The Annals of Statistics
2000, Vol. 28, No. 5, 1452–1471

OPTIMAL SEQUENTIAL DESIGNS OF CASE-CONTROL STUDIES[1]

BY KANI CHEN

Hong Kong University of Science and Technology

Fixed case-control studies separately collect a case sample and a control sample with the two sample sizes being fixed prior to studies and sometimes arbitrarily chosen. This often results in loss of efficiency of case-control designs in terms of cost-saving or time-saving of the studies. We study sequential case-control designs and, in connection with treatment allocation and stochastic approximation, derive a simple sampling rule that leads to optimal case-control designs. Some important issues such as fixed-width confidence intervals and sequential tests of hypotheses with possible early stopping to save time or costs, which cannot be answered with fixed case-control designs, are shown to be naturally solved with the derived optimal sequential case-control designs.

1. Introduction. The popularity of case-control methodology is mostly attributed to its two major advantages: cost-saving and time-saving. However, despite extensive research and practice on case-control studies, the cost-saving (or sample saving) advantage has not been fully explored or justified and such an impression only remains vague. In addition, in terms of time-saving, fixed case-control designs may be inferior when an early stopping of studies is required for some reason.

Classical fixed case-control studies are carried out by sampling separately from case and control populations with the two sample sizes being fixed and often arbitrary. In fact, how much cost-saving a case-control sampling can achieve depends mostly on the appropriate choice of the ratio of case and control sample sizes. An arbitrarily chosen ratio may result in loss of cost efficiency since such an arbitrary choice is rarely optimal. Fixed case-control designs usually cannot fully achieve cost-saving advantage to our expectation since the optimal ratio is not known prior to the studies. In addition, fixed case-control studies are also inferior, in terms of time-saving, to properly designed sequential case-control studies when cases occur sequentially in time or when there are ethical motivations for trying to terminate case-control studies early for reasons such as saving of samples (see the following example of Nurses Health Study) or reaching a conclusion that requires expeditious public health policy decisions [see O'Neill (1998)].

The above discussed issue is perhaps most clearly reflected in some cost conscious matched case-control studies where cases occur sequentially in time

Received November 1998; revised April 2000.
[1]Supported by Hong Kong RGC Grant HKUST 702/96P.
AMS 1991 *subject classifications*. Primary 62F12; secondary 62I05.
Key words and phrases. Sequential sampling, logistic regression, fixed-width confidence interval, sequential test of hypotheses, treatment allocation, stochastic approximation.

and one must decide a proper number of controls to be selected to match each newly available case. In two-phase or multiphase case-control studies, the proper ratio of case and control sizes must also be determined at each phase. In practice, many case-control studies have the common characteristic that the information on cases and controls are collected sequentially. See, for example, Sartwell, Masi and Arthes (1969), Vessey and Doll (1968), Boston Collaborative Drug Surveillance Project (1973), O'Neill and Anello (1978), Pasternack and Shore (1981), O'Neill (1983) and O'Neill (1998). O'Neill and Anello (1978) first pointed out the advantages of sequential case-control designs. Based on a review of published studies on the relation between breast cancer and reserpine (a widely used antihypertensive agent), including six fixed case-control studies that produced inconclusive results [see, e.g., Heinonen, Shapiro and Tuominen (1974) and Mack, Handerson and Gerkins (1975)], O'Neill and Anello (1978) concluded that fixed case-control studies may not be efficient designs for studies with confirmatory (e.g., relation/no relation) purposes. Thereupon they proposed a simple stopping rule for case-control sampling based on the Wald SPRT. However, their statistical analysis is limited to only matched case-control studies with one dichotomous covariable (i.e., exposure/nonexposure) and fixed matching ratio. The critical problem of how many controls should be matched with each case was not studied there [see also O'Neill (1983) and Pasternack and Shore (1981)]. To the best of our knowledge, there has been no further development in statistical theory or methodology regarding optimal sequential case-control designs.

The ethics and advantages of sequential case-control designs may be best illustrated through a specific example as follows. Consider the Nurses Health Study [Stampfer, Willett, Coldits, Roser, Speizer and Hennekens (1985); see also Robins, Rotnitzky and Zhao (1994)]. At the beginning of the study, a blood serum sample was obtained from each of the 100,000 study subjects and frozen for later analysis. After a follow-up, some of the subjects developed coronary artery disease, which are thus classified as cases, and the rest are disease-free, which are controls. Some coinvestigators wish to study the effect upon the disease development of the antioxidants serum vitamins A and F recorded at the beginning of the study. Due to the high cost of the laboratory analysis and to the small amount of stored serum per subject, only 2% of the stored serum (namely 2000 subjects) is allowed to be used by the coinvestigators. A natural problem is how they should decide the sizes of case and control samples so that the most accurate estimation of the regressor-related (namely, vitamins A and B) parameters can be obtained. Or, more specifically, if they are allowed to take one or a number of samples at a time till 2000 samples are taken, how should they design a rule of allocating the number of case and control samples at each time so that the final estimator based on all 2000 samples would be the most accurate? More practically, suppose the goal of the study is to obtain a confidence interval with a certain confidence level (say 95%) and a fixed-width for the vitamin A related parameter, then with an appropriate sequential sampling rule, they may achieve the purpose without using up the allowed 2000 samples and thus make a considerable saving of blood serum for other studies. Only sequential case-control designs rather than fixed case-control designs can deal with these important issues appropriately.

In this paper, we provide an extensive analysis of sequential case-control designs. We derive a simple rule of sequential case-control sampling that leads to the optimal ratio of case and control sample sizes and thus ensures the optimality in terms of the accuracy of the estimation of regression parameters per sample or per cost unit. The theoretical foundation of such a sequential design is the extension of the asymptotic normality of the semiparametric maximum likelihood estimator (MLE) for fixed case-control designs to sequential case-control designs based on the now popular martingale theory. We show that the optimality problem is essentially to design the sampling so that the ratio of the case and control samples approximates the root of an equation. In this regard, it bears a spiritual resemblance to stochastic approximation [see, e.g., Robbins and Monro (1951), Lai and Robbins (1979, 1981) and Wu (1985a, 1985b)]. By so doing we further prove the proposed simple sequential design is indeed optimal, along with a discussion of an important issue related to treatment allocation problems [cf. Efron (1971) and Wei (1977)]. The proposed rule is based on a key convex property of the asymptotic variance of the semiparametric MLE as a function of the ratio of case and control sample sizes. For practical reasons, the sequential sampling rules are extended to group sequential designs. Fixed-width confidence intervals and sequential tests of hypotheses are then naturally derived from the optimal sequential case-control designs. Furthermore, we consider the general cost efficiency with the cost of collecting a case and a control being possibly unequal.

In the next section, we describe the case-control logistic regression model and extend the classical asymptotic normality of the semiparametric MLE for fixed case-control designs [Anderson (1972), Prentice and Pyke (1979), Breslow and Day (1980) and Breslow (1996)] to sequential case-control designs. In Section 3, we derive a simple sequential sampling rule and prove its optimality. We also discuss the connection between the problem considered and the classical statistical problems of treatment allocation and stochastic approximation. To cope with practical situations, in Section 4 we extend the results in Section 3 to sequential matched case-control designs and group sequential case-control designs. Fixed-width confidence intervals and sequential tests of hypotheses are addressed in Section 5. For simplicity of presentation, Sections 3–5 deal with scalar covariables. Generalizations to covariables of high dimension are considered in Section 6, along with a discussion of cost efficiency. All proofs are deferred to the Appendix.

2. Case-control logistic regression model. Let δ be the dichotomous response given a p-dimensional covariable X with $\delta = 1$ indicating disease and $\delta = 0$ disease-free. A logistic regression model assumes

$$P(\delta = 1|X = x) = u(\alpha + \beta' x) = 1 - P(\delta = 0|X = x), \tag{2.1}$$

where u is the logistic function, that is, $u(t) = e^t/(1+e^t)$, and α and β are the intercept and p-dimensional regressor-related parameters. Throughout the paper, we let α_0 and β_0 be the true values of α and β and denote $\bar{u} = 1 - u$ and $\dot{u} = u\bar{u}$. Let ϕ be the population density of X with respect to a certain measure μ or R^p and let ϕ_1 and ϕ_0 be, respectively, population case and control densities (i.e., the conditional densities of X given $\delta = 1$ and given $\delta = 0$). Then the case-control logistic model assumes

$$\phi_1(x) \propto u(\alpha_0 + \beta_0' x)\phi(x) \quad \text{and} \quad \phi_0(x) \propto \bar{u}(\alpha_0 + \beta_0' x)\phi(x), \tag{2.2}$$

or, equivalently,

$$\phi_1(x) = \exp(\alpha^* + \beta_0' x)\phi_0(x) \quad \text{with } e^{-\alpha^*} = \int_{R^p} \exp(\beta_0' x)\phi_0(x)\mu(dx). \tag{2.3}$$

Notice that the intercept parameter α vanishes and thus is unidentifiable.

