Biometrika (2010), **97**, 4, *pp*. 935–946
doi: 10.1093/biomet/asq055

Advance Access publication 15 November 2010
Printed in Great Britain

Compound optimal allocation for individual and collective ethics in binary clinical trials

BY ALESSANDRO BALDI ANTOGNINI AND ALESSANDRA GIOVAGNOLI

Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, 40126, Bologna, Italy

a.baldi@unibo.it   alessandra.giovagnoli@unibo.it

SUMMARY

In recent years, several authors have investigated response-adaptive allocation rules for comparative clinical trials, in order to favour, at each stage of the trial, the treatment that appears to be best. In this paper, we define admissible allocations, namely treatment assignments that cannot be simultaneously improved upon with respect to both a specific design criterion, reflecting the inferential properties of the experiment, and the proportion of patients assigned to the best treatment or treatments; we survey existing designs from this viewpoint. We also suggest combining information and ethical considerations by taking a suitable weighted mean of two corresponding standardized criteria, with weights that depend on the actual treatment effects. This compound criterion leads to a locally optimal allocation that can be targeted by some response-adaptive randomization rule. The paper mainly deals with the case of two treatments, but the suggested methodology is shown to extend to more than two.

Some key words: Admissibility; Balance; Neyman's allocation; Optimality; Standardized compound criterion.

1. INTRODUCTION

The design of a clinical trial is affected both by the expected information gain from the experiment and by care considerations for the subjects participating in it; these demands often come into conflict. This paper deals with comparative clinical trials with binary outcomes and faces the problem of allocating patients to treatments so as to achieve a good trade-off between statistical efficiency and risk to the subjects' health. Recently, there has been a growing statistical interest in sequential procedures for treatment comparison that at each stage use the available information with the ethical aim of skewing allocations towards the best treatment. Starting from the so-called play-the-winner (Zelen, 1969), other examples are the randomized play-the-winner (Wei & Durham, 1978) and the drop-the-loser (Ivanova, 2003) rules, which use response-adaptive randomization (Hu & Rosenberger, 2006).

At the start of this paper, we discuss the following issue: some treatment assignments, depending on the true state-of-nature, may turn out to be worse than others with respect to some inferential criterion, say D-optimality or trace-optimality, and in terms of the expected percentage of patients successfully treated. We say that a treatment allocation is admissible if it is free of this shortcoming and we give a necessary and sufficient condition for this type of admissibility. It is well known that some treatment allocations do not always meet this requirement: in this framework, we discuss classical allocations as well as recent additions to the literature by Rosenberger et al. (2001) and by Geraldes et al. (2006).

In this paper, the question of mediating between informativity and possible ethical skewing of the allocations is approached via the optimization of a compound criterion, obtained by taking a weighted average of a common design optimality measure and a measure of the subjects' risk. Since it is reasonable to suppose that the more the effects of the treatments differ from one another, the more important it is to increase the chances of assigning the best treatment, we propose to let the relative weights depend on the success probabilities of the treatments. The allocation obtained by solving the compound optimization problem is a function of the unknown parameters and can be targeted by an appropriate response-adaptive procedure, namely one in which the available outcomes are used at each stage for redirecting the assignments in order to converge to it.

The purpose of the large majority of Phase III trials is to compare just two distinct treatments, so the present paper deals mainly with exactly two treatments. In § 7, the results are generalized to the case of several treatments, which is also of interest.

Our approach complements existing ones. For binary trials, Geraldes et al. (2006) have proposed the doubly adaptive weighted difference design, a procedure in which at each step the allocation is the result of a compromise between the demand for balance and the wish to assign the better treatment more often. Bandyopadhyay et al. (2007) suggest a two-stage design adjusting for prognostic factors, where stage 1 is balanced and at stage 2 randomization is skewed towards the treatment that performs better at the end of the first stage. In the randomized Pólya urn (Durham et al., 1998) and the randomly reinforced urn design (Muliere et al., 2006; May & Flournoy, 2009), the authors adopt an even stronger ethical viewpoint and introduce sequential procedures under which the proportion of patients allocated to the better treatment tends to 1 asymptotically. Assuming a different approach, Rosenberger et al. (2001) and Tymofyeyev et al. (2007) cast ethical and inferential demands into a constrained optimization problem, leading to interesting target allocations.

2. BINOMIAL MODEL AND TARGET ALLOCATIONS

We start with some notation. Let $K \geqslant 2$ independent populations of Bernoulli response data represent patient outcomes under treatments $T_1, \ldots, T_K$. Let $p = (p_1, \ldots, p_K)^t$ and $q = (q_1, \ldots, q_K)^t$, with $q_j = 1 - p_j$ for $j = 1, \ldots, K$, be the vectors of success/failure probabilities. Let Y_{ij} be the response of patient i $(i \geqslant 1)$ to treatment T_j:

$$E(Y_{ij}) = \Pr(Y_{ij} = 1) = p_j. \tag{1}$$

After n subjects are recruited into the trial, out of which n_j are assigned to T_j, let $\pi = (\pi_1, \ldots, \pi_K)^t$ be the vector of allocation proportions $\pi_j = n_j/n$ to treatment T_j, with $\sum_{j=1}^K \pi_j = 1$. Let $\hat{p}_j$ be the maximum likelihood estimator of p_j $(j = 1, \ldots, K)$; conditional on the design, the variance-covariance matrix of $(\hat{p}_1, \ldots, \hat{p}_K)$ is

$$\mathrm{var}(\hat{p}_1, \ldots, \hat{p}_K) = \mathrm{diag}\left(\frac{p_1 q_1}{n_1}, \ldots, \frac{p_K q_K}{n_K}\right) = n^{-1}\mathrm{diag}\left(\frac{p_1 q_1}{\pi_1}, \ldots, \frac{p_K q_K}{\pi_K}\right)$$

and the average per observation Fisher information matrix is

$$M(\pi; p) = \mathrm{diag}\left(\frac{\pi_j}{p_j q_j}\right)_{(j=1,\ldots,K)}. \tag{2}$$

