



**From the SelectedWorks of Dunlei Cheng**

---

January 2010

# Accounting for Response Misclassification and Covariate Measurement Error Improves Powers and Reduces Bias in Epidemiologic Studies

Contact  
Author

Start Your Own  
SelectedWorks

Notify Me  
of New Work

---

Available at: [http://works.bepress.com/dunlei\\_cheng/3](http://works.bepress.com/dunlei_cheng/3)

# **Accounting for Response Misclassification and Covariate Measurement Error Improves Power and Reduces Bias in Epidemiologic Studies**

Running Title: power calculations with measurement error

DUNLEI CHENG<sup>1</sup>, ADAM J. BRANSCUM<sup>2</sup>, and JAMES D. STAMEY<sup>3</sup>

<sup>1</sup>*Institute for Health Care Research and Improvement, Baylor Health Care System,  
Dallas, TX 75206, USA*

<sup>2</sup>*Departments of Biostatistics, Statistics, and Epidemiology, University of Kentucky,  
Lexington, KY 40536, USA*

<sup>3</sup>*Department of Statistical Science, Baylor University, Waco TX 76798, USA*

*Word Count: 2,949*

*Abstract Word count: 161*

*Number of Tables: 2*

*Number of Figures: 3*

## **Accounting for response misclassification and covariate measurement error improves power and reduces bias in epidemiologic studies**

**Purpose:** To quantify the impact of ignoring misclassification of a response variable and measurement error in a covariate on statistical power, and to develop software for sample size and power analysis that accounts for these flaws in epidemiologic data.

**Methods:** A Monte Carlo simulation-based procedure is developed to illustrate the differences in design requirements and inferences between analytic methods that properly account for misclassification and measurement error to those that do not in regression models for cross-sectional and cohort data.

**Results:** We found that failure to account for these flaws in epidemiologic data can lead to a substantial reduction in statistical power, over 25% in some cases. The proposed method substantially reduced bias by up to a ten-fold margin compared to naive estimates obtained by ignoring misclassification and mismeasurement.

**Conclusion:** We recommend as routine practice that researchers account for errors in measurement of both response and covariate data when determining sample size, performing power calculations, or analyzing data from epidemiological studies.

Keywords: Power calculations; Measurement error; Simulation

List of Abbreviations: OR: odds ratio  
Se: sensitivity  
Sp: specificity

---

### **INTRODUCTION**

Measurement error and misclassification commonly occur in epidemiologic data.

Measurement error refers to inaccuracy in quantifying a continuous variable. For example, the biomarker CD4 count that has been used to predict time to occurrence of AIDS is known to exhibit measurement error (1, 2). Misclassification refers to incorrectly assigning group-membership for a categorical variable. For instance, in prevalence and test evaluation studies, true infection status may be unavailable because it is too expensive or invasive to ascertain (3, 4). Failure to account for misclassification and/or measurement error is well known to potentially bias estimates, generally towards

the null hypothesis. Both are forms of measurement error, so in this paper we use this term to include both types of error.

Research on the effects of exposure measurement error (e.g., nutrition, radiation, blood pressure) (5–8) or the effects of response misclassification (e.g., breast cancer, stroke, myocardial infarction) (9, 10) have separately appeared in the literature. Recent interest in related epidemiologic methods has focused on appropriately adjusting logistic regression inferences in the presence of measurement error in covariates and misclassification of a binary response (11, 12). In this research, the authors assessed the effect of radiation exposure on cancer mortality among atomic bomb survivors in Japan. Cause of death (cancer related or not) had the potential to be misclassified because autopsies were carried out on only a subset of the atomic bomb surviving cohort. In addition, an individual's radiation exposure was based on recall of their location and shielding information on the bombing day, which is subject to measurement error.

In this paper, we illustrate that by ignoring measurement error with misclassified logistic regression data there is the potential to miscalculate statistical power and to enroll fewer subjects than are actually needed for the future study. Therefore, when planning an investigation into the association between a mismeasured exposure and a misclassified health outcome, sample size and power analysis should address both error types in the design phase of the study. To do this we adopt a Bayesian approach. As a result, the methods and conclusions reported in this study regarding power and sample size calculations apply to scenarios where a Bayesian analysis of the future data will be conducted.

