*The Annals of Statistics*
1992, Vol. 20, No. 4, 1929–1952

# EFFECTS OF MISMODELING ON TESTS OF ASSOCIATION BASED ON LOGISTIC REGRESSION MODELS[1]

By Melissa Dowd Begg and Stephen Lagakos

*Columbia University and Harvard University*

We consider the effects of misspecifying the covariate vector on tests of association between a particular covariate and the response variable in logistic regression models. Misspecification can include random measurement error, discretizing a continuous explanatory variable and mismodeling the functional form of an explanatory variable. The effects of misspecification are examined by deriving the asymptotic distribution of the test statistic under a sequence of models. Applications of this result are discussed.

**1. Introduction.** In using logistic regression models to assess the relationship between a binary response variable and one or more explanatory variables, misspecification of the explanatory variables is common and can adversely affect hypothesis tests of association between the response and exposure variables. The purpose of this paper is to derive an asymptotic result that can be used for studying the consequences of misspecification on the validity and efficiency of tests of association between exposure and response.

The literature contains many references to the effects of variable misspecification, but many of these, such as Cochran (1968) and Fuller (1980), discuss errors of measurement only for linear models. Several authors have begun to study the effects of mismodeling on parameter estimation in nonlinear regression. For example, Prentice (1982) studies variable misspecification for failure-time data. Stefanski (1985) and Armstrong (1985) examine the consequences of misspecification for the class of generalized linear models. Michalek and Tripathi (1980), Carroll, Spiegelman, Lan, Bailey and Abbott (1984), Stefanski and Carroll (1985) and Burr (1988) focus in particular on binary data. With regard to the consequences of mismodeling on the size and power characteristics of tests of association, Lagakos (1988a) and Lagakos and Schoenfeld (1984) look at the efficiency of tests of association when covariates have been mismodeled or omitted from the proportional hazards model for failure-time data. Tosteson and Tsiatis (1988) and Stefanski and Carroll (1990, 1991) consider the effects of using surrogate variables for exposure in the generalized linear model. Gail, Wieand and Piantadosi (1984) study how covariate omission in a randomized clinical trial biases the estimated treatment effect. In related work, Gail (1988) studies the effect of pooling in

Received January 1990; revised December 1991.
[1]Research supported by Grant CA39640 from the National Cancer Institute.
*AMS* 1980 *subject classifications*. Primary 62F05; secondary 62J02.
*Key words and phrases*. Asymptotic relative efficiency, asymptotic validity, logistic regression, model misspecification, tests of association.

perfectly balanced trials, and Gail, Tan and Piantadosi (1988) demonstrate via simulations that, under simple randomization, score tests from logistic models lose efficiency when needed covariates are omitted. Lagakos (1988b) explores the efficiency of tests of association in which a single exposure of interest is mismodeled in the absence of other covariates for linear, logistic and proportional hazards regression models. In this paper, we wish to expand upon these earlier results to study the consequences of simultaneously mismodeling both the exposure of interest and other covariates on the asymptotic validity and efficiency of tests of association between exposure and response in logistic regression. For the special case of omitted covariates, we refer the reader to Begg and Lagakos (1993).

Suppose that $Q(x, \mathbf{z})$ represents the likelihood score test statistic based on the correct exposure variable $\mathbf{x}$ and on the correct $(p \times 1)$ vector of covariates $\mathbf{z}$. Further, let $Q(x^*, \mathbf{z}^*)$ denote a test statistic which is identical in form to $Q(x, \mathbf{z})$, but in which $x$ and $\mathbf{z}$ are replaced by $x^*$ and $\mathbf{z}^*$, respectively. Here, $x^*$ denotes a misspecified version of $x$, and $\mathbf{z}^*$ denotes a misspecified version of $\mathbf{z}$. By misspecification, we mean simple measurement error, mismodeling the functional form of an explanatory variable [e.g., using age instead of log(age)] or discretizing a continuous variable. The only restriction imposed on misspecification is that the distribution of the response variable given $(x, x^*, \mathbf{z}, \mathbf{z}^*)$ must be equivalent to the distribution of response given $x$ and $\mathbf{z}$. Loosely speaking, this means that $x^*$ and $\mathbf{z}^*$ do not add to the information about outcome already furnished by $x$ and $\mathbf{z}$. The statistic $Q(x^*, \mathbf{z}^*)$ does not, in general, have the same distributional properties as $Q(x, \mathbf{z})$. Its properties depend on the type and the extent of mismodeling that has occurred. The goal of this paper is to develop a convenient way to assess the consequences of using $Q(x^*, \mathbf{z}^*)$ instead of $Q(x, \mathbf{z})$.

There are various ways of assessing the ramifications of using $Q(x^*, \mathbf{z}^*)$. We will focus on the asymptotic distributions of $Q(x, \mathbf{z})$ and $Q(x^*, \mathbf{z}^*)$ because their exact distributions are, in general, intractable. Under a specific sequence of models, we show that as the sample size goes to infinity $Q(x, \mathbf{z})$ converges in distribution to $N(\mu, 1)$ random variable, while $Q(x^*, \mathbf{z}^*)$ converges to $N(\mu^*, 1)$. Since $\mu = 0$ under $H_0$, we can assess the asymptotic validity of $Q(x^*, \mathbf{z}^*)$ by studying $\mu^*$ when the null hypothesis holds. By comparing the magnitudes of $\mu$ and $\mu^*$ for alternatives to $H_0$, we can assess asymptotic relative efficiency (under conditions when both tests are valid).

We state the theoretical problem and main results in Sections 2 and 3. In Section 4 we discuss its applications to assessing the effects of misspecification of both the covariate of interest and other covariates on tests of association. In Section 5 we present an example which applies the theoretical result to a problem in AIDS research. Proofs of the main results are provided in Section 6.

**2. Model and test statistic.** Suppose that the response variable $Y$ is related to covariates $x$ and $\mathbf{z}$ by the logistic model

$$\text{logit}\,\Pr(Y = 1 | x, \mathbf{z}) = \theta_0 + \alpha x + \beta_0' \mathbf{z}, \tag{2.1}$$

where $\theta_0$ and $\alpha$ are unknown scalar parameters and $\beta_0$ is an unknown $(p \times 1)$ vector parameter. We focus on the hypothesis of no association between $x$ and $Y$, that is, $H_0$: $\alpha = 0$. Given $n$ independent observations of the form $(Y_{ni}, x_{ni}, \mathbf{z}_{ni})$, $i = 1, 2, \dots, n$, $H_0$: $\alpha = 0$ can be assessed by one of three asymptotically equivalent tests based on the maximum likelihood estimator (MLE) of $\alpha$, the likelihood ratio statistic or the likelihood score statistic [Cox and Hinkley (1974)]. We will focus on the score test, which has the form

$$Q(x, \mathbf{z}) = \sum_{i=1}^{n} x_{ni} \left( Y_{ni} - \frac{\exp(\hat{\theta} + \hat{\beta}'\mathbf{z}_{ni})}{1 + \exp(\hat{\theta} + \hat{\beta}'\mathbf{z}_{ni})} \right) \Bigg/ \sqrt{w}, \tag{2.2}$$

where $\hat{\theta}$ and $\hat{\beta}$ are the restricted MLEs of the nuisance parameters $\theta_0$ and $\beta_0$ when $\alpha = 0$, and $w$ is the reciprocal of the (1, 1) element from the inverse sample information matrix for $(\alpha, \theta_0, \beta_0)$ [Cox and Hinkley (1974)]. For contiguous alternatives to $\alpha = 0$, the limiting distribution of $Q(x, \mathbf{z})$ can be shown [Cox and Hinkley (1974)] to be $N(\mu, 1)$, where $\mu$ depends on $\theta_0$, $\alpha$, $\beta_0$ and the joint distribution of $x$ and $\mathbf{z}$. In particular, $\mu = 0$ whenever $\alpha = 0$.

Now suppose that $x$ and $\mathbf{z}$ are misspecified and consider the test, say, $Q(x^*, \mathbf{z}^*)$, obtained by replacing $x$ and $\mathbf{z}$ in $Q$ by their misspecified versions, say, $x^*$ and $\mathbf{z}^*$. The test statistic now takes the following form:

$$Q(x^*, \mathbf{z}^*) = \sum_{i=1}^{n} x_{ni}^* \left( Y_{ni} - \frac{\exp(\tilde{\theta} + \tilde{\beta}'\mathbf{z}_{ni}^*)}{1 + \exp(\tilde{\theta} + \tilde{\beta}'\mathbf{z}_{ni}^*)} \right) \Bigg/ \sqrt{w^*}, \tag{2.3}$$

where $w^*$ is identical in form to $w$ in (2.2), but with $x$ and $\mathbf{z}$ replaced by $x^*$ and $\mathbf{z}^*$, and where $\tilde{\theta}$ and $\tilde{\beta}$ denote the maximizing solutions to estimating equations based on the misspecified versions of $x$ and $\mathbf{z}$. For a particular sequence of contiguous alternatives to model (2.1), we derive that

$$Q(x^*, \mathbf{z}^*) \to_L N(\mu^*, 1),$$

where $\mu^*$ depends on $\theta_0$, $\alpha$, $\beta_0$ and the joint distribution of $x$, $x^*$, $\mathbf{z}$ and $\mathbf{z}^*$.

## 3. Asymptotic distribution of $Q(x^*, \mathbf{z}^*)$.

3.1. *Sequence of models.* We derive the asymptotic distribution of $Q(x^*, \mathbf{z}^*)$ by considering a sequence of contiguous models indexed by $n$. Suppose that $\mathbf{z}_{ni}^*$, the misspecified version of $\mathbf{z}_{ni}$, is equal to $\mathbf{z}_{ni}$ plus a term of order $O_p(n^{-1/2})$, say $n^{-1/2}\mathbf{V}_{ni}$. Thus $\mathbf{V}_{ni}$ is a random variable whose relationship to $\mathbf{z}_{ni}^*$ is described by

$$\mathbf{z}_{ni}^* = \mathbf{z}_{ni} + n^{-1/2}\mathbf{V}_{ni}. \tag{3.1}$$

For a given $n$, assume that $(Y_{ni}, x_{ni}, \mathbf{z}_{ni}, x_{ni}^*, \mathbf{V}_{ni})$, $i = 1, 2, \dots, n$, are independent and identically distributed (iid), where $(x_{ni}, \mathbf{z}_{ni}, x_{ni}^*, \mathbf{V}_{ni})$ has cumulative distribution function $F$ that is independent of $n$. Furthermore, suppose that the conditional distribution of $Y_{ni}$ given $(x_{ni}, \mathbf{z}_{ni}, x^*_{ni}, \mathbf{V}_{ni})$ is given by

$$\begin{aligned}\operatorname{logit}\{\Pr(Y_{ni} = 1|x_{ni}, \mathbf{z}_{ni}, x^*_{ni}, \mathbf{V}_{ni})\} &= \operatorname{logit}\{\Pr(Y_{ni} = 1|x_{ni}, \mathbf{z}_{ni})\} \\ &= \theta_0 + \frac{\alpha_0}{\sqrt{n}} x_{ni} + \beta_0' \mathbf{z}_{ni}.\end{aligned}$$