DEFINITION 2.1. A sequential case-control sampling takes samples (x_i, δ_i), $i \geq 1$, with x_i being the observed covariate of the ith sample and $\delta_i = 1$ or 0 indicating the ith sample being a case or a control, and the sampling scheme is such that δ_i is $\mathscr{F}_{i-1}$ measurable and x_i is sampled from case or control population according as $\delta_i = 1$ or 0, where $\mathscr{F}_i \equiv \sigma(\{(x_j, \delta_j), j \leq i\}, \mathscr{F}_0)$ for $i \geq 1$ and $\mathscr{F}_0$ is the trivial σ-algebra.

Fixed case-control designs can be viewed as special cases of sequential case-control designs with δ_i being nonrandom. For a sequential case-control design, δ_i is possibly random but must be predictable with respect to the filtration $\{\mathscr{F}_n\}$, that is, must not be dependent on future samples $\{x_i, (x_j, \delta_j), j > i\}$. Throughout the paper, we let $n_1 = \Sigma_{i=1}^n \delta_i$ and $n_0 = n - n_1$ be the sizes of cases and controls up to the nth sample, and let $r_n = n_1/n$ be the case percentage in the first n samples. Observe that n_1, n_0 and r_n are adapted to $\mathscr{F}_{n-1}$ and can all be random in a sequential case-control design. The full likelihood of the first n observations $\{(x_1, \delta_1), \ldots, (x_n, \delta_n)\}$ is

$$L_n^f(\beta, \phi_0) = \prod_{i=1}^n \exp(\delta_i(\alpha^* + \beta' x_i))\phi_0(x_i), \tag{2.4}$$

which is formally the same as fixed case-control designs. Profiling over ϕ_0, one obtains the profile likelihood of β, denoted by L_n, and the log-profile likelihood of β, denoted by l_n, ignoring a constant multiplier,

$$\begin{aligned} L_n(\beta) &= \prod_{i=1}^n u^{\delta_i}(\hat{\alpha}_n(\beta) + \beta' x_i)\bar{u}^{1-\delta_i}(\hat{\alpha}_n(\beta) + \beta' x_i), \\ l_n(\beta) &= \sum_{i=1}^n [\delta_i(\hat{\alpha}_n(\beta) + \beta' x_i) - \log(1 + \exp(\hat{\alpha}_n(\beta) + \beta' x_i))], \end{aligned} \tag{2.5}$$

where $\hat{\alpha}_n(\beta)$ is a function of β, being the unique solution to

$$\sum_{i=1}^n [\delta_i - u(\hat{\alpha}_n(\beta) + \beta' x_i)] = 0. \tag{2.6}$$

Differentiating the log-profile likelihood $l_n(\beta)$ with respect to β, one obtains the semiparametric maximum likelihood estimating equation

$$\dot{l}_n(\beta) = \sum_{i=1}^{n}[x_i(\delta_i - u(\hat{\alpha}_n(\beta) + \beta' x_i))] = 0, \tag{2.7}$$

with $\hat{\alpha}_n(\beta)$ satisfying (2.6). Let $\hat{\beta}_n$ be the semiparametric MLE of β [i.e., solution to (2.7)] and let $\hat{\alpha}_n \equiv \hat{\alpha}_n(\hat{\beta}_n)$. Then, by combining (2.6) and (2.7), we know $(\hat{\alpha}_n, \hat{\beta}_n)$ is the unique solution to

$$G_n(\alpha, \beta) \equiv \sum_{i=1}^{n}\left[\begin{pmatrix}1\\ x_i\end{pmatrix}(\delta_i - u(\alpha + \beta' x_i))\right] = 0. \tag{2.8}$$

This gives rise to the remarkable finding in Anderson (1972) and Prentice and Pyke (1979) that they are formally identical to the prospective maximum likelihood estimating equation with parameter (α, β) although the two sampling schemes are totally different. However, one should notice that $\dot{l}_n$ in (2.7) instead of G_n in (2.8), is the semiparametric maximum likelihood estimating function for case-control sampling and that (2.6) should be regarded as only a restriction condition.

Let E_1 and E_0 denote the expectation with respect to functions of x with densities ϕ_1 (case) and ϕ_0 (control), respectively. Throughout the sequel we define

$$\alpha_r = \log(r/(1-r)) - \log(p_1/(1-p_1)) + \alpha_0, \tag{2.9}$$

where p_1 is the population case percentage [i.e., $p_1 = \int_{R^p} u(\alpha_0 + \beta_0' x) \times \phi(x)\mu(dx)$]. Then α_r is the unique value of α satisfying

$$rE_1[g(x, r)\bar{u}(\alpha_r + \beta_0' x)] = (1-r)E_0[g(x, r)u(\alpha_r + \beta_0' x)] \tag{2.10}$$

for any function g.

PROPOSITION 2.1. *Assume $\int \|x\|^2\phi(x)\mu(dx) < \infty$. If $\min(n_1, n_0) \to \infty$ a.s, $\hat{\alpha}_n - \alpha_{r_n} \to 0$ and $\hat{\beta}_n \to \beta_0$ a.s. If $r_n \to \gamma$ in probability for some constant $\gamma \in (0,1)$, $\binom{\hat{\alpha}_n}{\hat{\beta}_n} - \binom{\alpha_{r_n}}{\beta_0}$ is asymptotically normal with mean 0 of the order of $n^{-1/2}$, and in particular,*

$$\begin{aligned}(n\Sigma(r_n))^{1/2}(\hat{\beta}_n - \beta_0) &\to N(0, I_p),\\ (-\ddot{l}_n(\hat{\beta}_n))^{1/2}(\hat{\beta}_n - \beta_0) &\to N(0, I_p),\end{aligned} \tag{2.11}$$

where I_p is the $p \times p$ identify matrix,

$$-\ddot{l}_n(\hat{\beta}_n) = \sum_{i=1}^{n}\left[\left(x_i - \frac{\sum_{j=1}^{n} x_j \dot{u}(\hat{\alpha}_n + \hat{\beta}_n' x_j)}{\sum_{j=1}^{n}\dot{u}(\hat{\alpha}_n + \hat{\beta}_n' x_j)}\right)^{\otimes 2}\dot{u}(\hat{\alpha}_n + \hat{\beta}_n' x_i)\right] \tag{2.12}$$

and $\Sigma(r)$ is defined in the following (2.13). Equation (2.11) is still true with $\ddot{l}_n(\hat{\beta}_n)$ replaced by $\ddot{l}_n(\beta_0)$ or with $\Sigma(r_n)$ replaced by $\Sigma(\gamma)$.

We now give the definition of $\Sigma(r)$. Define

$$\Sigma(r) = r\Sigma_1(r) + (1-r)\Sigma_0(r),$$

with

$$\Sigma_1(r) = E_1[(x - A(r))^{\otimes 2}\dot{u}(\alpha_r + \beta_0' x)],$$
$$\Sigma_0(r) = E_0[(x - A(r))^{\otimes 2}\dot{u}(\alpha_r + \beta_0' x)]$$

and

$$A(r) = \frac{rE_1(x\dot{u}(\alpha_r + \beta_0' x)) + (1-r)E_0(x\dot{u}(\alpha_r + \beta_0' x))}{rE_1(\dot{u}(\alpha_r + \beta_0' x)) + (1-r)E_0(\dot{u}(\alpha_r + \beta_0' x))}. \tag{2.13}$$

An estimator of $\Sigma(r)$ as a function of r can be naturally derived by plugging in the estimators of α_r, β and the empirical analogues of ϕ_1 and ϕ_0. This will be presented in (3.3).

3. Optimal case-control designs. For simplicity of illustration, we consider a one-dimensional parameter β and covariate X in this section and the following Sections 4 and 5. A sequential case-control design requires $\delta_n \in \mathscr{F}_{n-1}$ which is equivalent to $r_n \in \mathscr{F}_{n-1}$. A sequential sampling rule is always characterized by the predictable random sequence $\{\delta_n\}$ or $\{r_n\}$.

DEFINITION 3.1. A sequential case-control design with sample case percentage $\{r_n\}$ is *asymptotically efficient (optimal)* if, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(\Sigma(\tilde{r}_n) \le (1+\varepsilon)\Sigma(r_n)) = 1 \tag{3.1}$$

for any predictable sequence $\{\tilde{r}_n\}$.