In optimal design theory, for fixed sample size n, the design problem consists in minimizing the loss of information as appropriately quantified by a suitable real function Φ_I, defined on the set of nonnegative definite variance-covariance matrices, convex in its argument and increasing with respect to the Loewner ordering (Pukelsheim, 2006). For model (1) this means finding the vector

$$\pi_I^* = (\pi_{I1}^*, \ldots, \pi_{IK}^*)^t = \arg\min_{\pi} \Phi_I\{M^{-1}(\pi; p)\}. \tag{3}$$

For simplicity, we write $\Psi_I(\pi) = \Phi_I\{M^{-1}(\pi; p)\}$, neglecting the dependence on p, and on the boundary we set $\Psi_I(\pi) = \infty$, stressing the penalty for extreme allocations. Observe that $\Psi_I(\pi)$ is convex in the vector π and if this convexity is strict, π_I^* satisfying (3) is unique. Clearly, convexity in the vector π implies convexity in each component. The best known optimality criteria are D-optimality, i.e., $\Psi_D(\pi) = \det\{M^{-1}(\pi; p)\}$, and trace-optimality $\Psi_{tr}(\pi) = \mathrm{tr}\{M^{-1}(\pi; p)\}$; other criteria when one is especially interested in the contrasts $p_1 - p_j$ $(j = 2, \ldots, K)$ are

$$\Psi_{D_A}(\pi) = \det\{A^t M^{-1}(\pi; p) A\}, \quad \Psi_{tr_A}(\pi) = \mathrm{tr}\{A^t M^{-1}(\pi; p) A\},$$

where $A^t = [1_{K-1} : -I_{K-1}]$, 1_{K-1} is the $(K-1)$-dimensional vector of ones and I_{K-1} denotes the $(K-1)$-dimensional identity matrix. Yet another interesting information criterion, to be maximized, is the power of the asymptotic chi-square test on the homogeneity of the contrasts $p_1 - p_j$, which leads to the maximization of the noncentrality parameter of the test (Tymofyeyev et al., 2007). This is a concave function of π, which can be easily translated into our convex setting.

A very popular allocation scheme balances the sizes of the treatment groups:

$$\pi_B^* = \left(\frac{1}{K}, \ldots, \frac{1}{K}\right), \tag{4}$$

which corresponds to minimizing the D-optimality criterion.

In general, one must find a compromise between the need to make the most efficient use of patient information and the need to achieve an overall benefit for the entire sample of patients involved in the trial. From an ethical viewpoint, a natural demand would be to reduce the expected number of failures. For any given sample size, this involves minimizing

$$\Psi_E(\pi) = \pi_1 q_1 + \cdots + \pi_K q_K, \tag{5}$$

where again the dependence of Ψ_E on p is neglected for simplicity. The optimal allocation π_E^* assigns all the subjects to the best treatment; when such treatment is not unique, π_E^* may be replaced by any convex combination of all the assignments to the best treatments. Clearly, when $p_1 = \cdots = p_K$, the treatment arms collapse as there is no longer a best or a worst treatment, so (5) is no longer useful.

In the rest of the paper, we refer to π as a target, meaning that it is either an actual proportion of assignments to $T_1, \ldots, T_K$ out of a total of n trials, or a vector of limit allocation proportions.

When $K = 2$, the aim of the experiment will be either to estimate the effects of T_1 and T_2 jointly or, more commonly, to compare them, namely estimate or test their difference. The information matrix (2) becomes

$$M(\pi; p) = \mathrm{diag}\left(\frac{\pi_1}{p_1 q_1}, \frac{1 - \pi_1}{p_2 q_2}\right)$$

and thus Ψ_I can be thought of as a convex function of just π_1, say $\Psi_I(\pi_1)$. The D-optimality criterion is

$$\Psi_D(\pi_1) = \frac{p_1 q_1 p_2 q_2}{\pi_1(1 - \pi_1)}$$

and the trace criterion is

$$\Psi_{tr}(\pi_1) = \frac{p_1 q_1}{\pi_1} + \frac{p_2 q_2}{1-\pi_1}, \tag{6}$$

which in this setting coincides with D_A- and tr_A-optimality. As is well known, the optimum of criterion (6) is achieved by the Neyman allocation

$$\pi_N^* = \left\{ \frac{\surd(p_1 q_1)}{\surd(p_1 q_1) + \surd(p_2 q_2)}, \frac{\surd(p_2 q_2)}{\surd(p_1 q_1) + \surd(p_2 q_2)} \right\}, \tag{7}$$

which also maximizes the power of Wald's test of equality of treatment effects (Hu & Rosenberger, 2003). Unlike (4), allocation (7) depends on the unknown parameters.

The expected proportion of failures is

$$\Psi_E(\pi) = \pi_1 q_1 + (1-\pi_1) q_2. \tag{8}$$

Let $q_{\min} = \min(q_1, q_2)$ and $q_{\max} = \max(q_1, q_2)$, note that $q_{\min} \leqslant \Psi_E(\pi) \leqslant q_{\max}$ for all $p_1, p_2, \pi_1 \in [0; 1]$. If $p_1 \neq p_2$, $\Psi_E(\pi)$ can be standardized to lie in [0, 1]:

$$\begin{aligned} \frac{\Psi_E(\pi) - q_{\min}}{q_{\max} - q_{\min}} &= \frac{\pi_1 q_1 + (1-\pi_1) q_2 - q_{\min}}{|q_1 - q_2|} \\ &= \frac{1}{2} + \left(\frac{1}{2} - \pi_1\right) \operatorname{sgn}(p_1 - p_2), \end{aligned} \tag{9}$$

where $\operatorname{sgn}(x)$ represents the sign of x. This standardized version (9) represents the proportion of patients who receive the worse treatment. It is obvious that the optimal allocation with respect to (8) and (9) assigns all the patients to the better treatment; it can be written formally as

$$\pi_E^* = \left(\frac{1}{2} + \frac{1}{2}\operatorname{sgn}(p_1 - p_2), \frac{1}{2} - \frac{1}{2}\operatorname{sgn}(p_1 - p_2) \right). \tag{10}$$

3. NONADMISSIBLE TARGETS

Assume the criteria of interest are Ψ_E in (8) and, at the same time, one particular inferential criterion Ψ_I. While we are dealing with just two treatments, we indicate the proportion of allocation to T_1 simply by π instead of π_1 and write $\Psi_E(\pi)$ and $\Psi_I(\pi)$ accordingly. We say that a given allocation is nonadmissible if there is another one which for some values of the unknown parameters is no less informative with respect to the same Ψ_I, while at the same time being more favourable to the better treatment.

