On optimal allocation in binary response trials; is adaptive design really necessary? *

David Azriel[1], Micha Mandel[1] and Yosef Rinott[1,2]

[1] Department of statistics, The Hebrew University of Jerusalem, Israel

[2] Center for the Study of Rationality, The Hebrew University of Jerusalem, Israel

March 3, 2011

Abstract

We consider the classical problem of selecting the best of two treatments in clinical trials with binary response. The target is to find the design that maximizes the power of the relevant test. Many papers use a normal approximation to the power function and claim that *Neyman allocation* that assigns subjects to treatment groups according to the ratio of the responses' standard deviations, should be used. As the standard deviations are unknown, an adaptive design is often recommended. The asymptotic justification of this approach is arguable, since it uses the normal approximation in tails where the error in the approximation is larger than the estimated quantity. We consider two different approaches for optimality of designs that are related to Pitman and Bahadur definitions of relative efficiency of tests. We prove that the optimal allocation according to the Pitman criterion is the balanced allocation and that the optimal allocation according to the Bahadur approach depends on the unknown parameters. Exact calculations reveal that the optimal allocation according to Bahadur is often close to the balanced design, and the powers of both are comparable to the Neyman allocation for small sample sizes and are generally better for large experiments. Our findings have important implications to the design of experiments, as the balanced design is proved to be optimal or close to optimal and the need for the complications involved in following an adaptive design for the purpose of increasing the power of tests is therefore questionable.

*We thank Amir Dembo for helpful comments and in particular for deriving the exact expression for ν^* in (4).

KEYWORDS: Neyman allocation, adaptive design, asymptotic power, Normal approximation, Pitman efficiency, Bahadur efficiency, large deviations.

1 Introduction

We consider the problem of optimal allocation of individuals to two treatment groups with the goal of selecting the better treatment. The problem arises frequently in phase III studies, though our original motivation came from adaptive sequential designs, such as those conducted in phase I clinical trials.

Let A and B be two treatments with unknown probabilities of success, p_A and p_B. A trial with n subjects is planned with $N_A(n)$ and $N_B(n)$ subjects assigned to treatment A and B, respectively, where $N_A(n) + N_B(n) = n$. For each subject, a binary response, success or failure, is observed. Let $\nu_n := N_A(n)/n$ be the proportion of subjects assigned to treatment A. We sometimes refer to ν_n as the *allocation*. The design problem considered here is of choosing the optimal allocation ν_n that maximizes the power of the standard test of the hypothesis $p_A = p_B$ versus one or two-sided alternatives. For given n, p_A, and p_B, the optimal allocation fraction ν_n can be found by a finite search over all possible allocations. However, much of the statistical literature suggests to study this problem for large n instead, and look for the asymptotically optimal allocation fraction ν^*. As ν^* depends on p_A and p_B, adaptive sequential designs have been proposed (e.g., Hu and Rosenberger (2006) and numerous references therein). However, as shown below, when aiming at maximizing the asymptotic power, $\nu = 1/2$ is optimal or close to optimal. Thus, equal allocation to the two treatments, which is indeed often used, is (almost) optimal, and adaptive allocation seems like an unjustified complication.

Let $Y_i(m) \sim Bin(m, p_i)$ be the number of successes if m patients are assigned to treatment i $(i = A, B)$. Let also $\hat{p}_A = \hat{p}_A(N_A(n)) = \frac{Y_A(N_A(n))}{N_A(n)}$ and $\hat{p}_B = \hat{p}_B(N_B(n)) = \frac{Y_B(N_B(n))}{N_B(n)}$ be the estimators of p_A and p_B; note that $\hat{p}_A$ and $\hat{p}_B$ depend on n and the allocation sequence ν_n, however they are suppressed for notational convenience.

The Neyman allocation rule, $\nu = \frac{\sqrt{p_A(1-p_A)}}{\sqrt{p_A(1-p_A)}+\sqrt{p_B(1-p_B)}}$, minimizes the variance of the estimator $\hat{p}_A(n) - \hat{p}_B(n)$ for the difference of probabilities (e.g., Melfi et. al (2001)). However, it is not clear that the Neyman allocation also maximizes the power of the Wald test for equality of proportions, as appears to be widely believed (e.g., Brittain and Schlesselman (1982); Rosenberger et. al (2001); Hu and Rosenberger (2003); Bandyopadhyay and Bhattacharya (2006); Hu et. al (2006); Hu and Rosenberger (2006); Tymofyeyev et. al (2007); Biswas et. al (2010); Zhu and Hu (2010);

Chambaz and van der Laan (2010)).

The standard Wald statistic for comparing p_A and p_B is

$$W := \{\hat{p}_B - \hat{p}_A\} \Big/ \sqrt{V(\hat{p}_A, \hat{p}_B, n, \nu_n)},$$

where $V(p_A, p_B, n, \nu_n) = \frac{p_A(1-p_A)}{\nu_n \cdot n} + \frac{p_B(1-p_B)}{(1-\nu_n)\cdot n}$. In the above papers, the power is often calculated by approximating the distribution of the squared Wald statistic by a non-central chi-square distribution; the Neyman allocation then maximizes the non-centrality parameter. The argument is based on the following normal approximation:

$$P_{p_A,p_B}(W > z_{1-\alpha}) = P_{p_A,p_B}\left(\frac{\hat{p}_B - \hat{p}_A - (p_B - p_A)}{\sqrt{V(p_A, p_B, n, \nu_n)}} > \frac{z_{1-\alpha} \cdot \sqrt{V(\hat{p}_A, \hat{p}_B, n, \nu_n)} - (p_B - p_A)}{\sqrt{V(p_A, p_B, n, \nu_n)}}\right)$$
$$\approx 1 - \Phi\left(\frac{z_{1-\alpha} \cdot \sqrt{V(\hat{p}_A, \hat{p}_B, n, \nu_n)} - (p_B - p_A)}{\sqrt{V(p_A, p_B, n, \nu_n)}}\right) \approx 1 - \Phi\left(z_{1-\alpha} - \frac{p_B - p_A}{\sqrt{V(p_A, p_B, n, \nu_n)}}\right),$$

where Φ is the standard normal distribution function, and $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$. The Normal approximation is valid only if $(p_B - p_A)/\sqrt{V(p_A, p_B, n, \nu_n)} = O(1)$, i.e., when $p_B - p_A \approx n^{-1/2}$. However, for fixed $p_B - p_A > 0$, the term $(p_B - p_A)/\sqrt{V(p_A, p_B, n, \nu_n)}$ is of order $\sqrt{n}$, and the expression $\Phi\left(z_{1-\alpha} - \frac{p_B - p_A}{\sqrt{V(p_A, p_B, n, \nu_n)}}\right)$ is of asymptotic order that is smaller than the precision of the normal approximation, and therefore its use is problematic. Thus, the claim that Neyman allocation maximizes the power seems theoretically questionable.

