For all experiments, the simulated expected value of $\widehat{IPR}_C$ was always within a few hundredths of the true value. We omit results for $IPR_T = 1$ in Tables 1 and 2 because in that case the null is true; hence there can be no bias towards the null, and both estimates always fall above (or below) the true null value ~50% of the time.

As expected, bias was towards the null in all the situations we examined. That is, the expected value of $\widehat{IPR}_M$ was always between the null value of 1 and IPR_T for our simulation results (Tables 1 and 2). The magnitude of bias depended on sensitivity, specificity, exposure prevalence, and the true value. When the true IPRs were above 1, for a balanced population structure (i.e. $P_E = 0.5$) the exposure probability was above 0.5 among cases and hence the magnitude of bias was more influenced by sensitivity than specificity. However, when the exposure was uncommon (i.e. $P_E = 0.1$), the bias was more influenced by specificity.

Nonetheless, even though bias was towards the null and the process of exposure classification was non-differential, estimates could often be greater than the true value ($\widehat{IPR}_M > IPR_T$)

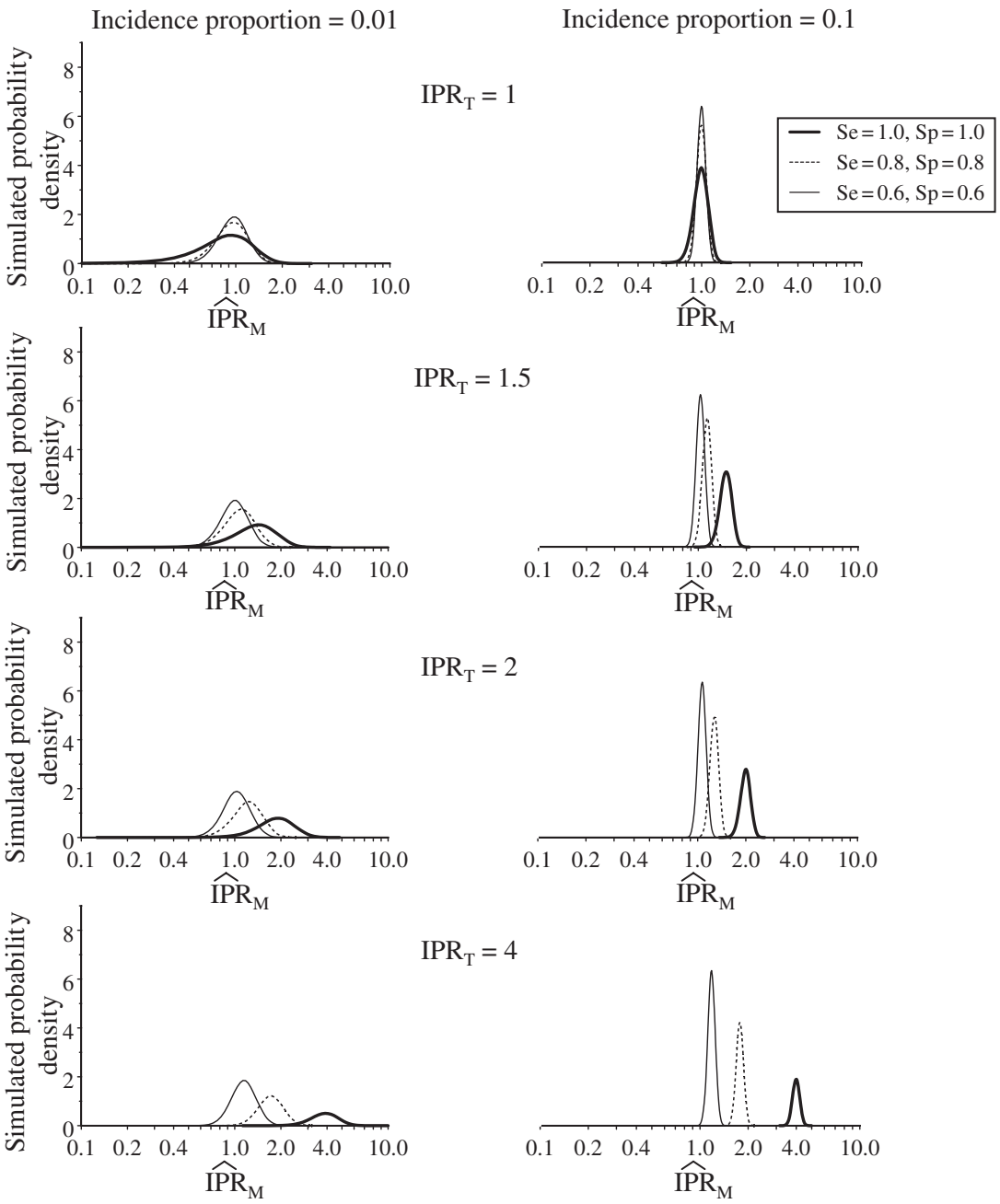

Figure 2 Distribution of $\widehat{IPR}_M$, by IPR_T, incidence proportion in unexposed subjects, sensitivity and specificity where exposure prevalence = 0.1. $\widehat{IPR}_M$, misclassified incidence-proportion ratio; IPR_T, true incidence-proportion ratio; Se, sensitivity; Sp, specificity; 10 000 simulation trials. Total number of exposed subjects = 1000, total number of unexposed subjects = 9000

(Figures 1 and 2, Tables 1 and 2). For example, when $P_E = 0.5$, $IP_0 = 0.01$, $IPR_T = 1.5$, and S_1 and $S_0 = 0.80$, $\widehat{IPR}_M > IPR_T$ in 19% of the simulation trials (Table 1). The proportion of times $\widehat{IPR}_M > IPR_T$ is indicated in Figures 1 and 2 as the area under the curves to the right of IPR_T. This proportion varied as a function of the true value, exposure prevalence, incidence proportion in the unexposed subjects, and sensitivity and specificity. It increased as the IPR_T decreased, P_E increased, IP_0 decreased, and S_1 and S_0 increased (Figures 1 and 2), situations in which the downward bias could easily be counterbalanced by upward random error. Larger sensitivity values had a greater influence on the proportion of times $\widehat{IPR}_M$ was greater than IPR_T for exposure prevalence $P_E = 0.5$, while specificity influenced this proportion more when exposure prevalence was 0.1.

