Journal of Statistical Research *https://doi.org/10.3329/jsr.v56i1.63944*
2022, Vol. 56, No. , pp. 11-36 *ISSN 0256 - 422 X*

THREE INFLUENTIAL DESIGN QUANTITIES ON THE POWER OF WALD-TYPE TESTS FOR TREATMENT COMPARISONS IN CLINICAL TRIALS

SELVAKKADUNKO SELVARATNAM

Department of Statistical Sciences, University of Toronto, Toronto, Canada
Email: selva.selvaratnam@utoronto.ca

ALWELL OYET*

Department of Mathematics and Statistics
Memorial University of Newfoundland, St. John's, NL, Canada
Email: aoyet@mun.ca

YANQING YI

Department of Community Health and Humanities
Memorial University of Newfoundland, St. John's, NL, Canada
Email: Yanqing.Yi@med.mun.ca

SUMMARY

In clinical trials, efficient statistical inference is critical to the well-being of future patients. We therefore construct Wald-type tests for the hypothesis of treatment-by-covariate interaction when treatments are assigned to patients by an adaptive design and the true model is a generalized linear model. Our measure of efficiency is the power of the test while ethics of a trial or well-being of participating patients is measured by the success rate of treatments. We demonstrate that the power of the test depends on the target allocation proportion, the bias of the randomization procedure from the target, and the variability induced by the randomization process (design variability) for adaptive designs. We prove that these quantities influence the power when the trial involves two treatments and a single covariate. We also show that, in this case, as design variability decreases the power increases. Due to the complexity of the problem, we demonstrate by simulation that this result still holds when more than one covariate is present in the model. In simulation studies, we compare the measures of efficiency and ethics under response-adaptive (RA), covariate-adjusted response-adaptive (CARA), and completely randomized (CR) designs. The methods are applied to data from a clinical trial on stroke prevention in atrial fibrillation (SPAF).

Keywords and phrases: Completely randomized design; covariate-adjusted response-adaptive designs; response-adaptive designs; statistical inference

AMS Classification: Primary 62L05; secondary 62F03, 62F05

* Corresponding author

1 Introduction

In clinical trials, researchers are often interested in comparing the effectiveness of treatments. They are also interested in adopting treatment assignment procedures which take into account the ethics of a trial or the well-being of participating patients. As a result, various adaptive designs for treatment assignments which satisfy the ethical principles of clinical trials have been developed and found to be useful alternatives to completely randomized designs. See for instance, Selvaratnam, Oyet, Yi and Gadag (2017) for a recent discussion of examples of adaptive designs such as response adaptive (RA), covariate adaptive (CA) and covariate-adjusted response adaptive (CARA) designs. Since patients react differently to a given treatment, it has become increasingly important to also account for the effect of certain characteristics or covariates of individual patients as well as treatment-by-covariate interaction on the response of patients during treatment comparisons. Ma, Hu and Zhang (2015) had established a theoretical foundation for hypothesis testing of parameters in linear models under a large class of covariate adaptive (CA) designs, which includes marginal methods and stratified permuted block design (Pocock and Simon, 1975). Although Ma, Hu and Zhang (2015) used all the important covariates at the design stage, they dropped some covariate information in the final stage of statistical inference. This will however lead to estimators of parameters that are generally inconsistent and biased. This will also affect the derivation of the distribution of the test statistic. In §2 of this paper, we will develop a theoretical foundation for hypotheses testing for a general class of adaptive designs when responses are binary.

Furthermore, we extend the results of Ma, Hu and Zhang (2015) and Selvaratnam, Yi and Oyet (2019) by constructing Wald-type test statistics for, (a) examining whether the effect of treatment-by-covariate interaction on the response is significant; and (b) comparing two treatments, say treatment A (new treatment) and treatment B (existing treatment), under a general class of adaptive designs when the true model is a GLM with treatment-by-covariate interaction. Let β_{A0} be the true effect of treatment A compared to treatment B. An experimenter may test the significance of the overall effect of treatment A compared to treatment B through the hypotheses

$$H_0 : \beta_{A0} = 0 \quad \text{vs} \quad H_A : \beta_{A0} > 0, \tag{1.1}$$

where H_0 and H_A are the null and alternative hypothesis, respectively. Suppose the experimenter decides to reject H_0, when in fact H_0 is true. The experimenter may decide to recommend the new treatment, though the new treatment does not lead to any improvement in the responses of patients compared to the existing treatment. This decision error is commonly referred to as Type 1 error. For the test we have constructed to be reliable in practice, it is important for it to be able to control the size of the test or the chance of committing this error. Thus, we will examine the size of the test through extensive simulation studies in §4 of this paper.

Concerning the power of a test, there are three main approaches, in the literature for power analysis. Exact methods use formulas which directly express power in terms of model parameters for power analysis. When exact formulas cannot be found due to the complicated nature of a problem, it may be possible to derive approximate or large sample approximations for power analysis. An approximation method for power analysis computation in multiple logistic regression was first introduced by Whittemore (1981). However, this approach is only suitable for binary responses with

rare events such as disease or death with single or multivariate covariates whose joint probability distribution must belong to the exponential family. Later, Self and Mauritsen (1988) developed an approximation procedure for power analysis based on the noncentrality parameter of the asymptotic χ^2 distribution of a score test statistic for testing the parameters in a GLM. They implemented their approach to categorical covariates with a finite number of distinct covariate configurations. They noted two potential problems with their approximation but observed that the problems are not likely to be serious for alternatives that are close to the null hypothesis. The problem with this scenario is that it may be difficult for a test to identify any significant differences between the null and alternative if they are close. Instead of power analysis for GLMs based on a score test, Self, Mauritsen and Ohara (1992) established a tool for power computation based on a likelihood ratio test statistic. Through simulation studies, they found that the analysis based on the likelihood ratio test was much easier to implement and also to be more accurate over a wider range of parameter values, than the approach based on the score test. Shieh (2000) carried out a simulation study to compare the method of Whittemore (1981) and Self, Mauritsen and Ohara (1992) with various combination of response probabilities and covariate distribution in logistic regression models. They found that the method of Self, Mauritsen and Ohara (1992) outperformed the Whittemore (1981) approach in a variety of situations. Later, Shieh (2005) developed an approach for power and sample size calculations based on the discrepancy between the noncentral and central χ^2 approximation of the distribution of a Wald-type statistic in GLMs. They discussed examples in logistic and Poisson regression. In their simulation studies they found that their method maintained a close agreement with the method of Shieh (2000).

In this paper, we have derived expressions for the power of the tests for interaction and main effects in §3.1 and §3.2, respectively, based on the noncentrality parameters of the asymptotic distributions of the Wald-type test statistics we have proposed. We observe that the design criteria used in data collection influence the power of tests through the noncentrality parameter (Hu and Rosenberger, 2003). Hu and Rosenberger (2003) identified three major influence factors: (i) the target allocation proportion, (ii) the randomization bias from target proportion, and (iii) the variance of randomization from target proportion. However, they derived an expression for the noncentrality parameter under the assumption of a simple homogeneous parametric structure. In this paper, we relax this assumption in deriving the expression for the noncentrality parameter. We also derive a concave relationship between the noncentrality parameter and target allocation proportion when a covariate is present in the logit model in §3.3.

2 Wald-type Tests and Asymptotic Distributions for Adaptive Designs

In this section, we assume that when a new patient enters a clinical trial, they are assigned to one of two possible treatments denoted by A or B based on their covariate profile and the outcome, treatment assignments, and covariate profile of previous patients. Suppose that following this process, binary responses and the treatment assignments for n patients have been observed and the $(n+1)$th patient is ready to enter the trial. Suppose further that K categorical covariates, say $u_1, u_2, \ldots, u_K$,

associated with the responses were also collected for each patient in the clinical trial. Let the number of levels for each covariate, u_k be L_k+1, say $u_{k0}, u_{k1}, \ldots, u_{k,L_k}$, for $k = 1, 2, \ldots, K$. Then, based on the reference category, say u_{k0}, $k = 1, 2, \ldots, K$ one can define a set of L_k dummy variables for each covariate. Without loss of generality, let $\mathbf{z}_i' = (z_{i1}, \ldots, z_{ip})$ be the $p = \sum_{k=1}^{K} L_k$ dimensional vector of dummy variables corresponding to the covariates of the ith patient, $i = 1, 2, \ldots, n$. Clearly, the covariate information $\mathbf{z}_i' = (z_{i1}, \ldots, z_{ip})$ of the ith patient is then known and also a binary vector variable. We assume that if u_k is continuous, the possible values of u_k can be grouped into a user defined categories by the experimenter. Let Y_i be the binary response of the ith, $i = 1, 2, \ldots, n$ patient, where $Y_i = 1$, if the treatment was a success and $Y_i = 0$, otherwise. Let $X_{iA} = 1$, if the ith patient received treatment A and $X_{iA} = 0$, if the patient received treatment B. Now, define $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$, $m = 2^p$, as the mutually exclusive configuration levels or strata of $\mathbf{z}$. For example, if $p = 2$, with $z_{i1} = 0, 1$ and $z_{i2} = 0, 1$, then $m = 4$ with $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} = \{(0,0), (0,1), (1,0), (1,1)\}$. Let $\pi_{iA}(\mathbf{v}_h)$, be the probability that the ith patient in stratum h was assigned to treatment A, n_h, be the number of patients with covariate configuration v_h and $N_{Ah}(n)$, the number of patients in stratum h assigned to treatment A. Then, $\hat{p}_{Ah} = N_{Ah}(n)/n_h$ is the sample proportion of patients in stratum h assigned to treatment A.

Suppose that the ideal model for the responses is the logit model with treatment-by-covariate interaction given by

$$\begin{aligned} \operatorname{logit}\Big[P(Y_i = 1|x_{iA}, \mathbf{z}_i)\Big] &= x_{iA}\beta_A + \gamma_0 + \mathbf{z}_i'\boldsymbol{\gamma} + x_{iA}\mathbf{z}_i'\boldsymbol{\delta} \\ &= \mathbf{w}_i'\boldsymbol{\theta}, \text{ for } i = 1, 2, \ldots, n, \end{aligned} \tag{2.1}$$

where γ_0 is an intercept, $\boldsymbol{\gamma}$ is a $p \times 1$ vector of the main effects of covariates and $\boldsymbol{\delta}$ is a $p \times 1$ vector of the effects of treatment by covariate interaction. In (2.1), $\mathbf{w}_i = (x_{iA}, 1, \mathbf{z}_i', x_{iA}\mathbf{z}_i')'$ and $\boldsymbol{\theta} = (\beta_A, \gamma_0, \boldsymbol{\gamma}', \boldsymbol{\delta}')' \in \boldsymbol{\Omega}(\boldsymbol{\theta_0}) \subseteq \Re^q$, where $\boldsymbol{\Omega}(\boldsymbol{\theta_0})$ is open and convex in $\Re^q$ and $q = 2(p+1)$. Let $\hat{\boldsymbol{\theta}}_n$ be the maximum likelihood estimator of the vector of true parameters $\boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0 = (\beta_{A0}, \gamma_{00}, \boldsymbol{\gamma}_0', \boldsymbol{\delta}_0')'$. Readers may refer to Selvaratnam, Yi and Oyet (2019) for detailed discussion and implementation of the maximum likelihood (ML) estimation of $\boldsymbol{\theta}_0$ with examples. We note that the non-random Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ associated with the model (2.1) can be partitioned, such that

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} I_{11}(\boldsymbol{\theta}) & I_{12}(\boldsymbol{\theta}) & \mathbf{I}_{13}(\boldsymbol{\theta}) & \mathbf{I}_{14}(\boldsymbol{\theta}) \\ I_{12}'(\boldsymbol{\theta}) & I_{22}(\boldsymbol{\theta}) & \mathbf{I}_{23}(\boldsymbol{\theta}) & \mathbf{I}_{24}(\boldsymbol{\theta}) \\ \mathbf{I}_{13}'(\boldsymbol{\theta}) & \mathbf{I}_{23}'(\boldsymbol{\theta}) & \mathbf{I}_{33}(\boldsymbol{\theta}) & \mathbf{I}_{34}(\boldsymbol{\theta}) \\ \mathbf{I}_{14}'(\boldsymbol{\theta}) & \mathbf{I}_{24}'(\boldsymbol{\theta}) & \mathbf{I}_{34}'(\boldsymbol{\theta}) & \mathbf{I}_{44}(\boldsymbol{\theta}) \end{bmatrix},$$

where $I_{12}(\boldsymbol{\theta}) = I_{11}(\boldsymbol{\theta})$, $I_{14}(\boldsymbol{\theta}) = I_{24}(\boldsymbol{\theta}) = I_{13}(\boldsymbol{\theta})$, and $I_{44}(\boldsymbol{\theta}) = I_{34}(\boldsymbol{\theta})$ (see Lemma 3.2(a) of Selva, Yi and Oyet, 2019). In what follows, we will outline in Theorem 1, some results that are required for establishing the asymptotic properties of the Wald-type test we have proposed in Theorem 2. First, we state the following assumptions.

