

# Chapter 9. The impact of misclassification in case-control studies of gene-environment interactions

Nathaniel Rothman, Montserrat Garcia-Closas, Walter T. Stewart and Jay Lubin

**In this Chapter we describe the impact of risk factor misclassification in case-control studies designed to estimate gene-environment interactions. We show that under certain scenarios even small amounts of exposure or genotype misclassification can substantially attenuate the interaction effect and, as a consequence, dramatically increase the sample size required to study these interactions. A consideration of how sample size is affected by exposure and genotype misclassification in the study design phase should help to identify situations where obtaining better risk factor information is crucial for the feasibility of studies.**

The misclassification of risk factor information is a common problem in epidemiological studies. Sources of exposure misclassification have been extensively described (Armstrong *et al.*, 1992) and misclassification errors of genetic polymorphisms, which can occur in phenotype or genotype assays, are discussed in the current volume by Vineis and Malats (Chapter 6), Benhamou *et al.* (Chapter 7), Pelkonen *et al.* (Chapter 8) and Blömeke and Shields (Chapter 13). We first consider the effects of misclassification in studies of a single risk factor (e.g. genotypes or environmental exposures, broadly defined as exogenous or endogenous carcinogenic agents), and then examine the effects of misclassification in studies of multiplicative interactions between factors. In each section we review the effects of misclassification on the estimated odds ratio (OR) and required sample size, and then illustrate the effects with examples.

## Misclassification in studies of one risk factor

The impact of the misclassification of risk factors on estimates of risk and sample size has been extensively addressed (Bross, 1954; Lillienfeld, 1962; Copeland *et al.*, 1977; Shy *et al.*, 1978; Gladen & Rogan, 1979; Greenland, 1980; Flegal *et al.*, 1986; Rothman *et al.*, 1993). In the present Chapter we focus on the effect of misclassification

of a dichotomous exposure or genotype (i.e. exposed versus unexposed or susceptible versus non-susceptible) on the estimation of the odds ratio measuring the association between the risk factor and disease and on the required sample size to study this association. The true odds ratio is denoted as  $OR_T$  and the observed or estimated odds ratio as  $OR_O$ . The evaluation of the effects of misclassifying variables with more than two categories or which are continuous measures is beyond the scope of this Chapter.

Misclassification of a dichotomous risk factor is defined by two probabilities: sensitivity (the probability of correctly classifying risk factor positive subjects) and specificity (the probability of correctly classifying risk factor negative subjects). When misclassification is non-differential with regard to disease status, that is, the sensitivity and specificity do not depend on case or control status, the  $OR_O$  is generally biased towards the null value of no association (Copeland *et al.*, 1977; Flegal *et al.*, 1986). As a consequence, for a given level of statistical power, a larger sample size is required to detect the attenuated  $OR_O$  (Armstrong *et al.*, 1992). The impact of misclassification depends on (1) the prevalence of the risk factor among the controls and (2) the magnitude of the  $OR_T$ . Reduced sensitivity tends to have a stronger impact on the magnitude of bias in the  $OR_O$

when the prevalence of the risk factor is high rather than low, and reduced specificity tends to have a stronger impact on the magnitude of the bias when the prevalence is low (Flegal *et al.*, 1986). For a given prevalence of the risk factor the bias in the  $OR_O$  increases with the magnitude of the  $OR_T$  (Flegal *et al.*, 1986).

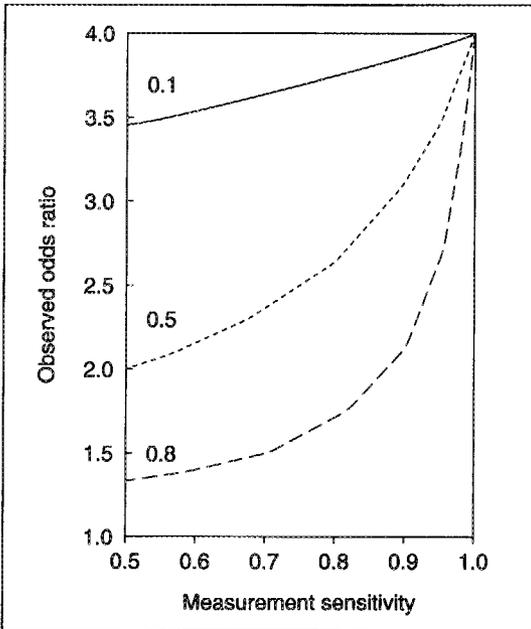
As pointed out by Flegal *et al.* (1986), the effects of exposure misclassification on measures of relative risk are complex and not easily generalized, and the potential degree of bias should therefore be evaluated in each particular situation. We illustrate below the effects of misclassification in particular examples which may be of interest in studies of genetic susceptibility to cancer.