Differentiating $\dot{l}_n(\beta)$ and taking expectation, one can see $\Sigma(r_n)$ approximates $-E\ddot{l}_n(\beta_0)/n$. Thus the above definition fits the general information criteria with the profile score $\dot{l}_n(\beta)$ treated as the true score by the semiparametric nature of the model. The above definition can also be viewed as motivated from the Pitman efficiency based on the asymptotic variance of the semiparametric MLF $\hat{\beta}_n$. To characterize optimal sequential designs, we examine the variance function $1/\Sigma(r)$ as a function of r.

PROPOSITION 3.1. *$1/\Sigma(r)$ as a function of r on the interval $[0, 1]$ is a strictly convex function with a unique minimum achieved at $r = \gamma_0$, which is the unique solution to*

$$D(r) \equiv \frac{r}{1-r}\Sigma_1(r) - \frac{1-r}{r}\Sigma_0(r) = 0. \tag{3.2}$$

A sequential case-control design with sample case percentage $\{r_n\}$ is asymptotically efficient if and only if $r_n \to \gamma_0$ in probability as $n \to \infty$.

To obtain an optimal sequential design, we essentially need to adaptively estimate γ_0. Based on the estimation, we then decide the sampling of a case or a control in the following step so that the combined sample case percentage approaches the estimated γ_0. A natural estimator of γ_0 arises as the solution to the sample analogue of $D(r) = 0$. One slight complication in the estimation of $D(r)$ occurs because $\Sigma_0(r)$ and $\Sigma_1(r)$ as defined in (2.13) depend on α_r which, by its definition in (2.9), involves the inestimable α_0 and p_1. Since $\hat{\alpha}_n - \alpha_{r_n} \to 0$ as $\min(n_1, n_0) \to \infty$ and $\alpha_r - \alpha_{r_n} = \log(r/(1-r)) - \log(r_n/(1-r_n))$, α_r can be estimated by

$$\hat{\alpha}_r \equiv \hat{\alpha}_n + \log(r/(1-r)) - \log(r_n/(1-r_n)) = \hat{\alpha}_n + \alpha_r - \alpha_{r_n}.$$

Now we can analogously define the estimators of $\Sigma_1(r), \Sigma_0(r), \Sigma(r)$ and $D(r)$. For simplicity, we suppress the subscript n and let

$$\widehat{\Sigma}(r) = r\widehat{\Sigma}_1(r) + (1-r)\widehat{\Sigma}_0(r),$$

with

$$\begin{aligned} \widehat{\Sigma}_1(r) &= \frac{1}{n_1}\sum_{i=1}^{n}[\delta_i(x_i - \widehat{A}(r))^{\otimes 2}\dot{u}(\hat{\alpha}_r + \hat{\beta}_n' x_i)], \\ \widehat{\Sigma}_0(r) &= \frac{1}{n_0}\sum_{i=1}^{n}[(1-\delta_i)(x_i - \widehat{A}(r))^{\otimes 2}\dot{u}(\hat{\alpha}_r + \hat{\beta}_n' x_i)], \end{aligned} \tag{3.3}$$

where

$$\widehat{A}(r) = \frac{(r/n_1)\sum_{i=1}^{n}[\delta_i x_i \dot{u}(\hat{\alpha}_r + \hat{\beta}_n' x_i)] + ((1-r)/n_0)\sum_{i=1}^{n}[(1-\delta_i)x_i\dot{u}(\hat{\alpha}_r + \hat{\beta}_n' x_i)]}{(r/n_1)\sum_{i=1}^{n}[\delta_i \dot{u}(\hat{\alpha}_r + \hat{\beta}_n' x_i)] + ((1-r)/n_0)\sum_{i=1}^{n}[(1-\delta_i)\dot{u}(\hat{\alpha}_r + \hat{\beta}_n' x_i)]}.$$

It is straightforward to check that $n\widehat{\Sigma}(r_n) = -\ddot{l}_n(\hat{\beta}_n)$. Define

$$\widehat{D}(r) = \frac{r}{1-r}\widehat{\Sigma}_1(r) - \frac{1-r}{r}\widehat{\Sigma}_0(r), \tag{3.4}$$

and let $\hat{\gamma}$ be the solution to

$$\widehat{D}(r) = 0. \tag{3.5}$$

We now present an efficient sequential case-control sampling rule. To avoid trivialities, we assume that an initial set of $N_1 \neq 0$ cases and $N_0 \neq 0$ controls are always available.

An optimal sequential case-control sampling rule. At stage n with n_1 cases and n_0 controls being already sample, if $\widehat{D}(r_n) < 0$, one takes the next sample as a case; if $\widehat{D}(r_n) > 0$, one takes the next sample as a control; if $\widehat{D}(r_n) = 0$, one can choose to take the next sample as either a case or a control; that is,

$$\delta_{n+1} = \begin{cases} 1, & \text{if } \widehat{D}(r_n) < 0; \\ 0, & \text{if } \widehat{D}(r_n) > 0; \\ 0 \text{ or } 1, & \text{if } \widehat{D}(r_n) = 0. \end{cases} \tag{3.6}$$

REMARK. The above defined sequential sampling rule has a computational virtue in that we need not find the solution to (3.5); one only needs to compute $\widehat{D}(r)$ at $r = r_n$. Notice that $\hat{\alpha}_{r_n} = \hat{\alpha}_n$. Thus no further computations are needed except for those involved in the computation of variance estimation [i.e., $-\ddot{l}_n(\hat{\beta}_n)$]. The resulting simplicity of the algorithm is due to the convexity of $1/\Sigma(r)$ given in Proposition 3.1.

A major issue of sequential analysis is the justification of the consistency of a sequential allocation rule which, in our case, is whether the above naturally derived sampling rule will indeed lead to an asymptotically efficient design, that is, whether $r_n \to \gamma_0$ as $n \to \infty$. This is not a trivial matter since it must be rigorously argued that the above defined sampling rule will at least ensure $\min(n_1, n_0) \to \infty$ as $n \to \infty$. Sequential case-control sampling appears to resemble the treatment allocation or multiarmed bandit problems although their ethics are quite different. In treatment allocation problems in clinical trials or multiarmed bandit problems in stochastic control, the consistency of an intrinsically well-defined allocation rule is often rather difficult to justify, mostly because the samples at early stages may be extremely atypical, leading to off-range estimation of parameters and derailing the intended sampling paths. For example, in sequential clinical trials with patients at their entries facing a choice of two or more treatments, an allocation rule may result in the persistence of assigning patients to some particular treatments. Hence the number of patients allocated to other treatments may remain bounded even when the total number converges to infinity. It then obviously leads to inconsistency since the information about the effects of these treatments cannot be consistently justified based on finite samples. To remedy this drawback, many ad hoc approaches were proposed to force the balance of allocation by sometimes *artificially* assigning patients to treatments that were not previously much used [see, e.g., Efron (1971)]. In our case, such a situation, if it occurs, would be particularly disturbing, since it leads to inconsistency of the estimation of β which is precisely the goal of the study. Fortunately, the design based on (3.6) automatically converges to the balanced asymptotically efficient design as shown in the following proposition and thus need not be artificially adapted to force design balance and consistency.

PROPOSITION 3.2. *Assume* $\int e^{\varepsilon|x|}\phi(x)\mu(dx) < \infty$ *for all* $\varepsilon > 0$. *The sequential case-control sampling based on rule* (3.6) *is asymptotically efficient. In fact,* $r_n \to \gamma_0$ *a.s as* $n \to \infty$.

REMARK. From the above analysis, it becomes clear that sequential case-control sampling can be viewed as a dual problem to stochastic approximation with binary response [cf. Wu (1985a,b)]. Classical stochastic approximation designs the covariate to approximate the root of a function whose response with an error is observed [see Robbins and Monro (1951), Lai and Robbins (1979, 1981)]. For the binary response model, Wu (1985a,b) considered a maximum likelihood recursion procedure to approximate the lethal dose of a certain percentile of response probability which is a solution to an

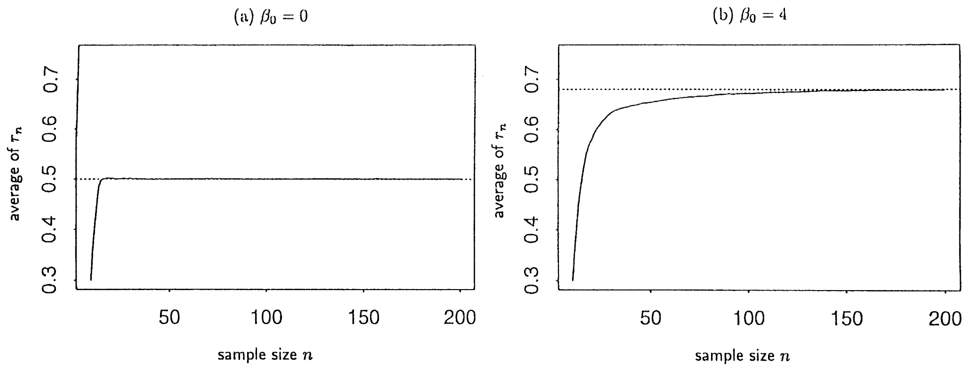

FIG. 1. *The average of r_n versus sample size n.*

equation. The similarities between those studied in Wu (1985a,b) and our optimal sequential case-control designs are that both designs are to approximate the roots of certain nonlinear equations and both are based on the (parametric or semiparametric) likelihood approach. The differences are that the former designs the covariates and the latter designs the response and, in addition, the value of the root is the goal of the former but not the latter. In other words, the precise value of γ_0 in (3.2) plays only an indirect role in our ultimate interest in accurately estimating β.

