Monte Carlo Simulation of Uncertainties in Epidemiological Studies: An Example of False-Positive Findings due to Misclassification

Alexander Shlyakhter and Richard Wilson

Department of Physics and Harvard Center for Risk Analysis
Harvard University, Cambridge, Ma 02138 USA

Abstract

The 95% confidence intervals for the Risk Ratios (RR) reported in epidemiological studies reflect only sampling errors and do not include uncertainty caused by misclassification and confounding. Analysis of uncertainties in epidemiological studies can be improved using Monte Carlo simulations. For case-control studies, we show how differential misclassification of exposure status increases the probability of getting a statistically significant false positive result. The misclassification error is relatively more important when several studies are pooled together. Simulations enable the uncertainties in epidemiologic results to be reported similarly to natural science where systematic and statistical uncertainties are carefully combined. We illustrate this by showing how false positives can result from misclassification.

1 Introduction

The 95% confidence intervals (95%CI) for Risk Ratios commonly reported in epidemiological literature account only for one component of uncertainty, namely the random sampling errors caused by the finite number of subjects. They describe the statistical precision but tell the reader nothing about the validity of the results. Investigators always try to minimize the uncertainties caused by various possible biases (coming from such sources as selection, differential misclassification, and confounding), but they do not attempt to quantify the effect of the residual biases on uncertainty in the results. Such biases can be viewed as analogues of systematic errors in physical measurements.

2 False-positive results in case-control studies

The quality of data available to researchers in observational studies is generally lower than the quality of data used in experimental science [1]. Unaccounted systematic errors probably occur more often in observational studies than in physical measurements [2-4]. This is particularly so in multiple studies with slightly elevated risks where the signal-to-noise ratio is small (e.g. environmental tobacco smoke). The relative magnitude of random and systematic errors can differ from study to study; one can expect that systematic errors are more important for large studies when 95% CI are narrow. To illustrate how residual biases can cause false-positive results in observational studies, Shlyakhter [4] has performed Monte Carlo simulations of case-control studies. Simulations were performed of 1,000 case-control studies "conducted" on the population exposed to a *nonharmful* agent so that the "true" odds ratio OR=1 (odds ratio is the measure of relative risk in case-control studies). The question asked was how often do the 95% confidence intervals cover this true value under different assumptions about the fraction of subjects for whom exposure status has been misclassified? A population sample was considered with 10,000 exposed and 10,000 nonexposed subjects and it was assumed that both groups have the same small probability, $p=0.01$, of contracting the disease. For each "case-control study," computer simulated the numbers in four cells of the 2x2 table: exposed cases, a, exposed controls, b, nonexposed cases, c, and nonexposed controls, d.

0-8186-7126-2/95 $4.00 © 1995 IEEE
Proceedings of ISUMA-NAFIPS '95

The effects of misclassification were simulated by moving a random fraction of subjects across the cells of the 2x2 table. More specifically, it was assumed that the *observed* numbers in each cell are a_1, b_1, c_1, d_1. Here $a_1=a+r_1c$, $c_1=c(1-r_1)$, $b_1=b+r_2d$, $d_1=d(1-r_2)$ and r_1, r_2 are random numbers representing the fractions of misclassified subjects among cases and controls.

It was assumed that r_1 follows normal distribution with zero mean and standard deviation ERR truncated at zero so that only positive values of r_1 were allowed. Therefore all truly exposed cases were classified correctly but some nonexposed cases were classified as exposed. For r_2 a non-truncated normal distribution with zero mean and same standard deviation *ERR* was assumed. This means that exposed and nonexposed controls are equally likely to give wrong answers about their exposure history. Different distributions for cases and controls account for the tendency of cases to better recall (and sometimes to exaggerate) their exposure history as compared with controls. Note that each simulated study has its own fraction of misclassified subjects; parameter *ERR* determines only the average rate of misclassification.

