The International Journal of Biostatistics

Volume 7, Issue 1

2011

Article 10

Targeting the Optimal Design in Randomized Clinical Trials with Binary Outcomes and No Covariate: Theoretical Study

Antoine Chambaz, *Laboratoire MAP5, Université Paris
Descartes and CNRS*

Mark J. van der Laan, *University of California, Berkeley*

Recommended Citation:

Chambaz, Antoine and van der Laan, Mark J. (2011) "Targeting the Optimal Design in Randomized Clinical Trials with Binary Outcomes and No Covariate: Theoretical Study," *The International Journal of Biostatistics*: Vol. 7: Iss. 1, Article 10.

DOI: 10.2202/1557-4679.1247

Targeting the Optimal Design in Randomized Clinical Trials with Binary Outcomes and No Covariate: Theoretical Study

Antoine Chambaz and Mark J. van der Laan

Abstract

This article is devoted to the asymptotic study of adaptive group sequential designs in the case of randomized clinical trials (RCTs) with binary treatment, binary outcome and no covariate. By adaptive design, we mean in this setting a RCT design that allows the investigator to dynamically modify its course through data-driven adjustment of the randomization probability based on data accrued so far, without negatively impacting on the statistical integrity of the trial. By adaptive group sequential design, we refer to the fact that group sequential testing methods can be equally well applied on top of adaptive designs. We obtain that, theoretically, the adaptive design converges almost surely to the targeted unknown randomization scheme. In the estimation framework, we obtain that our maximum likelihood estimator of the parameter of interest is a strongly consistent estimator, and it satisfies a central limit theorem. We can estimate its asymptotic variance, which is the same as that it would feature had we known in advance the targeted randomization scheme and independently sampled from it. Consequently, inference can be carried out as if we had resorted to independent and identically distributed (iid) sampling. In the testing framework, we obtain that the multidimensional t-statistic that we would use under iid sampling still converges to the same canonical distribution under adaptive sampling. Consequently, the same group sequential testing can be carried out as if we had resorted to iid sampling.

Furthermore, a comprehensive simulation study that we undertake in a companion article validates the theory.

A three-sentence take-home message is “Adaptive designs do learn the targeted optimal design and inference, and testing can be carried out under adaptive sampling as they would under the targeted optimal randomization probability iid sampling. In particular, adaptive designs achieve the same efficiency as the fixed oracle design. This is confirmed by a simulation study, at least for moderate or large sample sizes, across a large collection of targeted randomization probabilities.”

KEYWORDS: adaptive design, canonical distribution, coarsening at random, contiguity, group sequential testing, maximum likelihood, randomized clinical trial

Author Notes: This collaboration took place while Antoine Chambaz was a visiting scholar at UC Berkeley, supported in part by a Fulbright Research Grant and the CNRS. The authors thank the reviewers for their interesting comments.

Erratum

Note from the authors: Mark van der Laan was partially supported by NIAID grant R01 AI074345-05.

1 Introduction

The present article and its companion article (Chambaz and van der Laan, 2010b) are devoted to the asymptotic study of *adaptive group sequential designs* in the case of randomized clinical trials with binary treatment, binary outcome and no covariate, *focusing here on its theoretical development and there on its study by simulations*. Thus, the experimental unit writes as $O = (A, Y)$ where the treatment A and the outcome Y are dependent Bernoulli random variables. Typical parameters of scientific interest are $\Psi_+ = E(Y|A = 1) - E(Y|A = 0)$ (additive scale) or $\Psi_\times = \log E(Y|A = 1) - \log E(Y|A = 0)$ (multiplicative scale, which we consider hereafter). One can interpret causally such parameters whenever one is willing to postulate the existence of a full data structure $X = (X(0), X(1))$ containing the two counterfactual outcomes under the two possible treatments and such that $Y = X(A)$ and A independent of X . If so indeed, $\Psi_+ = E(X(1)) - E(X(0))$ and $\Psi_\times = \log E(X(1)) - \log E(X(0))$. Let us now explain what we mean by adaptive group sequential design.

1.1 The notion of adaptive group sequential designs.

By *adaptive design*, we mean in this setting a clinical trial design that allows the investigator to dynamically modify its course through data-driven adjustment of the randomization probability based on data accrued so far, without negatively impacting on the statistical integrity of the trial. This definition is slightly adapted from (Golub, 2006), the introductory article to the proceedings (to which many articles cited below belong) of a workshop entitled “Adaptive clinical trial designs: Ready for prime time?” held in October 2004, and jointly sponsored by the FDA and Harvard-MIT Division of Health Science and Technology. Using the definition of prespecified sampling plans given in (Emerson, 2006), let us emphasize that we assume that, prior to collection of the data, the trial protocol specifies: the parameter of scientific interest, and

- estimation framework: the confidence level to be used in constructing frequentist confidence intervals for the latter parameter, the related inferential method;
- testing framework: the null and alternative hypotheses regarding the latter parameter, the wished type I and type II errors, the rule for determining the maximal statistical information to be accrued, the frequentist testing procedure (including conditions for early stopping).

Furthermore, we assume that the protocol specifies a user-supplied optimal unknown choice of randomization scheme: our adaptive design does not belong to the class of prespecified sampling schemes in that it targets the latter optimal unknown choice of randomization scheme, learning it based on accrued data. We focus in this article on maximum likelihood estimation and testing. The considered user-supplied optimal unknown choice of randomization scheme is that which minimizes the asymptotic variance of our maximum likelihood estimator (MLE) of the parameter of interest. This choice, known in the literature as the *Neyman allocation*, is interesting because minimizing the asymptotic variance of our estimator guarantees narrower confidence intervals and earlier decision to reject the null for its alternative or not. Yet, targeting this treatment mechanism in real-life clinical trials may raise ethical issues, since this may result in more patients assigned to the inferior treatment arm. But we emphasize that there is nothing special about targeting the Neyman allocation, the whole methodology applying equally well to any choice of targeted treatment mechanism. To be more specific, we could also have chosen to target that treatment mechanism which minimizes the expected number of treatment failures subject to power constraint (the *optimal allocation* proposed by Rosenberger, Stallard, Ivanova, Harper, and Ricks (2001)), or any other treatment mechanism of interest (see for instance (Hu and Rosenberger, 2006, Chapter 2) or (Tymofeyev, Rosenberger, and Hu, 2007)). Quoting James Hung of the FDA (about design adaptation in general—see (Hung, 2006)), our adaptive design meets clearly stated objectives, it is certainly a “more careful planning, not sloppy planning”.

By adaptive *group sequential* design, we refer to the fact that group sequential testing methods can be equally well applied on top of adaptive designs. According to Hu and Rosenberger (2006), “The basic statistical formulation of a sequential testing procedure requires determining the joint distribution of the sequentially computed test statistics. Under response-adaptive randomization, this is a difficult task. There has been little theoretical work done to this point, nor has there been any evaluation of sequential monitoring in the context of sequential estimation procedures [*i.e.* targeted adaptive designs] such as the double adaptive biased coin design.” These authors end their final chapter with a quote from (Rosenberger, 2002): “Surprisingly, the link between response-adaptive randomization and sequential analysis has been tenuous at best, and this is perhaps the logical place to search for open research topics.” Indeed, we determine the limit joint distribution of the sequentially computed test statistics based on our results, and provide a theoretical background to rely upon for adaptive design group sequential testing procedures. Simultaneously, a similar result is obtained in (Zhu and Hu, 2010) through a different approach based on the study of the limit distribution of a stochastic process defined over $(0, 1]$.

1.2 Bibliography.

The literature on adaptive designs is vast and we apologize for not including all of it. Quite misleadingly, the expression “adaptive design” has also been used in the literature for sequential testing and, in general, for designs that allow data-adaptive stopping times for the whole study (or for certain treatment arms) which achieve the wished type I and type II errors requirements when testing a null hypothesis against its alternative.

In the literature dedicated to what corresponds to our definition of adaptive design, such adaptive designs are referred to as “response-adaptive randomization” designs (see the quote from (Hu and Rosenberger, 2006) above). Of course, data-adaptive randomization schemes have a long history, that goes back to the 1930s, and we refer to (Hu and Rosenberger, 2006, Section 1.2) and (Jennison and Turnbull, 2000, Section 17.4) to provide the reader with a comprehensive historical perspective.

The organization of the Section 1.1 illustrates the fact that we have decided to tackle separately the group sequential testing problem from the data-adaptive determination of the randomization probability in response to data collected so far—a choice justified by the fact that separating the characterization of a group sequential testing procedure from the adaptation of the randomization probability makes perfect sense from a methodological point of view. Resorting to the same organization here is more delicate, because response-adaptive treatment allocation is indebted to early studies in the context of sequential statistical analysis, such as (Armitage, 1975, Chernoff and Roy, 1965, Flehinger and Louis, 1971) among many others. However, Hu and Rosenberger (2006) manage to trace back the idea of incorporating randomization in the context of adaptive treatment allocation designs to (Wei and Durham, 1978).

On the one hand, regarding the adaptation of the randomization probability, data-adaptive randomization schemes belong to either the “urn model” or “double adaptive biased coin” families (see the quote from (Hu and Rosenberger, 2006) above). Adaptive designs based on urn models (so called because the randomization scheme can be modeled after different ways of pulling various colored balls from an urn) notably include the seminal “randomized play-the-winner rule” from the aforementioned article (Wei and Durham, 1978) or the more recent “drop-the-loser rule” (Ivanova, 2003). The theory of adaptive designs based on *homogeneous* urn models (referring to the fact that the updating rule does not evolve through time) is presented in detail in (Hu and Rosenberger, 2006, Chapter 4), with a comprehensive bibliography. *Nonhomogeneous* urn models are adaptive designs based on urn models which target a randomization scheme. Also known in the literature as estimation-adjusted urn models, these adaptive designs involve updating rules of

the urn which rely on the estimation of some parameters based on data accrued so far. Recently, Zhang, Hu, and Cheung (2006) established the consistency and asymptotic normality for both the randomization scheme and the estimator of the parameter of interest in this challenging framework, under weak assumptions. Targeting a specific user-supplied optimal unknown choice of randomization scheme is at the core of adaptive designs based on flipping a (data-adaptively) biased coin. More precisely, the latter targeted randomization scheme is expressed as a function $f(\theta)$ of an unknown parameter θ of the response model, and the adaptive design is characterized by the sequence $f(\theta_n)$ which is based on updated estimates θ_n of θ as the data accrue. For instance, the targeted randomization scheme we consider (namely, the minimizer of the asymptotic variance of our MLE of the parameter of interest in clinical trials with binary treatment, binary outcome and no covariate) is a function of the two marginal probabilities of success. Again, Hu and Rosenberger (2006) manage to trace back this kind of procedure to (Eisele, 1994). A series of articles including (Rosenberger et al., 2001, Hu and Zhang, 2004b) address the theoretical study of such adaptive designs, or investigate their properties based on simulations (Hu and Rosenberger, 2003). Overall, the most relevant references for our present article certainly are (Hu and Rosenberger, 2006) (already cited many times), which concerns asymptotic theory for likelihood-based estimation (not testing) based on data-adaptive randomization schemes in clinical trials, and (Zhu and Hu, 2010). In the latter article, Zhu and Hu mainly derive the limit joint distribution of their sequentially computed test statistic and carry out a simulation study of its properties. We compare their results with ours in the appropriate sections of (Chambaz and van der Laan, 2010b).