Two simulation examples are presented in each of the following figures. In both examples, we let the density of controls be $\phi_0(t) = 3(2t-1)^2$, $t \in [0,1]$. The true value of β, β_0, is chosen to be 0 in example (a) and 4 in example (b). By some theoretical calculation, it is known that the optimal case percentage γ_0 is 0.5 in example (a) and is 0.68 in example (b). In the simulations, we start with an initial random sample which contains three cases and seven controls, sequentially select additional cases or controls by the optimal sampling rule (3.6) and stop when the sample size n reaches 200. Over 500 simulations, we compute, for every $10 \le n \le 200$, the average of the case percentage r_n, the empirical standard deviation of r_n, the average of $\hat{\beta}_n$, the empirical and (the average of) estimated standard deviations of $\hat{\beta}_n$ and the coverage probabilities of confidence intervals for β at nominal levels of 95% and 90%. These quantities are plotted against the sample size n in the following figures.

The dotted horizontal bars in Figure 1 indicate the optimal case percentage γ_0, which equals 0.5 in (a) and 0.68 in (b). Figures 1 and 2 clearly demonstrate that r_n converges to γ_0, which is theoretically announced in Proposition 3.2. Figures 3–5 show the empirical evidence of the consistency of $\hat{\beta}_n$ and the validity of its normal based inferences, which are established in Proposition 2.1.

4. Optimal group sequential case-control and sequential matched case-control designs. The standard sequential design discussed in the preceding section assumes one takes one case or control at a time and each time

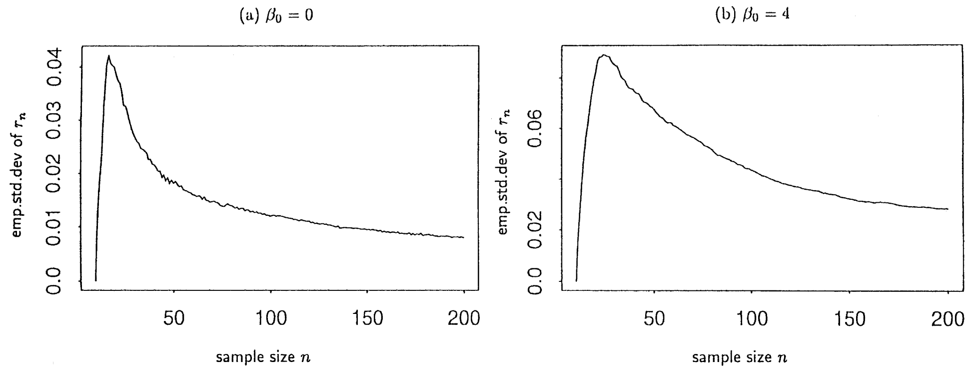

FIG. 2. *The empirical standard deviation of r_n versus sample size n.*

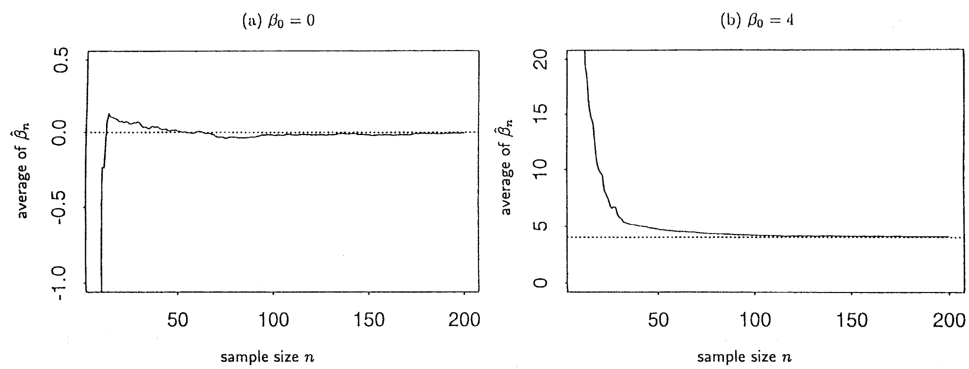

FIG. 3. *The average of $\hat{\beta}_n$ versus sample size n.*

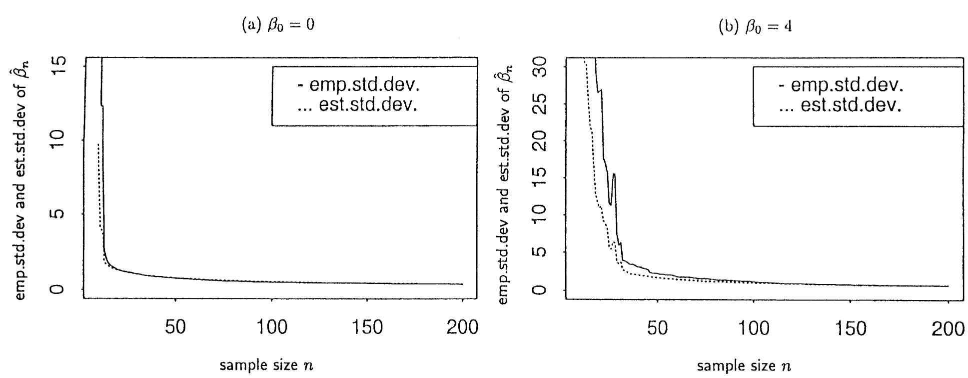

FIG. 4. *The empirical and estimated standard deviations of $\hat{\beta}_n$ versus sample size n.*

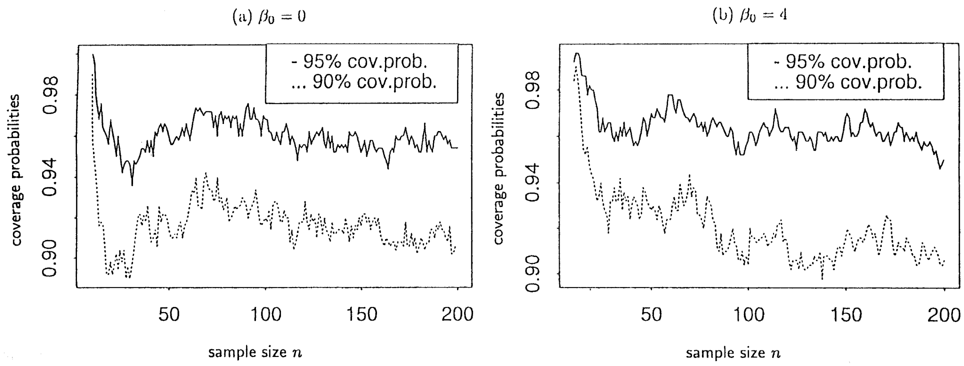

FIG. 5. *The coverage probabilities at* 95% *and* 90% *nominal levels versus sample size* n.

the sampling decision is made based on previously collected samples. It is useful when the allowed total sample size is expectedly of a moderate number (say 20) due to the high cost of covariable ascertainment or different availability of samples. In some situations, such a standard sequential design may not be convenient since it may require many steps of sampling if the allowed total sample size is fairly large. Practitioners often appeal to group sequential sampling, that is, one takes a batch of samples at a time with the size of a batch being possibly random but it must be predictable. Consider a group sequential design with sizes of batches $\{m_i\}$ where m_i is the size of samples at stage i for $i \geq 1$. Then $m_i \in \mathscr{F}_{k_i}$ where $k_i = \sum_{j=0}^{i-1} m_j$ (with $m_0 \equiv 0$). Group sequential designs are also special cases of sequential case-control designs defined in Definition 2.1.

