Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

SCIENZE STATISTICHE

Ciclo 33

Settore Concorsuale: 13/D1 - STATISTICA

Settore Scientifico disciplinare: SECS-S/01 - STATISTICA

DESIGN OF EXPERIMENTS IN DRUG DEVELOPMENT: OPTIMAL ALLOCATIONS IN MULTI-ARM CLINICAL TRIALS AND DESIGN SPACE FOR MULTI-STEP PROCESSES

Presentata da: Rosamarie Frieri

Cordinatore Dottorato

Monica Chiogna

Supervisore

Maroussa Zagoraiou

Co-Supervisori

Marco Mariti

Marilena Paludi

Esame finale anno 2021

Abstract

In this thesis, we deal with the design of experiments in the drug development process, focusing on the design of clinical trials for treatment comparisons (Part I) and the design of preclinical laboratory experiments for proteins development and manufacturing (Part II).

In Part I, we propose a multi-purpose design methodology for sequential clinical trials. We derived optimal allocations of patients to treatments for testing the efficacy of several experimental groups by also taking into account ethical considerations. We first consider exponential responses for survival trials and we then present a unified framework for heteroscedastic experimental groups that encompasses the general ANOVA set-up. The very good performance of the suggested optimal allocations, in terms of both inferential and ethical characteristics, are illustrated analytically and through several numerical examples, also performing comparisons with other designs proposed in the literature.

Part II concerns the planning of experiments for processes composed of multiple steps in the context of preclinical drug development and manufacturing. Following the Quality by Design paradigm, the objective of the multi-step design strategy is the definition of the manufacturing design space of the whole process and, as we consider the interactions among the subsequent steps, our proposal ensures the quality and the safety of the final product, by enabling more flexibility and process robustness in the manufacturing.

Contents

Chapter 1

Introduction

In the last decades, the pharmaceutical industry is facing big challenges, not only due to the novel COVID-19 outbreak. Because of the increasing competition at a global scale and the need to speed up the drug development process, pharmaceutical industry is undergoing an accelerated structural change with the objective of improving its operational performance (e.g., to reduce time to market, waste and costs) and the quality of its products.[1] Clearly all the procedures have to be compliant with the regulatory guidelines provided by the health authorities. Indeed, the pharmaceutical industry is one of the most regulated in the world: the consequences of releasing harmful drugs would be devastating.[2] The process from the discovery of a new drug candidate to its introduction to the market is very long, complex and expensive: on average 10–12 years are required for this process with a cost that was estimated to be more than 1 billion US dollars in the early 2000s (note that this includes the amortized cost of failed drugs). According to the statistics, among the 5000–10000 promising compounds, 5 will pass into clinical trials, and only one will be approved for sales.[2]

1.1 The process of drug development

The drug development process consists of the following phases:

Discovery $\longrightarrow$ Preclinical $\longrightarrow$ Clinical $\longrightarrow$ Marketing Application and Approval.

As discussed in the next paragraphs, each phase of this process is supervised by the regulatory authorities in order to ensure safety, efficacy and consistency of drugs for human subjects' use.[2]

Drug discovery

Drug development starts with the identification of a "druggable target" (e.g. proteins, receptors, enzymes,...) that is involved in the pathology and causes or leads to the disease. Once a target has been selected, the next step is the research of any substances that might result in some therapeutic effect on it[3] ending up with a small number of substances that may lead to a candidate drug and eventually to a usable pharmaceutical. This is the second major stage of the research and development (R&D) process and marks the transition from research into development.[2]

Preclinical Drug Development

In the preclinical research activities, the lead compounds are tested *in vitro* and *in vivo* (recently also *in silico*) to check their safety and efficacy. These compounds are also modified to increase their effectiveness and to reduce toxicity. The resulting optimized compound becomes the drug candidate for clinical trials in humans. Preclinical studies have to be compliant with the guidelines dictated by Good Laboratory Practice (GLP) to ensure reliable results and required by authorities such as the Food and Drug Administration (FDA) or the European Medicines Agency (EMA).[4] This period is also dedicated to the design of the process that will be used to manufacture trial batches of the substances for use in the clinical trials (and eventually for full-scale manufacture).

Despite preclinical studies are not usually very large, they must provide detailed information on dosing and toxicity levels which are essential to determine whether it is reasonable to test the drug in people.[4]

Clinical Research

In clinical trials the drug dosage form intended to be commercialized is administered on human subjects. Clinical trials must be performed in accordance with the Good Clinical Practice (GCP). GCP is "an international ethical and scientific quality standard for designing, conducting, recording and reporting trials that involve the participation of human subjects".[5] When clinical trial data are intended to be submitted to the regulatory agencies, its generation should be in accordance with GCP. In addition, every trial has to be approved by the regulatory authority in the country or region where the clinical trial is to be conducted (from the United States FDA, from European Union EMA, ...).[2]

Clinical trials are conducted into four Phases.

- *Phase I*. These studies represent the first time that a drug is administered to humans. The population of interest is usually a small number of healthy volunteers or people with the disease and the primary goal is to investigate the drug's safety and tolerability, its dosage and, if possible, the pharmacokinetics (how the body affects the drug) and pharmacodynamics (how the drug affects the body) of the product.[6] Phase I studies involve from 20 to 100 subjects and last several months. Around the 70% of drugs move to Phase II.[7]

- *Phase II*. The goal of the Phase II is to examine the safety and effectiveness of the drug in a group of patients with the disease or condition for which the drug has been designed. Studies of Phase II supply further safety data that researchers use to find the effective dose and the dosing regimen, refine the research questions and to prepare the design of the next Phase research protocols.[2,7] According to the disease and whether one or more experimental therapies are available, comparisons may be performed between experimental groups and/or against a control group and/or a placebo group.[8] In Phase II trials are recruited up to several hundred of subjects and the length of the study ranges from several months up to around 2 years. In this case, approximately the 33% of drugs move to Phase III.[7]

- *Phase III*. This is the final stage before product registration. The main objective for which Phase III studies are designed is to assess whether the drug delivers any therapeutic effects. As for Phase II, Phase III trials may be multi-arm. The number of participants of Phase III trials goes from 300 to 3000 volunteers affected by the disease. The duration of the study is 1-4 years and around the

25-30% of drugs move to Phase IV. As these studies are larger and last longer, the data are more likely to show previously undetected long-term or rare adverse events.[7]

- *Phase IV*. These studies are performed after the drug approval. Commonly Phase IV trials may regard additional drug-drug interaction, dose-response or safety studies.[9]

Manufacturing

The drug must be manufactured in accordance to the current Good Manufacturing Practice (cGMP) following FDA, EMA or International Conference on Harmonization (ICH) directives. Regulatory authorities can inspect manufacturing plants to ensure they follow cGMP guidelines to guarantee that the drug manufactured is safe and effective for patients.

Marketing, Approval, and Postapproval

In the case of successful clinical trials results for a drug candidate, the next step is the submission of the marketing authorization application to the regulatory authorities, which has to be reviewed and approved. Even after the drug approval and its marketing, its safety and performance are still monitored to guarantee that it is prescribed correctly and that adverse events are reported and investigated.[2] Also the advertising of drugs is checked by regulatory authorities, in order to ensure that correctly statements on the product are delivered to patients.

1.1.1 How COVID-19 pandemic is changing pharmaceutical development

In the last years, the need to speed up the development process has been one of the main objectives of regulators, researchers and industry leaders but the COVID-19 pandemic has incredibly push the challenge beyond the limits: the rapid development of a therapeutics or a vaccine has become a global imperative in 2020.

Since the end of 2019, a respiratory disease caused by a novel coronavirus has leaded to a pandemic outbreak that is still ongoing. From the first cases reported in China, it rapidly spread all over the world. The name of the virus is "SARS-CoV-2" and it lead to a disease that has been named "COVID-19".[10] On January 31, 2020, the Secretary of Health and Human Services has declared the state of public health emergency related to COVID-19 and on March 11, 2020 the World Health Organization announced COVID-19 outbreak as a pandemic.[11] Such pandemic has caused widespread morbidity, mortality and economic disruption, and the consequences of repeated epidemic waves could be unacceptably serious, worldwide.[12] Therefore, a safe and effective vaccine is urgently needed to control and hopefully stop the pandemic. With this aim, pharmaceutical companies, academic researchers and government agencies are working together and making an unprecedented effort to compress within months the process of the COVID-19 vaccine development.[13] Several national regulatory agencies have released a plethora of documents with recommendations and guidelines on the procedures to facilitate the clinical development[10,14–16]. For example FDA has created the Coronavirus Treatment Acceleration Program,[17] a special emergency plan for possible therapies with the aim of moving new treatments to patients as quickly as possible, still preserving their safety. FDA has also provided the Emergency Use Authorization for COVID-19 vaccines[18] that contains recommendations regarding the data and information needed to support the issuance of an Emergency Use Authorization during the public health emergency. EMA outlined

the rapid formal review procedures[16] including e.g. the rolling reviews (data for an upcoming vaccine are sent to EMA as they become available so that their assessment can start before the formal submission). Moreover EMA introduced other regulatory mechanisms as the Conditional Marketing Authorisation (less comprehensive data may also be accepted for its request in the case of emergency situations in which the benefit of availability of the product on public health overcomes the risks) and compassionate use programmes (the unauthorised medicine can be used under strict conditions for patients with a disease for which no approved therapies exist).[15]

Also thanks to these accelerated regulatory procedures, new promising vaccines are emerging faster than before and promising results seems to be close to come.[19] However, the demand of obtaining timely response on the vaccine performance should be balanced with the concern of scientific rigour that prove its effectiveness but firstly its safeness. Indeed vaccines, unlike therapeutics, are administered to a large number of otherwise healthy individuals so that safety must be the primary goal.[12]

1.1.2 The role of Statistics in pharmaceutical development

Nowadays, there is no aspect of pharmaceutical development in which Statistics is not involved: from drug discovery to forecasting sales of the final product.[20] In particular, the application of Statistics in pharmaceutical research is gaining importance due to the growing number of new challenges, the need to speed up the pharmaceutical development process and the high expectations of regulatory agencies.[21] The significant role of statistical thinking in scientific and clinical research has been widely recognized and discussed in the literature.[20–24] Now that the COVID-19 pandemic situation is affecting everyone's life, the importance of scientific data is becoming the talk of the town. The role of data is crucial: starting from the drugs approval, which requires the submission of data demonstrating its safety and effectiveness in both preclinical and clinical stages, up to the communication to the public, for which is essential to translate the information coming from data into user-friendly materials referencing publicly available scientific studies.[25] However, besides the manner in which data are analysed, the scientific validity and reliability of any result is related to the way in which experimental and clinical data are collected: the quality of the information extrapolated from data depends on the quality of the data itself.[26] The *design of experiment* is the branch of Statistics that deals with the statistical and mathematical methods aimed at planning experimental study to effectively gain information form empirical data. Depending on the field in which the experiments are performed, the research has been developed in several directions in order to address problem-specific issues and to achieve purposeful experimental goals. Even in the pharmaceutical development process itself, the design strategies for preclinical experiments substantially differ from the planning of clinical studies in which humans are involved.