On the other hand, regarding the group sequential testing problem, let us emphasize that we consider the case where one starts with a large up-front commitment sample size and uses group sequential testing to allow early stopping rather than starting out with a small commitment of sample size and extending it if necessary—the latter distinction is taken from (Mehta and Patel, 2006). Therefore, negative results obtained *e.g.* in (Jennison and Turnbull, 2003, Tsiatis and Mehta, 2003) for such procedures (inconveniently referred to as adaptive designs methods in (Mehta and Patel, 2006, Tsiatis and Mehta, 2003)) that start out with a small commitment do not apply at all to our procedure. On the contrary, we can build upon the thorough understanding of group sequential methods as exposed in (Jennison and Turnbull, 2000, Proschan, Lan, and Wittes, 2006), and more recently explored in (Lokhnygina and Tsiatis, 2008).

Furthermore, there is also a rich literature on the Bayesian approach to adaptive designs. The reader is referred to (Berry and Stangl, 1996, J., Abrams, and Myles, 2004, Berry, 2006, Banerjee and Tsiatis, 2006) for further details.

Finally, this article builds upon the seminal technical report (van der Laan, 2008) which paves the way to robust and efficient estimation in randomized clinical trials thanks to adaptation of the design in a variety of settings.

1.3 Forthcoming results in words.

Following the same presentation as in Section 1.1, we obtain that the adaptive design converges almost surely to the targeted unknown randomization scheme (Theorem 1), and that

- estimation framework: the MLE of the parameter of interest is a strongly consistent estimator (Theorem 1), it satisfies a central limit theorem (Theorem 2); we can estimate its asymptotic variance, which is the same as that it would feature had we known in advance the targeted randomization scheme and independently sampled from it (Theorem 2); consequently, inference can be carried out as if we had resorted to independent and identically distributed (iid) sampling;
- testing framework: the multidimensional t -statistics that we would use under iid sampling still converges to the same canonical distribution under adaptive sampling (Theorem 3); consequently, the same group sequential testing can be carried out as if we had resorted to iid sampling.

Furthermore, the comprehensive simulation study that we undertake in (Chambaz and van der Laan, 2010b) validates the theory, notably showing that the confidence intervals we obtain achieve the desired coverage even for moderate sample sizes and that type I error control at the prescribed level is guaranteed, and that all sampling procedures only suffer from a very slight increase of the type II error—see (Chambaz and van der Laan, 2010b) for the details.

A three-sentence take-home message is “Adaptive designs do learn the targeted optimal design and inference and testing can be carried out under adaptive sampling as they would under the targeted optimal randomization probability iid sampling. In particular, adaptive designs achieve the same efficiency as the fixed oracle design. This is confirmed by a simulation study, at least for moderate or large sample sizes, across a large collection of targeted randomization probabilities.”

In essence, everything works as predicted by theory. However, theory also warns us that gains cannot be dramatic in the particular setting of clinical trials with binary treatment, binary outcome and no covariate. Nonetheless, this article and its companion article (Chambaz and van der Laan, 2010b) are important: they provide

a theoretical template and tools for asymptotic analysis of robust adaptive designs in less constrained settings, which we will consider in future work. This notably includes the setting of clinical trials *with covariate*, binary treatment, and *discrete or continuous outcome*, or the setting of clinical trials *with covariate*, binary treatment, and *possibly censored time-to-event* among others. Resorting to targeted maximum likelihood estimation (van der Laan and Rubin, 2006) along with adaptation of the design provides substantial gains in efficiency.

Finally, we want to emphasize that the whole adaptive design methodology that we develop here is only relevant for clinical trials in which a substantial number of observations are available before all patients are randomized. From now on, we assume that the clinical trial's time scale permits the application of the adaptive design methodology. It must be noted that one can find in the literature some articles devoted to the study of response-adaptive clinical trials based on possibly delayed responses (see Section 7.1 in (Hu and Rosenberger, 2006), (Hu and Zhang, 2004a, Zhang, Chan, Cheung, and Hu, 2007)). A typical sufficient condition which allows handling the possible delay is the following: the probability of the event "the outcome of the n th patient is unavailable when the $(n+m)$ th patient is randomized" is upper-bounded by a constant times $m^{-\gamma}$ for some $\gamma > 0$.

1.4 Organization of the article.

The article is organized as follows. We define the targeted optimal design in Section 2, and describe how to adapt to it in Section 3. The asymptotic study of the MLE of the parameter of interest under adaptive design is addressed in Section 4, where we obtain that the MLE is strongly consistency and asymptotically Gaussian. In Section 5 we show how a group sequential testing procedure can be applied on top of the adaptive design methodology. In Appendix A.1 we present an important building block for consistency results. It consists of a uniform Kolmogorov strong law of large numbers for martingales sums that essentially relies on a maximal inequality for martingales. Another important building block for central limit theorems is presented in Appendix A.2, where we derive a central limit theorem for discrete martingales.

Finally, in order to ease the reading, we highlight throughout this article and its companion article (Chambaz and van der Laan, 2010b) the most important results. We notably point out here how to construct confidence intervals and how to apply a group sequential testing procedure while targeting the optimal design and thus accruing observations data-adaptively, see the two highlights entitled

1. *pointwise estimation and confidence interval* (Section 4),
2. *targeted optimal design adaptive group sequential testing* (Section 5.1).

2 Balanced versus optimal treatment mechanisms

2.1 The observed data structure and related likelihood.

We consider the simplest example of randomized trials, where an observation writes as $O = (A, Y)$, A being a binary treatment of interest and Y a binary outcome of interest. We postulate the existence of a full data structure $X = (X(0), X(1))$ containing the two counterfactual (or potential) outcomes under the two possible treatments. The observed data structure $O = (A, X(A)) = (A, Y)$ only contains the outcome corresponding to the treatment the experimental unit did receive. Therefore O is a missing data structure on X with missingness variable A .

We denote the conditional probability distributions of treatment A by

$$g(a|x) = P(A = a|X = x).$$

We assume that the *coarsening at random* (abbreviated to *CAR*) assumption holds: for all $a \in \mathcal{A} = \{0, 1\}, x \in \mathcal{X} = \{0, 1\}^2$,

$$g(a|x) = P(A = a|X(a) = x(a)), \quad (1)$$

We denote by $\mathcal{G}$ the set of such CAR conditional distributions of A given X , referred to as the set of fixed designs. In the framework of this article, (1) is equivalent to

$$g(a|x) = g(a)$$

for all $a \in \mathcal{A}, x \in \mathcal{X}$: $g \in \mathcal{G}$ if and only if the random variables A and X are independent. We only consider such treatment mechanisms in the rest of Section 2.

The distribution P_X of the full data structure X has two marginal Bernoulli laws characterized by $\theta = (\theta_0, \theta_1) \in]0, 1]^2$ with $\theta_0 = E_{P_X} X(0)$ and $\theta_1 = E_{P_X} X(1)$ (the only identifiable part of P_X). Therefore, introducing $\mathcal{X}(O) = \{x \in \mathcal{X} : x(A) = Y\}$ (the set of full data structure realizations compatible with O), the likelihood of O writes as

$$\begin{aligned} L(O) &= \sum_{x \in \mathcal{X}(O)} P(O, X = x) = \sum_{x \in \mathcal{X}(O)} P(O|X = x)P(X = x) \\ &= \sum_{x \in \mathcal{X}(O)} g(A|x)P(X = x) = g(A)P(X \in \mathcal{X}(O)) \\ &= \theta_A^Y (1 - \theta_A)^{1-Y} g(A) = \theta(O)g(A), \end{aligned}$$

using the convenient shorthand notation $\theta(O) = \theta_A^Y (1 - \theta_A)^{1-Y}$. Because of the form of the likelihood, we can say that the observed data structure O is obtained under (θ, g) .

2.2 Efficient influence curve and efficient asymptotic variance for the log-relative risk.

Say that the parameter of scientific interest is

$$\Psi(\theta) = \log \frac{\theta_1}{\theta_0},$$

the log-relative risk; of course, the sequel applies to other choices, such as the excess risk.

In a classical randomized trial, we would determine a fixed treatment mechanism g (therefore complying with CAR) and sample as many iid copies of O as necessary.

The theory of semiparametric statistics teaches us that the efficient influence curves for parameters θ_0 and θ_1 under (θ, g) are respectively

$$D_0^*(\theta, g)(O) = (Y - \theta_0) \frac{\mathbb{I}\{A = 0\}}{g(0)}, D_1^*(\theta, g)(O) = (Y - \theta_1) \frac{\mathbb{I}\{A = 1\}}{g(1)}. \quad (2)$$

Then the delta-method (and page 386 in (van der Vaart, 1998)) implies that the efficient influence curve for parameter $\Psi(\theta)$ under (θ, g) writes as

$$\text{IC}(\theta, g)(O) = -\frac{\mathbb{I}\{A = 0\}}{\theta_0 g(0)}(Y - \theta_0) + \frac{\mathbb{I}\{A = 1\}}{\theta_1 g(1)}(Y - \theta_1) \quad (3)$$

so that the efficient asymptotic variance under (θ, g) is

$$E_{\theta, g} \text{IC}(\theta, g)^2(O) = \frac{\sigma^2(\theta)(0)}{\theta_0^2 g(0)} + \frac{\sigma^2(\theta)(1)}{\theta_1^2 g(1)} = \frac{1 - \theta_0}{\theta_0 g(0)} + \frac{1 - \theta_1}{\theta_1 g(1)}$$

with notation $\sigma^2(\theta)(a) = \text{Var}_\theta(Y|A = a) = \theta_a(1 - \theta_a)$, the conditional variance of Y given $A = a$.

2.3 A relative efficiency criterion.

Defining $\text{OR}(\theta) = \frac{\theta_1/(1-\theta_1)}{\theta_0/(1-\theta_0)}$, the efficient asymptotic variance as a function of the treatment mechanism g is minimized at the *optimal treatment mechanism* characterized by

$$g^*(\theta)(1) = \frac{1}{1 + \sqrt{\text{OR}(\theta)}} = \frac{\sqrt{\theta_0(1 - \theta_1)}}{\sqrt{\theta_0(1 - \theta_1)} + \sqrt{\theta_1(1 - \theta_0)}}, \quad (4)$$

known as the Neyman allocation (Hu and Rosenberger, 2006, page 13). Interestingly, $g^*(\theta)(1) \leq \frac{1}{2}$ whenever $\theta_0 \leq \theta_1$, meaning that the Neyman allocation $g^*(\theta)$

favors the inferior treatment. The corresponding optimal efficient asymptotic variance $v^*(\theta)$ then satisfies

$$v^*(\theta) = \left(\sqrt{\frac{1-\theta_0}{\theta_0}} + \sqrt{\frac{1-\theta_1}{\theta_1}} \right)^2 \geq 2 \left(\frac{1-\theta_0}{\theta_0} + \frac{1-\theta_1}{\theta_1} \right) = v^b(\theta),$$

where $v^b(\theta)$ denotes the efficient asymptotic variance associated with the standard balanced treatment characterized by $g^b(1) = \frac{1}{2}$, hence the relative efficiency criterion

$$R(\theta) = \frac{v^*(\theta)}{v^b(\theta)} = \frac{1}{2} + \frac{\sqrt{\text{OR}(\theta)}}{1 + \text{OR}(\theta)} \in \left(\frac{1}{2}, 1\right]. \quad (5)$$

The definition of our relative efficiency criterion illustrates the fact that we decide to consider the balanced treatment mechanism as a benchmark. We emphasize that *any* fixed design could be chosen as benchmark treatment mechanism, with minor impact on the study we expose below.

