Research Article

Statistics in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Adaptive sampling in two-phase designs: A biomarker study for progression in arthritis

Michael. A. McIsaac[a]*and Richard. J. Cook[b]

Response-dependent two-phase designs are used increasingly often in epidemiological studies to ensure sampling strategies offer good statistical efficiency while working within resource constraints. Optimal response-dependent two-phase designs are difficult to implement, however, since they require specification of unknown parameters. We propose adaptive two-phase designs which exploit information from an internal pilot study to approximate the optimal sampling scheme for an analysis based on mean score estimating equations. The frequency properties of estimators arising from this design are assessed through simulation and they are shown to be similar to those from optimal designs. The design procedure is then illustrated through application to a motivating biomarker study in an ongoing rheumatology research program.

Keywords: adaptive design; asymptotic efficiency; incomplete data; mean-score analysis; response-dependent sampling.

1. Introduction

1.1. Two-Phase Designs with Response-Dependent Sampling

Epidemiological studies with multi-phase sampling designs [1] yield efficient estimators of exposure effects when one or more of the exposure variables are expensive or difficult to measure [2–4]. Key early papers in examining two-phase designs for logistic regression include Breslow and Cain [5] and Scott and Wild [6]. The considerable appeal of such designs has led to their use in many areas of health research including dementia [7], myocardial infarction [8], and nephroblastoma [9]. Two-phase designs are perhaps the most widely employed multi-phase design. In the first phase of sampling (*phase-I*), characteristics which are relatively inexpensive to measure are recorded for all individuals in a large sample; these characteristics may include the response of interest (e.g. disease status) as well as auxiliary covariates. This information is then used to inform the sampling scheme for a second phase (*phase-II*) in which a subsample of individuals from the phase-I sample is selected for measurement of the expensive exposure variable. Efficiency gains over use of simple random sampling at phase-II are realized when phase-I data are used effectively to inform the phase-II sampling design [10, 11] and partial information available from phase-I is exploited in analyses.

Reilly and Pepe [12] develop mean score estimating equations for settings with incomplete covariate data which can be easily implemented in two-phase studies if phase-I data are discrete. This approach exploits information from individuals

¡[a] Queen's University, 99 University Avenue, Kingston, Ontario Canada K7L 3N6
[b] University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1
** Correspondence to: Department of Public Health Sciences, Carruthers Hall, Queen's University Kingston, ON, Canada K7L 3N6. E-mail: mcisaacm@queensu.ca*

selected at both phases of sampling to achieve greater efficiency than is possible based only on 'complete' data (i.e. data collected only from the subsample selected at phase-II). In terms of design, the mean score is particularly attractive since closed-form expressions are available for the phase-II sampling probabilities which minimize the asymptotic variance of the resulting mean score estimator [10, 12]. General use of these asymptotically-optimal phase-II mean score designs has been advocated because of the ease with which they can be implemented [3] and because these designs have been shown to be highly efficient in a variety of settings, even when analyses are not based on the mean score approach [13, 14]. Implementation of asymptotically-optimal designs, however, requires specification of underlying population characteristics including population variances and values of the parameter(s) of interest. Often these necessary design parameters are estimated using data from an external pilot study [3, 10, 15], but external pilot data is also expensive to obtain and the extra cost may not be justified or recovered by the subsequent efficiency gains.

We consider a multi-phase sampling design in which each sampling phase uses internal pilot data collected previously for estimating the asymptotically-optimal sampling procedure for the next phase of sampling. We implement and study a fully adaptive phase-II sampling design in which the asymptotically-optimal design is re-estimated between the selection of every individual in phase-II, and we implement a more practical adaptive design in which the phase-II sample is divided into an internal pilot sub-sample (a *phase-IIa* sample) which exploits the phase-I data, and an approximately optimal sub-sample (a *phase-IIb* sample) which exploits information gathered at both phase-I and at phase-IIa. By using such adaptive procedures, an asymptotically-optimal sampling design can be well-approximated without the need for external pilot data. The use of internal pilot data to modify study design was advocated by Wittes and Brittain [16]. Lohr [17] used a similar approach, called *triple sampling*, in estimation of a multivariate mean. Fedorov *et al.* [18] consider related adaptive designs in the design of dose-finding trials. Pepe *et al.* [15] acknowledged the possibility of such an adaptive approach to approximate optimal mean score designs and stated that this approach needs further exploration. This manuscript addresses this need.

This work is motivated by the need to derive a sampling scheme for a biomarker study involving a cohort of patients with psoriatic arthritis (PsA). We describe the particular details of this problem in the next subsection. The remainder of this article is organized as follows. In Section 2 we define notation, give a derivation of the mean score estimating equations and the limiting distribution of the resulting estimator. The adaptive two-phase design is then described in detail in Section 3. The results of empirical studies are discussed in Section 4 for the case of a binary response and binary and continuous exposure variables. Concluding remarks are made in Section 5.

1.2. A Biomarker Study in Psoriatic Arthritis

Psoriatic arthritis (PsA) is an inflammatory, potentially debilitating, form of arthritis associated with psoriatic skin involvement [19]. The course of this disease is highly variable; many patients do not experience any joint destruction, while as many as 20% suffer extensive joint destruction and consequently experience significantly poorer quality of life and decreased functional ability after developing a more serious form of the condition called arthritis mutilans [20]. Mild cases of PsA are often successfully treated with nonsteroidal anti-inflammatory drugs, but more severe cases require third-line therapies involving disease-modifying antirheumatic drugs and biologic agents [21]. The long-term use of these stronger therapies incurs considerable cost to the health care system and may be associated with increased risk of liver toxicity [22]. It is therefore important to determine in which patients early and aggressive intervention is warranted to alter the disease course and maintain quality of life and functional ability. To this end, there is great interest in identifying biomarkers that are associated with rapid disease progression and disability [21, 22].

Researchers at the University of Toronto PsA Clinic maintain a registry of over 1000 patients with psoriatic arthritis [20]. Disease activity and progression in this registry cohort is tracked through annual clinic visits and bi-annual radiographic examinations. In addition, blood and urine samples from these individuals have been stored in a biobank. Samples from this biobank can be drawn for measurement of potentially important biomarkers. There is particular interest in examining the effect of the matrix metalloproteinase 3 (MMP-3) biomarker, which has been shown to be both associated with presence of PsA in patients with psoriasis, and correlated with disease activity in patients with other forms of arthritis [19].

Processing and testing of the serum samples is expensive and it is not feasible to do this for all patients in the cohort. We therefore require design strategies so that a subset of patients can be selected for measurement of MMP-3 to provide optimal information regarding the association between levels of the expensive biomarker MMP-3 and disease progression. Since the erythrocyte sedimentation rate (ESR), a generic marker of inflammation, is commonly used for prognostic inferences, the goal is to assess the relationship between MMP-3 levels and progression while controlling for the ESR level. Pilot data from the University of Toronto PsA Clinic are used to frame the simulation studies reported in Section 4.

2. Response-Dependent Sampling with the Mean Score Method

2.1. Notation and Model Assumptions

Let Y_i denote a discrete response which could be binary, categorical, or multivariate in which case $Y_i = (Y_{i1}, \ldots, Y_{iK})'$ is a $K \times 1$ vector for individual i, $i = 1, 2 \ldots$. We let X_i denote an exposure variable of primary interest which is difficult or expensive to observe, and let V_i denote a discrete, inexpensive auxiliary covariate. Suppose that interest lies in estimation of the $p \times 1$ parameter β indexing the conditional, possibly joint, probability mass function for Y_i given (X_i, V_i) denoted

$$f(Y_i|X_i, V_i; \beta). \tag{1}$$

The conditional distribution of X_i given V_i, denoted $g(X_i|V_i; \alpha)$, is indexed by a $q \times 1$ parameter α, and we suppose that the marginal probability mass function of V is indexed by an $r \times 1$ parameter γ. The full vector of random variables is therefore governed by the joint model $f(Y, X, V; \beta, \alpha, \gamma) = f(Y|X, V; \beta)g(X|V; \alpha)h(V; \gamma)$, where α and γ are nuisance parameters which are routinely eliminated by conditioning on (X, V) when data are complete.