In this thesis we focus on two main aspects of the design of experiments in the drug development process: the design of clinical trials for treatment comparisons and the design of preclinical laboratory experiments for proteins development and manufacturing.

1.2 Content and thesis outline

This thesis is organized into two parts. Part I concerns the design of clinical experiments for testing the efficacy of several heterogeneous treatment groups. We propose a methodology that is suitable for the design of clinical Phase II/III sequential trials to derive optimal allocations of patients to treatments by taking also into account ethical considerations. This topic is presented in Chapter 2 and 3 and the

results are collected into two papers: paper A and paper B. The research addressed in Part II is placed in the context of preclinical drug development and manufacturing. The aim of the work was the design of experiments for processes composed of multiple subsequent steps, with the final objective of deriving the manufacturing design space, which is one of the fundamental concept in the Quality by Design paradigm.[27] This is the subject of paper C, reported in Chapter 4.

1.2.1 Overview on Part I

The fundamental feature of clinical trials is randomization of patients to treatments.[26] The large majority of randomized clinical trials have been designed to achieve the equal allocation among the treatment groups. The motivations behind the use of the balanced allocation is that it maximizes the inferential precision in the estimation of the treatment effects in many set-ups. In addition, the balanced allocation is consistent with the state of equipoise, i.e. a genuine uncertainty about which experimental therapy is more effective.[28] Traditionally, the state of equipoise was the ethical requirement to accept that patients are assigned to treatments through randomization.[26] However, the equal allocation may not be efficient and strongly inappropriate for clinical trials for rare or lethal diseases. Therefore, especially in these scenarios, the design of a clinical experiment should be aimed at finding a compromise[29] between

(i) the ethical demand of maximizing patients' care, by favouring the assignments to the superior treatment(s) according to the evidence accrued thus far,

(ii) the inferential goal of deriving statistical conclusions with high precision.

This is the so called dilemma of individual versus collective ethics, i.e., what is best for the current patient enrolled in the trial (i) versus what is best for reaching medical advances from which future patients may benefit (ii).[29,30] The trade-off between these conflicting goals can be formalized into a combined/constrained optimization problem, whose solution is a target allocation. The target allocation is the proportion of patients to be assigned to each treatment group in order to achieve the desired compromise. The target generally depends on the unknown model parameters so that it is a-priori unknown. In the context of sequential experiments, Response Adaptive Randomization procedures can be implemented under suitable conditions. According to Response Adaptive rules, as each subject joins the trials, model parameters are estimated from previously accumulated responses and past assignments, and the next allocation probabilities is randomly changed in order to asymptotically approach the target.[29] The use of adaptive designs has been widely promoted to increase clinical trial efficiency and flexibility, also by the regulatory authorities.[2,6,31] Moreover, the clinical development programs for COVID-19 vaccines encourage the implementation of adaptive clinical trial designs, whose use may accelerate the drug development process.[10]

While most of the literature on the design of clinical experiments is focused on the estimation of treatment effects, the problem of hypothesis testing has received limited attention. In this thesis we address the problem of testing the null-hypothesis of equality among treatment effects in the presence of several heteroscedastic experimental groups. More specifically, we derived the unconstrained optimal target, namely the allocation proportion that maximizes the power of the multivariate Wald test. In addition, to take into account ethical considerations, we also derived the constrained optimal target. This target maximizes the power subject to a constraint on the allocation proportions, reflecting the treatment efficacies. The global null-hypothesis considered in Paper A and B was adopted to test the equality of all the treatment effects. Besides being a cornerstone in the statistical literature, it constitutes the first stage

of multiple comparison methodologies for several stepwise procedures, like the well-known Fisher least significant difference method.[32] In addition, due to the strong increase of new potential drugs, mainly in cancer research, one of the primary Phase II goals may be to evaluate the effectiveness of new treatments and to identify which one(s) most warrants additional evaluation in a larger Phase III.[8,33] Particularly relevant to the present pandemic situation, the overall hypothesis of homogeneity may be appropriate for cases in which no treatments with demonstrated efficacy exist and so no standard of care is available to be set as an appropriate/fixed control. Indeed for rapidly emerging novel infectious diseases, such as Ebola[34] or COVID-19, clinical trials must start as soon as possible and the global null-hypothesis, as a first step, allows to evaluate several candidate treatments at once.

Our research has started with paper A, in which we derived optimal designs for hypothesis testing in the case of exponential responses. Among parametric models, the exponential one is typically assumed for the lifetime distribution in survival trials, making our proposal very relevant in the context of clinical studies. Indeed, in paper A, we deeply studied the properties of the constrained optimal design, matching analytical results and numerical examples. We included comparisons with other designs proposed in the literature from which a remarkable gain in terms of both power and ethics arise from our proposal. Additionally, extensive simulation studies have been performed, including robustness studies to model misspecification, robustness to misspecification of patients' recruitment pattern and the redesign of a real lung cancer trial. We addressed several natural complications of the design of adaptive survival trials by also considering in the model the presence of censored observations, as well as delayed responses and patients staggered entries. We widely discussed their impact in the implementation through Response Adaptive Randomization. Paper A has been published as *Frieri R, Zagoraiou M. Optimal and ethical designs for hypothesis testing in multi-arm exponential trials. Statistics in Medicine. 2021;1–26. https://doi.org/10.1002/sim.8919.*

Later, in paper B, we expanded our results by presenting a unified framework, that also encompass the general ANOVA. The results of paper B own a particular theoretical relevance and fill the gap in the literature on the optimal designs of experiments for comparing several treatments when the inferential aim is testing the homogeneity of the treatment effects in the ANOVA set-up with heteroscedastic errors. Due to the generality of the framework in which we analytically derived the optimal targets, the results apply to the parametric models belonging to the so-called minimal exponential family, like, Binary, Poisson and Exponential models with and without censoring. We finally illustrated both theoretically and numerically the interesting properties of our targets also in comparisons with other designs previously obtained.

1.2.2 Overview on Part II

The aim of pharmaceutical development is to design a product and its manufacturing process to consistently obtain a drug with the intended quality.[35] As previously mentioned, the manufacturing of drugs must be carried out in compliance with cGMP and in 2002, FDA introduced a new initiative to cGMP: *Pharmaceutical cGMPs for the 21st Century: A Risk-Based Approach.*[36] The main motivation was to modernize the regulation of pharmaceutical manufacturing and product quality in order to promote faster and more consistent product and process development.[37] In addition, FDA has published as part of the International Conference on Harmonization (ICH) guidelines several documents (Q8 on Pharmaceutical Development[27], Q9 on Quality Risk Management[35] and Q10 on pharmaceutical quality system[38]) defining the Quality by Design (QbD) principles and its implementation. The underlying philosophy

of QbD is that quality cannot be tested into a product (Quality by Testing), but must be built into it and ensured since its design. QbD principles provide a science-based rather than empirical-based approach to design and develop a product and its manufacturing process to guarantee that the quality of the products consistently attains the desired clinical performance. This new approach to pharmaceutical development replaces the traditional Quality by Testing, by introducing more flexible approaches from both the manufacturing and the regulatory perspective.[39]

One of the key stage in the implementation of the QbD is the identification of the critical quality attributes that are properties or characteristics (physical, chemical, biological or microbiological) that should be within an appropriate range to guarantee the desired product quality. The relationship between the process inputs (material attributes and process parameters/input factors) and the critical quality attributes can be described in the manufacturing *design space* which is defined as the "multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality".[27] The manufacturing design space is proposed by scientists as a region of process capability and once it has been approved by the regulatory agencies provide an useful tool for the applicants. According to the guidelines, movements within the manufacturing design space are not considered as a change. For instance, by using the manufacturing design space, the pharmaceutical manufacturers are allowed to make small adjustments in the operating conditions (inside the design space), without requiring long regulatory approval process. Notice that the changes within the manufacturing design space of the manufacturing conditions may also potentially save from batch failure.[21] Whereas, operating outside of the manufacturing design space would normally initiate a regulatory post approval change process. A manufacturing design space is based on mathematical models that can be derived from first principles reflecting physical laws or from experimental data (or from their combination).[40] The manufacturing design space is then established in terms of ranges of material attributes and process parameters, or through more complex mathematical relationships.

In practice, most biochemical processes are composed of multiple subsequent steps. In each step, the operations and the resources involved are controlled by multiple process parameters/input factors and the outcome of each step affects the result of the subsequent steps. On this purpose, QbD guidelines[27] highlight that rather than defining a separate independent manufacturing design space for each unit operation (often simpler to develop), a single manufacturing design space that spans the entire process can increase the operational flexibility. However, in the case of multi-step processes, the relationship between the process inputs and the critical quality attributes is really complex and, in general, models based on first principles are not available.

In the context of the application of the QbD principles in the Technical R&D of GSK Vaccines (Siena), in Paper C we present a multi-step experimental strategy to design experiments and derive the manufacturing design space of a multi-step process. As we take into account the interactions among subsequent steps in the definition of the manufacturing design space, our proposal guarantees the quality and the safety of the final product with more flexibility and process robustness in the manufacturing. We considered the application of the multi-step methodology to the biochemical process of expression and purification of a recombinant protein. Paper C has been published as *Frieri R, Mariti M, Paludi M. Design of experiments and manufacturing design space for multi-step processes. Applied Stochastic Models in Business and Industry. 2021; 1-19. https://doi.org/10.1002/asmb2620.*

Bibliography

[1] Grangeia HB, Silva C, Simões SP, and Reis MS. Quality by design in pharmaceutical manufacturing: A systematic review of current status, challenges and future perspectives. *Eur J Pharm Biopharm*, 147:19 – 37, 2020.

[2] Ng R. *Drugs: From Discovery to Approval*. Wiley, 2005.

[3] Hughes JP, Rees S, Kalindjian SB, and Philpott KL. Principles of early drug discovery. *Br J Pharmacol*, 162(6):1239–1249, 2011.

[4] U.S. Food and Drug Administration. The drug development process. `https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process`, 2017. Available Online.

[5] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). ICH E6 (R2) Good Clinical Practice, 2016. Available online.

[6] Salazar ED and Gormley G. Modern drug discovery and development. In *Clinical and Translational Science, Principles of Human Research*, pages 719–743. Academic Press, 2017.

[7] US Food and Drug Administration. Step 3: Clinical research. `https://www.fda.gov/patients/drug-development-process/step-3-clinical-research`, 2018.

[8] Jung SH and George SL. Between-arm comparisons in randomized Phase II trials. *J Biopharm Stat*, 19(3):456–468, 2009.

[9] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). ICH topic E8 General Considerations for Clinical Trials, 1998. Available online.

[10] US Department of Health and Human Service (FDA), Center for Biologics Evaluation and Research. Development and licensure of vaccines to prevent COVID-19, 2020. Available online.

[11] World Health Organization. Coronavirus disease (COVID-19) pandemic. `https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/novel-coronavirus-2019-ncov`.

[12] Graham BC. Rapid COVID-19 vaccine development. *Science*, 368(6494):945–946, 2020.