The first part of this equation implies that $x^*_{ni}$ and $\mathbf{V}_{ni}$ add no information about $Y_{ni}$ beyond that supplied by $x_{ni}$ and $\mathbf{z}_{ni}$. The replacement of $\alpha$ in (2.1) by $\alpha_0/\sqrt{n}$ is the usual form for the contiguous alternative to the hypothesis of interest. The assumption in (3.1), similar in spirit and purpose to replacing $\alpha$ by $\alpha_0/\sqrt{n}$, is intended to allow the misspecification of $\mathbf{z}$ to affect the test of $H_0$ but still give a tractable result. As we see later, this rate of convergence of $\mathbf{z}^*_{ni}$ to $\mathbf{z}_{ni}$ ensures this. Had we instead let $\mathbf{z}^*_{ni}$ and $\beta$ be fixed, then the distribution of the test statistic would, in general, move toward infinity and, therefore, not converge to a proper distribution as $n \to \infty$. Had we let $\mathbf{z}^*_{ni}$ be fixed and let $\beta = O(n^{-1/2})$, the test statistic would have a stable limiting distribution, but the result would be of no value because it essentially would be assuming that the covariate was not important. However, by allowing $\beta$ to be fixed and $\mathbf{z}^*_{ni}$ to converge to $\mathbf{z}_{ni}$, we get a limiting distribution for the test statistic which is a proper distribution and which reflects the misspecification of $\mathbf{z}_{ni}$. It is important to note that the assumption of local alternatives has no direct physical significance; it is employed merely to guarantee tractability [Cox and Hinkley (1974)]. Finally, although the joint distribution $F$ can be arbitrary, we assume that $E[x]$, $E[x\mathbf{V}]$, $E[x^*\mathbf{V}]$, $E[x^{*2}\mathbf{V}]$, $E[xx^*\mathbf{z}]$, $E[xx^*\mathbf{V}]$, $E[\mathbf{V}\mathbf{V}']$, $E[x^*\mathbf{z}\mathbf{z}']$, $E[x'\mathbf{V}\mathbf{V}']$ and $E[x^*\mathbf{V}\mathbf{z}']$ are finite.

3.2. *Main results.* For the specified sequence of contiguous alternative models described in Section 3.1, it can be shown that $\tilde{\theta}$ and $\tilde{\beta}$ are strongly consistent for $\theta_0$ and $\beta_0$ and are asymptotically multivariate normal.

THEOREM 3.1. *For the sequence of models described in Section* 3.1,

$$\left(\tilde{\theta}, \tilde{\beta}\right) \to (\theta_0, \beta_0) \quad a.s.$$

THEOREM 3.2. *For the sequence of models described in Section* 3.1,

$$\sqrt{n}\left\{\begin{bmatrix}\tilde{\theta} \\ \tilde{\beta}\end{bmatrix} - \begin{bmatrix}\theta_0 \\ \beta_0\end{bmatrix}\right\} \to_L N_{(p+1)}\left(i_{22}^{-1}\mathbf{m}, i_{22}^{-1}\right) \quad as\ n \to \infty,$$

*where*

$$\mathbf{m} = E\left\{\begin{bmatrix}1 \\ \mathbf{z}\end{bmatrix}(\alpha_0 x - \beta_0'\mathbf{V})\frac{\exp(\theta_0 + \beta_0'\mathbf{z})}{[1 + \exp(\theta_0 + \beta_0'\mathbf{z})]^2}\right\} \tag{3.2}$$

*and*

$$i_{22} = E\left\{\begin{bmatrix} 1 & \mathbf{z}' \\ \mathbf{z} & \mathbf{z}\mathbf{z}' \end{bmatrix} \frac{\exp(\theta_0 + \beta_0'\mathbf{z})}{[1 + \exp(\theta_0 + \beta_0'\mathbf{z})]^2}\right\}.$$

Proofs of these theorems can be found in Section 6.

It can also be shown (Theorem 3.3) that the misspecified test statistic $Q(x^*, \mathbf{z}^*)$ has an asymptotic normal distribution.

THEOREM 3.3. *For the sequence of models specified in Section* 3.1,

$$Q(x^*, \mathbf{z}^*) \to_L N(\mu^*, 1),$$

*where*

$$\mu^* = \delta^* / \sqrt{\xi^*}$$

*and*

$$\delta^* = c - \mathbf{m}' i_{22}^{-1} \mathbf{i}_{21}^*, \qquad \xi^* = i_{11}^* - \mathbf{i}_{12}^* i_{22}^{-1} \mathbf{i}_{21}^*,$$

*where*

$$c = E\left\{x^*(\alpha_0 x - \beta_0'\mathbf{V}) \frac{\exp(\theta_0 + \beta_0'\mathbf{z})}{[1 + \exp(\theta_0 + \beta_0'\mathbf{z})]^2}\right\},$$

$$\text{(3.3)} \qquad i_{11}^* = E\left\{x^{*2} \frac{\exp(\theta_0 + \beta_0'\mathbf{z})}{[1 + \exp(\theta_0 + \beta_0'\mathbf{z})]^2}\right\},$$

$$\mathbf{i}_{21}^* = \mathbf{i}_{12}^{*\prime} = E\left\{\begin{bmatrix} 1 \\ \mathbf{z} \end{bmatrix} x^* \frac{\exp(\theta_0 + \beta_0'\mathbf{z})}{[1 + \exp(\theta_0 + \beta_0'\mathbf{z})]^2}\right\},$$

*and where* $\mathbf{m}$ *and* $i_{22}$ *are defined in* (3.2).

The test statistic $Q(x^*, \mathbf{z}^*)$ is invariant to scale or location changes in $x$, $x^*$, $\mathbf{z}$ and $\mathbf{z}^*$. Similarly, it can be shown that the expression $\mu^*$ is invariant to scale or location changes to $x$, $x^*$, $\mathbf{z}$ and $\mathbf{z}^*$. For a proof of Theorem 3.3, see Section 6.

**4. Discussion of applications.** Given the joint moments of the random variables $x$, $x^*$, $\mathbf{z}$ and $\mathbf{z}^*$, Theorem 3.3 gives the limiting distribution of the misspecified test statistic $Q(x^*, \mathbf{z}^*)$. The limiting distribution of the "correct" test, $Q(x, \mathbf{z})$, can be obtained by setting $x = x^*$ and $\mathbf{z} = \mathbf{z}^*$ in the general result. Similarly, the asymptotic distributions of $Q(x^*, \mathbf{z})$ and $Q(x, \mathbf{z}^*)$ follow as special cases of Theorem 3.3. In general, the main result has no closed-form solution, but is readily evaluable by numerical techniques. Although expressions for asymptotic relative efficiencies can be complicated in general, they simplify in several special cases.

When there are no covariates $\mathbf{z}$ in the model, the misspecified test is always valid, since $\mu^* = 0$ whenever $\alpha = 0$. For this special case, it is easily shown that the asymptotic relative efficiency of $Q(x^*)$ to $Q(x)$ simplifies to the squared correlation of $x$ and $x^*$, say, $\rho^2(x, x^*)$, which is the same as the result derived by Lagakos (1988b) and Tosteson and Tsiatis (1988). This simple expression can be numerically evaluated under a wide variety of conditions, giving rise to guidelines for test selection and design. For example, Lagakos (1988b) evaluates this expression to study the effects of choosing the wrong dose metameter in a test for trend, mismodeling the functional form of a continuous exposure variable, discretizing a measured exposure, misclassifying a categorical exposure and mismeasuring a continuous exposure variable.

When correctly specified covariates are present ($\mathbf{z}^* = \mathbf{z}$), it follows from Theorem 3.3 that $\mu^* = 0$ and hence $Q(x^*, \mathbf{z})$ is asymptotically valid, although inefficient. When $\mathbf{z}$ is independent of $x$ and $x^*$ (as in a randomized clinical trial), the formula for asymptotic relative efficiency reduces to $\rho^2(x, x^*)$, just as when there were no covariates in the model. Hence the reduction in efficiency caused by misspecifying $x$ in the absence of covariates is the same as the reduction in efficiency in the presence of correctly specified covariates.

When both the exposure and covariates are misspecified simultaneously, the asymptotic size of $Q(x^*, \mathbf{z}^*)$ can be distorted. Asymptotic validity is preserved, however, in certain special cases, such as when $x$ and $x^*$ are independent of $\mathbf{z}$ and $\mathbf{V}$, when $\beta_0 = \mathbf{0}$ or when less intuitive, algebraic conditions hold, for example, when

$$E\left[x^* \mathbf{V} \frac{\exp(\theta_0 + \beta_0' \mathbf{z})}{[1 + \exp(\theta_0 + \beta_0' \mathbf{z})]^2}\right], \qquad E\left[x^* \mathbf{z} \frac{\exp(\theta_0 + \beta_0' \mathbf{z})}{[1 + \exp(\theta_0 + \beta_0' \mathbf{z})]^2}\right]$$

and

$$E\left[x^* \frac{\exp(\theta_0 + \beta_0' \mathbf{z})}{[1 + \exp(\theta_0 + \beta_0' \mathbf{z})]^2}\right]$$

are all equal to zero.

For the special case when $x$ and $x^*$ are independent of $\mathbf{z}$ and $\mathbf{V}$, the asymptotic relative efficiency of $Q(x^*, \mathbf{z}^*)$ to $Q(x, \mathbf{z})$ simplifies to $\rho^2(x, x^*)$, just as when there are no covariates or when there are correctly specified covariates in the model. This implies that misspecification of covariates that are independent of exposure does not further reduce efficiency. We find this result counterintuitive. For the special case of independence, a slower rate of convergence in (3.1) or a different formulation of the problem may be needed to give a result that is more relevant to finite-sample conditions.

Similarly, this paper's main result, although valid, is not useful in assessing the effects of omitted covariates; when $\mathbf{z}^* = \mathbf{0}$, the assumption that $(\mathbf{z}^* - \mathbf{z})$ is $O_p(n^{-1/2})$ forces covariate effects to be negligible in the limit. Under these circumstances, an asymptotic approach that deals specifically with omitted covariates provides a more meaningful result. Results in a separate paper [Begg and Lagakos (1993)] show that, for the special case of omitted covariates when $\mathbf{z}$ is independent of $(x, x^*)$, asymptotic test size is retained. However, a loss in efficiency results. The manuscript provides a formula for the asymptotic relative efficiency and evaluates the formula for several examples. Because of the independence of $(x, x^*)$ and $\mathbf{z}$ and because of the simple form of $Q(x^*, \mathbf{0})$, we were able to derive this formula without assuming contiguous alternatives for $\mathbf{z}$. Hence the results in Begg and Lagakos (1991) are quite distinct from the results given here, as well as distinct from related work by Gail, Tan and Piantadosi (1988). These authors considered the problem of omitting relevant covariates in a randomized trial. They show that when $\mathbf{z}$ is omitted ($\mathbf{z}^* = \mathbf{0}$), $Q(x, \mathbf{0})$ retains nominal size. When there is partial adjustment for covariates (i.e., when only a subset of the important covariates is included), the authors find only a minute distortion of test size and conclude that "for practical purposes, logistic regression yields tests of $H_0$ with nominal size, whether no covariates or a few covariates are included in the model." Because the authors assume independence and restrict attention to the case of covariate omission (not general misspecification), they did not have to appeal to any assumptions like condition (3.1) to reach a tractable result. Thus the results in Begg and Lagakos (1991) verify the findings of Gail, Tan and Piantadosi (1988) about test size and extend these findings by providing a formula for asymptotic relative efficiency for the special case of omitted covariates under independence.