In a two-phase study, $\{(Y_i, V_i), i = 1, 2, \ldots, N\}$ are observed for N individuals selected in a phase-I sample, but X_i is observed only for the n individuals selected for inclusion in the phase-II sub-sample, where $n < N$. We let R_i indicate selection of individual i into the phase-II sample so that X_i is known only for those n individuals for whom $R_i = 1$. The response and auxiliary covariate are discrete in the present setting, which means that the phase-I sample can be sub-divided into strata defined by (Y, V); we let $N_{YV} = \sum_i I(Y_i = Y, V_i = V)$ denote the size of the corresponding strata, where the indicator function $I(\mathcal{C})$ is 1 if condition $\mathcal{C}$ is true and is 0 otherwise. We let $n_{YV} = \sum_i I(Y_i = Y, V_i = V, R_i = 1)$ represent the number of individuals randomly sub-sampled from the phase-I strata, and we suppose that these phase-II stratum sample sizes can be specified by the investigator; by design, a simple random sample of size n_{YV} will be selected for the measurement of X from the available N_{YV} individuals in stratum (Y, V). The resulting data will be *missing at random* [23] since $X \perp R | Y, V$ (i.e. missingness is conditionally independent of the potentially missing covariate given the phase-I data). We wish to employ optimal phase-II sampling designs which allocate resources to select individuals such that the asymptotic variance is minimized for the estimator of interest; ultimately we wish to determine the optimal choices of n_{YV} for selecting the $n = \sum_{YV} n_{YV}$ individuals for whom the expensive covariate X will be observed.

2.2. The Mean Score Method

Upon selection of the phase-II sample, X is unobserved for $N - n$ individuals, so estimation of β must occur jointly with estimation of the nuisance parameter α in likelihood analysis. Estimation can be carried out by maximizing the observed data likelihood directly or by implementing an EM algorithm [24] in which the equations

$$\sum_{i=1}^{N} \left\{ R_i U_\beta(Y_i|X_i, V_i) + (1 - R_i) E_{X|Y,V}[U_\beta(Y_i|X, V_i)] \right\} = 0 \tag{2}$$

and

$$\sum_{i=1}^{N} \left\{ R_i U_\alpha(X_i|V_i) + (1 - R_i) E_{X|Y,V}[U_\alpha(X|V_i)] \right\} = 0$$

are iteratively solved, where $U_\beta(Y_i|X_i, V_i) = \partial \log f(Y_i|X_i, V_i; \beta)/\partial\beta$ and $U_\alpha(X_i|V_i) = \partial \log g(X_i|V_i; \alpha)/\partial\alpha$ [23,25]. The resulting maximum likelihood estimator for β will be consistent and fully efficient if all model assumptions are valid. It can be challenging to specify the nuisance probability model for $X|V$, and it is undesirable for inferences about β to be sensitive to misspecification of this model [26].

To address this, Reilly and Pepe [12] propose a pseudolikelihood method which leads to a mean score estimating equation for β obtained through a non-iterative, empirical approximation to the expectation in (2). When Y and V are discrete and the data are missing at random, the conditional expectation of the pseudoscore is estimated as

$$\widehat{E}_{X|Y,V}[U_\beta(Y_j|X,V_j)] = n^{-1}_{Y_jV_j} \sum_{i:R_i=1} U_\beta(Y_i|X_i,V_i) I(Y_i = Y_j, V_i = V_j).$$

and the mean score estimating equations reduce to

$$\bar{U}(\beta) = \sum_{i=1}^{N} R_i \pi(Y_i,V_i)^{-1} U_\beta(Y_i|X_i,V_i) = 0,$$

where $\pi(Y,V) = n_{YV}/N_{YV}$ [12, 25]. The mean-score method, therefore, can be viewed as a Horvitz-Thompson-type inverse probability weighted estimating equation for which the selection probabilities are estimated empirically [8].

Under mild regularity conditions [27], a consistent estimator $\widehat{\beta}$ is obtained as the solution to the mean score equation $\bar{U}(\beta) = 0$. Moreover,

$$\sqrt{N}(\widehat{\beta} - \beta) \xrightarrow{p} N(0, \mathcal{A}^{-1} + \mathcal{A}^{-1}\mathcal{B}\mathcal{A}^{-1}), \tag{3}$$

where $\mathcal{A} = E_{YXV}[-\partial U_\beta(Y_i|X_i,V_i)/\partial\beta']$,

$$\mathcal{B} = \sum_{YV} P(Y,V)\big[\frac{N_{YV}}{n_{YV}} - 1\big] \cdot \text{var}_{X|Y,V}[U_\beta(Y_i|X_i,V_i))],$$

and $P(Y,V)$ is the joint probability mass function for (Y,V); see the Supplementary Material for details. Reilly and Pepe [12] show that when N and n are fixed, the form of this asymptotic variance can be exploited to give closed-form solutions for the optimal phase-II sampling design which selects individuals in order to minimize the asymptotic variance of an estimator of interest. Reilly and Pepe [12] considered fixing the expected phase-II sample size through the constraint

$$n = N \sum_{YV} \big[n_{YV}/N_{YV} \cdot P(Y,V)\big]. \tag{4}$$

If we wish to minimize the $[k,k]$ entry of the covariance matrix defined in (3) corresponding to the estimator of a parameter of interest with this constraint, then the optimal stratum-specific sample sizes are

$$n^{(\text{opt})}_{yv} = \frac{N_{yv} \cdot n/N \cdot \big\{\mathcal{A}^{-1}\text{var}_{X|y,v}\big[U_\beta(Y_i|X_i,V_i)\big]\mathcal{A}^{-1}\big\}^{1/2}_{[k,k]}}{\sum_{YV} P(Y,V) \cdot \big\{\mathcal{A}^{-1}\text{var}_{X|Y,V}\big[U_\beta(Y_i|X_i,V_i)\big]\mathcal{A}^{-1}\big\}^{1/2}_{[k,k]}},$$

where the $[k,k]$ entry of a matrix $\mathcal{V}$ is denoted by $\{\mathcal{V}\}_{[k,k]}$. Here we focus on minimizing the asymptotic variance for the estimator of a particular parameter, however, researchers are free to define optimality criteria in a variety of ways and designs can naturally be modified to accommodate optimality criteria based on linear functions of the elements of the asymptotic variance matrix (e.g. analogues to A-optimality and C-optimality [28] could be considered as described in McIsaac and Cook [13]).

An obvious challenge in implementing an optimal design is that computation of these optimal stratum-specific sample sizes hinges on specification of the matrix $\mathcal{A}$, $\text{var}_{X|Y,V}\big[U_\beta(Y_i|X_i,V_i)\big]$, and the joint probability $P(Y,V)$ [12]. McIsaac and Cook [13] also noted that the expected sample size constraint of Reilly and Pepe [12] may not be appropriate when the sizes of the phase-I strata are known. In the next section, we derive an adaptive procedure which uses a more suitable constraint and which does not require external information to specify the approximately optimal selection probabilities.

3. Adaptive Response-Dependent Multi-Phase Sampling Designs

The proposed adaptive sampling strategy involves partitioning the phase-II sampling into a series of distinct sub-phases. Individuals in the first sub-phase of phase-II data collection cannot be selected optimally since at this point there is insufficient data to estimate the components necessary for approximating the optimal mean score design (i.e. $\mathcal{A}$, $\mathrm{var}_{X|Y,V}\left[U_\beta(Y_i|X_i,V_i)\right]$, and $P(Y,V)$ cannot all be estimated using only phase-I data). Subsequent sub-phases, however, can make use of the accumulating data to estimate the necessary design components and, therefore, one can approximate the asymptotically-optimal design [18, 29, 30].

As an extreme implementation of the proposed approach we consider a *fully adaptive* sampling strategy in which, following a small initial sample, selection probabilities are updated after each individual is sampled in phase-II. This fully adaptive design therefore features almost as many sampling sub-phases as there are individuals selected in phase-II. We also consider a *two-stage* adaptive sampling strategy which uses only two sub-phases of sampling individuals for measurement of the expensive covariates; we refer to these as *phase-IIa* in which $n^{(a)}$ individuals are selected for measurement of X without information about the optimal design components, and *phase-IIb* in which we sample the remaining $n^{(b)}$ phase-II individuals by exploiting the phase-I and phase-IIa data to approximate the asymptotically-optimal design.

In what follows, we describe how the two-stage adaptive sampling procedure enables one to approximate asymptotically-optimal sampling at the second sub-phase of phase-II (i.e. phase-IIb); this process can be generalized to a sampling design with an arbitrary number of stages. The impact on efficiency of the number of sampling sub-phases and the relative sizes of each sampling sub-phase are explored through simulations in the next section, where we also examine the impact of using empirical estimators for $\mathcal{A}$, $\mathrm{var}_{X|Y,V}[U_\beta(Y_i|X_i,V_i)]$, and $P(Y,V)$.