For asymptotic power comparisons and evaluation of the relative asymptotic efficiency of certain tests, two different criteria are often used, related to the notions of Pitman and Bahadur efficiency (see e.g., van der Vaart (1998), Chapter 14). In our context, the Pitman approach looks at sequences of probabilities $p_B^k > p_A^k$ that tend to a common limit at a suitable rate. The Bahadur approach considers fixed probabilities p_A and p_B and approximates the power using large deviations theory.

We show in the next sections that the optimal allocation corresponding to the Pitman approach is always $\nu^* = 0.5$ while the Bahadur optimal allocation depends on p_A and p_B and can be calculated in a way described below. Interestingly, computation of the Bahadur criterion for different values of p_A and p_B reveals that the optimal allocation is often close to 0.5. In disagreement with some of the papers mentioned above, these results cast doubts on the asymptotic justification of adaptive designs and show that, at best, such designs can lead to a practically negligible improvement over a non-sequential balanced design.

The paper is organized as follows: Sections 2 and 3 describe the approaches of Pitman and Bahadur for maximizing the power, and find the corresponding optimal rules. In Section 4, the optimal allocation according to the Bahadur criterion is calculated for different parameters and compared to the Neyman allocation. Exact calculations are performed for a wide range of parameters. A related problem that arises in dose findings experiments is discussed in Section 5; the Neyman allocation is shown to be optimal or close to optimal in this case. Section 6 extends the Bahadur approach to general (rather than binary) responses; concluding remarks are given in Section 7. All proofs are given in the Appendix.

2 The Pitman Approach

Pitman relative efficiency provides an asymptotic comparison of two families of tests applied to a sequence of statistical problems. Here we utilize the same idea to compare different allocation fractions.

Consider a sequence of statistical problems indexed by k, where $p_A^k = p + \frac{\delta_A}{\sqrt{k}}$, $p_B^k = p + \frac{\delta_B}{\sqrt{k}}$, for $\delta_A < \delta_B$ and $0 < p < 1$. Let $n_k = n_k(\delta_A, \delta_B, p, \alpha, \beta, \{\nu_n\})$ be the minimal number of observations required for a one-sided Wald test at significance level α and power at least β (for $\beta > \alpha$) at the point p_A^k, p_B^k, where the observations are allocated to the two groups according to the fraction ν_n. Set $n_k = \infty$ if no finite number of observations satisfies these requirements. The next theorem implies that the balanced allocation is asymptotically optimal.

Theorem 1. *Fix $\delta_A < \delta_B$, $\alpha < \beta$ and $0 < p < 1$. Let $\{\nu_n\}$ be a any sequence of allocations and let $\{\tilde{\nu}_n\}$ be another sequence of allocations satisfying $\tilde{\nu}_n \to 1/2$. Then*

$$\liminf_{k\to\infty} \frac{n_k(\delta_A, \delta_B, p, \alpha, \beta, \{\nu_n\})}{n_k(\delta_A, \delta_B, p, \alpha, \beta, \{\tilde{\nu}_n\})} \geq 1.$$

The theorem follows readily from the following lemma, proved in the Appendix.

Lemma 1. *I. If $\nu_n \to \nu$ for $0 < \nu < 1$ then*

$$\lim_{k\to\infty} \frac{n_k}{k} = \left(\frac{z_{1-\alpha} - z_{1-\beta}}{(\delta_B - \delta_A)} \cdot \sqrt{\frac{p(1-p)}{\nu(1-\nu)}} \right)^2. \qquad (1)$$

II. If $\nu_n \to 0$ or $\nu_n \to 1$ then

$$\lim_{k\to\infty} \frac{n_k}{k} = \infty$$

III. For any sequence of allocations $\{\nu_n\}$

$$\liminf_{k\to\infty} \frac{n_k}{k} \geq \left(\frac{z_{1-\alpha} - z_{1-\beta}}{(\delta_B - \delta_A)} \cdot \sqrt{\frac{p(1-p)}{\frac{1}{4}}} \right)^2.$$

Theorem 1 holds also when considering a two-sided test. The theorem shows that the balanced design is asymptotically optimal in the Pitman sense, and as a consequence, one cannot gain efficiency (in the above sense) by considering sequential adaptive designs. The key point here is that when p_A^k and p_B^k converge to the same value p, the variances of their estimators converge to the same value and hence the limiting Neyman allocation is $1/2$ regardless of p. This phenomenon is not observed in problems concerning the Normal distribution or similar cases where the variance is not a function of the mean.

It can be argued that rather than considering sequences of statistical problems as above, one should optimize for fixed p_A and p_B. The next section deals with this case.

3 The Bahadur Approach

In this section, large deviations theory is used to approximate the power of the Wald test for fixed p_A and p_B. This power increases exponentially to one with n at a rate that depends on the allocation fraction ν. Recall that $\hat{p}_A$ and $\hat{p}_B$ depend on both n and an allocation ν_n. The aim is to find the optimal limiting allocation fraction ν^* for which the rate is maximized. We prove the following large deviations result:

Theorem 2. *Define*

$$H(t,\nu) := \nu \log(1 - p_A + p_A e^{t/\nu}) + (1-\nu)\log(1 - p_B + p_B e^{-t/(1-\nu)}),$$

and let $g(\nu) := \inf_{t>0} H(t,\nu)$.

I. One sided test: assume that $p_B > p_A$ *and* $\nu_n \to \nu$, *where* $0 < \nu < 1$, *then for any constant* $K \geq 0$

$$\lim_n \frac{1}{n} \log \left\{ 1 - P\left(\frac{\hat{p}_B - \hat{p}_A}{\sqrt{V(\hat{p}_A, \hat{p}_B, n, \nu_n)}} > K \right) \right\} = g(\nu). \tag{2}$$

II. Two sided test: assume that $p_B \neq p_A$ *and* $\nu_n \to \nu$, *where* $0 < \nu < 1$, *then for any constant* $K > 0$

$$\lim_n \frac{1}{n} \log \left\{ 1 - P\left(\frac{\{\hat{p}_B - \hat{p}_A\}^2}{V(\hat{p}_A, \hat{p}_B, n, \nu_n)} > K \right) \right\} = g(\nu). \tag{3}$$

III. If $\nu_n \to 0$ *or 1 then (2) and (3) hold with* $g(0) = g(1) = 0$.

Note that $\hat{p}_B - \hat{p}_A$ is not an average of n i.i.d random variables and, therefore, Theorem 2 does not follow directly from the Cramér-Chernoff theorem (see e.g., van der Vaart (1998), p. 205), however, its proof uses similar ideas.

For each fixed n, let $\nu_n^{*(1)} = \nu_n^{*(1)}(p_A, p_B, K)$ be the allocation that maximizes the power of the one sided test for a total sample size of n subjects, i.e,

$$\nu_n^{*(1)} = \arg\max_{\nu_n \in \{\frac{1}{n},\dots,\frac{n-1}{n}\}} P\left(\frac{\hat{p}_B - \hat{p}_A}{\sqrt{V(\hat{p}_A, \hat{p}_B, n, \nu_n)}} > K\right);$$

similarly, $\nu_n^{*(2)} = \nu_n^{*(2)}(p_A, p_B, K)$ is the optimal allocation of the two-sided test.