It is worth noting that $v^b(\theta) = v^*(\theta)$, or in other words that the so-called balanced treatment mechanism is actually optimal, if and only if $\theta_0 = \theta_1$. In particular, there is no gain to expect from adapting the treatment mechanism in terms of type I error control when testing the null “ $\Psi(\theta) = 0$ ” against its negation. In addition, the following bound involving the relative efficiency criterion on one side and the log-relative risk on the other holds:

$$R(\theta) \leq \frac{1}{2} + \frac{\sqrt{e^{\Psi(\theta)}}}{1 + e^{\Psi(\theta)}} \in \left(\frac{1}{2}, 1\right].$$

We present in Figure 1 three curves $\theta_1 \mapsto R(\theta)$ for three different values of θ_0 . It notably illustrates that when θ_0 is small, $R(\theta)$ can be significantly lower than 1 for values of $\Psi(\theta)$ which are not very large. For instance, $\theta = (\frac{1}{100}, \frac{5}{100})$ yields $\Psi(\theta) \simeq 1.609$, $R(\theta) \simeq 0.868$ and optimal treatment mechanism characterized by $g^*(\theta)(1) \simeq 0.305$. Were we given the optimal treatment mechanism in advance, we would obtain confidence intervals (based on the central limit theorem and Slutsky’s lemma) whose widths are approximately $\sqrt{R(\theta)} \simeq 0.931$ times those of the corresponding confidence intervals we would have got using the balanced treatment mechanism.

However the gain could actually be more dramatic than the previous example suggests.

Let us consider again the testing setting: we want to test the null “ $\Psi(\theta) = 0$ ” against the alternative “ $\Psi(\theta) > 0$ ” with type I error α and power $(1 - \beta)$ at some

Figure 1: Plot of the relative efficiency $R(\theta)$ as a function of θ_1 for different values of θ_0 ($\theta = (\theta_0, \theta_1)$). The solid, dashed and dotted curves respectively correspond to $\theta_0 = \frac{1}{2}, \frac{1}{10}, \frac{1}{100}$.

user-defined alternative $\psi > 0$. By the delta-method, we know that the MLE of $\Psi(\theta)$

$$\Psi_n = \log \frac{\sum_{i=1}^n Y_i \mathbb{I}\{A_i = 1\}}{\sum_{i=1}^n \mathbb{I}\{A_i = 1\}} - \log \frac{\sum_{i=1}^n Y_i \mathbb{I}\{A_i = 0\}}{\sum_{i=1}^n \mathbb{I}\{A_i = 0\}}$$

based on n iid copies $O_i = (A_i, Y_i)$ of O is asymptotically efficient. It is furthermore natural to refer to $I_n = n/s_n^2$, the inverse of the estimated variance of Ψ_n at time n , as the *statistical information* available at that time. Under ψ , the central limit theorem applies and teaches us that $\sqrt{I_n}(\Psi_n - \psi)$ converges in distribution, as n grows to infinity, to the standard normal distribution.

Deciding to reject the null if $\sqrt{I_n}\Psi_n \geq \xi_{1-\alpha}$ yields a test with asymptotic type I error α . In order to ensure that its asymptotic power at alternative ψ is $(1 - \beta)$, it is sufficient that

$$n = \inf \left\{ t \geq 1 : I_t \geq \left(\frac{\xi_{1-\alpha} - \xi_\beta}{\psi} \right)^2 = I_{\max} \right\}, \quad (6)$$

$I_{\max}$ being the so-called *maximum committed information*.

For n large enough, $I_n \simeq n/v^b(\theta)$ if we use the balanced treatment mechanism, while I_n would have been approximately equal to $n/v^*(\theta)$, had we used the optimal treatment mechanism. Substituting bluntly $n/v^b(\theta)$ or $n/v^*(\theta)$ to I_n in (6), we see that the ratio of the testing times n^b (using the balanced treatment mechanism) and n^* (using the optimal one) satisfies

$$\frac{n^*}{n^b} \simeq \frac{v^*(\theta)}{v^b(\theta)} = R(\theta),$$

the relative efficiency criterion. In other words, were we given the optimal treatment mechanism in advance, we would in average need to sample $R(\theta) \in (\frac{1}{2}, 1)$ times the number of observations required when using the balanced treatment mechanism. In the previous example where $\theta = (\frac{1}{100}, \frac{5}{100})$, setting $\alpha = 0.05$, $\beta = 0.1$ and the alternative parameter $\psi = \Psi(\theta) \simeq 1.609 > 0$, the maximal committed information is $I_{\max} \simeq 3.306$, $n^* \simeq 676.901$ and $n^b \simeq 780.248$.

In summary, resorting to the balanced treatment mechanism may be a very poor (inefficient) choice. Since the optimal treatment mechanism g^* can be learned from the data, why not use it? Once again, we emphasize that there is nothing special about targeting this specific treatment mechanism. One could have also chosen to target that treatment mechanism which minimizes the expected number of failures subject to power constraint (Rosenberger et al., 2001).

Of course, targeting the optimal treatment mechanism on the fly implies losing independence between successive observations, making the study of the design more involved. However, we present and study here such a methodology. It is built on the seminal technical report (van der Laan, 2008).

3 Targeting the optimal design

In this section, we describe how we carry out the adaptation of the sampling scheme. For simplicity's sake, the randomization probability is updated at each step (*i.e.* each time a new observation is sampled). However, it is important to understand that all our results (and their proofs) still hold when adaptation only occurs once c new observations are accrued, where c is a pre-determined integer. In this view, we use $c = 1$ in the whole article as well as in the companion article (Chambaz and van der Laan, 2010b)—although our R code can handle the general case.

3.1 Adaptive coarsening at random assumption.

We denote by A_i , $X_i = (X_i(0), X_i(1))$, $Y_i = X_i(A_i)$ and $O_i = (A_i, Y_i)$ the treatment assignment, full data structure, outcome, and observation for experimental unit i . Whereas $X_1, \dots, X_n$ are assumed iid, the random variables $A_1, \dots, A_n$ are not independent anymore since we want to adapt the treatment mechanism based on past observations. Defining

$$\mathbf{A}_n = (A_1, \dots, A_n), \quad \mathbf{X}_n = (X_1, \dots, X_n), \quad \mathbf{O}_n = (O_1, \dots, O_n),$$

and for every $i = 0, \dots, n$

$$\mathbf{A}_n(i) = (A_1, \dots, A_i), \quad \mathbf{X}_n(i) = (X_1, \dots, X_i), \quad \mathbf{O}_n(i) = (O_1, \dots, O_i)$$

(with convention $\mathbf{A}_n(0) = \mathbf{X}_n(0) = \mathbf{O}_n(0) = \emptyset$), let $\mathbf{g}_n(\cdot|\mathbf{X}_n)$ denote the conditional distribution of the design settings $\mathbf{A}_n$ given the full data $\mathbf{X}_n$: by the chain rule,

$$\mathbf{g}_n(\mathbf{A}_n|\mathbf{X}_n) = \prod_{i=1}^n P(A_i|\mathbf{A}_n(i-1), \mathbf{X}_n), \quad (7)$$

hence the additional notation

$$g_i(a_i|\mathbf{a}(i-1), \mathbf{x}) = P(A_i = a_i|\mathbf{A}_n(i-1) = \mathbf{a}(i-1), \mathbf{X}_n = \mathbf{x})$$

for all $1 \leq i \leq n$, $\mathbf{a} = (a_1, \dots, a_n) \in \mathcal{A}^n$, $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$.

In this new setting, we state the following adaptive counterpart of the CAR assumption (1): for all $1 \leq i \leq n$, $\mathbf{a} \in \mathcal{A}^n$, $\mathbf{x} \in \mathcal{X}^n$, letting $o_i = (a_i, x_i(a_i))$ be the corresponding realization of observation O_i ,

$$\begin{aligned} g_i(a_i|\mathbf{a}(i-1), \mathbf{x}) &= P(A_i = a_i|X_i = x_i, \mathbf{O}_n(i-1) = \mathbf{o}(i-1)) \\ &= P(A_i = a_i|X_i(a_i) = x_i(a_i), \mathbf{O}_n(i-1) = \mathbf{o}(i-1)). \end{aligned}$$

With obvious convention, the new *adaptive CAR* assumption also writes as

$$g_i(a_i|\mathbf{a}(i-1), \mathbf{x}) = g_i(a_i|x_i, \mathbf{o}(i-1)) = g_i(a_i|x_i(a_i), \mathbf{o}(i-1)) \quad (8)$$

for all $1 \leq i \leq n$, $\mathbf{a} \in \mathcal{A}^n$, $\mathbf{x} \in \mathcal{X}^n$; it states that for each i A_i is conditionally independent of the full data $\mathbf{X}_n$ given the observed data $\mathbf{O}_n(i-1)$ for the first $(i-1)$ experimental units and the full data X_i for the i th experimental unit, and in addition that the conditional probability of $A_i = a_i$ given X_i and $\mathbf{O}_n(i-1)$ actually only depends on the observed part $X_i(a_i)$ and $\mathbf{O}_n(i-1)$. In particular, (7) reduces to

$$\mathbf{g}_n(\mathbf{A}_n|\mathbf{X}_n) = \prod_{i=1}^n g_i(A_i|X_i(A_i), \mathbf{O}_n(i-1)), \quad (9)$$

which justifies the notation $\mathbf{g}_n = (g_1, \dots, g_n)$. Note that in the framework of this article, a treatment mechanism $\mathbf{g}_n$ complies with the adaptive CAR assumption (9) if and only if, for all $1 \leq i \leq n$, $\mathbf{a} \in \mathcal{A}^n$, $\mathbf{x} \in \mathcal{X}^n$,

$$g_i(a_i|\mathbf{a}(i-1), \mathbf{x}) = g_i(a_i|\mathbf{o}(i-1)).$$

Note finally that we find useful to consider g_i satisfying (8) as random element (through $\mathbf{O}_n(i-1)$) of the set $\mathcal{G}$ of CAR designs.

3.2 Data generating mechanism for adaptive design and likelihood.

Given the available data $\mathbf{O}_n(i-1) = (O_1, \dots, O_{i-1})$ at step i , one first draws X_i from P_X independently of $\mathbf{X}_n(i-1)$, then one calculates the conditional distribution $g_i(\cdot | X_i, \mathbf{O}_n(i-1))$ and one samples A_i given X_i from it, the next observation finally being $O_i = (A_i, X_i(A_i))$. Regarding the likelihood of $\mathbf{O}_n$, if

$$\mathcal{X}(\mathbf{O}_n) = \{\mathbf{x} \in \mathcal{X}^n : x_i(A_i) = Y_i, i \leq n\} = \otimes_{i=1}^n \mathcal{X}(O_i)$$

(the set of those realizations $\mathbf{x}$ of $\mathbf{X}_n$ compatible with $\mathbf{O}_n$), then

$$\begin{aligned} L(\mathbf{O}_n) &= \sum_{\mathbf{x} \in \mathcal{X}(\mathbf{O}_n)} P(\mathbf{O}_n, \mathbf{X}_n = \mathbf{x}) = \sum_{\mathbf{x} \in \mathcal{X}(\mathbf{O}_n)} \mathbf{g}_n(\mathbf{A}_n | \mathbf{x}) P(\mathbf{X}_n = \mathbf{x}) \\ &= \prod_{i=1}^n g_i(A_i | Y_i, \mathbf{O}_n(i-1)) P(\mathbf{X}_n \in \mathcal{X}(\mathbf{O}_n)) \\ &= \prod_{i=1}^n g_i(A_i | Y_i, \mathbf{O}_n(i-1)) \prod_{i=1}^n \theta(O_i), \end{aligned} \quad (10)$$

the third and fourth equalities being derived from the adaptive CAR equality (9) and from independence of $X_1, \dots, X_n$ respectively. Thus, the likelihood remarkably factorizes into the product of a θ -factor and a $\mathbf{g}_n$ -factor. Thanks to the form of the likelihood, we can say that $\mathbf{O}_n$ is obtained under $(\theta, \mathbf{g}_n)$. For convenience, we also write sometimes that $\mathbf{O}_n$ is *obtained under $\mathbf{g}_n$ -adaptive sampling scheme* without specifying the parameter θ . Likewise, we later refer to data *obtained under iid g^b -balanced or under iid g^* -optimal sampling schemes*.