The simulated odds ratio was calculated as $OR=a_1d_1/b_1c_1$; the upper and lower bounds of 95% CI were calculated as: $OR\exp(\pm 1.96(1/a_1+1/b_1+1/c_1+1/d_1)^{1/2})$. It is easily seen that the comparable x value for each study becomes $ln(RR)/SE(ln(RR))$. Accordingly x was calculated as $x=ln(OR)/SE(ln(OR))=ln(OR)/(1/a_1+1/b_1+1/c_1+1/d_1)^{1/2}$ and the cumulative distribution of x values was plotted. The results of 1,000 trials are presented in Figure 1. For ERR=0.01 the Monte Carlo calculation follows a Gaussian distribution - as it should. Even for a relatively small fraction of misclassification (ERR=0.1) the tails extend far beyond the Gaussian distribution and the 95% CI corresponds to Z=3.8 standard deviations instead of Z=1.96 for the Gaussian curve. Moreover, Gaussian distribution gives only $1.45\cdot10^{-4}$ for probability of errors larger than 3.8 standard deviations.

3 Study Size

The effects of differential misclassification are more important for large studies where random errors are relatively small. This is illustrated in Figure 2 where the probability of false positive findings is shown for individual studies with different numbers of cases and controls. Fixed 10% rate of differential misclassification was assumed (ERR=0.1). For n=10, random errors dominate and the distribution is close to Gaussian. However, for n>100, random errors fall below 10% and systematic errors become most important.

4 Pooled studies

Monte Carlo simulations can also help in understanding of uncertainties when several studies are pooled together. As before, we simulate studies conducted on the population exposed to a *nonharmful* agent so that the "true" odds ratio $OR=1$. We ask the following question: for a given fraction of subjects with misclassified exposure status, how does pooling several studies together affect the probability that the 95% confidence intervals cover this true value?

We consider a population sample with 10,000 exposed and 10,000 nonexposed subjects and assume that both groups have the same small probability, $p=0.01$, of contracting the disease. For the *i-th* "case-control study," we simulate the "true" numbers of exposed cases, a_i, exposed controls, b_i, nonexposed cases, c_i, and nonexposed controls, d_i and the "apparent" numbers, a_{1i}, b_{1i}, c_{1i}, d_{1i}; the difference between "true" and "apparent" numbers accounts for exposure misclassification. For each study, we then calculate the simulated odds ratio, $OR_i=a_{1i}d_{1i}/b_{1i}c_{1i}$, the upper and lower bounds of 95% CI: $OR_i\exp(\pm 1.96(1/a_{1i}+1/b_{1i}+1/c_{1i}+1/d_{1i})^{1/2})$.
Summary estimate of *OR* from n studies pooled together is calculated as follows. First, we assign to each study a weight, $w_i=1/var(ln(OR_i))$; w_i is inverse of the squared

width of the 95% CI on the log scale. These weights are used in calculation of the summary odds ratio *OR* and 95% *CI* [5]: $ln(OR) = (\Sigma w_i ln(OR_i))/\Sigma w_i$, $1/var(ln(OR)) = \Sigma(1/var(ln(OR_i))$.

We calculate the normalized deviation from the null value, $x = ln(OR)/SE(ln(OR))$, and plot the cumulative distribution of x values. A set of pooled studies will produce false-positive results if the apparent value of pooled estimate *ln(OR)* is more than its two standard errors away from the null value. Results of 1,000 trials for individual studies (n=1) and combinations of n=5, n=10 and n=30 studies assuming 5% misclassification rate (ERR=0.05) are presented in Figure 3. Probability of a statistically significant false positive finding, $x>2$, increases from 12% for n=1 to 28% for n=5, to 40% for n=10, and to 70% for n=30. For comparison, a second set of simulations was conducted assuming 2% misclassification rate (ERR=0.02). Results are presented in Figure 4. The effect of pooling several studies together is less dramatic than in Figure 3 but still large. Probability of a statistically significant false positive finding, increases from 4% for n=1 to 6% for n=5, to 10% for n=10, and to 24% for n=30.