In a Bayesian analysis, prior information is combined with current data to produce updated posterior inferences about model parameters. Prior information is quantified through a probability distribution, which describes how likely or unlikely it would be to find the “true” parameter values in various sets of the parameter space. Independent of the current data, prior distributions are generally constructed based on historical data and/or expert knowledge. Absent such information, reference priors that are essentially uniform over the parameter space can be used. Bayesian inference proceeds by using Bayes theorem to determine a posterior distribution, which is a weighted combination of information from the prior and the data. The growing popularity of Bayesian methods in epidemiology can be attributed to the flexibility in modeling complicated scenarios without having to rely on asymptotic inference (13–15), among other reasons. For instance, the Bayesian approach to power calculations used here accounts for uncertainty in the value of an effect size (e.g., odds ratios) through simulation.

Sample size requirements for generalized regression models with covariate measurement error were investigated by (16, 17), and a Bayesian approach for sample size calculations with misclassified logistic regression response data and correctly measured covariates was considered in (18) . The present study extends those methods to estimate power for logistic regression data subject to both response and covariate measurement error.

---

## **METHODS**

### **Statistical Model**

We begin by discussing the model. For ease of exposition we assume that only one covariate is measured with error, but the techniques discussed can be easily extended to handle multiple mismeasured covariates. The logistic regression analysis might also involve a vector  $Z$  of correctly measured covariates, which can include confounders and effect modifiers. Let  $\tilde{Y}$  denote the observed response variable and let  $Y$  denote the true, but potentially misclassified, response. Suppose that a continuous variable  $X$  cannot be observed exactly and instead the surrogate predictor  $\tilde{X}$  is used in its place. The observed response data are obtained from use of an imperfect classification instrument. For concreteness, we assume the response variable represents disease status or cause of death, but it could be other binary outcomes. We also assume that a diseased subject is correctly classified with probability  $Se$  (sensitivity) and a non diseased subject is correctly classified with probability  $Sp$  (specificity). When  $Se = Sp = 1$ , no misclassification of the response occurs, but measurement error of  $X$  remains. Hence, the methods developed here apply to scenarios where only measurement error or only misclassification is present, and where both are present.

The relationship between the true binary response variable,  $Y$ , and the true risk factors  $X$  and  $Z$  is described according to the logistic regression model:

$$\text{logit}[\Pr(Y_j = 1)] = \beta_0 + \beta_1 X_j + Z_j' \theta_z, \quad (1)$$

where  $\beta_1$  characterizes the (true) effect of the (unobserved) covariate  $X_j$  on disease occurrence, and  $\theta_z$  is an unknown coefficient vector associated with the predictors in  $Z_j'$ .

Assuming homoscedastic  $Se$  and  $Sp$  across subpopulations defined by  $X$  and  $Z$ , the probability of an affirmative observed response for subject  $j$  is given by

$$\Pr(\tilde{Y}_j = 1) = \Pr(Y_j = 1) \times \text{Se} + [1 - \Pr(Y_j = 1)] \times (1 - \text{Sp}). \quad (2)$$

Equations (1) and (2) are directly connected through the term  $\Pr(Y_j = 1)$ . From (1), this term depends on the unobserved  $X_j$ . We need to impute  $X_j$  using information from the surrogate measurement  $\tilde{X}_j$ . Thus, we require a prediction model that relates these two values.

We first consider the Berkson measurement error model, which assumes that the true covariate value is known up to a random deviation. Specifically, the Berkson measurement error model stipulates that

$$X_j = \tilde{X}_j + \varepsilon_j, \quad (3)$$

where  $\varepsilon_j$  is a normal random variable with mean 0 and variance  $\sigma_\varepsilon^2$ . In formula (3), the measurement error  $\varepsilon_j$  is independent of the surrogate  $\tilde{X}_j$ , and the  $\varepsilon_j$ 's are independent of each other. The Berkson model has been found to be a reasonable choice for measurement errors in many settings (7, 19).

We also consider a classical measurement error model that uses the covariates  $Z_j$  to predict  $X_j$  through a hierarchical regression structure:

$$\tilde{X}_j = X_j + \eta_j, \quad (4)$$

where  $\eta_j \sim \text{Normal}(0, \sigma_\eta^2)$  and the unknown  $X_j$  is calibrated through  $Z_j$  via a regression, for instance the normal linear regression with

$$E[X_j] = \alpha_0 + Z_j' \alpha_z. \quad (5)$$

This type of error model has been used for many dietary studies where each error-prone covariate is unique to an individual and can be replicated (5, 7), and a modified version has been used as a model for time course trends of CD4 count (1).

