Multipurpose optimal designs for hypothesis testing in normal response trials

Disegni ottimi multi-obiettivo per la verifica di ipotesi

Marco Novelli and Maroussa Zagoraiou

Abstract This work deals with the problem of designing multiarm clinical trials for comparing treatments in order to achieve a compromise between the power of the classical Wald test of homogeneity of the treatment effects and ethical demands. In [5] the authors derived the target allocation maximizing the non-centrality parameter of Wald test for normal responses under a suitable ethical constraint reflecting the treatment effects. Starting from these results, in this paper we provide some important properties of this constrained optimal allocation, like e.g. its D_A-admissibility and its efficiency with respect to ethical and inferential criteria, taking into account estimation precision as well. Comparisons with some allocation proportions proposed in the literature are also presented.

Abstract *Questo lavoro riguarda il problema della pianificazione ottimale di esperimenti comparativi volti ad ottenere validi compromessi tra precisione inferenziale ed esigenze etiche. Prendendo in considerazione il modello normale, in [5] è stata derivata l'allocazione ideale dei trattamenti che massimizza la potenza del test di Wald basato sui contrasti, sotto opportuni vincoli etici legati agli effetti dei singoli trattamenti. L'obiettivo di questo articolo è quello di fornire alcune importanti proprietà di tale allocazione, ossia la D_A-ammissibilità e la sua efficienza rispetto a criteri sia etici che inferenziali, riguardanti anche la precisione di stima, effettuando inoltre opportuni confronti con altre allocazioni target proposte in letteratura.*

Key words: asymptotic inference; ethics; power; multiarm clinical trials

Marco Novelli
Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, Bologna
e-mail: marco.novelli4@unibo.it

Maroussa Zagoraiou
Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, Bologna
e-mail: maroussa.zagoraiou3@unibo.it

1 Introduction

The large majority of randomized clinical trials for treatment comparisons have been designed in order to achieve balanced allocation among the treatment groups, with the aim of maximizing inferential precision about the estimation of the treatment effects. The main justification concerns the so-called "universal optimality" of the balanced design (see e.g. [8]), especially in the context of the linear homoscedastic model, since it optimizes the usual design criteria for the estimation of the treatment contrasts, (like the well-known D-optimality minimizing the volume of the confidence ellipsoid of the contrasts), and it is nearly optimal under several optimality criteria, also under heteroscedasticity [6, 7].

Taking into account the problem of testing statistical hypothesis about the equality of the treatment effects, balance is still optimal in the case of two treatments, since it maximizes the power of the test for normal homoscedastic responses and it is asymptotically optimal in the case of binary outcomes (see e.g. [2, 3]). However, in the case of several treatments the balanced allocation may not be efficient, since it is significantly different from the optimal design for hypothesis testing and could be strongly inappropriate for phase III-trials, in which the ethical demand of individual care often induces to skew the allocations to more efficacious (or less toxic) treatments. To derive a suitable compromise between these goals, Baldi Antognini et al. [5] suggested a constrained optimal target which maximizes the power of the classical Wald test of homogeneity, subject to an ethical constraint on the allocation proportions reflecting the efficacy of the treatments. The aim of the present work is to push forward the results in [5], by providing some important properties of this constrained optimal allocation like, e.g., the D_A-admissibility, and its efficiency with respect to both ethical and inferential criteria, taking into account estimation precision as well. Comparisons with some targets proposed in the literature are also presented.