5 Summary

Interpretation of the results of observational studies and their use in regulatory risk assessment becomes progressively more difficult as epidemiologists deal with smaller risk ratios. Currently, the reported 95% confidence intervals reflect only sampling errors and do not formally include uncertainty caused by misclassification and confounding. This makes it hard to describe an overall uncertainty in the epidemiological result. We apply techniques of Monte Carlo simulation to the analysis of the effects of systematic uncertainties in case-control studies with slightly elevated risk ratios. We show that even a small fraction of subjects with misclassified exposure status (differential among cases and controls) can cause a considerable fraction of statistically significant, but false positive, results. The effect of differential misclassification is more important for large studies where random errors are relatively small and when several studies are pooled: upon pooling, the statistical uncertainty is reduced but the misclassification uncertainty stays approximately constant.

Acknowledgements

This research was funded by a number of gifts to Harvard University. E-mail addresses:
shlyakhter@physics.harvard.edu
wilson@physics.harvard.edu

References

[1] Feinstein A.R. "Scientific standards in epidemiologic studies of the menace of daily life," *Science* 242:1257-1263 (1988)

[2] Shlyakhter A.I. "Uncertainty Estimates in Scientific Models: Lessons from Trends in Physical Measurements, Population and Energy Projections," in *Uncertainty Modelling and Analysis: Theory and Applications*, editors: B.M. Ayyub and M.M.Gupta, p.477-496. Elsevier Science B.V., 1994

[3] Shlyakhter A.I. An improved framework for uncertainty analysis: accounting for unsuspected errors. *Risk Analysis* 14:441-447, 1994.

[4] Shlyakhter A. (1995) "Analysis of Multiple Epidemiologic Studies Using Frequency Distributions of Relative Risks," Submitted to the *Journal of Clinical Epidemiology*

[5] Greenland S. (1987) "Quantitative methods in the review of epidemiologic literature," *Epidemiologic Reviews* 9:1-30.

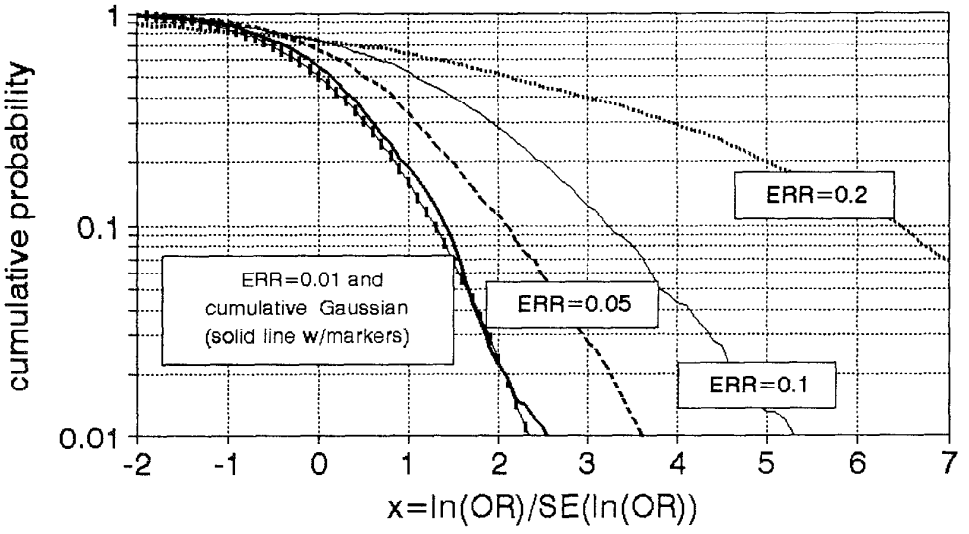

Figure 1: Results of Monte Carlo simulations illustrating the effect of exposure misclassification on the frequency of false-positive results in case-control studies. Cumulative probability that apparent normalized logarithm of the odds ratio, *(OR), x=ln(OR)/SE(ln(OR))* exceeds given value is shown for several values of the parameter *ERR*. This parameter represents the average fraction of subjects with misclassified exposure status (see Section 2 for details). The true value of *OR=1* (the risk is not elevated). Studies that produce values *x>2* are false-positive because the lower bound of the 95% confidence interval (95%CI) for *OR* is above the true value *OR=1*.