DEFINITION 1. *A target $\tilde{\pi}$ is said to be nonadmissible with respect to the chosen pair (Ψ_E, Ψ_I) if there exists another allocation π' such that, for some values of the unknown parameters (p_1, p_2) with $p_1 \neq p_2$,*

$$\Psi_E(\pi') \leqslant \Psi_E(\tilde{\pi}), \quad \textit{and} \quad \Psi_I(\pi') \leqslant \Psi_I(\tilde{\pi}),$$

where at least one inequality is strict.

The design of a clinical trial with a nonadmissible target may induce losses on both ethical and inferential grounds. We now give a necessary and sufficient condition for a target to be admissible.

THEOREM 1. *A target $\tilde{\pi}$ is admissible with respect to (Ψ_E, Ψ_I), with Ψ_I strictly convex, if and only if*

$$(\tilde{\pi} - \pi_I^*) \quad (p_1 - p_2) \geqslant 0, \qquad (p_1, p_2), \tag{11}$$

where π_I^* is the target that minimizes Ψ_I.

Proof. Assume $\tilde{\pi}$ is (Ψ_E, Ψ_I)-admissible and $(\tilde{\pi} - \pi_I^*)(p_1 - p_2) < 0$ for at least one pair $(\tilde{p}_1, \tilde{p}_2)$. Let $\tilde{p}_1 > \tilde{p}_2$, so that $\tilde{\pi} < \pi_I^*$. Then $\Psi_I(\tilde{\pi}) > \Psi_I(\pi_I^*)$, by the strict convexity of Ψ_I, and $\Psi_E(\tilde{\pi}) > \Psi_E(\pi_I^*)$ from the linearity of Ψ_E in π; therefore π_I^* dominates $\tilde{\pi}$, thus contradicting admissibility. Similarly if $\tilde{p}_1 < \tilde{p}_2$. Conversely, assume $(\tilde{\pi} - \pi_I^*)(p_1 - p_2) \geqslant 0$ for all (p_1, p_2). Since admissibility of π_I^* with respect to the pair (Ψ_E, Ψ_I) is trivial, let $\tilde{\pi} \neq \pi_I^*$. Then either $p_1 > p_2$ and $\tilde{\pi} > \pi_I^*$ or $p_1 < p_2$ and $\tilde{\pi} < \pi_I^*$. If $p_1 > p_2$ and $\tilde{\pi} > \pi_I^*$, observe that for all $\pi < \tilde{\pi}$, $\Psi_E(\pi) > \Psi_E(\tilde{\pi})$, and for all $\pi > \tilde{\pi}$, $\Psi_E(\pi) < \Psi_E(\tilde{\pi})$, but by the convexity of $\Psi_I(\tilde{\pi})$, it follows that $\Psi_I \geqslant \Psi_I(\tilde{\pi}) > \Psi_I(\pi_I^*)$, and thus Definition 1 is never contradicted. Similarly when $p_1 < p_2$ and $\tilde{\pi} < \pi_I^*$. Therefore $\tilde{\pi}$ is admissible. □

As a consequence, an allocation $\tilde{\pi}$ is admissible with respect to the pair (Ψ_E, Ψ_D), D-admissible for short, if and only if it assigns the majority of subjects to the better treatment, i.e., for any (p_1, p_2),

$$\left(\tilde{\pi} - \frac{1}{2}\right)(p_1 - p_2) \geqslant 0. \tag{12}$$

The balanced allocation π_B^* and the target π_{PW}^* are clearly D-admissible. Similarly, an allocation $\tilde{\pi}$ is admissible with respect to (Ψ_E, Ψ_{tr}), or trace-admissible for short, if and only if, for any (p_1, p_2),

$$\left\{\tilde{\pi} - \frac{\surd(p_1 q_1)}{\surd(p_1 q_1) + \surd(p_2 q_2)}\right\}(p_1 - p_2) \geqslant 0. \tag{13}$$

Example 1. The target

$$\pi_{\text{PW}}^* = \frac{q_2}{q_1 + q_2}, \tag{14}$$

which is the limit allocation of the play-the-winner sequential experiment (Zelen, 1969) and of several others, e.g. (Ivanova, 2003), is known not to be trace-admissible, since it is dominated by π_N^* in (7) when $p_1 + p_2 < 1/2$. To see this, without loss of generality let $p_1 > p_2$. From (13) it is enough to show that $\pi_{\text{PW}}^* < \pi_N^*$ for some values of p_1 and p_2. This is equivalent to showing $p_2 q_2^3 < p_1 q_1^3$, namely $q_1^2 + q_1 q_2 + q_2^2 < (q_1 + q_2)(q_1^2 + q_2^2)$, i.e., $q_1 q_2 < (q_1^2 + q_2^2)(q_1 - p_2)$, which is satisfied when $p_1 + p_2 < 1/2$, since in this case

$$\frac{q_1 q_2}{q_1^2 + q_2^2} = \left\{\frac{q_1}{q_2} + \left(\frac{q_1}{q_2}\right)^{-1}\right\}^{-1} \leqslant \frac{1}{2} < q_1 - p_2.$$

Example 2. The balanced target π_B^* is not trace-admissible, since it is dominated by π_N^* when $p_1 + p_2 < 1$. It is easy to check that $\pi_N^* > 1/2$, namely $p_2 q_2 < p_1 q_1$, if and only if $p_1 + p_2 < 1$.