Assumption 2.1. *Let $\pi_A(\mathbf{v}_h)$ be the target proportion of patients assigned to treatment A in stratum h and $J_n^{(h)} = \{i : \boldsymbol{z}_i = \mathbf{v}_h;\ \ i = 1, 2, \ldots, n\}$ be the set of all indices of patients with covariate*

configuration $\mathbf{v}_h$ *(*$h = 1, 2, \ldots, m$*). For each* $h = 1, 2, \ldots, m$*, we assume that*

(a) $(1/n_h)\sum_{i \in J_n^{(h)}} \pi_{iA}(\mathbf{v}_h) \xrightarrow{a.s.} \pi_A(\mathbf{v}_h)$ *as* $n_h \to \infty$*, where* $0 < \pi_A(\mathbf{v}_h) < 1$*.*

(b) $(n_h/n) \xrightarrow{a.s} \rho_h$ *as* $n \to \infty$*, where* $P(\boldsymbol{Z} = \mathbf{v}_h) = \rho_h$*,* $\sum_{h=1}^{m} \rho_h = 1$*,* $m < \infty$*.*

Theorem 1. *Let* $\boldsymbol{I}(\boldsymbol{\theta})$ *be a positive definite non-random Fisher information matrix associated with the model (2.1). Define* $g(\boldsymbol{w}_i'\boldsymbol{\theta}) = \exp(-\boldsymbol{w}_i'\boldsymbol{\theta})[1+\exp(-\boldsymbol{w}_i'\boldsymbol{\theta})]^{-2}$ *and* $\boldsymbol{F}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} g(\boldsymbol{w}_i'\boldsymbol{\theta})\boldsymbol{w}_i\boldsymbol{w}_i'$*. Then, under Assumption 2.1, we have*

(a) $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s} \boldsymbol{\theta_0}$ *as* $n \to \infty$*, where* $\hat{\boldsymbol{\theta}}_n$ *is the MLE of the true vector of parameters* $\boldsymbol{\theta_0}$*,*

(b) $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})$ *is asymptotically multivariate normal in distribution with mean* $\mathbf{0}$ *and variance-covariance matrix* $\boldsymbol{I}(\boldsymbol{\theta}_0)^{-1}$*,*

(c) $(1/n)\boldsymbol{F}_n(\boldsymbol{\theta}) \to \boldsymbol{I}(\boldsymbol{\theta})$*, for any* $\boldsymbol{\theta} \in \boldsymbol{\Omega}(\boldsymbol{\theta_0}) \subseteq \Re^q$*.*

The proof of Theorem 1 follows directly from Lemma 3.2 and Theorem 3 of Selvaratnam, Yi and Oyet (2019). Now, in general, the hypotheses for testing the significance of any of the effects in the model (2.1) can be written as

$$H_0 : \mathbf{D}\boldsymbol{\theta}_0 = \mathbf{d}_0 \quad \text{vs} \quad H_A : \mathbf{D}\boldsymbol{\theta}_0 \neq \mathbf{d}_0, \tag{2.2}$$

where $\mathbf{D}$ is a $d \times q$ matrix of full row rank and $\mathbf{d}_0$ is a $d \times 1$ constant vector. In a series of theorems that follow, we will outline our main results on the Wald-type test statistics for testing the significance of interaction and main effects in the model (2.1). The asymptotic distributions of the test statistics and expressions for the noncentrality parameters we need for analyzing the powers of the tests are also established in the theorems. We begin by summarizing our results on the Wald-type test statistic for the general hypotheses (2.2) in Theorem 2. The asymptotic distributions of the test statistics under the null and alternative hypothesis are also established in this theorem. The proofs of all theorems in this paper can be found in the Appendix.

Theorem 2. *Let* $\hat{\boldsymbol{\theta}}_n$ *be the unrestricted MLE of* $\boldsymbol{\theta_0}$*. Define*

$$\phi^{(a)} = n[\boldsymbol{D}\boldsymbol{\theta}_0 - \boldsymbol{d}_0]'[\boldsymbol{D}\boldsymbol{I}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{D}']^{-1}[\boldsymbol{D}\boldsymbol{\theta}_0 - \boldsymbol{d}_0],$$

and the Wald-type test statistic

$$T_W = [\boldsymbol{D}\hat{\boldsymbol{\theta}}_n - \boldsymbol{d}_0]'[\boldsymbol{D}\boldsymbol{F}_n(\hat{\boldsymbol{\theta}}_n)^{-1}\boldsymbol{D}']^{-1}[\boldsymbol{D}\hat{\boldsymbol{\theta}}_n - \boldsymbol{d}_0],$$

where $\boldsymbol{F}_n(\hat{\boldsymbol{\theta}}_n)$ *is the observed Fisher information matrix evaluated at* $\hat{\boldsymbol{\theta}}_n$*. Then, under Assumption (2.1),*

(a) T_W *converges to the central chi-square distribution,* χ^2_d*, with* d *degrees of freedom, if* H_0 *is true,*

(b) T_W *is asymptotically distributed as a noncentral chi-square distribution,* $\chi^2_d(\phi^{(a)})$ *with* d *degrees of freedom, and noncentrality parameter* $\phi^{(a)}$*, if* H_A *is true.*

2.1 Testing for the interaction effect

Let $\mathbf{d}_0$ and $\mathbf{D}$ in the hypotheses (2.2) be given by $\mathbf{D} = \begin{pmatrix} \mathbf{0}_{p\times(q-p)} & \mathbf{I}_{p\times p} \end{pmatrix}$ and $\mathbf{d}_0 = \mathbf{0}_{p\times 1}$, respectively, where $\mathbf{I}_{p\times p}$ is an identity matrix of dimension p. Then, the hypothesis (2.2) reduces to

$$H_{0I} : \boldsymbol{\delta}_0 = \mathbf{0}_{p\times 1} \quad \text{vs} \quad H_{AI} : \boldsymbol{\delta}_0 \neq \mathbf{0}_{p\times 1}. \tag{2.3}$$

It is clear that (2.3) is a hypotheses for testing for the significance of the effect of interaction in model (2.1). Now, partition $\mathbf{F}_n(\boldsymbol{\theta})$, such that

$$(1/n)\mathbf{F}_n(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Delta}_{11}^{[n]}(\boldsymbol{\theta}) & \boldsymbol{\Delta}_{12}^{[n]}(\boldsymbol{\theta}) \\ \boldsymbol{\Delta}_{21}^{[n]}(\boldsymbol{\theta}) & \boldsymbol{\Delta}_{22}^{[n]}(\boldsymbol{\theta}) \end{pmatrix},$$

where

$$\boldsymbol{\Delta}_{11}^{[n]}(\boldsymbol{\theta}) = (1/n)\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta}) \begin{pmatrix} x_{iA} & x_{iA} & x_{iA}\mathbf{z}_i' \\ x_{iA} & 1 & \mathbf{z}_i' \\ x_{iA}\mathbf{z}_i & \mathbf{z}_i & \mathbf{z}_i\mathbf{z}_i' \end{pmatrix}, \quad \boldsymbol{\Delta}_{21}^{[n]}(\boldsymbol{\theta}) = \boldsymbol{\Delta}_{12}^{[n]}(\boldsymbol{\theta})',$$

$$\boldsymbol{\Delta}_{12}^{[n]}(\boldsymbol{\theta}) = (1/n)\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta}) x_{iA} \begin{pmatrix} \mathbf{z}_i & \mathbf{z}_i & \mathbf{z}_i\mathbf{z}_i' \end{pmatrix}', \quad \boldsymbol{\Delta}_{22}^{[n]}(\boldsymbol{\theta}) = (1/n)\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta}) x_{iA}\mathbf{z}_i\mathbf{z}_i'$$

$$g(\mathbf{w}_i'\boldsymbol{\theta}) = \left[1 + \exp(-\mathbf{w}_i'\boldsymbol{\theta})\right]^{-2} \exp(-\mathbf{w}_i'\boldsymbol{\theta}).$$

Then, it follows from Theorem 2 that the Wald-type statistic, T_{WI} given by

$$T_{WI} = n[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]'\{\mathbf{D}[(1/n)\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)]^{-1}\mathbf{D}'\}^{-1}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0] = n\hat{\boldsymbol{\delta}}_n'[S_{\boldsymbol{\Delta}_{11}^{[n]}}(\hat{\boldsymbol{\theta}}_n)]\hat{\boldsymbol{\delta}}_n,$$

asymptotically follows a central χ_p^2 distribution, when H_{0I} is true, where

$$S_{\boldsymbol{\Delta}_{11}^{[n]}}(\hat{\boldsymbol{\theta}}_n) = \boldsymbol{\Delta}_{22}^{[n]}(\hat{\boldsymbol{\theta}}_n) - \boldsymbol{\Delta}_{21}^{[n]}(\hat{\boldsymbol{\theta}}_n)[\boldsymbol{\Delta}_{11}^{[n]}(\hat{\boldsymbol{\theta}}_n)]^{-1}\boldsymbol{\Delta}_{12}^{[n]}(\hat{\boldsymbol{\theta}}_n). \tag{2.4}$$

2.2 Testing for the main effect of treatments

Following our approach in §2.1, we let $d_0 = 0$ and $\mathbf{D} = \begin{pmatrix} 1 & \mathbf{0}_{1\times(q-1)} \end{pmatrix}$ in (2.2). Hypotheses (2.2) then reduces to

$$H_{0T} : \beta_{A0} = 0 \qquad H_{AT} : \beta_{A0} \neq 0, \tag{2.5}$$

which is the hypotheses for testing the significance of the main effect of treatments. Now, define

$$\frac{1}{n}\mathbf{F}_n(\boldsymbol{\theta}) = \begin{pmatrix} \Delta_{11t}^{[n]}(\boldsymbol{\theta}) & \boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}) \\ \boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}) & \boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}) \end{pmatrix},$$

where

$$\Delta_{11t}^{[n]}(\boldsymbol{\theta}) = (1/n)\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta})x_{iA},$$

$$\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}) = \boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta})' = (1/n)\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta})\left(x_{iA}, x_{iA}\mathbf{z}_i', x_{iA}\mathbf{z}_i'\right),$$

$$\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{w}_i'\boldsymbol{\theta})\begin{pmatrix} 1 & \mathbf{z}_i' & x_{iA}\mathbf{z}_i' \\ \mathbf{z}_i & \mathbf{z}_i\mathbf{z}_i' & x_{iA}\mathbf{z}_i\mathbf{z}_i' \\ x_{iA}\mathbf{z}_i' & x_{iA}\mathbf{z}_i\mathbf{z}_i' & x_{iA}\mathbf{z}_i\mathbf{z}_i' \end{pmatrix}, \text{ and}$$

$$g(\mathbf{w}_i'\boldsymbol{\theta}) = \exp(-\mathbf{w}_i'\boldsymbol{\theta})[1+\exp(-\mathbf{w}_i'\boldsymbol{\theta})]^{-2}.$$

It then follows, from Theorem 2 that under H_{0T}, the test statistic

$$T_W = n[\mathbf{D}\hat{\boldsymbol{\theta}}_n]'\{\mathbf{D}[(1/n)\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)]^{-1}\mathbf{D}'\}^{-1}[\mathbf{D}\hat{\boldsymbol{\theta}}_n] = n[\hat{\beta}_{An}]^2[S_{\boldsymbol{\Delta}_{22t}^{[n]}}(\hat{\boldsymbol{\theta}}_n)],$$

follows the chi-square distribution with 1 degree of freedom, χ_1^2 asymptotically, where

$$S_{\boldsymbol{\Delta}_{22t}^{[n]}}(\hat{\boldsymbol{\theta}}_n) = \Delta_{11t}^{[n]}(\hat{\boldsymbol{\theta}}_n) - \boldsymbol{\Delta}_{12t}^{[n]}(\hat{\boldsymbol{\theta}}_n)[\boldsymbol{\Delta}_{22t}^{[n]}(\hat{\boldsymbol{\theta}}_n)]^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\hat{\boldsymbol{\theta}}_n). \tag{2.6}$$

3 Statistical Power Computation

We mentioned earlier that there are three main approaches in the literature for computing the power of a test. In this section, we discuss the noncentrality parameters of the asymptotic distributions of the test statistics we developed in §2.