### Power Calculations

We assume principal interest is in whether there is a relationship between  $Y$  and  $X$  in equation (1). Thus, we evaluate  $\Pr(\beta_1 > 0 \mid \text{data})$  if the risk of event increases with an increase in  $X$ . When both the response and primary predictor are subject to error, the power criterion selects a sample size  $N$  for which this parameter, averaged over the conditions that exist in the population, exceeds a user-defined threshold,  $\gamma$ , i.e.,

$$E[\Pr(\beta_1 > 0 \mid \text{data})] \geq \gamma. \quad (6)$$

Typical choices of  $\gamma$  are 0.8, 0.9, 0.95, or others depending on the nature of the problem. From (6) we see that an average power is computed; in particular the average is taken over multiple simulated data sets and across distributions of possible “true” values for  $\beta_1$  and any other regression coefficients in the model.

If disease occurrence and the risk factor are negatively related, expression (6) would be changed to  $E[\Pr(\beta_1 < 0 \mid \text{data})] \geq \gamma$ . More generally, if detecting an effect size of at least  $b^*$  (say 0.2) is the goal of a particular study, then a sample size  $N$  is determined such that

$$E[\Pr(\beta_1 > b^* \mid \text{data})] \geq \gamma. \quad (7)$$

For example, to mimic the calculations reported in (20) and (21) for predicting coronary heart disease risk from serum cholesterol level, we would use  $b^* = 0.5$  and  $b^* = 0.1$  in deciding the optimal sample size required. In practice, selecting  $b^*$  may be difficult and

instead it might be easier to think about the value of the odds ratio,  $e^{b^*}$ , and then make the necessary log transformation to determine  $b^*$ .

The computational approach that we employ involves the power criterion proposed by (22), which is a Monte Carlo procedure originally used in the setting of ordinary linear and logistic regression. For the latest discussion and application of this two-prior scheme, see (23–25). Briefly, distributions for the parameters used to simulate data sets are elicited. These are termed sampling priors. The purpose of sampling priors is to account for uncertainty in the “true” values of population parameters, which differs from use of one set of fixed values as is commonly done in frequentist sample size determination. For instance, instead of fixing  $\beta_1$  to be 0.2 we might generate it from a uniform distribution over the range between 0.1 and 0.3, allowing for uncertainty in our presumed “true” value. In addition to these, prior distributions for the actual Bayesian data analysis are also required. These distributions are referred to as fitting priors.

The Monte Carlo power analysis procedure used here is composed of two major steps: data simulation and data analysis. The steps are summarized by a flow chart in Figure 1. We require three sets of distributions. To generate data sets, we need (i) the population distribution of  $X$  and  $Z$  and (ii) the sampling priors for (a)  $\beta_1$ , (b) the other regression coefficients, (c) the variance parameters in the measurement error models, and (d) the Se and Sp of misclassification. For data analysis, we also need (iii) fitting prior distributions for all regression coefficients and any other parameters (e.g., Se, Sp, and variance components) in the model. The fitting priors on the regression coefficients are generally taken to be diffuse so as to not unduly impact inferences. However, the sampling priors of the regression coefficients and other parameters used to simulate data

sets will contain substantive information. For instance, for data simulation the distribution of each regression coefficient in equation (1) can be taken as a narrow uniform or normal distribution centered on the most probable value.

Figure 1 about here.

---

## **RESULTS**

### **Atomic Bomb Example**

We consider an example motivated by data reported in (11) from a study that evaluated the impact of radiation exposure on cancer mortality in a cohort of 86,520 atomic bomb survivors. As stated previously, the cause of death, i.e., cancer related or not, recorded from death certificates might be misdiagnosed for a subset of individuals. The predictor, radiation exposure, was mismeasured and we assumed it followed a Berkson error process, consistent with the analyses in (11, 12).

We conducted a power analysis to detect a positive relationship between death due to cancer and the amount of radiation absorption. Radiation exposure is measured in Gray (Gy). Most of the survivors had an observed radiation level below 0.1 Gy and there were few survivors whose observed radiation exposure was above 6.0 Gy. Therefore, we used a Weibull distribution to simulate (observed) radiation exposure in the population that had the appropriate skewness of radiation exposure based on recommendations in (26). Under a Berkson measurement error process, true radiation exposure varies according to a lognormal with median value set equal to an individual's reported amount

plus random variation, where the variance parameter  $\sigma_\varepsilon^2$  was modeled as Uniform(0.4, 0.6).