In group sequential designs, m_i is usually chosen by practitioners for reasons concerning the expeditions collection or processing of a sample. Consider first the case of fixed m_i. Assume at the completion of stage $i-1$ we have totally sampled $n = \sum_{j=0}^{i-1} m_j$ subjects which contains n_1 cases and n_0 controls. Recall that $\hat{\gamma}$ is the solution to $\widehat{D}(r) = 0$. Then at stage i, one takes m_{1i} cases and $m_i - m_{1i}$ controls with $m_{1i} = \max(0, \min((n + m_i)\hat{\gamma} - n_1, m_i))$. In this way, at the completion of stage i, we will have case percentage $(n_1 + m_{1i})/(n + m_i) = \max(n_1/(n+m_i), \min(\hat{\gamma}, (n_1+m_i)/(n+m_i)))$ which is the closest possible to $\hat{\gamma}$. Here and in the following, we assume sample sizes can be fractional numbers to avoid trivial but notationally difficult truncations. In practice, the step of solving $\widehat{D}(r) = 0$ may not be necessary if the sizes of batches at each stage are relatively small (say 10) as compared with the practically limited or allowed total sample size (say 1000). In this case, one can decide the sampling of the next stage as being all cases or all controls depending on whether $\widehat{D}(r_n)$ is less than 0 or greater than 0. Thus, the computational load of the procedure may be lessened by skipping the step of solving $\widehat{D}(r) = 0$ or minimizing $1/\widehat{\Sigma}(r)$.

For group sequential sampling with random sizes of batches chosen by practitioners, we propose a sampling rule based on the principle of minimizing $1/\Sigma(r)$. Specifically, suppose n samples containing n_1 cases and n_0 controls have been collected by the end of stage $i-1$. One solves $\widehat{D}(r)=0$ for the solution $r=\hat{\gamma}$ and decides the sampling at stage i such that $(n_1+m_{1i})/(n+m_i)=\hat{\gamma}$, where m_{1i} and m_{0i} are the sizes of cases and controls to be taken at stage i and $m_i = m_{1i} + m_{0i}$. Obviously there are infinite solutions to this equation. To avoid the deviation of $(n_1 + m_{1i})/(n + m_i)$ from γ_0 resulting from the inaccuracy of $\hat{\gamma}$ as an estimator of γ_0, we propose to use the smallest m_i such that (m_{1i}, m_i) solves the above equation. Simple algebra shows, at stage i, that one should take $m_{1i} = (n\hat{\gamma} - n_1)/(1-\hat{\gamma})$ cases and no controls if $n_1/n < \hat{\gamma}$, $m_{0i} = -(n\hat{\gamma} - n_1)/\hat{\gamma}$ controls and no cases if $n_1/n > \hat{\gamma}$, and any moderate sizes of cases and controls if $n_1/n = \hat{\gamma}$.

Sequential matched case-control studies are essentially special cases of group sequential case-control studies. The feature of matched case-control design is that there is one new case at each stage and we must decide the number of controls to be selected to match the case. An optimal design can be found by the same principal. For sequential matched case-control designs, the number of cases at every stage is 1. With the same notation and argument as in the above paragraph, we have $m_{1i} = 1$ and m_{0i} should be such that $(n_1+1)/(n+1+m_{0i}) = \hat{\gamma}$.

The above proposed group sequential case-control designs and sequential matched case-control designs can be proved to be asymptotically efficient in a way similar to the proof of Proposition 3.2. Here we omit the details.

5. Fixed-width confidence intervals and sequential tests of hypotheses.

Suppose now the goal of a study is to obtain a fixed-width confidence interval or a test of hypotheses at a fixed significance level for the parameter β. For this purpose, an early stopping of sampling may be desired. The advantage of an optimal sequential sampling rule along with a proper stopping rule is that it achieves the goal with the smallest total sample size.

In sequential analysis, the fixed-width confidence interval is a classical topic dating back to Stein (1945) [cf. Blum and Rosenblatt (1966) and Chow and Robbins (1965), among others] and has since been extensively studied. In particular, Chang and Martinsek (1992) addressed fixed-size confidence regions for prospective logistic regression models. Although our problem in sequential case-control designs is of a semiparametric nature, with the available optimal design, we can adopt with slight modifications the asymptotic consistency and efficiency for fixed-width confidence intervals given by Khan (1969) for parametric models. Along the lines of Khan (1969), we present the definitions in the following.

DEFINITION 5.1. A fixed-width confidence interval $(\hat{\beta}_T - d, \hat{\beta}_T + d)$ for β associated with a sequential case-control sampling $\{\delta_n\}$ and a stopping time $T = T(d)$ at length $2d$ and confidence level $(1-\bar{\alpha})$ is *asymptotically consistent* if

$$\lim_{d\to 0} P(\beta_0 \notin (\hat{\beta}_T - d, \hat{\beta}_T + d)) \leq \bar{\alpha}, \tag{5.1}$$

and is *asymptotically efficient* if, in addition to (5.1),

$$\lim_{d\to 0} d^2 E(T)\Sigma(\gamma_0)/z^2(\bar{\alpha}/2) = 1, \tag{5.2}$$

where γ_0 is defined in (3.2), $z(\bar{\alpha}/2)$ is the $1-\bar{\alpha}/2$ percentile of the standard normal distribution.

Slightly different from the analogous classical definitions associated only with the stopping rule, the above definition of asymptotic efficiency is also associated with sequential sampling rules. In fact, an inefficient sequential sampling will never produce an asymptotically efficient fixed-width confidence interval, regardless of how the stopping time is defined. Now we present a fixed-width confidence interval with stopping time

$$N = \inf\left\{n \ge 1 \colon n \ge \frac{z^2(\bar{\alpha}/2)}{d^2\widehat{\Sigma}(\hat{\gamma})}\right\}. \tag{5.3}$$

The following proposition shows that a fixed-width confidence interval based on sampling rule (3.6) and stopping time (5.3) is indeed asymptotically consistent and asymptotically efficient.

PROPOSITION 5.1. *In a sequential case-control sampling with sampling rule* (3.6), *a fixed-width confidence interval* $(\hat{\beta}_N - d, \hat{\beta}_N + d)$ *based on the stopping time N defined in* (5.3) *is asymptotically consistent and asymptotically efficient under the condition of Proposition* 3.2.

Another important issue in sequential analysis is sequential tests of hypothesis which may also lead to an early stopping of sampling with desired significance level and power and thus save samples or costs of the study. There are at least three classical asymptotically equivalent test procedures (Wald test, score test and χ^2 test) that are widely used and studied in the literature and extended in sequential analysis. In sequential tests of hypotheses, it is often convenient to consider the score function-based test procedures. For sequential case-control sampling, since the full likelihood is unknown by the semiparametric nature of the model, it is natural to use the profile score function $\dot{l}_n(\beta)$ defined in (2.7).

There are rich literatures on a variety of sequential test procedures, and the important issues such as identification or approximation of the power function and expected sample sizes are rather well studied, especially for the score function-based test procedures. In our case, the problem is slightly complicated by the sequential sampling rules; however, with the optimal sequential sampling rule in (3.6), the classical theory of sequential tests can largely be carried over with no further difficulty. To ensure the validity of such applications, the essential ingredient is Gaussian approximation as follows. For $r_n \to \gamma$ in probability for some nonrandom $\gamma \in (0,1)$, as a process of n,

$$G_n(\alpha_{r_n}, \beta_0) \approx \sum_{i=1}^{n}\left[\begin{pmatrix}1\\x_i\end{pmatrix}(\delta_i - u(\alpha_\gamma + \beta_0 x_i)) - E\left[\begin{pmatrix}1\\x_i\end{pmatrix}(\delta_i - u(\alpha_\gamma+\beta_0 x_i))\middle|\mathscr{F}_{i-1}\right]\right].$$

The approximation implies

$$\dot{l}_n(\beta_0) \approx \sum_{i=1}^{n} \Big[(x_i - A(\gamma))(\delta_i - u(\alpha_\gamma + \beta_0 x_i)) \\ - E\Big[(x_i - A(\gamma))(\delta_i - u(\alpha_\gamma + \beta_0 x_i)) \Big| \mathscr{F}_{i-1} \Big] \Big],$$

where the function A is defined in (2.13). Thus, $\dot{l}_n(\beta_0)$, as a process of n, is approximately a Brownian motion on a time scale of $n\Sigma(\gamma)$. The above approximations can be derived similarly to the proof of Proposition 2.1.

