The Annals of Statistics
1985, Vol. 13, No. 4, 1335–1351

COVARIATE MEASUREMENT ERROR IN LOGISTIC REGRESSION*

BY LEONARD A. STEFANSKI AND RAYMOND J. CARROLL

Cornell University and University of North Carolina, Chapel Hill

In a logistic regression model when covariates are subject to measurement error the naive estimator, obtained by regressing on the observed covariates, is asymptotically biased. We introduce a bias-adjusted estimator and two estimators appropriate for normally distributed measurement errors—a functional maximum likelihood estimator and an estimator which exploits the consequences of sufficiency. The four proposals are studied asymptotically under conditions which are appropriate when the measurement error is small. A small Monte Carlo study illustrates the superiority of the measurement-error estimators in certain situations.

1. Introduction and motivation. Logistic regression is the most used form of binary regression [see Berkson (1951), Cox (1970), Efron (1975), and Pregibon (1981)]. Independent observations (y_i, x_i) are observed where (x_i) are fixed p-vector predictors and (y_i) are Bernoulli variates with

$$\Pr\{y_i = 1|x_i\} = F(x_i^T\beta_0) \triangleq \left(1 + \exp(-x_i^T\beta_0)\right)^{-1}, \qquad i = 1, \dots, n. \tag{1.1}$$

Subject to regularity conditions, the large-sample distribution of the maximum likelihood estimator of β_0 is approximately normal with mean zero and covariance matrix $(1/n)S_n^{-1}(\beta_0)$, where $S_n(\cdot)$ is defined for $\gamma \in R^p$ as

$$S_n(\gamma) = n^{-1}\sum_1^n F^{(1)}(x_i^T\gamma)x_ix_i^T. \tag{1.2}$$

Motivation for our paper comes from the Framingham Heart Study (Gordon and Kannel, 1968), a prospective study of the development of cardiovascular disease. This ongoing investigation has had an important impact on the epidemiology of heart disease. Much of the analysis is based on the logistic regression model with y an indicator of heart disease and x a vector of baseline risk factors such as systolic blood pressure, serum cholesterol, smoking, etc. It is well known that many of these baseline predictors are measured with substantial error, e.g., systolic blood pressure. When a person's "true" blood pressure is defined as a long-term average, then individual readings are subject to temporal as well as reader-machine variability. In one group of 45–54 year old Framingham males it was estimated that one fourth of the observed variability in blood pressure readings was due to within-subject variability. The second author was asked by some Framingham investigators to assess the impact of such substantial measure-

Received July 1983; revised July 1985.
*Supported by the Air Force Office of Scientific Research under grant AFOSR-F49620-82-C-0009.
AMS 1980 *subject classifications.* Primary 62J05; secondary 62H25.
Key words and phrases. Errors-in-variables, functional maximum likelihood, logistic regression, measurement error, sufficiency.

ment error and to suggest alternatives to usual logistic regression which account for this error. The present study is an outgrowth of these questions.

When covariates are measured with error the usual logistic regression estimator of β_0 is asymptotically biased, [see Clark (1982) and Michalik and Tripathi (1980)]. As a consequence of bias there is generally a tendency to underestimate the disease probability for high-risk cases and overestimate for low-risk cases; it will be said that measurement error attenuates predicted probabilities. Also, bias creates a problem with hypothesis testing; in Section 2 it is shown that the usual asymptotic tests for individual regression components can have levels different than expected. An example of this occurs in an unbalanced two-group analysis of covariance where interest lies in testing for treatment effect but the covariable is measured with error.

The severity of these problems depends, of course, on the magnitude of the measurement error. In some situations ordinary logistic regression might perform satisfactorily. However, when measurement error is substantial, alternative procedures are necessary. In addition, the availability of techniques which correct for measurement error can make clear the need for better measurement, e.g., more blood-pressure readings over a period of days.

In Section 2 our measurement-error model is defined and the asymptotic bias in the usual logistic-regression estimator is studied. Section 3 presents some alternative estimators; results of a Monte Carlo study are outlined in Section 4; proofs of the asymptotic results are given in Section 5.

Until recently the study of measurement-error models has focused primarily on linear models; see the review article by Madansky (1959) and the papers by Fuller (1980) and Gleser (1981). Interest in nonlinear models is increasing with recent contributions by Prentice (1982), Wolter and Fuller (1982a, 1982b), Carroll, Spiegelman, Lan, Bailey, and Abbott (1984), Armstrong (1984), Amemiya (1982), and Clark (1982). Of these articles Clark (1982) and Carroll et al. (1984) focus specifically on logistic regression. The asymptotic methods employed in this paper are similar to those used by Wolter and Fuller (1982a) and Amemiya (1982) in their studies of nonlinear functional relationships.

2. A measurement error model for logistic regression.

2.1. The model. Our measurement-error model starts with (1.1), but rather than observing the p-vector x_i we observe

$$X_i = x_i + \sigma v_i \quad \textit{where } v_i = \Sigma^{1/2}\varepsilon_i. \tag{2.1}$$

In (2.1) $\Sigma^{1/2}$ is the square root of a symmetric positive semidefinite matrix Σ scaled so that $\|\Sigma\| = 1$ and (ε_i) are independent and identically distributed random vectors with zero mean and identity covariance; also ε_i is independent of y_i, $i = 1, \ldots, n$. The scale factor σ dictates the magnitude of the measurement error, e.g., if X_i is a mean of m independent replicate measurements of x_i then $\sigma \propto m^{-1/2}$. The asymptotic theory presented in this paper requires that $\sigma \to 0$ as $n \to \infty$, i.e., large sample, small measurement-error asymptotics. The asymptotics are relevant for two situations: (i) when X_i is an average of m-independent measurements of x_i, in which case the Central Limit Theorem suggests that (ε_i) should be viewed as normal random variates, and (ii) when measurement error is small but nonnegligible. In the latter case the moments of order greater than two of (ε_i) generally differ from those of a normal variate.

Our methods of correcting for bias require knowledge of the error covariance matrix $V \triangleq \sigma^2\Sigma$. Since this information is seldom available all asymptotic results are derived for the case in which V is replaced by an estimator $\hat{V}$ satisfying

$$n^{1/2}(\hat{V} - V) = O_p(\sigma^2). \tag{2.2}$$

Condition (2.2) is satisfied, for example, when V is estimated by replication. It is convenient to write $\hat{V} = \hat{\sigma}^2\hat{\Sigma}$ where $\hat{\sigma}^2 = \|\hat{V}\|$ and $\hat{\Sigma} = \hat{V}/\|\hat{V}\|$. Note that (2.2) then implies $n^{1/2}(1 - \hat{\sigma}^2/\sigma^2) = O_p(1)$.