Example 3. The Neyman allocation π_N^* in (7) is not D-admissible, since it is dominated by π_B^* in (4) when $p_1 + p_2 > 1$. To see this, without loss of generality let $p_1 > p_2$. If $p_1 + p_2 > 1$, it is straightforward to show that $\pi_N^* < 1/2$, since in this case $p_2 q_2 > p_1 q_1$. Thus, Neyman's target may assign more subjects to the inferior treatment (Rosenberger et al., 2001; Hu et al., 2006).

As another example, we show that the target allocation of the doubly adaptive weighted difference design (Geraldes et al., 2006) is not trace-admissible. This design can be briefly described as follows. Let $g(\cdot)$ and $h(\cdot)$ be two real functions, both standing for the probability of assigning treatment T_1 at the next step; $g(\cdot)$ is dictated by ethics and is an increasing function of the difference $p_1 - p_2$, with $g(0) = 1/2$; $h(\cdot)$ is dictated by balance and is a decreasing function of the relative imbalance, with $h(0) = 1/2$; $g(\cdot)$ and $h(\cdot)$ are combined by weights $\lambda \in [0; 1)$ and $1 - \lambda$. The target allocation to T_1 is the unique solution π^*_{dawd} in $(0, 1)$ of the equation:

$$L(\pi) = \pi - \lambda g(p_1 - p_2) + (1 - \lambda)h(2\pi - 1) = 0.$$

Example 4. The target π^*_{dawd} is (Ψ_E, Ψ_D)-admissible, but nonadmissible with respect to (Ψ_E, Ψ_{tr}). To see this, without loss of generality let $p_1 > p_2$. D-admissibility means $\pi^*_{\text{dawd}} > 1/2$. Suppose now that $\pi^*_{\text{dawd}} < 1/2$, then $h(2\pi^*_{\text{dawd}} - 1) > 1/2$; furthermore, $g(p_1 - p_2) > 1/2$ since $p_1 > p_2$. Thus $\lambda g(p_1 - p_2) + (1 - \lambda)h(2\pi^*_{\text{dawd}} - 1) > 1/2$ for any $\lambda \in [0; 1)$, namely a contradiction because $\lambda g(p_1 - p_2) + (1 - \lambda)h(2\pi^*_{\text{dawd}} - 1) = \pi^*_{\text{dawd}}$. Assume now that (Ψ_E, Ψ_{tr})-admissibility holds, namely $\pi^*_{\text{dawd}} \geqslant \pi^*_N$ for all $p_1 > p_2$. Since $L(\cdot)$ is monotonically increasing in π, then $L(\pi^*_N) \leqslant L(\pi^*_{\text{dawd}}) = 0$, namely

$$\pi^*_N \leqslant \lambda g(p_1 - p_2) + (1 - \lambda)h(2\pi^*_N - 1) \quad \text{for all } p_1 > p_2. \tag{15}$$

This condition is clearly satisfied when $p_1 + p_2 \geqslant 1$, since in this case $\pi^*_N \leqslant 1/2$. But if $p_1 + p_2 < 1$, then $h(2\pi^*_N - 1) < 1/2$, so that $\lambda g(p_1 - p_2) + (1 - \lambda)h(2\pi^*_N - 1) < (\lambda + 1)/2$. For any given $\lambda < 1$ it is always possible to find values of p_1 and p_2, with $p_1 > p_2$ and $p_1 + p_2 < 1$, for which $\pi^*_N > (\lambda + 1)/2$, thus contradicting (15).

After these examples, it might seem that no target will ever be admissible with respect to both the D- and the trace criterion, but this is not so. For instance, $\tilde{\pi} = p_1/(p_1 + p_2)$ assigns more subjects to the superior treatment, D-admissibility, and is also trace-admissible, since $\{p_1/(p_1 + p_2) - \pi^*_N\}(p_1 - p_2) \geqslant 0$ for any (p_1, p_2). Another example is the target allocation proposed by Rosenberger et al. (2001):

$$\pi^*_R = \frac{\sqrt{p_1}}{\sqrt{p_1} + \sqrt{p_2}}. \tag{16}$$

A general way of obtaining admissible targets is given in § 4.

4. AN ADAPTIVE COMPOUND CRITERION

When compromising between Ψ_E and Ψ_I, several approaches are possible.

Method 1. *Trade off the experiments.* Find two sequential designs Δ_E and Δ_I with optimal targets derived from Ψ_E and Ψ_I, respectively, and take a mixture of these designs with weights λ and $1 - \lambda$, namely at each step assign treatment T_1 with the probability rule given by design Δ_E chosen with probability λ, or by Δ_I chosen with probability $1 - \lambda$.

Method 2. *Trade off the targets.* Take a mixture of the two optimal allocations with respect to criteria Ψ_E and Ψ_I and find a sequential design converging to this target. Method 1 described above can be seen as a special case of Method 2.

Method 3. *Trade off the criteria.* This can be done in two different ways:

Method 3·1. Optimize either criterion under a constraint on the other one. Given the convexity of Ψ_I and the linearity of Ψ_E in π, it is easy to see the equivalence of either way.

Method 3·2. Derive the optimal target with respect to a compromise of the criteria Ψ_E and Ψ_I, and find a sequential design converging to this target.

As shown by several authors (Cook & Wong, 1994; Clyde & Chaloner, 1996), a constrained optimal design obtained by Method 3·1 can be seen as the solution of a combined optimization problem, namely it can be derived by a suitable application of Method 3·2. An example of Method 3·1 is the target π_R^* in (16), proposed by Rosenberger et al. (2001) and further generalized to the case of several treatments by Tymofyeyev et al. (2007), which aims at minimizing the expected number of failures for a given variance of the estimated difference $p_1 - p_2$, or a fixed power of the test of equality of treatments effects. The allocation (16) is also obtained by minimizing the geometric mean of criteria (5) and (6) with equal weights.