The fact that the score function at the true β_0, $\dot{l}_n(\beta_0)$, behaves like a Brownian motion in a changed time scale makes the sequential test problem fall in the framework of classical sequential analysis. In particular, the approximation of power function becomes regular. For β in a neighborhood of the true β_0 of distance $O(n^{-1/2})$, the test statistic based on $\dot{l}_n(\beta)$ can be written as

$$\dot{l}_n(\beta) \approx \dot{l}_n(\beta_0) + \ddot{l}_n(\beta_0)(\beta - \beta_0) \approx \dot{l}_n(\beta_0) - n\Sigma(\gamma)(\beta - \beta_0),$$

which is approximately a Brownian motion on time scale $n\Sigma(\gamma)$ with a drift $-n\Sigma(\gamma)(\beta - \beta_0)$. To illustrate the point, consider the hypotheses $H_0: \beta = \beta^o$ versus $H_a: \beta > \beta^o$. Then one can typically choose a stopping time

$$T = \inf\{n \geq N^o: \dot{l}_n(\beta^o) > b_0 - b_1 \ddot{l}_n(\beta^o)\},$$

for some fixed integer N^o and positive constants b_0 and b_1, and decide to stop sampling and reject H_0 at T if $T < M$, and stop sampling at M and accept H_0 otherwise, for some large but fixed M. Then the power function at β in the neighborhood of β^o of distance $O(n^{-1/2})$ can be found through the analogous test for a Brownian motion with a drift. In fact, for this particular typical test, the power functions are tabulated in the literature [see, e.g., Siegmund (1985)]. In addition, the expected sample sizes $E(\min(T, M))$ can also be approximated.

In sequential case-control sampling with $r_n \to \gamma$, the power function of the sequential test described above is associated with $\Sigma(\gamma)$ by the Brownian approximation. Since $\Sigma(r)$ is minimized at $r = \gamma_0$, it is clear that the asymptotically efficient sequential sampling given in (3.6) which ensures $r_n \to \gamma_0$ gives the best power function for typical sequential test procedures such as those above. This also highlights yet another advantage of employing the asymptotically efficient sequential sampling rule in (3.6).

6. Cost efficiency and generalizations. One major concern in designing a case-control sampling is to acquire the desired accuracy while keeping to a minimum the cost of sampling for covariable ascertainment. A special feature is that the cost of collection of a case and of a control is often different. As seen in the nurse study example, the cases are often required by many investigators for different purposes and are thus more precious than controls. While the cost of collecting each case can be reasonably assumed equal as is that of each control, the different collection cost of a case and a control may be difficult to be quantitatively described. In spite of this, we choose a reasonable way of defining the cost of collecting a case to be a quantitative number c, and of comparatively defining the cost of collecting a control to be 1. Then, for n samples containing n_1 cases and n_0 controls, the total cost is $n_1c + n_0$. We shall discuss in the following the asymptotic cost efficiency by considering the asymptotic variance $\Sigma(r_n)^{-1}/n$ associated with the cost $n_1c + n_0$.

When the regression parameter β is a scalar, the definition of the efficient case-control design in (3.1) is natural. However, when dealing with a general p-dimensional parameter β, Σ becomes a $p \times p$ matrix and a straightforward generalization of (3.1) by considering $\Sigma(\tilde{r}_n)$ as smaller than $\Sigma(r_n)$ is inadequate. Here the inequality between two $p \times p$ matrices, $B_1 \leq B_2$, means $b'B_1b \leq bB_2b$ for all $b \in R^p$. This is because no such optimal $\{r_n\}$ may exist. It can be seen from the expression of the matrix $\Sigma(r)$ as a matrix function of r that there may not in general exist a γ such that $\Sigma(\gamma) \leq \Sigma(r)$ for all $r \in [0, 1]$. We choose to consider a fairly general criterion through minimizing

$$\text{trace}(H'E[(\hat{\beta}_n - \beta_0)^{\otimes 2}]H) \tag{6.1}$$

for some fixed $p \times p$ matrix $H \neq 0$. A typical choice of H is a diagonal matrix with the diagonal elements being nonnegative real numbers, denoted by $(w_1, \dots, w_p)$, where w_i can be viewed as a weight attached to the accuracy of estimating the ith component of β. If $w_i = 1$ for all $i = 1, \dots, p$ (i.e., $H = I_p$), (6.1) becomes $E\|\hat{\beta}_n - \beta_0\|^2$. If one is only concerned with estimating the ith component of β, he can choose $w_i = 1$ and $w_j = 0$ for all $j \neq i$. We also note that if the criterion is to minimize $Eg(\hat{\beta}_n - \beta_0)$ for a smooth and strictly convex function g on R^p with minimum at the origin, asymptotically it can be reduced to minimizing (6.1) with H being the second derivative matrix of g at the origin.

Now consider the cost efficiency of minimizing (6.1) at a level of cost $n_1c + n_0$. Similarly to Definition 3.1, we can define an asymptotically cost efficient sequential case-control design $\{r_n\}$ as one satisfying, for any $\varepsilon > 0$,

$$\begin{aligned} \lim_{n\to\infty} P[&\text{trace}(H'\Sigma(r_n)^{-1}H)(cr_n + (1 - r_n)) \\ &\leq (1+\varepsilon)\text{trace}(H'\Sigma(\tilde{r}_n)^{-1}H)(c\tilde{r}_n + (1 - \tilde{r}_n))] = 1 \end{aligned} \tag{6.2}$$

for all $\{\tilde{r}_n\}$ with $\tilde{r}_n \in \mathscr{F}_{n-1}$. If $c = 1$, that is, the costs of collecting a case and a control are equal, the terms $cr_n + 1 - r_n$ and $c\tilde{r}_n + 1 - \tilde{r}_n$ vanish in (6.2). If, furthermore, β and x are scalars, (6.2) agrees with (3.1). The strict convexity of $\text{trace}(H'\Sigma^{-1}(r)H)(cr + 1 - r)$ as a function of r is still retained and its derivative function is

$$D_H(r) \equiv \text{trace}\left(H'\Sigma(r)^{-1}\left(\frac{cr}{1-r}\Sigma_1(r) - \frac{1-r}{r}\Sigma_0(r)\right)\Sigma(r)^{-1}H\right),$$

which has a unique 0 solution γ_H. Now define

$$\widehat{D}_H(r) = \operatorname{trace}\biggl(H'\widehat{\Sigma}(r)^{-1}\biggl(\frac{cr}{1-r}\widehat{\Sigma}_1(r) - \frac{1-r}{r}\widehat{\Sigma}_0(r)\biggr)\widehat{\Sigma}(r)^{-1}H\biggr).$$

Let $\hat{\gamma}_H$ be the solution to $\widehat{D}_H(r) = 0$. All the efficient sequential case-control designs defined in Sections 3–5 for scalar β can be proved to be cost efficient with $\hat{\gamma}$ replaced by $\hat{\gamma}_H$ and $\widehat{D}$ replaced by $\widehat{D}_H$.

For fixed-width confidence intervals for one component of β, say the ith component (or, more generally but without further technical difficulty, a linear function of β), one can choose H to be the diagonal matrix with the ith diagonal element being 1 and the rest being 0. The asymptotic consistency can be defined by (5.1) with β_0 and $\hat{\beta}_T$ replaced by their ith components, and the asymptotic cost efficiency by (5.2) with $\Sigma(r)$ replaced by $1/\operatorname{trace}(H'\Sigma^{-1}(\gamma_H)H)$. With the cost efficient case-control sampling and a stopping rule defined in (5.3) with $\widehat{\Sigma}(\hat{\gamma})$ replaced by $1/\operatorname{trace}(H'\widehat{\Sigma}^{-1}(\hat{\gamma}_H)H)$, the fixed-width confidence interval is also asymptotically consistent and asymptotically cost efficient. Notice that, in this case, $\operatorname{trace}(H'\widehat{\Sigma}^{-1}(r)H)$ is simply the ith diagonal element of $\widehat{\Sigma}^{-1}(r)$.

For sequential tests of hypothesis, if the null and alternative hypotheses are divided by a $p-1$ dimension hyperplane, the score test discussed in Section 5 can actually be straightforwardly adapted. For example, consider the hypotheses $H_0\colon \beta_1 = 0$ versus $H_a\colon \beta_1 > 0$, where β_1 is the first coordinate of β. Treating the other components $\beta_2, \dots, \beta_p$ of β as nuisance parameters, one can obtain the profile score of β_1 at $\beta_1 = 0$,

$$\dot{l}_n(\beta_1)\Big|_{\beta_1=0} = \sum_{i=1}^{n}[x_{i1}(\delta_i - u(\hat{\alpha}_n^* + x_{i2}\hat{\beta}_{n2} + \cdots + x_{ip}\hat{\beta}_{np}))],$$

where x_{ik} is the kth coordinate of x_i, and $\hat{\beta}_{nk}$, $k = 2, \dots, p$ are the values that maximize $L_n(\beta)$ with $\beta = (0, \beta_2, \dots, \beta_p)'$ in (2.5) and $\hat{\alpha}_n^* \equiv \hat{\alpha}_n((0, \hat{\beta}_{n2}, \dots, \hat{\beta}_{np})')$ as defined in (2.6). The asymptotic analysis in Section 5 can also be applied with like modifications. However, more general form hypotheses with null and alternative not divided by a $p-1$ dimension hyperplane can be difficult to analyze except for special cases such as $H_0\colon \beta = \beta^o$ versus $H_a\colon \beta \neq \beta^o$ which can be handled by the χ^2 test. This is essentially because problems associated with boundary crossing for high-dimensional Brownian motion is rather complicated to investigate.