2.2. The effects of measurement error. Our investigation starts with a study of the estimator obtained by regressing y_i on the observed X_i. This estimator, to be called $\hat{\beta}$, maximizes

$$L_n(\gamma) \triangleq n^{-1}\sum_1^n \{ y_i \log F(c_i^T\gamma) + (1 - y_i)\log F(-c_i^T\gamma)\} \tag{2.3}$$

and satisfies

$$\sum_1^n (y_i - F(c_i^T\hat{\beta}))c_i = 0, \tag{2.4}$$

when $c_i = X_i$, $i = 1, \dots, n$. Our interest lies in the behavior of $\hat{\beta}$ as $\max(\sigma, n^{-1}) \to 0$. In addition to assumptions on the errors ε_i, some design conditions are necessary to insure weak consistency of $\hat{\beta}$. We shall work with the following assumptions where $\|\cdot\|$ denotes the Euclidean norm:

(C1) $G_n(\gamma)$ converges pointwise to a function $G(\gamma)$ possessing a unique maximum at β_0, where $G_n(\cdot)$ is defined as

$$G_n(\gamma) \triangleq n^{-1}\sum_1^n \{F(x_i^T\beta_0)\log F(x_i^T\gamma) + F(-x_i^T\beta_0)\log F(-x_i^T\gamma)\};$$

(C2) $\sum_1^n (\|x_i\|)^2 = o(n^2)$;

(C3) $E(\|\varepsilon_1\|) < \infty$.

The condition (C1) is an assumption of convenience since for each n, $G_n(\cdot)$ is concave with a maximum at β_0. Weaker conditions could thus be employed by studying subsequences of $G_n(\cdot)$ [see Theorem 10.9, Rockafellar (1970)].

Consistency of $\hat{\beta}$ is proved in Theorem 5.1. This result is necessary to establish the following asymptotic expansion which is crucial to our investigation. Theorem 1 gives conditions such that

$$\begin{aligned}\hat{\beta} = \beta_0 &+ n^{-1/2}S_n^{-1}(\beta_0)Z_n \\ &+ \sigma^2 S_n^{-1}(\beta_0)(J_{n,1} + J_{n,2})\beta_0 + o_p(\max(\sigma^2, n^{-1/2})),\end{aligned} \tag{2.5}$$

where

$$Z_n = n^{-1/2}\sum_1^n \left(y_i - F(x_i^T\beta_0)\right)x_i$$

$$J_{n,1} = -(2n)^{-1}\sum_1^n F^{(2)}(x_i^T\beta_0)x_i\beta_0^T\Sigma$$

$$J_{n,2} = -n^{-1}\sum_1^n F^{(1)}(x_i^T\beta_0)\Sigma.$$

THEOREM 1. (Asymptotic expansion of $\hat{\beta}$). *Assume that $\hat{\beta}$ is a consistent estimator of β_0 satisfying* (2.4). *Also assume*:

(A1) *There exists a positive definite matrix M, $\delta > 0$ and $N_0 < \infty$, such that $S_n(\gamma) \geq M$ whenever $n \geq N_0$ and $\|\gamma - \beta_0\| \leq \delta$*;

(A2) $n^{-1}\Sigma_1^n\|x_i\|^2 \to x^2 < \infty$, $\max_{1\leq i\leq n}\|x_i\| = o(\sigma^{-2})$;

(A3) $E(\varepsilon_1) = 0$, $E(\varepsilon_1\varepsilon_1^T) = I$, $E(\|\varepsilon_1\|^{2+\alpha}) < \infty$ *for some* $\alpha > 0$.

Then $\hat{\beta}$ has the expansion given in (2.5).

Note that the first part of (A2) implies $\max_{1\leq i\leq n}\|x_i\| = o(n^{1/2})$. This fact is used repeatedly in the proofs in Section 5. Assumptions (A1) and (A2) are sufficient to prove asymptotic normality for $S_n^{-1/2}(\beta_0)Z_n$ by using the Cramér–Wold device (Billingsley, 1979, Theorem 29.4) and appealing to Proposition 5.3.2 of Laha and Rohatgi (1979). Thus Theorem 1 indicates that with $\lambda = n^{1/2}\sigma^2$, we can expect $n^{1/2}(\hat{\beta} - \beta_0)$ to be approximately normally distributed with mean $\lambda S_n^{-1}(\beta_0)(J_{n,1} + J_{n,2})\beta_0$ and covariance $S_n^{-1}(\beta_0)$, when n is large and σ is small. When X_i is a mean of m replicates, $\sigma^2 \propto m^{-1}$ and λ describes the relationship between the sample size and the rate of replication. The asymptotic bias obviously decreases with increasing replication.

We can use expansion (2.5) to construct a corrected estimator, $\hat{\beta}_c$, which has smaller asymptotic bias. Before doing so we comment on the problems with $\hat{\beta}$ alluded to in the introduction.

BIAS AND ATTENUATION. Consider simple logistic regression through the origin with $\beta_0 > 0$. One expects to see attenuation, i.e., a negative first-order bias term. For most designs this is true. Somewhat surprisingly and completely at odds with the linear regression case, $S_n^{-1}(\beta_0)(J_{n,1} + J_{n,2})\beta_0$ can be positive. One design in which this occurs arises when most cases have very high or very low risk, i.e., $|x_i^T\beta_0|$ is large for most i.

HYPOTHESIS TESTING. Consider a two-group analysis of covariance, $x_i^T = (1, (-1)^i, d_i)$, $\beta_0 = (\beta_0, \beta_1, \beta_2)$. The covariance d_i is measured with error having variance σ^2. Often interest lies in testing hypotheses about the treatment effect β_1. A standard method to test $\beta_1 = 0$ is to compute its logistic regression estimate compared to the usual estimate of its asymptotic standard error. When the asymptotics of Theorem 1 are relevant and $n^{1/2}\sigma \to \lambda > 0$, this test approaches its nominal level only if the second component of $S_n^{-1}(\beta_0)(J_{n,1} + J_{n,2})\beta_0$

approaches zero. Letting s_2 denote the second row of $S_n^{-1}(\beta_0)$ this is achieved only if

$$n^{-1}\sum_1^n s_2^T x_i F^{(2)}(x_i^T\beta_0)\sigma^2\beta_2^2 \to 0.$$

This will not hold in the common epidemiologic situation in which the true covariables are not balanced across the two treatments. Thus, when substantial measurement error occurs in a nonrandomized study, there will be bias in the asymptotic levels of the usual tests.

3. Accounting for measurement error. In this section three alternative approaches to estimatation are studied. The first is based on expansion (2.5) and is distribution free in the sense that only moment assumptions are made about the measurement errors. The second two methods are based on an assumption of normally distributed errors; their asymptotic properties are then studied under more general conditions.

3.1. Adjusting for bias in $\hat{\beta}$. Write $b_n = S_n^{-1}(\beta_0)(J_{n,1} + J_{n,2})\beta_0$ and $\hat{b}_n = \hat{S}_n^{-1}(\hat{\beta})(\hat{J}_{n,1} + \hat{J}_{n,2})\hat{\beta}$, where

$$\hat{S}_n(\gamma) = n^{-1}\sum_1^n F^{(1)}(X_i^T\gamma)X_iX_i^T$$

$$\hat{J}_{n,1} = -(2n)^{-1}\sum_1^n F^{(2)}(X_i^T\hat{\beta})X_i\hat{\beta}^T\hat{\Sigma}, \tag{3.1}$$

$$\hat{J}_{n,2} = -n^{-1}\sum_1^n F^{(1)}(X_i^T\hat{\beta})\hat{\Sigma};$$

$\hat{b}_n$ depends only on the observed data and, under the conditions of Theorem 1 and (2.2), approximates b_n in the sense that $\hat{b}_n - b_n = o_p(1)$ as $\min(n, \sigma^{-1}) \to \infty$. This result suggests that the bias-corrected estimator $\hat{\beta}_c \triangleq \hat{\beta} - \hat{\sigma}^2\hat{b}_n$ should have smaller asymptotic bias for large n and small σ. We state these results as a theorem.