Let $\nu^* = \nu^*(p_A, p_B) := \arg\min_\nu g(\nu)$. It is easy to prove directly that g is strictly convex, and the minimum is attained uniquely. More generally, it is readily shown by differentiation that if $M(t) = Ee^{tX}$ is a moment generating function, then $\nu M(t/\nu)$ is a convex function of ν. Theorem 2 suggests the use of ν^* as the design fraction. However, for a given n, the optimal allocation, is not necessarily ν^*, but the fraction $\nu_n^{*(1)}$ or $\nu_n^{*(2)}$ for the one or two-sided test, respectively. Therefore, it is reasonable to use ν^* as the design fraction only if $\nu_n^{*(i)} \to \nu^*$ for $i = 1, 2$. The following theorem shows that this is indeed the case.

Theorem 3. *I. If $p_B > p_A$ then for any $K \geq 0$, $\nu_n^{*(1)} \to \nu^*$.*

II. If $p_B \neq p_A$ then for any $K > 0$, $\nu_n^{*(2)} \to \nu^*$.

Remark 1. *Another formulation of these results, for the one-sided case, say, is the following: assume that $p_B > p_A$ then for any sequence ν_n and constant $K \geq 0$*

$$\liminf_n \frac{1}{n} \log\left\{1 - P\left(\frac{\hat{p}_B - \hat{p}_A}{\sqrt{V(\hat{p}_A, \hat{p}_B, n, \nu_n)}} > K\right)\right\} \geq g(\nu^*),$$

and the infimum is attained for sequences $\nu_n \to \nu^*$.

Remark 2. *When $p_A < p_B$ represent success probabilities of two treatments, and treatment B is selected as better if $\hat{p}_B(n) > \hat{p}_A(n)$, then the expression in (2) with $K = 0$ approximates the probability of incorrect selection.*

4 Numerical Illustration

Some tedious calculations show that

$$\nu^* = \log\left\{\frac{p_B \log(\frac{p_B}{p_A})}{(1-p_B)\log(\frac{1-p_A}{1-p_B})}\right\} \Big/ \log\left\{\frac{p_B(1-p_A)}{p_A(1-p_B)}\right\}. \tag{4}$$

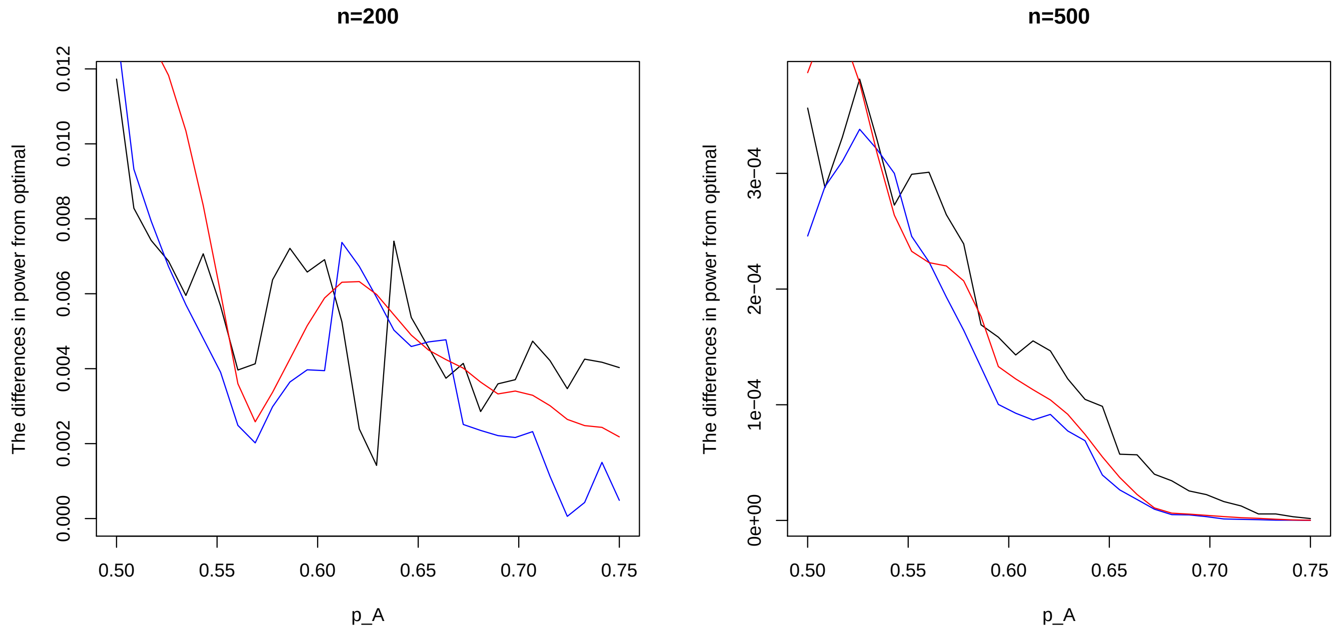
Table 1: The optimal Bahadur allocation ν^* for different parameters compared to Neyman allocation.

p_A	p_B	ν^*	Neyman allocation
0.5	0.8	0.518	0.556
0.5	0.65	0.504	0.512
0.6	0.75	0.510	0.531
0.7	0.75	0.505	0.514
0.7	0.85	0.521	0.562
0.7	0.9	0.535	0.604
0.85	0.95	0.541	0.621
0.5	0.9	0.542	0.625

Table 1 compares the asymptotic Bahadur optimal allocation and the Neyman allocation for several pairs (p_A, p_B). The table and further systematic numerical calculations indicate that the Bahadur allocation is closer to 0.5 than the Neyman allocation and that it is quite close to 0.5 unless p_A and p_B are very far apart (e.g., $p_A = 0.5, p_B = 0.9$). In the latter case, the power is close to 1 for any reasonable allocation. These findings justify the use of the balanced allocation and question the utility of more complicated adaptive sequential designs.

We preformed some exact calculations to compare the Bahadur allocation, the balanced allocation and the Neyman allocation. Figure 1 compares the difference between the maximal possible power for sample size 200 and 500, and the power under the different allocation methods for the two-sided test with $\alpha = 0.05$ and for different parameters. The power is calculated exactly using R. While for moderate sample size ($n = 200$) no allocation is better for all the parameters we considered, for large sample size ($n = 500$), Bahadur is better for almost all parameters, and the balanced allocation is usually better than Neyman; however, the differences in power are relatively small.

Figure 2 shows the power of the two-sided test for different allocations where $p_A = 0.7, p_B = 0.9$; it is clearly seen that the Neyman allocation, which is widely recommended for maximizing the power, is far from being optimal. Thus, the exact calculations presented in this section support the theoretical results: the balanced allocation is usually better than the Neyman allocation for large samples, and they are indistinguishable for small samples. In all cases, the differences are quite negligible, and therefore the balanced allocation should be preferred due to its simplicity.