The sampling plan involves deciding how to distribute the fixed phase-II sample size of $n = n^{(a)} + n^{(b)}$ individuals amongst the strata defined by phase-I data while adhering to the phase-II sample size constraint

$$n = \sum_{YV} [n^{(a)}_{YV} + n^{(b)}_{YV}] \;, \tag{5}$$

where $n^{(a)}_{YV}$ represents the number of individuals selected from stratum (Y,V) at phase-IIa and $n^{(b)}_{YV}$ is defined analogously for phase-IIb. We assume that the stratum-sizes of the internal pilot study, $n^{(a)}_{YV}$, are fixed and we select the remaining individuals using the optimal design which minimizes the asymptotic variance of the estimator of interest (the $[k,k]$ entry of the asymptotic covariance matrix in (3)). Assuming that the necessary optimal design components are estimated as $\widehat{\mathcal{A}}$, $\widehat{\mathrm{var}}_{X|Y,V}[U_\beta(Y_i|X_i,V_i)]$, and $\widehat{P}(Y,V)$ using data collected at phase-I and phase-IIa, we wish to find the $\{n^{(b)}_{YV}\}$ that are stationary points of the Lagrangian

$$\Lambda = \left\{ \widehat{\mathcal{A}}^{-1} + \widehat{\mathcal{A}}^{-1} \left\{ \sum_{YV} \widehat{P}(Y,V) \left[\frac{N_{YV}}{n^{(a)}_{YV} + n^{(b)}_{YV}} - 1 \right] \cdot \widehat{\mathrm{var}}_{X|Y,V}[U_\beta(Y_i|X_i,V_i)] \right\} \widehat{\mathcal{A}}^{-1} \right\}_{[k,k]} + \lambda \left[\sum_{Y,V} (n^{(a)}_{YV} + n^{(b)}_{YV}) - n \right].$$

The optimal $n^{(b)}_{yv}$ is the solution to

$$\frac{\partial \Lambda}{\partial n^{(b)}_{yv}} = \left\{ \widehat{\mathcal{A}}^{-1} \left\{ \frac{N_{yv}}{(n^{(a)}_{yv} + n^{(b)}_{yv})^2} \widehat{P}(y,v) \widehat{\mathrm{var}}_{X|y,v}[U_\beta(Y_i|X_i,V_i)] \right\} \widehat{\mathcal{A}}^{-1} \right\}_{[k,k]} + \lambda = 0,$$

and the optimal stratum-specific phase-IIb sampling sizes are therefore best selected as

$$n^{(b,\mathrm{opt})}_{yv} = \frac{n \cdot N_{yv}^{1/2} \widehat{P}(y,v)^{1/2} \cdot \left\{ \widehat{\mathcal{A}}^{-1} \widehat{\mathrm{var}}_{X|y,v}\left[U_\beta(Y_i|X_i,V_i)\right] \widehat{\mathcal{A}}^{-1} \right\}^{1/2}_{[k,k]}}{\sum_{YV} N_{YV}^{1/2} \widehat{P}(Y,V)^{1/2} \cdot \left\{ \widehat{\mathcal{A}}^{-1} \widehat{\mathrm{var}}_{X|Y,V}\left[U_\beta(Y_i|X_i,V_i)\right] \widehat{\mathcal{A}}^{-1} \right\}^{1/2}_{[k,k]}} - n^{(a)}_{yv} \tag{6}$$

under the constraints $0 \le n^{(b)}_{yv} \le N_{yv} - n^{(a)}_{yv}$. If $n^{(b)}_{yv} > N_{yv} - n^{(a)}_{yv}$ for some (y,v) then one can set $n^{(b)}_{yv} = N_{yv} - n^{(a)}_{yv}$ and solve for the other phase-IIb sample sizes using an updated version of (6) [12]. Here we need not worry about the degenerate case $n^{(b)}_{yv} = 0$, but must simply ensure $n^{(a)}_{yv} + n^{(b)}_{yv} > 0$. Note that $n^{(a)}_{yv} + n^{(b,\mathrm{opt})}_{yv}$ is slightly different than $n^{(\mathrm{opt})}_{yv}$ of Reilly and Pepe [12] due to the different phase-II sample size constraints we adopt (i.e. (5) as opposed to (4)), although the two will align when $P(Y, V)$ is estimated empirically as $\widehat{P}(Y, V) = N_{YV}/N$.

4. Empirical Properties of Adaptive Two-Phase Mean-Score Designs

4.1. Design of Simulation Studies

In this section, we examine empirically the stratum-specific sampling probabilities and the efficiency of adaptive mean-score designs in the context of the psoriatic arthritis biomarker study. In each simulation, we implement a non-adaptive sampling design based on proportional stratified sampling (i.e. where $n_{YV} \propto N_{YV}$), a non-adaptive sampling designs based on balanced stratified sampling (i.e. where all n_{YV} are equal), a fully adaptive sampling design, two-stage adaptive designs with varying choices of $n^{(b)}$, and the local asymptotically-optimal design based on the true (in practice, unknown) parameters. The latter is considered as a benchmark to assess how close to optimal the adaptive designs can behave. We also consider estimation of the design components $\mathcal{A}$, $\text{var}_{X|Y,V}\left[U_\beta(Y_i|X_i, V_i)\right]$, and $P(Y, V)$ in two ways. The first is by assuming parametric forms of the component models and finding interim estimates of necessary parameter values; $\mathcal{A}$, for example, would be estimated parametrically as $\widehat{\mathcal{A}} = \mathcal{A}(\widehat{\beta}^{(a)}, \widehat{\alpha}^{(a)}, \widehat{\gamma}^{(a)})$ where parameters are estimated based on the currently available (phase-I and phase-IIa) data. The second is through empirical estimation; for example, using

$$\widehat{\mathcal{A}} = -N^{-1} \sum_i R_i^{(a)} \pi^{(a)}(Y_i, V_i)^{-1} \partial U_\beta(Y_i|X_i, V_i)/\partial \beta' \big|_{\widehat{\beta}^{(a)}} ,$$

where $R_i^{(a)}$ indicates that individual i was selected in phase-IIa and $\pi^{(a)}(Y, V) = n_{YV}^{(a)}/N_{YV}$.

In our simulation studies, we assume the correct parametric form when estimating design components; it would be impossible to guarantee that an assumed parametric form is correct in practice, but other studies have demonstrated that the efficiency of optimal mean score designs can be fairly robust to misspecification of such auxiliary models [13, 31]. In addition, we note that misspecification of these auxiliary models at the design stage will not affect the consistency of the robust mean-score estimation procedure. As suggested by Pepe *et al.* [15], we stabilized the estimation of the design components by adding two artificial observations to each phase-I stratum.

We consider here results from simulation studies which focus on Y and V as univariate binary variables, but consider separately binary and continuous versions of the covariate X. Again, for the two-stage adaptive sampling we consider phase-II selection in two steps. First, we select $n^{(a)}$ individuals using proportional or balanced sampling. Second, we select the remaining $n^{(b)} = n - n^{(a)}$ individuals according to the approximate optimal mean score design found by using both the phase-I and phase-IIa data to estimate the necessary design components parametrically or empirically. The fully adaptive sampling is similar, but it uses a small balanced sample at phase-IIa to allow for an initial estimate of the design components and then individuals are selected one-at-a-time with estimates of the design components being updated after the selection of each individual.

For the two-stage adaptive sampling, we consider ten possible choices for the proportion of the phase-II sample selected at phase-IIb: we allow $n^{(b)}/n$ to range from 0% to 90% in 10% increments. Note that when $n^{(b)}/n = 0\%$, $n^{(b)} = 0$ which represents a scenario where the entire phase-II sample is chosen through either proportional or balanced sampling (i.e. the design is non-adaptive). We frame the design in the context of the motivating two-phase biomarker study in psoriatic arthritis. Specifically, this study aims to examine optimal two-phase designs for selecting patients for measurement of the biomarker MMP-3 in the University of Toronto Psoriatic Arthritis Clinic with a view to examining the relationship between baseline MMP-3 levels and disease progression. We let Y denote a binary response indicating the fact that the disease has progressed over the period of follow-up, let V denote an inexpensive binary covariate indicating an elevated (abnormal) ESR measurement at the baseline assessment, and let X denote the baseline MMP-3 levels, the expensive covariate of interest.