THEOREM 2. *Assume the conditions of Theorem* 1 *and* (2.2). *Then $\hat{\beta}_c$ is consistent and*

$$\hat{\beta}_c = \beta_0 + n^{-1/2}S_n^{-1}(\beta_0)Z_n + o_p(\max(\sigma^2, n^{-1/2})).$$

REMARKS. Theorem 2 follows from Theorems 5.1 and 5.2 which are proved using the following characterization of $\hat{\beta}_c$. Note that $\hat{\beta}_c = (I - \hat{\sigma}^2\hat{B}_n)\hat{\beta}$ where $\hat{B}_n = \hat{S}_n^{-1}(\hat{\beta})(\hat{J}_{n,1} + \hat{J}_{n,2})$. Since $X_i^T\hat{\beta} = X_i^T(I - \hat{\sigma}^2\hat{B}_n)^{-1}\hat{\beta}_c$ it follows that $\hat{\beta}_c$ maximizes (2.3) when $c_i = \hat{x}_{i,c}$, defined as

$$\hat{x}_{i,c} = X_i + \hat{\sigma}^2(I - \hat{\sigma}^2\hat{B}_n^T)^{-1}\hat{B}_n^T X_i. \tag{3.2}$$

In this sense $\hat{\beta}_c$ is a type of two-stage estimator obtained by doing logistic regression with $\hat{x}_{i,c}$ replacing X_i.

The estimator $\hat{\beta}_c$ is not unbiased just less biased. The Monte Carlo study of Section 4 shows that in realistic sampling situations the reduction in bias can be substantial.

3.2. Normal measurement error. When measurement error is present there is an added source of variability which is not accounted for by model (1.1). We now expand this model by assuming that (ε_i) are normally distributed, an assumption which is not unreasonable in some situations. The log-likelihood for estimating β_0 and $x_1, \dots, x_n$ is then

$$\sum_1^n \left\{ y_i \log\left(F(x_i^T\beta)\right) + (1 - y_i)\log\left(F(-x_i^T\beta)\right) - (2\sigma^2)^{-1}(X_i - x_i)^T \Sigma^{-1}(X_i - x_i) \right\}. \tag{3.3}$$

The vectors $\tilde{\beta}_f$, $\tilde{c}_i$ maximizing (3.3) satisfy

$$\sum_1^n \left(y_i - F\left(\tilde{c}_i^T \tilde{\beta}_f\right)\right)\tilde{c}_i = 0$$

$$\tilde{c}_i = X_i + \left(y_i - F\left(\tilde{c}_i^T \tilde{\beta}_f\right)\right)\sigma^2 \Sigma \tilde{\beta}_f, \qquad i = 1, \dots, n.$$

There are two problems with this estimator—it depends on the unknown matrix $\sigma^2\Sigma$ and solving for $\tilde{\beta}_f$ and $(\tilde{c}_i)$ is difficult. For these reasons we suggest an approximate version of $\tilde{\beta}_f$. Noting the form of $\tilde{c}_i$ we let

$$\hat{x}_{i,f} = X_i + \left(y_i - F\left(X_i^T \hat{\beta}\right)\right)\hat{\sigma}^2 \hat{\Sigma} \hat{\beta} \tag{3.4}$$

and define $\hat{\beta}_f$ as the estimator obtained by maximizing (2.3) with $c_i = \hat{x}_{i,f}$; $\hat{\beta}_f$ is consistent and has an asymptotic expansion given in the next theorem. The assumption of normal errors is not necessary for Theorem 3.

THEOREM 3. *Assume the conditions of Theorem* 1 *and* (2.2). *Then* $\hat{\beta}_f$ *is consistent and*

$$\hat{\beta}_f = \beta_0 + n^{-1/2} S_n^{-1}(\beta_0) Z_n + \sigma^2 S_n^{-1}(\beta_0) J_{n,1} \beta_0 + o_p\left(\max(\sigma^2, n^{-1/2})\right). \tag{3.5}$$

REMARKS. A comparison of (2.5) and (3.5) indicates that our approximate functional maximum likelihood estimator, $\hat{\beta}_f$, and the uncorrected estimator, $\hat{\beta}$, have first-order biases of the same magnitude. It can be shown (Stefanski, 1983) that the bias term in $\hat{\beta}_f$ is not due to our one-step modification nor to use of $\hat{V}$ in place of V, i.e., when V is known the full functional maximum likelihood estimator, $\tilde{\beta}_f$, also has the expansion given in (3.5) even in the case of simple logistic regression. This is in contrast to linear regression where, if the ratio of error variances is known or if there is finite replication of the predictors, the functional maximum likelihood estimator is consistent.

Our final estimator starts with an assumption of normal errors and exploits the consequences of sufficiency. Given $\sigma^2\Sigma$ and β_0, a sufficient statistic for estimating x_i is $\bar{c}_i(\beta_0) = X_i + \sigma^2(y_i - \frac{1}{2})\Sigma\beta_0$. It follows that the distribution of

y_i given $\bar{c}_i(\beta_0)$ does not depend on x_i. The reason for using this particular sufficient statistic is that

(3.6) $$P\{y_i = 1|\bar{c}_i(\beta_0)\} = F(\bar{c}_i^T(\beta_0)\beta_0)$$

and hence the score equation

(3.7) $$\sum_1^n (y_i - F(\bar{c}_i^T(\beta)\beta))\bar{c}_i(\beta) = 0$$

is unbiased for β_0. The conditional probability (3.6) also suggests another approach—replace c_i by $\bar{c}_i(\gamma)$ in (2.3) and maximize the resulting expression as a function of γ. However, a simple calculation indicates that the resulting score equation is not unbiased for β_0, thus we will confine our attention to (3.7).

Equation (3.7) can have multiple solutions not all which produce a consistent sequence of estimators. Since $\bar{c}_i(\beta)$ also depends on the unknown matrix $\sigma^2\Sigma$, we propose the following modification: Let

(3.8) $$\hat{x}_{i,s} = X_i + \hat{\sigma}^2(y_i - \tfrac{1}{2})\hat{\Sigma}\hat{\beta}$$

and define $\hat{\beta}_s$, the sufficiency estimator, as the maximizer of (2.3) when c_i is replaced by $\hat{x}_{i,s}$. This estimator is consistent and has the expansion given in the next theorem.

THEOREM 4. *Assume the conditions of Theorem* 1 *and* (2.2). *Then* $\hat{\beta}_s$ *is consistent and*

(3.9) $$\hat{\beta}_s = \beta_0 + n^{-1/2}S_n^{-1}(\beta_0)Z_n + o_p(\max(\sigma^2, n^{-1/2})).$$

REMARKS 1. Theorem 4 does not require the assumption of normal measurement error. Also, $\hat{\beta}$ can be replaced by any consistent estimator in the definition of $\hat{x}_{i,s}$. The effects of nonnormal measurement error and our particular choice of $\hat{x}_{i,s}$ become apparent only when $\hat{\beta}_s$ is expanded through terms of order $\max^2(\sigma^2, n^{-1/2})$. This analysis is lengthy and is not presented here [see Stefanski (1983)].

2. It is possible to define a sufficiency estimator for a large class of measurement-error models. In particular, we have in mind the generalized linear models with canonical link functions (McCullagh and Nelder, 1983). A complete exposition of this theory will appear elsewhere.