[13] Schwartz JL. Evaluating and deploying covid-19 vaccines – the importance of transparency, scientific integrity, and public trust. *N Engl J Med*, 383(18):1703–1705, 2020.

[14] US Food and Drug Administration. COVID-19-related guidance documents for industry, FDA staff, and other stakeholders. https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-related-guidance-documents-industry-fda-staff-and-other-stakeholders, 2020.

[15] European Medicines Agency. Guidance for medicine developers and other stakeholders on COVID-19. https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19/guidance-medicine-developers-other-stakeholders-covid-19, 2020.

[16] European Medicines Agency. EMA initiatives for acceleration of development support and evaluation procedures for COVID-19 treatments and vaccines, 2020. Available online.

[17] US Food and Drug Administration. Coronavirus treatment acceleration program (CTAP). https://www.fda.gov/drugs/coronavirus-covid-19-drugs/coronavirus-treatment-acceleration-program-ctap#FDAvoices, 2020.

[18] US Department of Health and Human Service (FDA), Center for Biologics Evaluation and Research. Emergency use authorization for vaccines to prevent COVID-19, 2020. Available online.

[19] Meier C, Rodriguez A, Steene A, and Bädeker M. How the pandemic is redefining clinical development. https://www.bcg.com/it-it/publications/2020/how-the-pandemic-is-redefining-clinical-development, September 2020.

[20] Senn S. *Statistical Issues in Drug Development*. John Wiley & Sons, Ltd, 2007.

[21] Peterson JJ, Snee RD, McAllister PR, Schofield TL, and Carella AJ. Statistics in pharmaceutical development and manufacturing. *J Qual Technol*, 41(2):111–134, 2009.

[22] Kuhn M., Yates P., and Hyde C. Statistical methods for drug discovery. In *Nonclinical Statistics for Pharmaceutical and Biotechnology Industries*, pages 53–81. Springer International Publishing, Cham, 2016.

[23] Grieve AP. Do statisticians count? a personal view. *Pharm Stat*, 1(1):35–43, 2002.

[24] Zapf A, Huebner M, Rauch G, and Kieser M. What makes a biostatistician? *Stat Med*, 38(4):695–701, 2019.

[25] Bloom BR, Nowak GJ, and Orenstein W. "When will we have a vaccine?" – understanding questions and answers about covid-19 vaccination. *N Engl J Med*, 383(23):2202–2204, 2020.

[26] Rosenberger WF and Lachin JM. *Randomization in clinical trials*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd, 2015.

[27] US Department of Health and Human Service (FDA, Rockville, MD). Guidance for industry: Q8 pharmaceutical development, 2006. Available online.

[28] Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med*, 317(3):141–145, 1987.

[29] A Baldi Antognini and A Giovagnoli. *Adaptive Designs for Sequential Treatment Allocation*. CRC Biostatistics Series. Chapman & Hall, CRC Press, Boca Raton, UK, 2015.

[30] SJ Pocock. *Clinical Trials*. John Wiley & Sons, Ltd, 1983.

[31] US Department of Health and Human Service (FDA). Adaptive designs for clinical trials of drugs and biologics, 2019.

[32] Christensen R. *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer Texts in Statistics. Springer, New York, 2002.

[33] Lee JJ and Feng L. Randomized Phase II designs in cancer clinical trials: current status and future directions. *J Clin Oncol*, 23(19):4450–4457, 2005.

[34] Magaret A, Angus DC, Adhikari NKJ, Banura P, Kissoon N, Lawler JV, and Jacob ST. Design of a multi-arm randomized clinical trial with no control arm. *Contemp Clin Trials*, 46:12–17, 2016.

[35] US Department of Health and Human Service (FDA, Rockville, MD). Guidance for industry: Q9 quality risk management. *US Food and Drug Administration*, 2006. Available online.

[36] US Department of Health and Human Service (FDA, Rockville, MD). Pharmaceutical cGMPs for the 21st century: a risk-based approach. `https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/pharmaceutical-quality-21st-century-risk-based-approach-progress-report`, 2002.

[37] Rathore A and Winkle H. Quality by design for biopharmaceuticals. *Nat Biotechnol*, 27:26–34, 2009.

[38] US Department of Health and Human Service (FDA, Rockville, MD). Guidance for industry: Q10 quality systems approach to pharmaceutical cGMP regulations. *US Food and Drug Administration*, 2006. Available online.

[39] Yu LX. Pharmaceutical quality by design: product and process development, understanding, and control. *Pharm Res*, 25(4):781–791, 2008.

[40] Tomba E, Facco P, Bezzo F, and Barolo M. Latent variable modeling to assist the implementation of quality-by-design paradigms in pharmaceutical development and manufacturing: A review. *Int J Pharm*, 457(1):283 – 297, 2013.

Part I

Chapter 2

Paper A

Optimal and ethical designs for hypothesis testing in multi-arm exponential trials

Rosamarie Frieri[1] *and Maroussa Zagoraiou*[1]

[1] Department of Statistical Sciences, University of Bologna

Abstract

Multi-arm clinical trials are complex experiments which involve several objectives. The demand for unequal allocations in a multi-treatment context is growing and adaptive designs are being increasingly used in several areas of medical research. For uncensored and censored exponential responses, we propose a constrained optimization approach in order to derive the design maximizing the power of the multivariate test of homogeneity, under a suitable ethical constraint. In the absence of censoring, we obtain a very simple closed-form solution that dominates the balanced design in terms of power and ethics. Our suggestion can also accommodate delayed responses and staggered entries, and can be implemented via response adaptive rules. While other targets proposed in the literature could present an unethical behaviour, the suggested optimal allocation is frequently unbalanced by assigning more patients to the best treatment, both in the absence and presence of censoring. We evaluate the operating characteristics of our proposal theoretically and by simulations, also redesigning a real lung cancer trial, showing that the constrained optimal target guarantees very good performances in terms of ethical demands, power and estimation precision. Therefore, it is a valid and useful tool in designing clinical trials, especially oncological trials and clinical experiments for grave and novel infectious diseases, where the ethical concern is of primary importance.

2.1 Introduction

In this paper, we deal with the design of randomized multi-arm clinical trials for treatment comparisons to achieve a suitable trade-off between inferential and ethical demands. Most of the randomized clinical trials have been designed to achieve balanced allocation among the treatment groups. Equal allocation frequently maximizes the inferential precision in the estimation of the treatment effects and reflects the condition of *equipoise*, that has been widely recognized as an ethically necessary condition that should hold at the beginning of each trial.[1] However, the balanced allocation may not be efficient and could be strongly inappropriate for clinical trials, in which the ethical concern of individual care could be of crucial importance. Indeed, it is becoming increasingly common the use of unequal allocations not only for ethical reasons,[2,3] and the absolute need of true *equipoise* is object of debate.[4] Moreover, for heterogeneous treatment groups, unequal randomization often outperforms the balanced design in terms of statistical efficiency. Advantages of unequal randomization can be ramped up in multi-arm trials with the promise of shortening drug development processes.[5]

Many clinical studies for severe/fatal diseases or oncological trials have time-to-event outcomes that can be modelled with the exponential distribution; this model can be used in trials for diseases with a very fast progression to death[6,7] or in combination with a censoring scheme.[8,9] In this context, the ethical demand of maximizing patient's care becomes prominent and the choice of the design should compromise between the conflicting goals of assigning more patients to the best performing treatment(s), while preserving power. These objectives can be formalized in a constrained/combined optimization problem,[10–12] whose solution is the so-called optimal *compromise* target. This framework has been adopted for example by Tymofyeyev et al[13] for the binary model, by Biswas et al[14] for both binary and continuous outcomes and by Baldi Antognini et al[15] for the linear homoscedastic model. In general, these optimal targets depend on the unknown model parameters and, under suitable conditions, response-adaptive randomization (RAR) procedures can be implemented to approach the desired target.[16–18] Principles and a variety of advantages of adaptive designs have been recently listed by FDA.[19] However, the application of RAR procedures in survival trials presents several complexities since: *(i)* the responses cannot be observed immediately but are naturally delayed, *(ii)* censored observations may be present and *(iii)* patients' enrolment is often staggered in time.

Indeed, literature on RAR procedures for survival outcomes is quite scarce. To the best of our knowledge, Zhang and Rosenberger[20] were the first authors suggesting to design a survival trial on the basis of optimality criteria. They derived targets for two-arm trials with exponential and Weibull distribution, by minimizing an approximation of the total expected hazard, subject to power constraints. For several treatments and in absence of censoring, Zhu and Hu[21] derived analytically the optimal allocation that maximizes power for fixed weighted sample size. On the other hand, Sverdlov et al[22] introduced two optimal allocations for censored exponentially distributed outcomes, $NP1$ and $NP2$, based on non linear programming; analytical solution is available only for $NP1$, while $NP2$ can be addressed numerically. However, both these works[21,22] are based on the same constrained optimization framework proposed by Tymofyeyev et al,[13] which requires the choice of two user-selected thresholds: one of them related to a minimum percentage of allocations to each treatment (to avoid degenerate scenarios) and the other regarding the chosen efficacy measure (which is, however, a priori unknown since it depends on the model parameters). A further downside is related to the structure of the ensuing targets, since they do not always send more patients to the best treatment. To overcome these drawbacks, Baldi Antognini et al[15] proposed a new multi-purpose design strategy for the normal homoscedastic model.

This work deals with the problem of how to allocate subjects to $K \geq 2$ treatments for exponential trials. After introducing notation in Section 2.2, Section 2.3 discusses the simple set-up without censored observations. Firstly, we derive analytically the design maximizing the power of the multivariate Wald test. In absence of treatments with the same efficacy, the optimal target is a Neyman allocation involving just the clinically best and the worst treatments. Clearly, this allocation presents undesirable properties for both inference and ethics and, on this purpose, we discuss the complex issue of how to take into account patients' health in the design of a trial for more than two treatments. Therefore, we formalize a constrained optimization problem in which the power function is maximized subject to an ethical constraint on the allocation proportions, reflecting the effectiveness of the treatments. We compare the ensuing optimal target to several targets proposed in the literature and we demonstrate that it is superior to the balanced design in terms of power and ethics. Then, in Section, 2.4 we generalize the results by taking into account censored observations, also including delayed responses and staggered entries. We implement our proposals with the Doubly adaptive Biased Coin Design[23] (DBCD) to discuss their operating characteristics in several experimental settings. We also perform robustness studies to model misspecifications and we redesign the three-arm KEYNOTE-010[8] clinical trial. In Section 2.5 we conclude the paper with a discussion and future developments, while mathematical details are reported in Section 2.6.