For the simulation study, 2000 phase-I samples of size $N = 800$ were generated. For each individual, the response Y had conditional mean

$$P(Y = 1|X, V; \beta) = \text{expit}(\beta_0 + \beta_x X + \beta_v V), \tag{7}$$

and the binary covariates satisfied

$$P(X = 1|V; \alpha) = \text{expit}(\alpha_0 + \alpha_v V),$$

and

$$P(V = 1; \gamma) = \text{expit}(\gamma_0), \tag{8}$$

where $\text{expit}(Z) = \exp(Z)/(1 + \exp(Z))$. The parameter values $(\beta_0, \beta_x, \beta_v) = (-1.95, 1.00, 0.90)$, $(\alpha_0, \alpha_v) = (1.05, -0.41)$, and $\gamma_0 = -0.04$ were obtained from fitting models to a small independent historical sample of the PsA patients where MMP-3 levels were dichotomized using a clinically relevant cut-point. We also considered the case in which X is a continuous variable representing the scenario in which MMP-3 levels were not dichotomized. Here Y and V were still Bernoulli variables generated according to (7) and (8), however, we supposed $X|V$ arose from a gamma distribution with shape α_0 and scale $\alpha_1 + \alpha_v V$, as in

$$g(X|V; \alpha) = \frac{X^{\alpha_0 - 1} \exp(-X/(\alpha_1 + \alpha_v V))}{\Gamma(\alpha_0)(\alpha_1 + \alpha_v V)^{\alpha_0}} \, .$$

Since Y and V were discrete, we could still define strata as before at phase-I. The parameter values used in generating these data were $(\beta_0, \beta_x, \beta_v) = (-2.18, 0.03, .84)$, $(\alpha_0, \alpha_1, \alpha_v) = (1.40, 10, 5)$, and $\gamma_0 = -0.04$. where values were chosen to reflect the empirical distribution of MMP-3 measurements given ESR status in the historical PsA data.

Optimal and adaptive designs were employed to minimize the asymptotic variance of the estimator of β_x, which was taken to be of special interest. Each design was executed for each of the 2000 simulated datasets, with $n = 200$ subjects sampled in phase-II. One way of contrasting the local asymptotically-optimal design and the adaptive designs is in terms of the proportion of individuals selected for measurement of X from each stratum, $(n_{YV} - n_{YV}^{\text{opt}})/N_{YV}$. These are displayed in Figures 1 and 3 for the case where X is binary and in Figures 2 and 4 for settings involving a continuous X. The local asymptotically-optimal design here utilized the true parameters that generate the data and is therefore not practically feasible, while the adaptive approaches attempted to approximate this asymptotically-optimal design by estimating necessary components using phase-I and phase-IIa data. Figures 1 and 2 display the differences in sampling proportions between the asymptotically-optimal design and the adaptive designs that resulted from both parametrically and empirically estimating $\mathcal{A}$, $\text{var}_{X|Y,V}[U_\beta(Y_i|X_i, V_i)]$, and $P(Y, V)$ when using proportional stratified sampling at phase-IIa. The corresponding figures for the designs employing balanced sampling in phase-IIa are presented in Figures 3 and 4.

The empirical relative efficiencies and empirical coverage probabilities of the estimators resulting from the different designs are presented in Table 1. The empirical relative efficiency (ERE) of design A was calculated as $\text{evar}_0/\text{evar}_A \cdot 100\%$, where evar_A is the empirical variance of the estimator for β_x arising from design A and evar_0 is the corresponding empirical variance under the asymptotically-optimal design which requires specification of the unknown parameter values. This definition is analogous with that of Fedorov et al. [18] who examined efficiency in the context of optimal design for dose finding experiments. As an additional measure of the variation of estimators over the 2000 simulations, we also present in Table 1 the empirical relative interquartile range (ERI) calculated as $\text{IQR}_0/\text{IQR}_A \cdot 100\%$, where IQR_A is the observed interquartile range among the estimates for β_x arising from design A and IQR_0 is defined analogously for the asymptotically-optimal design. Values of ERE or ERI less than 100% reflect a loss of precision relative to the optimal design and values greater than 100% indicate greater local precision than realized with the asymptotically-optimal design using the true parameter values. Empirical coverage probabilities (ECPs) show the percentage of the simulations in which the true β_x is contained in 95% confidence intervals, which are constructed using sample estimates of the asymptotic variance in (3) upon completion of phase-II sampling.

4.2. Empirical Findings from Simulation Studies

The adaptive designs (i.e. those that selected some individuals in phase-IIb) resulted in sampling fractions which were much closer to the optimal designs on average than were the non-adaptive sampling designs (see Figures 1-2 and, similarly, Figures 3-4). However, the adaptive designs generally displayed greater variability in the sampling rates between simulations. As the fraction of individuals selected in phase-IIb increased: (i) the average two-stage adaptive designs became closer to the asymptotically-optimal design, and (ii) the between-simulation variability in the two-stage adaptive designs increased. Selecting more individuals at phase-IIa allows for greater precision in estimating design components and, therefore, more stable designs. With more individuals selected at phase-IIa, however, fewer individuals can be selected according to the approximately optimal design at phase-IIb. This demonstrates a trade-off between precision and accuracy in the two-stage adaptive design when determining the fraction of individuals to select at each sub-phase of the phase-II sampling. The best properties of the adaptive sampling design (in terms of designs that are centred near the asymptotically-optimal design and have small between-simulation variability) resulted from balancing the number of phase-II individuals selected in the internal pilot study (phase-IIa) and the number of individuals that were selected according to the approximately optimal design (phase-IIb).

In the setting where X was binary, there was little difference between the adaptive approaches that resulted from estimating design components parametrically and those arising from estimating design components empirically either in terms of the designs themselves (see Figures 1 and 3) or in terms of observed efficiency (see Table 1). However, in the setting with a continuous expensive covariate X, the greater efficiency of the parametric estimation of the design components translated into much less variability in the two-stage adaptive design (see Figures 2 and 4) and, in turn, much greater efficiency in the estimator of interest (Table 1). In fact, when considering a continuous expensive covariate, the adaptive designs based on empirical estimation of design components offered no consistent efficiency gains over the non-adaptive sampling design; this was true for both two-stage adaptive and fully adaptive sampling designs based on empirical estimates of design components and this result held both in terms of direct measures of empirical relative efficiency (ERE) and in terms of the relative size of interquartile ranges (ERI), which is an alternate measure of the spread of resulting estimates that gives less weight to outliers. The adaptive designs (both two-stage and fully adaptive) based on parametric estimation of design components, on the other hand, consistently performed significantly better than the non-adaptive designs in terms of the efficiency achieved for the estimator of interest. In terms of the relative size of interquartile range, the parametric adaptive designs generally offered improvements over the non-adaptive designs, however, the non-adaptive balanced sampling design was very close to the asymptotically-optimal design for binary X and was not consistently improved by the adaptive procedures. The efficiency of the adaptive designs was often quite similar to the efficiency achieved by the local asymptotically-optimal sampling design, which of course cannot be implemented in practice. In fact, the adaptive designs often exceeded the asymptotically-optimal designs in terms of efficiency; we note that this apparently contradictory result is possible here because the benchmark designs, while obviously very efficient in these finite-sample settings, are only truly optimal as sample sizes tend to infinity.

In all cases, the empirical coverage probabilities were compatible with the nominal 95% level, which indicates that the estimated asymptotic standard errors for the adaptive design closely tracked the empirical standard errors. Therefore, the regular variance estimators worked well for these adaptive designs, which achieved efficiency that was very similar to the asymptotically-optimal design.

The non-adaptive balanced design was more efficient than the non-adaptive proportional stratified sampling design, but the non-adaptive designs were generally not as efficient as their adaptive counterparts. The fully adaptive designs based on parametric estimation of the design components generally led to the most precise estimates of effects. However, the two-stage adaptive designs are easier to implement in practice and were also very efficient when based on parametric estimation of design components. These two-stage adaptive designs were particularly effective when balancing the number of individuals selected in each of the two sub-phases of sampling and when using balanced sampling design for the internal pilot study. Although not presented here, these adaptive designs also generally resulted in an increase in efficiency for estimation of β_0 over non-adaptive designs; there was little impact on the efficiency for estimation of β_v.