(a) Moderate sample size ($n = 200$)

(b) Large sample size ($n = 500$)

Figure 1: The differences between the maximal power of the two-sided test with critical value $K = 1.96$, attainable (by $\nu_n^{*(2)}$) and the Neyman allocation (black), the balanced allocation (red) and the Bahadur allocation (blue) for $p_A = 0.5, \ldots, 0.75$, $p_B = p_A + 0.2$; for moderate ($n = 200$) and large ($n = 500$) sample size.

5 A Related Problem

Dose finding studies are conducted as part of phase I clinical trials in order to find the maximal tolerated dose (MTD) among a finite, usually very small, number of potential doses. The MTD is defined as the dose with the closest probability of toxic reaction to a pre-specified probability p_0. Recently, we showed that under certain natural assumptions, in order to estimate the desired dose consistently, one can consider experiments that eventually concentrate on two doses (Azriel et. al (2010)). Thus, asymptotically, the allocation problem in MTD studies reduces to the problem of finding which of two probabilities of toxic reaction $p_A < p_B$ (corresponding to the doses $d_A < d_B$) is closer to p_0.

Let $\hat{p}_A$ and $\hat{p}_B$ denote the proportions of toxic reactions in doses d_A and d_B based on a total sample size of n individuals and an allocation ν_n. For large n, $\hat{p}_A < \hat{p}_B$, and a natural estimator for the MTD is $\widehat{MTD} = d_A$ if $(\hat{p}_A + \hat{p}_B)/2 > p_0$ and $\widehat{MTD} = d_B$ otherwise. Similar to the problems discussed in previous sections, an optimal design is an allocation rule of $n\,\nu_n$ and $n(1-\nu_n)$ individuals to doses d_A and d_B, respectively, such that $P(\widehat{MTD} = d_A) = P((\hat{p}_A + \hat{p}_B)/2 > p_0)$ is maximized if

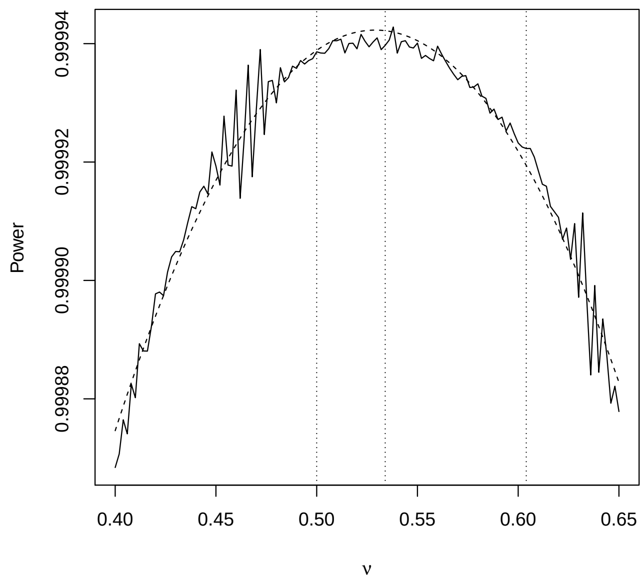


Figure 2: The power of Wald tests with critical value $K = 1.96$ for different allocations ν where $p_A = 0.7, p_B = 0.9$ and $n = 500$. The smooth line is a parabolic fit to the function. The vertical lines show the balanced allocation ($\nu = 0.5$), the Bahadur allocation ($\nu = 0.5349374$) and the Neyman allocation ($\nu = 0.6043561$).

d_A is indeed the MTD.

For the current problem, the Pitman approach is translated to a comparison of designs under sequences of parameters p_A^k, p_B^k and p_0^k such that $|(p_A^k + p_B^k)/2 - p_0^k| = K/\sqrt{k}$, for fixed $0 < K < \infty$, and $p_A^k \to p_A$, $p_B^k \to p_B$. Let $0 < \nu < 1$ and let $n_k = n_k(p_A^k, p_B^k, p_0^k, \alpha, \{\nu_n\})$ be the minimal number of observations required such that the probability of incorrect estimation of the MTD is smaller than α for the given parameters when the allocation for dose d_A is $n \cdot \nu_n$. As in Lemma 1, it can be shown that if $\nu_n \to \nu$ then

$$\lim_{k\to\infty} \frac{n_k}{k} = \left\{\frac{z_{1-\alpha}}{2K}\right\}^2 \left\{\frac{p_A(1-p_A)}{\nu} + \frac{p_B(1-p_B)}{1-\nu}\right\}.$$

Thus, the asymptotically optimal design uses Neyman allocation, $\nu = \frac{\sqrt{p_A(1-p_A)}}{\sqrt{p_A(1-p_A)}+\sqrt{p_B(1-p_B)}}$, as it minimizes the limit of n_k/k. Unlike the previous problem, now p_A^k and p_B^k do not converge to the same value under the Pitman approach as defined here, and hence the Neyman allocation does not reduce to the balanced design.

For the case of fixed p_A, p_B, and p_0, assume that p_B is nearer than p_A to p_0, and consider the problem of minimizing the probability of selecting d_A. The following theorem, analogous to Theorems 2 and 3, gives the asymptotic optimal allocation rule in the current setting.

Table 2: Comparison of the Bahadur and Pitman (here Neyman) allocation rules for different parameters.

p_A	p_B	p_0	Bahadur	Pitman
0.1	0.3	0.28	0.420	0.396
0.2	0.35	0.3	0.460	0.456
0.22	0.33	0.3	0.471	0.468
0.25	0.35	0.33	0.479	0.476
0.2	0.4	0.33	0.455	0.449
0.1	0.4	0.3	0.400	0.380

Theorem 4. *Let $\nu_n = N_A(n)/n$, $0 < \nu < 1$, and assume that $\nu_n \to \nu$, then,*

$$\lim_{n\to\infty} \frac{1}{n} \log P[\{\hat{p}_A + \hat{p}_B\}/2 \geq p_0] = \psi(\nu),$$

where $\psi(\nu) = \inf_t\{\nu \log(1 - p_A + p_A e^{t/\nu}) + (1-\nu)\log(1 - p_B + p_B e^{t/(1-\nu)}) - 2p_0 t\}$.

Moreover, let $\nu^* = \arg\min \psi(\nu)$, and let ν_n^* be the value of the allocation minimizing $P[\{\hat{p}_A + \hat{p}_B\}/2 \geq p_0]$ for a given n. Then, $\nu_n^* \to \nu^*$.

We calculated ν^* for several values of p_A and p_B and found that ν^* is often close to the Neyman allocation, see Table 2. Both criteria, Bahadur and Pitman, yield quite similar results in this problem. Allocating subjects according to the Neyman or Bahadur improves the probability of correct MTD estimation compared to the balanced allocation for very large samples, as the optimal allocations according to Bahadur or Pitman are far from 0.5. Calculations not presented here, show that for practical sample sizes for the MTD problem, all three methods differ in a negligible way.