3.1 Power of the test for the interaction effect

Following Theorem 2, it can be shown that when H_{AI} is true, the asymptotic distribution of the Wald-type statistic T_{WI}, is the noncentral chi-square distribution with p degrees of freedom and noncentrality parameter $\phi^{(a)}$, given by

$$\begin{aligned}\phi^{(a)} &= n\big[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0\big]'\Big\{\mathbf{D}[(1/n)\mathbf{F}_n(\boldsymbol{\theta_0})]^{-1}\mathbf{D}'\Big\}^{-1}\big[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0\big] \\ &= n\boldsymbol{\delta}_0'[S_{\boldsymbol{\Delta}_{11}^{[n]}}(\boldsymbol{\theta}_0)]\boldsymbol{\delta}_0,\end{aligned} \tag{3.1}$$

where $S_{\boldsymbol{\Delta}_{11}^{[n]}}(\boldsymbol{\theta}_0)$ is given by (2.4) and $\hat{\boldsymbol{\theta}}_n$ is replaced by $\boldsymbol{\theta}_0$. When the test for interaction effect in (2.3) is performed, the vector of unknown true parameters, $\boldsymbol{\theta}_{0I} = (\beta_{A0}, \gamma_{00}, \boldsymbol{\gamma_0'})'$, in model (2.1) is typically considered a vector of nuisance parameters. In practice, the values of these nuisance parameters needed to compute the noncentrality parameter, hence the power of the test, for a real data, will be unknown to an experimenter. Thus, the value of the noncentrality parameter $\phi^{(a)}$ in (3.1), is commonly approximated by replacing $\mathbf{F}_n(\boldsymbol{\theta_0})$ with $\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)$. This is the approach adopted by Demidenko (2007).

3.2 Power of the test for the main effect

When testing for main effects under H_{AT} in (2.5), the asymptotic distribution of the statistic T_W becomes the noncentral chi-square distribution, with 1 degree of freedom and noncentrality parameter $\phi^{(a)}$, given by

$$\begin{aligned}\phi^{(a)} &= n\big[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0\big]'\Big\{\mathbf{D}[(1/n)\mathbf{F}_n(\boldsymbol{\theta_0})]^{-1}\mathbf{D}'\Big\}^{-1}[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0] \\ &= n[\beta_{A0}]^2[S_{\boldsymbol{\Delta}_{22t}^{[n]}}(\boldsymbol{\theta}_0)],\end{aligned}$$

where $S_{\boldsymbol{\Delta}_{22t}^{[n]}}(\boldsymbol{\theta}_0)$ is given by (2.6), and $\hat{\boldsymbol{\theta}}_n$ is replaced by $\boldsymbol{\theta}_0$. Here, the true treatment effect β_{A0}, is the main parameter of interest. The other parameters, $\boldsymbol{\theta}_{0N} = (\gamma_{00}, \boldsymbol{\gamma}_{\mathbf{0}}', \boldsymbol{\delta}_{\mathbf{0}}')'$, in the model (2.1) are then considered to be nuisance parameters. The nuisance parameters, $\boldsymbol{\theta}_{0N}$ are however required when computing the power of the test (2.5) for treatment effect. Following Demidenko (2007), we will replace $\mathbf{F}_n(\boldsymbol{\theta_0})$ by $\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)$ in our computation of the value of the noncentrality parameter $\phi^{(a)}$ in (3.2).

Recall that the sample proportion of patients assigned to treatment A in stratum h, is $\hat{p}_{Ah} = N_{Ah}(n)/n_h, h = 1, 2, \ldots, m$. Define the sample proportion vector $\hat{\mathbf{p}}_A = (\hat{p}_{A1}, \hat{p}_{A2}, \ldots, \hat{p}_{Am})'$ and target proportion vector $\boldsymbol{\pi_A} = (\pi_A(\mathbf{v}_1), \pi_A(\mathbf{v}_2), \ldots, \pi_A(\mathbf{v}_m))'$. In Theorem 3 that follows, we establish the result that the noncentrality parameter defined in (3.2) can be expressed in terms of the design variability induced by the randomization process, the bias, $\hat{\mathbf{p}}_A - \boldsymbol{\pi_A}$ and the target allocation proportions, $\boldsymbol{\pi_A}$. The proof of Theorem 3 can be found in the Appendix of this paper.

Theorem 3. *Let $m = 2^p$ and consider the noncentrality parameter $\phi^{(a)}$ defined in (3.2). Following the notations introduced in this section, we have that*

(a) $\phi^{(a)}$ is a function of $\hat{\boldsymbol{p}}_A$, where $\hat{\boldsymbol{p}}_A = (\hat{p}_{A1}, \hat{p}_{A2}, \ldots, \hat{p}_{Am})'$ is the vector of sample proportions of patients assigned to treatment A for given n. Let this function be $\phi(\hat{\boldsymbol{p}}_A)$.

(b) $\phi(\hat{\boldsymbol{p}}_A)$ can be expressed in terms of the target allocation proportion, the bias of the randomization procedure from the target, and the variability induced by the randomization process, in the following way

$$\begin{aligned}n^{-1}\phi(\hat{\boldsymbol{p}}_A) = n^{-1}\phi(\boldsymbol{\pi_A}) + n^{-1}\phi^{(1)}(\boldsymbol{\pi_A})[\hat{\boldsymbol{p}}_A - \boldsymbol{\pi_A}] + 2^{-1}[\hat{\boldsymbol{p}}_A - \boldsymbol{\pi_A}]' \\ n^{-1}\phi^{(2)}(\boldsymbol{\pi_A})[\hat{\boldsymbol{p}}_A - \boldsymbol{\pi_A}] + \boldsymbol{o}(\| \hat{\boldsymbol{p}}_A - \boldsymbol{\pi_A} \|^m),\end{aligned}$$

where

$$\phi^{(1)}(\boldsymbol{\pi_A}) = \Big(\partial\phi(\hat{\boldsymbol{p}}_A)/\partial\hat{\boldsymbol{p}}_A \Big)_{\hat{\boldsymbol{p}}_A = \boldsymbol{\pi_A}} \text{ and } \phi^{(2)}(\boldsymbol{\pi_A}) = \Big(\partial^2\phi(\hat{\boldsymbol{p}}_A)/\partial\hat{\boldsymbol{p}}_A^2 \Big)_{\hat{\boldsymbol{p}}_A = \boldsymbol{\pi_A}}.$$

We remark that if it can be shown that the matrix of second order derivatives, $\phi^{(2)}(\hat{\mathbf{p}}_A)$ is a negative definite matrix, then $\phi(\hat{\mathbf{p}}_A)$ is a concave down function. This will then lead to the significant result that the power of the tests increases as design variability decreases. Now, due to the complexity of the function $\phi^{(2)}(\hat{\mathbf{p}}_A)$ (see proof of Theorem 4), we will begin with the theoretical proof of the concavity of the noncentrality parameter $\phi(\hat{\mathbf{p}}_A)$ for a simple logit model with $p = 1$. We will then demonstrate, by simulation, that the result also holds for $p > 1$.

3.2.1 Concaveness of the noncentrality narameter

In this section, we consider a special case of the model (2.1) given by

$$\begin{aligned}\text{logit}\big[P(Y_i = 1|x_{iA}, z_i)\big] &= x_{iA}\beta_A + \gamma_0 + z_i\gamma_1, \\ &= \mathbf{w}_i'\boldsymbol{\theta}, \text{ for } i = 1, 2, \ldots, n,\end{aligned}$$

where γ_1 is the main effect of the covariate, β_A is the effect of treatment A compared to treatment B, γ_0 is the intercept term, $\mathbf{w}_i = (x_{iA}, 1, z_i)'$, and $\boldsymbol{\theta} = (\beta_A, \gamma_0, \gamma_1)'$. In this case, the design matrix $\mathbf{X}$ becomes $\mathbf{X} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n)'$ and the vector of true parameters is $\boldsymbol{\theta}_0 = (\beta_{A0}, \gamma_{00}, \gamma_{10})'$. We note that the third term in the expression (3.2), for the noncentrality parameter $\phi(\hat{\mathbf{p}}_A)$, is the design variability induced by the randomization process. However, the term $\phi^{(2)}(\boldsymbol{\pi_A})$ depends on the unknown model parameters. So, for the purpose of our simulation studies and for practical computations, we now introduce an expression for estimating the design variability. Let S be the number of simulations and define $\hat{\mathbf{p}}_{Ai} = (\hat{p}_{A1i}, \hat{p}_{A2i}, \ldots, \hat{p}_{Ami})'$ as the proportions from the ith simulation. The design variability D_{v}, is defined by $D_{\text{v}} = (S-1)^{-1}\sum_{i=1}^{S}[\hat{\mathbf{p}}_{Ai} - \boldsymbol{\pi}_A]'[\hat{\mathbf{p}}_{Ai} - \boldsymbol{\pi}_A]$. It can be estimated by $\hat{D}_{\text{v}} = (S-1)^{-1}\sum_{i=1}^{S}[\hat{\mathbf{p}}_{Ai} - \hat{\boldsymbol{\pi}}_A]'[\hat{\mathbf{p}}_{Ai} - \hat{\boldsymbol{\pi}}_A]$, where $\hat{\boldsymbol{\pi}}_A = S^{-1}\sum_{i=1}^{S}\hat{\mathbf{p}}_{Ai}$.

In what follows, we shall prove theoretically, that when only a single covariate ($p = 1$) is involved in a clinical trial, the noncentrality parameter (3.2) is a concave down function. We will also show that the power of the test for main effect increases as the design variability decreases. The proof of these results outlined in Theorem 4 can be found in the Appendix. Through our simulation studies, we have also demonstrated that these results hold even when $p > 1$. It is clear that when $p = 1$, the vector of sample proportion of patients assigned to treatment A becomes $\hat{\mathbf{p}}_A = (\hat{p}_{A1}, \hat{p}_{A2})'$, and $\boldsymbol{\pi_A} = (\pi_A(\text{v}_1), \pi_A(\text{v}_2))'$.

Theorem 4. *Let $p = 1$ be the number of covariates in a clinical trial and consider the noncentrality parameter defined by (3.2). Then, we have that*

(a) $\phi^{(2)}(\hat{\boldsymbol{p}}_A)$ is a negative definite matrix, hence $\phi(\hat{\boldsymbol{p}}_A)$ is a concave down function,

(b) the power of the hypothesis test increases as the design variability decreases.

In Figure 1, we display a three dimensional graph of $n^{-1}\phi(\boldsymbol{\pi_A})$ versus $\boldsymbol{\pi_A} = (\pi_A(\text{v}_1), \pi_A(\text{v}_2))'$ for the purpose of illustration. The graph was constructed with the following combination of model parameters: $\beta_A = 1.5$, $\gamma_0 = 0.6$, and $\gamma_1 = -0.4$ and $n_1 = 300$ and $n_2 = 200$. Figure 1 confirms the concaveness of the noncentrality parameter as a function of the proportion of treatment assignment.