5. Concluding Remarks

We presented an adaptive approach to selecting individuals for measurement of an expensive covariate in a two-phase sampling design. This extension to Reilly and Pepe's [12] optimal mean-score sampling design allows for explicit specification of near-optimal designs without *a priori* knowledge of the parameters and without the necessity for costly external pilot studies. These adaptive procedures are particularly important for elaborate response models where many parameters must be specified in order to derive the optimal design. The efficiencies of estimators under adaptive sampling based on parametric estimation of design components were generally very similar to those found under the theoretical optimal design, which can not be implemented in practice. Previous studies have shown that the optimal mean score design that is being emulated by this adaptive procedure has desirable properties even when analyses are to be conducted using other methods [3, 13]. Additionally, there was no evidence of bias among the adaptive sampling designs and the empirical coverage probabilities were all compatible with the nominal levels.

Balanced sampling designs, as advocated by Breslow and colleagues [5,9], were more efficient than proportional stratified sampling, but these non-adaptive designs could not achieve the same efficiency as the adaptive approaches. The fully adaptive designs could be very efficient, but would be difficult to implement in practice. The two-stage adaptive procedure may be the most useful approach as it often offered near-optimal efficiency and would be relatively easy to implement in practice as it does not require re-estimation of the design components following the selection and response ascertainment of each individual. These findings of good efficiency and practical appeal in using a small number of phases as opposed to a fully adaptive sampling design are similar to the results seen in the setting of optimal dose-finding designs [18] and are analogous to the appeal of group sequential monitoring over sequential monitoring in clinical trials [32].

Other simulation studies showed very similar results and so have not been presented in depth here. Simulations involving clustered response vectors (based on [14]) similarly showed that the adaptive design achieved levels of efficiency on a par with the true optimal design; in those simulations, however, the non-adaptive balanced design was not nearly as efficient, so the benefit of the adaptive design was even more obvious.

One important finding here is that these two-stage adaptive designs should be based on parametric assumptions about necessary design components, rather than less precise empirical estimation. Focusing on parametric estimation at the design stage is reasonable as our goal is to employ the estimator from the most efficient design possible. This is especially true given that model misspecification at the design stage will not affect the consistency of the robust mean-score estimation procedure. Additionally, the efficiency resulting from asymptotically optimal mean-score designs has been shown to be quite robust to model misspecification at the design stage [13].

There is also a trade-off between the potential efficiency of a large phase-IIb sample and the decreased precision of estimates of the design parameter values when phase-IIa is small in two-stage adaptive sampling. The two-stage adaptive designs that collected nearly equal numbers of individuals at phase-IIa and at phase-IIb were generally quite effective at generating near-optimal sampling fractions that were very efficient in finite samples.

We have focussed here on the setting of a two-phase design in which phase-I data are available and one wishes to optimally select a fixed number of individuals for measurement of an expensive covariate during the second phase; this is precisely the setting of the motivating biomarker study. The simulations used were based on parameters derived from analyses based on a pilot study consisting of complete data from 20 arthritis patients with (Y,V) = (0,0), 7 patients with (Y,V) = (0,1), 17 patients with (Y,V) = (1,0), and 9 patients with (Y,V) = (1,1). If these data are taken as an internal pilot sub-sample (i.e. as a phase-IIa sample) in this setting where $N_{00}/N = 0.39, N_{01}/N = 0.12, N_{10}/N = 0.28$, and $N_{11}/N = 0.21$, then the two-stage adaptive sampling strategy would recommend the phase-IIb design $(n_{00}^{(b,\text{opt})}, n_{01}^{(b,\text{opt})}, n_{10}^{(b,\text{opt})}, n_{11}^{(b,\text{opt})}) = (17, 40, 36, 54)$ in order to select the allotted 200 individuals for efficient estimation of β_x when MMP-3 measurements are to be dichotomized; if MMP-3 is to be considered continuously, then the optimal phase-IIb design is $(n_{00}^{(b,\text{opt})}, n_{01}^{(b,\text{opt})}, n_{10}^{(b,\text{opt})}, n_{11}^{(b,\text{opt})}) = (35, 42, 32, 38)$ for selection of the remaining 147 patients. Note that in order to achieve a balanced or proportional stratified sampling design, we would use $(n_{00}^{(b)}, n_{01}^{(b)}, n_{10}^{(b)}, n_{11}^{(b)}) = (30, 33, 43, 41)$ and $(n_{00}^{(b)}, n_{01}^{(b)}, n_{10}^{(b)}, n_{11}^{(b)}) = (58, 40, 17, 32)$, respectively, regardless of whether or not MMP-3 is to be dichotomized. The efficiency gained through this two-stage adaptive sampling strategy will depend on the accuracy of the parameters estimated at phase-IIa.

More generally, we recommend that the phase-II sampling be carried out adaptively in two stages as follows:

(i) with a fixed number of individuals to be sampled in phase-II, half of them should be sampled using balanced stratified sampling with strata defined based on the phase-I data,
(ii) the remaining individuals to be sampled in phase-II should be chosen to satisfy the approximately optimal stratum-specific sampling sizes derived using (6) with available data being used in parametric estimation of necessary design components.

This approach would not be much more complex to implement than the non-adaptive two-phase designs in the motivating biomarker study, but it would lead to efficiency gains in the range of 5%-20% when estimating the parameter of interest.

In settings different from the one we considered, it could be that the phase-I or phase-II sample sizes are not fixed in advance; further study of such settings is required to determine the best strategies for the use of adaptive two-phase designs.

Acknowledgements

This research was supported by an Alexander Graham Bell Canada Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) to MAM and a Discovery Grant from NSERC to RJC (RGPIN 155849) and a grant to RJC from the Canadian Institutes for Health Research (FRN 13887). Richard Cook is a Canada Research Chair in Statistical Methods for Health Research.

References

1. Neyman J. Contribution to the theory of sampling from human populations. *Journal of the American Statistical Association* 1938; **33**:101–116.
2. Pickles A, Dunn G, Vazquez-Barquero JL. Screening for stratification in two-phase ('two- stage') epidemiological surveys. *Statistical Methods in Medical Research* 1995; **4**:73–89.
3. Whittemore AS, Halpern J. Multi-stage sampling in genetic epidemiology. *Statistics in Medicine* 1997; **16**:153–167.
4. Chatterjee N, Chen Y, Breslow NE. A pseudoscore estimator for regression problems with two-phase sampling. *Journal of the American Statistical Association* 2003; **98**(461):158–168.
5. Breslow NE, Cain KC. Logistic regression for two-stage case-control data. *Biometrika* 1988; **75**(1):11–20.
6. Scott A, Wild C. Fitting logistic regression models in stratified case-control studies. *Biometrics* 1991; 497–510.
7. Clayton D, Spielgelhalter D, Dunn G, Pickles A. Analysis of longitudinal binary data from multi-phase sampling. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 1998; **60**(1):71–87.
8. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 1994; **89**(427):846–866.
9. Breslow NE, Chatterjee N. Design and analysis of two-phase studies with binary outcome applied to Wilm's tumour prognosis. *Applied Statistics* 1999; **48**(4):457–468.
10. Reilly M. Optimal sampling strategies for two phase studies. *American Journal of Epidemiology* 1996; **143**:92–100.
11. Schaubel D, Hanley J, Collet JP, Boivin JF, Sharpe C, Morrison HI, Mao Y. Two-stage sampling for etiologic studies sample size and power. *American Journal Of Epidemiology* 1997; **146**(5):450–458.
12. Reilly M, Pepe MS. A mean score method for missing and auxiliary covariate data in regression models. *Biometrika* 1995; **82**(2):299–314.
13. McIsaac MA, Cook RJ. Response-dependent two-phase sampling designs for biomarker studies. *Canadian Journal of Statistics* 2014; **42**(2):268–284.
14. McIsaac MA, Cook RJ. Response-dependent sampling with clustered and longitudinal data. In Sutradhar B, ed., *ISS-2012 Proceedings Volume On Longitudinal Data Analysis Subject To Measurement Errors, Missing Values, And/Or Outliers*, 157–181. Springer, 2013; .
15. Pepe MS, Reilly M, Fleming TR. Auxiliary outcome data and the mean-score method. *Journal of Statical Planning and Inference* 1994; **42**:137–160.
16. Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Statistics in Medicine* 1990; **9**(1-2):65–72.
17. Lohr SL. Accurate multivariate estimation using triple sampling. *The Annals of Statistics* 1990; **18**(4):1615–1633.
18. Fedorov V, Wu Y, Zhang R. Optimal dose-finding designs with correlated continuous and discrete responses. *Statistics in Medicine* 2012; **31**(3):217–234.
19. Chandran V, Cook RJ, Edwin J, Shen H, Pellett FJ, Shanmugarajah S, Rosen CF, Gladman DD. Soluble biomarkers differentiate patients with psoriatic arthritis from those with psoriasis without arthritis. *Rheumatology* 2010; **49**(7):1399–1405.
20. Gladman DD, Chandran V. Observation cohort studies: Lessons learnt from the University of Toronto Psoriatic Arthritis Program. *Rheumatology* 2011; **50**:25–31.
21. Gottlieb A, Korman NJ, Gordon KB, Feldman SR, Lebwohl M, Koo JY, . . ., Menter A. Guidelines of care for the management of psoriasis and psoriatic arthritis: Section 2. psoriatic arthritis: Overview and guidelines of care for treatment with an emphasis on the biologics. *Journal of the American Academy of Dermatology* 2008; **58**(5):851–886.
22. Ruderman EM. Evaluation and management of psoriatic arthritis: The role of biologic therapy. *Journal of the American Academy of Dermatology* 2003; **49**(2):125–132.
23. Little RJA, Rubin DB. *Statistical Analysis with Missing Data, Second Edition*. John Wiley & Sons, New York, 2002.
24. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B (Methodological)* 1977; **39**(1):1–38.
25. Lawless JF, Kalbfleisch JD, Wild CJ. Semiparametric methods for response-selective and missing data problems in regression. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 1999; **61**(2):413–438.
26. Lawless JF. Likelihood and pseudo likelihood estimation based on response-biased observations. In Basawa IS, Godambe VP, Taylor RL, eds., *Studies in Item Analysis and Prediction*, 43–55. Institute of Mathematical Statistics, 1997; .
27. Wild CJ. Fitting prospective regression models to case-control data. *Biometrika* 1991; **78**:705–717.
28. Emery A, Nenarokomov AV. Optimal experiment design. *Measurement Science and Technology* 1999; **9**(6):864.
29. Lai TL. Sequential analysis: Some classical problems and new challenges. *Statistica Sinica* 2001; **11**(2):303–350.
30. Rosenberger WF, Hughes-Oliver JM. Inference from a sequential design: Proof of a conjecture by Ford and Silvey. *Statistics & Probability Letters* 1999; **44**(2):177–180.
31. McIsaac MA, Cook RJ. Biomarkers for disease progression in rheumatology: A review and empirical study of two-phase designs. Research Paper Series 2013-03, University of Waterloo, 2013.