Method 3·2 requires a compromise function to combine the criteria Ψ_E and Ψ_I, for instance a weighted arithmetic, geometric or harmonic mean. The simplest idea is to take a weighted average of Ψ_E and Ψ_I, but Ψ_E and Ψ_I are not homogeneous measures and need to be standardized to a comparable scale. Put $\Psi_E^* = \Psi_E(\pi_E^*)$ and $\Psi_I^* = \Psi_I(\pi_I^*)$, where π_I^* (3) is the optimal target for inference and π_E^* (10) for ethics. A standardized compound criterion is

$$\Psi_\omega(\pi) = \omega \left\{ \frac{\Psi_E(\pi)}{\Psi_E^*} \right\} + (1-\omega) \left\{ \frac{\Psi_I(\pi)}{\Psi_I^*} \right\}. \tag{17}$$

This can be seen (Dette, 1997; Biedermann et al., 2007) as the reciprocal of the weighted harmonic mean of the efficiencies associated with criteria Ψ_E and Ψ_I and is a convex function of π, since it is a linear combination of a linear and a convex function. From now on our aim is to minimize (17).

We shall choose the ethical weight ω to be a suitable continuous function of the unknown parameters, i.e., $\omega = \omega(p_1, p_2) : (0; 1)^2 \to [0; 1)$. For instance, if the treatment effects differ in a substantial way, we may want to increase the importance of Ψ_E and, on the other hand, if p_1 and p_2 are close it is harder to discriminate between the two treatments and less attention may be paid to the ethical skewing. Alternatively, the relative importance of Ψ_E and Ψ_I may vary on the basis of how near p_1 and p_2 are to 0, 1 or 0·5. Observe that the degenerate cases $p_1, p_2 = 0, 1$, i.e., either or both treatments leading to success or failure with probability 1, are excluded, since they are of no practical interest. Moreover, we assume that the ethical impact should not completely overcome the inferential goal, so $\omega < 1$.

In conclusion, the target $\pi_\omega^* = \pi_\omega^*(p_1, p_2)$, obtained by minimizing the compound criterion Ψ_ω in (17), will depend on the chosen pair (Ψ_E, Ψ_I), the weight ω and the unknown parameters.

THEOREM 2. *If Ψ_I is continuously differentiable in π, the optimal target $\pi_\omega^*(p_1, p_2)$ minimizing the compound criterion Ψ_ω in* (17) *is the unique solution in* $(0; 1)$ *of the equation*

$$\frac{\partial \Psi_I(\pi)}{\partial \pi} = \frac{\omega(p_1, p_2)}{1 - \omega(p_1, p_2)} \left(\frac{p_1 - p_2}{q_{\min}} \right) \Psi_I^*. \tag{18}$$

This target $\pi_\omega^*(p_1, p_2)$ is admissible with respect to the pair (Ψ_E, Ψ_I).

Proof. Since the compound criterion (17) is strictly convex in π, there exists a unique $\pi_\omega^* \in [0; 1]$ achieving the minimum of Ψ_ω. Furthermore, from the properties of the information criterion Ψ_I, π_ω^* lies in $(0; 1)$ and differentiation shows that it satisfies (18). Admissibility with respect to (Ψ_E, Ψ_I) is straightforward: if there were $\pi' \in [0; 1]$ such that for some values of

(p_1, p_2), we have $\Psi_E(\pi') \leqslant \Psi_E(\pi^*_\omega)$ and $\Psi_I(\pi') \leqslant \Psi_I(\pi^*_\omega)$, then

$$\omega \left\{ \frac{\Psi_E(\pi')}{\Psi^*_E} \right\} + (1-\omega) \left\{ \frac{\Psi_I(\pi')}{\Psi^*_I} \right\} \leqslant \omega \left\{ \frac{\Psi_E(\pi^*_\omega)}{\Psi^*_E} \right\} + (1-\omega) \left\{ \frac{\Psi_I(\pi^*_\omega)}{\Psi^*_I} \right\},$$

which is impossible, since π^*_ω is the unique minimum of Ψ_ω. □

The following result shows the promised link between admissibility and compound criteria of the type above.

THEOREM 3. *For all continuously differentiable Ψ_I, a design is (Ψ_E, Ψ_I)-admissible if and only if it is optimal with respect to a compound criterion Ψ_ω, obtained from Ψ_E and Ψ_I as defined in (17), for a suitable choice of the weight function ω.*

Proof. Let $\tilde{\pi}$ satisfy (11), so that when $p_1 \geqslant p_2$ then $\tilde{\pi} \geqslant \pi^*_I$, and vice versa. We must show that $\tilde{\pi}$ is a solution of

$$\frac{\partial \Psi_I(\pi)}{\partial \pi} = \frac{\omega(p_1, p_2)}{1-\omega(p_1, p_2)} \left(\frac{p_1 - p_2}{q_{\min}} \right) \Psi^*_I.$$

Since the left-hand side is an increasing function of π, taking

$$\frac{\omega(p_1, p_2)}{1-\omega(p_1, p_2)} = \left. \frac{\partial \Psi_I(\pi)}{\partial \pi} \right|_{\tilde{\pi}} \Big/ \left\{ \left(\frac{p_1 - p_2}{q_{\min}} \right) \Psi^*_I \right\}$$

solves the problem. □

Example 5. As shown in § 3, the target $\tilde{\pi} = p_1/(p_1 + p_2)$ is D-admissible. We check that it is also Ψ_ω-optimal when Ψ_I is D-optimality. It is straightforward to see that $\tilde{\pi}$ minimizes (17) when the weights are chosen in the following ratio

$$\frac{\omega(p_1, p_2)}{1-\omega(p_1, p_2)} = \frac{(p_1 + p_2)^3 q_{\min}}{4 p_1^2 p_2^2},$$

since $\pi^*_I = 1/2$, $\Psi^*_I = 4 p_1 q_1 p_2 q_2$ and $\partial \Psi_I(\pi)/\partial \pi = p_1 q_1 p_2 q_2 (2\pi - 1)[\pi(1-\pi)]^{-2}$.

Suitable choices of the weight function $\omega(\cdot, \cdot)$ allow us to derive admissible targets with respect to both trace- and D-optimality, as the following result shows.