2 Notation and model

Consider a clinical trial where patients come sequentially and are assigned to one of K available treatments. At each step n, let $\delta_{kn} = 1$ if the nth patient is allocated to the kth ($k = 1, \dots, K$) treatment and 0 otherwise, where $\sum_{k=1}^{K} \delta_{kn} = 1$. Let Y_n be the normally distributed response of the corresponding subject, with $E(Y_n \mid \delta_{kn} = 1) = \mu_k$ denoting the treatment effect and $V(Y_n \mid \delta_{kn} = 1) = \sigma^2$ the unknown common variance; conditionally on the allocations, the responses are assumed to be independent. Furthermore, we denote by $\boldsymbol{\pi}_n^\top = (\pi_{1n}, \dots, \pi_{Kn})$ the vector collecting the proportion of patients assigned to the treatments up to that stage, where $\pi_{kn} = n^{-1} \sum_{i=1}^{n} \delta_{ki}$ ($k = 1, \dots, K$) and $\sum_{k=1}^{K} \pi_{kn} = 1$; also let $\hat{\mu}_{kn}$ ($k = 1, \dots, K$) be the MLE of μ_k, i.e. the sample mean, so $\boldsymbol{\mu}^\top = (\mu_1, \dots, \mu_K)$ and $\hat{\boldsymbol{\mu}}_n^\top = (\hat{\mu}_{1n}, \dots, \hat{\mu}_{Kn})$ are the vectors of the treatment effects and their estimates, respectively. In what follows we assume

"the larger the better" scenario and the following ordering regarding the treatment effects $\mu_1 > \mu_2 > \ldots > \mu_K$.

After n steps the Fisher information matrix (conditional on the design) associated with $\boldsymbol{\mu}$ is $\mathbf{M} = \mathbf{M}(\boldsymbol{\mu} \mid \boldsymbol{\pi}_n) = \sigma^{-2}\text{diag}\,(\pi_{kn})_{k=1,\ldots,K}$. Several authors suggested target allocations $\boldsymbol{\rho}^\top = (\rho_1, \ldots, \rho_K)$ (with $\rho_k \geq 0$ and $\sum_{k=1}^K \rho_k = 1$) in order to optimize the estimation of the treatment effects by choosing suitable criteria regarding $\mathbf{M}(\boldsymbol{\mu} \mid \boldsymbol{\rho})$.

In the context of multiarm clinical trials, the inferential attention is usually devoted to the contrasts. So, letting $\mathbf{A}^\top = [\mathbf{1}_{K-1} \mid -\mathbf{I}_{K-1}]$, where $\mathbf{1}_r$ and $\mathbf{I}_r$ represent the r-dim vector of ones and the identity matrix, respectively, then the vector of contrasts wrt the first treatment (considered as the reference) is $\boldsymbol{\mu}_c = \mathbf{A}^\top \boldsymbol{\mu} = (\mu_1 - \mu_2, \ldots, \mu_1 - \mu_K)^\top$. Under well-known regularity conditions, the corresponding MLE $\hat{\boldsymbol{\mu}}_{cn} = \mathbf{A}^\top \hat{\boldsymbol{\mu}}_n$ is strongly consistent and asymptotically normal with $\sqrt{n}(\hat{\boldsymbol{\mu}}_{cn} - \boldsymbol{\mu}_{cn}) \hookrightarrow_d \mathscr{N}\left(\mathbf{0}, \mathbf{A}^\top \mathbf{M}^{-1} \mathbf{A}\right)$. Within this framework, the balanced design $\boldsymbol{\rho}^B$, namely $\rho_k = K^{-1}$ for every $k = 1, \ldots, K$ is the so-called D_A-optimal allocation, since it minimizes $\det[\mathbf{A}^\top \mathbf{M}^{-1} \mathbf{A}]$.

Whereas, taking into account the problem of testing hypothesis on the equality of the treatments effects, i.e., $H_0 : \boldsymbol{\mu}_c = \mathbf{0}_{K-1}$, versus the alternative $H_A : \boldsymbol{\mu}_c \neq \mathbf{0}_{K-1}$, where $\mathbf{0}_{K-1}$ is the $(K-1)$-dim vector of zeros, then the optimal design maximizing the power of the classical Wald test is $\boldsymbol{\rho}^* = (1/2, 0, \ldots, 0, 1/2)^\top$ (see [5]). Clearly, this optimal allocation is unsuitable both from the ethical and the inferential point of views.