According to (27), the estimates of sensitivity and specificity for recording cause of death were 0.78 and 0.965, respectively, so we used Beta(39, 11) and Beta(193, 7) distributions for data simulation and analysis. Based on results in (11), Normal(-0.69, 0.1<sup>2</sup>) and Uniform(0.3, 0.4) distributions were selected for  $\beta_0$  and  $\beta_1$  for data simulation. For model fitting, independent normal distributions with mean 0 and variance 10 were used for regression coefficients.

The parameters and data were generated using the software package R. Posterior distributions were approximated using WinBUGS. Complete programs to perform the power analysis and sample size procedures in this paper are available upon request from the corresponding author. Figure 2 provides the estimated power at sample sizes of 100 to 600 in steps of 100, each based on averaging 1000 data sets, one from each of 1000 plausible populations. After interpolation, we conclude that to detect  $\beta_1 > 0$  in this setting, at least 218 atomic bomb survivors are required for 80% power, 314 subjects for 85% power, and 469 for 90% power.

Figure 2 about here.

In the following two subsections we report results from two simulation studies. We first consider the effect of ignoring measurement error on power and sample size calculations for a future study. We then illustrate the increase in bias that can occur when failing to account for measurement error in data analysis.

### Simulation Related to Power

We studied the impact of not adjusting for misclassification and/or measurement error on statistical power using Monte Carlo simulation in a Bayesian framework. Consider misclassified binary response data with a single continuous covariate subject to classical measurement error. Since there are no other predictors in this scenario, the expected value of the true covariate mean in equation (5) is simply  $\alpha_0$ . Distributional information for  $\alpha_0$ ,  $\sigma_v^2$ , and  $\sigma_\eta^2$  can be found in Table 1. We assume the classification instrument has relatively high specificity (mean of 0.92) and moderate sensitivity (mean of 0.8). For data generation, a Uniform(0.2, 0.7) distribution is used to simulate  $\beta_1$ .

Table 1 about here.

We are particularly interested in the posterior probability of detecting an effect size of OR = 1 or OR = 1.2, which implies that  $\beta_1$  is at least 0 or 0.2. Although only modest gains in power are seen in the former case when adjusting for measurement error, increasing the effect size slightly to OR=1.2 yields large gains. An effect size of OR = 1.2 leads to substantially lower power at each fixed sample size (from 30 to 150 subjects) when errors in the covariate and/or response are ignored, as shown in Table 2. Compared with a model that includes structure that accounts for response classification error and predictor measurement error, the power on average decreases by 25.5%, 17.1%, and 3.6%, respectively, when both misclassification and measurement error are ignored, only misclassification is ignored, or only measurement error is ignored. Also, to obtain a power of 0.8 with OR = 1, 17 more subjects will be sampled if one ignores both

measurement error and misclassification, compared with a total sample size of only 38 if one were to account for these errors.

Separate comparisons of columns 2 and 3 to column 4 in Table 2 reveal that ignoring response misclassification is more costly than ignoring measurement error in this setting. For instance, at  $N = 60$ , the power increases from 0.61 (ignore misclassification) to 0.73 (ignore measurement error) to 0.75 (ignore neither) for  $\beta_1 > 0.2$ .

Table 2 about here.

### **Simulation Related to Bias**

We also investigated the reduction in bias achieved by accounting for misclassification and measurement error simultaneously. For data generation, we fixed  $\beta_1$  at 0.4 and the other parameters were set equal to the mean values used in the previous subsection. We focused on estimation of the odds ratio, where the true OR is  $e^{\beta_1=0.4} \approx 1.5$ . The histogram in the top panel of Figure 3 shows that by modeling classical measurement error and misclassification, the estimated ORs from 1000 simulated data sets are clustered around the true value of 1.5, with average bias as small as 0.023. On the other hand, the bottom panel of Figure 3 shows how poorly the OR is estimated when no account is taken for either misclassification or measurement error; the naïve statistical analysis results in a severe downward bias of 0.227 for the OR.

Figure 3 about here.