THEOREM 4. *Let our compound criterion be as in (17) with Ψ_I =trace-optimality. If the weight function is chosen so that when $x + y > 1$,*

$$\omega(x, y) \geqslant \frac{1}{2}, \tag{19}$$

then the optimal target π^*_ω assigns the majority of subjects to the better treatment.

Proof. Without loss of generality let $p_1 > p_2$. Note that $\Psi_{tr}(\pi^*_N) = \{\surd(p_1 q_1) + \surd(p_2 q_2)\}^2$. From (18) the optimal target π^*_ω satisfies the following condition:

$$\frac{p_2 q_2 (\pi^*_\omega)^2 - p_1 q_1 (1-\pi^*_\omega)^2}{\{\pi^*_\omega (1-\pi^*_\omega)\}^2} = \frac{\omega}{1-\omega} \left(\frac{p_1 - p_2}{q_1} \right) \{\surd(p_1 q_1) + \surd(p_2 q_2)\}^2.$$

From (12) it is sufficient to check that when (19) holds $\pi^*_\omega \geqslant 1/2$ for all (p_1, p_2). Since Ψ_{tr} is convex, then $\partial \Psi_{tr}(\pi)/\partial \pi$ is increasing in π, so it is equivalent to

$$\left. \frac{\partial \Psi_{tr}(\pi)}{\partial \pi} \right|_{\pi = \frac{1}{2}} = 4(p_2 q_2 - p_1 q_1) \leqslant \frac{\omega}{1-\omega} \left(\frac{p_1 - p_2}{q_1} \right) \{\surd(p_1 q_1) + \surd(p_2 q_2)\}^2.$$

If $p_1 + p_2 \leqslant 1$ then $p_2q_2 \leqslant p_1q_1$, so that the above condition is satisfied for any choice of ω. If $p_1 + p_2 > 1$ then $p_1 > q_2$ and $p_1q_1 - p_2q_2 < 0 < 2\surd(p_1q_1)\{\surd(p_2q_2) - \surd(p_1q_1)\} + 4q_1q_2$. It follows that $4q_1(p_1 - q_2) < \{\surd(p_1q_1) + \surd(p_2q_2)\}^2$ and hence $4q_1(p_2q_2 - p_1q_1)/(p_1 - p_2) = 4q_1(p_1 - q_2) < \{\surd(p_1q_1) + \surd(p_2q_2)\}^2$. Thus, if $\omega(p_1, p_2)$ satisfies condition (19) then

$$4q_1(p_2q_2 - p_1q_1)(p_1 - p_2)^{-1}\{\surd(p_1q_1) + \surd(p_2q_2)\}^{-2} < 1 \leqslant \omega(p_1, p_2)\{1 - \omega(p_1, p_2)\}^{-1},$$

which proves the assertion. □

The optimal target allocation π^*_ω will clearly depend on the unknown parameters. To converge to it, several sequential procedures may be employed. From Theorem 2 and (18), $\pi^*_\omega(p_1, p_2)$ is a continuous function of p_1 and p_2, since $\partial\Psi_I(\pi)/\partial\pi$ is continuous and strictly increasing in π, see for instance the proof of Lemma 7 in Geraldes et al. (2006). Thus the sequential maximum likelihood design can be implemented to get $\lim_{n\to\infty} \pi(n) = \pi^*_\omega(p_1, p_2)$ almost surely, where $\pi(n)$ is the proportion of allocations to T_1 after n assignments (Baldi Antognini & Giovagnoli, 2005; Hu & Rosenberger, 2006).

5. The choice of the weights in the compound criterion

The choice of the best weight function in a given applied context is open to discussion, but it should satisfy the following requirements.

Condition 1. Let $\omega(x, y) = \omega(y, x)$.

Condition 2. As $|x - y|$ increases, $\omega(x, y)$ is nondecreasing.

Condition 1 guarantees that T_1 and T_2 are treated symmetrically, whereas Condition 2 implies that the ethical impact will be more crucial the more different the effects of the two treatments are.

By way of example, the weight ω can be chosen to be a power function of the absolute difference $|p_1 - p_2|$, i.e., $\omega(p_1, p_2) = (p_1 - p_2)^{2m}$ with $m \geqslant 1$. Furthermore, often a very small difference between p_1 and p_2, say up to a value δ, is of no practical interest. Then we may assume $\omega(p_1, p_2) = 0$ if $|p_1 - p_2| \leqslant \delta$ and also $\omega(p_1, p_2) \to 0$ as $|p_1 - p_2| \to \delta^+$, so that $\Psi_\omega \to \Psi_I$, as well as $\omega(p_1, p_2) \to 1$ as $|p_1 - p_2| \to 1$, i.e. $\Psi_\omega \to \Psi_E$. Furthermore, by Theorem 4, the weight function can be chosen so as to derive target allocations that are admissible with respect to both trace and D-optimality; thus, the above proposals can be modified according to (19), for instance

$$\omega(p_1, p_2) = \frac{(p_1 - p_2)^{2m} + 1}{2}, \quad m \geqslant 1. \tag{20}$$

As well as a function of the absolute difference of the probabilities, a suitable function of their ratio would serve the same purpose and Condition 2 may be modified accordingly.

6. Comparison of different target allocations

Designs competing with our own have different target allocations. Let us compare different targets: when nonadmissible, for some values of the unknown parameters there will be losses on inferential grounds and also in terms of a higher percentage of expected failures and of patients who receive the worse treatment. The magnitude of the losses may be small or large, depending on the true values of p_1 and p_2. For instance, with the trace criterion Ψ_{tr}, if $p_1 + p_2 \leqslant 1/2$,