4 Simulation Studies

In order to secure approval for a new drug at the end of a clinical trial, scientific evidence from the trial has to be provided to the approval agency. The evidence will include the statistical power for treatment effectiveness as well as the p-values from results of statistical tests. In our extensive simulation studies, we will first examine the performance of the Wald-type tests in estimating the true size of the test. We will also examine the power of the test by comparing the simulated power

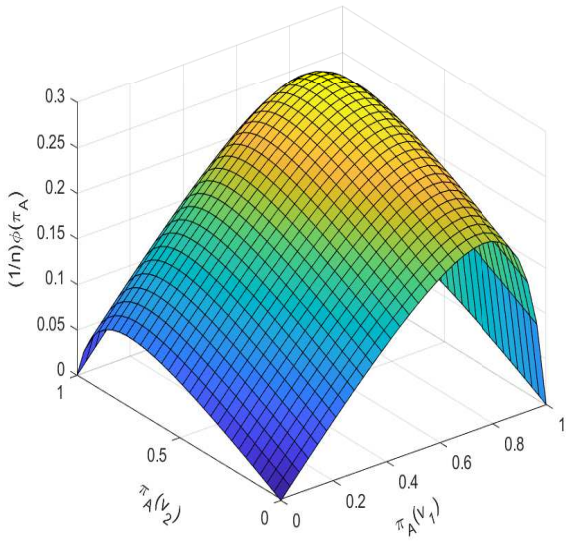


Figure 1: A plot of $(1/n)\phi(\boldsymbol{\pi_A})$ as a function of $\boldsymbol{\pi_A} = (\pi_A(\mathrm{v}_1), \pi_A(\mathrm{v}_2))'$.

obtained by computing the proportion of rejections of the null hypothesis when the alternative hypothesis is true, with the conventional power computed directly based on the methods discussed in §3.

Suppose it is of interest to control for the covariates, gender, age and chronic conditions, in a clinical trial for comparing two treatments, denoted by A and B, involving 500 or 1000 participating patients. Let the covariates be defined by

$$Z_{i1} = \begin{cases} 1 & \text{Male (55\%),} \\ 0 & \text{Female.} \end{cases} \quad \text{and} \quad Z_{i2} = \begin{cases} 1 & \text{at least one chronic condition (60\%),} \\ 0 & \text{otherwise.} \end{cases}$$

$$Z_{i3} = \begin{cases} 1 & 20 \leq \text{Age} \leq 50 \text{ (30\%),} \\ 0 & \text{otherwise.} \end{cases} \quad \text{and} \quad Z_{i4} = \begin{cases} 1 & 50 < \text{Age} \leq 65 \text{ (30\%),} \\ 0 & \text{otherwise.} \end{cases}$$

Though, the results in this paper apply to a general class of designs, we will assume that treatments were assigned to the 500 or 1000 patients based on RA or CARA or CR design. To generate the data for our simulation studies based on these 3 designs, we have followed the detailed process for treatment assignment and data generation discussed by Selvaratnam, Yi and Oyet (2019). The combination of true parameter values we have used for data generation in this paper were:

Scenarios without treatment-by-covariate interaction:

(a) $\beta_{A0} = 1.50, \gamma_{00} = 0.50, \gamma_{10} = -0.60, \gamma_{20} = -0.30, \gamma_{30} = 0.25, \gamma_{40} = 0.10, \delta_{10} = 0.00, \delta_{20} = 0.00, \delta_{30} = 0.00, \delta_{40} = 0.00.$

(b) $\beta_{A0} = 0.50, \gamma_{00} = 0.25, \gamma_{10} = -0.20, \gamma_{20} = -0.40, \gamma_{30} = 0.35, \gamma_{40} = 0.20, \delta_{10} = 0.00, \delta_{20} = 0.00, \delta_{30} = 0.00, \delta_{40} = 0.00.$

Scenario with treatment-by-covariate interaction:

(c) $\beta_{A0} = 0.50, \gamma_{00} = 0.25, \gamma_{10} = -0.20, \gamma_{20} = -0.40, \gamma_{30} = 0.35, \gamma_{40} = 0.20, \boldsymbol{\delta} = (\delta_{10}, \delta_{20}, \delta_{30}, \delta_{40})$, with

(i) $(\delta_{10}, \delta_{20}, \delta_{30}, \delta_{40}) = (0.10, -1.50, 0.00, 0.00)$.

(ii) $(\delta_{10}, \delta_{20}, \delta_{30}, \delta_{40}) = (0.10, -1.50, 0.35, 0.00)$.

(iii) $(\delta_{10}, \delta_{20}, \delta_{30}, \delta_{40}) = (0.30, -1.20, 0.35, 0.20)$.

(iv) $(\delta_{10}, \delta_{20}, \delta_{30}, \delta_{40}) = (0.20, -2.00, 0.07, 0.05)$.

Once data was generated based on a given design, we computed a value of the Wald-test statistic we proposed in §2.1,

$$T_{WI} = n\hat{\boldsymbol{\delta}}_n'[S_{\boldsymbol{\Delta}_{11}^{[n]}}(\hat{\boldsymbol{\theta}}_n)]\hat{\boldsymbol{\delta}}_n,$$

for testing the hypotheses (2.3) of covariate-by-treatment interaction. For a fixed value α^*, of the significance level α, we then made a decision to reject H_0 in (2.3) if the computed value of T_{WI} was within the rejection region

$$\left\{T_{WI} : T_{WI} < \chi^2_{4,\alpha^*/2} \text{ or } T_{WI} > \chi^2_{4,(1-\alpha^*/2)}\right\}, \tag{4.1}$$

where $P\left(T_{WI} < \chi^2_{4,\alpha^*/2}\right) = \alpha^*/2$ and $P\left(T_{WI} < \chi^2_{4,1-\alpha^*/2}\right) = 1 - \alpha^*/2$. For each simulation, we also noted the treatment assigned to each patient and the outcome of the treatment which is the response Y. This process was repeated 3000 times under the null hypothesis to obtain the size estimates $\hat{\alpha}^*$ shown in Tables 1, 2 and 3. Similarly, the simulated power shown in Tables 4 and 5 were obtained by generating the data under the alternative hypothesis and repeating the process 3000 times.

4.1 Discussion of results

Size of the test

In Tables 1, 2 and 3, we display the overall success rate, the proportion of patients assigned to treatment A, the proportion of rejections $\hat{\alpha}^*$ in 3000 simulations and the average design variability.

The results clearly show that under CARA, RA and CR designs the true value of the significance level was correctly estimated by the test in all the cases we considered. Thus, the statistic was able to control the size of the test. For instance, under CARA design, when $\alpha^* = 0.1$, we obtained $\hat{\alpha}^* = 0.1053$ (Table 1), $\hat{\alpha}^* = 0.1017$ (Table 2) and $\hat{\alpha}^* = 0.0937$ (Table 3). Also, when $\alpha^* = 0.05$, we obtained $\hat{\alpha}^* = 0.0507$ (Table 1), $\hat{\alpha}^* = 0.0507$ (Table 2) and $\hat{\alpha}^* = 0.0537$ (Table 3). The results obtained under RA, CARA and CR designs were similar. Tables 1, 2 and 3 also show that both the overall success rate and the proportion of patients that were assigned the better treatment were significantly higher under CARA and RA designs. This result further confirms that using CARA

Table 1: Estimated size, $\hat{\alpha}^*$ of test for interaction, overall success rate of treatment A (OSR), proportion of treatment A assigned (AP) and design variability (DV) under CARA, RA, and CR designs obtained from 3000 simulations involving 500 patients with $\beta_{A0} = 1.50$, $\gamma_{00} = 0.50$, $\gamma_{10} = -0.60$, $\gamma_{20} = -0.30$, $\gamma_{30} = 0.25$, $\gamma_{40} = 0.10$, $\delta_{10} = 0.00$, $\delta_{20} = 0.00$, $\delta_{30} = 0.00$, $\delta_{40} = 0.00$.

Design	α^*	$\hat{\alpha}^*$	OSR	AP	DV
CARA	0.10	0.1053	0.7535	0.7587	0.0458
	0.05	0.0507	0.7535	0.7587	0.0458
	0.01	0.0083	0.7535	0.7587	0.0458
RA	0.10	0.0947	0.7695	0.8124	0.0403
	0.05	0.0453	0.7695	0.8124	0.0403
	0.01	0.0103	0.7695	0.8124	0.0403
CR	0.10	0.0927	0.6742	0.4996	0.0224
	0.05	0.0443	0.6742	0.4996	0.0224
	0.01	0.0080	0.6742	0.4996	0.0224

Table 2: Column headings are as in Table 1. Results were obtained under CARA, RA, and CR designs from 3000 simulations involving 1000 patients with $\beta_{A0} = 1.50$, $\gamma_{00} = 0.50$, $\gamma_{10} = -0.60$, $\gamma_{20} = -0.30$, $\gamma_{30} = 0.25$, $\gamma_{40} = 0.10$, $\delta_{10} = 0.00$, $\delta_{20} = 0.00$, $\delta_{30} = 0.00$, $\delta_{40} = 0.00$.

Design	α^*	$\hat{\alpha}^*$	OSR	AP	DV
CARA	0.10	0.1017	0.7630	0.7923	0.0348
	0.05	0.0507	0.7630	0.7923	0.0348
	0.01	0.0067	0.7630	0.7923	0.0348
RA	0.10	0.0927	0.7691	0.8131	0.0283
	0.05	0.0463	0.7691	0.8131	0.0283
	0.01	0.0103	0.7691	0.8131	0.0283
CR	0.10	0.0957	0.6741	0.4996	0.0157
	0.05	0.0457	0.6741	0.4996	0.0157
	0.01	0.0087	0.6741	0.4996	0.0157

Table 3: Column headings are as in Table 1. Results were obtained under CARA, RA, and CR designs from 3000 simulations involving 500 patients with $\beta_{A0} = 0.50$, $\gamma_{00} = 0.25$, $\gamma_{10} = -0.20$, $\gamma_{20} = -0.40$, $\gamma_{30} = 0.35$, $\gamma_{40} = 0.20$, $\delta_{10} = 0.00$, $\delta_{20} = 0.00$, $\delta_{30} = 0.00$, $\delta_{40} = 0.00$.

Design	α^*	$\hat{\alpha}^*$	OSR	AP	DV
CARA	0.10	0.0937	0.5890	0.6103	0.0599
	0.05	0.0537	0.5890	0.6103	0.0599
	0.01	0.0113	0.5890	0.6103	0.0599
RA	0.10	0.1073	0.5901	0.6227	0.0477
	0.05	0.0507	0.5901	0.6227	0.0477
	0.01	0.0100	0.5901	0.6227	0.0477
CR	0.10	0.1013	0.5756	0.4996	0.0223
	0.05	0.0527	0.5756	0.4996	0.0223
	0.01	0.0113	0.5756	0.4996	0.0223

or RA design will ensure that the trial achieves the ethically desirable outcome of maintaining the well-being of participating patients. Furthermore, the maximum value of the design variability was approximately 0.06, which is quite reasonable.

Power of the test

Concerning the power of the test, we generated data under the alternative hypothesis and computed both the simulated power and the conventional power. The conventional power under each simulation, was computed using the asymptotic noncentral χ^2 distribution discussed in §3.1 with non-centrality parameter (3.1). Since this is an approximate method, we also computed the deviation between the simulated power and the average conventional power. The maximum and minimum values of the deviations, shown in Tables 4 and 5, were 0.097 and 0.000 respectively. The deviation between the average conventional power and the simulated power become smaller in magnitude as the number of participating patients increases. This clearly shows that the conventional power we proposed in this paper, can be used to closely approximate the power of the test. In Theorem 4, we had established a theoretical inverse relationship between the power of the test and the design variability when controlling for only $p = 1$ covariate. In our simulation studies, we examined whether this relationship is also valid if the ideal model contains more than a single covariate. The results, reported in Tables 4 and 5 for the four scenarios we considered, show that even when $p > 1$, the average conventional power of the tests were generally higher for smaller values of the design variability. For instance, when the RA design was implemented under scenario (c)(i), the design variability was 0.0480 in Table 4 and the average conventional power was 78.57%. However, the

Table 4: Simulated power (SP) and average conventional power (ACP) of test for interaction, overall success rate of treatment A (OSR), average proportion of treatment A assigned (AP) and design variability (DV) under CARA, RA, and CR designs obtained from 3000 simulations involving 500 patients with $\alpha^* = 0.05$, $\beta_{A0} = 0.50$, $\gamma_{00} = 0.25$, $\gamma_{10} = -0.20$, $\gamma_{20} = -0.40$, $\gamma_{30} = 0.35$, $\gamma_{40} = 0.20$, $\boldsymbol{\delta}_0 = (\delta_{10}, \delta_{20}, \delta_{30}, \delta_{40})'$.