The utility of the main result depends on how well the asymptotics approximate what happens in finite samples. The asymptotic results in this paper have been shown to reflect accurately small-sample results for at least one application. We have applied Theorem 3.3 to a problem in animal carcinogencity experiments [Begg and Lagakos (1991)]. Very briefly, because of problems with treatment lethality, tumor lethality and unobservable time to tumor, the test procedures used in this setting often make strong assumptions about the mechanisms underlying tumor development and death [see Lagakos (1982), McKnight and Crowley (1984), Lagakos and Ryan (1985) and their references]. We compare the limiting behaviors of various carcinogenicity tests via asymptotic results such as Theorem 3.3 and the results of Schoenfeld (1981). There is some overlap between the cases we consider and earlier work on the same problem, which relied on simulation results to compare the various test procedures [Bailer and Portier (1988)]. Preliminary findings indicate that our asymptotic results are very close to the results obtained by simulation. Thus, there is some indication that the asymptotic formula can reflect what actually happens in finite samples.

## 5. An illustrative application.

5.1. *General setting.* In general, the formula for the mean ($\mu^*$) of the misspecified test statistic is complex, but can be evaluated numerically to assess the consequences of misspecification in a particular setting. To demonstrate this point, we have applied Theorem 3.3 to an example arising from a phase-3 AIDS clinical trial comparing two treatments: zidovudine (AZT) and dideoxyinosine (ddI). Entry criteria required that all subjects be diagnosed with clinical AIDS and that all had previously taken AZT for at least 4 months, but less than 2 years. After randomization to one of the two treatment arms, each subject was followed to see whether there was a response in CD4 cell count.

The results of such a trial can be analyzed via a logistic model, with a binary treatment indicator (T) as the exposure of interest ($T = 1$ if AZT, $T = 0$ if ddI). However, a comparison of the two treatments is complicated by the fact that prior exposure to AZT may introduce a resistance to its effects among those randomized to AZT in the current trial. Consequently, even if the two treatments were equally effective, AZT may look worse than ddI simply because patients have developed a resistance to AZT. Similarly, if AZT is truly better than ddI, resistance to AZT will diminish the treatment difference; or if ddI is better than AZT, resistance to AZT will exaggerate this difference in treatments.

In an effort to control for the possibility of resistance to AZT, we can adjust for duration of prior exposure to AZT (call this D) in the analysis. Hence the logistic model fitted can include variables for treatment effect ($T$), duration effect ($D$), and a treatment–duration interaction term ($TD$). The inclusion of an interaction term allows for a differential effect of length of prior exposure to AZT on the AZT and ddI arms of the study. The logistic model can be cast as follows:

$$\text{logit}\,\Pr(Y = 1|T, D) = \theta_0 + \beta_{01}T + \beta_{02}D + \alpha_0 TD, \tag{5.1}$$

in which the hypothesis $H_0$: $\alpha_0 = 0$ represents the null hypothesis of no treatment–duration interaction effect (i.e., no resistance to AZT).

Theorem 3.3 allows us to assess the consequences of misspecifying prior duration of AZT exposure on the asymptotic validity and efficiency of tests of the treatment–duration interaction term. To shape this problem in terms of variables defined earlier, set $x$, the exposure of interest, equal to the interaction term $TD$, and set the vector of covariates $\mathbf{z}$ equal to vector $[T, D]'$. The misspecified version of $x$ ($x^*$) is equal to $TD^*$, and $\mathbf{z}^*$ (the misspecified version of $\mathbf{z}$) is equal to $[T, D^*]'$. Finally, by (3.1), $\mathbf{V}$ is equal to $n^{1/2}[0, D^* - D]'$. Having fashioned the AIDS clinical trial problem in terms of Theorem 3.3 [i.e., in terms of the variables appearing in equations (3.3)], we must specify the sample size $n$, values for each of the unknown parameters $(\alpha_0, \theta_0, \beta_{01}, \beta_{02})$ and the joint distribution of random variables $T$ and $D^*$.

All calculations assume a total sample size of 100 subjects. Since this is a randomized trial, $T$ is assumed to be independent of $D$ and to follow a Bernoulli distribution with parameter 0.5. Judging by actual data from the trial, the distribution of duration is roughly uniform from four months to two years. The choices of the unknown parameters depend on the type of misspecification considered. For each type of misspecification, $D^*$ is equal to the number of years exposed, while the correct parameterization $D$ is some function of $D^*$. The four functions considered are $D = \ln D^*$, $D = \exp D^*$, $D = (D^*)^{1/2}$ and $D = (D^*)^2$.

TABLE 1

*Choices of $\beta_{02}$ for four different types of misspecification. In the following, $\beta_{02}$ was computed by taking the ratio of the true duration* ln(*odds ratio*) *to* $[f(2) - f(1/3)]$, *where f is one of the four functions listed*

| True OR ($D^* = 2 : D^* = \frac{1}{3}$) | Type of misspecification $D = \ln D^*$ | $D = \exp D^*$ | $D = (D^*)^{1/2}$ | $D = (D^*)^2$ |
|---|---|---|---|---|
| $\frac{1}{2}$ | −0.3868 | −0.1157 | −0.8282 | −0.1782 |
| $\frac{1}{3}$ | −0.6131 | −0.1833 | −1.3127 | −0.2825 |
| $\frac{1}{5}$ | −0.8982 | −0.2685 | −1.9231 | −0.4139 |
| $\frac{1}{10}$ | −1.2851 | −0.3842 | −2.7513 | −0.5921 |

5.2. *Asymptotic validity.* In assessing the effects of model misspecification on test size, parameter $\alpha_0$ was set to zero. Parameter $\theta_0$, the constant term, was allowed to take on values in the set $\{-1, 0, 1\}$, translating to "baseline" probabilities of outcome of 27, 50 and 73%. Parameter $\beta_{01}$, representing the effect of treatment when $D = 0$, could assume values $\{-\ln 2, 0, \ln 2\}$, giving treatment odds ratios of 0.5, 1.0 and 2.0. The coefficient of duration, $\beta_{02}$, was chosen to make the effects of mismodeling across the four different types of misspecification comparable, as we now explain. Under $H_0$: $\alpha_0 = 0$, the model states that

$$\operatorname{logit} \Pr(Y = 1 | T, D) = \theta_0 + \beta_{01} T + \beta_{02} D = \theta_0 + \beta_{01} T + \beta_{02} f(D^*),$$

since $D$ is a function of $D^*$. Hence the true duration ln(odds ratio) (ln OR) for a subject with $D^* = 2$ versus a subject with $D^* = 1/3$ is given by

$$\ln \mathrm{OR} = \beta_{02}\left[ f(2) - f\left(\tfrac{1}{3}\right)\right].$$

Thus the actual values of $\beta_{02}$ vary from one type of mismodeling to another; but all represent the same true duration odds ratios: {0.5, 0.33, 0.2, 0.1} (see Table 1).

Having chosen sample size, the distributions of $T$ and $D^*$, and ranges of plausible values for the unknown parameters, $\mu^*$ was numerically computed using equations (3.3). Actual test size was calculated assuming a two-sided test with nominal significance level 0.05. We note that because we consider two-sided tests which assume $\mu^* = 0$, a shift away from zero causes test size to be inflated (i.e., tests become anticonservative). Had we taken a one-sided approach, we would observe both increases and decreases in test size, giving tests that are both conservative and anticonservative. In addition, these distortions would be more extreme than those observed here. Table 2 gives the values of actual test size for different combinations of parameter values and for four different types of mismodeling of $D^*$. Note that we present values only for $\theta_0 = 0$, since varying this value did not qualitatively alter the results. As can be seen in the table entries, distortion of test size is almost negligible. We conclude from these results that distortion of two-sided test size is not a serious problem in this particular setting.

TABLE 2

*The effect of misspecification of duration of prior AZT use on asymptotic validity. Table entries give actual test size. In the following, $\theta_0$ represents the constant term, $\beta_{01}$ represents the coefficient of the binary treatment indicator and $\beta_{02}$ (the coefficient of duration) is chosen such that the true duration odds ratio of* 2 *years prior AZT use versus* 4 *months prior AZT use takes on values* 1/2, 1/3, 1/5 *and* 1/10. *Calculations assume a sample size of* 100 *and a nominal test size of* 0.05.

| | | **Type of misspecification** | | | |
|---|---|---|---|---|---|
| $\beta_{01}$ | **True OR** | $D = \ln D^*$ | $D = \exp D^*$ | $D = (D^*)^{1/2}$ | $D = (D^*)^2$ |
| $-\ln 2$ | $\frac{1}{2}$ | 0.0501 | 0.0500 | 0.0500 | 0.0500 |
| | $\frac{1}{3}$ | 0.0503 | 0.0500 | 0.0500 | 0.0501 |
| | $\frac{1}{5}$ | 0.0504 | 0.0501 | 0.0500 | 0.0501 |
| | $\frac{1}{10}$ | 0.0506 | 0.0501 | 0.0500 | 0.0502 |
| 0 | $\frac{1}{2}$ | 0.0502 | 0.0500 | 0.0500 | 0.0500 |
| | $\frac{1}{3}$ | 0.0505 | 0.0500 | 0.0501 | 0.0500 |
| | $\frac{1}{5}$ | 0.0510 | 0.0501 | 0.0501 | 0.0501 |
| | $\frac{1}{10}$ | 0.0520 | 0.0501 | 0.0501 | 0.0502 |
| $\ln 2$ | $\frac{1}{2}$ | 0.0502 | 0.0500 | 0.0501 | 0.0500 |
| | $\frac{1}{3}$ | 0.0506 | 0.0500 | 0.0503 | 0.0500 |
| | $\frac{1}{5}$ | 0.0515 | 0.0500 | 0.0504 | 0.0501 |
| | $\frac{1}{10}$ | 0.0532 | 0.0501 | 0.0505 | 0.0501 |

5.3. *Asymptotic relative efficiency.* To assess the asymptotic efficiency of the misspecified test statistic relative to the ideal test, we set $\beta_{02}$ equal to zero to ensure that both tests being compared were valid. We allowed $\theta_0$, the constant term, to take on values $\{-1, 0, 1\}$, and $\beta_{01}$, the coefficient of treatment, to assume values $\{-\ln 2, 0, \ln 2\}$, just as before. The parameter of interest, $\alpha_0$, is calculated in a way similar to the way we computed $\beta_{02}$ for test size. We know that the treatment ln(OR) ($T = 1$ versus $T = 0$), based on (5.1), is given by

$$\ln \mathrm{OR} = (\beta_{01} + \alpha_0 D) = (\beta_{01} + \alpha_0 f(D^*)).$$

Hence, given values for ln(OR) [2, 3, 5, 10] and $\beta_{01}$, we can solve for $\alpha_0$ by taking $D^* = 2$ to assess the interaction effect at the maximum duration of prior exposure. When $\beta_{02} = 0$, however, $\alpha_0$ cancels out of the ratio of $\mu^*$ to $\mu$; hence choice of $\alpha_0$ does not affect calculation of the asymptotic relative efficiency.