3.3 Strategy.

Let $\theta_i = (\theta_{i,0}, \theta_{i,1})$ denote for each $i \leq n$ the MLE of $\theta = (\theta_0, \theta_1) \in]0, 1[^2$ based on $\mathbf{O}_n(i)$ (with convention $\theta_{i,a} = \frac{1}{2}$ as long as no relevant observation is available). Thanks to the form of the log-likelihood exhibited in (10) and as soon as $\sum_{j=1}^i \mathbb{1}\{A_j = a\} > 1$, $\theta_{i,a}$ is the empirical mean

$$\theta_{i,a} = \frac{\sum_{j=1}^i Y_j \mathbb{1}\{A_j = a\}}{\sum_{j=1}^i \mathbb{1}\{A_j = a\}},$$

as if we used a deterministic treatment mechanism (and observations were iid).

These empirical means yield plug-in estimates $\sigma_i^2(a) = \theta_{i,a}(1 - \theta_{i,a})$ of $\sigma^2(\theta)(a)$, as well as plug-in estimates of the optimal treatment mechanism $g^*(\theta)$ introduced in (4), characterized by $g_1^s(1|\mathbf{O}_n(0)) = \frac{1}{2}$ and for $i \geq 1$,

$$g_{i+1}^s(1|\mathbf{O}_n(i)) = \frac{\sqrt{\theta_{i,0}(1 - \theta_{i,1})}}{\sqrt{\theta_{i,0}(1 - \theta_{i,1})} + \sqrt{\theta_{i,1}(1 - \theta_{i,0})}} \quad (11)$$

(sometimes abbreviated to $g_{i+1}^s(1)$), hence a first adaptive CAR treatment mechanism $\mathbf{g}_n^s = (g_1^s, \dots, g_n^s)$. Another interesting choice is also considered here, which we characterize iteratively by $g_1(1|\mathbf{O}_n(0)) = \frac{1}{2}$ and for $i \geq 1$,

$$g_{i+1}(1|\mathbf{O}_n(i)) = \arg \min_{\gamma \in (0,1)} \left\{ \frac{1}{i+1} \sum_{j=1}^i g_j(1|\mathbf{O}_n(j-1)) + \gamma \right\} - g_i^s(1) \Bigg\}^2. \quad (12)$$

This alternative choice aims at obtaining a balance between the two treatments which, at experiment i , closely approximates $g^*(\theta)$, in the sense that $\frac{1}{i} \sum_{j=1}^i \mathbb{I}\{A_j = 1\} \simeq g_{i-1}^s(1)$, the current best guess. This second definition is more *aggressive* in the pursuit of the optimal treatment mechanism, as it tries to compensate on the fly for early sub-optimal sampling.

A technical condition was actually left aside in the definition of $\mathbf{g}_n^s$. Because we want to exclude the possibility that the adaptive design stops a treatment arm with probability tending to 1, we impose that $g_{i+1}(1|\mathbf{O}_n(i)) \in [\delta; 1 - \delta]$ for a small $\delta > 0$ (such that $\delta < \min_{a \in \mathcal{A}} g^*(\theta)(a)$ and $1 - \delta > \max_{a \in \mathcal{A}} g^*(\theta)(a)$) by letting $g_1^s(1|\mathbf{O}_n(0)) = \frac{1}{2}$ and for $i \geq 1$,

$$g_{i+1}^*(1|\mathbf{O}_n(i)) = \min \{1 - \delta, \max \{\delta, g_{i+1}^s(1|\mathbf{O}_n(i))\}\} \quad (13)$$

(sometimes abbreviated to $g_{i+1}^*(1)$), thus characterizing the adaptive CAR treatment mechanism $\mathbf{g}_n^* = (g_1^*, \dots, g_n^*)$. Similarly, we substitute $g_i^*(1)$ to $g_i^s(1)$ and allow γ to vary in $[\delta, 1 - \delta]$ only in (12), yielding another adaptive CAR treatment mechanism $\mathbf{g}_n^a$ (where the superscript a stands for *aggressive*). The same kind of δ -thresholding, which makes perfect sense from an applied point of view too, was already suggested in (Tymofyeyev et al., 2007).

It is worth noting that Hu, Zhang, and He (2009) address a very interesting problem that we do not consider here: they show how to target the user-supplied treatment mechanism in such a way that the Cramér-Rao lower bounds *on the allocation variances* are achieved, thanks to a clever generalization of the celebrated Efron's biased coin design (Efron, 1971). Of course, resorting to the same kind of strategy would be possible here, but is outside the scope of this article.

In the rest of this article, we investigate theoretically the properties of the data-adaptive designs based on $\mathbf{g}_n^*$ only. The theoretical study is backed by the

simulation carried out in the companion article (Chambaz and van der Laan, 2010b), where $\mathbf{g}_n^a$ is also considered.

4 Asymptotic study

We address in this section the asymptotic statistical study of the method presented in the previous section. We state strong consistency results then a central limit theorem, from which we derive a rule for constructing confidence intervals.

The following consistency result holds, which teaches us that the method does learn what is the optimal design. Its proof mainly relies on Theorem 6, and we refer to (Chambaz and van der Laan, 2010a) for the details.

Theorem 1. *Let θ_n be the MLE of $\theta \in]0, 1[^2$ based on $\mathbf{O}_n$ sampled from $(\theta, \mathbf{g}_n^*)$, for the adaptive CAR treatment mechanism $\mathbf{g}_n^*$ characterized by (13). Then θ_n converges almost surely to θ . Consequently, Ψ_n is a strongly consistent estimate of $\Psi(\theta)$ and $\mathbf{g}_n^*$ converges to the optimal design $g^*(\theta)$ in the sense that $g_n^*(1)$ and $\frac{1}{n} \sum_{i=1}^n g_i^*(1)$ both converge almost surely to $g^*(\theta)(1)$.*

In order to provide statistical inference, we need now to establish that the vector $\sqrt{n}(\theta_n - \theta)$ converges in distribution so that one can construct confidence intervals for θ and come up with valid testing procedures. The following central limit theorem actually holds. Its proof mainly involves the multivariate central limit theorem stated in Theorem 8 and resorting to the delta-method, see (Chambaz and van der Laan, 2010a) for the details.

Theorem 2. *Let θ_n be the MLE of $\theta \in]0, 1[^2$ based on $\mathbf{O}_n$ sampled from $(\theta, \mathbf{g}_n^*)$, for the adaptive CAR treatment mechanism $\mathbf{g}_n^*$ defined in (13). Let $D_0^*(\theta, g^*(\theta))$, $D_1^*(\theta, g^*(\theta))$ and $\text{IC}(\theta, g^*(\theta))$ be as defined in (2) and (3) with $g = g^*(\theta)$. Let $D^*(\theta, g^*(\theta)) = (D_0^*(\theta, g^*(\theta)), D_1^*(\theta, g^*(\theta)))$. Then under $(\theta, \mathbf{g}_n^*)$*

$$\sqrt{n}(\theta_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n D^*(\theta, g^*(\theta))(O_i)(1 + o_P(1)). \quad (14)$$

Furthermore, $\frac{1}{\sqrt{n}} \sum_{i=1}^n D^*(\theta, g^*(\theta))(O_i)$ is a normalized discrete martingale which converges under $(\theta, \mathbf{g}_n^*)$ to a centered Gaussian distribution with covariance matrix

$$\begin{aligned} \Sigma^* &= P_{\theta, g^*(\theta)} D^*(\theta, g^*(\theta))^\top D^*(\theta, g^*(\theta)) \\ &= \text{diag} \left(\frac{\theta_0(1 - \theta_0)}{g^*(\theta)(0)}, \frac{\theta_1(1 - \theta_1)}{g^*(\theta)(1)} \right). \end{aligned} \quad (15)$$

The latter is consistently estimated with its empirical counterpart

$$\frac{1}{n} \sum_{i=1}^n \text{diag} \left((Y_i - \theta_{n,0})^2 \frac{\mathbb{I}\{A_i = 0\}}{g_n^*(0)^2}, (Y_i - \theta_{n,1})^2 \frac{\mathbb{I}\{A_i = 1\}}{g_n^*(1)^2} \right) \quad (16)$$

as if the sampling was iid.

Thus under $(\theta, \mathbf{g}_n^*)$, we also have

$$\sqrt{n}(\Psi_n - \Psi(\theta)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{IC}(\theta, g^*(\theta))(O_i) + o_P(1), \quad (17)$$

and convergence in distribution of $\sqrt{n}(\Psi_n - \Psi(\theta))$ to a centered Gaussian distribution with variance $v^*(\theta)$, the optimal efficient asymptotic variance. The latter is finally consistently estimated with either $v^*(\theta_n)$ or

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \text{IC}(\theta_n, g_n^*)(O_i)^2 \\ &= \frac{1}{n} \sum_{i=1}^n (Y_i - \theta_{n,0})^2 \frac{\mathbb{I}\{A_i = 0\}}{\theta_{n,0}^2 g_n^*(0)^2} + (Y_i - \theta_{n,1})^2 \frac{\mathbb{I}\{A_i = 1\}}{\theta_{n,1}^2 g_n^*(1)^2} \end{aligned}$$

as if sampling was iid.

We wish to construct a confidence interval for $\Psi(\theta)$ based on $\mathbf{O}_n$ sampled under $(\theta, \mathbf{g}_n^*)$. Let us denote by s_n^2 either consistent estimates of $v^*(\theta)$ based on $\mathbf{O}_n$ as introduced in Theorem 2, and let $\xi_{1-\alpha/2}$ be the $(1 - \alpha/2)$ -quantile of the standard normal distribution. By the latter theorem,

Highlight 1 (pointwise estimation and confidence interval). *In view of Theorems 1 and 2, the estimator Ψ_n of $\Psi(\theta)$ obtained under $(\theta, \mathbf{g}_n^*)$ -adaptive sampling scheme is strongly consistent, the estimated probability of being treated $g_n^*(1)$ also converging almost surely to the optimal probability of being treated $g^*(\theta)(1)$. In addition, the confidence interval*

$$\left[\Psi_n \pm \frac{s_n}{\sqrt{n}} \xi_{1-\alpha/2} \right]$$

has asymptotic coverage $(1 - \alpha)$.