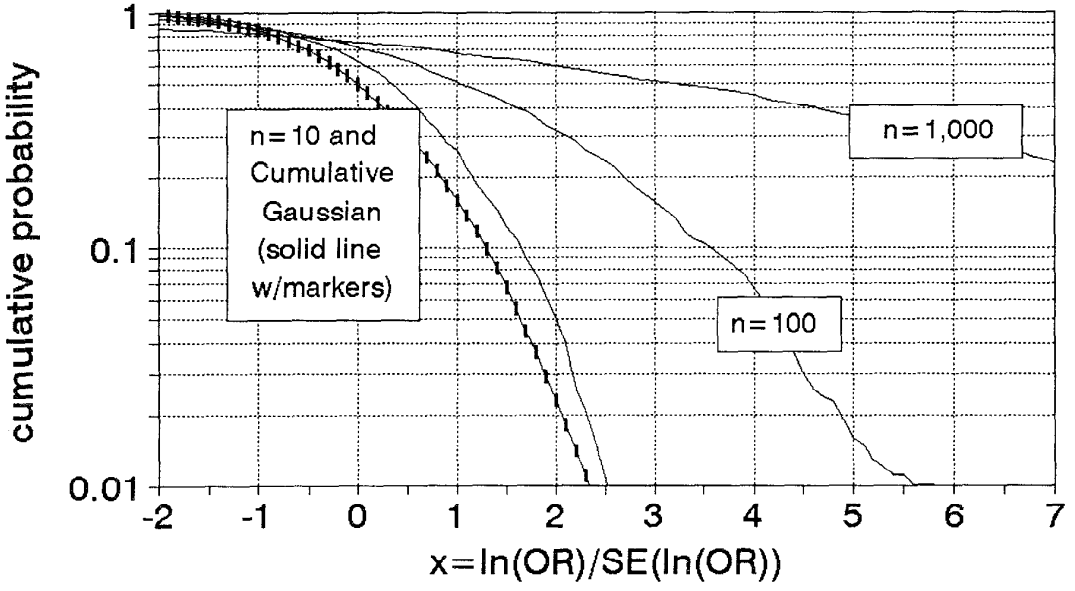

Figure 2: Probability of false positive findings for individual studies with n cases and n controls. Fixed 10% rate of differential misclassification was assumed (ERR=0.1). Systematic errors are more important for large studies where errors caused by random sampling are relatively small.