Efficiency calculations are presented in Table 3. As with test size, we see that misspecification of duration does not greatly affect asymptotic relative efficiency. At most, the misspecified test loses about 10% efficiency relative to the ideal test. In planning such a trial, then one would only need to increase the proposed sample size by about 11% at most to compensate for any loss in efficiency due to the misspecification of duration. We observe that the consequences of misspecification are somewhat greater for the transcendental

TABLE 3

*Asymptotic relative efficiency of the misspecified test statistic relative to the ideal test when there is no main effect of prior duration* ($\beta_{02} = 0$). *Calculations assume* 100 *subjects per treatment group. Note that the asymptotic relative efficiencies do not depend on* $\alpha_0$, *the coefficient of the treatment–duration interaction term*

| | | Type of misspecification | | | |
|---|---|---|---|---|---|
| $\theta_0$ | $\beta_{01}$ | $D = \ln D^*$ | $D = \exp D^*$ | $D = (D^*)^{1/2}$ | $D = (D^*)^2$ |
| −1 | $-\ln 2$ | 0.9237 | 0.9296 | 0.9815 | 0.9463 |
| | 0 | 0.9098 | 0.9167 | 0.9779 | 0.9363 |
| | $\ln 2$ | 0.9000 | 0.9076 | 0.9753 | 0.9291 |
| 0 | $-\ln 2$ | 0.9144 | 0.9210 | 0.9791 | 0.9396 |
| | 0 | 0.9098 | 0.9167 | 0.9779 | 0.9363 |
| | $\ln 2$ | 0.9144 | 0.9210 | 0.9791 | 0.9396 |
| 1 | $-\ln 2$ | 0.9000 | 0.9076 | 0.9753 | 0.9291 |
| | 0 | 0.9098 | 0.9167 | 0.9779 | 0.9363 |
| | $\ln 2$ | 0.9237 | 0.9296 | 0.9815 | 0.9463 |

functions than for the power functions. This is not surprising, since the transcendental functions represent a greater departure from a linear relationship (in the logit scale) than do the power functions. In sum, there appears to be little loss in efficiency due to misspecifying duration. We note, however, that for all of the situations considered, both tests suffer from very low power for detecting a treatment–duration effect. Rejection probabilities for different values of $\alpha_0$ range from 5 to 8%. These probabilities increase if one considers larger values of $n$ or $\alpha_0$. This result emphasizes that either very large effect sizes or very large sample sizes are needed to achieve reasonable power for detecting interaction effects, as has been noted by others [Smith and Day (1984)].

In developing this example, we considered the possibility that the minimal distortion of test size and small losses in efficiency may be due to condition (3.1) (the $\mathbf{z}^*$ assumption), rather than to the choice of example. However, the formula for $\mu^*$ [see equations (3.3)] clearly indicates that the difference between the mean of $Q(x, \mathbf{z})$ and the mean of $Q(x^*, \mathbf{z}^*)$ (and, therefore, asymptotic validity and efficiency) depends on the size of the covariate effect ($\beta_0$) and the degree of misspecification of $\mathbf{z}$ (as represented by $\mathbf{V} = n^{1/2}[\mathbf{z}^* - \mathbf{z}]$). There does not appear to be any upper limit on the distortion of $\mu^*$; in fact, the distortion may be substantial. We conclude from this that the amount of distortion possible depends on the particular example chosen.

**6. Proofs of theorems.** Theorem 3.1 states that estimators $\tilde{\theta}$ and $\tilde{\beta}$ are strongly consistent for $\theta_0$ and $\beta_0$. Theorem 3.2 claims that vector $[\tilde{\theta}, \tilde{\beta}]'$ has a limiting multivariate normal distribution. These two results are needed to prove Theorem 3.3 [the asymptotic normality of $Q(x^*, \mathbf{z}^*)$].

6.1. *Proof of Theorem* 3.1. Define

$$L^*(\theta, \beta) = \prod_{i=1}^{n} \frac{\exp(Y_{ni}(\theta + \beta' \mathbf{z}^*_{ni}))}{1 + \exp(\theta + \beta' \mathbf{z}^*_{ni})},$$

and note that this would be the correct likelihood under $H_0$ if $\mathbf{z}^*_{ni} = \mathbf{z}_{ni}$. Also, let

$$\mathbf{W}^*(\theta, \beta) = \sum_{i=1}^{n} \begin{bmatrix} 1 \\ \mathbf{z}^*_{ni} \end{bmatrix} \left( Y_{ni} - \frac{\exp(\theta + \beta' \mathbf{z}^*_{ni})}{1 + \exp(\theta + \beta' \mathbf{z}^*_{ni})} \right). \tag{6.1}$$

Define the estimators $\tilde{\theta}$ and $\tilde{\beta}$ to be the values of $\theta$ and $\beta$ which maximize $L^*(\theta, \beta)$. The proof of Theorem 3.1 begins with the following strong law.

LEMMA 6.1. *$n^{-1}\mathbf{W}^*(\theta, \beta) \to \mathbf{A}_0(\theta, \beta)$ a.s., where*

$$\mathbf{A}_0(\theta, \beta) = E\left[ \begin{bmatrix} 1 \\ \mathbf{z} \end{bmatrix} \left( \frac{\exp(\theta_0 + \beta_0' \mathbf{z})}{1 + \exp(\theta_0 + \beta_0' \mathbf{z})} - \frac{\exp(\theta + \beta' \mathbf{z})}{1 + \exp(\theta + \beta' \mathbf{z})} \right) \right]. \tag{6.2}$$

PROOF. Note that

$$\begin{aligned} \frac{1}{n}\mathbf{W}^*(\theta, \beta) &= \frac{1}{n} \sum_{i=1}^{n} \begin{bmatrix} 1 \\ \mathbf{z}_{ni} \end{bmatrix} \left( Y_{ni} - \frac{\exp(\theta + \beta' \mathbf{z}^*_{ni})}{1 + \exp(\theta + \beta' \mathbf{z}^*_{ni})} \right) \\ &\quad + \frac{1}{n^{3/2}} \sum_{i=1}^{n} \begin{bmatrix} 0 \\ \mathbf{V}_{ni} \end{bmatrix} \left( Y_{ni} - \frac{\exp(\theta + \beta' \mathbf{z}^*_{ni})}{1 + \exp(\theta + \beta' \mathbf{z}^*_{ni})} \right) \\ &= \mathbf{M}_n + \mathbf{N}_n. \end{aligned}$$

We first show that $\mathbf{N}_n \to \mathbf{0}$ a.s. For any vector $\mathbf{V}$, let $\langle \mathbf{V} \rangle$ be the vector whose elements are the absolute values of the elements of $\mathbf{V}$. Furthermore, define $\mathbf{V}_1 \le \mathbf{V}_2$ to mean that each element of $\mathbf{V}_1$ is less than or equal to the corresponding element of $\mathbf{V}_2$.

Clearly,

$$\langle \mathbf{N}_n \rangle \le \frac{1}{n^{3/2}} \sum_{i=1}^{n} \langle \mathbf{V}_{ni} \rangle \left| Y_{ni} - \frac{\exp(\theta + \beta' \mathbf{z}^*_{ni})}{1 + \exp(\theta + \beta' \mathbf{z}^*_{ni})} \right|.$$

Since the second term in absolute value on the right-hand side falls between 0 and 1,

$$\langle \mathbf{N}_n \rangle \le \frac{1}{n^{3/2}} \sum_{i=1}^{n} \langle \mathbf{V}_{ni} \rangle.$$

Thus, since the $\mathbf{V}_{ni}$'s are independent and identically distributed (iid) random variables with finite first moments, $n^{-3/2}\Sigma_i \mathbf{V}_{ni} \to \mathbf{0}$ a.s. It follows that $\mathbf{N}_n \to \mathbf{0}$ a.s.

Thus $n^{-1}\mathbf{W}^*(\theta, \beta)$ has the same almost sure limit as $\mathbf{M}_n$. Adding and subtracting the expected value of $Y_{ni}$ given $x_{ni}$ and $\mathbf{z}_{ni}$ within the parentheses in $\mathbf{M}_n$, we can write $\mathbf{M}_n = \mathbf{M}_{1n} + \mathbf{M}_{2n}$, where

$$\mathbf{M}_{1n} = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}1\\ \mathbf{z}_{ni}\end{bmatrix}(Y_{ni} - E[Y_{ni}|x_{ni}, \mathbf{z}_{ni}]),$$

$$\mathbf{M}_{2n} = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}1\\ \mathbf{z}_{ni}\end{bmatrix}\left(E[Y_{ni}|x_{ni}, \mathbf{z}_{ni}] - \frac{\exp(\theta +' \mathbf{z}_{ni}^*)}{1 + \exp(\theta + \beta'\mathbf{z}_{ni}^*)}\right).$$

Because each of the terms in the summand of $\mathbf{M}_{1n}$ has expectation zero, it follows that $\mathbf{M}_{1n} \to \mathbf{0}$ a.s. by a variant of the strong law of large numbers [Puri and Sen (1971)].

Next consider $\mathbf{M}_{2n}$. By expanding the right-hand term in the summand in a Taylor series, $\mathbf{M}_{2n}$ can be expressed as the sum of three terms:

$$\mathbf{M}_{2n} = \mathbf{M}_{21n} + \mathbf{M}_{22n} + \mathbf{M}_{23n},$$

where

$$\mathbf{M}_{21n} = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}1\\ \mathbf{z}_{ni}\end{bmatrix}\left(\frac{\exp(\theta_0 + \beta_0'\mathbf{z}_{ni})}{1 + \exp(\theta_0 + \beta_0'\mathbf{z}_{ni})} - \frac{\exp(\theta + \beta'\mathbf{z}_{ni})}{1 + \exp(\theta + \beta'\mathbf{z}_{ni})}\right),$$

$$\mathbf{M}_{22n} = \frac{1}{n^{3/2}}\sum_{i=1}^{n}\begin{bmatrix}1\\ \mathbf{z}_{ni}\end{bmatrix}\left(\alpha_0 x_{ni}\frac{\exp(\theta_0 + \beta_0'\mathbf{z}_{ni})}{(1 + \exp(\theta_0 + \beta_0'\mathbf{z}_{ni}))^2} - \beta'\mathbf{V}_{ni}\frac{\exp(\theta + \beta'\mathbf{z}_{ni})}{(1 + \exp(\theta + \beta'\mathbf{z}_{ni}))^2}\right),$$

$$\mathbf{M}_{23n} = \frac{1}{n}\sum_{i=1}^{n} O_{\mathrm{as}}\left(\frac{1}{n}\right).$$

By Kolmogorov's strong law of large numbers, $\mathbf{M}_{21n}$ converges almost surely to a constant, $\mathbf{A}_0(\theta, \beta)$, which is defined in (6.2). Similarly, since $\mathbf{M}_{22n}$ is $n^{-1/2}$ times an average of $n$ iid random variables, each having finite expectation, $\mathbf{M}_{22n} \to \mathbf{0}$ a.s. By similar arguments, $\mathbf{M}_{23n} \to \mathbf{0}$ a.s. Lemma 6.1 follows. □