6 A General Response

In previous sections, we dealt with the very important, though specific, case of a binary response. In this section, we consider the more general case where the response of an individual treated in group A (B) follows a distribution F_A (F_B) having moment generating function $M_A(t)$ ($M_B(t)$), and find the optimal allocation according to the Bahadur approach. Let $\bar{Y}_A(m)$ ($\bar{Y}_B(m)$) denote the average of m responses of subjects having treatment A (B). Assume that the treatment with the largest mean response is declared better at the end of the experiment. The following theorem, which can

Table 3: The optimal allocation ν^* for different distributions compared to the Neyman allocation.

F_A	F_B	Bahadur allocation	Neyman allocation
Poisson(1)	Poisson(2)	0.471	0.414
Poisson(2)	Poisson(3)	0.483	0.449
Poisson(3)	Poisson(4)	0.488	0.464
Poisson(4)	Poisson(5)	0.491	0.472
Gamma(0.5,0.5)	Gamma(0.5,0.6)	0.515	0.590
Gamma(0.5,0.5)	Gamma(0.5,0.7)	0.528	0.662
Gamma(0.5,0.5)	Gamma(0.5,0.8)	0.539	0.719
Gamma(0.5,0.5)	Gamma(0.5,0.9)	0.549	0.764

be proved in a similar way as Theorems 2 and 3, provides the Bahadur optimal allocation rule for correct selection:

Theorem 5. *Assume that treatment B is better, i.e,* $\int xF_B(dx) > \int xF_A(dx)$*, and that* $\nu_n \to \nu$*. Then,*

$$\lim_{n\to\infty} \frac{1}{n} \log P\left\{\bar{Y}_A(n\,\nu_n) \geq \bar{Y}_B(n(1-\nu_n))\right\} = h(\nu),$$

where

$$h(\nu) = \inf_t [\nu \log\{M_A(t/\nu)\} + (1-\nu)\log\{M_B[-t/(1-\nu)]\}]. \tag{5}$$

Moreover, let $\nu^* = \arg\min_\nu h(\nu)$*, and* ν_n^* *be the value of the allocation minimizing* $P\left\{\bar{Y}_A(n\,\nu_n) \geq \bar{Y}_B(n(1-\nu_n))\right\}$*. Then* $\nu_n^* \to \nu^*$*.*

When the responses in the two treatments are normally distributed, then the Bahadur allocation agrees with the Neyman allocation. This can be easily verified by using the moment generating functions of Normal variables in (5). However, for other distributions, the allocations suggested by the Bahadur and the Neyman criteria may differ considerably. Table 3 compares the Bahadur and the Neyman allocations for different Poisson and Gamma distributions. The two rules clearly differ. As in the Binomial case, the Bahadur allocation is closer to 0.5 than to the Neyman allocation. Further study is required to determine if the improvement over the balanced allocation, in terms of power or probability of correct selection, is significant. Anyway, optimality of the Neyman allocation for non-normal distributions should be questioned, and may hold only under restrictive conditions.

7 Conclusions

We discussed asymptotic approximations of power and probability of correct selection in testing and selecting the best treatment, and in MTD finding, and related optimal allocation of subjects to treatments.

Neyman allocation is optimal when the response is Normal, and it is asymptotically optimal in the Pitman sense, that is, for converging sequences of alternatives as described above. In the binary response selection problem in which p_A and p_B become closer, Neyman allocation reduces to a balanced allocation, independent of the parameters p_A and p_B. The Bahadur allocation for fixed p_A and p_B turns out to be close to balanced, and therefore, by both criteria, our conclusion is that adaptive allocation seems unwarranted, and the simpler, non-sequential balanced allocation should be preferred.

Our findings are partly in contrast with the literature that bases allocations on the noncentrality parameter appearing in a Normal or Chi-Square approximation (e.g., Rosenberger et. al (2001); Tymofyeyev et. al (2007)). These designs minimize or control the variance of the difference but need not be efficient in the sense of controlling or maximizing the power.

A Appendix

Proof of Lemma 1 part I. The proof, included here for completeness, uses arguments as in Theorem 14.19 in van der Vaart (1998) (p. 205), which is stated in terms of relative efficiency rather than allocation.

First note that $\lim_k n_k = \infty$; otherwise, there exists a bounded subsequence of n_k on which the power converges to a value $\leq \alpha$, since as $k \to \infty$ we have $p_A^k - p_B^k \to 0$. This contradicts the definition of n_k and the assumption that $\alpha < \beta$.

By the Berry-Esseen theorem we have

$$\frac{\hat{p}_A^k - (p + \frac{\delta_A}{\sqrt{k}})}{\sqrt{\frac{(p+\frac{\delta_A}{\sqrt{k}})[1-(p+\frac{\delta_A}{\sqrt{k}})]}{\nu_{n_k} \cdot n_k}}} \xrightarrow{D} N(0,1)$$

since the third moment is bounded; a similar limit holds for $\hat{p}_B^k$. Here we use the notation $\hat{p}_A^k = \hat{p}_A(\nu_{n_k} n_k) = Y_A^k(\nu_{n_k} n_k)/(\nu_{n_k} n_k)$, where $Y_A^k(m) \sim Bin(m, p_A^k)$ is the sum of m independent binary responses with probability p_A^k.

Now, if $\nu_{n_k} \to \nu$ we have

$$U_k := \frac{\sqrt{n_k}(\hat{p}_B^k - \hat{p}_A^k) - (\delta_B - \delta_A)\sqrt{\frac{n_k}{k}}}{\sqrt{\frac{p(1-p)}{\nu(1-\nu)}}} \xrightarrow{D} N(0,1). \tag{6}$$

Since $n_k \to \infty$, the critical value for the level α one-sided Wald test is $z_{1-\alpha} + o(1)$; then

$$P_{p_A^k, p_B^k}(W > z_{1-\alpha} + o(1)) = P\left(\frac{\hat{p}_B^k - \hat{p}_A^k}{\sqrt{V(\hat{p}_A^k, \hat{p}_B^k, n_k, \nu_{n_k})}} > z_{1-\alpha} + o(1)\right)$$
$$= P\left(U_k > \frac{(z_{1-\alpha} + o(1))\sqrt{V(\hat{p}_A^k, \hat{p}_B^k, n_k, \nu_{n_k})n_k}}{\sqrt{\frac{p(1-p)}{\nu(1-\nu)}}} - \frac{(\delta_B - \delta_A)\sqrt{\frac{n_k}{k}}}{\sqrt{\frac{p(1-p)}{\nu(1-\nu)}}}\right).$$

Also,

$$\frac{\sqrt{V(\hat{p}_A^k, \hat{p}_B^k, n_k, \nu_{n_k})n_k}}{\sqrt{\frac{p(1-p)}{\nu(1-\nu)}}} \xrightarrow{a.s.} 1,$$

and since the limiting power is exactly β we have due to (6)

$$z_{1-\alpha} - \frac{(\delta_B - \delta_A)\{\lim_k \sqrt{\frac{n_k}{k}}\}}{\sqrt{\frac{p(1-p)}{\nu(1-\nu)}}} = z_{1-\beta};$$

hence (1) holds.

Proof of part II. We only prove the case $\nu_n \to 0$, as $\nu_n \to 1$ is similar. If $n\nu_n$ is bounded, then the power converges to α and $n_k = \infty$ for large k.