We also estimated the bias when only one type of error is accounted for in the analysis. Adjustment for classical error gave bias -0.160 while adjustment for misclassified response data gave a much smaller bias of -0.036. Figure 3 also indicates that the variance of the distribution of estimated ORs may be underestimated from a naïve analysis, which is in part a consequence of not accounting for uncertainty about the sensitivity, specificity, and classical measurement error.

---

## **Discussion**

The current study has used simulation to investigate the influence of two types of errors in data, misclassification and measurement error, on power estimation in epidemiological studies. Response misclassification was modeled by the parameters Se and Sp, and the amount of measurement error in the covariate was characterized by  $\sigma_{\eta}^2$  or  $\sigma_{\varepsilon}^2$ , depending on classical or Berkson error structure, respectively. The criterion adopted was average power, a Monte Carlo simulation procedure that takes account of the conditions that exist in a population and uncertainty about effect size through distributions on data and model parameters. We illustrated how the use of historical information can be used to construct required distributions in the atomic bomb example that investigated the relationship between cancer related deaths and exposure to radiation.

A referee pointed out a particularly scenario in which adjustment for misclassification error does not alter the outcome of a likelihood ratio test for exposure assessment. However, the scenario covered in (28, 29) differs from ours in that it pertains to case-control studies with a perfectly measured binary response variable (case or control) and a binary predictor variable that is subject to misclassification. We

considered cross-sectional and cohort designs with a continuous surrogate predictor and responses that are subject to misclassification error. Moreover, (28) considered a non Bayesian approach to exposure assessment using likelihood ratio tests whereas our study focused on calculating power to detect a given effect size. In future research we plan to consider case-control studies and issues related to confounding using the Bayesian power analytic method presented here.

---

The authors thank two anonymous reviewers and the Editor for valuable suggestions that improved the quality of the paper.

---

## REFERENCES

1. Wang Y, Taylor JMG. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *J Am Stat Assoc* 2001;96:895-905.
2. Prentice RL. Covariate measurement error and parameter estimation in Cox's regression model. *Biometrika* 1982;69:331-342.
3. Chu H, Wang Z, Cole SR, Greenland S. Sensitivity analysis of misclassification: A graphical and a Bayesian approach. *Ann Epidemiol.* 2006;16:834-841.
4. Johnson WO, Gastwirth JL, Pearson LM. Screening without a "gold standard": the Hui-Walter paradigm revisited. *Am J Epidemiol* 2001;153:921-924.
5. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition.* Boca Raton: Chapman & Hall, 2006.
6. Rosner B, Spiegelman, D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol* 1990;132:734-745.

7. Richardson S, Gilks WR. Conditional independence models for epidemiological studies with covariate measurement error. *Stat Med* 1993;12:1703-1722.
8. Thoresen M. Correction for measurement error in multiple logistic regression: a simulation study. *J Stat Comput Sim* 2006;76:475-487.
9. Yamagishi K, Ikeda A, Iso H, Inoue M, Tsugane S, JPHC Study Group. Self-reported stroke and myocardial infarction had adequate sensitivity in a population-based prospective study JPHC (Japan Public Health Center)-based Prospective Study. *J Clin Epidemiol* 2009; 62: 667-673.
10. Couris CM, Polazzi S, Olive F, Remontet L, Bossard N, Gomez F, Schott AM, Mitton N, Colonna M, Trombert B. Breast cancer incidence using administrative data: correction with sensitivity and specificity. *J Clin Epidemiol* 2009; 62: 660-666.
11. Roy S, Banerjee T, Maiti T. Measurement error model for misclassified binary responses. *Stat Med* 2005;24:269-283.
12. McGlothlin A, Stamey JD, Seaman JW. Binary regression with misclassified response and covariate subject to measurement error: a Bayesian approach. *Biometrical J* 2008;50:123-134.
13. Stamey JD, Bekele BN, Powers S. Bayesian modeling of follow-up studies with missing data. *Ann Epidemiol.* 2009;19:416-422.
14. Bellara CA, Hanley JA, Joseph L., Albertsen PC. Hierarchical changepoint models for biochemical markers illustrated by tracking postradiotherapy prostate-specific antigen series in men with prostate cancer. *Ann Epidemiol.*, 2008; 18:270-282.
15. Dennis LK, Vanbeek MJ, Beane Freeman LE, Smith BJ, Dawson DV, Coughlin JA. Sunburns and risk of cutaneous melanoma: Does age matter? A comprehensive meta-analysis. *Ann Epidemiol.*, 2008; 18:614-627.
16. Devine OJ, Smith JM. Estimating sample size for epidemiologic studies: the impact of ignoring exposure measurement uncertainty. *Stat Med*1998;17:1375-1389.
17. Tosteson TD, Buzas JS, Demidenko E, Karagas, M. Power and sample size calculations for generalized regression models with covariate measurement error. *Stat Med* 2003;22:1069-1082.
18. Cheng D, Stamey JD, Branscum AJ. Bayesian approach to average power calculation for binary regression with misclassified outcomes. *Stat Med* 2009;28:848-863.
19. Wang L. Estimation of nonlinear models with Berkson measurement errors. *Ann Stat* 2004;32:2559-2579.