PROOF OF THEOREM 3.1. Assume without loss of generality that each component of vector $\mathbf{z}_{ni}$ is nonnegative and let

$$\phi = \begin{bmatrix}\theta\\ \beta\end{bmatrix}, \qquad \phi_0 = \begin{bmatrix}\theta_0\\ \beta_0\end{bmatrix} \quad\text{and}\quad \tilde{\phi} = \begin{bmatrix}\tilde{\theta}\\ \tilde{\beta}\end{bmatrix}.$$

Because each component of $\mathbf{A}_0(\phi)$ is continuous in $\phi$ and monotone decreasing in $\phi$, and $\mathbf{A}_0(\phi_0) = \mathbf{0}$, we have for all $\delta > 0$,

$$\mathbf{A}_0(\phi - \delta\mathbf{1}) > \mathbf{0}, \qquad \mathbf{A}_0(\phi + \delta\mathbf{1}) < \mathbf{0},$$

where $\mathbf{1}$ is the $(p + 1)$ vector of ones. (As before, $\mathbf{V}_1 < \mathbf{V}_2$ means that each element of $\mathbf{V}_1$ is less than the corresponding element of $\mathbf{V}_2$.) Since $n^{-1}\mathbf{W}^*(\phi) \to \mathbf{A}_0(\phi)$ a.s., we know that for all $\delta > 0$ and for all $\varepsilon > 0$, there exist integers $N_1(\delta, \varepsilon)$ and $N_2(\delta, \varepsilon)$ such that

$$\Pr\left[\frac{1}{n}\mathbf{W}^*(\phi_0 - \delta\mathbf{1}) > 0 \text{ for all } n > N_1(\phi, \varepsilon)\right] > 1 - \frac{\varepsilon}{2}$$

$$\Pr\left[\frac{1}{n}\mathbf{W}^*(\phi_0 + \delta\mathbf{1}) < 0 \text{ for all } n > N_2(\phi, \varepsilon)\right] > 1 - \frac{\varepsilon}{2}.$$

Define $N(\delta, \varepsilon) = \max(N_1, N_2)$. Then, for all $\delta > 0$ and all $\varepsilon > 0$,

$$\Pr\left[\frac{1}{n}\mathbf{W}^*(\phi_0 - \delta\mathbf{1}) > \mathbf{0} \text{ and } \frac{1}{n}\mathbf{W}^*(\phi_0 + \delta\mathbf{1}) < \mathbf{0} \text{ for all } n > N(\delta, \varepsilon)\right] > 1 - \varepsilon. \tag{6.3}$$

Since $n^{-1}\mathbf{W}^*(\phi)$ is continuous in $\phi$, positive for $(\phi - \delta\mathbf{1})$ and negative for $(\phi + \delta\mathbf{1})$, there exists some value of $\phi$, say, $\tilde{\phi}$, such that $\phi_0 - \delta\mathbf{1} \le \tilde{\phi} \le \phi_0 + \delta\mathbf{1}$ and $n^{-1}\mathbf{W}^*(\tilde{\phi}) = \mathbf{0}$. If there is more than one such value, define $\tilde{\phi}$ to be the one closest to $\phi_0$. Then statement (6.3) is equivalent to the following: for all $\delta > 0$ and for all $\varepsilon > 0$, there exists an integer $N(\delta, \varepsilon)$ such that

$$\Pr\left[\phi_0 - \delta\mathbf{1} \le \tilde{\phi} \le \phi_0 + \delta\mathbf{1}, \text{ for all } n > N(\delta, \varepsilon)\right] > 1 - \varepsilon.$$

But this statement by definition means that $\tilde{\phi}$ is strongly consistent for $\phi_0$, and Theorem 3.1 is proven. □

6.2. *Proof of Theorem* 3.2 (*Asymptotic normality of* $\tilde{\theta}$ *and* $\tilde{\beta}$). To prove Theorem 3.2, we must first define $T^*(\theta, \beta)$, the $((p + 1) \times (p + 1))$ matrix of negative second derivatives of $L^*$ with respect to $\theta$ and $\beta$, that is,

$$T^*(\theta, \beta) = \sum_{i=1}^{n} \begin{bmatrix} 1 \\ \mathbf{z}^*_{ni} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{z}^{*\prime}_{ni} \end{bmatrix} \frac{\exp(\theta + \beta'\mathbf{z}^*_{ni})}{[1 + \exp(\theta + \beta'\mathbf{z}^*_{ni})]^2}.$$

By expanding $\mathbf{W}^*(\tilde{\theta}, \tilde{\beta})$ in Taylor series about $\tilde{\theta} = \theta_0$ and $\tilde{\beta} = \beta_0$, we get

$$\mathbf{W}^*(\tilde{\theta}, \tilde{\beta}) = \mathbf{W}^*(\theta_0, \beta_0) - T^*(\theta_*, \beta_*)\left(\begin{bmatrix} \tilde{\theta} \\ \tilde{\beta} \end{bmatrix} - \begin{bmatrix} \theta_0 \\ \beta_0 \end{bmatrix}\right),$$

where $(\theta_*, \beta_*)$ satisfies

$$\left\| \begin{bmatrix} \tilde{\theta} \\ \tilde{\beta} \end{bmatrix} - \begin{bmatrix} \theta_* \\ \beta_* \end{bmatrix} \right\| \le \left\| \begin{bmatrix} \tilde{\theta} \\ \tilde{\beta} \end{bmatrix} - \begin{bmatrix} \theta_0 \\ \beta_0 \end{bmatrix} \right\|.$$

Since $\mathbf{W}^*(\tilde{\theta}, \tilde{\beta}) = \mathbf{0}$,

$$\frac{T^*(\theta_*, \beta_*)}{n}\sqrt{n}\left\{\begin{bmatrix} \tilde{\theta} \\ \tilde{\beta} \end{bmatrix} - \begin{bmatrix} \theta_0 \\ \beta_0 \end{bmatrix}\right\} = \sqrt{n}\left(\frac{\mathbf{W}^*(\theta_0, \beta_0)}{n}\right). \tag{6.4}$$

Given (6.4), the proof of Theorem 3.2 involves two steps. The first is to prove that $n^{-1}T^*(\theta_*, \beta_*)$ converges in probability to a constant (Lemma 6.2).

We then prove that $n^{-1/2}\mathbf{W}^*(\theta_0, \beta_0)$ has an asymptotic multivariate normal distribution (Lemma 6.3). The desired result (Theorem 3.2) will follow by Slutsky's theorem [Bickel and Doksum (1977)].

LEMMA 6.2. *As $n \to \infty$,*

$$\frac{T^*(\theta_*, \beta_*)}{n} \to_p i_{22},$$

*where $i_{22}$ is defined as in* (3.2).

PROOF. Replacing $\mathbf{z}^*_{ni}$ by $\mathbf{z}_{ni} + (1/\sqrt{n})\mathbf{V}_{ni}$ in $n^{-1}T^*(\theta_*, \beta_*)$ gives

$$\begin{aligned}\frac{1}{n}T^*(\theta_*, \beta_*) &= \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} 1 & \mathbf{z}'_{ni} + \frac{1}{\sqrt{n}}\mathbf{V}'_{ni} \\ \mathbf{z}_{ni} + \frac{1}{\sqrt{n}}\mathbf{V}_{ni} & \left[\mathbf{z}_{ni} + \frac{1}{\sqrt{n}}\mathbf{V}_{ni}\right]\left[\mathbf{z}'_{ni} + \frac{1}{\sqrt{n}}\mathbf{V}'_{ni}\right] \end{bmatrix} \\ &\qquad \times \frac{\exp(\theta_* + \beta'_*\mathbf{z}^*_{ni})}{[1 + \exp(\theta_* + \beta'_*\mathbf{z}^*_{ni})]^2} \\ &= \begin{bmatrix} T^*_1 & \mathbf{T}^{*\prime}_2 \\ \mathbf{T}^*_2 & T^*_3 \end{bmatrix},\end{aligned}$$

where $T^*_1$ is a scalar, and can be written as

$$T^*_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{\exp(\theta_* + \beta'_*\mathbf{z}^*_{ni})}{[1 + \exp(\theta_* + \beta'_*\mathbf{z}^*_{ni})]^2}.$$

Using Taylor expansions, it can be shown that

$$T^*_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{\exp(\theta_0 + \beta'_0\mathbf{z}_{ni})}{[1 + \exp(\theta_0 + \beta'_0\mathbf{z}_{ni})]^2} + o_p(1).$$

Thus, by the weak law of large numbers [Puri and Sen (1971)],

$$T^*_1 \to_p E\left\{\frac{\exp(\theta_0 + \beta'_0\mathbf{z})}{[1 + \exp(\theta_0 + \beta'_0\mathbf{z})]^2}\right\}.$$

To find the probability limit of $\mathbf{T}^*_2$, we express this term as

$$\begin{aligned}\mathbf{T}^*_2 &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_{ni}\frac{\exp(\theta_* + \beta'_*\mathbf{z}_{ni})}{[1 + \exp(\theta_* + \beta'_*\mathbf{z}_{ni})]^2} \\ &\quad + \frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbf{V}_{ni}\frac{\exp(\theta_* + \beta'_*\mathbf{z}_{ni})}{[1 + \exp(\theta_* + \beta'_*\mathbf{z}_{ni})]^2}.\end{aligned}$$

By similar arguments as used earlier, we can show that the first term above converges in probability to

$$E\left\{\mathbf{z}\frac{\exp(\theta_0+\beta_0'\mathbf{z})}{[1+\exp(\theta_0+\beta_0'\mathbf{z})]^2}\right\}, \tag{6.5}$$

and that the second term converges in probability to zero. Thus $\mathbf{T}_2^*$ converges in probability to (6.5).

The third term, $T_3^*$, is given by

$$T_3^*=\frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{z}_{ni}+\frac{1}{\sqrt{n}}\mathbf{V}_{ni}\right]\left[\mathbf{z}'_{ni}+\frac{1}{\sqrt{n}}\mathbf{V}'_{ni}\right]\frac{\exp(\theta_*+\beta'_*\mathbf{z}^*_{ni})}{[1+\exp(\theta_*+\beta'_*\mathbf{z}^*_{ni})]^2}.$$

Using arguments similar to those used above, it is easily shown that

$$T_3^*\to_p E\left\{\mathbf{z}\mathbf{z}'\frac{\exp(\theta_0+\beta_0'\mathbf{z})}{[1+\exp(\theta_0+\beta_0'\mathbf{z})]^2}\right\}.$$

Combining these results, Lemma 6.2 is proven. □

LEMMA 6.3. *As $n\to\infty$,*

$$\sqrt{n}\,\frac{\mathbf{W}^*(\theta_0,\beta_0)}{n}\to_L N_{p+1}(\mathbf{m},i_{22}),$$

*where $\mathbf{m}$ and $i_{22}$ are defined in* (3.2).