This theoretical result is validated with simulations in Section 3 of the companion article (Chambaz and van der Laan, 2010b), focusing on the empirical distribution of the MLE in Section 3.2, on the empirical coverage guaranteed by the confidence intervals in Section 3.3 and on their empirical widths in Section 3.4. Moreover, an illustrating example is developed in Section 3.5 of (Chambaz and van der Laan, 2010b).

5 Targeted optimal design group sequential testing

Obviously, the sequence of estimators $(\Psi_n)_{n \geq 1}$ can be used to carry out the test of the null “ $\Psi(\theta) = \psi_0$ ” against its unilateral alternative “ $\Psi(\theta) > \psi_0$ ” for some $\psi_0 \in \mathbb{R}$. We build in this section a *group sequential testing procedure*, that is a testing procedure which repeatedly tries to make a decision at intervals rather than once all data are collected, or than after every new observation is obtained (such a testing procedure would be said fully sequential). We refer to (Jennison and Turnbull, 2000, Proschan et al., 2006) for a general presentation of group sequential testing procedures.

5.1 The targeted optimal design group sequential testing procedure.

Formal description of the targeted optimal design group sequential testing procedure.

We wish to test the null “ $\Psi(\theta) = \psi_0$ ” against “ $\Psi(\theta) > \psi_0$ ” with asymptotic type I error α and asymptotic type II error β at some $\psi_1 > \psi_0$. We intend to proceed group sequentially with $K \geq 2$ steps, and we wish to rely on a multidimensional t -statistic of the form

$$(\tilde{T}_1, \dots, \tilde{T}_K) = \left(\frac{\sqrt{N_k}(\Psi_{N_k} - \psi_0)}{s_{N_k}} \right)_{k \leq K}, \quad (18)$$

where each N_k is a carefully chosen (random) sample size and where s_n^2 estimates the asymptotic variance of $\sqrt{n}(\Psi_n - \Psi(\theta))$ under $(\theta, \mathbf{g}_n^*)$ sampling (see Theorem 2).

To this end, let $0 < p_1 < \dots < p_K = 1$ be increasingly ordered proportions. Consider the α -spending and β -spending strategies $(\alpha_1, \dots, \alpha_K)$ and $(\beta_1, \dots, \beta_K)$, i.e. K -tuples of positive numbers such that $\sum_{k=1}^K \alpha_k = \alpha$ and $\sum_{k=1}^K \beta_k = \beta$. One could for instance choose α -spending and β -spending functions f_α, f_β , that are increasing functions from $[0, 1]$ to $[0, 1]$ such that $f_\alpha(0) = f_\beta(0) = 0$ and $f_\alpha(1) = f_\beta(1) = 1$, and set $\sum_{l=1}^k \alpha_l = f_\alpha(p_k)\alpha$, $\sum_{l=1}^k \beta_l = f_\beta(p_k)\beta$ for all $k \leq K$.

Now, let $(Z_1, \dots, Z_K)$ follow the centered Gaussian distribution with covariance matrix $\mathcal{C} = (\sqrt{p_{k \wedge l}/p_{k \vee l}})_{k, l \leq K}$ and let us assume that there exists a unique value $I > 0$, the so-called *maximum committed information* from now on denoted by $I_{\max}$, such that there exist a rejection boundary $(a_1, \dots, a_K)$ and a futility boundary $(b_1, \dots, b_K)$ satisfying $a_K = b_K$, $P(Z_1 \geq a_1) = \alpha_1$, $P(Z_1 + (\psi_1 - \psi_0)\sqrt{p_1 I} \leq b_1) = \beta_1$, and for every $1 \leq k < K$,

$$P(\forall j \leq k, b_j < Z_j < a_j \text{ and } Z_{k+1} \geq a_{k+1}) = \alpha_{k+1},$$

$$P(\forall j \leq k, b_j < Z_j + (\psi_1 - \psi_0)\sqrt{p_j I} < a_j$$

$$\text{and } Z_{k+1} + (\psi_1 - \psi_0)\sqrt{p_{k+1} I} \leq b_{k+1}) = \beta_{k+1}.$$

Note that the closer ψ_1 is to ψ_0 , the larger is $I_{\max}$ (actually, $\psi_1 \mapsto (\psi_1 - \psi_0)\sqrt{I_{\max}}$ is both upper bounded and bounded away from zero). Heuristically, the closer ψ_1 is to ψ_0 , the more difficult it is to decide between the null and its alternative while preserving the required type II error at ψ_1 , the more information is needed to proceed. In this setting, it is natural to refer to the inverse of the variance of Ψ_n as an amount of statistical information collected so far. The latter information writes as n/s_n^2 , notably making clear that the product $s_n^2 I_{\max}$ and number of observations n are on the same scale. In this spirit, let us finally define for each $k \leq K$

$$N_k = \inf \left\{ n \geq 1 : \frac{n}{s_n^2} \geq p_k I_{\max} \right\}.$$

Under $(\theta, \mathbf{g}_n^*)$, if $v^*(\theta)I_{\max}$ is large, then N_k tend to be large too.

The targeted optimal design group sequential testing rule finally writes as follows:

starting from $k = 1$,

if $\tilde{T}_k \geq a_k$ then reject the null and stop accruing data,
 if $\tilde{T}_k \leq b_k$ then fail rejecting the null and stop accruing data,
 if $b_k < \tilde{T}_k < a_k$ then set $k \leftarrow k + 1$ and repeat.

If $(\tilde{T}_1, \dots, \tilde{T}_K)$ had the same distribution as $(Z_1, \dots, Z_K)$, then the latter rule would yield a testing procedure with the required type I error and type II error at the specified alternative parameter.

The merit is clear of targeting the fixed design $g^*(\theta)$ that makes the estimate Ψ_n have the optimal (*i.e.* smallest) asymptotic variance, the random variables N_k (*i.e.* successive number of observations required for testing) then being, at least informally, stochastically smaller than they would have been had another fixed design been used (or targeted).

Carrying out the targeted optimal design group sequential testing procedure.

We wish to test the null “ $\Psi(\theta) = \psi_0$ ” against “ $\Psi(\theta) > \psi_0$ ” with asymptotic type I error α and asymptotic type II error β at some $\psi_1 > \psi_0$. We intend to proceed group sequentially with $K \geq 2$ steps.

Highlight 2 (targeted optimal design adaptive group sequential testing). *To do so:*

1. Choose increasingly ordered proportions $0 < p_1 < \dots < p_K = 1$.
2. Compute numerically the maximum committed information $I_{\max}$, rejection and futility boundaries $(a_1, \dots, a_K)$ and $(b_1, \dots, b_K)$.
3. Starting from $k = 1$,
 - (a) keep sampling under $(\theta, \mathbf{g}_n^*)$; as soon as N_k data are collected,
 - (b) compute $\tilde{T}_k$,
 - (c) apply the following rule:
 - if $\tilde{T}_k \geq a_k$ then reject the null and stop accruing data,
 - if $\tilde{T}_k \leq b_k$ then fail rejecting the null and stop accruing data,
 - if $b_k < \tilde{T}_k < a_k$ then set $k \leftarrow k + 1$ and repeat.

We investigate the behavior of this group sequential testing procedure from a theoretical perspective in Section 5.2. The procedure is studied by simulations in Section 4 of the companion article (Chambaz and van der Laan, 2010b), focusing on empirical type I and type II errors in Section 4.2 and on empirical sample sizes at decision in Section 4.3.

5.2 Asymptotic study of the targeted optimal design group sequential testing procedure powered at local alternatives.

In order to tackle the asymptotic study of the targeted optimal design group sequential testing procedure, we resort to contiguity arguments. According to (van der Vaart, 1998, Chapter 6), “contiguity arguments are a technique to obtain the limit distribution of a sequence of statistics under underlying laws Q_n from a limiting distribution under laws P_n .” Here the laws P_n describe a null distribution under investigation (the distribution of the test statistic under $(\theta, \mathbf{g}_n^*)$), and the laws Q_n correspond to an alternative hypothesis.

Proving the validity of the targeted optimal design group sequential testing procedure (as defined in Section 5.1) powered at local alternatives is outside the scope of this article. We rather consider a slightly simpler version where the random N_k are replaced by deterministic n_k . We conjecture that the theorem we prove in the simpler deterministic sample sizes context can be extended to the random sample sizes context. The simulation study that we undertake in Section 4 of the companion article (Chambaz and van der Laan, 2010b) confirms the conjecture.

Once again, let $0 < p_1 < \dots < p_K = 1$ be increasingly ordered proportions for some integer $K \geq 2$, and define $n_k = \lceil np_k \rceil$ (the smallest integer not smaller than

$np_k)$ for each $k \leq K$. We wish to test the null against its alternative based on the multidimensional t -statistic

$$(T_1, \dots, T_K) = \left(\frac{\sqrt{n_k}(\Psi_{n_k} - \psi_0)}{s_{n_k}} \right)_{k \leq K} \quad (19)$$

(s_n^2 estimates the asymptotic variance of $\sqrt{n}(\Psi_n - \Psi(\theta))$ under $(\theta, \mathbf{g}_n^*)$ —see Theorem 2). Before going any further, we state a crucial theorem which describes how the test statistic converges towards the so-called *canonical distribution*. A similar result is obtained by Zhu and Hu (2010) through a different approach based on the study of the limit distribution of a stochastic process defined over $(0, 1]$.

Theorem 3. Consider $h = (h_0, h_1) \in \mathbb{R}^2$ satisfying both $h_1 > h_0$ and $\gamma h_1 + \gamma^{-1} h_0 \neq 0$ where $\gamma = \text{OR}(\theta)g^*(\theta)(1)/g^*(\theta)(0)$. Define

$$\theta_{h/\sqrt{n}} = (\theta_0(1 + h_0/\sqrt{n}), \theta_1(1 + h_1/\sqrt{n}))$$

for all $n \geq n_0$ large enough to ensure $\theta_{h/\sqrt{n}} \in]0, 1]^2$. The sequence $(\theta_{h/\sqrt{n}})_{n \geq n_0}$ defines a sequence $(\psi_n)_{n \geq n_0}$ of contiguous parameters (“from direction h ”), with $\psi_n = \Psi(\theta_{h/\sqrt{n}}) > \Psi(\theta)$.

Introduce the mean vector $\mu(h) = (h_1 - h_0)(\sqrt{p_1}, \dots, \sqrt{p_K})/\sqrt{v^*(\theta)}$ and the covariance matrix $\mathcal{C} = (\sqrt{p_{k \wedge l}/p_{k \vee l}})_{k, l \leq K}$. Then:

- (i) under $(\theta, \mathbf{g}_n^*)$, $(T_1, \dots, T_K)$ converges in distribution, as n tends to infinity, to the centered Gaussian distribution with covariance matrix $\mathcal{C}$;
- (ii) under $(\theta_{h/\sqrt{n}}, \mathbf{g}_n^*)$, $(T_1, \dots, T_K)$ converges in distribution, as n tends to infinity, to the Gaussian distribution with mean $\mu(h)$ and covariance matrix $\mathcal{C}$.

Say we want to perform a test such with asymptotic type I error α and asymptotic power $(1 - \beta)$ at the limit of the sequence of contiguous parameters $(\psi_n)_{n \geq n_0}$, i.e. such that (a) the probability of rejecting the null for its alternative under $(\theta, \mathbf{g}_n^*)$, and (b) the probability of failing to reject the null for its alternative under $(\theta_{h/\sqrt{n}}, \mathbf{g}_n^*)$ converge (as n tends to infinity) towards α and β , respectively.