In the discussion following Theorem 1, it was noted that $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically normal with nonzero mean provided $n^{1/2}\sigma^2 \to \lambda$. It follows from Theorems 2 and 4 that both $n^{1/2}(\hat{\beta}_c - \beta_0)$ and $n^{1/2}(\hat{\beta}_s - \beta_0)$ are asymptotically normal with zero means under the same conditions. Furthermore, it can be shown that for $\hat{\beta}_c$ and $\hat{\beta}_s$ asymptotic normality is obtained under the weaker condition $n^{1/2}\sigma^4 \to \lambda$ [see Stefanski (1983) for details].

In the next section results from a small Monte Carlo study are presented.

4. Monte Carlo. We conducted a small simulation experiment to determine the relative merits of the four estimators $\hat{\beta}$, $\hat{\beta}_c$, $\hat{\beta}_f$, and $\hat{\beta}_s$. The model for the study was

$$\Pr\{y_i = 1|d_i\} = F(\alpha + \beta d_i), \qquad i = 1, \ldots, n, \tag{4.1}$$

where $F(\cdot)$ is defined in (1.1).

As our estimators are derived for the functional case, one possible Monte Carlo study would have consisted of generating for fixed $(d_1, \ldots, d_n)$ a sequence of response vectors $(y_1, \ldots, y_n)$ according to (4.1), and a sequence of measurement-error vectors. This would allow evaluation of the estimators' performance for the particular design $(d_1, \ldots, d_n)$. However, several different designs would have to be studied in order to obtain a useful overall measure of performance. We opted instead for a study in which at each step the design $(d_1, \ldots, d_n)$ is generated at random and, in turn, a single response vector and measurement-error vector are generated. After a number of such steps are completed, the overall performance of the estimators is investigated [c.f. Olkin, Petkau, and Zidek (1981) and Dempster, Schatzoff, and Wermuth (1977)]. We believe this approach better indicates the estimators' performance in a wide variety of settings.

We considered these sampling situations where χ_1^2 denotes a chi-squared random variable with one degree of freedom:

$$\text{(I)} \qquad (\alpha, \beta) = (-1.4, 1.4), \qquad (d_i) \sim \text{Normal}(0, \sigma_d^2);$$

$$\text{(II)} \qquad (\alpha, \beta) = (-1.4, 1.4), \qquad (d_i) \sim \sigma_d(\chi_1^2 - 1)/\sqrt{2};$$

where,

$$\sigma_d^2 = 0.10, \qquad n = 300, 600.$$

For each case, we considered two sampling distributions for the measurement errors: (a) Normal$(0, \tau^2)$ and (b) a contaminated normal distribution, which is Normal$(0, \tau^2)$ with probability 0.90 and Normal$(0, 25\tau^2)$ with probability 0.10. For both cases, τ^2 was one third the variance of the true predictors ($\tau^2 = \sigma_d^2/3$).

We believe these two sampling situations are realistic, but their representativeness is limited by the size of our study. The sample sizes $n = 300, 600$ may seem large, but our primary interest is in larger epidemiologic studies where such sample sizes are common. For example, Clark (1982) was motivated by a study with $n = 2580$, Hauck (1983) quotes a partially completed study with $n \geq 340$, and we have analyzed Framingham data for males aged 45–54 with $n = 589$. In addition, for the particular designs in our study, the unconditional probability of response $(y = 1)$ is only about 0.10. As in the case of Bernoulli trials, an estimator's variance decreases more like $1/np(1-p)$ than $1/n$ and for this reason $np(1-p)$ is sometimes called the effective sample size. In our study the effective sample sizes are only about 30 and 60 respectively. Furthermore, the results of the study suggest that correcting for measurement error when the effective sample size is small is unwarranted, except possibly when measurement error is larger than what we have studied.

The values of the predictor variance σ_d^2 and the normal measurement error variance τ^2 are similar to those found in the Framingham cohort mentioned in the previous paragraph when the predictor was $\log_e\{$(systolic blood pressure $-$ 75)/3$\}$, a standard transformation. The choice of (α, β) comes from Framingham data as well. All experiments were repeated 100 times.

In each experiment, we sampled two independent measurements $(D_{i,1}, D_{i,2})$ of each d_i; the observed covariate was $X_i = (1, \bar{D}_i)^T$, where $\bar{D}_i = (D_{i,1} + D_{i,2})/2$. Thus σ^2, the variance of $\bar{D}_i$, was equal to $(1/6)\sigma_d^2$ for the case of normal measurement error while for the contaminated normal error distribution $\sigma^2 = (3.4/6)\sigma_d^2$. The matrix $\sigma^2\Sigma$ has only one nonzero entry which was estimated by the sample variance of $(D_{i,1} - D_{i,2})/2$.

In addition to the four estimators presented in this paper, we included in the study a proposal due to Clark (1982). She suggests the estimator $\hat{\beta}_N$ obtained by maximizing (2.3) when c_i is replaced by $\hat{x}_{i,N} = X_i - \hat{\sigma}^2\hat{\Sigma}\hat{\Sigma}_X^{-1}(X_i - \hat{\mu})$ where $\hat{\mu}$ and $\hat{\Sigma}_X$ are the sample mean and covariance of the observed data. Motivation for this estimator derives from an assumption of normal errors and normal covariates x_i. In this case $E(x_i|X_i) = X_i - \sigma^2\Sigma\Sigma_X^{-1}(X_i - \mu)$ and hence $\hat{x}_{i,n}$ is a natural estimator of x_i. Theorems 5.1 and 5.2 can be used to prove consistency and derive an asymptotic expansion for this estimator. Like $\hat{\beta}$ and $\hat{\beta}_f$, $\hat{\beta}_N$ has a nonzero first-order bias although it is too lengthy to present here.

Sweeping conclusions cannot be made from such a small study. However, we can make the following qualitative suggestions. First $\hat{\beta}$ is less variable but more biased than the others. Sample sizes such as $n = 600$ as in the study or Clark's $n = 2580$ are such that bias dominates and hence are candidates for using corrected estimators. An opposite conclusion holds for small sample sizes where variance dominates. A second suggestion from Table 1 is that when $\mathrm{Var}(\hat{\beta})$ is small relative to its bias [Case I(b), II(b), and when $n = 600$], the corrected estimators perform quite well.

Both $\hat{\beta}_s$ and $\hat{\beta}_f$ were defined via an assumption of normal errors yet they also performed well when the errors were contaminated normal [Cases I(b), II(b)]. Clark's estimator proved to be sensitive to the assumption of normal covariates; $\hat{\beta}_N$ performed very well in our study when the predictors were normally distributed, but it did have a noticeable drop in efficiency when the predictors were highly skewed (Case II). Finally, the corrected estimator $\hat{\beta}_c$, which was derived with no distributional assumptions for either the predictors or errors, performed well throughout the study.

In summary, the Monte Carlo results suggest that the estimators $\hat{\beta}_c$, $\hat{\beta}_f$, $\hat{\beta}_s$, and Clark's $\hat{\beta}_N$ are useful alternatives to $\hat{\beta}$ when covariates are measured with error. The pressing practical problem now appears to be how to delineate those situations in which ordinary logistic regression should be corrected for its bias. Studies of inference and more detailed comparisons of alternative estimators will be enhanced by the identification of those problems where measurement error severely affects the usual estimation and inference.