PROOF. From (6.2), we have

$$\begin{aligned}\mathbf{W}^*(\theta_0,\beta_0)&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{bmatrix}1\\ \mathbf{z}_{ni}\end{bmatrix}(Y_{ni}-E[Y_{ni}|x_{ni},\mathbf{z}_{ni}])\\ &\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{bmatrix}1\\ \mathbf{z}_{ni}\end{bmatrix}\left(E[Y_{ni}|x_{ni},\mathbf{z}_{ni}]-\frac{\exp(\theta_0+\beta_0'\mathbf{z}^*_{ni})}{1+\exp(\theta_0+\beta_0'\mathbf{z}^*_{ni})}\right)\\ &\quad+\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}0\\ \mathbf{V}_{ni}\end{bmatrix}\left(Y_{ni}-\frac{\exp(\theta_0+\beta_0'\mathbf{z}^*_{ni})}{1+\exp(\theta_0+\beta_0'\mathbf{z}^*_{ni})}\right)\\ &=\mathbf{D}_n+\mathbf{E}_n+\mathbf{F}_n.\end{aligned} \tag{6.6}$$

The terms, say, $\mathbf{d}_{ni}$, in the summand of $\mathbf{D}_n$ have mean zero and covariance matrix

$$\operatorname{Cov}(\mathbf{d}_{ni})=\Sigma_n=E\left\{\begin{bmatrix}1 & \mathbf{z}'\\ \mathbf{z} & \mathbf{z}\mathbf{z}'\end{bmatrix}\frac{\exp\left(\theta_0+\{\alpha_0/\sqrt{n}\}x+\beta_0'\mathbf{z}\right)}{\left[1+\exp\left(\theta_0+\{\alpha_0/\sqrt{n}\}x+\beta_0'\mathbf{z}\right)\right]^2}\right\}.$$

Expanding the fractional part in the brackets in Taylor series about $\alpha_0=0$ and taking the limit of $\Sigma_{n=1}^{m}\Sigma_n/\mathbf{m}$, we find that this sequence of matrices converges to the matrix $i_{22}$. It follows from a variant of the multivariate central limit theorem [Puri and Sen (1971)] that

$$\mathbf{D}_n\to_L N_{p+1}(\mathbf{0},i_{22}).$$

By expanding the first term in parentheses about $\theta_0 = 0$ and the second term in parentheses about $\beta_0'\mathbf{z}_{ni}^* = \beta_0'\mathbf{z}_{ni}$, we can re-express $\mathbf{E}_n$ as

$$\begin{aligned}\mathbf{E}_n &= \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}1\\ \mathbf{z}_{ni}\end{bmatrix}\left\{(\alpha_0 x_{ni} - \beta_0'\mathbf{V}_{ni})\frac{\exp(\theta_0 + \beta_0'\mathbf{z}_{ni})}{[1+\exp(\theta_0+\beta_0'\mathbf{z}_{ni})]^2}\right\}\\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} O_p\left(\frac{1}{n}\right)\\ &= \mathbf{E}_{1n} + \mathbf{E}_{2n}.\end{aligned}$$

By direct application of Khintchin's rule, $\mathbf{E}_{1n}$ converges in probability to $\mathbf{m}$. Since $\mathbf{E}_{2n}$ converges in probability to the zero vector, we have that $\mathbf{E}_n \to_p \mathbf{m}$.

Now consider $\mathbf{F}_n$ defined in (6.6). By adding and subtracting the conditonal mean of $Y_{ni}$ within the parentheses, we get

$$\begin{aligned}\mathbf{F}_n &= \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}0\\ \mathbf{V}_{ni}\end{bmatrix}(Y_{ni} - E[Y_{ni}|x_{ni},\mathbf{z}_{ni}])\\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}0\\ \mathbf{V}_{ni}\end{bmatrix}\left\{E[Y_{ni}|x_{ni},\mathbf{z}_{ni}] - \frac{\exp(\theta_0+\beta_0'\mathbf{z}_{ni}^*)}{1+\exp(\theta_0+\beta_0'\mathbf{z}_{ni}^*)}\right\}.\\ &= \mathbf{F}_{1n} + \mathbf{F}_{2n}.\end{aligned}$$

Since $\mathbf{F}_{1n}$ is the average of $n$ iid random variables, each having mean zero, $\mathbf{F}_{1n} \to_p \mathbf{0}$. To find the probability limit of $\mathbf{F}_{2n}$, expand the first term in parentheses of $\mathbf{F}_{2n}$ about $\alpha_0 = 0$ and the second term about $\beta_0'\mathbf{z}_{ni}^* = \beta_0'\mathbf{z}_{ni}$. This leads to

$$\begin{aligned}\mathbf{F}_{2n} &= \frac{1}{n^{3/2}}\sum_{i=1}^{n}\begin{bmatrix}0\\ \mathbf{V}_{ni}\end{bmatrix}\left\{(\alpha_0 x_{ni} - \beta_0'\mathbf{V}_{ni})\frac{\exp(\theta_0+\beta_0'\mathbf{z}_{ni})}{[1+\exp(\theta_0+\beta_0'\mathbf{z}_{ni})]^2}\right\}\\ &\quad + \sum_{i=1}^{n} O_p\left(\frac{1}{n^2}\right).\end{aligned}$$

Clearly both terms converge in probability to the zero vector, proving Lemma 6.3. □

Proof of Theorem 3.2. From (6.4), we have

$$\frac{T^*(\theta_*,\beta_*)}{n}\sqrt{n}\left\{\begin{bmatrix}\tilde\theta\\ \tilde\beta\end{bmatrix} - \begin{bmatrix}\theta_0\\ \beta_0\end{bmatrix}\right\} = \sqrt{n}\left(\frac{\mathbf{W}^*(\theta_0,\beta_0)}{n}\right).$$

The theorem follows from Lemmas 3.2 and 3.3 and Slutsky's theorem. □

6.3. *Proof of Theorem* 3.3 [*asymptotic normality of* $Q(x^*,\mathbf{z}^*)$]. To prove Theorem 3.3, we first re-express $Q(x^*,\mathbf{z}^*)$ as a quotient. From (2.3), we can write $Q(x^*,\mathbf{z}^*) = U^*(0)/\sqrt{I^*(0)}$, where

$$(6.7)\quad U^*(0) = U^*(\alpha_0)|_{\alpha_0=0} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_{ni}^*\left(Y_{ni} - \frac{\exp\left(\tilde\theta + \tilde\beta'\mathbf{z}_{ni}^*\right)}{1+\exp\left(\tilde\theta+\tilde\beta'\mathbf{z}_{ni}^*\right)}\right)$$

and

$$I^*(0) = I^*(\alpha_0)\big|_{\alpha_0=0} = \frac{1}{n} w^*.$$

The proof is completed in three steps. First, we derive the asymptotic distribution of the numerator term $U^*(0)$. Next, we show that the denominator $I^*(0)$ converges in probability to a constant which is equal to the asymptotic variance of the numerator $U^*(0)$. The theorem then follows from an application of Slutsky's theorem.

Given (6.7), we can rewrite $U^*(0)$ as follows:

$$\begin{aligned} U^*(0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{ni}^* \{Y_{ni} - E[Y_{ni}|x_{ni}, \mathbf{z}_{ni}]\} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{ni}^* \left\{ E[Y_{ni}|x_{ni}, \mathbf{z}_{ni}] - \frac{\exp\left(\tilde{\theta} + \tilde{\beta}' \mathbf{z}_{ni}^*\right)}{1 + \exp\left(\tilde{\theta} + \tilde{\beta}' \mathbf{z}_{ni}^*\right)} \right\} \\ &= A_n + B_n. \end{aligned}$$

By expanding the first term in brackets in $B_n$ in Taylor series about $\alpha_0 = 0$ and the second term about $\tilde{\theta} = \theta_0$ and $\tilde{\beta}' \mathbf{z}_{ni}^* = \beta_0' \mathbf{z}_{ni}$, we get

$$\begin{aligned} B_n &= n^{-1} \sum_{i=1}^{n} x_{ni}^* \alpha_0 x_{ni} \frac{\exp(\theta_0 + \beta_0' \mathbf{z}_{ni})}{[1 + \exp(\theta_0 + \beta_0' \mathbf{z}_{ni})]^2} + \sum_{i=1}^{n} x_{ni}^* O_p(n^{-3/2}) \\ &\quad - n^{-1/2} \sum_{i=1}^{n} \left[ \left(\tilde{\theta} - \theta_0\right) + \left(\tilde{\beta} - \beta_0\right)' \mathbf{z}_{ni} \right] x_{ni}^* \frac{\exp(\theta_0 + \beta_0' \mathbf{z}_{ni})}{[1 + \exp(\theta_0 + \beta_0' \mathbf{z}_{ni})]^2} \\ &\quad - n^{-1} \sum_{i=1}^{n} x_{ni}^* \tilde{\beta}' \mathbf{V}_{ni} \frac{\exp(\theta_0 + \beta_0' \mathbf{z}_{ni})}{[1 + \exp(\theta_0 + \beta_0' \mathbf{z}_{ni})]^2} \\ &\quad - n^{3/2} \sum_{i=1}^{n} x_{ni}^* \left(\tilde{\beta}' \mathbf{V}_{ni}\right)^2 \frac{\exp(\theta_0 + \beta_0' \mathbf{z}_{ni})[1 - \exp(\theta_0 + \beta_0' \mathbf{z}_{ni})]}{[1 + \exp(\theta_0 + \beta_0' \mathbf{z}_{ni})]^3} \\ &\quad - n^{-1/2} \sum_{i=1}^{n} x_{ni}^* \left[ \left(\tilde{\theta} - \theta_0\right) + \left(\tilde{\beta} - \beta_0\right)' \mathbf{z}_{ni} \right]^2 \\ &\qquad\qquad \times \frac{\exp(\theta_0 + \beta_0' \mathbf{z}_{ni})[1 - \exp(\theta_0 + \beta_0' \mathbf{z}_{ni})]}{[1 + \exp(\theta_0 + \beta_0' \mathbf{z}_{ni})]^3} \\ &\quad - n^{-1} \sum_{i=1}^{n} 2 x_{ni}^* \left(\tilde{\beta}' \mathbf{V}_{ni}\right) \left[ \left(\tilde{\theta} - \theta_0\right) + \left(\tilde{\beta} - \beta_0\right)' \mathbf{z}_{ni} \right] \\ &\qquad\qquad \times \frac{\exp(\theta_0 + \beta_0' \mathbf{z}_{ni})[1 - \exp(\theta_0 + \beta_0' \mathbf{z}_{ni})]}{[1 + \exp(\theta_0 + \beta_0' \mathbf{z}_{ni})]^3} \\ &\quad - n^{-1/2} \sum_{i=1}^{n} x_{ni}^* O_p\left[ \left\{ \left(\tilde{\theta} - \theta_0\right) + \left(\tilde{\beta} - \beta_0\right)' \mathbf{z}_{ni} + n^{-1/2} \tilde{\beta}' \mathbf{V}_{ni} \right\}^3 \right]. \end{aligned}$$