Let us consider the α -spending and β -spending strategies $(\alpha_1, \dots, \alpha_K)$ and $(\beta_1, \dots, \beta_K)$. It is usually assumed that the next lemma holds:

Lemma 4. In the framework of Theorem 3, let $(Z_1, \dots, Z_K)$ follow the centered Gaussian distribution with covariance matrix $\mathcal{C}$ (as defined in Theorem 3). Assume that $\alpha + \beta < 1$. There exists a unique $\varepsilon > 0$, a unique rejection boundary $(a_1, \dots, a_K)$, a unique futility boundary $(b_1, \dots, b_K)$ such that $a_K = b_K$, $P(Z_1 \geq a_1) = \alpha_1$, $P(Z_1 + \mu(\varepsilon h)_1 \leq b_1) = \beta_1$, and for every $1 \leq k < K$,

$$P(\forall j \leq k, b_j < Z_j < a_j \text{ and } Z_{k+1} \geq a_{k+1}) = \alpha_{k+1},$$

$$P(\forall j \leq k, b_j < Z_j + \mu(\varepsilon h)_j < a_j \text{ and } Z_{k+1} + \mu(\varepsilon h)_{k+1} \leq b_{k+1}) = \beta_{k+1}.$$

Given such rejection and futility boundaries, we proceed as follows: starting from $k = 1$,

if $T_k \geq a_k$ then reject the null and stop accruing data,
 if $T_k \leq b_k$ then fail rejecting the null and stop accruing data,
 if $b_k < T_k < a_k$ then set $k \leftarrow k + 1$ and repeat

Theorem 3 and Lemma 4 teach us that the group sequential testing procedure described above satisfies the requirements on stepwise type I and type II error control, once h is replaced with εh (which actually corresponds to a shift in n in the definition of the sequence of contiguous alternatives $(\theta_{h/\sqrt{n}})_{n \geq n_0}$).

5.3 Proof of Theorem 3.

Theorem 3 is a corollary of the following

Lemma 5. Denote by Λ_n the log-likelihood ratio of the $(\theta_{h/\sqrt{n}}, \mathbf{g}_n^*)$ experiment with respect to the $(\theta, \mathbf{g}_n^*)$ experiment, as defined in Theorem 3. There exists a constant $\tau^2 > 0$ such that, under $(\theta, \mathbf{g}_n^*)$, the vector $(\sqrt{n_1}(\Psi_{n_1} - \Psi(\theta)), \dots, \sqrt{n_K}(\Psi_{n_K} - \Psi(\theta)), \Lambda_n)$ converges in distribution, as n tends to infinity, to the Gaussian distribution with mean $(0, \dots, 0, -\frac{1}{2}\tau^2)$ and covariance matrix

$$\begin{pmatrix} \left(\sqrt{\frac{p_{k\wedge l}}{p_{k\vee l}}} v^*(\theta) \right)_{k,l \leq K} & \sqrt{v^*(\theta)} \mu(h)^\top \\ \sqrt{v^*(\theta)} \mu(h) & \tau^2 \end{pmatrix}.$$

In particular, the $(\theta_{h/\sqrt{n}}, \mathbf{g}_n^*)$ and $(\theta, \mathbf{g}_n^*)$ experiments are mutually contiguous.

It is easy to obtain the limiting distribution of $(T_1, \dots, T_K)$ under $(\theta, \mathbf{g}_n^*)$ from Lemma 5. Le Cam's third lemma solves the problem of obtaining the limiting distribution of $(T_1, \dots, T_K)$ under $(\theta_{h/\sqrt{n}}, \mathbf{g}_n^*)$ from the convergence under $(\theta, \mathbf{g}_n^*)$ exhibited in Lemma 5. The second limiting distribution is still Gaussian, has the same asymptotic covariance matrix as under $(\theta, \mathbf{g}_n^*)$, but differs by its asymptotic mean which is no longer 0.

Proof of Theorem 3. By the continuous mapping theorem, we readily obtain from Lemma 5 the convergence in distribution under $(\theta, \mathbf{g}_n^*)$ of

$$(\sqrt{n_1}(\Psi_{n_1} - \Psi(\theta)), \dots, \sqrt{n_K}(\Psi_{n_K} - \Psi(\theta)))$$

to the centered Gaussian distribution with covariance matrix

$$(v^*(\theta) \sqrt{p_{k \wedge l} / p_{k \vee l}})_{k, l \leq K}.$$

Then Slutsky's lemma straightforwardly yield the first convergence (i).

Regarding (ii), we first invoke Lemma 5 and Le Cam's third lemma (see Example 6.7 in (van der Vaart, 1998)) in order to obtain that, under $(\theta_{h/\sqrt{n}}, \mathbf{g}_n^*)$, the vector $(\sqrt{n_1}(\Psi_{n_1} - \Psi(\theta)), \dots, \sqrt{n_K}(\Psi_{n_K} - \Psi(\theta)))$ converges in distribution to the Gaussian distribution with mean $\sqrt{v^*(\theta)}\mu(h)$ and covariance matrix

$$(v^*(\theta) \sqrt{p_{k \wedge l} / p_{k \vee l}})_{k, l \leq K}.$$

In addition, the $(\theta_{h/\sqrt{n}}, \mathbf{g}_n^*)$ and $(\theta, \mathbf{g}_n^*)$ experiments are mutually contiguous, implying that if s_n^2 estimates $v^*(\theta)$ under $(\theta, \mathbf{g}_n^*)$, then it also estimates $v^*(\theta)$ under $(\theta_{h/\sqrt{n}}, \mathbf{g}_n^*)$. We apply again Slutsky's lemma in order to obtain the second convergence (ii). $\square$

Proof of Lemma 5. Let us consider first the log-likelihood ratio of the experiments $(\theta_{h/\sqrt{n}}, \mathbf{g}_n^*)$ with respect to $(\theta, \mathbf{g}_n^*)$. The shorthand notations $\theta(O) = \theta_A^Y (1 - \theta_A)^{1-Y}$ and $\theta_{h/\sqrt{n}}(O) = [\theta_A(1 + h_A/\sqrt{n})]^Y [1 - \theta_A(1 + h_A/\sqrt{n})]^{1-Y}$ and (10) readily yield

$$\begin{aligned} \Lambda_n &= \log \prod_{i=1}^n \frac{g_i^*(A_i | Y_i, \mathbf{O}_n(i-1))}{g_i^*(A_i | Y_i, \mathbf{O}_n(i-1))} \prod_{i=1}^n \frac{\theta_{h/\sqrt{n}}(O_i)}{\theta(O_i)} \\ &= \sum_{i=1}^n \mathbb{1}\{A_i = 0\} \left[Y_i \log \left(1 + \frac{h_0}{\sqrt{n}} \right) + (1 - Y_i) \log \left(1 - \frac{\theta_0 h_0}{(1 - \theta_0)\sqrt{n}} \right) \right] \\ &\quad + \mathbb{1}\{A_i = 1\} \left[Y_i \log \left(1 + \frac{h_1}{\sqrt{n}} \right) + (1 - Y_i) \log \left(1 - \frac{\theta_1 h_1}{(1 - \theta_1)\sqrt{n}} \right) \right] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n L_1(O_i) - \frac{1}{2n} \sum_{i=1}^n L_2(O_i) + o_P(1), \end{aligned}$$

with

$$\begin{aligned} L_1(O_i) &= \mathbb{1}\{A_i = 0\} \frac{Y_i - \theta_0}{1 - \theta_0} h_0 + \mathbb{1}\{A_i = 1\} \frac{Y_i - \theta_1}{1 - \theta_1} h_1, \\ L_2(O_i) &= \mathbb{1}\{A_i = 0\} \left[Y_i + (1 - Y_i) \left(\frac{\theta_0}{1 - \theta_0} \right)^2 \right] h_0^2 \end{aligned}$$

$$+ \mathbb{I}\{A_i = 1\} \left[Y_i + (1 - Y_i) \left(\frac{\theta_1}{1 - \theta_1} \right)^2 \right] h_1^2.$$

First, we observe that, since L_2 is bounded (and measurable), $\frac{1}{n} \sum_{i=1}^n L_2(O_i) = \frac{1}{n} \sum_{i=1}^n P_{\theta, g_i^*} L_2 + \frac{1}{n} \sum_{i=1}^n [L_2(O_i) - P_{\theta, g_i^*} L_2] = \frac{1}{n} \sum_{i=1}^n P_{\theta, g_i^*} L_2 + o_P(1)$ by virtue of the Kolmogorov law of large numbers. Now, $P_{\theta, g_i^*} L_2 = g_i^*(0) \frac{\theta_0 h_0^2}{1 - \theta_0} + g_i^*(1) \frac{\theta_1 h_1^2}{1 - \theta_1}$, hence

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n P_{\theta, g_i^*} L_2 &= \frac{1}{n} \sum_{i=1}^n g_i^*(0) \frac{\theta_0 h_0^2}{1 - \theta_0} + \frac{1}{n} \sum_{i=1}^n g_i^*(1) \frac{\theta_1 h_1^2}{1 - \theta_1} \\ &= g^*(\theta)(0) \frac{\theta_0 h_0^2}{1 - \theta_0} + g^*(\theta)(1) \frac{\theta_1 h_1^2}{1 - \theta_1} + o_P(1) \end{aligned}$$

by virtue of Theorem 1. In summary, we obtain that

$$\Lambda_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n L_1(O_i) - \frac{1}{2} \tau^2 + o_P(1) \quad (20)$$

for $\tau^2 = g^*(\theta)(0) \frac{\theta_0 h_0^2}{1 - \theta_0} + g^*(\theta)(1) \frac{\theta_1 h_1^2}{1 - \theta_1}$.

Second, we define $Z_i = (\mathbb{I}\{i \leq n_1\}, \dots, \mathbb{I}\{i \leq n_K\})$ and we introduce the bounded (and measurable) function f such that

$$f(O_i, Z_i) = (Z_i \text{IC}(\theta, g^*(\theta))(O_i), L_1(O_i)).$$

Let us show that $M_n = \frac{1}{n} \sum_{i=1}^n [f(O_i, Z_i) - P_{\theta, g_i^*} f] = \frac{1}{n} \sum_{i=1}^n f(O_i, Z_i)$ (this equality holds because for all $i \leq n$, one has $P_{\theta, g_i^*} \text{IC}(\theta, g^*(\theta)) = P_{\theta, g_i^*} L_1 = 0$) is asymptotically Gaussian. In view of Theorem 8, let $W_n = \sum_{i=1}^n P_{\theta, g_i^*} f^\top f$ and $\Sigma_n = \frac{1}{n} E W_n$. The entries of W_n take one of the following forms: $A_n = \sum_{i=1}^{n_k} P_{\theta, g_i^*} L_1 \text{IC}(\theta, g^*(\theta))$, or $B_n = \sum_{i=1}^{n_k \wedge n_l} P_{\theta, g_i^*} \text{IC}(\theta, g^*(\theta))^2$, or $C_n = \sum_{i=1}^n P_{\theta, g_i^*} L_1^2$. Now,

- $A_n = \frac{h_1}{g^*(1)} \sum_{i=1}^{n_k} g_i^*(1) - \frac{h_0}{g^*(0)} \sum_{i=1}^{n_k} g_i^*(0)$, so that $\frac{1}{n} E A_n = p_k(h_1 - h_0) + o(1)$ and $\frac{1}{n} A_n - \frac{1}{n} E A_n = o_P(1)$ since the almost sure convergence of the bounded sequence $\frac{1}{n} \sum_{i=1}^n g_i^*(a)$ towards $g^*(\theta)(a)$ (see Theorem 1) implies its convergence in L^1 norm to the same limit;
- $B_n = \frac{1 - \theta_1}{\theta_1 g^*(\theta)(1)^2} \sum_{i=1}^{n_k \wedge n_l} g_i^*(1) + \frac{1 - \theta_0}{\theta_0 g^*(\theta)(0)^2} \sum_{i=1}^{n_k \wedge n_l} g_i^*(0)$, from which it follows that $\frac{1}{n} E B_n = p_{k \wedge l} \left(\frac{1 - \theta_1}{\theta_1 g^*(\theta)(1)} + \frac{1 - \theta_0}{\theta_0 g^*(\theta)(0)} \right) + o(1) = p_{k \wedge l} v^*(\theta) + o(1)$, hence $\frac{1}{n} B_n - \frac{1}{n} E B_n = o_P(1)$ for the same reasons as above;

- $C_n = \frac{\theta_1 h_1^2}{1-\theta_1} \sum_{i=1}^n g_i^*(1) + \frac{\theta_0 h_0^2}{1-\theta_0} \sum_{i=1}^n g_i^*(0)$, hence $\frac{1}{n}EC_n = \tau^2 + o(1)$, and $\frac{1}{n}C_n - \frac{1}{n}EC_n = o_P(1)$ for the same reasons as above.