Regarding ethics, Atkinson [1] proposed a target intended to skew the assignments towards the best treatment in order to minimize the exposure of patients to toxic (or inefficacious) treatments. In particular, denoting by $\bar{\mu} = K^{-1} \sum_{k=1}^K \mu_k = \boldsymbol{\mu}^\top \boldsymbol{\rho}^B$ the overall treatment mean, the target $\boldsymbol{\rho}^A_{(\gamma)}$ proposed by Atkinson is

$$\rho^A_{(\gamma)k} = \Phi\left(\frac{\mu_k - \bar{\mu}}{\gamma}\right) \Big/ \left[\sum_{i=1}^K \Phi\left(\frac{\mu_i - \bar{\mu}}{\gamma}\right)\right], \quad k = 1, \ldots, K.$$

In the same spirit, instead of $\Phi(\cdot)$ any non-negative increasing function can be used. An example is the exponential target $\boldsymbol{\rho}^E_{(\gamma)}$ given by

$$\rho^E_{(\gamma)k} = e^{\frac{\mu_k - \bar{\mu}}{\gamma}} \Big/ \left(\sum_{i=1}^K e^{\frac{\mu_i - \bar{\mu}}{\gamma}}\right) = e^{\frac{\mu_k}{\gamma}} \Big/ \left(\sum_{i=1}^K e^{\frac{\mu_i}{\gamma}}\right), k = 1, \ldots, K.$$

Clearly, small values of γ induce a strong ethical skew, while as γ increases more emphasis is given to inferential purposes. In particular, adopting $\boldsymbol{\rho}^E_{(\gamma)}$, the allocation proportion $\rho^E_{(\gamma)1}$ to the best treatment is decreasing as γ grows, since

$$\frac{\partial \rho^E_{(\gamma)1}}{\partial \gamma} = \frac{\sum_{i=1}^K e^{\frac{\mu_i + \mu_1}{\gamma}} (\mu_i - \mu_1)}{\left(\sum_{i=1}^K e^{\frac{\mu_i}{\gamma}}\right)^2 \gamma^2} < 0.$$

Note that such monotonicity property does not hold, in general, for $\boldsymbol{\rho}^A$ as we shall show in the last section.

3 Constrained optimal allocation and its D_A-admissibility

Adopting a constrained optimization framework, Baldi Antognini et al. in [5] derived the allocation maximizing the power of Wald test under a suitable ethical constraint reflecting the efficacy of the treatments. In particular, the optimal target $\tilde{\boldsymbol{\rho}}^\top=(\tilde{\rho}_1,\ldots,\tilde{\rho}_K)$ maximizing the non-centrality parameter $\phi(\tilde{\boldsymbol{\rho}}) = n\sigma^{-2}\boldsymbol{\mu}_c^\top \left[\mathbf{A}^\top \text{diag}(\tilde{\boldsymbol{\rho}})^{-1}\mathbf{A}\right]^{-1}\boldsymbol{\mu}_c$ of the multivariate Wald test subject to the ethical constraint $\tilde{\rho}_1 \geq \tilde{\rho}_2 \geq \ldots \geq \tilde{\rho}_K$ is $\tilde{\boldsymbol{\rho}} = (1-t[K-1],t,\ldots,t)^\top$ if $t \leq K^{-1}$, while $\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}^B$ if $t > K^{-1}$, where $t = \sum_{k=2}^K (\mu_1-\mu_k)^2 / \left\{2\left[\sum_{k=2}^K(\mu_1-\mu_k)\right]^2\right\}$. Table 1 shows how the allocation $\tilde{\boldsymbol{\rho}}$ moves away from the balanced design as the distance between μ_1 and μ_2 increases, skewing the assignments to the superior treatment.

Table 1: The behaviour of the optimal constrained target $\tilde{\boldsymbol{\rho}}$ with $K=3$ as μ_2 varies.