$\boldsymbol{\delta}_0$	Design	SP	ACP	Error	OSR	AP	DV
$(0.10, -1.50, 0.00, 0.00)'$	CARA	0.8163	0.7195	0.0968	0.5086	0.4256	0.0596
	RA	0.8050	0.7857	0.0193	0.4825	0.4233	0.0480
	CR	0.8303	0.8111	0.0192	0.4770	0.4996	0.0224
$(0.10, -1.50, 0.35, 0.00)'$	CARA	0.8183	0.7320	0.0863	0.5172	0.4462	0.0601
	RA	0.8140	0.7963	0.0177	0.4904	0.4446	0.0485
	CR	0.8353	0.8140	0.0213	0.4878	0.4996	0.0223
$(0.30, -1.20, 0.35, 0.20)'$	CARA	0.6273	0.5595	0.0678	0.5436	0.5194	0.0605
	RA	0.6510	0.6135	0.0375	0.5252	0.5200	0.0484
	CR	0.6463	0.6148	0.0315	0.5253	0.4996	0.0223
$(0.20, -2.00, 0.07, 0.05)'$	CARA	0.9677	0.8707	0.0970	0.5125	0.3936	0.0541
	RA	0.9687	0.9494	0.0193	0.4718	0.3891	0.0479
	CR	0.9783	0.9713	0.0070	0.4595	0.4996	0.0224

design variability reduces to 0.0224 under the CR design and the average conventional power increases to 81.11%. This pattern can be seen throughout Tables 4 and 5. In general, we found that design variability was smallest under CR designs and largest under CARA designs. Thus, the average conventional power for CR designs was higher than that of RA and CARA designs and smallest for CARA designs. The difference in magnitude were however not very large. The maximum difference computed from values of the average conventional power with 500 participating patients, in Table 4 was 0.1006. The maximum difference reduces to 0.0326 when the number of participating patients was increased to 1000 in Table 5. This shows that the efficiency of the designs will be about the same if a sufficiently large number of patients participate in the trial. We also note that as the number of participating patients increased from 500 in Table 4 to 1000 in Table 5, the simulated and average conventional power of the Wald-type test increased. For instance, the simulated power for scenario (c)(iii) were 62.73%, 65.10% and 64.63% under CARA, RA and CR designs respectively, while the average conventional power were 55.95%, 61.35% and 61.48% (see Table 4) respectively, when 500 patients participated in the clinical trial. However, when the number of patients increased to 1000, the simulated powers increased significantly to 94.10%, 94.87%, and 94.80%, respectively and the average conventional power increased to 90.80%, 94.06% and 94.06% respectively (see Table 5). These results demonstrate that the efficiency of the Wald-type test we proposed increases as

the number of participating patients increases.

Table 5: Column headings are as in Table 4. Results were obtained under CARA, RA, and CR designs from 3000 simulations involving 1000 patients with $\alpha^* = 0.05$, $\beta_{A0} = 0.50$, $\gamma_{00} = 0.25$, $\gamma_{10} = -0.20$, $\gamma_{20} = -0.40$, $\gamma_{30} = 0.35$, $\gamma_{40} = 0.20$, $\boldsymbol{\delta}_0 = (\delta_{10}, \delta_{20}, \delta_{30}, \delta_{40})'$.

$\boldsymbol{\delta}_0$	Design	SP	ACP	Error	OSR	AP	DV
$(0.10, -1.50, 0.00, 0.00)'$	CARA	0.9887	0.9734	0.0153	0.5108	0.4239	0.0455
	RA	0.9913	0.9894	0.0019	0.4831	0.4226	0.0339
	CR	0.9953	0.9924	0.0029	0.4772	0.4996	0.0157
$(0.10, -1.50, 0.35, 0.00)'$	CARA	0.9913	0.9768	0.0145	0.5191	0.4461	0.0450
	RA	0.9940	0.9907	0.0033	0.4909	0.4440	0.0340
	CR	0.9960	0.9927	0.0033	0.4879	0.4996	0.0157
$(0.30, -1.20, 0.35, 0.20)'$	CARA	0.9410	0.9080	0.0330	0.5447	0.5236	0.0453
	RA	0.9487	0.9406	0.0081	0.5253	0.5188	0.0348
	CR	0.9480	0.9406	0.0074	0.5252	0.4996	0.0157
$(0.20, -2.00, 0.07, 0.05)'$	CARA	1.0000	0.9888	0.0112	0.5167	0.3877	0.0430
	RA	1.0000	0.9998	0.0002	0.4722	0.3878	0.0338
	CR	1.0000	1.0000	0.0000	0.4598	0.4996	0.0157

Furthermore, the results in Tables 4 and 5 show that the overall success rates for CARA design were higher (i.e. more ethical) than that of RA design and that of RA design higher than that of CR design, when the effect of treatment-by-covariate interaction is significant. However, when the effect of treatment-by-covariate interaction is not significant as in Tables 1, 2, and 3, the RA design was more ethical than CARA design and CARA design more ethical than CR design. Previously, many clinical trials have been conducted using CR and covariate-adjusted (CA) designs for treatment assignments and data collection with the expectation that it will lead to a more efficient statistical inference. Our results have shown that if the number of participating patients is sufficiently large, CARA and RA designs can be as equally efficient as CR designs and also more ethical than CR designs. We will therefore recommend implementing RA or CARA designs in a clinical trial if a researcher wishes to achieve a balance between ethics and efficiency.

4.2 Application

Selvaratnam, Yi and Oyet (2019) noted that increasing discoveries of biomarkers, which can be used as covariates in a GLM, and observed diversity among responses of patients to treatments has led to a growing interest in the development of personalized medicine. Since patients in a clinical trial participate voluntarily and magnanimously in order to assist in identifying a better treatment for

future generations, it is highly important to consider their well-being at the planning stage and also during the trial. Previous authors have used only simulation studies to document the advantages RA and CARA designs have over CR designs in achieving this important ethical objective of any clinical trial. Recently, however, Selvaratnam, Yi and Oyet (2019) introduced a new approach for using data from an actual clinical trial to mimick the trial under a different design in order to obtain real data which can be used to compare the performance of designs. They implemented their approach on data obtained from a clinical trial on Stroke Prevention in Atrial Fibrillation (SPAF) (Hart et al., 2003), which was obtained based on a CR design. The data from the clinical trial is summarized in Table 6.

Table 6: Data obtained from 1120 patients in a CR clinical trial on Stroke Prevention in Atrial Fibrillation (SPAF).

Treatment	Covariate	Response		Total
x_A	z_1	Success	Failure	
1	1	205	1	206
1	0	321	25	346
0	1	193	18	211
0	0	329	28	357

Using the approach they introduced, they mimicked the clinical trial while maintaining the true covariate information of patients in the trial to obtain responses based on RA and CARA designs. In this section, we follow their approach to obtain responses from the SPAF clinical trial under CR, RA and CARA designs for the purpose of comparing the efficiencies of the Wald-type tests we have constructed under these designs. The average number of patients assigned to treatments in 100 simulations mimicking the SPAF clinical trial are shown in Table 7. We also investigated the overall success rates of the treatments which is a measure of the ability of the designs to achieve the ethical objective of the trial.

In this paper, we have considered anticoagulation status from the SPAF study as a covariate. Now, let $Z_{i1} = 1$, if the ith patient received anticoagulation therapy, and $Z_{i1} = 0$, otherwise. Also, let v_1 and v_2 be the stratum of patients with covariate $Z_1 = 1$ and $Z_1 = 0$, respectively. We will also consider two treatments from the study namely, aspirin, we shall refer to as treatment A and placebo we shall refer to as treatment B. We have assumed that the ideal model for the data is the logit model,

$$\begin{aligned}\operatorname{logit}\big[P(Y_i = 1|x_{iA})\big] &= x_{iA}\beta_A + \gamma_0 + \gamma_1 z_{i1} + \delta_1 x_{iA} z_{i1} \\ &= \mathbf{w}_i'\boldsymbol{\theta} \text{ for } i = 1, 2, \ldots, n,\end{aligned}$$

where $\boldsymbol{\theta} = (\beta_A, \gamma_0, \gamma_1, \delta_1)'$ and $\mathbf{w}_i = (x_{iA}, 1, z_{i1}, x_{iA}z_{i1})'$. Thus, the hypotheses for treatment-by-covariate interaction can be written as

$$H_0 : \delta_{10} = 0 \quad \text{vs} \quad H_A : \delta_{10} \neq 0, \tag{4.2}$$

Table 7: Average number of patients assigned to treatments obtained from 100 simulations mimicking the SPAF study under CR, RA and CARA designs.

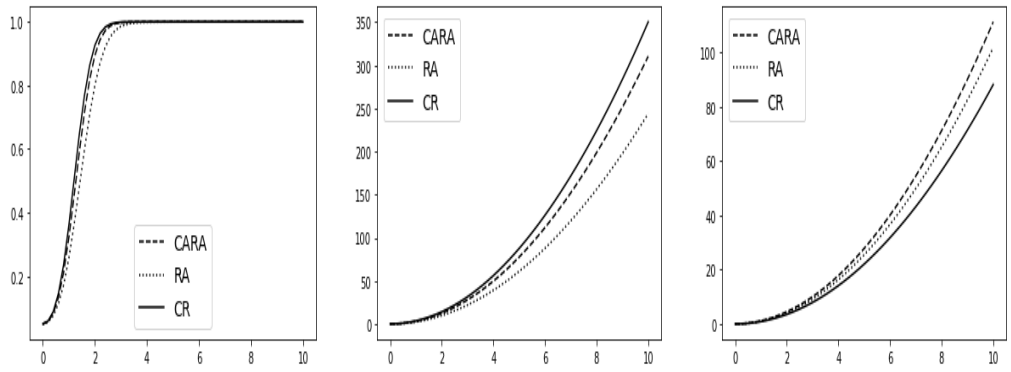
Treatment	Covariate	The number of patients		
x_A	z_1	CR	RA	CARA
1	1	209	253	302
1	0	352	430	512
0	1	208	164	115
0	0	351	273	191

where δ_{10} is the true interaction effect between treatment and covaraiate.

Table 8: The headings of columns 2, 3, and 4 are as in Table 1. The results were obtained from 100 simulations mimicking the SPAF study under CARA, RA, and CR designs.