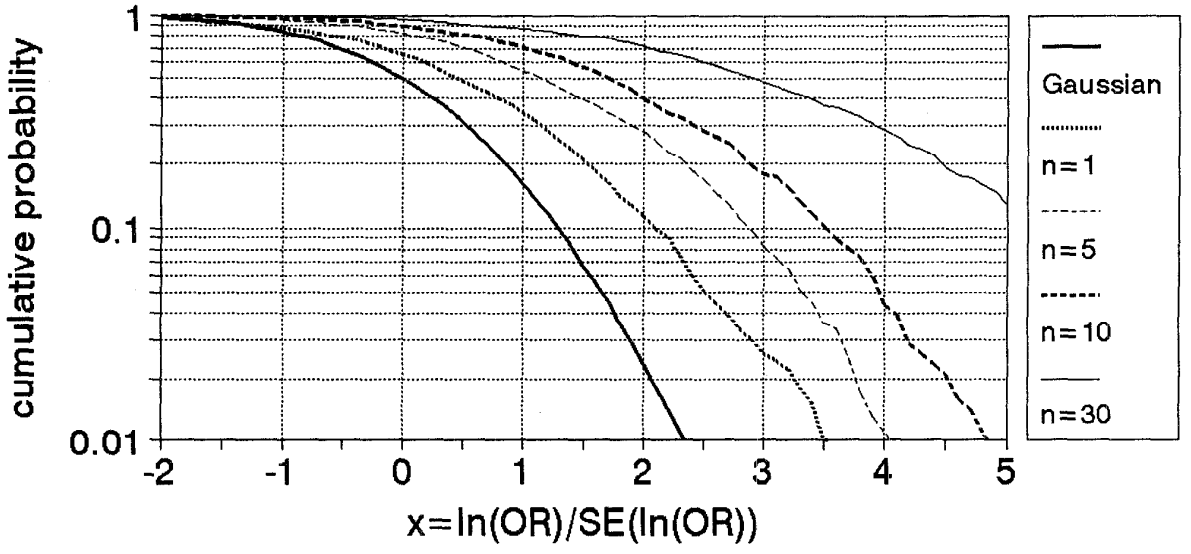

Figure 3: Probability of false positive findings in case control studies for individual studies (n = 1) and combinations of n = 5, n = 10, and n = 30 studies assuming that exposure status was differentially misclassified for 5% of subjects (ERR = 0.05).

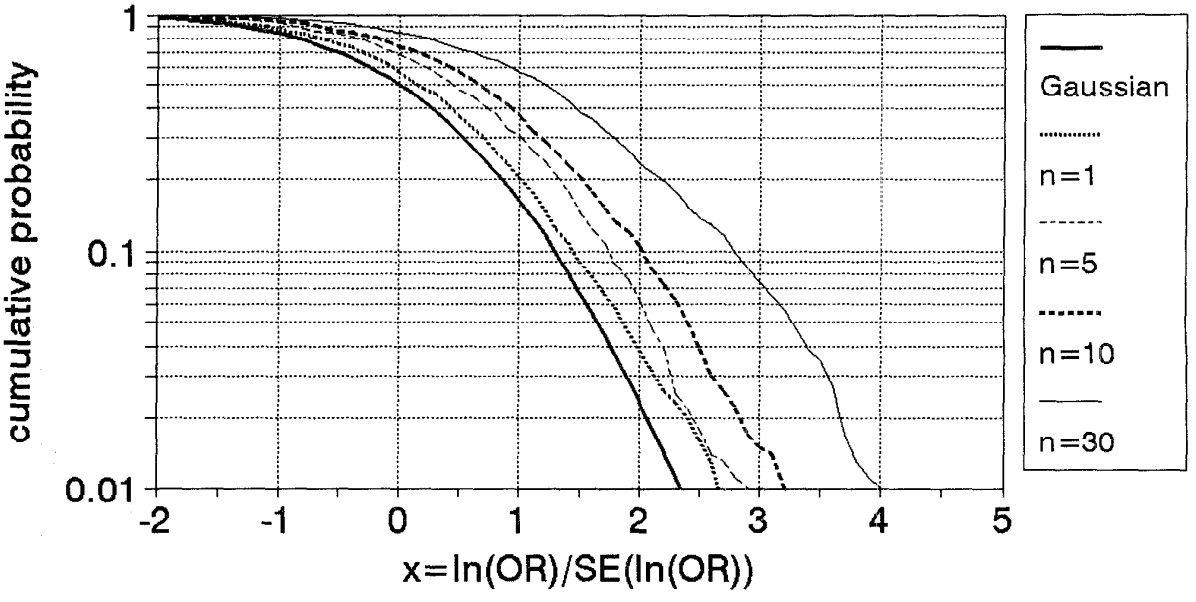

Figure 4: Probability of false positive findings in case control studies for individual studies (n = 1) and combinations of n = 5, n = 10, and n = 30 studies assuming that exposure status was differentially misclassified for 2% of subjects (ERR = 0.02).