2.2 Framework and notation

Consider a clinical trial in which patients are allocated sequentially to $K \geq 2$ treatments and let Y_{ij} be the response of the j-th patient assigned to the i-th treatment where Y_{ij} follows an exponential distribution with mean $\theta_i \in \mathbb{R}^+$, for $i = 1, \dots, K$. Let $\mathbf{N}_n = (N_{1n}, \dots, N_{Kn})^\top$ be the random allocation vector, whose i-th component is the number of patients assigned to treatment i up to step n, where $n = \mathbf{N}_n^\top \mathbf{1}_K$ and $\mathbf{1}_K$ is the K-dimensional vector of ones. Let $\pi_{in} = N_{in}/n$, then $\boldsymbol{\pi}_n = (\pi_{1n}, \dots, \pi_{Kn})^\top$ is the vector of allocation proportions such that $\boldsymbol{\pi}_n^\top \mathbf{1}_K = 1$.

After n steps, letting $\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_{1n}, \dots, \hat{\theta}_{Kn})^\top$ be the vector of the MLEs of the treatment effects, by a well-known result $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}$ and $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathrm{N}(\mathbf{0}_K, \mathbf{M}^{-1})$, where $\mathbf{M} = \mathrm{diag}\left(\pi_{in}\theta_i^{-2}\right)_{i=1,\dots,K}$ is the Fisher information matrix and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)^\top$. In multi-arm trials, the inferential interest is usually focused on the contrasts, so let us define $\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\theta}$ where $\mathbf{A} = [\mathbf{1}_{K-1} | -\mathbf{I}_{K-1}]$ and $\mathbf{I}_{K-1}$ is the $(K-1)$-dim identity matrix so that $\boldsymbol{\gamma} = (\theta_1 - \theta_2, \dots, \theta_1 - \theta_K)^\top$. By denoting with $\hat{\boldsymbol{\gamma}}_n = \mathbf{A}\hat{\boldsymbol{\theta}}_n$ the corresponding MLE, it is known that $\hat{\boldsymbol{\gamma}}_n$ is strongly consistent and asymptotically normal with $\sqrt{n}(\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}) \xrightarrow{d} \mathrm{N}(\mathbf{0}_{K-1}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{M}^{-1}\mathbf{A}^\top$.

We define the target as the desirable treatment allocation proportion $\boldsymbol{\rho} = (\rho_1, \dots, \rho_K)^\top$, where $\boldsymbol{\rho}^\top \mathbf{1}_K = 1$ and $\rho_i \geq 0$, for $i = 1, \dots, K$. If the latter inequality is strict, namely if $\rho_i > 0$ for all $i = 1, \dots, K$, $\boldsymbol{\rho}$ is called non degenerate. The target can be seen, in a finite set-up, as the actual desired proportion of treatment assignments. Otherwise, it can be found as a limit, for increasing n, to which the allocation proportion should ideally converge. In general, the targets could depend on the unknown parameters and, under suitable conditions, RAR procedures can be carried out to sequentially estimate the model parameters, and then force the assignments to asymptotically approach the chosen target.[17,18]

Without loss of generality, in this work we assume that higher values for the response are more desirable and we let $\theta_1 \geq \theta_2 \geq \dots \geq \theta_K$, i.e. the best treatment is indicated with label 1 and the worst one with label K, admitting also groups of treatments with the same efficacy. This assumption

is not restrictive since it is simply a label-coding choice. Indeed, the ordering of treatments is a priori unknown, but for RAR rules the treatment effects are estimated step by step and then their ranking is sequentially updated. Clearly, the contrasts can be defined with respect to any treatment (not necessary the best one) by re-defining the matrix $\mathbf{A}$. For instance, if we set $\mathbf{A} = [\mathbf{I}_{K-1}| - \mathbf{1}_{K-1}]$, then $\boldsymbol{\gamma} = (\theta_1 - \theta_K, \dots, \theta_{K-1} - \theta_K)^\top$ so that the contrasts are defined with respect to the worst performing treatment.

2.3 Optimal allocations for hypothesis testing for the exponential model

In this section we derive optimal targets for testing the hypothesis of homogeneity among treatment effects. Such overall null-hypothesis is a milestone in the statistical literature and it is the first stage of multiple comparison techniques for many stepwise procedures[24,25] (see also Section 2.5). We then compare the ensuing target with the optimal allocations for the exponential model proposed in the literature both theoretically and numerically.

2.3.1 Single-objective optimal allocation for hypothesis testing

Let us consider the problem of testing the null-hypothesis of equality among the treatment effects,

$$\begin{cases} \mathrm{H}_0 : \boldsymbol{\gamma} = \mathbf{0}_{K-1} \\ \mathrm{H}_1 : \boldsymbol{\gamma} \neq \mathbf{0}_{K-1}, \end{cases}$$

where $\mathbf{0}_{K-1}$ is the $K-1$ dimensional vector of zeros. After n steps, let $\hat{\mathbf{M}}_n$ and $\hat{\boldsymbol{\Sigma}}_n = \mathbf{A}\hat{\mathbf{M}}_n^{-1}\mathbf{A}^\top$ be consistent estimators of $\mathbf{M}$ and $\boldsymbol{\Sigma}$, respectively. Under H_0, Wald's statistic $W_n = n \cdot \hat{\boldsymbol{\gamma}}_n^\top \, \hat{\boldsymbol{\Sigma}}_n^{-1} \, \hat{\boldsymbol{\gamma}}_n \xrightarrow{d} \chi^2_{K-1}(0)$, while under H_1, $W_n \xrightarrow{d} \chi^2_{K-1}(n\phi)$, where $\chi^2_{K-1}(n\phi)$ is a chi-squared r.v. with $K-1$ degrees of freedom and non centrality parameter (NCP) $n\phi$, where $\phi = \phi(\boldsymbol{\pi}) = \boldsymbol{\gamma}^\top \, \boldsymbol{\Sigma}^{-1} \, \boldsymbol{\gamma}$ is given by (see for example Zhu and Hu[21])

$$\phi(\boldsymbol{\pi}) = \sum_{i=1}^{K} \left(\frac{\theta_1 - \theta_i}{\theta_i} \right)^2 \pi_i - \frac{1}{\sum_{i=1}^{K} \frac{\pi_i}{\theta_i^2}} \left(\sum_{i=1}^{K} \frac{\theta_1 - \theta_i}{\theta_i^2} \, \pi_i \right)^2 . \tag{2.3.1}$$

For every sample size, the power of Wald's homogeneity test monotonically increases as ϕ grows; as a shorthand we shall often refer to ϕ as the NCP. In the next Theorem, following a similar set-up to Tymofyeyev et al,[13] we derive the allocation proportion maximizing the NPC.

Theorem 1. *The target allocation $\tilde{\boldsymbol{\rho}} = (\tilde{\rho}_1, \dots, \tilde{\rho}_K)^\top$ maximizing the power of Wald's test is such that $\phi(\tilde{\boldsymbol{\rho}}) = \left(\frac{\theta_1 - \theta_K}{\theta_1 + \theta_K} \right)^2$. Given $\theta_1 = \cdots = \theta_j > \theta_{j+1} \geq \cdots \geq \theta_h = \cdots = \theta_K$, with $1 \leq j < h \leq K$,*

(i) if $\theta_{j+1} > \theta_h$, then every allocation $\tilde{\boldsymbol{\rho}}$ such that $\sum_{i=1}^{j} \tilde{\rho}_i = \frac{\theta_1}{\theta_1 + \theta_K}$, $\tilde{\rho}_{j+1}, \dots, \tilde{\rho}_{h-1} = 0$ and $\sum_{i=h}^{K} \tilde{\rho}_i = \frac{\theta_K}{\theta_1 + \theta_K}$ is optimal;

(ii) if $\theta_{j+1} = \theta_h$ i.e. in presence of only two clusters of treatments, every allocation $\tilde{\boldsymbol{\rho}}$ such that $\sum_{i=1}^{j} \tilde{\rho}_i = \frac{\theta_1}{\theta_1 + \theta_K} = 1 - \sum_{i=j+1}^{K} \tilde{\rho}_i$ is optimal.

Proof. See Appendix 2.6.1. □

The targets in *(i)* are degenerate, since every $\tilde{\boldsymbol{\rho}}$ provides at least an empty treatment arm. In particular, in presence of a single superior and inferior treatments ($j = 1$ and $h = K$),

$$\tilde{\boldsymbol{\rho}} = \left(\frac{\theta_1}{\theta_1 + \theta_K}, 0, \ldots, 0, \frac{\theta_K}{\theta_1 + \theta_K} \right)^\top \tag{2.3.2}$$

i.e. it is a generalization of Neyman allocation involving just the best and the worst treatments, not collecting informations on the intermediates. Target in (2.3.2) is always a possible solution for both *(i)* and *(ii)*, and for $K = 2$ we retrieve the usual Neyman allocation. Notice that the only non degenerate optimal targets $\tilde{\boldsymbol{\rho}}$ are those obtained under scenario (ii) of Theorem 1.

Example 1. *If $\boldsymbol{\theta} = (4, 4, 4, 1)^\top$, then $\tilde{\boldsymbol{\rho}} = \left(\frac{4}{5}, 0, 0, \frac{1}{5}\right)^\top$ is optimal with $\phi(\tilde{\boldsymbol{\rho}}) = \frac{9}{25}$. Moreover, every combination of $\tilde{\rho}_1, \tilde{\rho}_2$ and $\tilde{\rho}_3$ such that $\sum_{i=1}^3 \tilde{\rho}_i = \frac{4}{5}$ is optimal, like e.g., $\left(\frac{2}{5}, \frac{2}{5}, 0, \frac{1}{5}\right)^\top$ or $\left(\frac{2}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)^\top$.*

Remark 2.3.1. *For trials comparing $K > 2$ treatments, the definition of the ethical issue is highly debated and controversial. For example, the trial may be designed with the requirement of maximizing patients' benefit by a) maximizing the number of subjects receiving the superior treatment(s) or b) minimizing the number of patients treated with the inferior arm(s). While for $K = 2$ these two ethical paradigms are equivalent, in the case of multi-arm trials the implementation of a) does not necessarily satisfy b) and vice versa; moreover, even if a) and b) hold simultaneously, the conclusions may be questionable. For instance, under the-larger-the-better scenario and assuming $\theta_1 > \theta_2 > \theta_3 > \theta_4$, the target $\left(\frac{7}{16}, \frac{2}{16}, \frac{6}{16}, \frac{1}{16}\right)^\top$ complies with a) and b), but how it can be considered as ethical? Wouldn't it be the target $\left(\frac{7}{16}, \frac{6}{16}, \frac{2}{16}, \frac{1}{16}\right)^\top$ more desirable for patients' health? In this paper, a target will be considered as ethical if its components are ordered according to the magnitude of the treatment effects (this definition was mentioned in Theorem 1 of Sverdlov et al[22]). In this way not only a) and b) hold, but the ranking of the ρ_is reflects the efficacy of $\boldsymbol{\theta}$, i.e. $\rho_i \geq \rho_{i+1} \iff \theta_i \geq \theta_{i+1}$ for all $i = 1, \ldots, K-1$. Clearly, for $K = 2$ the Neyman allocation for the exponential model is ethical since $\rho_1 \geq \rho_2 \iff \theta_1 \geq \theta_2$.*

Despite the optimal unconstrained design $\tilde{\boldsymbol{\rho}}$ maximizes power, it presents undesirable characteristics from both ethical and inferential perspective. In general, this target does not assign patients to the intermediate treatments, giving unreliable variance of the estimate of model parameters. Moreover, $\tilde{\boldsymbol{\rho}}$ is not attractive from an ethical point of view, since it always allocates a fraction of patients to the worst treatment arm.