**Examples**

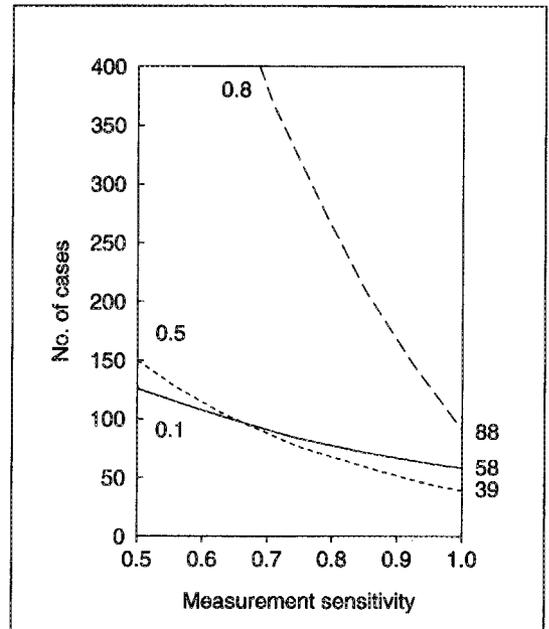
The examples below illustrate the outlined principles in the study of an exposure or genetic factor associated with a fourfold increase in risk of disease ( $OR_T = 4$ ). The  $OR_O$  in the presence of misclassification is a function of the  $OR_T$ , the true

prevalence of the risk factor in the controls, and the sensitivity and specificity of the risk factor classification (Kleinbaum *et al.*, 1982; Flegal *et al.*, 1986). To illustrate the effects of misclassification we calculated the expected  $OR_O$  for a range of values for prevalence, sensitivity and specificity using previously published formulae (Kleinbaum *et al.*, 1982; Flegal *et al.*, 1986). We then estimated the number of subjects necessary to achieve 80% power to detect the expected  $OR_O$ , using a two-sided test at the 5% level as described by Schlesselman (1974).

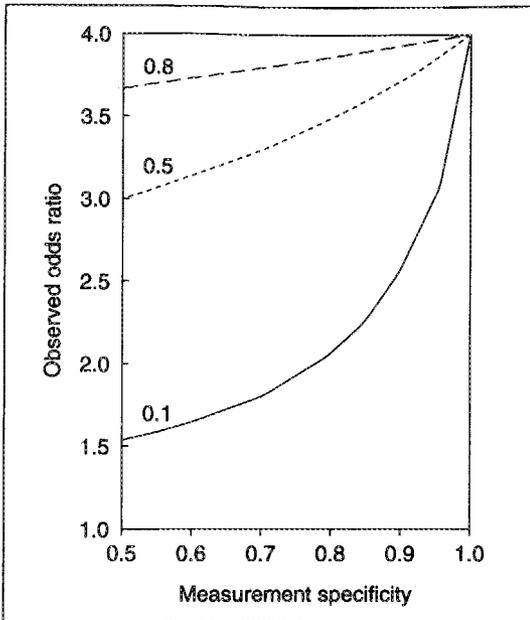
Figs. 1a and 1b show how the  $OR_O$  and sample size requirement change as risk factor sensitivity ranges from 0.5 to 1.0 and as the prevalence of the risk factor varies over 0.1, 0.5, and 0.8 (given  $OR_T = 4$ , a 1:1 case to control ratio, non-differential misclassification, perfect specificity and a desired 80% power). The three lines represent the three different risk factor prevalences. Under these conditions, sensitivity has the greatest impact on the  $OR_O$  and sample size requirements at medium to high prevalences of the risk factor