32. Pocock SJ. Interim analyses for randomized clinical trials: The group sequential approach. *Biometrics* 1982; 153–162.

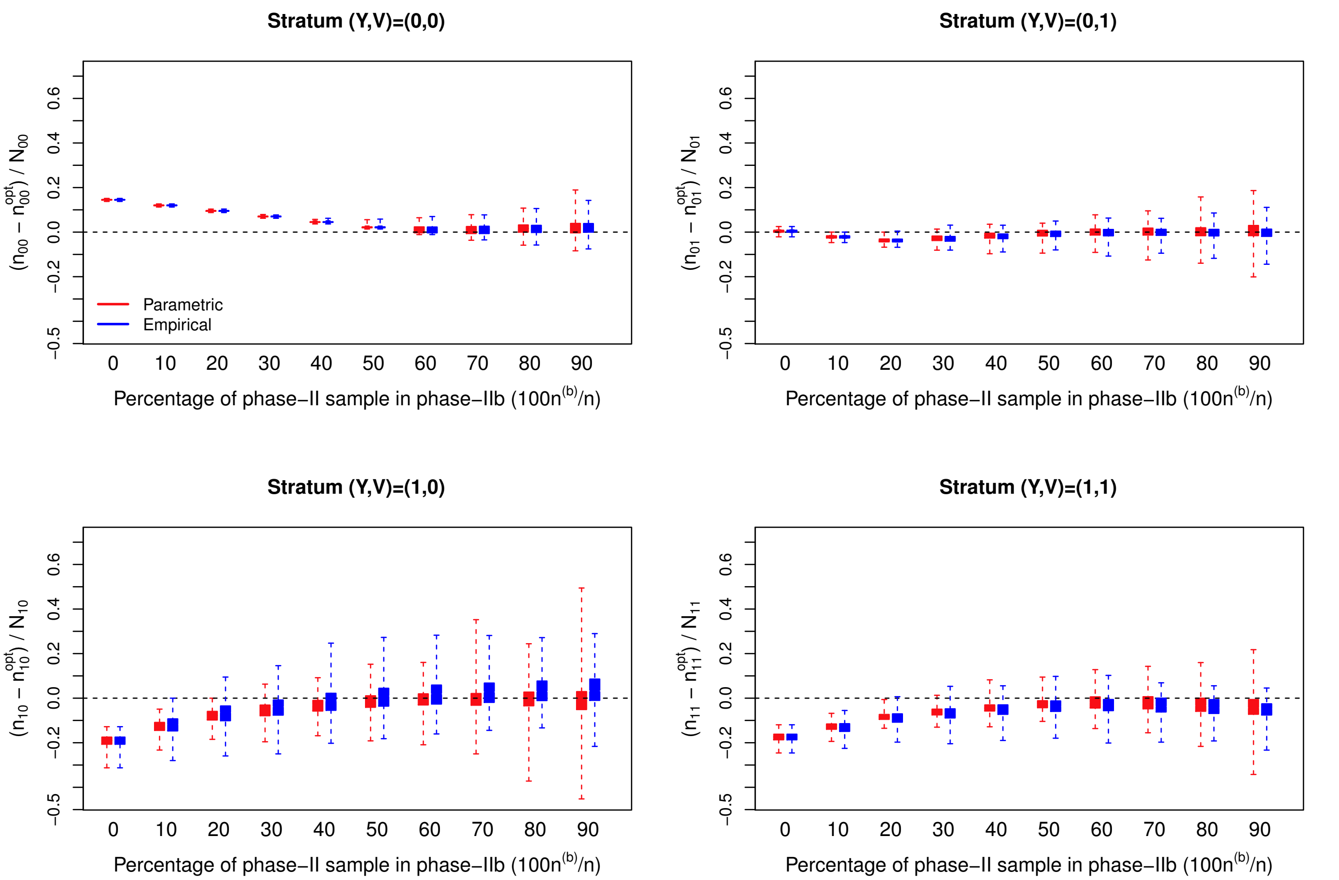

Figure 1. Empirical differences in sampling fractions between the asymptotically-optimal design and the two-stage adaptive designs employing proportional stratified sampling in phase-IIa according to the percentage of the phase-II sample that is selected at phase-IIb ($n^{(b)}/n$) for a binary expensive covariate; $N = 800$; $n = 200$; $(\beta_0, \beta_x, \beta_v) = (-1.95, 1.00, 0.90)$, $(\alpha_0, \alpha_v) = (1.05, -0.41)$, and $\gamma_0 = -0.04$; n_{YV} represents the number of individuals selected for the measurement of X from the available N_{YV} individuals in stratum (Y, V) in the two-stage adaptive design, while n_{YV}^{opt} represents the corresponding sample size for the asymptotically-optimal design. This two-stage adaptive design was derived for selecting individuals for measurement of a binary X and relied on either parametric or empirical estimation of design components from data collected in phase-I and phase-IIa.