μ_1	μ_2	μ_3	$\tilde{\rho}_1$	$\tilde{\rho}_2$	$\tilde{\rho}_3$	t
15	14	6	0.333	0.333	0.333	0.410
15	12	6	0.375	0.312	0.312	0.312
15	10	6	0.459	0.270	0.270	0.270
15	8	6	0.492	0.254	0.254	0.254

Following the definition of admissibility proposed in [4], it is easy to show that $\tilde{\boldsymbol{\rho}}$ is D_A-admissible, i.e. it does not exist another allocation which is simultaneously superior wrt both ethics and D_A-optimality. Indeed, when $t > 1/K$, $\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}^B$ and the D_A-admissibility is trivially satisfied, while for $t \leq 1/K$, $\boldsymbol{\mu}^\top(\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho}^B) \geq 0 \iff \boldsymbol{\mu}^\top\tilde{\boldsymbol{\rho}} \geq \bar{\mu} \iff \mu_1(1-Kt) \geq \bar{\mu}(1-Kt)$ which is always true since $\mu_1 > \bar{\mu}$.

4 Comparisons

We now compare the performance of $\tilde{\boldsymbol{\rho}}$, $\boldsymbol{\rho}^B$, Atkinson's target $\boldsymbol{\rho}^A_\gamma$ and the exponential one $\boldsymbol{\rho}^E_\gamma$ both with $\gamma=1$ and $\gamma=3$. In particular, in Table 2 we consider the following criteria: i) an ethical measure of efficiency given by the ratio between the total expected outcomes and its optimal value, i.e., $E_E(\boldsymbol{\rho}) = \sum_{k=1}^K \mu_k\rho_k/\mu_1$, ii) an efficiency measure of statistical power $E_P(\boldsymbol{\rho}) = \phi(\boldsymbol{\rho})/\phi(\boldsymbol{\rho}^*)$ and the D_A-efficiency $E_{D_A}(\boldsymbol{\rho}) = \left\{\det\left[\mathbf{A}^\top\mathbf{M}^{-1}\left(\boldsymbol{\rho}^B\right)\mathbf{A}\right]/\det\left[\mathbf{A}^\top\mathbf{M}^{-1}(\boldsymbol{\rho})\mathbf{A}\right]\right\}^{\frac{1}{K-1}}$ for estimation.