Assume now that $n\,\nu_n \to \infty$; by the Berry-Esseen theorem and Slutsky's Lemma we have

$$\sqrt{n_k\,\nu_{n_k}}(\hat{p}_A^k - (p + \frac{\delta_A}{\sqrt{k}})) \xrightarrow{D} N(0, p(1-p)) \text{ and } \sqrt{n_k\,\nu_{n_k}}(\hat{p}_B^k - (p + \frac{\delta_B}{\sqrt{k}})) \xrightarrow{D} 0.$$

This implies that

$$\sqrt{n_k\,\nu_{n_k}}(\hat{p}_B^k - \hat{p}_A^k) - (\delta_B - \delta_A)\sqrt{\frac{n_k\,\nu_{n_k}}{k}} \xrightarrow{D} N(0, p(1-p))$$

and by arguments as in the first part we have

$$\lim_k \frac{n_k\,\nu_{n_k}}{k} = \left(\frac{z_{1-\alpha} - z_{1-\beta}}{\delta_B - \delta_A}\right)^2 p(1-p).$$

Because $\nu_{n_k} \to 0$, $\lim_k \frac{n_k}{k} = \infty$.

Proof of part III. There exists a subsequence $\{k'\}$ such that $\nu_{n_{k'}} \to \nu'$ for some ν' and

$$\liminf_{k\to\infty} \frac{n_k}{k} = \lim_{k'\to\infty} \frac{n_{k'}}{k'} = \left(\frac{z_{1-\alpha} - z_{1-\beta}}{(\delta_B - \delta_A)} \cdot \sqrt{\frac{p(1-p)}{\nu'(1-\nu')}}\right)^2,$$

where the second equality follows by part I. If $\nu'(1-\nu') = 0$ we interpret the limit as ∞; since $\nu'(1-\nu') \le \frac{1}{4}$ the third part of the lemma follows. $\square$

Proof of Theorem 2 parts I and II. The proof follows known large deviations ideas; however, certain variations are needed for the present non-standard case. Notice that the probability in part I is larger than the probability of part II (for $\sqrt{K}$). Therefore, it is enough to show that for any $K \ge 0$

$$\limsup_n \frac{1}{n} \log \left\{ 1 - P\left(\frac{\hat{p}_B - \hat{p}_A}{\sqrt{V(\hat{p}_A, \hat{p}_B, n, \nu_n)}} > K \right) \right\} \le g(\nu), \tag{7}$$

and for any $K > 0$

$$\liminf_n \frac{1}{n} \log \left\{ 1 - P\left(\frac{\{\hat{p}_B - \hat{p}_A\}^2}{V(\hat{p}_A, \hat{p}_B, n, \nu_n)} > K \right) \right\} \ge g(\nu).$$

In fact, instead of the latter inequality we prove in the sequel a stronger result, namely

$$\liminf_n \frac{1}{n} \log P\left(0 \le \frac{\hat{p}_A - \hat{p}_B}{\sqrt{V(\hat{p}_A, \hat{p}_B, n, \nu_n)}} \le K' \right) \ge g(\nu), \tag{8}$$

for all $K' > 0$, which is also used for the case of $K = 0$ in part I, when $K' = \infty$.

For the upper bound (7), define $S(n) := \sqrt{V(\hat{p}_A, \hat{p}_B, \nu_n, n) \cdot n}$; notice that $S(n)$ is bounded. Hence, for any $\varepsilon > 0$ and for large enough n we have

$$1 - P\left(\frac{\hat{p}_B - \hat{p}_A}{\sqrt{V(\hat{p}_A, \hat{p}_B, n, \nu_n)}} > K \right) = P\left(\hat{p}_A - \hat{p}_B \ge -\frac{K}{\sqrt{n}} S(n) \right) \le P\left(\hat{p}_A - \hat{p}_B \ge -\varepsilon \right).$$

Now, for any $t > 0$,

$$P(\hat{p}_A - \hat{p}_B \ge -\varepsilon) = P(e^{t(\frac{Y_A(n\,\nu_n)}{N_A(n)/n} - \frac{Y_B(n(1-\nu_n))}{N_B(n)/n})} \ge e^{-nt\varepsilon}) \le E[e^{t(\frac{Y_A(n\,\nu_n)}{\nu_n} - \frac{Y_B(n(1-\nu_n))}{1-\nu_n})}]e^{nt\varepsilon},$$

by Markov's inequality. We can write the latter term as

$$(1 - p_A + p_A e^{t/\nu_n})^{n\,\nu_n} \cdot (1 - p_B + p_B e^{-t/(1-\nu_n)})^{n(1-\nu_n)} e^{nt\varepsilon}.$$

Since $\nu_n \to \nu$, and the inequality holds for all $t > 0$,

$$\limsup_n \frac{1}{n} \log P\left(\hat{p}_A - \hat{p}_B \ge -\frac{K}{\sqrt{n}} S(n) \right) \le g_\varepsilon(\nu),$$

where $g_\varepsilon(\nu) := \inf_{t>0}\{\varepsilon t + H(t, \nu)\}$. This is true for any $\varepsilon > 0$, and by the continuity of $g_\varepsilon(\nu)$ in ε we have for any $K \ge 0$

$$\limsup_n \frac{1}{n} \log P\left(\hat{p}_A - \hat{p}_B \ge -\frac{K}{\sqrt{n}} S(n) \right) \le g(\nu),$$

which verifies (7).

To prove (8), assume without loss of generality that $p_B > p_A$; define

$$T_n := \hat{p}_A(n\,\nu_n) - \hat{p}_B(n(1-\nu_n)) = \frac{Y_A(n\,\nu_n)}{n\,\nu_n} - \frac{Y_B(n(1-\nu_n))}{n(1-\nu_n)}.$$

The log of the moment generating function of T_n is

$$\log E[e^{tT_n}] = n\,\nu_n \log(1 - p_A + p_A e^{\frac{t}{n\,\nu_n}}) + n(1-\nu_n)\log(1 - p_B + p_B e^{\frac{-t}{n(1-\nu_n)}}) = nH(\frac{t}{n}, \nu_n). \tag{9}$$

Since $E[T_n] = p_A - p_B < 0$, by (9) we have $\frac{d}{dt}H(0,\nu_n) < 0$. Also, $H(0,\nu_n) = 0$ and $H(\cdot,\nu_n)$ is strictly convex being the log of a moment generating function, up to a constant. Since $P(T_n > 0) > 0$ it follows that $H(t,\nu_n) \to \infty$ as $t \to \infty$ and therefore, $\arg\min_{t>0} H(t,\nu_n) =: t_0^{(n)}$ is a unique interior point and $\frac{\partial}{\partial t}H(t_0^{(n)},\nu_n) = 0$. Let t_0 be the minimizer of $H(\cdot,\nu)$; we show that $t_0^{(n)} \to t_0$. If there is a subsequence $\{t_0^{(n_k)}\}$ that converges to $t_1 \le \infty$ then $H(t_0^{(n_k)},\nu_{n_k}) \le H(t_0,\nu_{n_k})$ (as $t_0^{(n_k)}$ is the minimizer) implies $H(t_1,\nu) \le H(t_0,\nu)$ and therefore $t_1 = t_0$ as the minimizer is unique and finite.