Hence, $(A_n, B_n)'$ can be re-expressed:

$$\begin{pmatrix} A_n \\ B_n \end{pmatrix} = \begin{pmatrix} A_n \\ B_1^* \end{pmatrix} + \begin{pmatrix} 0 \\ B_2^* \end{pmatrix} + o_p(1),$$

where

$$B_1^* = -\begin{bmatrix} \tilde{\theta} - \theta_0 \\ \tilde{\beta} - \beta_0 \end{bmatrix}' \left\{ n^{-1/2} \sum_{i=1}^{n} x_{ni}^* \begin{bmatrix} 1 \\ \mathbf{z}_{ni} \end{bmatrix} \frac{\exp(\theta_0 + \beta_0'\mathbf{z}_{ni})}{1 + \exp(\theta_0 + \beta_0'\mathbf{z}_{ni})} \right\}$$

and

$$\begin{aligned} B_2^* &= \left\{ n^{-1} \sum_{i=1}^{n} \alpha_0 x_{ni} x_{ni}^* \frac{\exp(\theta_0 + \beta_0'\mathbf{z}_{ni})}{[1 + \exp(\theta_0 + \beta_0'\mathbf{z}_{ni})]^2} \right\} \\ &\quad - \tilde{\beta}' \left\{ n^{-1} \sum_{i=1}^{n} x_{ni}^* \mathbf{V}_{ni} \frac{\exp(\theta_0 + \beta_0'\mathbf{z}_{ni})}{[1 + \exp(\theta_0 + \beta_0'\mathbf{z}_{ni})]^2} \right\} \\ &= B_{21}^* + B_{22}^*. \end{aligned} \tag{6.8}$$

All other terms from the expansion of $B_n$ converge to zero in probability. We wish to find the distribution of the vector

$$\begin{aligned} \begin{pmatrix} A_n \\ B_1^* \end{pmatrix} &= n^{-1/2} \sum_{i=1}^{n} \begin{bmatrix} x_{ni}^*[Y_{ni} - E(Y_{ni} | x_{ni}, \mathbf{z}_{ni})] \\ -\begin{bmatrix} \tilde{\theta} - \theta_0 \\ \tilde{\beta} - \beta_0 \end{bmatrix}' \begin{bmatrix} 1 \\ \mathbf{z}_{ni} \end{bmatrix} x_{ni}^* \frac{\exp(\theta_0 + \beta_0'\mathbf{z}_{ni})}{[1 + \exp(\theta_0 + \beta_0'\mathbf{z}_{ni})]^2} \end{bmatrix} \\ &= n^{-1/2} \sum_{i=1}^{n} \begin{bmatrix} a_{ni} \\ b_{1i}^* \end{bmatrix}, \end{aligned}$$

where the subscript $n$ reminds us that the distribution of $Y_{ni}$ depends on $n$. Note that the mean of $a_{ni}$ is zero, and its variance is given by

$$\operatorname{Var}(a_{ni}) = E\left\{ x^{*2} \frac{\exp\left(\theta_0 + \{\alpha_0/\sqrt{n}\}x + \beta_0'\mathbf{z}\right)}{\left[1 + \exp\left(\theta_0 + \{\alpha_0/\sqrt{n}\}x + \beta_0'\mathbf{z}\right)\right]^2} \right\}.$$

As $n \to \infty$, $\operatorname{Var}(a_{ni}) \to i_{11}^*$, defined in (3.3).

Also, note that $B_1^*$ is the product of two terms: By Theorem 3.2, the first term is asymptotically normal, with mean $-\mathbf{m}' i_{22}^{-1}$ and covariance matrix $i_{22}^{-1}$; the second term converges in probability to $\mathbf{i}_{12}^*$ by the weak law of large numbers. Therefore, the expected value of $b_{1i}$ is $-\mathbf{m}' i_{22}^{-1} \mathbf{i}_{21}^*$, and its variance is $\mathbf{i}_{12}^* i_{22}^{-1} \mathbf{i}_{21}^*$.

The covariance of $A_n$ and $B_1^*$ is equal to

$$\operatorname{Cov}(A_n, B_1^*) = \operatorname{Cov}\left\{ n^{-1/2} \sum_{i=1}^{n} x_{ni}^* [Y_{ni} - E(Y_{ni} | x_{ni}, \mathbf{z}_{ni})], \right.$$
$$\left. -\begin{bmatrix} \tilde{\theta} - \theta_0 \\ \tilde{\beta} - \beta_0 \end{bmatrix}' n^{-1/2} \sum_{i=1}^{n} x_{ni}^* \begin{bmatrix} 1 \\ \mathbf{z}_{ni} \end{bmatrix} \frac{\exp(\theta_0 + \beta_0' \mathbf{z}_{ni})}{[1 + \exp(\theta_0 + \beta_0' \mathbf{z}_{ni})]^2} \right\}.$$

To make computation of the limit of this expression easier, we re-express $B_1^*$ as a summation involving $\{Y_{ni} - E(Y_{ni} | x_{ni}, \mathbf{z}_{ni})\}$. Recall the following equality from (6.4):

$$\begin{bmatrix} \tilde{\theta} - \theta_0 \\ \tilde{\beta} - \beta_0 \end{bmatrix} = \left[ \frac{T^*(\theta_*, \beta_*)}{n} \right]^{-1} n^{-1} \mathbf{W}^*(\theta_0, \beta_0).$$

Substituting this in $B_1^*$ yields

$$B_1^* = -n^{-1/2} [\mathbf{W}^*(\theta_0, \beta_0)]' \left[ \frac{T^*(\theta_*, \beta_*)}{n} \right]^{-1}$$
$$\times n^{-1} \sum_{i=1}^{n} x_{ni}^* \begin{bmatrix} 1 \\ \mathbf{z}_{ni} \end{bmatrix} \frac{\exp(\theta_0 + \beta_0' \mathbf{z}_{ni})}{[1 + \exp(\theta_0 + \beta_0' \mathbf{z}_{ni})]^2}.$$

$B_1^*$ is the product of three terms. The second term converges in probability to $i_{22}^{-1}$ by Lemma 6.2. The third converges to $\mathbf{i}_{21}^*$ by the weak law of large numbers. Hence,

$$\begin{aligned} B_1^* = &-\left\{ n^{-1/2} \sum_{i=1}^{n} \begin{bmatrix} 1 \\ \mathbf{z}_{ni}^* \end{bmatrix}' (Y_{ni} - E[Y_{ni} | x_{ni}, \mathbf{z}_{ni}]) \right\} i_{22}^{-1} \mathbf{i}_{21}^* \\ &- \left\{ n^{-1/2} \sum_{i=1}^{n} \begin{bmatrix} 1 \\ \mathbf{z}_{ni}^* \end{bmatrix}' \left( E[Y_{ni} | x_{ni}, \mathbf{z}_{ni}] - \frac{\exp(\theta_0 + \beta_0' \mathbf{z}_{ni}^*)}{1 + \exp(\theta_0 + \beta_0' \mathbf{z}_{ni}^*)} \right) \right\} i_{22}^{-1} \mathbf{i}_{21}^* \\ &+ o_p(1) \\ = &-B_{11}^* - B_{12}^* + o_p(1). \end{aligned}$$

Since $B_{12}^*$ converges in probability to $\mathbf{m}' i_{22}^{-1} \mathbf{i}_{21}^*$, it contributes nothing to covariance and can be ignored. Thus, $(A_n, B_1^*)'$ has the same limit as $C_n$, where

$$\begin{aligned} C_n &= \operatorname{Cov}\left[ n^{-1/2} \sum_{i=1}^{n} x_{ni}^* (Y_{ni} - E[Y_{ni} | x_{ni}, \mathbf{z}_{ni}]), \right. \\ &\qquad \left. -n^{-1/2} \sum_{i=1}^{n} \begin{bmatrix} 1 \\ \mathbf{z}_{ni}^* \end{bmatrix}' (Y_{ni} - E[Y_{ni} | x_{ni}, \mathbf{z}_{ni}]) \right] i_{22}^{-1} \mathbf{i}_{21}^* \\ &= -n^{-1} E\left[ \sum_{i=1}^{n} x_{ni}^* \begin{bmatrix} 1 \\ \mathbf{z}_{ni}^* \end{bmatrix}' (Y_{ni} - E[Y_{ni} | x_{ni}, \mathbf{z}_{ni}])^2 \right] i_{22}^{-1} \mathbf{i}_{21}^*. \end{aligned}$$

Substituting for $\mathbf{z}_{ni}^*$ and taking iterated expectations,

$$C_n = -E\left[x^*\begin{bmatrix}1\\ \mathbf{z} + n^{-1/2}\mathbf{V}\end{bmatrix}\frac{\exp\left(\theta_0 + \{\alpha_0/\sqrt{n}\}x + \beta_0'\mathbf{z}\right)}{\left[1 + \exp\left(\theta_0 + \{\alpha_0/\sqrt{n}\}x + \beta_0'\mathbf{z}\right)\right]^2}\right]i_{22}^{-1}\mathbf{i}_{21}^*.$$

After expanding in Taylor series about $\alpha_0 = 0$, it is easy to see that $C_n$ converges to $\mathbf{i}_{12}^* i_{22}^{-1}\mathbf{i}_{21}^*$. Hence, $\operatorname{Cov}(A_n, B_1^*) \to -\mathbf{i}_{12}^* i_{22}^{-1}\mathbf{i}_{21}^*$ as $n \to \infty$.

Given these results, the distribution of $(A_n, B_1^*)' = n^{-1/2}\Sigma(a_{ni}, b_{1i}^*)'$ follows. Since $\operatorname{Var}(a_{ni}) \to i_{11}^*$, clearly

$$\lim_{n\to\infty}\max_{1\le i\le n}\Pr\left\{\left|\frac{a_{ni}}{\left(n\operatorname{Var}(a_{ni})\right)^{1/2}}\right| > \varepsilon\right\} = 0.$$

Also, $\operatorname{Var}(b_{1i}^*) < \infty$ by Lemma 6.2, and $\operatorname{Cov}(A_n, B_1^*) < \infty$. Therefore, by the multivariate central limit theorem [Chung (1974)], $(A_n, B_1^*)'$ is asymptotically normal, with mean $(0, -\mathbf{m}' i_{22}^{-1}\mathbf{i}_{21}^*)'$ and covariance matrix $\Sigma_1$, where

$$\Sigma_1 = \begin{bmatrix} i_{11}^* & -\mathbf{i}_{12}^* i_{22}^{-1}\mathbf{i}_{21}^* \\ -\mathbf{i}_{12}^* i_{22}^{-1}\mathbf{i}_{21}^* & \mathbf{i}_{12}^* i_{22}^{-1}\mathbf{i}_{21}^* \end{bmatrix}.$$

Recall $B_2^*$ from (6.8). By the weak law of large numbers,

$$B_{21}^* \to_p E\left[\alpha_0 x x^* \frac{\exp(\theta_0 + \beta_0'\mathbf{z})}{[1 + \exp(\theta_0 + \beta_0'\mathbf{z})]^2}\right]$$

and

$$B_{22}^* \to_p E\left[x^*\beta_0'\mathbf{V}\frac{\exp(\theta_0 + \beta_0'\mathbf{z})}{[1 + \exp(\theta_0 + \beta_0'\mathbf{z})]^2}\right].$$

Thus, $B_2^*$ converges in probability to $c$, defined in (3.3). Thus, $(A_n, B_n)'$ converges in distribution to a normal random variable with mean $(0, c - \mathbf{m}' i_{22}^{-1}\mathbf{i}_{21}^*)'$ and covariance matrix $\Sigma_1$ by Slutsky's theorem [Bickel and Doksum (1977)]. By the Cramér–Wold result [Rao (1987)], it follows that $A_n + B_n$ is asymptotically normal, with mean $c - \mathbf{m}' i_{22}^{-1}\mathbf{i}_{21}^*$ and variance $i_{11}^* - \mathbf{i}_{12}^* i_{22}^{-1}\mathbf{i}_{21}^*$.