APPENDIX

PROOF OF PROPOSITION 2.1. The proof uses the established strong consistency and asymptotic normality of MLE for fixed case-control designs. Details of the proofs of related results may be found in Anderson (1978), Prentice and Pyke (1979), Qin and Zhang (1997) and Chen, Jing and Ying (1999).

Let x_{1i} and x_{0i} denote the ith case sample and the ith control sample, respectively, that is, for $s = 0, 1$, $x_{si} = x_j$ if $\sum_{l=1}^{j} \delta_l = i$ and $\delta_j = s$. Since x_n given δ_n is independent of $\mathscr{F}_{n-1}$, it can be shown $\{x_{1i}\}$ and $\{x_{0i}\}$ are two independent sets of iid random variables with, respectively, case and control population densities ϕ_1 and ϕ_0 as their common densities. Let $\{k_1\}, \{k_0\}$ be two sequences of nonrandom integers. Write the semiparametric maximum likelihood estimating equation based on the independent case samples x_{1i}, $1 \le i \le k_1$, and control samples x_{0i}, $1 \le i \le k_0$, as

$$\sum_{i=0}^{1} \sum_{j=1}^{k_i} \binom{1}{x_{ij}} (i - u(\alpha + \beta x_{ij})) = 0.$$

Let $k = k_1 + k_0$, $r_k^* = k_1/k$ and denote by $(\hat{\alpha}_k^*, \hat{\beta}_k^*)$ the solution to the above equation. Then, by the strong consistency of semiparametric MLE for fixed designs, we know for any $\varepsilon > 0$, with probability 1, $|\hat{\alpha}_k^* - \alpha_{r_k^*}| < \varepsilon$, $\|\hat{\beta}_k^* - \beta_0\| < \varepsilon$ for all large k_1 and k_0. It is clear that $\hat{\alpha}_k^* - \alpha_{r_k^*} \to 0$ and $\hat{\beta}_k^* \to \beta_0$ a.s as $\min(k_1, k_0) \to \infty$. Consequently, the same convergence still holds if $\{k_1\}$ and $\{k_0\}$ are two sequences of random integers such that $\min(k_1, k_0) \to \infty$ a.s. Set $k_1 = n_1$ and $k_0 = n_0$. Then apparently $(\hat{\alpha}_k^*, \hat{\beta}_k^*)$ is identical to $(\hat{\alpha}_n, \hat{\beta}_n)$. Therefore $\hat{\alpha}_n - \alpha_{r_n} \to 0$ and $\hat{\beta}_n \to \beta_0$ a.s if $\min(n_1, n_0) \to \infty$ a.s.

To show asymptotic normality, the key step is to show the asymptotic normality of $n^{-1/2} G_n(\alpha_{r_n}, \beta_0)$ [G_n is defined in (2.8)] and the rest, with the help of the strong consistency proved above, can be proved through the mean value theorem and law of large numbers, following the proofs in, for example, Qin and Zhang (1997) and Chen, Jing and Ying (1999). Consider the sequence

$$Q_n(\alpha, \beta_0) \equiv n^{-1/2} \sum_{i=1}^{n} \left[\binom{1}{x_i} (\delta_i - u(\alpha + \beta_0' x_i)) - E\left(\binom{1}{x_i} (\delta_i - u(\alpha + \beta_0' x_i)) \Big| \mathscr{F}_{i-1} \right) \right].$$

It is clear that $n^{1/2} Q_n(\alpha, \beta_0)$, $n \ge 1$, is an L^2 martingale sequence for every fixed α. Then it can be straightforwardly verified that $Q_n(\alpha_\gamma, \beta_0)$ converges to a normal distribution by the martingale central limit theorem. Furthermore, one can show with a little calculation that there exists a constant $c_0 > 0$ such that

$$E\|Q_n(\alpha_1, \beta_0) - Q_n(\alpha_2, \beta_0)\|^2 \le c_0(\alpha_1 - \alpha_2)^2$$

for all $-M \le \alpha_1, \alpha_2 \le M$ and $n \ge 1$, where $\|\cdot\|$ is the Euclidean norm. Notice that $Q_n(\alpha, \beta_0)$ is a continuous function of α. It follows from Pisier (1983) [see also Pollard (1990), page 13] that there exists a constant $c_0^* > 0$ such that

$$E \sup_{\substack{-M \le \alpha_1, \alpha_2 \le M \\ |\alpha_1 - \alpha_2| \le \varepsilon}} \|Q_n(\alpha_1, \beta_0) - Q_n(\alpha_2, \beta_0)\| \le c_0^* \varepsilon^{1/2}$$

for all $n \ge 1$ and $\varepsilon > 0$. Thus $Q_n(\alpha, \beta_0)$ as a sequence of random functions of $\alpha \in [-M, M]$ is stochastically equicontinuous. This implies that $Q_n(\alpha_{r_n}, \beta_0)$ has the same limiting distribution as $Q_n(\alpha_\gamma, \beta_0)$ if $r_n \to \gamma$ in probability. Applying (2.10), one can show the second term in the definition of $Q_n(\alpha, \beta_0)$

equals 0 when $\alpha = \alpha_{r_n}$. Therefore $n^{-1/2}G_n(\alpha_{r_n}, \beta_0) = Q_n(\alpha_{r_n}, \beta_0)$. The variance computations in (2.11) are nothing more than regular.

PROOF OF PROPOSITION 3.1. By a direct calculation with an application of (2.10), we obtain

$$\frac{\partial}{\partial r}(1/\Sigma(r)) = \left(\frac{r}{1-r}\Sigma_1(r) - \frac{1-r}{r}\Sigma_0(r)\right)\Big/\Sigma^2(r) = D(r)/\Sigma^2(r).$$

Further differentiation shows that

$$\frac{\partial}{\partial r}D(r) = \frac{2}{r^2(1-r)}\bigg[E_0\{(x-A(r))^2 u(\alpha_r+\beta_0 x)\dot{u}(\alpha_r+\beta_0 x)\} + \left(\frac{\partial}{\partial r}A(r)\right)^2 E_0 u(\alpha_r+\beta_0 x)\bigg] > 0,$$

for all $r \in (0,1)$. Thus,

$$\frac{\partial^2}{\partial r^2}(1/\Sigma(r)) = \frac{2D^2(r)}{\Sigma^3(r)} + \left(\frac{\partial}{\partial r}D(r)\right)\Big/\Sigma^2(r) > 0,$$

all $r \in (0,1)$. The second claim of the proposition is then obvious.

PROOF OF PROPOSITION 3.2. We first show that the sequential sampling rule in (3.6) ensures $\min(n_1, n_0) \to \infty$ a.s. Assume there is a collection S of sample paths with positive probability such that the sequence of n_1 remains bounded. This implies, in S, $r_n \to 0$ at the order of $1/n$ and $\widehat{D}(r_n) \ge 0$ for all large n. Without loss of generality, we assume, by excluding a probability 0 set, that almost sure convergences in the rest of the proof hold on every sample path in S. It is seen via the empirical approximation that, for any fixed $M > 0$,

$$G_n(\alpha+\alpha_{r_n}, \beta) \to \sum_{i=1}^{\infty}\left[\binom{1}{x_i}\delta_i\right] - \left(\sum_{i=1}^{\infty}\delta_i\right)\frac{(1-p_1)e^{\alpha_0}}{p_1}E_0\left(\binom{1}{x}\exp(\alpha+\beta' x)\right), \tag{A.1}$$

uniformly over $\{(\alpha,\beta): |\alpha| \le M, |\beta| \le M\}$ for every sample path in S. Notice that $\sum_{i=1}^{\infty}\delta_i$ is finite in S. Equate the right-hand side of (A.1) to 0 and let $(\tilde{\alpha}, \tilde{\beta})$ denote the solution of the equation. Then clearly $(\tilde{\alpha}, \tilde{\beta})$ is finite. It follows from the uniqueness of the solution $(\hat{\alpha}_n, \hat{\beta}_n)$ to $G_n(\alpha,\beta) = 0$ that, in S, $\hat{\alpha}_n - \alpha_{r_n} \to \tilde{\alpha}$ and $\hat{\beta}_n \to \tilde{\beta}$. Since $\hat{\alpha}_n - \alpha_{r_n}$ and $\hat{\beta}_n$ are bounded, one can apply the law of large numbers to show that, in S, $r_n/(1-r_n)\widehat{\Sigma}_1(r_n) \to 0$ and $(1-r_n)/r_n\widehat{\Sigma}_0(r_n)$ is bounded below away from 0 for all large n. Hence $\widehat{D}(r_n) < 0$ for all large n for every sample path in S. This clearly contradicts the preceding statement, $\widehat{D}(r_n) \ge 0$ in S for all large n. We conclude that S must be a set with zero probability and, equivalently, that $n_1 \to \infty$ a.s. Similar arguments show $n_0 \to \infty$ a.s. Thus we have shown $\min(n_1, n_0) \to \infty$ a.s.