Table 2: the case $K = 5$ treatments

$\boldsymbol{\mu}^\top$	Targets	$E_E(\boldsymbol{\rho})$	$E_P(\boldsymbol{\rho})$	$E_{D_A}(\boldsymbol{\rho})$
(21,20,19,18,16)	$\boldsymbol{\rho}^A_{(1)} = (0.37, 0.332, 0.217, 0.08, 0.001)^\top$	0.952	0.147	0.282
	$\boldsymbol{\rho}^A_{(3)} = (0.305, 0.26, 0.209, 0.157, 0.07)^\top$	0.929	0.321	0.867
	$\tilde{\boldsymbol{\rho}} = (0.355, 0.161, 0.161, 0.161, 0.161)^\top$	0.916	0.503	0.930
	$\boldsymbol{\rho}^E_{(1)} = (0.641, 0.236, 0.087, 0.032, 0.004)^\top$	0.975	0.112	0.274
	$\boldsymbol{\rho}^E_{(3)} = (0.359, 0.257, 0.184, 0.132, 0.068)^\top$	0.935	0.324	0.830
	$\boldsymbol{\rho}^B$	0.895	0.474	1
(23,20,19,18,16)	$\boldsymbol{\rho}^A_{(1)} = (0.43, 0.339, 0.181, 0.05, 0)^\top$	0.913	0.264	0.186
	$\boldsymbol{\rho}^A_{(3)} = (0.364, 0.246, 0.192, 0.14, 0.058)^\top$	0.886	0.392	0.813
	$\tilde{\boldsymbol{\rho}} = (0.452, 0.137, 0.137, 0.137, 0.137)^\top$	0.887	0.554	0.840
	$\boldsymbol{\rho}^E_{(1)} = (0.93, 0.046, 0.017, 0.006, 0.001)^\top$	0.989	0.068	0.059
	$\boldsymbol{\rho}^E_{(3)} = (0.522, 0.192, 0.137, 0.099, 0.051)^\top$	0.914	0.406	0.680
	$\boldsymbol{\rho}^B$	0.835	0.438	1
(25,20,19,18,16)	$\boldsymbol{\rho}^A_{(1)} = (0.504, 0.33, 0.138, 0.028, 0)^\top$	0.893	0.367	0.112
	$\boldsymbol{\rho}^A_{(3)} = (0.41, 0.235, 0.179, 0.126, 0.049)^\top$	0.857	0.479	0.760
	$\tilde{\boldsymbol{\rho}} = (0.476, 0.131, 0.131, 0.131, 0.131)^\top$	0.859	0.618	0.814
	$\boldsymbol{\rho}^E_{(1)} = (0.99, 0.007, 0.002, 0.001, 0)^\top$	0.998	0.015	0.009
	$\boldsymbol{\rho}^E_{(3)} = (0.68, 0.128, 0.092, 0.066, 0.034)^\top$	0.922	0.427	0.486
	$\boldsymbol{\rho}^B$	0.784	0.446	1
(27,20,19,18,16)	$\boldsymbol{\rho}^A_{(1)} = (0.595, 0.297, 0.094, 0.014, 0)^\top$	0.890	0.428	0.060
	$\boldsymbol{\rho}^A_{(3)} = (0.449, 0.227, 0.168, 0.115, 0.041)^\top$	0.836	0.550	0.710
	$\tilde{\boldsymbol{\rho}} = (0.486, 0.129, 0.129, 0.129, 0.129)^\top$	0.833	0.669	0.803
	$\boldsymbol{\rho}^E_{(1)} = (0.999, 0.001, 0, 0, 0)^\top$	1	0.003	0.001
	$\boldsymbol{\rho}^E_{(3)} = (0.805, 0.078, 0.056, 0.04, 0.021)^\top$	0.941	0.352	0.309
	$\boldsymbol{\rho}^B$	0.741	0.463	1
(25,20,19,18,11)	$\boldsymbol{\rho}^A_{(1)} = (0.351, 0.323, 0.23, 0.096, 0)^\top$	0.853	0.155	$\to 0$
	$\boldsymbol{\rho}^A_{(3)} = (0.372, 0.257, 0.209, 0.159, 0.002)^\top$	0.853	0.175	0.382
	$\tilde{\boldsymbol{\rho}} = (0.402, 0.149, 0.149, 0.149, 0.149)^\top$	0.809	0.467	0.890
	$\boldsymbol{\rho}^E_{(1)} = (0.99, 0.007, 0.002, 0.001, 0)^\top$	0.998	0.006	0.002
	$\boldsymbol{\rho}^E_{(3)} = (0.699, 0.132, 0.095, 0.068, 0.007)^\top$	0.928	0.165	0.332
	$\boldsymbol{\rho}^B$	0.744	0.413	1
(25,20,19,18,13)	$\boldsymbol{\rho}^A_{(1)} = (0.4, 0.337, 0.2, 0.063, 0)^\top$	0.867	0.213	0.007
	$\boldsymbol{\rho}^A_{(3)} = (0.391, 0.252, 0.2, 0.148, 0.009)^\top$	0.856	0.252	0.537
	$\tilde{\boldsymbol{\rho}} = (0.436, 0.141, 0.141, 0.141, 0.141)^\top$	0.831	0.498	0.857
	$\boldsymbol{\rho}^E_{(1)} = (0.99, 0.007, 0.002, 0.001, 0)^\top$	0.998	0.008	0.004
	$\boldsymbol{\rho}^E_{(3)} = (0.695, 0.131, 0.094, 0.067, 0.013)^\top$	0.926	0.233	0.389
	$\boldsymbol{\rho}^B$	0.760	0.411	1
(25,20,19,18,15)	$\boldsymbol{\rho}^A_{(1)} = (0.465, 0.337, 0.16, 0.038, 0)^\top$	0.884	0.303	0.052
	$\boldsymbol{\rho}^A_{(3)} = (0.406, 0.243, 0.187, 0.134, 0.03)^\top$	0.857	0.384	0.693
	$\tilde{\boldsymbol{\rho}} = (0.464, 0.134, 0.134, 0.134, 0.134)^\top$	0.850	0.562	0.827
	$\boldsymbol{\rho}^E_{(1)} = (0.99, 0.007, 0.002, 0.001, 0)^\top$	0.998	0.012	0.007
	$\boldsymbol{\rho}^E_{(3)} = (0.686, 0.13, 0.093, 0.067, 0.024)^\top$	0.923	0.345	0.453
	$\boldsymbol{\rho}^B$	0.776	0.426	1
(25,20,19,18,17)	$\boldsymbol{\rho}^A_{(1)} = (0.547, 0.317, 0.116, 0.02, 0.001)^\top$	0.903	0.453	0.204
	$\boldsymbol{\rho}^A_{(3)} = (0.411, 0.226, 0.17, 0.118, 0.075)^\top$	0.857	0.598	0.813
	$\tilde{\boldsymbol{\rho}} = (0.485, 0.129, 0.129, 0.129, 0.129)^\top$	0.866	0.700	0.803
	$\boldsymbol{\rho}^E_{(1)} = (0.99, 0.007, 0.002, 0.001, 0)^\top$	0.998	0.020	0.011
	$\boldsymbol{\rho}^E_{(3)} = (0.671, 0.127, 0.091, 0.065, 0.047)^\top$	0.920	0.536	0.520
	$\boldsymbol{\rho}^B$	0.792	0.485	1