2.3.2 Multi-objective optimal allocation for hypothesis testing

In this section we introduce the multi-purpose optimal target $\boldsymbol{\rho}^C$ maximizing the NCP subject to an ethical constraint reflecting the order among treatments. Specifically, in the following Theorem we derive the closed form solution of the constrained optimization problem

$$\begin{cases} \max \phi(\boldsymbol{\rho}) \\ \text{s.t.} \quad \rho_i \geq \rho_{i+1} \text{ for } i = 1, \ldots, K-1 \quad \text{and} \quad \sum_{i=1}^K \rho_i = 1. \end{cases} \tag{2.3.3}$$

Theorem 2. *Given $\theta_1 = \cdots = \theta_j > \theta_{j+1} \geq \cdots \geq \theta_h = \cdots = \theta_K$ ($1 \leq j < h \leq K$), let us define*

$$x = \frac{\frac{1}{\theta_1} \sum_{i=1}^K \left(\frac{1}{\theta_i} - \frac{1}{\theta_1}\right)^2}{\left[\sum_{i=1}^K \left(\frac{1}{\theta_i} - \frac{1}{\theta_1}\right)\right] \left[\sum_{i=1}^K \left(\frac{1}{\theta_i^2} - \frac{1}{\theta_1^2}\right)\right]}. \tag{2.3.4}$$

If $x > K^{-1}$, then the balanced design $\boldsymbol{\rho}^B = \left(\frac{1}{K}, \ldots, \frac{1}{K}\right)^\top$ is optimal. Otherwise, when $x \leq K^{-1}$, the solution of (2.3.3) is $\boldsymbol{\rho}^C = (\rho_1^C, \ldots, \rho_K^C)^\top$ (with $\rho_i^C \geq \rho_{i+1}^C$ for $i = 1, \ldots, K-1$ and $\mathbf{1}_K^\top \boldsymbol{\rho}^C = 1$), where

(i) if $\theta_{j+1} > \theta_h$, then $\rho_{j+1}^C = \ldots = \rho_K^C = x$ (and, clearly, $\sum_{i=1}^j \rho_i^C = 1 - (K-j)x$);

(ii) in the case of just two clusters of treatments, namely when $\theta_{j+1} = \theta_h$, then $\sum_{i=1}^j \rho_i^C = 1 - (K-j)x = \theta_1/(\theta_1 + \theta_K) = 1 - \sum_{i=j+1}^K \rho_i^C$.

Proof. See Appendix 2.6.2. □

The constrained optimal target presents very appealing properties: $\boldsymbol{\rho}^C$ has a very simple form and it is non degenerate, so that there is no need to fix a lower bound for the treatment allocation proportion to the worst treatment(s). Furthermore, it assigns the same proportion of patients to all the treatment arms or it skews the allocations in favour of the best performing treatment(s).

The behaviour of $\boldsymbol{\rho}^C$ for $K = 3$ and several experimental settings is displayed in Table 2.1. The optimal constrained target allocates a higher proportion of subjects to the superior treatment and this proportion increases as the difference between θ_1 and θ_2 grows (Table 2.1a). Moreover, ρ_1^C increases as the magnitude of the superior treatment increases (Table 2.1b). Note that, in the last scenario of Table 2.1a $x = 1/6$, so that all the targets with $\rho_2^C + \rho_3^C = 1/3$ are optimal.

Table 2.1: Behaviour of the optimal constrained target.

(a) Fixed θ_1 and θ_3 and decreasing θ_2.

$\boldsymbol{\theta}$	${\rho_1}^C$	${\rho_2}^C$	${\rho_3}^C$
$(10, 9, 5)^\top$	0.436	0.282	0.282
$(10, 7, 5)^\top$	0.590	0.205	0.205
$(10, 5, 5)^\top$	0.667	0.1665	0.1665

(b) Fixed θ_2 and θ_3 and increasing θ_1.

$\boldsymbol{\theta}$	${\rho_1}^C$	${\rho_2}^C$	${\rho_3}^C$
$(10, 8, 4)^\top$	0.546	0.227	0.227
$(15, 8, 4)^\top$	0.706	0.147	0.147
$(20, 8, 4)^\top$	0.774	0.113	0.113

2.3.3 Comparison of optimal targets for the exponential model

The aim of this section is to compare, both theoretically and with numerical examples, the constrained optimal target with those proposed in the literature for exponential outcomes. By taking into account the well known A-optimal criterion, Sverdlov and Rosenberger[26] derived the optimal allocation $\boldsymbol{\rho}^A$ minimizing the trace of $\boldsymbol{\Sigma}$,

$$\rho_1^A = \frac{\theta_1\sqrt{K-1}}{\theta_1\sqrt{K-1} + \sum_{k=2}^K \theta_k} \quad \text{and} \quad \rho_i^A = \frac{\theta_i}{\theta_1\sqrt{K-1} + \sum_{k=2}^K \theta_k} \text{ for } i = 2, \ldots, K,$$

while Wong and Zhu[27] found the D-optimal allocation $\boldsymbol{\rho}^D$ minimizing the determinant of $\boldsymbol{\Sigma}$ (this target is not available in closed form but is derived as the unique root of a non linear system of equations). Although for linear contrasts these criteria are usually called A_A and D_A, for the sake of notation, we simply refer to them as $\boldsymbol{\rho}^A$ and $\boldsymbol{\rho}^D$.

For exponential outcomes without censoring, Zhu and Hu[21] derived the optimal allocation $\boldsymbol{\rho}^Z$ maximizing the NCP for fixed n, subject to the constraint that $\rho_i^Z \geq T$ for all $i = 1, \ldots, K$ where

$T \in [0, K^{-1}]$ is a user-selected threshold. Nevertheless, no guidelines are provided by the authors to help the choice of T; moreover, Theorem 1 of Zhu and Hu[21] does not include trials in which the treatments are grouped into two clusters (like, e.g., $K = 3$ with $\theta_1 = \theta_2 > \theta_3$).

To measure the effectiveness of such targets in terms of both statistical and ethical performances, we evaluate:

1. the efficiency in terms of power of a given target $\boldsymbol{\rho}$ as $E_\phi(\boldsymbol{\rho}) = \frac{\phi(\boldsymbol{\rho})}{\phi(\tilde{\boldsymbol{\rho}})}$ where $\tilde{\boldsymbol{\rho}}$ is the unconstrained optimal target defined in Theorem 1;

2. D_A and A_A efficiency, defined by $E_{D_A}(\boldsymbol{\rho}) = \left[\frac{|\mathbf{A}\mathbf{M}^{-1}(\boldsymbol{\rho}^D)\mathbf{A}^\top|}{|\mathbf{A}\mathbf{M}^{-1}(\boldsymbol{\rho})\mathbf{A}^\top|}\right]^{\frac{1}{K-1}}$ and $E_{A_A}(\boldsymbol{\rho}) = \frac{tr\left(\mathbf{A}\mathbf{M}^{-1}(\boldsymbol{\rho}^A)\mathbf{A}^\top\right)}{tr\left(\mathbf{A}\mathbf{M}^{-1}(\boldsymbol{\rho})\mathbf{A}^\top\right)}$, respectively. Accordingly, values of $E_{D_A}(\boldsymbol{\rho})$ or $E_{A_A}(\boldsymbol{\rho})$ close to 1 point out that $\boldsymbol{\rho}$ has similar performance in terms of estimation precision to $\boldsymbol{\rho}^D$ or $\boldsymbol{\rho}^A$;

3. besides the ethical measures provided by ρ_1 and ρ_K representing the assignments to the best and worst treatments, to assess a global ethical performance we compute $E_e(\boldsymbol{\rho}) = \frac{\sum_{i=1}^K \theta_i \rho_i}{\theta_1}$, that is the ratio between the total expected responses under a given target $\boldsymbol{\rho}$ and the total expected outcomes obtained by assigning all the subjects to the best treatment.

In the comparisons we also take into account the balanced design as a benchmark. However, equal allocation may be suboptimal in trials with heterogeneous variance.[3] In the following Theorem we demonstrate that the constrained optimal target proposed in this paper has always better performance in terms of power and ethical efficiency with respect to the balanced design.

Theorem 3. *The optimal constrained target $\boldsymbol{\rho}^C$ dominates the balanced allocation $\boldsymbol{\rho}^B$, namely $E_\phi(\boldsymbol{\rho}^C) \geq E_\phi(\boldsymbol{\rho}^B)$ and $E_e(\boldsymbol{\rho}^C) \geq E_e(\boldsymbol{\rho}^B)$, simultaneously.*

Proof. See Appendix 2.6.3. □

Remark 2.3.2. *While $\boldsymbol{\rho}^D$ satisfies the property of Remark 2.3.1 (shown by Sverdlov et al[22]), i.e. $\theta_1 \geq ... \geq \theta_K \Leftrightarrow \rho_1^D \geq ... \geq \rho_K^D$, targets $\boldsymbol{\rho}^Z$ and $\boldsymbol{\rho}^A$ could present a controversial behaviour in terms of ethics. In general, $\boldsymbol{\rho}^A$ assigns more subjects to the reference treatment (that does not always coincide with the best one, as in our framework). For instance, if $\boldsymbol{\theta} = (25, 29, 30)^\top$ then $\boldsymbol{\rho}^A = (0.375, 0.307, 0.318)^\top$, so that more subjects receive the worst treatment. Moreover, also $\boldsymbol{\rho}^Z$ is unethical since it does not always satisfy a) and/or b) of Remark 2.3.1. Indeed, in the same scenario, by setting $T = 0.2$, we obtain $\boldsymbol{\rho}^Z = (0.425, 0.2, 0.375)^\top$, so it assigns the higher proportion of patients to the less effective treatment.*

Further numerical comparisons are reported in Tables 2.2 and 2.3 for $K = 3$ and 5 and several values of $\boldsymbol{\theta}$. We have included $\tilde{\boldsymbol{\rho}}$ for completeness, even if it is strongly inadequate, as widely discussed in Section 2.3.1. Let us first consider the case of $K = 3$. As far as power is concerned, the proposed constrained target $\boldsymbol{\rho}^C$ always presents higher power than the competitors except in the scenario $\boldsymbol{\theta} = (8, 5, 4)^\top$, with a loss of 0.7% with respect to $\boldsymbol{\rho}^Z$. In the remaining cases, the gain of $\boldsymbol{\rho}^C$ in terms of power efficiency on the second best is up to 3.4%. With regard to the ethical concern, $\boldsymbol{\rho}^C$ presents values closest to one, except in the first scenario, in which $\boldsymbol{\rho}^A$ has essentially equivalent ethical efficiency. This latter target represents the second best choice in terms of ethical demand. In general, the D_A efficiency of $\boldsymbol{\rho}^C$ is higher for narrower values of the relative differences among θ_1 and θ_2 and it is higher than 77% in all these settings. Very good performances in terms of A_A efficiency are achieved by the constrained

target, with values always greater than 90.6%. Note that $\boldsymbol{\rho}^Z$ is nearly constant as the vector of treatment effects changes. Indeed, the allocation proportion to the worst treatments is equal, or very close to T. The results for $K = 5$ treatments enhance the value of our proposal and similar considerations to the case of $K = 3$ treatments still hold. Note that the gain in terms of power efficiency reaches the 19% with respect to other designs (see the last scenario). The undesirable behaviour in terms of patients' benefit of $\boldsymbol{\rho}^Z$, discussed in Remark 2.3.2, holds for both $K = 3$ and $K = 5$.