5. Proofs of theorems. Consider the estimator $\tilde{\beta}$ obtained by maximizing (2.3) when c_i is replaced with $\tilde{x}_i$ where

$$\tilde{x}_i = x_i + \sigma v_i + \sigma^2 g_{in}. \tag{5.1}$$

In Theorem 5.1 we prove weak consistency of $\tilde{\beta}$ under conditions (C1), (C2), (C3),

TABLE 1

Results from our Monte Carlo study of the simple logistic regression model $\Pr\{y_i = 1|d_i\} = F(\alpha + \beta d_i)$. *Observed covariates are* $X_i = (1, \overline{D}_i)^T$ *where* $\overline{D}_i$ *is the mean of two independent measurements of* d_i. *The normal measurement errors have variance* $\sigma_d^2 / 3$; *the contaminated normal errors have distribution function* $G(x) = 0.9\Phi(x/\tau) + 0.1\Phi(x/5\tau)$ *and variance* $(3.4/3)\sigma_d^2$. *"Efficiency" refers to mean-squared error efficiency with respect to ordinary logistic regression.*

	$\hat{\beta}$	$\hat{\beta}_c$	$\hat{\beta}_f$	$\hat{\beta}_N$	$\hat{\beta}_s$
CASE I(a). $(\alpha, \beta) = (-1.4, 1.4)$, $(d_i) \sim N(0, \sigma_d^2 = 0.1)$, *normal measurement error.*					
$n = 300$ Bias	−0.21	−0.04	−0.05	−0.02	−0.06
Std. Dev.	0.52	0.61	0.61	0.61	0.60
Efficiency	100%*	85%	85%	84%	88%
$n = 600$ Bias	−0.22	−0.05	−0.05	−0.02	−0.06
Std. Dev.	0.33	0.38	0.38	0.38	0.38
Efficiency	100%*	108%	106%	107%	108%
CASE I(b). Same as Case I(a) but measurement errors have the contaminated normal distribution.					
$n = 300$ Bias	−0.49	−0.16	−0.19	0.02	−0.20
Std. Dev.	0.34	0.48	0.48	0.54	0.46
Efficiency	100%*	143%	139%	121%	143%
$n = 600$ Bias	−0.53	−0.20	−0.21	−0.03	−0.22
Std. Dev.	0.24	0.33	0.34	0.38	0.33
Efficiency	100%*	223%	215%	234%	216%
CASE II(a). $(\alpha, \beta) = (-1.4, 1.4)$, $(d_i) \sim \sigma_d(\chi_1^2 - 1)/\sqrt{2}$, $\sigma_d^2 = 0.1$, *normal measurement error.*					
$n = 300$ Bias	−0.28	−0.05	−0.07	0.10	−0.08
Std. Dev.	0.47	0.58	0.57	0.66	0.56
Efficiency	100%*	90%	91%	69%	93%
$n = 600$ Bias	−0.27	−0.03	−0.04	0.11	−0.05
Std. Dev.	0.33	0.41	0.41	0.45	0.40
Efficiency	100%*	111%	110%	85%	112%
CASE II(b). Same as Case II(a) but measurement errors have the contaminated normal distribution.					
$n = 300$ Bias	−0.43	−0.13	−0.15	0.12	−0.17
Std. Dev.	0.33	0.44	0.45	0.53	0.43
Efficiency	100%*	141%	134%	103%	141%
$n = 600$ Bias	−0.46	−0.15	−0.16	0.10	−0.18
Std. Dev.	0.25	0.33	0.34	0.40	0.33
Efficiency	100%*	201%	190%	159%	194%

*By definition.

and

$$\text{(P1)} \qquad \sum_1^n \|g_{in}\|^2 = O_p(n).$$

In Theorem 5.2 an asymptotic expansion for $\tilde{\beta}$ is given. The consistency and asymptotic expansions of $\hat{\beta}$, $\hat{\beta}_c$, $\hat{\beta}_f$, and $\hat{\beta}_s$ follow from these general results by noting that X_i, $\hat{x}_{i,c}$, $\hat{x}_{i,f}$, and $\hat{x}_{i,s}$ all have the representation given in (5.1). We remind the reader that all the asymptotic expressions hold as $\max(\sigma, n^{-1}) \to 0$.

Theorem 5.1 (Consistency). *Assume* (C1), (C2), (C3), *and* (P1), *then* $\tilde{\beta} - \beta_0 = o_p(1)$.

Proof. Define $\tilde{L}_n(\gamma)$ to be the function obtained by taking $c_i = \tilde{x}_i$ in (2.3). The identity $\log(F(t)/(1 - F(t))) = t$ is used to show $\tilde{L}_n(\gamma) - G_n(\gamma) = R_{n,1} + R_{n,2}$, where

$$R_{n,1} = n^{-1}\sum_1^n \left(y_i - F\left(x_i^T\beta_0\right)\right)x_i^T\gamma$$

$$R_{n,2} = n^{-1}\sum_1^n \left\{ y_i\left(\tilde{x}_i^T\gamma - x_i^T\gamma\right) + \log F\left(-\tilde{x}_i^T\gamma\right) - \log F\left(-x_i^T\gamma\right)\right\}.$$

Under (C2), $R_{n,1}$ has mean zero and asymptotically negligible variance, and by (C3) and (P1),

$$\|R_{n,2}\| \le 2\sigma\|\gamma\|n^{-1}\sum_1^n \|v_i + \sigma g_{in}\| = o_p(1).$$

Consequently (C1) implies that $\tilde{L}_n(\cdot)$ converges pointwise in probability to $G(\cdot)$. An appeal to Corollary II.2 of Andersen and Gill (1982) concludes the proof.

The consistency results follow by applying Theorem 5.1 first to $\hat{\beta}$, ($g_{in} = 0$) and then to $\hat{\beta}_c$, $\hat{\beta}_f$, and $\hat{\beta}_s$. Next we derive the asymptotic expansions for these estimators.

Theorem 5.2 (Asymptotic expansion). *Assume* (P1) *and the conditions of Theorem* 1, *then*

$$\tilde{\beta} = \beta_0 + n^{-1/2}S_n^{-1}(\beta_0)Z_n + \sigma^2 S_n^{-1}(\beta_0)\{(J_{n,1} + J_{n,2})\beta_0 + b_{n,3} + b_{n,4}\} + o_p\left(\max(\sigma^2, n^{-1/2})\right),$$

where

$$b_{n,3} = n^{-1}\sum_1^n \left(y_i - F\left(x_i^T\beta_0\right)\right)g_{in},$$

$$b_{n,4} = -n^{-1}\sum_1^n F^{(1)}\left(x_i^T\beta_0\right)x_i g_{in}^T\beta_0,$$

where $S_n(\cdot)$ *is given in* (1.2), *and* Z_n, $J_{n,1}$, *and* $J_{n,2}$ *are defined in* (2.5).