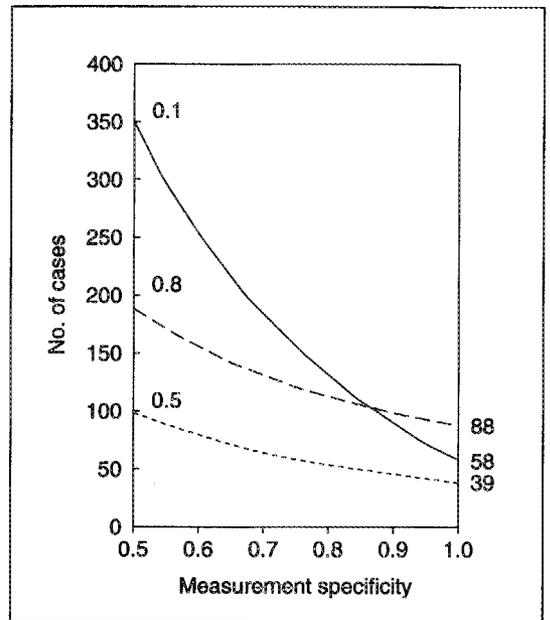
**Figure 1a.** Observed odds ratio as a function of measurement sensitivity for different prevalences of the risk factor (\_\_\_\_ 0.1 prevalence, .... 0.5 prevalence, - - - 0.8 prevalence). True odds ratio equals 4.0.



**Figure 1b.** Number of cases required to have 80% power to detect an odds ratio of 4.0 using a two-sided test at the 5% level, as a function of measurement sensitivity for different prevalences of the risk factor (\_\_\_\_ 0.1 prevalence, .... 0.5 prevalence, - - - 0.8 prevalence). Case:control ratio is 1:1.



**Figure 2a.** Observed odds ratio as a function of measurement specificity for different prevalences of the risk factor (— 0.1 prevalence, .... 0.5 prevalence, --- 0.8 prevalence). True odds ratio equals 4.0.



**Figure 2b.** Number of cases required to have 80% power to detect an odds ratio of 4.0 using a two-sided test at 5% level, as a function of measurement specificity for different prevalences of the risk factor (— 0.1 prevalence, .... 0.5 prevalence, --- 0.8 prevalence). Case:control ratio is 1:1.

and this impact decreases as the prevalence becomes lower. For example, when the prevalence is 0.5, a sensitivity of 0.9 results in an  $OR_0$  of 3.1 and requires 52 cases and 52 controls, while a sensitivity of 0.7 results in an  $OR_0$  of 2.4 and requires 88 cases and 88 controls.

Figs. 2a and 2b illustrate the same scenario except that sensitivity is perfect (1.0) and specificity varies from 0.5 to 1.0. Under these conditions, specificity has the greatest impact on the  $OR_0$  and sample size requirements at low prevalences. For example, when the risk factor prevalence is 0.1, a specificity of 0.9 results in an  $OR_0$  of 2.6 and requires 90 cases and 90 controls, while a specificity of 0.7 results in an  $OR_0$  of 1.8 and requires 183 cases and 183 controls.

The knowledge that risk factor prevalence determines the relative importance of sensitivity and specificity of exposure assessment or genetic analysis can assist investigators to select techniques that minimize the impact of misclassification in their studies. For example, a genotype assay with essentially perfect specificity but with

slightly less than perfect sensitivity (e.g. due to alleles not detected by the assay) would have a minimal impact on the  $OR_0$  for a relatively low prevalence allele but could have a more substantial impact for a high prevalence allele.

#### Misclassification in studies of a multiplicative interaction between two risk factors

This section illustrates the effects of misclassification in case-control studies which seek to determine if a disease-exposure association, as measured by the OR, varies for subjects with and without a hypothesized at-risk genotype. A multiplicative interaction implies that the OR for people exposed to both the at-risk genotype and the exposure (i.e. the joint OR) is greater than the product of the OR for the genotype and exposure alone. The interaction effect is the factor by which the joint OR is different from the multiplication of the genetic and exposure effects individually. We focus on the study of multiplicative interaction. However, studies of departures from additive models may also be of

interest (Pearce, 1989), especially when the goal is to estimate disease frequency reduction (Kleinbaum *et al.*, 1982).