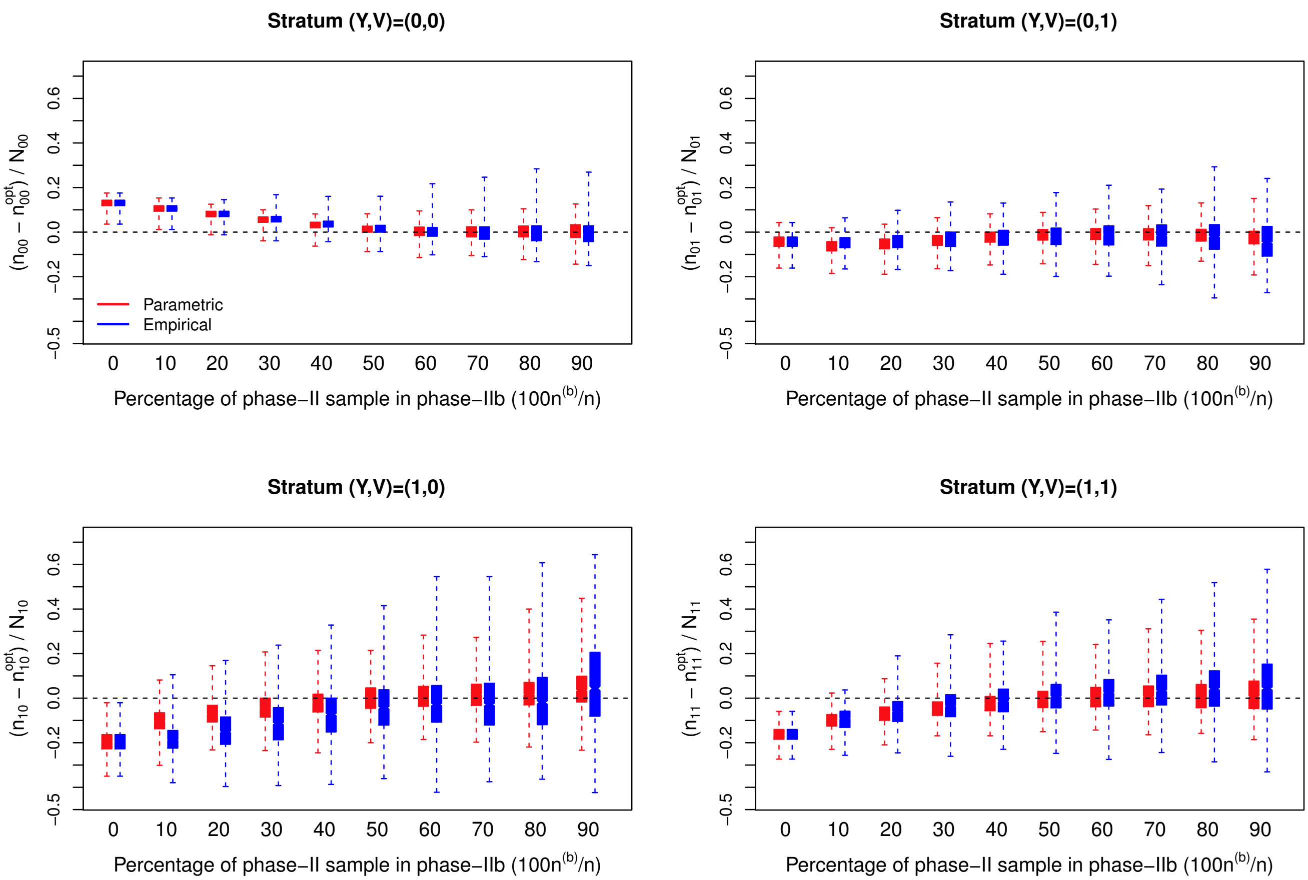

Figure 2. Empirical differences in sampling fractions between the asymptotically-optimal design and the two-stage adaptive designs employing proportional stratified sampling in phase-IIa according to the percentage of the phase-II sample that is selected at phase-IIb ($n^{(b)}/n$) for a continuous expensive covariate; $N = 800$; $n = 200$; $(\beta_0, \beta_x, \beta_v) = (-2.18, 0.03, .84)$, $(\alpha_0, \alpha_1, \alpha_v) = (1.40, 10, 5)$, and $\gamma_0 = -0.04$. n_{YV} represents the number of individuals selected for the measurement of X from the available N_{YV} individuals in stratum (Y, V) in the two-stage adaptive design, while n_{YV}^{opt} represents the corresponding sample size for the asymptotically-optimal design. This two-stage adaptive design was derived for selecting individuals for measurement of a continuous X and relied on either parametric or empirical estimation of design components from data collected in phase-I and phase-IIa.

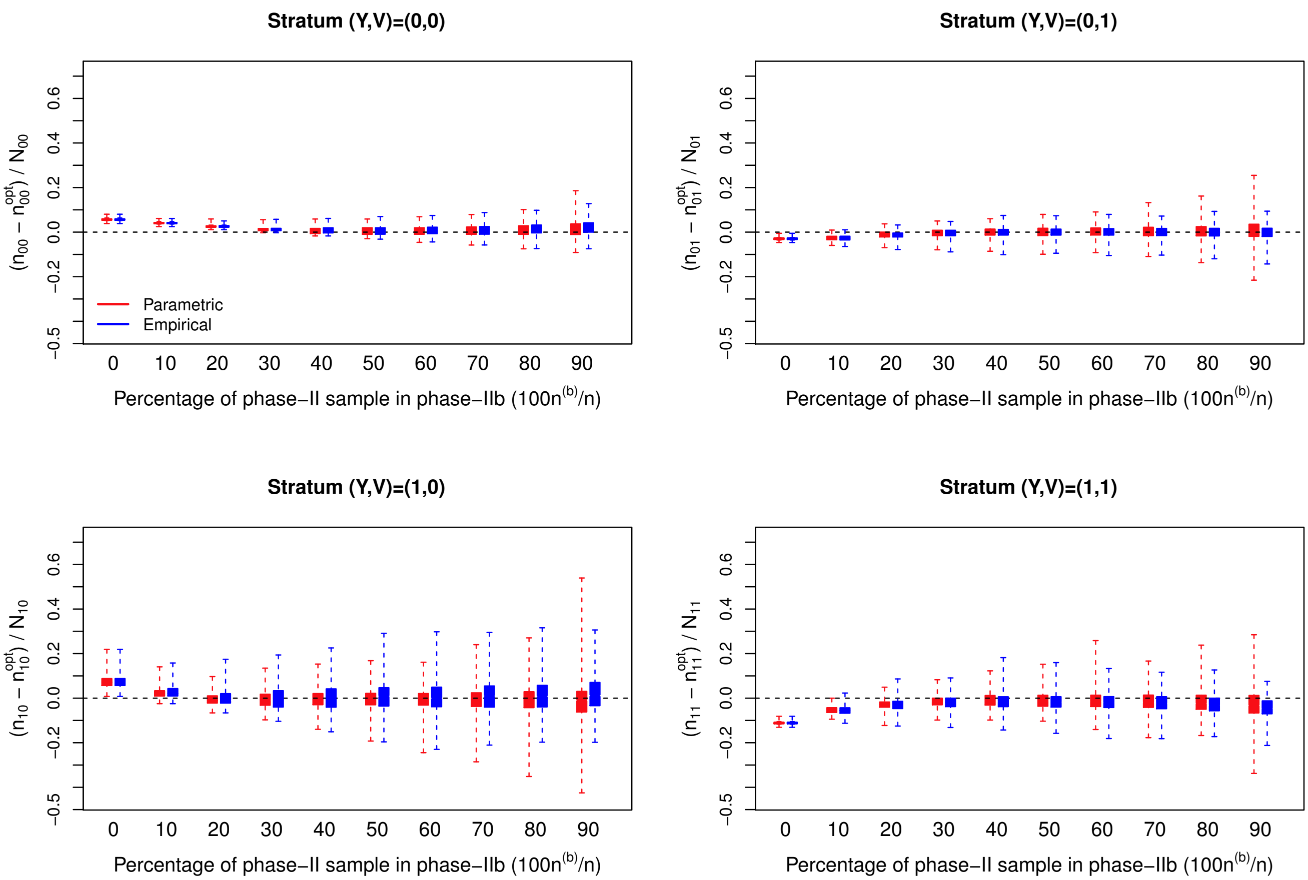

Figure 3. Empirical differences in sampling fractions between the asymptotically-optimal design and the two-stage adaptive designs employing balanced sampling in phase-IIa according to the percentage of the phase-II sample that is selected at phase-IIb $(n^{(b)}/n)$ for a binary expensive covariate; $N = 800$; $n = 200$; $(\beta_0, \beta_x, \beta_v) = (-1.95, 1.00, 0.90)$, $(\alpha_0, \alpha_v) = (1.05, -0.41)$, and $\gamma_0 = -0.04$; n_{YV} represents the number of individuals selected for the measurement of X from the available N_{YV} individuals in stratum (Y, V) in the two-stage adaptive design, while n_{YV}^{opt} represents the corresponding sample size for the asymptotically-optimal design. This two-stage adaptive design was derived for selecting individuals for measurement of a binary X and relied on either parametric or empirical estimation of design components from data collected in phase-I and phase-IIa.