Table 2.2: Comparison of optimal targets for $K = 3$ and 5 treatments in different experimental scenarios with respect to power efficiency $E_\phi(\boldsymbol{\rho})$, ethical efficiency $E_e(\boldsymbol{\rho})$, D_A efficiency $E_{D_A}(\boldsymbol{\rho})$, and A_A efficiency $E_{A_A}(\boldsymbol{\rho})$.

$\boldsymbol{\theta}^\top$	$\boldsymbol{\rho}$	$E_\phi(\boldsymbol{\rho})$	$E_e(\boldsymbol{\rho})$	$E_{D_A}(\boldsymbol{\rho})$	$E_{A_A}(\boldsymbol{\rho})$
$(30, 20, 8)$	$\boldsymbol{\rho}^A = (0.602, 0.284, 0.114)^\top$	0.761	0.822	0.933	1
	$\boldsymbol{\rho}^D = (0.441, 0.385, 0.174)^\top$	0.765	0.744	1	0.905
	$\tilde{\boldsymbol{\rho}} = (0.789, 0, 0.211)^\top$	1	0.845	$\to 0$	$\to 0$
	$\boldsymbol{\rho}^C = (0.664, 0.168, 0.168)^\top$	0.889	0.821	0.836	0.906
	$\boldsymbol{\rho}^B$	0.740	0.644	0.903	0.730
	$\boldsymbol{\rho}^Z = (0.591, 0.200, 0.209)^\top$	0.881	0.780	0.888	0.927
$(30, 10, 8)$	$\boldsymbol{\rho}^A = (0.702, 0.165, 0.133)^\top$	0.868	0.792	0.864	1
	$\boldsymbol{\rho}^D = (0.464, 0.295, 0.241)^\top$	0.657	0.627	1	0.815
	$\tilde{\boldsymbol{\rho}} = (0.789, 0, 0.211)^\top$	1	0.845	$\to 0$	$\to 0$
	$\boldsymbol{\rho}^C = (0.768, 0.116, 0.116)^\top$	0.900	0.839	0.770	0.973
	$\boldsymbol{\rho}^B$	0.501	0.533	0.954	0.620
	$\boldsymbol{\rho}^Z = (0.600, 0.200, 0.200)^\top$	0.807	0.720	0.953	0.956
$(12, 5, 4)$	$\boldsymbol{\rho}^A = (0.653, 0.193, 0.154)^\top$	0.838	0.785	0.899	1
	$\boldsymbol{\rho}^D = (0.449, 0.303, 0.248)^\top$	0.668	0.658	1	0.856
	$\tilde{\boldsymbol{\rho}} = (0.750, 0, 0.250)^\top$	1	0.833	$\to 0$	$\to 0$
	$\boldsymbol{\rho}^C = (0.726, 0.137, 0.137)^\top$	0.872	0.828	0.805	0.968
	$\boldsymbol{\rho}^B$	0.535	0.583	0.962	0.683
	$\boldsymbol{\rho}^Z = (0.600, 0.200, 0.200)^\top$	0.822	0.750	0.942	0.985
$(8, 5, 4)$	$\boldsymbol{\rho}^A = (0.557, 0.246, 0.197)^\top$	0.760	0.809	0.944	1
	$\boldsymbol{\rho}^D = (0.411, 0.321, 0.268)^\top$	0.669	0.746	1	0.919
	$\tilde{\boldsymbol{\rho}} = (0.667, 0, 0.333)^\top$	1	0.834	$\to 0$	$\to 0$
	$\boldsymbol{\rho}^C = (0.628, 0.186, 0.186)^\top$	0.801	0.837	0.876	0.973
	$\boldsymbol{\rho}^B$	0.603	0.708	0.979	0.814
	$\boldsymbol{\rho}^Z = (0.555, 0.200, 0.245)^\top$	0.808	0.802	0.931	0.980

$\boldsymbol{\rho}^A$, A_A optimal design; $\boldsymbol{\rho}^D$, D_A optimal design; $\tilde{\boldsymbol{\rho}}$, optimal unconstrained target; $\boldsymbol{\rho}^C$, optimal constrained target; $\boldsymbol{\rho}^B$, balanced design; $\boldsymbol{\rho}^Z$, Zhu and Hu[21] optimal target (with minimum allocation proportion for each arm T equal to 0.2 for $K = 3$).

Table 2.3: Comparisons of optimal targets for $K = 5$ treatments in different scenarios with respect to power efficiency $E_\phi(\boldsymbol{\rho})$, ethical efficiency $E_e(\boldsymbol{\rho})$, D_A efficiency $E_{D_A}(\boldsymbol{\rho})$, and A_A efficiency $E_{A_A}(\boldsymbol{\rho})$.

$\boldsymbol{\theta}^\top$	$\boldsymbol{\rho}$	$E_\phi(\boldsymbol{\rho})$	$E_e(\boldsymbol{\rho})$	$E_{D_A}(\boldsymbol{\rho})$	$E_{A_A}(\boldsymbol{\rho})$
$(12, 11, 10, 5, 3)$	$\boldsymbol{\rho}^A = (0.453, 0.208, 0.189, 0.094, 0.057)^\top$	0.660	0.854	0.858	1
	$\boldsymbol{\rho}^D = (0.235, 0.232, 0.229, 0.182, 0.123)^\top$	0.719	0.745	1	0.775
	$\tilde{\boldsymbol{\rho}} = (0.800, 0, 0, 0, 0.200)^\top$	1	0.850	$\to 0$	$\to 0$
	$\boldsymbol{\rho}^C = (0.540, 0.115, 0.115, 0.115, 0.115)^\top$	0.810	0.818	0.791	0.856
	$\boldsymbol{\rho}^B$	0.723	0.683	0.966	0.676
	$\boldsymbol{\rho}^Z = (0.373, 0.150, 0.150, 0.150, 0.177)^\top$	0.796	0.742	0.912	0.869
$(12, 10, 8, 6, 4)$	$\boldsymbol{\rho}^A = (0.462, 0.192, 0.154, 0.115, 0.077)^\top$	0.574	0.808	0.865	1
	$\boldsymbol{\rho}^D = (0.231, 0.224, 0.211, 0.189, 0.144)^\top$	0.562	0.701	1	0.763
	$\tilde{\boldsymbol{\rho}} = (0.750, 0, 0, 0, 0.250)^\top$	1	0.833	$\to 0$	$\to 0$
	$\boldsymbol{\rho}^C = (0.548, 0.113, 0.113, 0.113, 0.113)^\top$	0.695	0.812	0.783	0.912
	$\boldsymbol{\rho}^B$	0.577	0.667	0.983	0.683
	$\boldsymbol{\rho}^Z = (0.348, 0.150, 0.150, 0.150, 0.202)^\top$	0.683	0.716	0.927	0.882
$(12, 8, 7, 6, 3)$	$\boldsymbol{\rho}^A = (0.500, 0.167, 0.146, 0.125, 0.062)^\top$	0.548	0.774	0.840	1
	$\boldsymbol{\rho}^D = (0.236, 0.221, 0.213, 0.202, 0.128)^\top$	0.526	0.640	1	0.718
	$\tilde{\boldsymbol{\rho}} = (0.800, 0, 0, 0, 0.200)^\top$	1	0.850	$\to 0$	$\to 0$
	$\boldsymbol{\rho}^C = (0.612, 0.097, 0.097, 0.097, 0.097)^\top$	0.716	0.805	0.716	0.898
	$\boldsymbol{\rho}^B$	0.565	0.600	0.973	0.628
	$\boldsymbol{\rho}^Z = (0.363, 0.150, 0.150, 0.150, 0.187)^\top$	0.674	0.672	0.918	0.877

$\boldsymbol{\rho}^A$, A_A optimal design; $\boldsymbol{\rho}^D$, D_A optimal design; $\tilde{\boldsymbol{\rho}}$, optimal unconstrained target; $\boldsymbol{\rho}^C$, optimal constrained target; $\boldsymbol{\rho}^B$, balanced design; $\boldsymbol{\rho}^Z$, Zhu and Hu[21] optimal target (with minimum allocation proportion for each arm T equal to 0.15).

In Figure 2.1 we show the behaviour of the above-mentioned efficiency measures for three treatments where θ_1 varies between 10 and 30, $\theta_2 = 9$ and $\theta_3 = 8$. The constrained target $\boldsymbol{\rho}^C$ dominates the other allocations in terms of power and ethics for values of $\theta_1 \geq 15$, whereas for smaller values of θ_1 the maximum loss with respect to $\boldsymbol{\rho}^Z$ is 8.8% for $\theta_1 = 10$. The ethical efficiency of $\boldsymbol{\rho}^C$ is increasing in θ_1 for $\theta_1 > 20$ (the same behaviour can be observed only for $\boldsymbol{\rho}^A$ when $\theta_1 > 22$, whereas $\boldsymbol{\rho}^D$, $\boldsymbol{\rho}^B$ and $\boldsymbol{\rho}^Z$ are always decreasing in θ_1), and the average gain of $\boldsymbol{\rho}^C$ on the second best ($\boldsymbol{\rho}^A$) is 3.5%. The best performance in terms of D_A efficiency is given by $\boldsymbol{\rho}^B$, while $\boldsymbol{\rho}^Z$ is the second best for high values of θ_1. On the other hand, $E_{D_A}(\boldsymbol{\rho}^C)$ and $E_{D_A}(\boldsymbol{\rho}^A)$ decrease as θ_1 increases, having lower values with respect to the other targets. Both $\boldsymbol{\rho}^Z$ and $\boldsymbol{\rho}^C$ present A_A efficiency close to 1 (in particular highest values are achieved by $\boldsymbol{\rho}^Z$ for $\theta_1 \in [13, 26]$ while in the remaining configurations of parameters the highest values are reached by $\boldsymbol{\rho}^C$). It is worth notice that $E_{A_A}(\boldsymbol{\rho}^C)$ is nearly constant w.r.t. θ_1 while A_A efficiency tends to decrease for the remaining targets.

The theoretical properties as well as the comparisons in Table 2.2 and Figure 2.1 show that the optimal constrained target $\boldsymbol{\rho}^C$ guarantees very good performance in terms of power, estimation precision and ethical concerns.