Theorem 5.2 is proved with a series of lemmas. First we show how Theorems 1–4 follow as corollaries. Theorem 1 is immediate since $g_{in} \equiv 0$ for $\hat{\beta}$. For $\hat{\beta}_c$, $g_{in} = (\hat{\sigma}^2/\sigma^2)(I - \hat{\sigma}^2\hat{B}_n^T)^{-1}\hat{B}_n^T X_i$ where $\hat{B}_n = \hat{S}_n^{-1}(\hat{\beta})(\hat{J}_{n,1} + \hat{J}_{n,2})$. Assumptions (A2), (A3), Lemma 5.1, and (2.2) imply $b_{n,3} = o_p(1)$, and

$$\begin{aligned}-b_{n,4} &= n^{-1}\sum_1^n F^{(1)}\left(x_i^T\beta_0\right)x_i X_i^T\hat{B}_n\left(I - \hat{\sigma}^2\hat{B}_n\right)^{-1}\beta_0 \\ &= S_n(\beta_0)\hat{B}_n\beta_0 + o_p(1) \\ &= (J_{n,1} + J_{n,2})\beta_0 + o_p(1),\end{aligned}$$

thus proving Theorem 2.

For $\hat{\beta}_f$, $g_{in} = (\hat{\sigma}^2/\sigma^2)(y_i - F(X_i^T\hat{\beta}))\hat{\Sigma}\hat{\beta}$ and (A2), (A3), Lemma 5.1, and (2.2) imply $b_{n,4} = o_p(1)$, and

$$b_{n,3} = n^{-1}\sum_1^n \left(y_i - F(x_i^T\beta_0)\right)^2 \Sigma\beta_0 + o_p(1)$$
$$= -J_{n,2}\beta_o + o_p(1).$$

Theorem 3 follows. Finally for $\hat{\beta}_s$, $g_{in} = (\hat{\sigma}^2/\sigma^2)(y_i - \frac{1}{2})\hat{\Sigma}\hat{\beta}$. (A2), (A3), Lemma 5.1, and (2.2) imply

$$b_{n,3} = n^{-1}\sum_1^n \left(y_i - F(x_i^T\beta_0)\right)\left(y_i - \tfrac{1}{2}\right)\Sigma\beta_0 + o_p(1)$$
$$= -J_{n,2}\beta_0 + o_p(1)$$
$$b_{n,4} = -n^{-1}\sum_1^n F^{(1)}(x_i^T\beta_0)\left(y_i - \tfrac{1}{2}\right)x_i\beta_0^T\Sigma\beta_0 + o_p(1)$$
$$= -n^{-1}\sum_1^n F^{(1)}(x_i^T\beta_0)\left(F(x_i^T\beta_0) - \tfrac{1}{2}\right)x_i\beta_0^T\Sigma\beta_0 + o_p(1)$$
$$= -J_{n,1}\beta_0 + o_p(1).$$

In the last step we use the identity $F^{(2)}(t) = F^{(1)}(t)(1 - 2F(t))$. This proves Theorem 4. Notice that in deriving these results we used only the fact that $\hat{\beta} - \beta_0 = o_p(1)$. Thus the conclusions of theorems 3 and 4 remain unchanged if $\hat{\beta}$ is replaced by any other *consistent* estimator in the definitions of $\hat{x}_{i,f}$ and $\hat{x}_{i,s}$. In particular, this can be shown to imply that the fully iterated versions of the functional and sufficiency estimators (provided consistent versions are chosen) also satisfy Theorems 3 and 4, respectively (Stefanski, 1983).

The proof of Theorem 5.2 starts with the following weak law.

LEMMA 5.1. *Let $u_1, u_2, \ldots$ be independent random vectors such that $E(u_i) = 0$ for all i, and $E(|u_{ij}|^{1+\alpha}) \le B$ for all i and j, and some $\alpha > 0$ and $B < \infty$, where u_{ij} is the jth element of u_i. If $\Sigma_i^n|a_i| = O(n)$ and $\max_{1\le i\le n}(|a_i|/n) = o(1)$ then $n^{-1}\Sigma_1^n a_i u_i = o_p(1)$.*

PROOF. The proof of the lemma entails a routine verification of the assumptions of Theorem 5.2.3 (Chung, 1974) and is not given here.

LEMMA 5.2. *Under the conditions of Theorem* 1,

$$n^{-1}\sum_1^n \left(y_i - F(X_i^T\beta_0)\right)X_i = n^{-1/2}Z_n + \sigma^2(J_{n,1} + J_{n,2})\beta_0 + o_p\left(\max(\sigma^2, n^{-1/2})\right).$$

PROOF. $n^{-1}\Sigma_1^n(y_i - F(X_i^T\beta_0))X_i = T_{n,1} + T_{n,2}$, where

$$T_{n,1} = n^{-1}\sum_1^n \left(y_i - F(X_i^T\beta_0)\right)x_i,$$
$$T_{n,2} = \sigma n^{-1}\sum_1^n \left(y_i - F(X_i^T\beta_0)\right)v_i.$$

A Taylor series expansion of $F(\cdot)$ shows that

$$T_{n,1} = n^{-1/2}Z_n + \sigma^2 J_{n,1}\beta_0 + n^{-1/2}Q_{n,1,\sigma} + \sigma^2(D_{n,1} + R_{n,1}),$$

where

$$Q_{n,1,\sigma} = -\sigma n^{-1/2}\sum_1^n F^{(1)}(x_i^T\beta_0)v_i^T\beta_0 x_i$$

$$D_{n,1} = -(2n)^{-1}\sum_1^n \left\{F^{(2)}(x_i^T\beta_0)\left((v_i^T\beta_0)^2 - \beta_0^T\Sigma\beta_0\right)x_i\right\}$$

$$R_{n,1} = -(2n)^{-1}\sum_1^n \left(F^{(2)}(\tilde{X}_i^T\beta_0) - F^{(2)}(x_i^T\beta_0)\right)(v_i^T\beta_0)^2 x_i,$$

and $\tilde{X}_i$ is on the line segment joining x_i to X_i. $Q_{n,1,\sigma}$ has mean zero and asymptotically negligible variance thus $n^{-1/2}Q_{n,1,\sigma} = o_p(n^{-1/2})$. Assumptions (A2) and (A3) and Lemma 5.1 are used to show $D_{n,1} = o_p(1)$. Also note that

$$\|R_{n,1}\| \le (2n)^{-1}\sum_1^n \|x_i\|(v_i^T\beta_0)^2 \min(1, \sigma|v_i^T\beta_0|) \le A_n A_n^*,$$

where

$$A_n = \left(n^{-1}\sum_1^n \|x_i\|^2(v_i^T\beta_0)^2\right)^{1/2},$$

$$A_n^* = \left(n^{-1}\sum_1^n (v_i^T\beta_0)^2 \min^2(1, \sigma|v_i^T\beta_0|)\right)^{1/2}.$$

Assumptions (A2) and (A3) and Lemma 5.1 imply $A_n = O_p(1)$ while (A3), the fact that $\max(n^{-1}, \sigma) \to 0$, and the Dominated Convergence Theorem imply $A_n^* = o_p(1)$. It follows that $\sigma^2(D_{n,1} + R_{n,1}) = o_p(\sigma^2)$. Combining these results we get

$$T_{n,1} = n^{-1/2}Z_n + \sigma^2 J_{n,1}\beta_0 + o_p(\max(\sigma^2, n^{-1/2})). \tag{5.2}$$