Design	AP	OSR	DV	$\pi_A(\mathrm{v}_1)$	$\pi_A(\mathrm{v}_2)$
CARA	0.72695	0.94548	0.00707	0.72458	0.72835
RA	0.60920	0.95859	0.00301	0.60597	0.61111
CR	0.50074	0.93646	0.00098	0.50010	0.50112

Following our approach in §4.1, we computed the simulated average of conventional power for various values of δ_{10}. A graph of the simulated averages for values of δ_{10} between 0 and 10 is displayed in Figure 2(a). We note that our results in this case study agree with the results of our simulation study in §4.1, in the sense that the conventional power under CARA and RA designs were smaller than the power for CR design. The conventional power was computed using the estimate of the noncentrality parameter $\hat{\phi}(\hat{\mathbf{p}}_A)$. In Figure 2(b), we display graphs of the simulated average of the estimates of the noncentrality parameters $\hat{\phi}(\hat{\mathbf{p}}_A)$ for various values of δ_{10} under the three designs. Taken together, Figures 2(a),(b) demonstrates that the conventional power increases as the noncentrality parameter increases. Similar to the pattern in Table 8, where design variabilities under RA design can be seen to be smaller than design variabilities for CARA design, the values of $\phi(\hat{\pi}_A)$ under RA design were also smaller than the values of $\phi(\hat{\pi}_A)$ for CARA design in Figure 2(c). Concerning success rates which we used as a measure of ethics, our results in Table 8 show that the success rate was also higher for CARA and RA designs than for CR design. These results clearly support our previous conclusion that applying RA and CARA designs in clinical trials will yield data that will lead to efficient statistical inference and also account for the well-being of patients. In fact, based on the overall success rates, the results indicate that if CARA or RA design had been applied in the SPAF study, the well-being of eight, under CARA or five, under RA, more patients



(a) (b) (c)

Figure 2: A plot of (a) Simulated average of conventional powers of the test in (4.2) against values of δ_{10}, (b) simulated average of the estimates of the noncentrality parameters against δ_{10}, (c) value of $\phi(\hat{\pi}_A)$ obtained from 100 simulations mimicking the SPAF clinical trial under CR, RA and CARA designs against δ_{10}.

could have being improved in the clinical trial.

5 Conclusion

In this paper, we discussed the importance of examining the effect of interaction between treatments and covariates, such as the personal characteristics of patients, in a clinical trial and developed a Wald-type test for testing for the significance of this effect. We also developed an approximate method for computing the statistical power of the test and found that as the number of participating patients increases, the power of the test also increases. We proved that our method depends on a general class of adaptive designs through the noncentrality parameter of the non-null asymptotic distribution of the test statistic. Furthermore, we found that the noncentrality parameter can be expressed as a function of the target allocation proportion, the bias of the randomization procedure from the target, and the variability induced by the randomization process. This finding is an extension of the work of Hu and Rosenberger (2003). For the purpose of illustration, we applied three designs, namely RA, CARA and CR, in intensive simulation studies, and found that the values of statistical power computed based on our approximate method were very close to values obtained by simulation under the three designs we considered. In terms of efficiency of the test, the power of the tests under all three designs were very close when a large number of patients participated in clinical trials. Thus, RA and CARA designs are useful alternatives to CR design that takes into account the well-being of patients in clinical trials.

We note that the well-being of participating patients is usually a top priority, in clinical trials, which cannot be traded for gain in power of the test. Results from our simulation studies and application to real data have shown that the twin objectives of ethics and achieving over 90% power

can be met by assigning treatments based on CARA or RA designs to a large number of participating patients in a trial. Thus, it is clear that a minimum sample size or number of participating patients will be required as proposed by Austin and Steyerberg (2017), Bujang et al. (2018), Smeden et al. (2019), among others. Consequently, when using our proposed method for power computation, one may proceed as follows.

Step 1: Following Bujang et al. (2018), start with the number of patients, $n = 100 + 50(q-1)$, where $(q-1)$ is the number of predictor variables in the model.

Step 2: Compute the conventional power of the test.

Step 3: Add 50 new patients to the clinical trial and obtain responses.

Step 4: Compute the conventional power of the test.

Step 5: Calculate the absolute deviation between the two recent successive powers.

Step 6: End the trial if the deviation is less than 1%. Otherwise, go to **Step 3**.

Furthermore, we derived a theoretical relationship between the power of the test for the main effect and the design variability when only a single binary covariate is in the logit model for the data. We found that as the design variability decreases, the conventional power increases. Through simulation studies, we confirmed that this inverse relationship is also true in the presence of more than one binary covariate. This empirical result creates the open problem of extending the results of Theorem 4 to more than one categorical covariates. However, the complicated structure of the noncentrality parameter makes this a challenging problem. The results of our case study further confirmed our theoretical and simulation results that RA and CARA designs are more ethical and can be equally efficient as CR designs if sufficiently large number of patients participate in a clinical trial.

Appendix

A Proof of Therem 2.2

Proof. Using Theorem 1(a) and the continuous mapping theorem, we obtain,

$$n^{-1}\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{a.s.} n^{-1}\mathbf{F}_n(\boldsymbol{\theta_0}). \tag{A.1}$$

Next, applying (A.1) and Theorem 1(c), we obtain

$$n^{-1}\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{a.s.} \mathbf{I}(\boldsymbol{\theta}_0). \tag{A.2}$$

(a) Under H_0, T_W can be written as

$$\begin{aligned} T_W &= [\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{D}\boldsymbol{\theta}_0]'[\mathbf{D}\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)^{-1}\mathbf{D}']^{-1}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{D}\boldsymbol{\theta}_0] \\ &= [\sqrt{n}\mathbf{D}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)]'\{\mathbf{D}[(1/n)\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)]^{-1}\mathbf{D}'\}^{-1}[\sqrt{n}\mathbf{D}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)]. \end{aligned}$$

By applying (A.2), we now have,

$$\begin{aligned}\{\mathbf{D}[(1/n)\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)]^{-1}\mathbf{D}'\}^{-1} &\xrightarrow{a.s.} \{\mathbf{D}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{D}'\}^{-1} \\ &= \mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}, \end{aligned} \tag{A.3}$$

where $\mathbf{D}^*(\boldsymbol{\theta}_0) = \mathbf{D}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{D}'$. Since $\mathbf{D}$ is a $(d \times q)$ matrix of full row rank and $\mathbf{I}(\boldsymbol{\theta}_0)^{-1}$ is a positive definite matrix, we have that $\mathbf{D}^*(\boldsymbol{\theta}_0)$ is a positive definite matrix (Seber and Lee, 2003). Therefore, there exists a unique square root matrix $\mathbf{D}^*(\boldsymbol{\theta}_0)^{1/2}$ of $\mathbf{D}^*(\boldsymbol{\theta}_0)$.

Now, $\mathbf{D}$ is a $(d \times q)$ matrix and $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_\mathbf{0}) \xrightarrow{d} N_q[\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}_0)^{-1}]$, by Theorem 1(b).Therefore, $\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}\sqrt{n}\mathbf{D}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N_d[\mathbf{0}, \mathbf{I}_d^*]$ (Srivastava, 2002), where $\mathbf{I}_d^*$ is an identity matrix of dimension d. Therefore,

$$\begin{aligned} T_W^* &= \{\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{D}\boldsymbol{\theta}_0]'\}[\mathbf{D}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{D}']^{-1}\{\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{D}\boldsymbol{\theta}_0]\} \\ &= \{\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{D}\boldsymbol{\theta}_0]'\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}\}\{\sqrt{n}\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{D}\boldsymbol{\theta}_0]\} \\ &\xrightarrow{d} \chi_d^2 \end{aligned} \tag{A.4}$$

where χ_d^2 is the central chi-square distribution with d degrees of freedom. It follows from (A.3) and (A.4), that T_W asymptotically follows the central chi-square distribution with d degrees of freedom.

(b) Under H_A

$$\begin{aligned} T_W &= [\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]'[\mathbf{D}\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)^{-1}\mathbf{D}']^{-1}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0] \\ &= \sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]'[\mathbf{D}\{(1/n)\mathbf{F}_n(\hat{\boldsymbol{\theta}}_n)\}^{-1}\mathbf{D}']^{-1}\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]. \end{aligned}$$

Considering the fact that,

$$\begin{aligned}\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0] &= \sqrt{n}[\mathbf{D}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + (\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0)] \\ &\xrightarrow{d} N_d[\sqrt{n}(\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0), \mathbf{D}^*(\boldsymbol{\theta}_0)]. \end{aligned}$$

It follows that,

$$\sqrt{n}\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0] \xrightarrow{d} N_d[\sqrt{n}\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}(\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0), \mathbf{I}_d^*].$$

Therefore, following Anderson (1966),

$$\begin{aligned} T_W^* &= \{\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]'\}[\mathbf{D}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{D}']^{-1}\{\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]\} \\ &= \{\sqrt{n}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]'\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}\}\{\sqrt{n}\mathbf{D}^*(\boldsymbol{\theta}_0)^{-1/2}[\mathbf{D}\hat{\boldsymbol{\theta}}_n - \mathbf{d}_0]\} \\ &\xrightarrow{d} \chi_d^2(\phi^{(a)}), \end{aligned} \tag{A.5}$$

where $\phi^{(a)} = n[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0]'[\mathbf{D}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{D}']^{-1}[\mathbf{D}\boldsymbol{\theta}_0 - \mathbf{d}_0]$. From (A.3) and (A.5), we have that T_W asymptotically follows the noncentral chi-square distribution with d degrees of freedom and noncentrality parameter $\phi^{(a)}$. □

B Proof of Theorem 3:

Proof. From (3.2), we have $n^{-1}\phi^{(a)} = [\beta_{A0}]^2[S_{\mathbf{\Delta}^{[n]}_{22t}}(\boldsymbol{\theta}_0)]$. First, we express the components of $n^{-1}\phi^{(a)}$, namely, $\Delta^{[n]}_{11t}(\boldsymbol{\theta}_0)$, $\mathbf{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_0)$, $\mathbf{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_0)$, and $\mathbf{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_0)$ as functions of $\hat{p}_{Ah}$ for $h = 1, 2, \ldots, m$ as follows:

$$\mathbf{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_0) = n^{-1}\sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta}_0)n_h\hat{p}_{Ah}\begin{pmatrix} 1 & \mathbf{v}_h' & \mathbf{v}_h' \end{pmatrix};$$
$$\mathbf{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_0) = \mathbf{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_0)';$$

$\Delta^{[n]}_{11t}(\boldsymbol{\theta}_0) = n^{-1}\sum_{h=1}^{m}\lambda_{Ah}(\boldsymbol{\theta}_0)n_h\hat{p}_{Ah}$; $\mathbf{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_0) = n^{-1}\sum_{h=1}^{m}\lambda_{Ah}n_h\hat{P}_{Ah}$
$$\begin{pmatrix} 1 & \mathbf{v}_h' & \mathbf{v}_h' \\ \mathbf{v}_h & \mathbf{v}_h\mathbf{v}_h' & \mathbf{v}_h\mathbf{v}_h' \\ \mathbf{v}_h & \mathbf{v}_h\mathbf{v}_h' & \mathbf{v}_h\mathbf{v}_h' \end{pmatrix} + n^{-1}\sum_{h=1}^{m}\lambda_{Bh}n_h\left(1-\hat{P}_{Ah}\right)\begin{pmatrix} 1 & \mathbf{v}_h' & \mathbf{0} \\ \mathbf{v}_h & \mathbf{v}_h\mathbf{v}_h' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$
where $\lambda_{Ah}(\boldsymbol{\theta}) = \exp(-\beta_A - \gamma_0 - \mathbf{v}_h'(\boldsymbol{\gamma}+\boldsymbol{\delta}))[1+\exp(-\beta_A-\gamma_0-\mathbf{v}_h'(\boldsymbol{\gamma}+\boldsymbol{\delta})]^{-2}$, and $\lambda_{Bh}(\boldsymbol{\theta}) = \exp(-\gamma_0-\mathbf{v}_h'\boldsymbol{\gamma})[1+\exp(-\gamma_0-\mathbf{v}_h'\boldsymbol{\gamma})]^{-2}$. Therefore, the noncentrality parameter $\phi^{(a)}$ is a function of $\hat{\mathbf{p}}_A$.