20. Whittemore AS. Sample size for logistic regression with small response probability. *J Am Stat Assoc* 1981;76:27-32.
21. Shieh G. Sample size calculations for logistic and Poisson regression models. *Biometrika* 2001;88:1193-1199.
22. Wang F, Gelfand AE. A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Stat Sci* 2002;17:193-208.
23. Sahu SK, Smith TMF. A Bayesian method of sample size determination with practical applications. *J Roy Stat Soc A* 2006;169:235-253.
24. Brutti P, De Santis F, Gubbiotti S. Mixture of prior distributions for predictive Bayesian sample size calculations in clinical trial. *Stat Med* 2009;28:2185-2201.
25. Cheng D, Branscum AJ, Stamey JD. A Bayesian approach to sample size determination for studies designed to evaluate continuous medical tests. *Comput Stat Data An* 2010;54:298-307.
26. Pierce DA, Stram DO, Vaeth, M, and Schafer, DW. The errors-in-variables problem: considerations provided by radiation dose-response analyses of the A-bomb survivor data. *J Am Stat Assoc* 1992;87:351-359.
27. Sposoto R, Preston DL, Shimizu Y, Mabuchi K. The effects of diagnostic misclassification on non cancer and cancer mortality dose response in A-bomb survivors. *Biometrics* 1992;48:605-617.
28. Greenland S, Gustafson P. Accounting for independent nondifferential misclassification does not increase certainty that an observed association is in the correct direction. *Am J Epidemiol* 2006;164:63-68.
29. Gustafson P, Greenland S. Curious phenomena in Bayesian adjustment for exposure misclassification. *Stat Med* 2006;25:87-103.

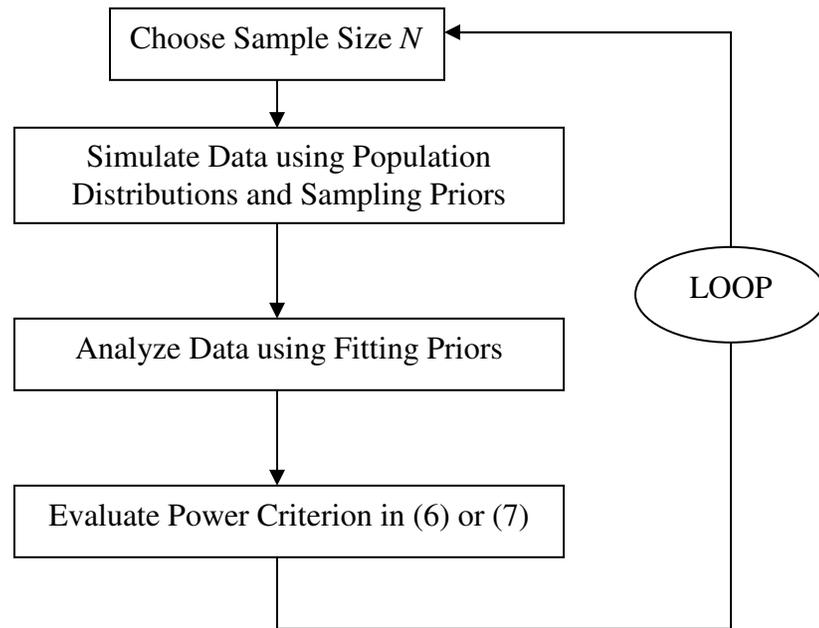


Figure 1: Flow chart of Monte Carlo simulation procedure.