Considering the statistical power, $\tilde{\boldsymbol{\rho}}$ has the best performance with a gain up to 13% with respect to any second best option. The rules $\boldsymbol{\rho}^A_{(1)}$ and $\boldsymbol{\rho}^E_{(1)}$ show the lowest statistical power but, at the same time, the highest ethical efficiency. Note that, as γ grows, more emphasis is devoted to inference. However, contrary to the exponential target, the Atkinson's allocation proportion to the best treatment, $\rho^A_{(\gamma)1}$, is not always decreasing in γ. Moreover, $\tilde{\boldsymbol{\rho}}$ performs very well also from the ethical point of view.

Regarding the D_A-efficiency, $\tilde{\boldsymbol{\rho}}$ is substantially superior with respect to $\boldsymbol{\rho}^A$ and $\boldsymbol{\rho}^E$ guaranteing at the same time an efficiency always greater than 80.3%. Note that, adopting $\boldsymbol{\rho}^A_{(1)}$ and $\boldsymbol{\rho}^E_{(1)}$ the D_A-efficiency often tends to zero and therefore the estimation precision may vanish.

Since ethics and inference are conflicting demands, a target showing high efficiency under one criterion may perform worst under other criteria. However, $\tilde{\boldsymbol{\rho}}$ represents a valid compromise between inferential (both in terms of power and estimation precision) and ethical concerns.

Acknowledgements Research supported by the Italian Ministry of Education, University and Research under PRIN 2015 "Environmental processes and human activities: capturing their interactions via statistical methods (EphaStat)"

References

1. Atkinson, A.C.: Adaptive biased-coin designs for clinical trials with several treatments. Discussiones Mathematicae Probability and Statistics **24**, 85–108 (2004)
2. Azriel, D., Mandel, M., Rinott, Y.: Optimal allocation to maximize power of two-sample tests for binary response. Biometrika **99**, 101–113 (2012)
3. Baldi Antognini, A.: A theoretical analysis of the power of biased coin designs. J Stat Plan Inference **138**, 1792–1798 (2008)
4. Baldi Antognini, A., Giovagnoli, A.: Compound optimal allocation for individual and collective ethics in binary clinical trials. Biometrika **97**, 935–946 (2010)
5. Baldi Antognini, A., Novelli, M., Zagoraiou, M.: Optimal designs for testing hypothesis in multiarm clinical trials. Submitted.
6. Begg, C.B., Kalish, L.A.: Treatment Allocation for Nonlinear Models in Clinical Trials: The Logistic Model. Biometrics **40**, 409–420 (1984)
7. Kalish, L.A., Harrington, D.P.: Efficiency of balanced treatment allocation for survival analysis. Biometrics **44**, 409–420 (1988)
8. Silvey, S.D.: Optimal Designs. Chapman & Hall, London (1980)