Next, we apply the multivariate version of Taylor's expansion to $\phi(\hat{\mathbf{p}}_A)$ in a neighborhood centered around $\boldsymbol{\pi}_{\boldsymbol{A}}$ to obtain

$$\begin{aligned} n^{-1}\phi(\hat{\mathbf{p}}_A) &= n^{-1}\phi(\boldsymbol{\pi}_{\boldsymbol{A}}) + n^{-1}\phi^{(1)}(\boldsymbol{\pi}_{\boldsymbol{A}})[\hat{\mathbf{p}}_A - \boldsymbol{\pi}_{\boldsymbol{A}}] + 2^{-1}n^{-1}[\hat{\mathbf{p}}_A-\boldsymbol{\pi}_{\boldsymbol{A}}]' \\ &\quad \phi^{(2)}(\boldsymbol{\pi}_{\boldsymbol{A}})[\hat{\mathbf{p}}_A-\boldsymbol{\pi}_{\boldsymbol{A}}] + \mathbf{o}(\parallel \hat{\mathbf{p}}_A - \boldsymbol{\pi}_{\boldsymbol{A}} \parallel^m), \text{ where} \qquad \text{(B.1)} \\ n^{-1}\phi^{(1)}(\boldsymbol{\pi}_{\boldsymbol{A}}) &= n^{-1}\left(\ \partial\phi(\hat{\mathbf{p}}_A)/\partial\hat{\mathbf{p}}_A\ \right)_{\hat{\mathbf{p}}_A=\boldsymbol{\pi}_{\boldsymbol{A}}} \\ &= n^{-1}\left(\ \partial\phi(\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}\ \right)_{\hat{\mathbf{p}}_A=\boldsymbol{\pi}_{\boldsymbol{A}}} \text{ for } h = 1, 2, \ldots, m. \qquad \text{(B.2)} \end{aligned}$$

The partial derivatives in (B.2) are given by
$n^{-1}(\partial\phi(\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}) = \beta_{A0}^2\{[\partial\Delta^{[n]}_{11t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}] - [\partial\mathbf{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}]$
$\mathbf{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A) - \mathbf{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)[\partial\mathbf{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}/\partial\hat{p}_{Ah}]\mathbf{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)$
$- \mathbf{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)\mathbf{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}[\partial\mathbf{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}]\}$, where
$[\partial\mathbf{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}] = n^{-1}\lambda_{Ah}(\boldsymbol{\theta}_0)n_h\begin{pmatrix} 1 & \mathbf{v}_h' & \mathbf{v}_h' \end{pmatrix}$, $[\partial\mathbf{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}] =$
$[\partial\mathbf{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}]'$, $[\partial\Delta^{[n]}_{11t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}] = n^{-1}\lambda_{Ah}(\boldsymbol{\theta}_0)n_h$, and
$[\partial\mathbf{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}/\partial\hat{p}_{Ah}] = -\mathbf{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}\mathbf{\Delta}^{[n]}_{22t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}$, with
$$\mathbf{M}_{Dh} = \frac{n_h}{n}\left[\lambda_{Ah}(\boldsymbol{\theta}_0)\begin{pmatrix} 1 & \mathbf{v}_h' & \mathbf{v}_h' \\ \mathbf{v}_h & \mathbf{v}_h\mathbf{v}_h' & \mathbf{v}_h\mathbf{v}_h' \\ \mathbf{v}_h & \mathbf{v}_h\mathbf{v}_h' & \mathbf{v}_h\mathbf{v}_h' \end{pmatrix} - \lambda_{Bh}(\boldsymbol{\theta}_0)\begin{pmatrix} 1 & \mathbf{v}_h' & \mathbf{0} \\ \mathbf{v}_h & \mathbf{v}_h\mathbf{v}_h' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}\right].$$
To evaluate the term $n^{-1}\phi^{(2)}(\boldsymbol{\pi}_{\boldsymbol{A}})$ in (B.1), we require the following second derivatives, $[\partial^2\mathbf{\Delta}^{[n]}_{12t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}] = [\partial^2\mathbf{\Delta}^{[n]}_{21t}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}]'$ =

$\mathbf{0}_{1\times(2p+1)}$, and $[\partial^2\Delta_{11t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}]=0$ for $h^*=h$ or $h^*\neq h$ and $[\partial^2\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}/\partial\hat{p}_{Ah}\partial\hat{p}_{Ah}]=2\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}$, and $[\partial^2\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}/\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}]=\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh^*}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}+\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh^*}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}$.

Thus, we have $n^{-1}[\partial^2\phi(\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}\partial\hat{p}_{Ah}]=4\beta_{A0}^2p_h\lambda_{Ah}(\boldsymbol{\theta}_0)\begin{pmatrix}1&\mathbf{v}_h'&\mathbf{v}_h'\end{pmatrix}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)\ -\ 2\beta_{A0}^2[p_h\lambda_{Ah}(\boldsymbol{\theta}_0)]^2\begin{pmatrix}1&\mathbf{v}_h'&\mathbf{v}_h'\end{pmatrix}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\begin{pmatrix}1&\mathbf{v}_h'&\mathbf{v}_h'\end{pmatrix}'\ -\ 2\beta_{A0}^2\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)[\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)]^{-1}\mathbf{M}_{Dh}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)$ for $h=1,2,\ldots m$, where $p_h=n_h/n$.

Also, we have $n^{-1}[\partial^2\phi(\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}]=2\beta_{A0}^2p_h\lambda_{Ah}(\boldsymbol{\theta}_0)\begin{pmatrix}1&\mathbf{v}_h'&\mathbf{v}_h'\end{pmatrix}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh^*}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)\ -\ 2\beta_{A0}^2p_h\lambda_{Ah}(\boldsymbol{\theta}_0)p_{h^*}\lambda_{Ah^*}(\boldsymbol{\theta}_0)\begin{pmatrix}1&\mathbf{v}_h'&\mathbf{v}_h'\end{pmatrix}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\begin{pmatrix}1&\mathbf{v}_{h^*}'&\mathbf{v}_{h^*}'\end{pmatrix}'\ +\ 2\beta_{A0}^2p_{h^*}\lambda_{Ah^*}(\boldsymbol{\theta}_0)\begin{pmatrix}1&\mathbf{v}_{h^*}'&\mathbf{v}_{h*}'\end{pmatrix}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)\ -\ 2\beta_{A0}^2\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)$ for $h,h^*=1,2,\ldots m$ and $h\neq h^*$. So $n^{-1}\phi^{(2)}(\boldsymbol{\pi_A})$ can be evaluated by

$$\frac{1}{n}\phi^{(2)}(\boldsymbol{\pi_A}) \quad = \quad \frac{1}{n}\begin{pmatrix}\dfrac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{\mathbf{p}}_A\partial\hat{\mathbf{p}}_A'}\end{pmatrix}_{\hat{\mathbf{p}}_A=\boldsymbol{\pi_A}}. \tag{B.3}$$

Hence the theorem holds. □

C Proof of Theorem 4

Proof. (a) Using the proof of Theorem 3, we have,

$$\begin{aligned}\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)=\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0,\ \hat{\mathbf{p}}_A)'&=n^{-1}\sum_{h=1}^{2}\lambda_{Ah}(\boldsymbol{\theta}_0)n_h\hat{p}_{Ah}\begin{pmatrix}1&\mathrm{v}_h\end{pmatrix},\ \boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)\\&=\sum_{h=1}^{2}\Upsilon_h(\boldsymbol{\theta}_0)\begin{pmatrix}1&\mathrm{v}_h\\\mathrm{v}_h&\mathrm{v}_h\end{pmatrix},and\\\Delta_{11t}^{[n]}(\boldsymbol{\theta}_0,\hat{\mathbf{p}}_A)&=n^{-1}\sum_{h=1}^{2}\lambda_{Ah}(\boldsymbol{\theta}_0)n_h\hat{p}_{Ah},\end{aligned}$$

where $\Upsilon_h(\boldsymbol{\theta}_0)=n^{-1}n_h\left\{\lambda_{Ah}(\boldsymbol{\theta}_0)\hat{p}_{Ah}+\lambda_{Bh}(\boldsymbol{\theta}_0)\left[1-\hat{p}_{Ah}\right]\right\}$, $\lambda_{Ah}(\boldsymbol{\theta}_0)=\exp(-\beta_{A0}-\gamma_{00}-\mathrm{v}_h\gamma_{10})[1+\exp(-\beta_{A0}-\gamma_{00}-\mathrm{v}_h\gamma_{10})]^{-2}$, and $\lambda_{Bh}(\boldsymbol{\theta}_0)=\exp(-\gamma_{00}-\mathrm{v}_h\gamma_{10})[1+\exp(-\gamma_{00}-\mathrm{v}_h\gamma_{10})]^{-2}$ for $h=1,2$; $\mathrm{v}_1=1$ and $\mathrm{v}_2=0$.

Again, applying the multivariate version of Taylor's expansion to $\phi(\hat{\mathbf{p}}_A)$ in a neighborhood centered around $\boldsymbol{\pi_A}$ we obtain,
$n^{-1}\phi(\hat{\mathbf{p}}_A) = n^{-1}\phi(\boldsymbol{\pi_A}) + n^{-1}\phi^{(1)}(\boldsymbol{\pi_A})[\hat{\mathbf{p}}_A - \boldsymbol{\pi_A}] - 2^{-1}n^{-1}[\hat{\mathbf{p}}_A - \boldsymbol{\pi_A}]'[-\phi^{(2)}(\boldsymbol{\pi_A})][\hat{\mathbf{p}}_A - \boldsymbol{\pi_A}]$.
Now, we consider, $n^{-1}\phi^{(1)}(\boldsymbol{\pi_A}) = n^{-1}\left(\partial\phi(\hat{\mathbf{p}}_A)/\partial\hat{\mathbf{p}}_A \right)_{\hat{\mathbf{p}}_A = \boldsymbol{\pi_A}}$, with,
$n^{-1}[\partial\phi(\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}] = \beta_{A0}^2\{[\partial\Delta_{11t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}] - [\partial\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}]\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}$
$\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A) - \boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)[\partial\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}/\partial\hat{p}_{Ah}]\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A) - \boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}$
$[\partial\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}]\}$, where $[\partial\Delta_{11t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}] = n^{-1}\lambda_{Ah}(\boldsymbol{\theta}_0)n_h$,
$[\partial\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}] = [\partial\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}]' = n^{-1}\lambda_{Ah}(\boldsymbol{\theta}_0)n_h \begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix}$,
$[\partial\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}/\partial\hat{p}_{Ah}] = -\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}$, and,
$\mathbf{M}_{Dh} = \left[n_h n^{-1}\{\lambda_{Ah}(\boldsymbol{\theta}_0) - \lambda_{Bh}(\boldsymbol{\theta}_0)\} \begin{pmatrix} 1 & \mathrm{v}_h \\ \mathrm{v}_h & \mathrm{v}_h \end{pmatrix} \right]$. We note that here, $h = 1, 2$, and $\mathrm{v}_1 = 1$, $\mathrm{v}_2 = 0$. Also, we have

$$\frac{1}{n}\phi^{(2)}(\boldsymbol{\pi_A}) = \frac{1}{n}\left(\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{\mathbf{p}}_A\partial\hat{\mathbf{p}}_A'} \right)_{\hat{\mathbf{p}}_A = \boldsymbol{\pi_A}}. \tag{C.1}$$

To evaluate the term $n^{-1}\phi^{(2)}(\boldsymbol{\pi_A})$ in (C.1), we require the following,
$[\partial^2\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}] = [\partial^2\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}] = \mathbf{0}_{1\times 2}$, and
$[\partial^2\Delta_{11t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}] = 0$ for $h^* = h$ or $h^* \neq h$ and
$[\partial^2\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}/\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}] = 2\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh^*}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}$
$\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}$.