Table 1. *Comparisons of targets with respect to the trace-optimality as p_1 and p_2 varies. The values of Ψ_{tr} are shown in brackets*

p_1	p_2	π^*_N	π^*_{PW}	π^*_R	π^*_{dawd}	$\pi^*_{\omega=1/2}$	$\pi^*_{\omega_p}$
0·1	0·05	0·579	0·514	0·586	0·508	0·586	0·586
		(0·268)	(0·273)	(0·268)	(0·274)	(0·268)	(0·268)
0·2	0·05	0·647	0·543	0·667	0·525	0·668	0·669
		(0·382)	(0·399)	(0·383)	(0·405)	(0·383)	(0·383)
0·2	0·1	0·571	0·529	0·586	0·517	0·586	0·587
		(0·490)	(0·493)	(0·490)	(0·496)	(0·490)	(0·490)
0·4	0·05	0·692	0·613	0·739	0·558	0·744	0·755
		(0·501)	(0·514)	(0·507)	(0·537)	(0·508)	(0·512)
0·4	0·2	0·551	0·571	0·586	0·533	0·590	0·593
		(0·792)	(0·793)	(0·796)	(0·793)	(0·797)	(0·798)
0·4	0·35	0·507	0·520	0·517	0·508	0·517	0·517
		(0·934)	(0·936)	(0·935)	(0·935)	(0·935)	(0·935)
0·65	0·4	0·493	0·632	0·560	0·542	0·578	0·588
		(0·935)	(1·012)	(0·952)	(0·944)	(0·962)	(0·970)
0·65	0·60	0·493	0·533	0·510	0·508	0·511	0·511
		(0·935)	(0·941)	(0·936)	(0·936)	(0·936)	(0·936)
0·95	0·65	0·314	0·875	0·547	0·550	0·724	0·747
		(0·483)	(1·874)	(0·589)	(0·592)	(0·890)	(0·963)
0·95	0·85	0·379	0·750	0·514	0·517	0·599	0·602
		(0·331)	(0·573)	(0·355)	(0·356)	(0·397)	(0·399)

the balanced target π^*_B is dominated by π^*_{PW}. The percentage increase of the estimated treatment difference variance, conditional on the balanced allocation, with respect to the target π^*_{PW}, is

$$\frac{\text{var}_{\pi^*_B}(\hat{p}_1-\hat{p}_2)}{\text{var}_{\pi^*_{\text{PW}}}(\hat{p}_1-\hat{p}_2)}-1=\left(\frac{p_1-p_2}{q_1+q_2}\right)\left(\frac{p_1q_1^2-p_2q_2^2}{p_1q_1^2+p_2q_2^2}\right),$$

while the increase in the proportion of patients who receive the inferior treatment is $|q_1-q_2|/2(q_1+q_2)$. When $p_1=0{\cdot}2$ and $p_2=0{\cdot}1$, the variance increases by 1·3% and 2·9% more subjects are allocated to the worse treatment. Whereas, if $p_1=0{\cdot}4$ and $p_2=0{\cdot}05$ the variance increases by 11·9% and 11·3% more patients receive the worse treatment, so the expected proportion of failures increases by 19·1%.

We now compare π^*_ω with some of the targets in the literature, namely π^*_{PW} (14), Neyman's allocation π^*_N (7), π^*_R (16) and also π^*_{dawd}, the target of the doubly adaptive weighted difference design obtained by setting $g(x)=(1+x)/2$, $h(x)=(1-x)/2$ and $\lambda=1/2$. Let the inferential criterion be trace-optimality and let the allocation π^*_ω be obtained with two different choices of the weight function, namely constant weight $\omega=1/2$, denoted by $\pi^*_{\omega=1/2}$, and the power function (20) with $m=1$ ($\pi^*_{\omega_p}$). Table 1 shows the values of the target allocations as (p_1,p_2) varies.

Both $\pi^*_{\omega=1/2}$ and $\pi^*_{\omega_p}$ assign the majority of subjects to the better treatment by (19) of Theorem 4, since they are both trace- and D-admissible; $\pi^*_{\omega_p}$ is more skewed than $\pi^*_{\omega=1/2}$, as the weight function (20) is greater than $1/2$ for all (p_1,p_2). As already pointed out, Neyman's allocation π^*_N shows a critical behaviour when $p_1+p_2>1$. For instance, if $p_1=0{\cdot}65$ and $p_2=0{\cdot}4$, the target $\pi^*_{\omega_p}$ brings a gain over π^*_N of about 9·5% patients treated by the better treatment, against an increase in the variance of the estimated p_1-p_2 by a mere 3·7%. Besides, if the values of p_1 and p_2 are small, π^*_{PW} and π^*_{dawd} do not behave well in terms of ethical skew and trace criterion and could be dominated by other targets. For instance, consider $\pi^*_{\omega=1/2}$. When $p_1=0{\cdot}2$ and $p_2=0{\cdot}05$, $\pi^*_{\omega=1/2}$ brings about a gain of 1·8% over π^*_{PW} in terms of variance and also a gain of

12·5% patients receiving the better treatment; $\pi^*_{\omega=1/2}$ brings about a gain of 5·4% over π^*_{dawd} in terms of variance and also a gain of 14·3% patients receiving the better treatment. Also note that $\pi^*_{\omega=1/2}$ induces a loss of 0·2% in terms of variance but a gain of 2·1% in terms of ethical skewing, in comparison with π^*_N. In general, the values of π^*_ω and π^*_R are close for both our choices of the weight function, especially if the treatment probabilities are close. The main differences arise when p_1 and p_2 are large: in this case π^*_ω tends to favour the better treatment more than π^*_R.

7. Extension to several treatments

Although the comparison between two competing treatments represents the most common situation in the practice of Phase III clinical trials, and also in the corresponding literature, in this section we generalize the suggested methodology to the case of several treatments.

Admissibility of a target allocation is defined as in Definition 1 and is characterized by the following straightforward extension of Theorem 1.

Theorem 5. *Given a strictly convex optimality criterion Ψ_I, a target π is admissible with respect to the pair (Ψ_E, Ψ_I) if and only if*

$$p^t(\pi - \pi^*_I) \geqslant 0, \tag{21}$$

for all p, with π^*_I given by (3).

For instance, assuming the D-criterion, the optimal inferential target is the balanced one (4), so from (21) an allocation π is D-admissible if $p^t\pi \geqslant K^{-1}\sum_{j=1}^K p_j$ for all p, namely π skews the assignments by favouring treatments with success probabilities greater than the average.