Estimates were sometimes less than the null ($\widehat{IPR}_M < 1$). For example, when $P_E = 0.5$, $IP_0 = 0.01$, $IPR_T = 1.5$, and S_1 and $S_0 = 0.8$, $\widehat{IPR}_M < 1$ in 9% of the simulation trials (Table 1). The proportion of times $\widehat{IPR}_M < 1$ is indicated in Figures 1 and 2 as the area under the curves to the left of the null value. This proportion varied as a function of IPR_T, P_E, IP_0, and S_1 and S_0.

It increased as the true value, exposure prevalence, incidence proportion in the unexposed subjects, and sensitivity and specificity all decreased (Figures 1 and 2), situations in which the downward random error could easily combine with downward bias to produce large downward total error.

Under some conditions, the misclassified estimates may lie almost entirely between the null and true value ($1 \leqslant \widehat{IPR}_M \leqslant IPR_T$), as illustrated by Figures 1 and 2. For example, when $P_E = 0.1$, $IPR_T = 2$, $IP_0 = 0.1$, and S_1 and $S_0 = 0.80$, all misclassified IPR estimates were between 1 and 2, that is the entire simulated distribution is between the null and true value (Figure 2). Thus, in some of the simulations, non-differential exposure misclassification did consistently lead to underestimation of the true value. This became more and more true as the true value increased, the baseline risk became larger, the exposure prevalence increased, and the misclassification probabilities became larger (Tables 1 and 2, Figures 1 and 2), situations in which the bias on the log scale would be nearer 50% and the random error would be small. This would also become more true as the sample size increased, for then the random error would decrease and bias would be the main determinant of the results.

Discussion

Because our simulation experiments satisfied the conditions for the non-differential misclassification rule, they all resulted in bias towards the null. With no misclassification (i.e. sensitivity and specificity = 1), a correctly classified estimate would exceed the true value in roughly half of the repetitions. When bias towards the null is added, it does become less probable that the estimate will exceed the true value, as long as any resulting increase in standard error does not exceed the bias. But it does not in general guarantee that the observed estimate will be an underestimate; in particular, non-differential exposure misclassification does not always result in an observed relative-risk estimate between the null and true value, even when it produces bias towards the null. Random error alone can cause an observed relative-risk estimate to be less than one or greater than the true value.