Another Taylor series expansion of $F(\cdot)$ shows that

$$T_{n,2} = \sigma^2 J_{n,2}\beta_0 + n^{-1/2}Q_{n,2,\sigma} + \sigma^2(D_{n,2} + R_{n,2}),$$

where

$$Q_{n,2,\sigma} = \sigma n^{-1/2}\sum_1^n (y_i - F(x_i^T\beta_0))v_i$$

$$D_{n,2} = -n^{-1}\sum_1^n F^{(1)}(x_i^T\beta_0)(v_i v_i^T - \Sigma)\beta_0$$

$$R_{n,2} = -n^{-1}\sum_1^n \left(F^{(1)}(\tilde{X}_i^T\beta_0) - F^{(1)}(x_i^T\beta_0)\right)v_i v_i^T\beta_0,$$

and $\tilde{X}_i$ lies on the line segment joining x_i to X_i. $Q_{n,2,\sigma}$, $D_{n,2}$ and $R_{n,2}$ are all $o_p(1)$, and the proofs are analogous to those for $Q_{n,1,\sigma}$, $D_{n,1}$, and $R_{n,1}$, respectively. Consequently,

$$T_{n,2} = \sigma^2 J_{n,2}\beta_0 + o_p(\max(\sigma^2, n^{-1/2})). \tag{5.3}$$

Combining (5.2) and (5.3) completes the proof of the lemma.

LEMMA 5.3. *Assume the conditions of Theorem* 1 *and* (P1) *and define* $\tilde{H}_n(\gamma) = n^{-1}\Sigma_1^n(y_i - F(\tilde{x}_i^T\gamma))\tilde{x}_i$. *Then*

$$\tilde{H}_n(\beta_0) = n^{-1/2}Z_n + \sigma^2((J_{n,1} + J_{n,2})\beta_0 + b_{n,3} + b_{n,4}) + o_p(\max(\sigma^2, n^{-1/2})).$$

PROOF. $\tilde{H}_n(\beta_0) = W_{n,1} + W_{n,2} + W_{n,3} + W_{n,4}$, where

$$W_{n,1} = n^{-1}\sum_1^n (y_i - F(X_i^T\beta_0))X_i,$$

$$W_{n,2} = \sigma n^{-1}\sum_1^n (F(X_i^T\beta_0) - F(\tilde{x}_i^T\beta_0))(v_i + \sigma g_{in}),$$

$$W_{n,3} = \sigma^2 n^{-1}\sum_1^n (y_i - F(X_i^T\beta_0))g_{in},$$

$$W_{n,4} = n^{-1}\sum_1^n (F(X_i^T\beta_0) - F(\tilde{x}_i^T\beta_0))x_i.$$

Note that in light of (A2) and (P1)

$$\|W_{n,2}\| \le \sigma^2 n^{-1}\sum_1^n \|g_{in}\|(\|v_i\| + \sigma\|g_{in}\|) = o_p(\sigma^2).$$

Also,

$$\begin{aligned}\|W_{n,3} - \sigma^2 b_{n,3}\| &\le \sigma^2 n^{-1}\sum_1^n \left|F(x_i^T\beta_0) - F(X_i^T\beta_0)\right|\|g_{in}\| \\ &\le \|\beta_0\|\sigma^3 n^{-1}\sum_1^n \|v_i\|\,\|g_{in}\| \\ &\le \|\beta_o\|\sigma^3\left(n^{-1}\sum_1^n\|v_i\|^2\right)^{1/2}\left(n^{-1}\sum_1^n\|g_{in}\|^2\right)^{1/2} \\ &= o_p(\sigma^2),\end{aligned}$$

using (A3) and (P1). One term in a Taylor series expansion of $F(\cdot)$ and Lemma 5.1, (A2), and (P1) show that

$$\begin{aligned}\|W_{n,4} - \sigma^2 b_{n,4}\| &\le \sigma^2\|\beta_0\|^2 n^{-1}\sum_1^n (\sigma\|v_i\| + \sigma^2\|g_{in}\|)\|x_i\|\,\|g_{in}\| \\ &\le \sigma^2\|\beta_0\|^2\left\{\sigma n^{-1}\sum_1^n\|v_i\|\,\|x_i\|\,\|g_{in}\| + \sigma^2 n^{-1}\sum_1^n\|x_i\|\,\|g_{in}\|^2\right\} \\ &\le \sigma^2\|\beta_0\|^2\left\{\sigma\left(n^{-1}\sum_1^n\|v_i\|^2\|x_i\|^2\right)^{1/2}\left(n^{-1}\sum_1^n\|g_{in}\|^2\right)^{1/2}\right. \\ &\qquad\left. +\sigma^2\left(\max_{1\le i\le n}\|x_i\|\right)n^{-1}\sum_1^n\|g_{in}\|^2\right\} \\ &= o_p(\sigma^2).\end{aligned}$$

An expansion for $W_{n,1}$ is given in Lemma 5.2. Combining the above results proves the lemma.

Define

$$\tilde{S}_n(\gamma) = n^{-1}\sum_1^n F^{(1)}(\tilde{x}_i^T\gamma)\tilde{x}_i\tilde{x}_i^T$$

and note that

$$\tilde{S}_n(\gamma) = (\partial/\partial\gamma)\tilde{H}_n(\gamma), \tag{5.4}$$

where $\tilde{H}_n(\cdot)$ is defined in Lemma 5.3.

LEMMA 5.4. *Conditions* (*A*2), (*A*3), *and* (*p*1) *imply* $\tilde{S}_n(\bar{\beta}) - S_n(\bar{\beta}) = o_p(1)$ *for any* $\bar{\beta}$ *on the line segment joining* β_0 *and* $\tilde{\beta}$.

PROOF. $\tilde{S}_n(\bar{\beta}) - S_n(\bar{\beta}) = H_{n,1} + H_{n,2}$, where

$$H_{n,1} = n^{-1}\sum_1^n F^{(1)}(\tilde{x}_i^T\bar{\beta})(\tilde{x}_i\tilde{x}_i^T - x_ix_i^T),$$

$$H_{n,2} = n^{-1}\sum_1^n \{F^{(1)}(\tilde{x}_i^T\bar{\beta}) - F^{(1)}(x_i^T\bar{\beta})\}x_ix_i^T.$$

The boundedness of $F^{(1)}(\cdot)$ and some elementary inequalities are used to show

$$\|H_{n,1}\| \le n^{-1}\sum_1^n \left(2\|x_i\|\,\|\sigma v_i + \sigma^2 g_{in}\| + \|\sigma v_i + \sigma^2 g_{in}\|^2\right)$$

$$\le 2\left(n^{-1}\sum_1^n \|x_i\|^2\right)^{1/2}\left(n^{-1}\sum_1^n\|\sigma v_i + \sigma^2 g_{in}\|^2\right)^{1/2} + n^{-1}\sum_1^n\|\sigma v_i + \sigma^2 g_{in}\|^2.$$

Assumption (A2) implies $n^{-1}\Sigma_1^n\|x_i\|^2 = O_p(1)$ and (A3) and (P1) imply $n^{-1}\Sigma_1^n\|\sigma v_i + \sigma^2 g_{in}\|^2 = o_p(1)$. Thus $\|H_{n,1}\| = o_p(1)$ as $\min(\sigma^{-1}, n) \to \infty$. A Taylor series expansion of $F^{(1)}(\cdot)$ and the boundedness of $F^{(2)}(\cdot)$ are used to show

$$\|H_{n,2}\| \le \|\bar{\beta}\| n^{-1}\sum_1^n \|\sigma v_i + \sigma^2 g_{in}\|\,\|x_i\|^2$$

$$\le \|\bar{\beta}\|\left\{\sigma n^{-1}\sum_1^n\|v_i\|\,\|x_i\|^2 + \sigma^2\left(\max_{1\le i\le n}\|x_i\|\right)\left(n^{-1}\sum_1^n g_{in}^2\right)^{1/2}\left(n^{-1}\sum_1^n\|x_i\|^2\right)^{1/2}\right\}. \tag{5.5}$$

Assumption (A2) and Lemma 5.1 imply $n^{-1}\Sigma_1^n\|v_i\|\,\|x_i\|^2 = O_p(1)$, and (A2) and (P1) imply that the second term in (5.5) is $o_p(1)$. Thus $\|H_{n,2}\| = o_p(1)$ as $\min(\sigma^{-1}, n) \to \infty$ and the proof is complete.