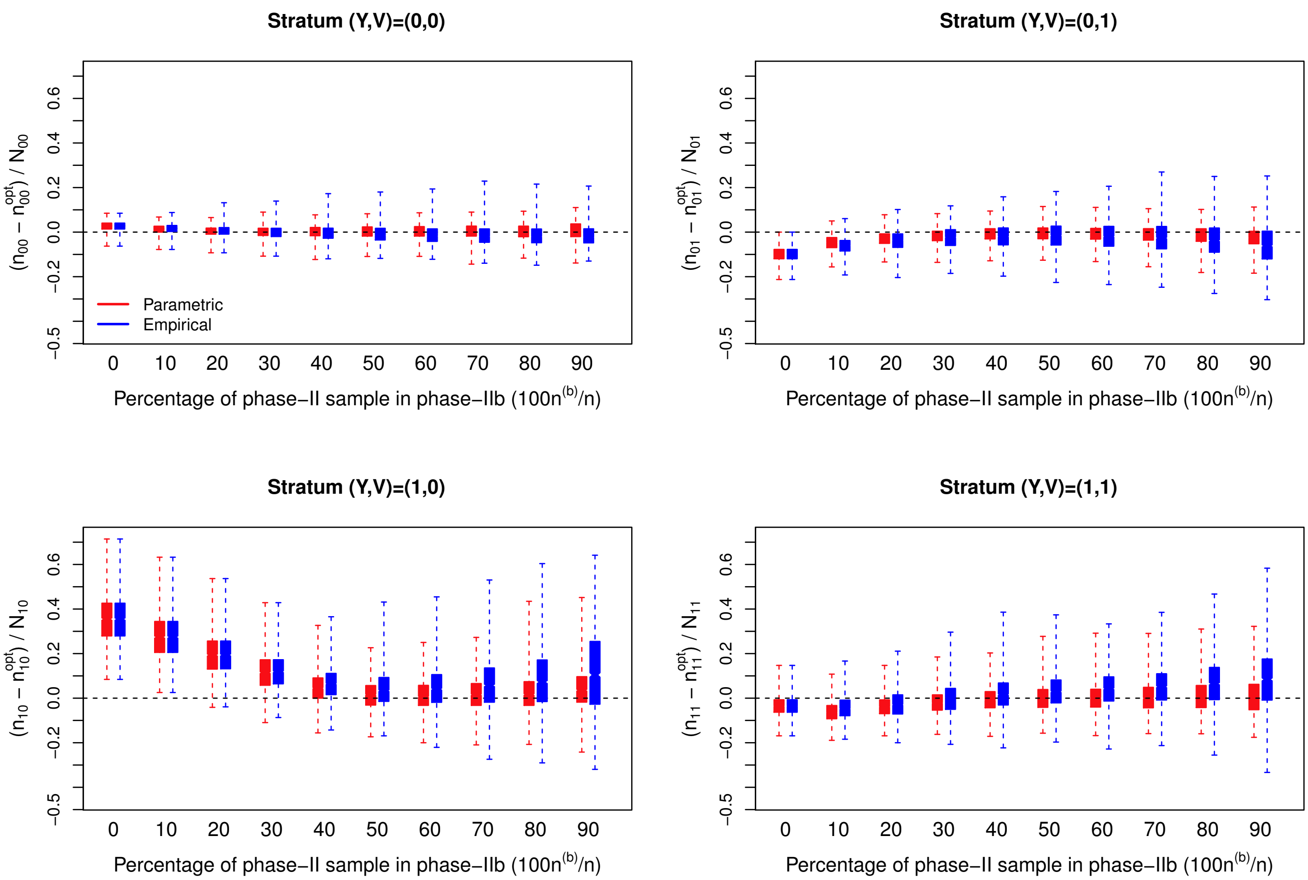

Figure 4. Empirical differences in sampling fractions between the asymptotically-optimal design and the two-stage adaptive designs employing balanced sampling in phase-IIa according to the percentage of the phase-II sample that is selected at phase-IIb ($n^{(b)}/n$) for a continuous expensive covariate; $N = 800$; $n = 200$; $(\beta_0, \beta_x, \beta_v) = (-2.18, 0.03, .84)$, $(\alpha_0, \alpha_1, \alpha_v) = (1.40, 10, 5)$, and $\gamma_0 = -0.04$. n_{YV} represents the number of individuals selected for the measurement of X from the available N_{YV} individuals in stratum (Y, V) in the two-stage adaptive design, while n_{YV}^{opt} represents the corresponding sample size for the asymptotically-optimal design. This two-stage adaptive design was derived for selecting individuals for measurement of a continuous X and relied on either parametric or empirical estimation of design components from data collected in phase-I and phase-IIa.

Table 1. Empirical relative efficiencies (*ERE*) and empirical relative interquartile ranges (ERI) compared to the asymptotically-optimal design based on true (unknown) parameters and empirical coverage probabilities (*ECP*) of estimators for β_x based on 2000 simulated datasets with $N = 800$ and $n = 200$.

	Binary X						Continuous X					
	Parametric			Empirical			Parametric			Empirical		
$100n^{(b)}/n$	ERE	ERI	ECP	ERE	ERI	ECP	ERE	ERI	ECP	ERE	ERI	ECP
TWO-STAGE ADAPTIVE – PROPORTIONAL SAMPLING IN PHASE-IIA												
0‡	76.6	88.7	95.5	76.6	88.7	95.5	88.9	94.4	95.2	88.9	94.4	95.2
10	85.0	93.4	95.3	84.3	90.2	95.2	94.7	97.7	94.8	89.4	97.9	94.8
20	86.3	89.9	94.7	84.6	90.9	94.8	94.0	96.9	94.7	87.8	92.2	94.0
30	94.0	97.8	95.2	89.3	97.9	93.8	102.6	99.8	95.2	93.2	97.1	94.6
40	96.3	96.3	95.0	103.1	100.8	95.4	100.7	104.3	94.8	86.0	91.1	93.2
50	98.2	95.0	94.8	92.6	96.3	94.2	103.2	102.4	95.1	88.5	94.7	93.3
60	103.1	100.0	95.2	95.2	96.5	94.8	105.0	102.0	95.5	88.2	97.0	93.7
70	97.2	98.2	95.0	95.4	98.8	94.4	103.0	103.3	94.6	93.9	96.9	94.8
80	100.7	103.1	95.1	100.8	101.3	95.0	107.9	105.2	95.6	85.6	96.6	94.5
90	98.6	102.2	95.4	97.7	97.5	94.5	103.7	103.8	95.0	82.8	94.0	94.7
TWO-STAGE ADAPTIVE – BALANCED SAMPLING IN PHASE-IIA												
0‡	95.2	99.6	95.2	95.2	99.6	95.2	90.6	90.6	95.0	90.6	90.6	95.0
10	101.6	102.3	95.3	102.9	102.5	95.5	98.3	98.6	94.6	95.2	99.2	95.0
20	96.8	101.5	93.8	98.4	98.9	95.0	102.0	99.4	94.7	95.9	99.8	94.6
30	105.2	103.7	95.3	99.9	100.7	95.0	107.5	103.1	95.9	88.2	92.2	93.8
40	100.1	103.2	95.0	102.5	98.4	94.9	109.7	105.4	95.5	93.9	95.3	94.0
50	103.8	99.3	95.3	100.0	100.1	94.8	101.0	101.8	95.0	96.7	95.5	95.5
60	99.8	97.6	95.0	98.9	96.4	95.2	100.8	98.3	94.8	86.5	90.8	94.2
70	103.7	98.9	95.5	96.6	98.5	94.3	103.2	103.6	95.0	85.0	90.7	93.8
80	100.2	101.1	94.7	97.7	98.0	94.7	103.9	104.0	95.2	86.6	91.5	94.8
90	95.8	98.5	95.2	96.0	96.2	94.8	102.3	100.6	95.2	84.0	95.9	94.8
FULLY ADAPTIVE												
	102.4	107.0	95.3	99.0	96.7	94.5	102.4	101.9	95.3	86.8	89.1	94.3

The parameters were set to $(\beta_0, \beta_x, \beta_v) = (-1.95, 1.00, 0.90)$, $(\alpha_0, \alpha_v) = (1.05, -0.41)$, and $\gamma_0 = -0.04$ for the case with binary X and to $(\beta_0, \beta_x, \beta_v) = (-2.18, 0.03, .84)$, $(\alpha_0, \alpha_1, \alpha_v) = (1.40, 10, 5)$, and $\gamma_0 = -0.04$ for the setting with continuous X.

‡ The non-adaptive design does not require estimation of the design components, so there is no distinction between design components being estimated "empirically" or "parametrically".

Two-stage adaptive designs select $n - n^{(b)}$ individuals using proportional or balanced sampling and use these individuals to estimate the design components either through parametric estimation (*Parametric*) or through empirical estimation (*Empirical*) to approximate optimal selection of the remaining $n^{(b)}$ individuals. Non-adaptive designs are a special case of the two-stage sampling where all individuals are selected in phase-IIa. Fully adaptive designs involved selecting an initial balanced sample of size 40 (corresponding to 20% of n) and then selecting the remaining individuals one at a time while updating estimates of the design component after each individual is selected.