In studies where the aim is to investigate the presence of a multiplicative interaction between two risk factors and disease, the sample size required is generally much larger than if the aim is to detect a single risk factor effect (Smith & Day, 1984; Lubin & Gail, 1990). The required sample size depends on the true magnitude of association with disease and the true prevalences of the two risk factors, as well as on the sensitivity and specificity for each risk factor (Greenland, 1983).

The effect of misclassification in the assessment of interactions has received only limited attention (Greenland, 1980; Flegal *et al.*, 1986; Cox & Elwood, 1991). If the genotype and environmental exposure are independent in the population and misclassification of either is non-differential with regard to both disease status and each other, the interaction effect tends to be biased towards the null value (Greenland, 1980). Moreover, in general when misclassification of exposure is differential with regard to disease status, which may occur in case-control studies, but is non-differential with regard to genotype, the interaction effect is also biased towards the null (Garcia-Closas *et al.*, 1997).

### Examples

Here we illustrate the impact of misclassification on the study sample size and the bias of both the interaction effect and the joint OR. To simplify matters the genotype and environmental exposures are defined as dichotomous variables and are assumed to be independent of each other in the population. Furthermore, misclassification of exposure and genotype are assumed to be independent of each other and disease status. The observed ORs in the presence of misclassification were calculated using previously published formulae which express the expected cell counts from a  $2 \times 2 \times 2$  table cross-classifying disease, exposure and genotype as a function of the true cell counts and the classification probabilities (i.e. sensitivity and specificity) (Kleinbaum *et al.*, 1982). Sample size calculations were performed as described by Lubin & Gail (1990).

For the examples presented here, genotype prevalence is fixed at 0.5, whereas exposure

**Table 1. True odds ratios used in an example of gene-environment interaction**

	Odds ratios	
	Genotype	
	-	+ (at-risk allele)
Exposure -	1.0	2.0
Exposure +	2.0	12.0

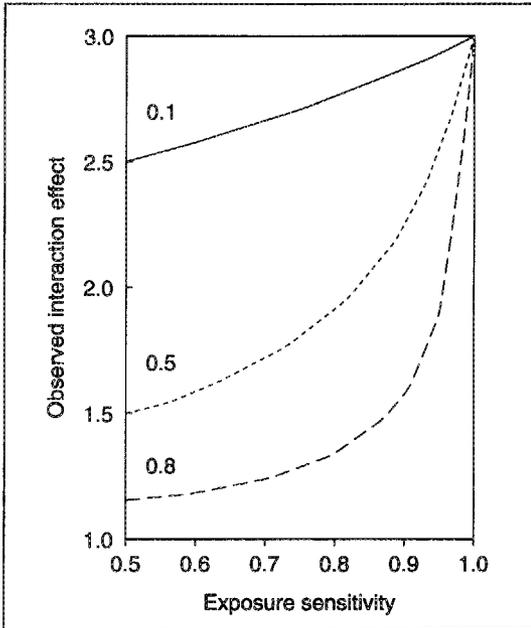
prevalence takes values of 0.1, 0.5 and 0.8. As shown in Table 1, the effect of the at-risk genotype in the absence of the environmental exposure and the effect of the exposure in the absence of the at-risk genotype have been arbitrarily set at 2.0, the joint effect (or joint OR) of the genotype and exposure has been set at 12.0, and the interaction effect has been set to 3.0 (i.e.  $12.0/(2.0 \times 2.0)$ ). The gene-specific exposure ORs are 6.0 (genotype (+)) =  $12.0/2.0$  and 2.0 (genotype (-)) =  $2.0/1.0$ . The numbers shown in Table 1 are true and expected parameters that do not depend on the number of cases and controls.