LEMMA 5.5. *Assume* (P1) *and the conditions of Theorem* 1, *then*

$$\tilde{\beta} - \beta_0 = O_p(\max(\sigma^2, n^{-1/2})).$$

PROOF. Let $\tilde{H}_n(\cdot)$ be the function defined in Lemma 5.3. Consider the real-valued function of γ defined as $\tilde{J}_n(\gamma) = \tilde{H}_n^T(\gamma)(\tilde{\beta} - \beta_0)$. The Mean Value Theorem proves the existence of some $\bar{\beta}$ on the line segment joining $\tilde{\beta}$ to β_0 such that

$$\tilde{H}_n^T(\beta_0)(\tilde{\beta} - \beta_0) = (\tilde{\beta} - \beta_0)^T \tilde{S}_n(\bar{\beta})(\tilde{\beta} - \beta_0),$$

where $\tilde{S}_n(\cdot)$ is defined in (5.4).

It follows that $\|\tilde{\beta} - \beta_0\| \le \|\tilde{H}_n(\beta_0)\| \lambda_{\min}^{-1}(\tilde{S}_n(\bar{\beta}))$ where $\lambda_{\min}(A) =$ minimum eigenvalue of A. By Lemma 5.4, $\tilde{S}_n(\bar{\beta}) - S_n(\bar{\beta}) = o_p(1)$ hence by (A1), $P\{\lambda_{\min}^{-1}(\tilde{S}_n(\bar{\beta})) \le 2\lambda_{\min}^{-1}(M)\} \to 1$. Thus $\|\tilde{\beta} - \beta_0\|$ and $\|\tilde{H}_n(\beta_0)\|$ have the same order which, from Lemma 5.3, is $O_p(\max(\sigma^2, n^{-1/2}))$.

We are now in a position to prove Theorem 5.2.

Proof of Theorem 5.2. By definition $n^{-1}\sum_1^n (y_i - F(\tilde{x}_i^T\tilde{\beta}))\tilde{x}_i = 0$; expanding $F(\cdot)$ in a Taylor series shows that $\tilde{S}(\tilde{\beta} - \beta_0) = \tilde{H}_n(\beta_0)$, where

$$\tilde{S} = n^{-1}\sum_1^n F^{(1)}(x_i^T\bar{\beta}_i)\tilde{x}_i\tilde{x}_i^T$$

and for each i, $\|\bar{\beta}_i - \beta_o\| \le \|\tilde{\beta} - \beta_0\|$. (A2), (A3), (P1), and the conclusion of Lemma 5.5 are used to show $\tilde{S} - S_n(\beta_0) = o_p(1)$. The theorem follows from Lemma 5.5.

Acknowledgment. This research was supported by the Air Force Office of Scientific Research under grant AFOSR-49620-82-C-0009. We thank Rob Abbott of the National Heart, Lung, and Blood Institute for suggesting the problem, David Ruppert for helpful discussions, and the referees for many useful comments on an earlier draft of this paper.

REFERENCES

AMEMIYA, Y. (1982). Estimators for the errors-in-variables model. Unpublished Ph.D. thesis, Iowa State University, Ames.

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120.

ARMSTRONG, B. (1984). Measurement error in the generalized linear model. To appear in *Comm. Statist.*

BERKSON, J. (1951). Why I prefer logits to probits. *Biometrics* **7** 327–339.

BILLINGSLEY, P. (1979). *Probability and Measure.* Wiley, New York.

CARROLL, R. J., SPIEGELMAN, C. H., LAN, K. K., BAILEY, K. T. and ABBOTT, R. D. (1984). On errors-in-variables for binary regression models. *Biometrika* **71** 19–25.

CHUNG, K. L. (1974). *A Course in Probability Theory.* Academic, New York,

CLARK, R. R. (1982). The errors-in-variables problem in the logistic regression model. Unpublished Ph.D. thesis, University of North Carolina, Chapel Hill.

COX, D. R. (1970). *Analysis of Binary Data.* Chapman and Hall, London.

DEMPSTER, A. P., SCHATZOFF, M. and WERMUTH, N. (1977). A simulation study of alternatives to ordinary least squares. *J. Amer. Statist. Assoc.* **72** 77–91.

EFRON, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *J. Amer. Statist. Assoc.* **70** 892–898.

FULLER, W. A. (1980). Properties of some estimators for the errors-in-variables model. *Ann. Statist.* **8** 407–422.

GLESER, L. R. (1981). Estimation in a multivariate "errors-in-variables" regression model: Large sample results. *Ann. Statist.* **9** 24–44.

GORDON, T. and KANNEL, W. E. (1968). *Introduction and General Background in the Framingham Study—The Framingham Study, Sections 1 and 2.* National Heart, Lung, and Blood Institute, Bethesda, Maryland.

HAUCK, W. W. (1983). A note on confidence bands for the logistic response curve. *Amer. Statist.* **37** 158–160.

LAHA, R. G. and ROHATGI, V. K. (1979). *Probability Theory.* Wiley, New York.

MADANSKY, A. (1959). The fitting of straight lines when both variables are subject to error. *J. Amer. Statist. Assoc.* **54** 173–205.

MCCULLAGH, P. and NELDER, J. A. (1983). *Generalized Linear Models.* Chapman and Hall, London.

MICHALIK, J. E. and TRIPATHI, R. C. (1980). The effect of errors in diagnosis and measurement on the estimation of the probability of an event. *J. Amer. Statist. Assoc.* **75** 713–721.

OLKIN, I. A., PETKAU, J. A. and ZIDEK, J. A. (1981). A comparison of n estimators for the binomial distribution. *J. Amer. Statist. Assoc.* **76** 637–642.

PREGIBON, D. (1981). Logistic regression diagnostics. *Ann. Statist.* **9** 705–724.

PRENTICE, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69** 331–42.

ROCKAFELLAR, R. T. (1970). *Convex Analysis.* Princeton University Press, Princeton.

STEFANSKI, L. A. (1983). Influence and measurement error in logistic regression. Institute of Statistics Mimeo Series No. 1548, University of North Carolina, Chapel Hill.

WOLTER, K. M. and FULLER, W. A. (1982a). Estimation of nonlinear errors-in-variables models. *Ann. Statist.* **10** 539–548.

WOLTER, K. M. and FULLER, W. A. (1982b). Estimation of the quadratic errors-in-variables model. *Biometrika* **69** 175–182.

DEPARTMENT OF ECONOMIC AND SOCIAL STATISTICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514