Define a new random variable Z_n, which is the Cramér transform of T_n

$$P(Z_n = z) := e^{-ng(\nu_n)} e^{z t_0^{(n)} n} P(T_n = z).$$

Now,

$$\begin{aligned} &P\left(0 \le \frac{\hat{p}_A - \hat{p}_B}{\sqrt{V(\hat{p}_A,\hat{p}_B,n,\nu_n)}} \le K\right) = P\left(0 \le T_n \le \frac{K}{\sqrt{n}}S(n)\right) \\ &= E[I\{0 \le Z_n \le \frac{K}{\sqrt{n}}S(n)\} e^{-Z_n t_0^{(n)} n}] e^{ng(\nu_n)} \ge P\big(0 \le Z_n \le \frac{K}{\sqrt{n}}S(n)\big) e^{-K\frac{1}{2}\sqrt{\frac{1}{\nu_n}+\frac{1}{1-\nu_n}}\, t_0^{(n)}\sqrt{n}} e^{ng(\nu_n)}, \end{aligned}$$

where the last inequality holds since $e^{-Z_n} \ge e^{-\frac{K}{\sqrt{n}}S(n)} \ge e^{-\frac{K}{\sqrt{n}}\frac{1}{2}\sqrt{\frac{1}{\nu_n}+\frac{1}{1-\nu_n}}}$. It follows that

$$g(\nu_n) - \frac{1}{n}\log P\left(0 \le \frac{\hat{p}_A - \hat{p}_B}{\sqrt{V(\hat{p}_A,\hat{p}_B,n,\nu_n)}} \le K\right) \le \frac{-\log P\big(0 \le Z_n \le \frac{K}{\sqrt{n}}S(n)\big)}{n} + \frac{K\frac{1}{2}\sqrt{\frac{1}{\nu_n}+\frac{1}{1-\nu_n}}\, t_0^{(n)}}{\sqrt{n}}.$$

Clearly, the second term on the right-hand side vanishes as n goes to infinity; for the first, we claim that $\sqrt{n}Z_n$ is asymptotically $N(0,\frac{\partial^2}{\partial t^2}H(t_0^{(n)},\nu_n))$ and consequently $P\big(0 \le Z_n \le \frac{K}{\sqrt{n}}S(n)\big) \to C$ for some constant $C > 0$. Indeed, the log of the moment generating function of $\sqrt{n}Z_n$ is

$$\log E[e^{s\sqrt{n}Z_n}] = -ng(\nu_n) + \log E[e^{T_n(s\sqrt{n}+t_0^{(n)}n)}] = n\{-H(t_0^{(n)},\nu_n) + H(t_0^{(n)} + \frac{s}{\sqrt{n}},\nu_n)\},$$

where the last equality follows from (9) and the identity $g(\nu_n) = H(t_0^{(n)},\nu_n)$. By Taylor expansion of $H(\cdot,\nu_n)$ around $t_0^{(n)}$ we obtain

$$H(t_0^{(n)} + \frac{s}{\sqrt{n}},\nu_n) - H(t_0^{(n)},\nu_n) = \frac{1}{2}\frac{s^2}{n}\frac{\partial^2}{\partial t^2}H(t_0^{(n)},\nu_n) + O(n^{-3/2})$$

since the first derivative is 0, and therefore,

$$\log E[e^{s\sqrt{n}Z_n}] \to \frac{s^2}{2}\frac{\partial^2}{\partial t^2}H(t_0^{(n)},\nu_n).$$

We conclude that

$$\limsup_n \left\{ g(\nu_n) - \frac{1}{n}\log P\left(0 \le \frac{\hat{p}_A - \hat{p}_B}{\sqrt{V(\hat{p}_A,\hat{p}_B,n,\nu_n)}} \le K\right)\right\} \le 0,$$

hence,

$$\liminf \frac{1}{n}\log P\left(0 \le \frac{\hat{p}_A - \hat{p}_B}{\sqrt{V(\hat{p}_A,\hat{p}_B,n,\nu_n)}} \le K\right) \ge g(\nu)$$

and part I and II follow.

Proof of part III. First note that (7) clearly holds with $g(\nu) = 0$ as $\log\{1 - P(\cdot)\} \le 0$, so it remains to prove (8) for $g(\nu) = 0$, that is, for any $K > 0$

$$\liminf_n \frac{1}{n}\log P\big(0 \le \frac{\hat{p}_A - \hat{p}_B}{\sqrt{V(\hat{p}_A,\hat{p}_B,\nu_n,n)}} \le K\big) \ge 0.$$

We only prove the case $\nu_n \to 0$, as $\nu_n \to 1$ is similar. If $n\,\nu_n \not\to \infty$ then $\hat{p}_A$ is inconsistent and the limit is easily seen to be zero. Assume now that $n\,\nu_n \to \infty$; since

$$V(\hat{p}_A,\hat{p}_B,\nu_n,n) = \frac{\hat{p}_A(1-\hat{p}_A)}{n\,\nu_n} + \frac{\hat{p}_B(1-\hat{p}_B)}{n(1-\nu_n)} \ge \frac{\hat{p}_A(1-\hat{p}_A)}{n\,\nu_n}$$

we have

$$P\big(0 \le \frac{\hat{p}_A - \hat{p}_B}{\sqrt{V(\hat{p}_A,\hat{p}_B,\nu_n,n)}} \le K\big) \ge P\big(0 \le \hat{p}_A - \hat{p}_B \le \frac{K\sqrt{\hat{p}_A(1-\hat{p}_A)}}{\sqrt{n\,\nu_n}}\big).$$

Now, for $\varepsilon := \frac{Kp_A(1-p_A)}{2}$,

$$P\big(0 \le \hat{p}_A - \hat{p}_B \le \frac{K\sqrt{\hat{p}_A(1-\hat{p}_A)}}{\sqrt{n\,\nu_n}}\big) \ge P\big(\{0 \le \hat{p}_A - \hat{p}_B \le \frac{K\sqrt{\hat{p}_A(1-\hat{p}_A)}}{\sqrt{n\,\nu_n}}\} \cap \{\hat{p}_B \in (p_B - \frac{\varepsilon}{\sqrt{n\,\nu_n}}, p_B)\}\big)$$
$$\ge P\big(p_B \le \hat{p}_A \le p_B + \frac{K\sqrt{\hat{p}_A(1-\hat{p}_A)} - \varepsilon}{\sqrt{n\,\nu_n}}\big)P\big(\hat{p}_B \in (p_B - \frac{\varepsilon}{\sqrt{n\,\nu_n}}, p_B)\big).$$