Perfect sensitivity and specificity are rarely attained in the measurement of environmental exposures. An environmental exposure assessment method that resulted in 0.8 sensitivity would generally be considered excellent. Table 2 shows the  $OR_T$  and  $OR_O$  for genotype and exposure when the prevalence of each is 0.5,

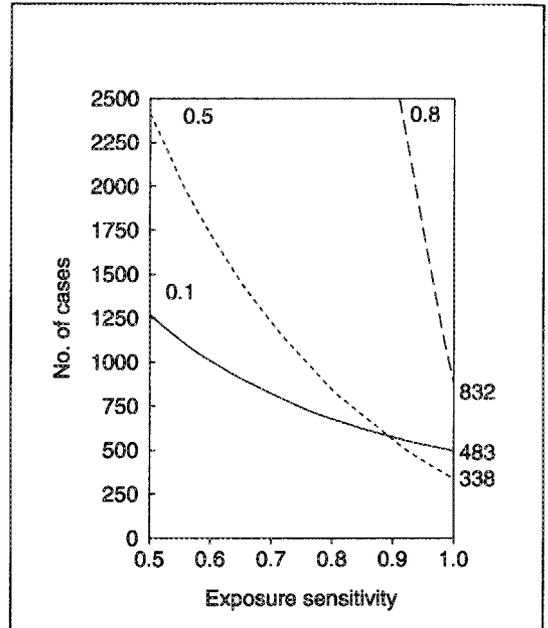
**Table 2. True odds ratios and observed odds ratios (shown in parentheses) when exposure sensitivity is 0.8<sup>1</sup>**

	True (observed) odds ratios	
	Genotype	
	-	+ (at-risk allele)
Exposure -	1.0	2.0 (3.1)
Exposure +	2.0 (1.7)	12.0 (10.3)

<sup>1</sup> Prevalence of both exposure and at-risk allele among controls = 50%; exposure specificity, genotype sensitivity and genotype specificity = 100%.



**Figure 3a.** Observed interaction effect as a function of exposure sensitivity for different exposure prevalences (— 0.1 prevalence, .... 0.5 prevalence, - - - 0.8 prevalence). True interaction effect is 3.0, the true ORs of disease given exposure among non-susceptibles and given genotype among unexposed are both 2.0, and the genotype prevalence is 0.5.

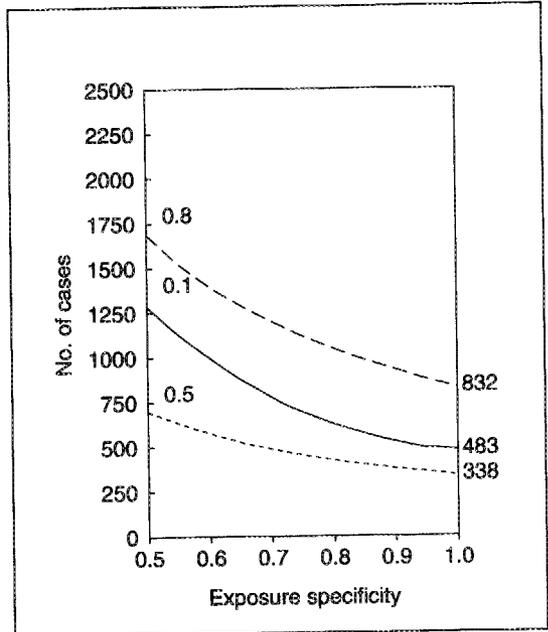
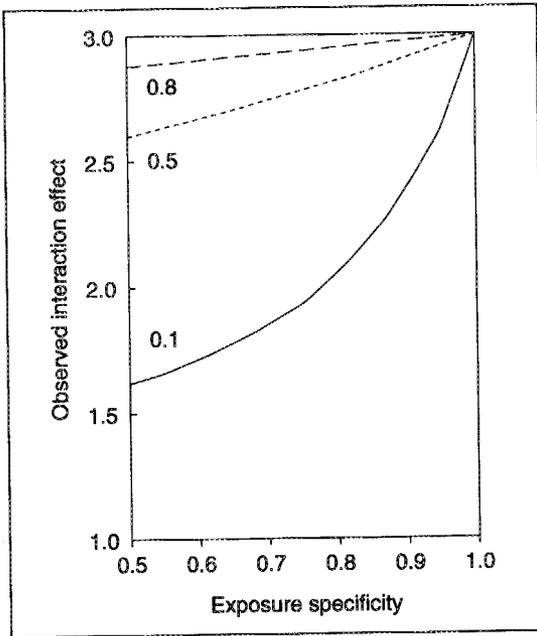