Those calculations notably teach us that, setting $m = (h_1 - h_0)(p_1, \dots, p_K)$ and $\Sigma_0 = (p_{k \wedge l})_{k, l \leq K}$, Σ_n converges to

$$\Sigma = \begin{pmatrix} v^*(\theta)\Sigma_0 & m^\top \\ m & \tau^2 \end{pmatrix}.$$

Is Σ a positive definite covariance matrix? Well, Σ_0 is a positive definite covariance matrix (that of the vector $(B_{p_1}, \dots, B_{p_K})$ where $(B_t)_{t \geq 0}$ is a standard Brownian motion), hence the symmetric matrix Σ is a positive definite covariance matrix if and only if its determinant $\det(\Sigma) > 0$. Subtracting $(h_1 - h_0)/v^*(\theta)$ times the K th row of Σ to its last row, we get that $\det(\Sigma) = v^*(\theta)^K \det(\Sigma_0) \times (\tau^2 - (h_1 - h_0)^2/v^*(\theta))$. Now, using $v^*(\theta) = \frac{1-\theta_0}{\theta_0 g^*(\theta)(0)} + \frac{1-\theta_1}{\theta_1 g^*(\theta)(1)}$ and $\gamma h_1 + \gamma^{-1} h_0 \neq 0$ (required in Theorem 3) yields

$$\begin{aligned} & v^*(\theta)\tau^2 - (h_1 - h_0)^2 \\ &= h_1^2 \left(\frac{\theta_1 v^*(\theta)}{1-\theta_1} g^*(\theta)(1) - 1 \right) + h_0^2 \left(\frac{\theta_0 v^*(\theta)}{1-\theta_0} g^*(\theta)(0) - 1 \right) + 2h_0 h_1 \\ &= (\gamma h_1 + \gamma^{-1} h_0)^2 > 0. \end{aligned}$$

In summary, Σ is a positive definite covariance matrix, the conditions of Theorem 8 are met, and therefore $\sqrt{n}M_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n f(O_i, Z_i)$ converges in distribution to the centered Gaussian distribution with covariance matrix Σ .

Let us define the diagonal matrices $\Delta_n = \text{diag}(\sqrt{n/n_1}, \dots, \sqrt{n/n_K}, 1)$ and $\Delta = \text{diag}(1/\sqrt{p_1}, \dots, 1/\sqrt{p_K}, 1)$; obviously, $\Delta_n = \Delta + o(1)$, $\sqrt{n}M_n \Delta_n = \sqrt{n}M_n \Delta + o_P(1)$. Invoking (17) in Theorem 2 and (20), it holds that, under $(\theta, \mathbf{g}_n^*)$,

$$\begin{aligned} & (\sqrt{n_1}(\Psi_{n_1} - \Psi(\theta)), \dots, \sqrt{n_K}(\Psi_{n_K} - \Psi(\theta)), \Lambda_n) \\ &= (0, \dots, 0, -\frac{1}{2}\tau^2) + \sqrt{n}M_n \Delta_n + o_P(1) \\ &= (0, \dots, 0, -\frac{1}{2}\tau^2) + \sqrt{n}M_n \Delta + o_P(1). \end{aligned}$$

This entails the convergence in distribution of $(\sqrt{n_1}(\Psi_{n_1} - \Psi(\theta)), \dots, \sqrt{n_K}(\Psi_{n_K} - \Psi(\theta)), \Lambda_n)$ under $(\theta, \mathbf{g}_n^*)$ to the Gaussian distribution with mean $(0, \dots, 0, -\frac{1}{2}\tau^2)$ and covariance matrix $\Delta \Sigma \Delta$. Simple calculations finally reveal that $\Delta \Sigma \Delta$ equals the positive definite covariance matrix given in the lemma. $\square$

6 Discussion

We have studied in this article the theoretical properties of a new adaptive group sequential design methodology for randomized clinical trials with binary treatment, binary outcome and no covariate (the experimental unit writes as $O = (A, Y) \in \{0, 1\}^2$, A being the assigned treatment and Y the corresponding outcome). In the companion article (Chambaz and van der Laan, 2010b), the study is carried out by simulations.

Prior to accruing data, the trial protocol must specify Ψ , the parameter of interest. Regarding the estimation of Ψ , the trial protocol must specify the confidence level to be used in constructing the confidence interval. Regarding the testing of Ψ , the trial protocol must specify the null and alternative hypotheses, the wished type I error, the alternative parameter at which the test is to be powered and the related wished type II error. If the investigator wants to resort to a group sequential testing procedure, then the trial protocol must also specify the number of intermediate tests, the related proportions, the α - and β -spending strategies (then the maximum committed information, rejection and futility boundaries are fully determined). Finally, the trial protocol must specify the (fixed) targeted design. We decided to focus in this article on the log-relative risk $\Psi = \log E(Y|A = 1) - \log E(Y|A = 0)$ and on that design g^* which minimizes the asymptotic variance of the MLE of Ψ . Other choices can be treated likewise.

The methodology is adaptive in the sense that the estimator of g^* , which appears to be strongly consistent (see Highlight 1), is alternatively used in the process of accruing new data, then updated and so on. The resulting MLE of Ψ , Ψ_n , is strongly consistent (see Highlight 1). It satisfies a central limit theorem, and performs as well (in terms of asymptotic variance) as its counterpart under iid sampling using g^* itself (see Highlight 1). Therefore, one easily constructs confidence intervals which are as narrow as the intervals one would get, has one known in advance g^* and used it to sample independently data (see Highlight 1). Those theoretical results are validated with simulations in the companion article (Chambaz and van der Laan, 2010b).

Furthermore, we explain how a group sequential testing procedure can be equally well applied on top of the adaptive sampling methodology (see Highlight 2). An accompanying theoretical result validates the adaptive group sequential testing procedure in the context of contiguous null and alternative hypotheses. It is supported by the simulations undertaken in the companion article (Chambaz and van der Laan, 2010b).

As stated in the abstract, a three-sentence take-home message is “Adaptive designs do learn the targeted optimal design and inference and testing can be carried out under adaptive sampling as they would under the targeted optimal randomiza-

tion probability iid sampling. In particular, adaptive designs achieve the same efficiency as the fixed oracle design. This is confirmed by a simulation study presented in the companion article (Chambaz and van der Laan, 2010b), at least for moderate or large sample sizes, across a large collection of targeted randomization probabilities.” In essence, everything works as predicted by theory. However, theory also warns us that gains cannot be dramatic in the particular setting of clinical trials with binary treatment, binary outcome and no covariate. Nonetheless, this article and its companion article (Chambaz and van der Laan, 2010b) are important: they provide a theoretical template and tools for asymptotic analysis of robust adaptive designs in less constrained settings, which we will consider in future work. This notably includes the setting of clinical trials *with covariate*, binary treatment, and *discrete or continuous outcome*, or the setting of clinical trials *with covariate*, binary treatment, and *possibly censored time-to-event* among others. Resorting to targeted maximum likelihood estimation (van der Laan and Rubin, 2006) along with adaptation of the design provides substantial gains in efficiency.

A Appendix

Normalized martingale sums of the form $M_n(f) = \frac{1}{n} \sum_{i=1}^n [f(O_i, Z_i) - P_{\theta, g_i} f]$ with $Z_i = Z_i(\mathbf{O}_n(i-1))$ play a central role in this study. The Kolmogorov strong law of large numbers (see *e.g.* (Sen and Singer, 1993, Theorem 2.4.2)) guarantees that $M_n(f)$ converges in probability to 0 almost surely for any uniformly bounded function f . However, in order to get consistency results, we need a uniform convergence result for $\sup_{f \in \mathcal{F}} |M_n(f)|$ for a certain class $\mathcal{F}$. This issue is addressed in A.1. Similarly, in order to get a central limit theorem, we need the convergence of $\sqrt{n}M_n(f)$ to a Gaussian random variable. We derive this result in Section A.2 from a standard central limit theorem for discrete martingales, see *e.g.* (Sen and Singer, 1993, Theorem 3.3.7).

A.1 Building block for consistency results.

Let $\mathbf{O}_n$ be a sequence of successive observations obtained as described in Section 3. We denote by $Z_i = Z_i(\mathbf{O}_n(i-1)) \in \mathcal{Z} \subset \mathbb{R}^d$ a summary measure of $\mathbf{O}_n(i-1)$ of fixed dimension d (for instance $Z_i = \theta_{i-1} \in \mathbb{R}^2$, the current MLE of θ at step i). Let $\mathcal{F}$ be a class of bounded (and measurable) functions of $(o, z) = (a, y, z)$ such that $\sup_{f \in \mathcal{F}} \|f\|_\infty = U < \infty$ (for instance, $f(O_i, Z_i) = D(\vartheta)(O_i) - P_{\theta, g_i^*} D(\vartheta)$ for some

$\vartheta \in [0, 1]^2$, where the dependency wrt $Z_i = \theta_{i-1}$ is conveyed through g_i^*). Defining

$$M_n(f) = \frac{1}{n} \sum_{i=1}^n [f(O_i, Z_i) - P_{\theta, g_i} f]$$

for all $f \in \mathcal{F}$ and $n \geq 1$, we note that $nM_n(f)$ is a discrete martingale sum.

Our uniform convergence result, Theorem 6, essentially relies on a maximal inequality for martingales (van Handel, 2010, Proposition A.2)—see (Chambaz and van der Laan, 2010a) for the details. Denote by $N(\mathcal{F}, \|\cdot\|_\infty, \varepsilon)$ the cardinality of the smallest finite collection $(\ell^j, u^j)_{j \leq N}$, where $\ell^j \leq u^j$ are (measurable) functions of $(o, z) = (a, y, z)$ satisfying $\|u^j - \ell^j\|_\infty \leq \varepsilon$ for all $j \leq N$, such that for every $f \in \mathcal{F}$, there exists $j \leq N$ for which $\ell^j \leq f \leq u^j$.