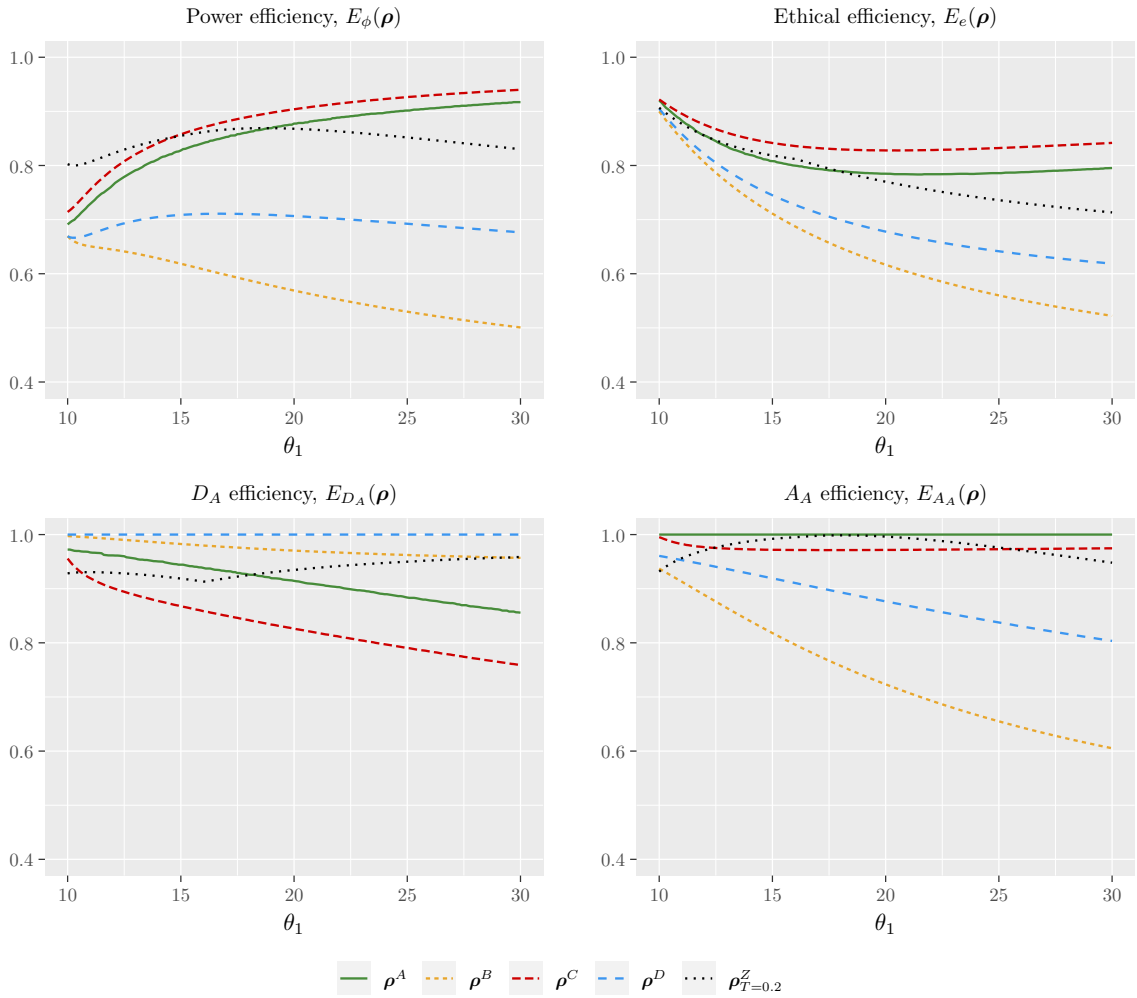


Figure 2.1: Comparisons of optimal allocations with respect to several measures of efficiency for $\boldsymbol{\theta} = (\theta_1, 9, 8)^\top$ as θ_1 varies from 10 to 30. $\boldsymbol{\rho}^A$, A_A optimal design; $\boldsymbol{\rho}^B$, balanced design; $\boldsymbol{\rho}^C$, optimal constrained target; $\boldsymbol{\rho}^D$, D_A optimal design; $\boldsymbol{\rho}^Z$, Zhu and Hu[21] optimal target (with minimum allocation proportion for each arm T equal to 0.2).

2.3.4 Implementing the optimal target with RAR procedures

In this section we apply the DBCD[23] to implement the constrained target, under the assumption that responses are available immediately (see e.g. Shulz et al[7]). This procedure starts with a sample of n_0 patients allocated to the K treatments (usually by restricted randomization) to obtain initial estimates of model parameters. After $j > n_0$ subjects are assigned and the responses are observed, the unknown parameters are estimated by $\hat{\boldsymbol{\theta}}_j = (\hat{\theta}_{1j}, \ldots, \hat{\theta}_{Kj})^\top$ and they are used to compute the estimated target $\hat{\boldsymbol{\rho}}_j = (\hat{\rho}_{1j}, \ldots, \hat{\rho}_{Kj})^\top$. Then, the $(j+1)$th patient is randomized to treatment i with probability $\Psi_{j+1,i} = \hat{\rho}_{ij} \left(\frac{\hat{\rho}_{ij}}{\pi_{ij}}\right)^\kappa \left[\sum_{i=1}^K \hat{\rho}_{ij} \left(\frac{\hat{\rho}_{ij}}{\pi_{ij}}\right)^\kappa\right]^{-1}$, where $\kappa \in [0, +\infty)$. We apply the above-mentioned procedure to target the constrained optimal design by setting $\kappa = 2$, that represents a good trade-off between randomness and optimality.[28] We take into account the scenarios of Tables 2.1a and 2.1b for different sample sizes, where each trial has been replicated 10000 times. The first $\frac{1}{10}$ of the total sample size is assigned to the treatments with restricted randomization, then allocations become response adaptive. The results are provided in Tables 2.4 and 2.5, in which we report for every scenario the estimates $\hat{\boldsymbol{\theta}}_n$, the theoretical target, the simulated allocation proportions $\boldsymbol{\pi}_n$ (with standard deviations in square brackets) and the simulated average power (W) of Wald's test.

The sequential procedure implementing $\boldsymbol{\rho}^C$ assigns a higher proportion of subjects to the more effective treatment. Clearly, the convergence to the optimal target improves for increasing n, even though, in many experimental scenarios, good convergence is also achieved for $n = 100$ and the simulated allocation proportions are equal or really close to their target values for $n = 250$. Moreover, the convergence improves as the efficacy of the best treatment grows (see Table 2.5) and the average power exhibits values over 72% in all the considered settings, while for $n \geq 150$ it is above 88%.

In Table 2.6 we summarize the results of the simulated type I error (W_α) of Wald's test, where the nominal value was set to 0.05. Clearly, under this setting $\boldsymbol{\rho}^C = \boldsymbol{\rho}^B$ and DBCD gives simulated allocation proportions that converge to the balanced design: even for $n = 100$, the type I error rate is really close to the nominal value.

Table 2.4: Simulation results (Scenarios Table 2.1a). Estimates of model parameters $\hat{\boldsymbol{\theta}}_n$, simulated allocation proportions $\boldsymbol{\pi}_n$ (standard deviation in square brackets) and simulated average power W of Wald's test; 10000 iterations.

n	$\boldsymbol{\theta}^\top$	$\hat{\boldsymbol{\theta}}_n^\top$	$\boldsymbol{\rho}^C$	$\boldsymbol{\pi}_n^\top$		W
100	$(10, 9, 5)$	$(9.6, 8.6, 4.9)$	$(0.44, 0.28, 0.28)^\top$	$(0.44, 0.32, 0.24)$	$[.164, .140, .072]$	0.721
150		$(9.7, 8.7, 5.0)$		$(0.44, 0.32, 0.24)$	$[.143, .122, .050]$	0.885
200		$(9.8, 8.8, 5.0)$		$(0.44, 0.31, 0.25)$	$[.129, .108, .044]$	0.965
250		$(9.8, 8.8, 5.0)$		$(0.44, 0.30, 0.26)$	$[.119, .098, .041]$	0.990
100	$(10, 7, 5)$	$(9.8, 6.8, 4.9)$	$(0.60, 0.20, 0.20)^\top$	$(0.55, 0.24, 0.21)$	$[.135, .093, .062]$	0.731
150		$(9.9, 6.8, 5.0)$		$(0.56, 0.23, 0.21)$	$[.110, .072, .049]$	0.894
200		$(9.9, 6.9, 5.0)$		$(0.57, 0.22, 0.21)$	$[.093, .058, .041]$	0.960
250		$(9.9, 6.9, 5.0)$		$(0.57, 0.22, 0.21)$	$[.083, .049, .038]$	0.987
100	$(10, 5, 5)$	$(9.9, 4.9, 4.9)$	$(0.66, 0.17, 0.17)^\top$	$(0.64, 0.18, 0.18)$	$[.096, .056, .049]$	0.880
150		$(10.0, 4.9, 4.9)$		$(0.65, 0.18, 0.17)$	$[.067, .038, .034]$	0.973
200		$(10.0, 4.9, 4.9)$		$(0.66, 0.17, 0.17)$	$[.053, .030, .027]$	0.996
250		$(10.0, 5.0, 5.0)$		$(0.66, 0.17, 0.17)$	$[.044, .024, .023]$	0.999

Table 2.5: Simulation results (Scenarios Table 2.1b). Estimates of model parameters $\hat{\boldsymbol{\theta}}_n$, simulated allocation proportions $\boldsymbol{\pi}_n$ (standard deviation in square brackets) and simulated average power W of Wald's test; 10000 iterations.

n	$\boldsymbol{\theta}^\top$	$\hat{\boldsymbol{\theta}}_n^\top$	$\boldsymbol{\rho}^C$	$\boldsymbol{\pi}_n^\top$		W
100	$(10, 8, 4)$	$(9.7, 7.6, 3.9)$	$(0.54, 0.23, 0.23)^\top$	$(0.51, 0.28, 0.21)$	$[.166, .136, .062]$	0.918
150		$(9.8, 7.8, 3.9)$		$(0.52, 0.27, 0.21)$	$[.142, .113, .048]$	0.988
200		$(9.8, 7.8, 4.0)$		$(0.53, 0.25, 0.22)$	$[.123, .095, .042]$	0.998
250		$(9.9, 7.9, 4.0)$		$(0.53, 0.25, 0.22)$	$[.110, .081, .039]$	1.000
100	$(15, 8, 4)$	$(14.9, 7.6, 3.9)$	$(0.70, 0.15, 0.15)^\top$	$(0.70, 0.16, 0.14)$	$[.106, .075, .044]$	0.998
150		$(14.9, 7.8, 3.9)$		$(0.70, 0.15, 0.15)$	$[.076, .049, .034]$	1.000
200		$(14.9, 7.8, 3.9)$		$(0.70, 0.15, 0.15)$	$[.058, .035, .028]$	1.000
250		$(15.0, 7.9, 4.0)$		$(0.70, 0.15, 0.15)$	$[.049, .028, .025]$	1.000
100	$(20, 8, 4)$	$(19.9, 7.6, 3.8)$	$(0.78, 0.11, 0.11)^\top$	$(0.77, 0.12, 0.11)$	$[.073, .047, .034]$	1.000
150		$(20.0, 7.7, 3.9)$		$(0.78, 0.11, 0.11)$	$[.053, .031, .027]$	1.000
200		$(20.0, 7.9, 3.9)$		$(0.78, 0.11, 0.11)$	$[.043, .024, .022]$	1.000
250		$(20.0, 7.9, 3.9)$		$(0.78, 0.11, 0.11)$	$[.037, .020, .019]$	1.000