**Figure 3b.** Number of cases required to have 80% power to detect a threefold interaction using a two-sided test at 5% level, as a function of exposure sensitivity for different exposure prevalences (— 0.1 prevalence, .... 0.5 prevalence, - - - 0.8 prevalence). The true ORs of disease given that exposure among non-susceptibles and genotype among unexposed are both 2.0, and the genotype prevalence is 0.5. Case:control ratio is 1:1.

exposure sensitivity is 0.8 and exposure specificity is 1.0. As we would expect from the non-differential nature of the misclassification, the gene-specific exposure ORs (genotype (+) = 10.3/3.1; genotype (-) = 1.7/1.0) are biased towards the null value. On the other hand, the genotype OR among unexposed subjects is biased away from the null (from 2.0 to 3.1). This reflects the fact that although the genotype is perfectly measured in this example, the observed genotype OR among the unexposed in the presence of exposure misclassification is a weighted average of the genotype OR among truly unexposed subjects and truly exposed subjects who were classified as unexposed. The interaction effect is attenuated from 3.0 to 1.95 ( $10.3/(1.7 \times 3.1)$ ), and the sample size required to have 80% power increases from 338 cases and 338 controls with perfect exposure assessment to 847 cases and 847 controls with 0.8 sensitivity.

Figs. 3a and 3b illustrate the impact on the interaction effect of varying exposure sensitivity from 0.5 to 1.0 while holding exposure specificity at 1.0. Again, the three lines represent three different exposure prevalences. The impact of sensitivity on the observed interaction effect is much greater for common exposures than for rare exposures (Fig. 3a). The attenuation in the observed interaction effect translates into an increased required sample size (Fig. 3b). For example, for an exposure prevalence of 0.5, lowering exposure sensitivity from 0.9 to 0.7 more than doubles the required sample size (from 560 to 1223 cases). At an exposure prevalence of 0.8, the sample size is dramatically increased even by small amounts of exposure inaccuracy, calling into question the feasibility of the study.

Figs. 4a and 4b illustrate results obtained when exposure sensitivity is held at 1.0 and exposure specificity is varied from 0.5 to 1.0.



**Figure 4a.** Observed interaction effect as a function of exposure specificity for different exposure prevalences (— 0.1 prevalence, .... 0.5 prevalence, - - - 0.8 prevalence). The true interaction effect is 3.0, the true ORs of disease given that exposure among non-susceptibles and genotype among unexposed are both 2.0, and genotype prevalence is 0.5.

**Figure 4b.** Number of cases required to have 80% power for detection of a threefold interaction using a two-sided test at 5% level, as a function of exposure specificity for different exposure prevalences (— 0.1 prevalence, .... 0.5 prevalence, - - - 0.8 prevalence). The true ORs of disease given that exposure among non-susceptibles and genotype among unexposed are both 2.0. Case:control ratio is 1:1.

They show that, under our assumptions, the effect of exposure specificity is greater for rare than for common exposures. For an exposure prevalence of 0.1, lowering exposure specificity from 0.9 to 0.7 increases the required sample

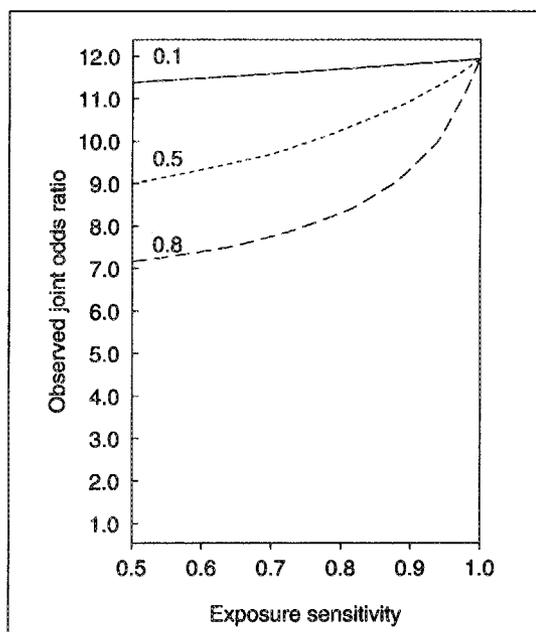
size by about 50% (from 523 to 784 cases), whereas the effect is smaller for exposure prevalences of 0.5 and 0.8.