Theorem 6. Recall that $\sup_{f \in \mathcal{F}} \|f\|_\infty = U < \infty$. If

$$\int_0^{\sqrt{2/3}U} \sqrt{\log N(\mathcal{F}, \|\cdot\|_\infty, x)} dx < \infty$$

then for all $\alpha > 0$ there exists $c > 0$ such that, for n large enough,

$$P \left(\sup_{f \in \mathcal{F}} M_n(f) \geq \alpha \right) \leq 2e^{-nc}.$$

Consequently, $\sup_{f \in \mathcal{F}} |M_n(f)|$ converges to 0 almost surely.

A.2 Building block for central limit theorems.

First, we obtain a central limit theorem for univariate discrete martingale sums as a by-product of a classical theorem, see *e.g.* (Sen and Singer, 1993, Theorem 3.3.7). Second, we rely on it and invoke the Cramér-Wold device, see *e.g.* (Sen and Singer, 1993, Theorem 3.2.4), in order to extend the result to the case of multivariate discrete martingale sums. We refer to (Chambaz and van der Laan, 2010a) for the details of the proofs.

Univariate case.

We use the same framework and notation as those exposed at the beginning of Section A.1. For a given *real-valued* $f \in \mathcal{F}$, let us introduce

$$w_n(f)^2 = \sum_{i=1}^n P_{\theta, g_i} f^2, \quad s_n(f)^2 = \sum_{i=1}^n E f(O_i, Z_i)^2, \quad \sigma_n(f)^2 = \frac{s_n(f)^2}{n}.$$

The following univariate central limit theorem holds:

Theorem 7. Assume that for all $i = 1, \dots, n$,

$$g_i(A_i | \mathbf{A}_n(i-1), \mathbf{X}_n) = g_i(A_i | X_i(A_i), \mathbf{O}_n(i-1))$$

(by virtue of the adaptive CAR assumption (8)) only depends on $\mathbf{O}_n(i-1)$ through Z_i . If $\liminf \sigma_n(f)^2 > 0$ and $\frac{1}{n}w_n(f)^2 - \frac{1}{n}Ew_n(f)^2 = \frac{1}{n}w_n(f)^2 - \sigma_n(f)^2$ converges in probability to 0, then $w_n(f)^2/s_n(f)^2$ converges to 1 in probability and $\sqrt{n}\frac{M_n(f)}{\sigma_n(f)}$ converges in distribution to the standard normal distribution.

Furthermore, the empirical mean $\hat{\sigma}_n(f)^2 = \frac{1}{n}\sum_{i=1}^n f(O_i, Z_i)^2$ mimics $\sigma_n(f)^2$ in the sense that $\hat{\sigma}_n(f)^2 - \sigma_n(f)^2$ converges to 0 in probability. Consequently, $\sqrt{n}\frac{M_n(f)}{\hat{\sigma}_n(f)}$ also converges in distribution to the standard normal distribution.

This result notably teaches us that we can estimate the asymptotic variance of $\sqrt{n}M_n(f)$ by considering $O_1, \dots, O_n$ as independent draws from P_{θ, g_i} , treating each g_i as a given deterministic fixed design in $\mathcal{G}$. Therefore, the parametric or non-parametric bootstrap ignoring the dependence structure of $\mathbf{O}_n$ consistently estimate the limiting variance.

Multivariate case.

Let us state the multivariate version of Theorem 7. It involves the following multi-dimensional counterparts of $w_n(f)^2$, $s_n(f)^2$ and $\sigma_n(f)^2$ in the case that $f \in \mathcal{F}$ takes values in $\mathbb{R}^r$ (expectations are taken componentwise):

$$W_n(f) = \sum_{i=1}^n P_{\theta, g_i} f f^\top, S_n(f) = \sum_{i=1}^n E f(O_i, Z_i) f(O_i, Z_i)^\top, \Sigma_n(f) = \frac{S_n(f)}{n}.$$

For every positive definite symmetric matrix Σ , let $\Sigma^{-1/2}$ be the positive definite symmetric matrix such that $(\Sigma^{-1/2})^2$ is the inverse of Σ . It holds that:

Theorem 8. Assume that for all $i = 1, \dots, n$,

$$g_i(A_i | \mathbf{A}_n(i-1), \mathbf{X}_n) = g_i(A_i | X_i(A_i), \mathbf{O}_n(i-1))$$

(by virtue of the adaptive CAR assumption (8)) only depends on $\mathbf{O}_n(i-1)$ through Z_i . If $\Sigma_n(f)$ converges to a positive definite covariance matrix $\Sigma(f)$ and $\frac{1}{n}W_n(f) - \frac{1}{n}EW_n(f) = \frac{1}{n}W_n(f) - \Sigma_n(f)$ converges componentwise in probability to 0, then $\sqrt{n}M_n(f)$ converges in distribution to the centered Gaussian law over $\mathbb{R}^r$ with covariance matrix $\Sigma(f)$.

Furthermore, the empirical mean $\hat{\Sigma}_n(f) = \frac{1}{n}\sum_{i=1}^n f(O_i, Z_i) f(O_i, Z_i)^\top$ is such that $\hat{\Sigma}_n(f)$ converges to $\Sigma(f)$ in probability. Consequently, $\hat{\Sigma}_n(f)$ is invertible with probability tending to 1 and $\sqrt{n}\hat{\Sigma}_n(f)^{-1/2}M_n(f)$ converges in distribution to the standard normal distribution over $\mathbb{R}^r$.

References

- Armitage, P. (1975): *Sequential medical trials*, New-York: Wiley.
- Banerjee, A. and A. A. Tsiatis (2006): "Adaptive two-stage designs in phase II clinical trials," *Statistics in Medicine*, 25, 3382–3395.
- Berry, D. A. (2006): "Bayesian clinical trials," *Nature Reviews*, 5, drug discovery.
- Berry, D. A. and D. K. Stangl (1996): *Bayesian biostatistics*, New-York: Marcel Dekker.
- Chambaz, A. and M. J. van der Laan (2010a): "Targeting the optimal design in randomized clinical trials with binary outcomes and no covariate," *U.C. Berkeley Division of Biostatistics Working Paper Series*, paper 258.
- Chambaz, A. and M. J. van der Laan (2011): "Targeting the optimal design in randomized clinical trials with binary outcomes and no covariate: simulation study," *Int. J. Biostat.*, Volume 7, Issue 1, Article 11.
- Chernoff, H. and S. N. Roy (1965): "A Bayes sequential sampling inspection plan," *Annals of Mathematical Statistics*, 36, 1387–1407.
- Efron, B. (1971): "Forcing a sequential experiment to be balanced," *Biometrika*, 58, 403–417.
- Eisele, J. R. (1994): "The doubly adaptive biased coin design for sequential clinical trials," *Journal of Statistical Planning and Inference*, 38, 249–261.
- Emerson, S. S. (2006): "Issues in the use of adaptive clinical trial designs," *Statistics in Medicine*, 25, 3270–3296.
- Flehinger, B. J. and T. A. Louis (1971): "Sequential treatment allocation in clinical trials," *Biometrika*, 58, 419–426.
- Golub, H. L. (2006): "The need for more efficient trial designs," *Statistics in Medicine*, 25, 3231–3235.
- Hu, F. and W. F. Rosenberger (2003): "Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons," *Journal of the American Statistical Association*, 98, 671–678.
- Hu, F. and L.-X. Zhang (2004a): "Asymptotic normality of urn models for clinical trials with delayed response," *Bernoulli*, 10, 447–463.
- Hu, F. and L.-X. Zhang (2004b): "Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials," *Annals of Statistics*, 32, 268–301.
- Hu, F., L.-X. Zhang, and X. He (2009): "Efficient randomized-adaptive designs," *The Annals of Statistics*, 37, 2543–2560.
- Hu, F. H. and W. F. Rosenberger (2006): *The theory of response-adaptive randomization in clinical trials*, Wiley.
- Hung, H.-M. J. (2006): "Discussion," *Statistics in Medicine*, 25, 3313–3314.
- Ivanova, A. (2003): "A play-the-winner type urn design with reduced variability," *Metrika*, 58, 1–13.

- J., D., K. R. Abrams, and J. P. Myles (2004): *Bayesian approaches to clinical trials and health care evaluation*, Chichester: Wiley.
- Jennison, C. and B. W. Turnbull (2000): *Group Sequential Methods with Applications to Clinical Trials*, Boca Raton, FL: Chapman & Hall/CRC.
- Jennison, C. and B. W. Turnbull (2003): "Mi-course sample size modification in clinical trials based on the observed treatment effect," *Statistics in Medicine*, 22, 971–993.
- Lokhnygina, Y. and A. A. Tsiatis (2008): "Optimal two-stage group-sequential designs," *J. Statist. Plann. Inference*, 138, 489–499, URL <http://dx.doi.org/10.1016/j.jspi.2007.06.011>.
- Mehta, C. R. and N. R. Patel (2006): "Adaptive, group sequential and decision theoretic approaches to sample size determination," *Statistics in Medicine*, 25, 3250–3269.
- Proschan, M. A., G. K. K. Lan, and J. T. Wittes (2006): *Statistical Monitoring of Clinical Trials: A Unified Approach*, Statistics for biology and health, New-York: Springer.
- Rosenberger, W. F. (2002): "Randomized urn models and sequential design," *Sequential Analysis*, 21, 1–41, (with discussion).
- Rosenberger, W. F., N. Stallard, A. Ivanova, C. N. Harper, and M. L. Ricks (2001): "Optimal adaptive designs for binary response trials," *Biometrics*, 57, 909–913.
- Sen, P. K. and J. M. Singer (1993): *Large sample methods in statistics*, New York: Chapman & Hall, an introduction with applications.
- Tsiatis, A. A. and C. R. Mehta (2003): "On the inefficiency of the adaptive design for monitoring clinical trials," *Biometrika*, 90, 367–368.
- Tymofyeyev, Y., W. F. Rosenberger, and F. Hu (2007): "Implementing optimal allocation in sequential binary response experiments," *J. Amer. Statist. Assoc.*, 102, 224–234.
- van der Laan, M. J. (2008): "The construction and analysis of adaptive group sequential designs," *U.C. Berkeley Division of Biostatistics Working Paper Series*, paper 232.
- van der Laan, M. J. and D. Rubin (2006): "Targeted maximum likelihood learning," *Int. J. Biostat.*, 2, Art. 11, 40pp.
- van der Vaart, A. W. (1998): *Asymptotic statistics, Cambridge Series in Statistical and Probabilistic Mathematics*, volume 3, Cambridge: Cambridge University Press.
- van Handel, R. (2010): "On the minimal penalty for markov order estimation," *Probab. Theory Relat. Fields*, to appear.
- Wei, L. J. and S. D. Durham (1978): "The randomized play-the-winner rule in medical trials," *Journal of the American Statistical Association*, 73, 840–843.

- Zhang, L.-X., W. S. Chan, S. H. Cheung, and F. Hu (2007): “A generalized drop-the-loser urn for clinical trials with delayed responses,” *Statist. Sinica*, 17, 387–409.
- Zhang, L.-X., F. Hu, and S. H. Cheung (2006): “Asymptotic theorems of sequential estimation-adjusted urn models,” *The Annals of Applied Probability*, 16, 340–369.
- Zhu, H. and F. Hu (2010): “Sequential monitoring of response-adaptive randomized clinical trials,” *The Annals of Statistics*.