The standardized compound criterion Ψ_ω is defined as in (17), where now $\omega = \omega(p) : (0,1)^K \to [0,1)$. It is straightforward to show that when the information criterion Ψ_I is differentiable, the target $\pi^*_\omega(p) = \{\pi^*_{\omega 1}(p), \ldots, \pi^*_{\omega K}(p)\}^t$ minimizing Ψ_ω is the solution of the following set of equations:

$$\frac{\partial \Psi_I(\pi)}{\partial \pi_1} - \frac{\partial \Psi_I(\pi)}{\partial \pi_j} = \frac{\omega(p)}{1-\omega(p)}\left(\frac{p_1 - p_j}{q_{\min}}\right)\Psi^*_I \quad (j = 2, \ldots, K).$$

If Ψ_I is strictly convex, the solution is unique; furthermore, π^*_ω belongs to $(0;1)^K$ and is admissible with respect to (Ψ_E, Ψ_I). However, for $K > 2$, it no longer seems possible to prove that any admissible allocation scheme is optimal with respect to the compound criterion Ψ_ω for a suitable choice of the weights.

8. Conclusions

The design of comparative binary clinical trials presented in this paper is intended to acheive a good trade-off between inferential efficiency and the ethical demand of favouring the better treatment by weighting their relative importance and allowing the relative weights to depend on the unknown success probabilities. The treatment allocation problem is solved in two steps, by first deciding which allocation appears to be a suitable target, and then implementing a sequential experiment converging to it. This method may be useful in contexts other than clinical trials, e.g., industrial experiments with very high costs. Further developments of the basic idea are needed. First of all, the choice of the best weight function $\omega(p_1, p_2)$ in a given applied context remains open. Furthermore, nonbinary observed responses should be considered. Moreover, it is likely that the experimenter would wish to take into account the patients' covariates. A compound criterion can be used in the presence of covariates, and for different types of responses, but with a different expression from (17) (Baldi Antognini & Zagoraiou, 2010). We stress that every comparison involves several subjective choices dependent on the experimenter's preferences, for instance the weight function ω of the present paper and the constraining constants for the optimal solution proposed by Tymofyeyev et al. (2007). In particular, the comparison between our approach and the constrained one by Tymofyeyev et al. (2007) does not necessarily lead to a clear superiority of either. This is true in the case of just two treatments, see § 6, and even more so in the more complicated scenarios when $K > 2$.

ACKNOWLEDGEMENT

This research was partly supported by the Italian Ministry for Research. The authors wish to thank the editor and two referees for their helpful remarks.

REFERENCES

BALDI ANTOGNINI, A. & GIOVAGNOLI, A. (2005). On the large sample optimality of sequential designs for comparing two or more treatments. *Seq. Anal.* **24**, 205–17.

BALDI ANTOGNINI, A. & ZAGORAIOU, M. (2010). Covariate adjusted designs for combining efficiency, ethics and randomness in normal response trials. In *mODa 9-Advances in Model-Oriented Design and Analysis*, A. C. Atkinson, A. Giovagnoli & B. Torsney, C. May (eds). Heidelberg: Springer-Verlag.

BANDYOPADHYAY, U., BISWAS, A. & BHATTACHARYA, R. (2007). A covariate adjusted two-stage allocation design for binary responses in randomized clinical trials. *Statist. Med.* **26**, 4386–99.

BIEDERMANN, S., DETTE, H. & ZHU, W. (2007). Compound optimal designs for percentile estimation in dose-reponse models with restricted design intervals. *J. Statist. Plan. Infer.* **137**, 3838–47.

CLYDE, M. & CHALONER, K. (1996). The equivalence of constrained and weighted designs in multiple objective design problems. *J. Am. Statist. Assoc.* **91**, 1236–44.

COOK, D. & WONG, W. (1994). On the equivalence of constrained and compound optimal designs. *J. Am. Statist. Assoc.* **89**, 687–92.

DETTE, H. (1997). Designing experiments with respect to standardized optimality criteria. *J. R. Statist. Soc.* B **59**, 97–110.

DURHAM, S. D., FLOURNOY, N. & LI, W. (1998). Sequential designs for maximizing the probability of a favorable response. *Can. J. Statist.* **3**, 479–95.

GERALDES, M., MELFI, V., PAGE, C. & ZHANG, H. (2006). The doubly adaptive weighted difference design. *J. Statist. Plan. Infer.* **136**, 1923–39.

HU, F. & ROSENBERGER, W. (2003). Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *J. Am. Statist. Assoc.* **98**, 671–8.

HU, F. & ROSENBERGER, W. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. New York: Wiley & Sons.

HU, F., ROSENBERGER, W. & ZHANG, L. (2006). Asymptotically best response-adaptive randomization procedures. *J. Statist. Plan. Infer.* **136**, 1911–22.

IVANOVA, A. (2003). A play-the-winner type urn model with reduced variability. *Metrika* **58**, 1–13.

MAY, C. & FLOURNOY, N. (2009). Asymptotics in response-adaptive designs generated by a two-color, randomly reinforced urn. *Ann. Statist.* **37**, 1058–78.

MULIERE, P., PAGANONI, A. & SECCHI, P. (2006). A randomly reinforced urn. *J. Statist. Plan. Infer.* **136**, 1853–74.

PUKELSHEIM, F. (2006). *Optimal Design of Experiments*. Philadelphia, PA: SIAM Classics.

ROSENBERGER, W., STALLARD, N., IVANOVA, A., HARPER, C. & RICKS, M. (2001). Optimal adaptive designs for binary response trials. *Biometrics* **57**, 909–13.

TYMOFYEYEV, Y., ROSENBERGER, W. & HU, F. (2007). Implementing optimal allocation in sequential binary response experiments. *J. Am. Statist. Assoc.* **102**, 224–34.

WEI, L. J. & DURHAM, S. (1978). The randomized play-the-winner rule in medical trials. *J. Am. Statist. Assoc.* **73**, 840–3.

ZELEN, M. (1969). Play-the-winner rule and the controlled clinical trials. *J. Am. Statist. Assoc.* **64**, 131–46.

[*Received August* 2009. *Revised May* 2010]