Figs. 5a and 5b illustrate the effect of exposure sensitivity and specificity on the observed joint

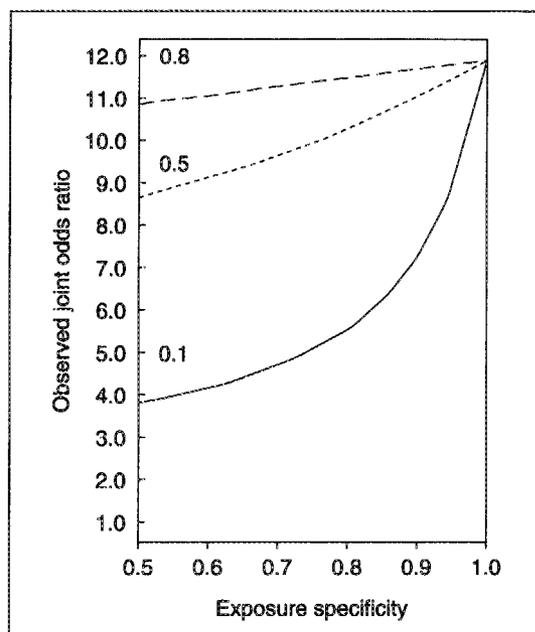
**Table 3. Impact of exposure and genotype misclassification on sample size requirements to detect gene-environment interaction<sup>1</sup>**

Example	Exposure accuracy Sensitivity	Genotype accuracy Sensitivity	# cases needed for 80% power to detect interaction
1	1.0	1.0	338
2	1.0	0.95	442
3	0.8	1.0	847
4	0.8	0.95	1156

<sup>1</sup> Interaction model described in Table 1; prevalence of both at-risk genotype and exposure = 0.5 among controls; genotype and exposure specificity = 1.0; 1:1 case to control ratio.



**Figure 5a.** Observed joint odds ratio as a function of exposure sensitivity for different exposure prevalences (— 0.1 prevalence, .... 0.5 prevalence, - - - 0.8 prevalence). The true joint odds ratio is 12.0, the true odds ratios of disease given that exposure among non-susceptibles and genotype among unexposed are both 2.0, and the genotype prevalence is 0.5.



**Figure 5b.** Observed joint odds ratio as a function of exposure specificity for different exposure prevalences (— 0.1 prevalence, .... 0.5 prevalence, - - - 0.8 prevalence). The true joint odds ratio is 12.0, the true ORs of disease given that exposure among non-susceptibles and genotype among unexposed are both 2.0, and the genotype prevalence is 0.5.

OR across exposure prevalences of 0.1, 0.5 and 0.8. The true joint OR has been set equal to 12 (Table 1). There is only a relatively small to modest decline in the observed joint OR as exposure sensitivity decreases to 0.5 for all three exposure prevalences. In contrast, imperfect exposure specificity can substantially reduce the observed joint OR at an exposure prevalence of 10%.

In exploring the effects of exposure misclassification we have assumed that the genotype was perfectly measured. However, misclassification of genetic status can occur (Cascorbi *et al.*, 1995; Blömeke & Shields, Chapter 13) and has an additional influence on sample size requirements. Table 3 illustrates the impact of changing genotype sensitivity from 1.0 to 0.95 on sample size, both in the absence and presence of non-differential exposure misclassification. In the absence of exposure misclassification, a genotype sensitivity of 0.95 increases the sample size by about 30% (from 338 to 442 cases). When exposure sensitiv-

ity is 0.8 rather than 1.0, the increase due to the genotype error is 36% (from 847 to 1156). Thus, even small genotype errors can substantially increase the required sample size to detect a gene-environment interaction, and this increase is greater when the exposure is also misclassified.