The denominator term of the misspecified score test $Q(x^*, \mathbf{z}^*)$ has exactly the same form as $I_{11.2}(0)$ [the reciprocal of the $(1,1)$ element of the inverse information matrix for $(\alpha_0, \theta_0, \beta_0)$, evaluated at $\alpha_0 = 0$], but replaces $x$ and $\mathbf{z}$ by their misspecified versions $x^*$ and $\mathbf{z}^*$:

(6.9) $$I^*(0) = I_{11}^*(0, \tilde{\theta}, \tilde{\beta}) - \mathbf{I}_{12}^*(0, \tilde{\theta}, \tilde{\beta})\left[I_{22}^*(0, \tilde{\theta}, \tilde{\beta})\right]^{-1}\mathbf{I}_{21}^*(0, \tilde{\theta}, \tilde{\beta}),$$

where

$$I_{11}^*(0,\tilde{\theta},\tilde{\beta}) = \frac{1}{n}\sum_{i=1}^{n} x_{ni}^{*2}\frac{\exp(\tilde{\theta}+\tilde{\beta}'\mathbf{z}_{ni}^*)}{\left[1+\exp(\tilde{\theta}+\tilde{\beta}'\mathbf{z}_{ni}^*)\right]^2},$$

$$\mathbf{I}_{12}^*(0,\tilde{\theta},\tilde{\beta}) = \frac{1}{n}\sum_{i=1}^{n} x_{ni}^*\begin{bmatrix}1 & \mathbf{z}_{ni}^{*\prime}\end{bmatrix}\frac{\exp(\tilde{\theta}+\tilde{\beta}'\mathbf{z}_{ni}^*)}{\left[1+\exp(\tilde{\theta}+\tilde{\beta}'\mathbf{z}_{ni}^*)\right]^2} = \mathbf{I}_{21}^*(0,\tilde{\theta},\tilde{\beta})'$$

$$I_{22}^*(0,\tilde{\theta},\tilde{\beta}) = \frac{1}{n}T^*(\tilde{\theta},\tilde{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}1 & \mathbf{z}_{ni}^{*\prime}\\ \mathbf{z}_{ni}^* & \mathbf{z}_{ni}^*\mathbf{z}_{ni}^{*\prime}\end{bmatrix}\frac{\exp(\tilde{\theta}+\tilde{\beta}'\mathbf{z}_{ni}^*)}{\left[1+\exp(\tilde{\theta}+\tilde{\beta}'\mathbf{z}_{ni}^*)\right]^2}.$$

Given these formulas for the components of the denominator term, we will derive their individual probability limits.

Substituting $\theta_0$ and $\beta_0$ for the consistent estimators $\tilde{\theta}$ and $\tilde{\beta}$ and using Taylor expansions, it can be shown that

$$I_{11}^*(0,\tilde{\theta},\tilde{\beta}) = \frac{1}{n}\sum_{i=1}^{n} x_{ni}^{*2}\frac{\exp(\theta_0+\beta_0'\mathbf{z}_{ni})}{[1+\exp(\theta_0+\beta_0'\mathbf{z}_{ni})]^2} + o_p(1).$$

Hence, $I_{11}^*(0,\tilde{\theta},\tilde{\beta})$ converges in probability to $i_{11}^*$ by Khintchin's rule.

Because $\tilde{\theta}$ and $\tilde{\beta}$ converge in probability to $\theta_0$ and $\beta_0$, $\mathbf{I}_{12}^*(0,\tilde{\theta},\tilde{\beta})$ has the same limit in probability as $\mathbf{I}_{12}^*(0,\theta_0,\beta_0)$:

$$\mathbf{I}_{12}^*(0,\theta_0,\beta_0) = n^{-1}\sum_{i=1}^{n} x_{ni}^*\begin{bmatrix}1 & \mathbf{z}_{ni}^{*\prime}\end{bmatrix}\frac{\exp(\theta_0+\beta_0'\mathbf{z}_{ni}^*)}{[1+\exp(\theta_0+\beta_0'\mathbf{z}_{ni}^*)]^2}.$$

Using a Taylor series expansion and replacing $\begin{bmatrix}1 & \mathbf{z}_{ni}^{*\prime}\end{bmatrix}$ by

$$\begin{bmatrix}1 & (\mathbf{z}_{ni}+n^{-1/2}\mathbf{v}_{ni})'\end{bmatrix},$$

one can show that

$$\mathbf{I}_{12}^*(0,\theta_0,\beta_0) = n^{-1}\sum_{i=1}^{n} x_{ni}^*\begin{bmatrix}1 & \mathbf{z}_{ni}'\end{bmatrix}\frac{\exp(\theta_0+\beta_0'\mathbf{z}_{ni})}{[1+\exp(\theta_0+\beta_0'\mathbf{z}_{ni})]^2} + o_p(1).$$

By the weak law of large numbers, $\mathbf{I}_{12}^*(0,\theta_0,\beta_0)$ converges in probability to $\mathbf{i}_{12}^*$. Hence, $\mathbf{I}_{12}^*(0,\tilde{\theta},\tilde{\beta}) \to_p \mathbf{i}_{12}^*$.

Earlier we showed that

$$I_{22}^*(0,\tilde{\theta},\tilde{\beta}) = \frac{1}{n}T^*(\tilde{\theta},\tilde{\beta}).$$

By Lemma 6.2, $n^{-1}T^*(\theta_*,\beta_*)$ converges in probability to the matrix $i_{22}$ [defined in (3.2)]. Since $\tilde{\theta}$ and $\theta_*$ are strongly consistent for $\theta_0$, and $\tilde{\beta}$ and $\beta_*$ are consistent for $\beta_0$, then $I_{22}^*(0,\tilde{\theta},\tilde{\beta})$ also converges in probability to $i_{22}$.

From (6.9), it follows that $I^*(0)$ converges in probability to $\xi^* = i_{11}^* - \mathbf{i}_{12}^* i_{22}^{-1}\mathbf{i}_{21}^*$.

It has been shown that the numerator of $Q(x^*, \mathbf{z}^*)$ has an asymptotic normal distribution, and the denominator converges in probability to a constant;

$$U^*(0) \to_L N(\delta^*, \xi^*), \qquad I^*(0) \to_p \xi^*,$$

where expressions for $\delta^*$ and $\xi^*$ can be found in (3.3). Hence we conclude by Slutsky's theorem that

$$Q(x^*, \mathbf{z}^*) \to_p N(\mu^*, 1),$$

where $\mu^* = \delta^*/\sqrt{\xi^*}$, proving Theorem 3.3. □

**Acknowledgments.** We are very grateful to John Orav, Louise Ryan, an Associate Editor and the referees for many valuable comments, and to Marcia Ybarsabal-Schoen for her help in preparing the manuscript.

## REFERENCES

ARMSTRONG, B. (1985). Measurement error in the generalized linear model. *Comm. Statist. Simulation Comput.* **14** 529–544.

BAILER, A. J. and PORTIER, C. J. (1988). Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics* **44** 417–431.

BEGG, M. D. and LAGAKOS, S. (1991). Properties of commonly used tests in animal carcinogenicity experiments. Technical report, Dept. Biostatistics, Harvard School of Public Health.

BEGG, M. D. and LAGAKOS, S. (1993). Loss in efficiency caused by omitting covariates and misspecifying exposure in logistic regression models. *J. Amer. Statist. Assoc.* To appear.

BICKEL, P. J. and DOKSUM, K. A. (1977). *Mathematical Statistics*. Holden-Day, San Francisco.

BURR, D. (1988). On errors-in-variables in binary regression—Berkson case. *J. Amer. Statist. Assoc.* **83** 739–743.

CARROLL, R. J., SPIEGELMAN, C. H., LAN, K. K. G., 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.

COCHRAN, W. G. (1968). Errors of measurement in statistics. *Technometrics* **10** 637–666.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman & Hall, London.

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

GAIL, M. H. (1988). The effect of pooling across strata in perfectly balanced studies. *Biometrics* **44** 151–162.

GAIL, M. H., TAN, W. Y. and PIANTADOSI, S. (1988). Tests for no treatment effect in randomized clinical trials. *Biometrika* **75** 57–64.

GAIL, M. H., WIEAND, S. and PIANTADOSI, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* **71** 431–444.

LAGAKOS, S. W. (1982). An evaluation of some two-sample tests used to analyze animal carcinogenicity experiments. *Utilitas Math.* **21B** 239–260.

LAGAKOS, S. W. (1988a). The loss in efficiency from misspecifying covariates in proportional hazards regression models. *Biometrika* **75** 156–160.

LAGAKOS, S. W. (1988b). Effects of mismodeling and mismeasuring explanatory variables on tests of association with a response variable. *Statistics in Medicine* **7** 257–274.

LAGAKOS, S. W. and RYAN, L. M. (1985). Statistical analysis of disease onset and lifetime data from tumorigenicity experiments. *Environmental Health Perspectives* **63** 211–216.

LAGAKOS, S. W. and SCHOENFELD, D. A. (1984). Properties of proportional-hazards score tests under misspecified regression models. *Biometrics* **40** 1037–1048.

MCKNIGHT, B. and CROWLEY, J. (1984). Tests for differences in tumor incidence based on animal carcinogenicity experiments. *J. Amer. Statist. Assoc.* **79** 639–648.

MICHALEK, 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.

PRENTICE, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69** 331–342.

PURI, M. L. and SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York.

RAO, B. L. S. P. (1987). *Asymptotic Theory of Statistical Inference*. Wiley, New York.

SCHOENFELD, D. (1981). The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* **68** 316–319.

SMITH, P. G. and DAY, N. E. (1984). The design of case-control studies: The influence of confounding and interaction effects. *International Journal of Epidemiology* **13** 356–365.

STEFANSKI, L. A. (1985). The effects of measurement error on parameter estimation. *Biometrika* **72** 583–592.

STEFANSKI, L. A. and CARROLL, R. J. (1985). Covariate measurement error in logistic regression. *Ann. Statist.* **13** 1335–1351.

STEFANSKI, L. A. and CARROLL, R. J. (1990). Score tests in generalized linear measurement error models. *J. Roy. Statist. Soc. Ser. B* **52** 345–359.

STEFANSKI, L. A. and CARROLL, R. J. (1991). Deconvolution-based score tests in measurement error models. *Ann. Statist.* **19** 249–259.

TOSTESON, T. D. and TSIATIS, A. A. (1988). The asymptotic relative efficiency of score tests in a generalized linear model with surrogate covariates. *Biometrika* **75** 507–514.

DIVISION OF BIOSTATISTICS
COLUMBIA UNIVERSITY
SCHOOL OF PUBLIC HEALTH
600 W. 168TH STREET
NEW YORK, NEW YORK 10032

DEPARTMENT OF BIOSTATISTICS
HARVARD SCHOOL OF PUBLIC HEALTH
677 HUNTINGTON AVENUE
BOSTON, MASSACHUSETTS 02115