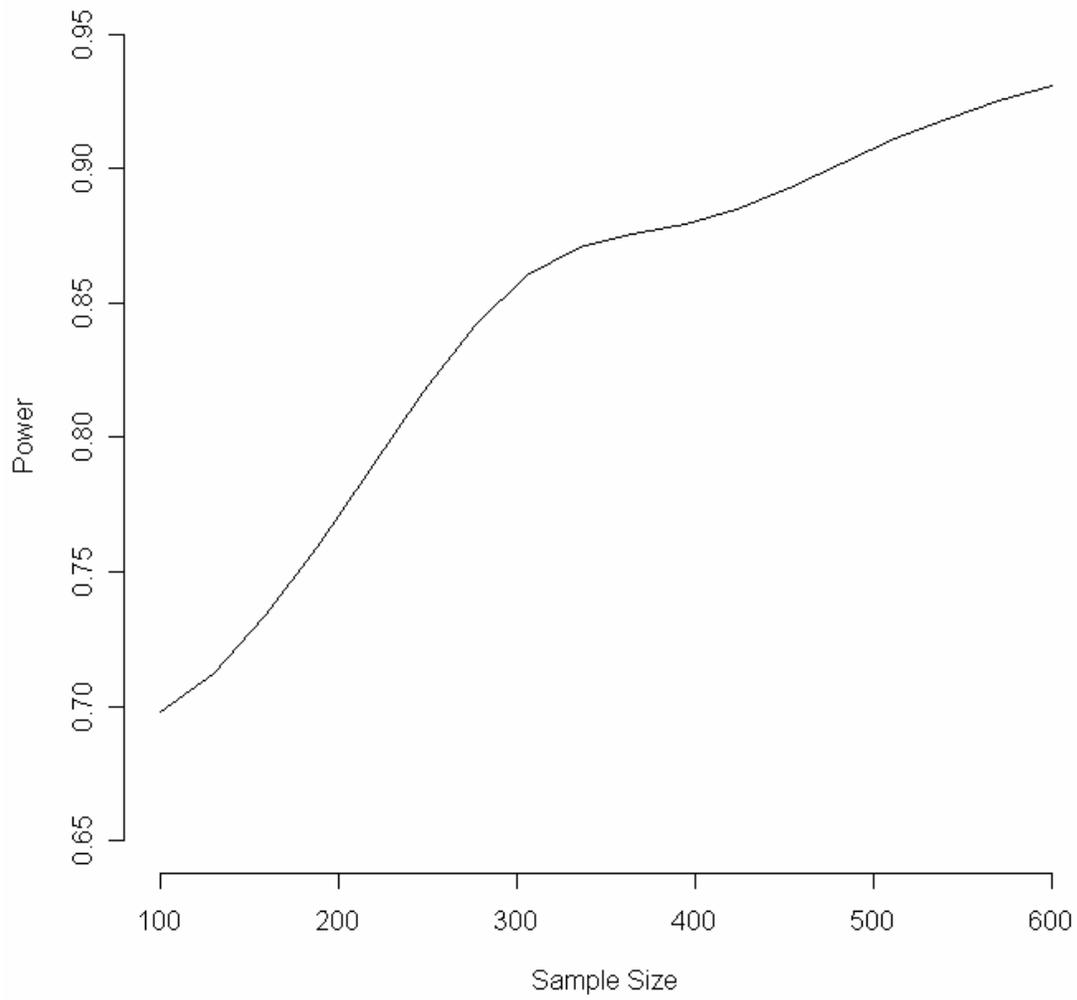


Figure 2: Power curve for atomic bomb example.