### Concluding remarks

These examples show that, in certain scenarios, even small amounts of exposure misclassification can substantially attenuate the interaction effect, and therefore substantially increase the already large sample size required to study gene-environment interactions. Errors in genotype determination further increase the required sample size. Considering how sample size is affected by exposure misclassification in the study design phase helps to identify situations where obtaining better exposure information is crucial for the feasibility of the study.

## References

- Armstrong, B.K., White, E. & Saracci R. (1992) *Principles of Exposure Measurement in Epidemiology*. Oxford, Oxford University Press, pp. 49-136
- Bross, I. (1954) Misclassification in 2 X 2 tables. *Biometrics*, 10, 478-486
- Cascorbi, I., Drakoulis, N., Brockmüller, J., Maurer, A., Sperling, K. & Roots, I. (1995) Arylamine N-acetyltransferase (NAT2) mutations and their allelic linkage in unrelated Caucasian individuals: correlation with phenotypic activity. *Am. J. Hum. Genet.*, 57, 581-592
- Copeland, K.T., Checkoway, H., McMichael A.J. & Holbrook, R.H. (1977) Bias due to misclassification in the estimation of relative risk. *Am. J. Epidemiol.*, 105, 488-495
- Cox, B. & Elwood, M.J. (1991) The effect on the stratum-specific odds ratios of non-differential misclassification of a dichotomous covariate. *Am. J. Epidemiol.*, 15, 202-207
- Diamond, E. & Lilienfeld, A.M. (1962) Effects of errors in classification and diagnosis in various types of epidemiological studies. *Am. J. Public Health*, 52(11), 37-44
- Flegal, K.M., Brownie, C. & Haas, J.D. (1986) The effects of exposure misclassification on estimates of relative risk. *Am. J. Epidemiol.*, 123, 736-751
- Garcia-Closas, M., Thompson, D.W. & Robins, J.M. (1998) Differential misclassification and the assessment of gene-environment interactions in case-control studies. *Am. J. Epidemiol.*, 147, 426-433
- Gladen, B. & Rogan, W.J. (1979) Misclassification and the design of environmental studies. *Am. J. Epidemiol.*, 109, 607-616
- Goldstein, A.M., Falk, R.T., Korczak, J.F. & Lubin, J.H. (1997) Detecting gene-environment interactions using a case-control design. *Am. J. Hum. Genet.*, 14, 1085-1089
- Greenland, S. (1980) The effect of misclassification in the presence of covariates. *Am. J. Epidemiol.*, 112, 564-569
- Greenland, S. (1983) Tests for interaction in epidemiologic studies: a review and a study of power. *Stat. Med.*, 2, 243-251
- Kleinbaum, D.G., Kupper, L.L. & Morgenstern, H. (1982) *Epidemiologic Research: Principles and Quantitative Methods*. New York, Van Nostrand Reinhold
- Lubin, J.H. & Gail, M.H. (1990) On power and sample size for studying features of the relative odds of disease. *Am. J. Epidemiol.*, 131, 552-566
- Pearce, N. (1989) Analytic implications of epidemiological concepts of interaction. *Int. J. Epidemiol.*, 18, 976-980
- Rothman, N., Stewart, W.F., Caporaso, N.E. & Hayes, R.B. (1993) Misclassification of genetic susceptibility biomarkers: implications for case-control studies and cross-population comparisons. *Cancer Epidemiol. Biomarkers Prev.*, 2, 299-303
- Schlesselman J.J. (1974) Sample size requirements in cohort and case-control studies of disease. *Am. J. Epidemiol.*, 99, 381-384
- Shy, C.M., Kleinbaum, D.G. & Morgenstern, H. (1978) The effect of misclassification of exposure in epidemiologic studies of air pollution health effects. *Bull. NY Acad. Med.*, 54, 1155-1156
- Smith, P.G. & Day, N.E. The design of case-control studies: the influence of confounding and interaction effects. *Int. J. Epidemiol.*, 13, 356-365

Corresponding author  
Nathaniel Rothman  
NIH/NCI EPS 8116,  
Bethesda, MD 20892,  
USA