Taking logs and limits in the above product, we have to consider two parts. For the first, we have by Lemma 2 below

$$\lim_n \frac{1}{n\,\nu_n}\log P\big(p_B \le \hat{p}_A \le p_B + \frac{K\sqrt{\hat{p}_A(1-\hat{p}_A)} - \varepsilon}{\sqrt{n\,\nu_n}}\big) = C.$$

for some constant C; therefore,

$$\lim_n \frac{1}{n}\log P\big(p_B \le \hat{p}_A \le p_B + \frac{K\sqrt{\hat{p}_A(1-\hat{p}_A)} - \varepsilon}{\sqrt{n\,\nu_n}}\big) = 0.$$

The limit of the log of the second part divided by n is 0, since

$$P\big(\hat{p}_B \in (p_B - \frac{\varepsilon}{\sqrt{n\,\nu_n}}, p_B)\big) \geq P(-\varepsilon \leq \sqrt{n(1-\nu_n)}(\hat{p}_B - p_B) \leq 0) \to C' > 0$$

by the CLT. □

Lemma 2. *Let $V_1, V_2, \ldots$ be i.i.d with $EV_1 < 0$ and moment generation function $M(t)$, and let X_n be positive and uniformly bounded random variables that satisfy $X_n \overset{a.s.}{\to} K$ for a constant $K > 0$; then,*

$$\lim_n \frac{1}{n} \log P\big(0 \leq \bar{V}_n \leq \frac{X_n}{\sqrt{n}}\big) = \inf_{t>0}\{\log M(t)\}.$$

Proof of Lemma 2. The lemma follows by the same argument as in van der Vaart (1998), p. 206 (replacing ε in that proof by $\frac{\tilde{K}}{\sqrt{n}}$, where $\tilde{K}$ is the bound of X_n); see also the proof of parts I and II of Theorem 2, where a similar argument is used. □

Proof of Theorem 3. We will prove part I; the proof of Part II is similar. Consider the sequence of allocations $\nu'_n = \frac{\lfloor n \cdot \nu^* \rfloor}{n}$; Theorem 2 implies that

$$\lim_n \frac{1}{n} \log \left\{ 1 - P\left(\frac{\hat{p}_B(n(1-\nu'_n)) - \hat{p}_A(n\,\nu'_n)}{\sqrt{V(\hat{p}_A, \hat{p}_B, n, \nu'_n)}} > K \right) \right\} = g(\nu^*). \tag{10}$$

Now, let $\tilde{\nu} \neq \nu^*$, $0 \leq \tilde{\nu} \leq 1$, be a limit of a certain subsequence $\{n_k\}$, i.e., $\nu^{*(1)}_{n_k} \to \tilde{\nu}$, and define $\varepsilon = (g(\tilde{\nu}) - g(\nu^*))/2$. By (10), there exists N such that for $n \geq N$

$$\frac{1}{n} \log \left\{ 1 - P\left(\frac{\hat{p}_B(n(1-\nu'_n)) - \hat{p}_A(n\,\nu'_n)}{\sqrt{V(\hat{p}_A, \hat{p}_B, n, \nu'_n)}} > K \right) \right\} < g(\nu^*) + \varepsilon.$$

Since $\nu^{*(1)}_{n_k} \to \tilde{\nu}$ we have by Theorem 2, for large enough k

$$\frac{1}{n_k} \log \left\{ 1 - P\left(\frac{\hat{p}_B(n_k(1-\nu^{*(1)}_{n_k})) - \hat{p}_A(n_k\,\nu^{*(1)}_{n_k})}{\sqrt{V(\hat{p}_A, \hat{p}_B, n_k, \nu^{*(1)}_{n_k})}} > K \right) \right\} > g(\tilde{\nu}) - \varepsilon = g(\nu^*) + \varepsilon,$$

where $g(0) = g(1) = 0$. Hence, there exists $n_k > N$ such that

$$P\left(\frac{\hat{p}_B(n_k(1-\nu^{*(1)}_{n_k})) - \hat{p}_A(n_k\,\nu^{*(1)}_{n_k})}{\sqrt{V(\hat{p}_A, \hat{p}_B, n_k, \nu^{*(1)}_{n_k})}} > K \right) < P\left(\frac{\hat{p}_B(n_k(1-\nu'_{n_k})) - \hat{p}_A(n_k\,\nu'_{n_k})}{\sqrt{V(\hat{p}_A, \hat{p}_B, n_k, \nu'_{n_k})}} > K \right)$$

in contradiction to the optimality of $\nu^{*(1)}_{n_k}$; therefore the limit of every converging sub-sequence is ν^*.

□

The proofs of Theorems 4 and 5 are omitted because they are very similar to the proofs of Theorems 2 and 3.

References

Azriel, D., Mandel, M., and Rinott Y. (2010). The treatment versus experimentation dilemma in dose-finding studies . *Center for Rationality Discussion paper*, **559**. To appear in *Journal of Statistical Planning and Inference.*

Bandyopadhyay U. and Bhattacharya, R. (2006). Adaptive Allocation and Failure Saving in Randomised Clinical Trials, *Journal of Biopharmaceutical Statistics*, **16** 817 - 829.

Biswas A., Mandal S., Bhattacharya R. (2010). Multi-treatment optimal response-adaptive designs for phase III clinical trials, *Journal of the Korean Statistical Society*, in press.

Brittain E., Schlesselman J.J. (1982). Optimal Allocation for the Comparison of Proportions, *Biometrics*, **38**, 1003 – 1009.

Brown L.D., Cai T.T., DasGupta A. (2001). Interval Estimation for a Binomial Proportion, *Statistical Science*, **16**, 101 – 117.

Chambaz A., van der Laan M.J. (2010). Targeting The Optimal Design In Randomized Clinical Trials With Binary Outcomes And No Covariate *U.C. Berkeley Division of Biostatistics Working Paper Series*, **258**.

Hu F., Rosenberger W.F. (2003), Optimality, Variability, Power: Evaluating Response-Adaptive Randomization Procedures for Treatment Comparisons, *Journal of the American Statistical Association*, **98**, 671–678.

Hu F., Rosenberger W.F. (2006). *The theory of response-adaptive randomization in clinical trials*, Wiley, New York.

Hu F., Rosenberger W.F. and Zhang L. (2006). Asymptotically best response-adaptive randomization procedures, *Journal of Statistical Planning and Inference*, **136** 1911 - 1922.

Melfi V.F., Page C. and Geraldes M. (2001). An adaptive randomized design with application to estimation, *The Canadian Journal of Statistics*, **29** 107 – 116.

Rosenberger W.F., Stallard N., Ivanova A., Harper C. N., and Ricks M. L. (2001). Optimal Adaptive Designs for Binary Response Trials, *Biometrics*, **57** 909 – 913.

Tymofyeyev Y., Rosenberger W.F. and Hu F. (2007). Implementing Optimal Allocation in Sequential Binary Response Experiments, *Journal of the American Statistical Association*, **102** 224 – 234.

van der Vaart A.W. (1998). *Asymptotic statistics*, Cambridge University Press, New York.

Zhu H. and Hu F. (2010). Sequential Monitoring of response-adaptive randomized clinical trials, *The Annals of Statistics* **38**, 2218 - 2241.