In summary, the belief that non-differential exposure misclassification always produces an underestimate of the true value is incorrect. One reason for this incorrect belief is a failure to understand that bias is not the only, or even necessarily, the main, source of error in an estimate. Bias is not the ratio of the observed estimate from one study to the true value, because the observed estimate also incorporates random errors. The latter errors diminish with sample size but remain substantial in most epidemiological studies, as revealed by the width of the confidence interval.

Another reason for the incorrect belief is the failure to recognize that non-differentiality is insufficient to guarantee the bias is towards the null; other conditions must be satisfied, especially independence of errors.[10,11,18,19] Furthermore, even small departures from non-differentiality can produce bias away from the null.[12] Finally, even when bias due to exposure misclassification is towards the null, other biases (such as confounding, selection bias, and mismeasurement of covariates) can cause the total bias to be away from the null. The combined effect from all biases must be considered when interpreting study results.[22–25]

When biases are a serious concern (the usual case in epidemiology) and the study results are of potential policy importance, we recommend using quantitative methods for evaluating the effect of not only exposure misclassification but also other systematic errors.[19,22–30] Methods such as sensitivity analysis, uncertainty analysis, and bias modelling provide a means to account for systematic error in ways that do not depend on traditional and often faulty qualitative heuristics. These methods can be extended and given empirical grounding by the addition of 'validation' or reproducibility data to the analysis,[31,32] although very large amounts of such data may be needed to provide effective corrections,[33] and use of incorrect non-differentiality assumptions in the ensuing analyses may worsen bias.[34,35] When such methods are not used, we recommend that results be presented in a very cautious and descriptive manner, rather than promoted by unfounded judgments that biases are small or errors are in a known direction. We believe a very descriptive approach to presenting results is often commendable, and need not detract from the scientific value of a research report.[36]

Simulation results such as ours are sensitive to the conditions chosen for the simulation. In practice one cannot show that a single epidemiological study has the conditions identical to those assumed in a simulation experiment. Thus, it may not be possible to generalize the details of our simulation results to many practical situations. They do not, for example, show how relaxing the strict non-differentiality or independence conditions would affect the behaviour of estimates. Nonetheless, they do show that even if the conditions are perfectly satisfied, the results may be an overestimate, with the probability of overestimation depending on the size of bias and the distribution of random errors. This point is important because any study is but one replication, and hence the results reflect random error as well as bias. Thus, unless the confidence intervals from the study are very narrow (suggesting random error is small), researchers should not infer that their results are an underestimate, even if they can persuasively argue that the net bias from all sources (not just exposure misclassification) is towards the null; the most that can be said is that the results are more likely an underestimate than an overestimate.

Acknowledgements

This research has been supported by a grant from the US Environmental Protection Agency's Science to Achieve Results (STAR) programme.

Disclaimer

Although the research described in the article has been funded in part by the US Environmental Protection Agency's STAR programme through grant (U-91615801-0), it has not been subject to any EPA review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

References

[1] Bross I. Misclassification in 2 × 2 tables. *Biometrics* 1954;**10:**478–86.

[2] Newell DJ. Errors in the interpretation of errors in epidemiology. *Am J Public Health* 1962;**52:**1925–28.

[3] Keys A, Kihlberg JK. Effect of misclassification on estimated relative prevalence of a characteristic: I. Two populations infallibly distinguished. II. Errors in two variables. *Am J Public Health* 1963;**53:**1656–65.

[4] Gullen WH, Bearman JE, Johnson EA. Effects of misclassification in epidemiologic studies. *Public Health Rep* 1968;**83:**914–18.

[5] Goldberg JD. The effects of misclassification on the bias in the difference between two proportions and the relative odds in the fourfold table. *J Am Stat Assoc* 1975;**70:**561–67.

[6] Weinberg CR, Umbach DM, Greenland S. When will nondifferential misclassification of an exposure preserve the direction of a trend? *Am J Epidemiol* 1994;**140:**565–71.

[7] Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol* 1990;**132:**746–48.