The second derivatives in the matrix (C.1) are then given by
$n^{-1}[\partial^2\phi(\hat{\mathbf{p}}_A)/\partial\hat{p}_{Ah}\partial\hat{p}_{Ah^*}] = 2\beta_{A0}^2\{n^{-1}\lambda_{Ah}(\boldsymbol{\theta}_0)n_h \begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix} \boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh^*}$
$\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A) - n^{-2}\lambda_{Ah}(\boldsymbol{\theta}_0)n_h \begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix} \boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\lambda_{Ah^*}(\boldsymbol{\theta}_0)$
$n_{h^*} \begin{pmatrix} 1 & \mathrm{v}_{h^*} \end{pmatrix}' + n^{-1}\lambda_{Ah^*}(\boldsymbol{\theta}_0)n_{h^*} \begin{pmatrix} 1 & \mathrm{v}_{h^*} \end{pmatrix} \boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}$
$\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A) - \boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh^*}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{Dh}$
$\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)\}$, for $h, h^* = 1, 2$.
Using the fact that $\mathrm{v}_1 = 1$ and $\mathrm{v}_2 = 0$, we obtain the following results
$\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A) = \begin{pmatrix} \Upsilon_1(\boldsymbol{\theta}_0) + \Upsilon_2(\boldsymbol{\theta}_0) & \Upsilon_1(\boldsymbol{\theta}_0) \\ \Upsilon_1(\boldsymbol{\theta}_0) & \Upsilon_1(\boldsymbol{\theta}_0) \end{pmatrix}$ with
$\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1} = \Upsilon_1(\boldsymbol{\theta}_0)^{-1}\Upsilon_2(\boldsymbol{\theta}_0)^{-1} \begin{pmatrix} \Upsilon_1(\boldsymbol{\theta}_0) & -\Upsilon_1(\boldsymbol{\theta}_0) \\ -\Upsilon_1(\boldsymbol{\theta}_0) & \Upsilon_1(\boldsymbol{\theta}_0) + \Upsilon_2(\boldsymbol{\theta}_0) \end{pmatrix}$. Recall that,

$\mathbf{M}_{Dh} = n^{-1}n_h\{\lambda_{Ah}(\boldsymbol{\theta}_0) - \lambda_{Bh}(\boldsymbol{\theta}_0)\}\begin{pmatrix} 1 & \mathrm{v}_h \\ \mathrm{v}_h & \mathrm{v}_h \end{pmatrix} = \mu_h(\boldsymbol{\theta}_0)\begin{pmatrix} 1 & \mathrm{v}_h \\ \mathrm{v}_h & \mathrm{v}_h \end{pmatrix}$, where $\mu_h(\boldsymbol{\theta}_0) = n^{-1}n_h\{\lambda_{Ah}(\boldsymbol{\theta}_0) - \lambda_{Bh}(\boldsymbol{\theta}_0)\}$ and $\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A) = n^{-1}\sum_{h=1}^{2}\lambda_{Ah}(\boldsymbol{\theta}_0)$ $n_h\hat{p}_{Ah}\begin{pmatrix} 1 & \mathrm{v}_h \end{pmatrix} = \begin{pmatrix} C_1(\boldsymbol{\theta}_0) + C_2(\boldsymbol{\theta}_0) & C_1(\boldsymbol{\theta}_0) \end{pmatrix}_{1\times 2}$, with $C_h(\boldsymbol{\theta}_0) = D_h(\boldsymbol{\theta}_0)\hat{p}_{Ah}$, and $D_h(\boldsymbol{\theta}_0) = n^{-1}\lambda_{Ah}(\boldsymbol{\theta}_0)n_h$. It follows that,

$\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{D1}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1} = \{\mu_1(\boldsymbol{\theta}_0)/\Upsilon_1(\boldsymbol{\theta}_0)^2\}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$; $\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}$ $\mathbf{M}_{D2}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1} = [\mu_2(\boldsymbol{\theta}_0)/\Upsilon_2(\boldsymbol{\theta}_0)^2]\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$; $\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{D1}$ $\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A) = [\mu_1(\boldsymbol{\theta}_0)C_1(\boldsymbol{\theta}_0)/\Upsilon_1(\boldsymbol{\theta}_0)^2]\begin{pmatrix} 0 & 1 \end{pmatrix}'$; $\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}$ $\mathbf{M}_{D2}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A) = [\mu_2(\boldsymbol{\theta}_0)C_2(\boldsymbol{\theta}_0)/\Upsilon_2(\boldsymbol{\theta}_0)^2]\begin{pmatrix} 1 & -1 \end{pmatrix}'$; $\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{D1}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{D1}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A) = [\mu_1(\boldsymbol{\theta}_0)C_1(\boldsymbol{\theta}_0)]^2/\Upsilon_1(\boldsymbol{\theta}_0)^3$; $\boldsymbol{\Delta}_{12t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{D2}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\mathbf{M}_{D2}$ $\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}\boldsymbol{\Delta}_{21t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A) = [\mu_2(\boldsymbol{\theta}_0)C_2(\boldsymbol{\theta}_0)]^2/\Upsilon_2(\boldsymbol{\theta}_0)^3$ and $\mathbf{M}_{D2}\boldsymbol{\Delta}_{22t}^{[n]}(\boldsymbol{\theta}_0, \hat{\mathbf{p}}_A)^{-1}$ $\mathbf{M}_{D1} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. The elements of (C.1) can be further simplified as follows,

$$\begin{aligned}
\frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{A1}} &= -\frac{2[\beta_{A0}]^2}{\Upsilon_1(\boldsymbol{\theta}_0)}\left\{D_1(\boldsymbol{\theta}_0) - \frac{\mu_1(\boldsymbol{\theta}_0)}{\Upsilon_1(\boldsymbol{\theta}_0)}C_1(\boldsymbol{\theta}_0)\right\}^2 < 0, \\
\frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}\partial\hat{p}_{A2}} &= -\frac{2[\beta_{A0}]^2}{\Upsilon_2(\boldsymbol{\theta}_0)}\left\{D_2(\boldsymbol{\theta}_0) - \frac{\mu_2(\boldsymbol{\theta}_0)}{\Upsilon_2(\boldsymbol{\theta}_0)}C_2(\boldsymbol{\theta}_0)\right\}^2 < 0, \\
\frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{A2}} &= \frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}\partial\hat{p}_{A1}} = 0.
\end{aligned}$$

So, the determinant of $n^{-1}\phi^{(2)}(\boldsymbol{\pi}_{\boldsymbol{A}})$ in (C.1) is positive and $n^{-1}[\partial^2\phi(\hat{\mathbf{p}}_A)/\partial\hat{p}_{A1}\partial\hat{p}_{A1}]$ is negative. That is, $n^{-1}\phi^{(2)}(\hat{\mathbf{p}}_A)$ is a negative definite matrix. Therefore, $\phi(\hat{\mathbf{p}}_A)$ is a concave down function.

(b) We consider the variability induced by the randomization process

$$\begin{aligned}
&(\hat{\mathbf{p}}_A - \boldsymbol{\pi}_{\boldsymbol{A}})'\, n^{-1}\phi^{(2)}(\boldsymbol{\pi}_{\boldsymbol{A}})\,(\hat{\mathbf{p}}_A - \boldsymbol{\pi}_{\boldsymbol{A}}) \\
&= \begin{pmatrix} \hat{p}_{A1} - \pi_A(\mathrm{v}_1) \\ \hat{p}_{A2} - \pi_A(\mathrm{v}_2) \end{pmatrix}' \begin{pmatrix} \frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{A1}} & 0 \\ 0 & \frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}\partial\hat{p}_{A2}} \end{pmatrix} \begin{pmatrix} \hat{p}_{A1} - \pi_A(\mathrm{v}_1) \\ \hat{p}_{A2} - \pi_A(\mathrm{v}_2) \end{pmatrix} \\
&= \frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A1}\partial\hat{p}_{A1}}(\hat{p}_{A1} - \pi_A(\mathrm{v}_1))^2 + \frac{1}{n}\frac{\partial^2\phi(\hat{\mathbf{p}}_A)}{\partial\hat{p}_{A2}\partial\hat{p}_{A2}}(\hat{p}_{A2} - \pi_A(\mathrm{v}_2))^2. \qquad \text{(C.2)}
\end{aligned}$$

Also, we have

$$(\hat{\mathbf{p}}_A - \boldsymbol{\pi}_{\boldsymbol{A}})' (\hat{\mathbf{p}}_A - \boldsymbol{\pi}_{\boldsymbol{A}}) \quad = \quad (\hat{p}_{A1} - \pi_A(\mathrm{v}_1))^2 + (\hat{p}_{A2} - \pi_A(\mathrm{v}_2))^2. \tag{C.3}$$

For $i = 1, 2$, $\dfrac{1}{n} \dfrac{\partial^2 \phi(\hat{\mathbf{p}}_A)}{\partial \hat{p}_{Ai} \partial \hat{p}_{Ai}}$, is negative and does not depend on $\hat{p}_{Ai}$. Therefore, from (C.2) and (C.3), we have that $(\hat{\mathbf{p}}_A - \boldsymbol{\pi}_{\boldsymbol{A}})' \left(-n^{-1}\phi^{(2)}(\boldsymbol{\pi}_{\boldsymbol{A}})\right) (\hat{\mathbf{p}}_A - \boldsymbol{\pi}_{\boldsymbol{A}})$ decreases if and only if $(\hat{\mathbf{p}}_A - \boldsymbol{\pi}_{\boldsymbol{A}})' (\hat{\mathbf{p}}_A - \boldsymbol{\pi}_{\boldsymbol{A}})$ decreases. Thus, the noncentrality parameter in (3.2) increases as the design variability decreases. As a result, the power of the hypothesis test increases as the design variability decreases. □

Acknowledgements

Thanks to the referees whose comments have led to improvements in the quality of this paper. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada.

References

Anderson, T. W. (1966), *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

Austin, P. C. and Steyerberg, E. W. (2017), "Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models," *Statistical Methods in Medical Research*, 26, 796-808.

Bujang, M. A., Sa'at, N., Ikhwan Tg, T., Sidik, A. B. and Joo1, L. C. (2018), "Sample Size Guidelines for Logistic Regression from Observational Studies with Large Population: Emphasis on the Accuracy Between Statistics and Parameters Based on Real Life Clinical Data," *Malays J Med Sci*, 25, 122-130.

Demidenko, E. (2007), "Sample size determination for logistic regression revisited," *Statistics in Medicine*, 26, 3385–3397.

Hart, R. G., Halperin, J. L., Pearce, L. A., Anderson, D. C., Kronmal, R. A., McBride, R.,..., Marler, J. R. (2003), "Lessons from the stroke prevention in atrial fibrillation trials," *Annals of Internal Medicine*, 138, 831–838.

Hu, F. and Rosenberger, W. F. (2003), "Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons," *Journal of the American Statistical Association*, 98, 671–678.

Ma, W., Hu, F. and Zhang, L. (2015), "Testing hypotheses of covariate-adaptive randomized clinical trials," *Journal of the American Statistical Association*, 110, 669–680.

Pocock, S. J. and Simon, R. (1975), "Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial," *Biometrics*, 31, 103–115.

Seber, G. A. F. and Lee, A. J. (2003), *Linear regression analysis*, Wiley, Hoboken, New Jersey.

Self, S. G. and Mauritsen, R. H. (1988), "Power/sample size calculations for generalized linear models," *Biometrics*, 44, 79–86.

Self, S. G., Mauritsen, R. H. and Ohara, J. (1992), "Power calculations for likelihood ratio tests in generalized linear models," *Biometrics*, 48, 31–39.

Selvaratnam, S., Oyet, A. J., Yi, Y. and Gadag, V. (2017), "Estimation of a generalized linear mixed model for response-adaptive designs in multi-center clinical trials," *Canadian Journal of Statistics*, 45, 310–325.

Selvaratnam, S., Yi, Y. and Oyet, A. J. (2019), "Maximum likelihood estimation of generalized linear models for adaptive designs: applications and asymptotics," *Biometrical Journal*, 61, 630–651.

Shieh, G. A. (2000), "Comparison of two approaches for power and sample size calculations in logistic regression models," *Communications in Statistics - Simulation and Computation*, 29, 763–791.

Shieh, G. A. (2005), "On power and sample size calculations for Wald tests in generalized linear models," *Journal of Statistical Planning and Inference*, 128, 43–59.

Smeden, M. V., Moons, K. GM., De Groot, J. AH., Collins, G. S., Altman, D. G., Eijkemans, M. JC. and Reitsma, J. B. (2019), "Sample size for binary logistic prediction models: Beyond events per variable criteria," *Statistical Methods in Medical Research*, 28, 2455-2474.

Srivastava, M. S. (2002, *Methods of Multivariate Statistics*, Wiley, Toronto.

Whittemore, A. S. (1981), "Sample size for logistic regression with small response probability," *Journal of the American Statistical Association*, 76, 27–32.

Received: *21 November 2021*

Accepted: *22 April 2022*