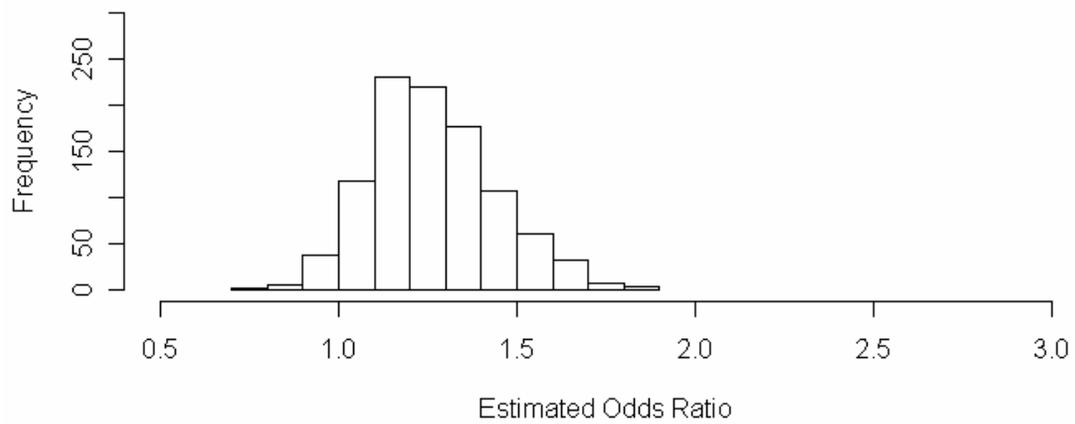
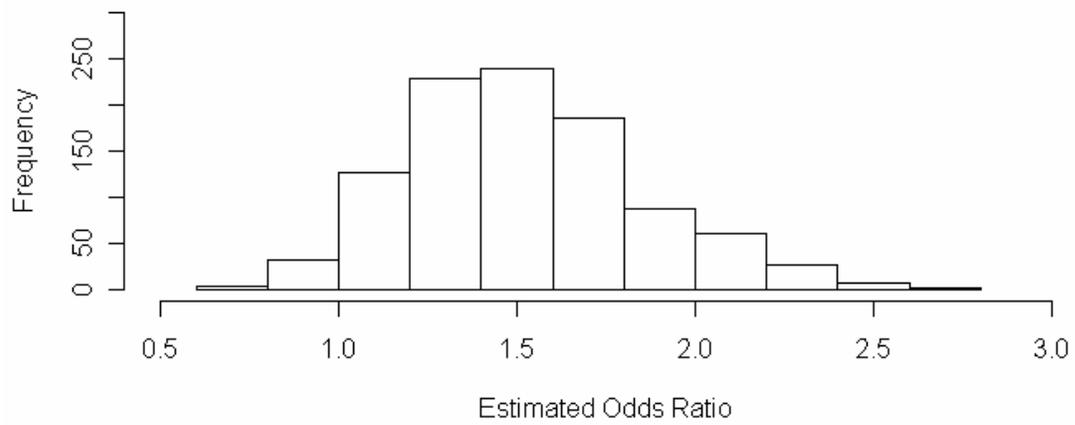


Figure 3: Top panel: Histogram of estimated odds ratio when accounting for misclassification and measurement error (the true odds ratio is 1.5). Bottom panel: Histogram of estimated odds ratios without accounting for misclassification and measurement error.

Table 1: Distributions for simulating and analyzing data for misclassified logistic regression with one mismeasured continuous covariate.

Parameter	Simulate data sets	Analyze data sets
$Se$	Beta(40,10)	Beta(40,10)
$Sp$	Beta(46, 4)	Beta(46, 4)
$\alpha_0$	Uniform(2.8, 3.2)	Normal(0, 10)
$\sigma_v^2$	Uniform(1.6, 2)	Uniform(1.6, 2)
$\sigma_\eta^2$	Uniform(0.3, 0.5)	Uniform(0.3, 0.5)
$\beta_0$	Normal(-1, 0.2 <sup>2</sup> )	Normal(0, 10)
$\beta_1$	Uniform(0.2, 0.7)	Normal(0, 10)

Table 2: Average power under four scenarios for detecting  $OR > 1$  and  $OR > 1.2$ .

Odds Ratio	Sample Size	Power <sup>1</sup>	Power <sup>2</sup>	Power <sup>3</sup>	Power <sup>4</sup>
OR > 1	30	0.744	0.745	0.782	0.783
	60	0.810	0.814	0.837	0.843
	90	0.861	0.861	0.875	0.879
	120	0.894	0.896	0.903	0.906
	150	0.914	0.915	0.919	0.922
OR > 1.2	30	0.561	0.596	0.708	0.716
	60	0.560	0.614	0.725	0.750
	90	0.581	0.643	0.746	0.774
	120	0.585	0.660	0.757	0.792
	150	0.589	0.669	0.762	0.802

1: Ignoring both misclassification and measurement error

2: Ignoring misclassification only

3: Ignoring measurement error only

4: Accounting for both misclassification and measurement error