[8] Wacholder S, Dosemeci M, Lubin JH. Blind assignment of exposure does not always prevent differential misclassification. *Am J Epidemiol* 1991;**134:**433–37.

[9] Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. *Am J Epidemiol* 1991;**134:**1233–44.

[10] Kristensen P. Bias from nondifferential but dependent misclassification of exposure and outcome. *Epidemiology* 1992;**3:**210–15.

[11] Chavance M, Dellatolas G, Lellouch J. Correlated nondifferential misclassifications of disease and exposure: application to a cross-sectional study of the relation between handedness and immune disorders. *Int J Epidemiol* 1992;**21:**537–46.

[12] Maldonado G, Greenland S, Phillips C. Approximately nondifferential exposure misclassification does not ensure bias toward the null [Abstract]. *Am J Epidemiol* 2000;**151:**S39.

[13] Thomas DC. Re: 'When will nondifferential misclassification of an exposure preserve the direction of a trend?' *Am J Epidemiol* 1995;**142:**782–83.

[14] Weinberg CR, Umbach DM, Greenland S. Weinberg *et al*. reply [Letter]. *Am J Epidemiol* 1995;**142:**784.

[15] Sorahan T, Gilthorpe MS. Non-differential misclassification of exposure always leads to an underestimate of risk: an incorrect conclusion. *Occup Environ Med* 1994;**51:**839–40.

[16] Wacholder S, Hartge P, Lubin JH, Dosemeci M. Non-differential misclassification and bias towards the null: a clarification. *Occup Environ Med* 1995;**52:**557–58.

[17] Sorahan T, Gilthorpe MS. Sorahan and Gilthorpe reply [Letter]. *Occup Environ Med* 1995;**52:**558.

[18] Lash TL, Fink AK. Re: 'Neighborhood environment and loss of physical function in older adults: evidence from the Alameda County study'. *Am J Epidemiol* 2003;**157:**472–73.

[19] Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd edn. Philadelphia, PA: Lippincott-Raven, 1998.

[20] Lilienfeld AM, Graham S. Validity of determining circumcision status by questionnaire as related to epidemiological studies of cancer of the cervix. *J Natl Cancer Inst* 1958;**21:**713–20.

[21] Guide to using @Risk. Newfield, NY: Palisade Corporation, 2002.

[22] Lash TL, Fink AK. Semi-Automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003;**14:** 451–58.

[23] Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology* 2003;**14:**459–66.

[24] Greenland S. Multiple bias modeling for analysis of epidemiologic data (with discussion). *J R Stat Soc A* 2005;**168:**267–308.

[25] Phillips CV, Maldonado G. Using Monte Carlo methods to quantify the multiple sources of error in studies [Abstract]. *Am J Epidemiol* 1999;**149:**S17.

[26] Eddy DM, Hasselblad V, Shachter R. *Meta-Analysis by the Confidence Profile Method*. Boston: Academic Press Inc., 1992.

[27] Morgan MG, Henrion M. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. New York: Cambridge University Press, 1990.

[28] Vose D. *Risk Analysis: A Quantitative Guide*. 2nd edn. New York: John Wiley & Sons, 2000.

[29] Greenland S. The impact of prior distributions for uncontrolled confounding and response bias. *J Am Stat Assoc* 2003;**98:**47–54.

[30] Steenland K, Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am J Epidemiol* 2004;**160:** 384–92.

[31] Carroll RJ, Ruppert D, Stefanski L. *Measurement Error in Nonlinear Models*. New York: Chapman and Hall, 1995.

[32] Espeland M, Hui SL. A general approach to analyzing epidemiologic data that contain misclassification errors. *Biometrics* 1987; **43:**1001–12.

[33] Greenland S. Statistical uncertainty due to misclassification: implications for validation substudies. *J Clin Epidemiol* 1988; **41:**1167–76.

[34] Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. *Am J Epidemiol* 1993;**137:**1251–58.

[35] Lagarde F, Alfredsson L. Re: 'Validation studies using an alloyed gold standard.' *Am J Epidemiol* 1996;**143:**1175–76.

[36] Greenland S, Gago-Domiguez M, Castellao JE. The value of risk-factor ('black-box') epidemiology (with discussion). *Epidemiology* 2004;**15:**519–35.