The almost sure convergence of $\min(n_1, n_0)$ to ∞ implies, by the strong consistency in Proposition 2.1, that $\hat{\alpha}_n - \alpha_{r_n} \to 0$ and $\hat{\beta}_n - \beta_0 \to 0$ a.s. Following the notation in the first part of the proof of Proposition 2.1, let $\{k_1\}$ and $\{k_0\}$ be two sequences of nonrandom integers. Based on independent iid case samples $\{x_{1i}, 1 \le i \le k_1\}$ and iid control samples $\{x_{0i}, 1 \le i \le k_0\}$, define $\widehat{\Sigma}_1^*(r)$, $\widehat{\Sigma}_0^*(r)$ and $\widehat{D}^*(r)$ in an obviously similar fashion to the definitions of $\widehat{\Sigma}_1(r)$, $\widehat{\Sigma}_0(r)$ and $\widehat{D}(r)$ in (3.3) and (3.4). Let again $k = k_1 + k_0$ and $r_k^* = k_1/k$. It is seen via the empirical approximation and the strong consistency of $(\hat{\alpha}_k^*, \hat{\beta}_k^*)$ that $((1-r)/r)(\widehat{\Sigma}_i^*(r) - \Sigma_i(r))$ converges to 0 a.s. uniformly over $r \in (0,1)$ for $i = 1, 0$ as $\min(k_1, k_0) \to \infty$. Observe that $((1-r)/r)\Sigma_i(r)$ is bounded above and bounded below away from 0 if r is bounded away from 1. Since $D(r)$ is increasing and negative for $r < \gamma_0$, we know that, with probability 1 for all large k_1 and k_0, $\widehat{D}^*(r)$ has the same sign as $D(r)$ if r is bounded above away from γ_0. The same assertion also holds for r bounded below away from γ_0 through an analogous argument.

Let $\{k_1\}$ and $\{k_0\}$ be two random sequences of integers. Similar to the proof of Proposition 2.1, we conclude that, with probability 1 for all large k_1 and k_0, $\widehat{D}^*(r)$ has the same sign as $D(r)$ if r is bounded away from γ_0. Set $k_1 = n_1$, $k_0 = n_0$ and notice that $\widehat{\Sigma}_i = \widehat{\Sigma}_i^*$, $i = 0, 1$ and $\widehat{D}^* = \widehat{D}$. The almost sure convergence of $\min(n_1, n_0)$ to ∞ proved before ensures that, with probability 1 for all large n, $\widehat{D}(r_n) < 0$ (> 0) if r_n is bounded above (below) away from γ_0. Hence, by the definition of the sequential sampling rule in (3.6), it is true that, with probability 1 for all large n, r_n strictly increases (decreases) if r_n is bounded above (below) away from γ_0. Therefore, $r_n \to \gamma_0$ a.s. as $n \to \infty$.

PROOF OF PROPOSITION 5.1. With the asymptotic normality of $\hat{\beta}_n$ proved in Proposition 2.1 and $r_n \to \gamma_0$ a.s. proved in Proposition 3.2, the proof of this proposition can be carried out along the lines of Khan (1969) with no additional difficulty. We omit the details.

REFERENCES

ANDERSON, J. A. (1972). Separate sample logistic discrimination. *Biometrika* **59** 19–35.

BLUM, J. R. and ROSENBLATT, J. (1966). On some statistical problems requiring purely sequential sampling schemes. *Ann Inst. Statist. Math.* **18** 351–355.

Boston Collaborative Drug Surveillance Project (1973). Oral contraceptives and venous thrombolic disease, surgically confirmed gallabladder disease and breast tumors. *Lancet* **1** 1399–1404.

BRESLOW, N. E. (1996). Statistics in epidemiology: the case-control study. *J. Amer. Statist. Assoc.* **91** 14–28.

BRESLOW, N. E. and DAY, N. E. (1980). *Statistical Methods in Cancer Research* **1**. *The Design and Analysis of Case-Control Studies*. IARC, Lyon.

CHANG, Y.-C. I. and MARTINSEK, A. T. (1992). Fixed size confidence regions for parameters of a logistic regression model. *Ann. Statist.* **20** 1953–1969.

CHEN, K., JING, B. Y. and YING, Z. (1999). An asymptotic theory for maximum likelihood estimator in case-control logistic regression. Unpublished manuscript.

CHOW, Y. S. and ROBBINS, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403–417.

HEINONEN, O. P., SHAPIRO, S., TUOMINEN, L. T. and TURVNEN, M. I. (1974). Reserpine use in relation to breast cancer. *Lancet* **2** 675–677.

KHAN, R. A. (1969). A general method of determining fixed-width confidence intervals. *Ann. Math. Statist.* **40** 704–709.

LAI, T. L. and ROBBINS, H. (1979). Adaptive design and stochastic approximation. *Ann. Statist.* **7** 1196–1221.

LAI, T. L. and ROBBINS, H. (1981). Consistency and asymptotic efficiency of slope estimates in stochastic approximation scheme. *Probab. Theory Related Fields* **56** 329–360.

MACK, T. M., HENDERSON, B. E., GERKINS, V. R., ARTHUR, M., BAPTISTA, J. and PIKE, M. C. (1975). Reserpine and breast cancer in a retirement community. *New. England J. Med.* **292** 1366–1371.

O'NEILL, R. T. (1983). Sample size for estimation of the odds ratio in unmatched case-control studies. *Amer. J. Epidemiology* **120** 145–153.

O'NEILL, R. T. (1998). Case-control study, sequential. *Encyclopedia of Biostatistics* (P. Armitage and T. Colton, eds.) **1** 528–532. Wiley, New York.

O'NEILL, R. T. and ANELLO, C. (1978). Case-control studies: a sequential approach. *Amer. J. Epidemiology* **108** 415–424.

PASTERNACK, B. S. and SHORE, R. E. (1981). Sample sizes for individually matched case-control studies. *Amer. J. Epidemiology* **115** 778–784.

PISIER, G. (1983). Some applications of the metric entropy condition to harmonic analysis. *Banach Spaces*, *Harmonic Analysis, and Probability Thoery. Lecture Notes in Math.* **995** 123–154. Springer, New York.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications.* IMS, Hayward, CA.

PRENTICE, R. L. and PYKE, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66** 403–411.

QIN, J. and ZHANG, B. (1997). A goodness-of-fit for logistic regression models based on case-control data. *Biometrika* **84** 609–618.

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **29** 400–407.

ROBINS, J. M., ROTNITZKY, A., and ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866.

SARTWELL, P. E., MASI, A. T., ARTHES, F. G., GREENE, G. R. and SMITH, H. E. (1969). Thromboembolism and oral contraceptives: an epidemiological case-control study. *Amer. J. Epidemiology* **90** 365–380.

SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* Springer, New York.

STEIN, C. (1945). A two-stage test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–258.

STAMPFER, M. J., WILLETT, W. C., COLDITS, G. A., ROSER, B., SPEIZER, F. E., and HENNEKENS, C. H. (1985). A prospective study of postmenopausal estrogen therapy and coronary heart disease. *New England J. Med.* **313** 1044–1049.

VESSEY, P. M. and DOLL, D. R. (1968). Investigation of relation between use of oral contraceptive and thromboembolic disease. *Brit. Med. J.* **2** 199–205.

WEI, L. J. (1977). A class of designs for sequential clinical trials. *J. Amer. Statist. Assoc.* **72** 382–386.

WU, C. F. J. (1985a). Efficient sequential designs with binary data. *J. Amer. Statist. Assoc.* **80** 974–984.

WU, C. F. J. (1985b). Maximum likelihood recursion and stochastic approximation in sequential designs. In *Adaptive Statistical Procedures and Related Topics* (J. van Ryzin, ed.) 298–313. IMS, Hayward, CA.

DEPARTMENT OF MATHEMATICS
HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
CLEAR WATER BAY, KOWLOON
HONG KONG
E-MAIL: makchen@ust.hk