Table 2.6: Simulation results of type I error W_α, estimates of model parameters $\hat{\boldsymbol{\theta}}_n$ and simulated allocation proportions $\boldsymbol{\pi}_n$ (standard deviation in square brackets); 10000 iterations.

n	$\boldsymbol{\theta}^\top$	$\hat{\boldsymbol{\theta}}_n^\top$	$\boldsymbol{\pi}_n^\top$		W_α
100	$(12, 12, 12)$	$(11.6, 11.9, 11.8)$	$(0.32, 0.33, 0.35)$	$[.114, .093, .108]$	0.052
150		$(11.7, 11.9, 11.9)$	$(0.33, 0.33, 0.34)$	$[.104, .083, .093]$	0.048
200		$(11.8, 11.9, 11.9)$	$(0.33, 0.33, 0.34)$	$[.099, .078, .087]$	0.048
250		$(11.8, 12.0, 12.0)$	$(0.33, 0.33, 0.34)$	$[.094, .075, .083]$	0.046
100	$(4, 4, 4)$	$(3.9, 4.0, 3.9)$	$(0.32, 0.33, 0.35)$	$[.114, .093, .108]$	0.052
150		$(3.9, 4.0, 4.0)$	$(0.33, 0.33, 0.34)$	$[.104, .083, .093]$	0.048
200		$(3.9, 4.0, 4.0)$	$(0.33, 0.33, 0.34)$	$[.099, .078, .087]$	0.048
250		$(3.9, 4.0, 4.0)$	$(0.33, 0.33, 0.34)$	$[.094, .075, .083]$	0.046

2.4 Optimal allocations for survival trials with right-censoring

Consider now a trial in which each patient has exponentially distributed survival time subjected to a random censoring time C. Assuming that his/her censoring is independent of the outcomes (this is considered a realistic assumption in many situations as stated, e.g., by Lawless[29]) and it is the same for each treatment group, let ϵ_i be the probability that a patient belonging to the ith treatment group experiences the event of interest (death/failure). Clearly, the probability ϵ depends on the censoring scheme adopted in the study and one of the most popular[20,22] has been introduced by Latta[30] and further studied by Rosenberger and Seshaiyer.[31] This scheme can be summarized as follows: let R be the total recruitment period and D the duration of the trial, subjects arrival times are assumed to be independent and uniformly distributed in $[0, R]$, while each patient is subjected to an independent censoring time over

$(0, D)$. In this setting, for any group $i = 1, \ldots, K$

$$\epsilon_i = \epsilon(\theta_i) = 1 - \frac{\theta_i}{D} - \frac{2\theta_i^2}{RD} \exp\left\{-\frac{D}{\theta_i}\right\} - \frac{\theta_i}{D}\left(1 - 2\frac{\theta_i}{R}\right) \exp\left\{-\frac{D-R}{\theta_i}\right\} \tag{2.4.1}$$

is a monotonically non increasing function of θ_i, following the idea that the longer the expected survival time is, the smaller the probability for a patient to fail before censoring. Clearly, the likelihood for $\boldsymbol{\theta}$ is modified (see e.g. Lawless[29]) and, after n assignments, the Fisher information matrix becomes $\mathbf{M}_\epsilon = \text{diag}\left(\pi_{in}\epsilon_i\theta_i^{-2}\right)_{i=1,\ldots,K}$ and, thus, $\sqrt{n}(\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}) \xrightarrow{d} \text{N}(\mathbf{0}_{K-1}, \boldsymbol{\Sigma}_\epsilon)$, where $\boldsymbol{\Sigma}_\epsilon = \mathbf{A}\mathbf{M}_\epsilon^{-1}\mathbf{A}^\top$. Accordingly, the NCP becomes $\phi_\epsilon(\boldsymbol{\pi}) = \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\gamma}$ and, in general, it is hard to obtain a closed-form expression of both the constrained and unconstrained optimal targets. Nevertheless, solutions can be found numerically with standard optimization software (R, Matlab).

Under the censoring scheme in equation (2.4.1), Table 2.7 shows the behaviour of the unconstrained optimal target, denoted by $\tilde{\boldsymbol{\rho}}^\epsilon$. As in the uncensored set-up, this target is degenerate and presents similar drawbacks to the one of Theorem 1, both from inferential and ethical viewpoints. Therefore, we apply the same constrained optimization framework in equation (2.3.3), in which NCP is maximized under the ethical constraint reflecting the effectiveness of the treatments. Since the closed-form solution of this target, denoted by $\boldsymbol{\rho}^{C_\epsilon}$, is not available, in what follows we take into account a smoothing transformation of it in order to obtain a continuous target function implementable via DBCD. In particular, we consider the convolution of $\boldsymbol{\rho}^{C_\epsilon}$ with a Gaussian kernel (see e.g. Tymofyeyev et al[13]) with $\sigma^2 = 1$.

Table 2.8 presents the constrained optimal target $\boldsymbol{\rho}^{C_\epsilon}$ for the same values of $\boldsymbol{\theta}$ reported in Table 2.1, with $R = 55$ and $D = 96$. We display both the theoretical and the smoothed version (in italics). The target $\boldsymbol{\rho}^{C_\epsilon}$ skews the assignments to the best performing treatment arm and $\rho_1^{C_\epsilon}$ is increasing in θ_1 for fixed θ_2 and θ_3 (see Table 2.8b), whereas it is decreasing in θ_2 for fixed θ_1 and θ_3 (Table 2.8a). Furthermore, $\boldsymbol{\rho}^{C_\epsilon}$ and its smoothed version substantially coincide: only small differences, of order 10^{-3}, are present. This behaviour was also confirmed by further computations not reported here for brevity and it suggests that the smoothed target should have similar performance to $\boldsymbol{\rho}^{C_\epsilon}$. Hence, from now on we will refer to the smoothed version of the constrained target with $\boldsymbol{\rho}^{C_\epsilon}$.

Table 2.7: Unconstrained optimal target $\tilde{\boldsymbol{\rho}}^\epsilon = (\tilde{\rho}_1^\epsilon, \tilde{\rho}_2^\epsilon, \tilde{\rho}_3^\epsilon)^\top$, under the right censoring scheme in equation (2.4.1) with $R = 55$ and $D = 96$.

$\boldsymbol{\theta}$	$\tilde{\rho}_1^\epsilon$	$\tilde{\rho}_2^\epsilon$	$\tilde{\rho}_3^\epsilon$
$(30, 10, 5)^\top$	0.876	0	0.124
$(20, 10, 5)^\top$	0.815	0	0.185
$(10, 10, 5)^\top$	0.336	0.336	0.328
$(10, 7, 5)^\top$	0.673	0	0.327
$(10, 5, 5)^\top$	0.674	0.163	0.163

Table 2.8: Theoretical and smoothed (in italics) optimal constrained target for $R = 55$, $D = 96$ and $\sigma = 1$.

(a) Fixed θ_1, θ_3 and decreasing θ_2.

$\boldsymbol{\theta}$	${\rho_1}^{C_\epsilon}$	${\rho_2}^{C_\epsilon}$	${\rho_3}^{C_\epsilon}$
$(10, 9, 5)^\top$	0.444	0.278	0.278
	0.444	*0.278*	*0.278*
$(10, 7, 5)^\top$	0.594	0.203	0.203
	0.594	*0.203*	*0.203*
$(10, 5, 5)^\top$	0.672	0.164	0.164
	0.672	*0.164*	*0.164*

(b) Fixed θ_2, θ_3 and increasing θ_1.

$\boldsymbol{\theta}$	${\rho_1}^{C_\epsilon}$	${\rho_2}^{C_\epsilon}$	${\rho_3}^{C_\epsilon}$
$(10, 8, 4)^\top$	0.552	0.224	0.224
	0.550	*0.225*	*0.225*
$(15, 8, 4)^\top$	0.714	0.143	0.143
	0.714	*0.143*	*0.143*
$(20, 8, 4)^\top$	0.786	0.107	0.107
	0.782	*0.109*	*0.109*

2.4.1 Comparisons of optimal targets under an independent right censoring scheme

In the presence of censoring, the following optimal allocations have been derived in the literature. The allocation minimizing the trace of $\boldsymbol{\Sigma}_\epsilon$ can be easily derived from the general result for heteroscedastic models,[26] from which we obtained (by simply substituting the variance with the appropriate variance in the case of censoring),

$$\rho_1^{A_\epsilon} = \frac{\frac{\theta_1}{\sqrt{\epsilon_1}}\sqrt{K-1}}{\frac{\theta_1}{\sqrt{\epsilon_1}}\sqrt{K-1} + \sum_{k=2}^{K}\frac{\theta_k}{\sqrt{\epsilon_k}}} \text{ and } \rho_i^{A_\epsilon} = \frac{\frac{\theta_i}{\sqrt{\epsilon_i}}}{\frac{\theta_1}{\sqrt{\epsilon_1}}\sqrt{K-1} + \sum_{k=2}^{K}\frac{\theta_k}{\sqrt{\epsilon_k}}} \quad \text{for } i = 2, \ldots, K. \quad (2.4.2)$$

The D-optimal design ($\boldsymbol{\rho}^{D_\epsilon}$) minimizing the determinant of $\boldsymbol{\Sigma}_\epsilon$ can be obtained as a solution of a non linear system of equations and it can be found numerically.[22] In addition, based on non linear programming, Sverdlov et al[22] proposed two optimal allocations, $\boldsymbol{\rho}^{NP1}$ and $\boldsymbol{\rho}^{NP2}$, adopting the same constrained optimization framework of Tymofyeyev et al.[13] In particular $\boldsymbol{\rho}^{NP1}$ and $\boldsymbol{\rho}^{NP2}$ were derived by minimizing the total sample size and the total expected hazard, respectively, under the constraints of a minimum desired (user-selected) proportion $T \in [0, K^{-1}]$ of subjects for each treatment group, and $\phi_\epsilon \geq U$, where U is a positive constant. Only $\boldsymbol{\rho}^{NP1}$ admits a closed-form expression, while $\boldsymbol{\rho}^{NP2}$ can be found numerically (the code was kindly provided by the corresponding author of Sverdlov et al[22]). Since both targets are discontinuous functions of the unknown model parameters, the authors apply a smoothing transformation to the targets using the above-mentioned multivariate Gaussian kernel.

Remark 2.4.1. *While $\boldsymbol{\rho}^{D_\epsilon}$ and $\boldsymbol{\rho}^{C_\epsilon}$ have their components ordered according to the treatment efficacies, $\boldsymbol{\rho}^{NP1}$ and $\boldsymbol{\rho}^{NP2}$ do not always fulfil this characteristic. Indeed, for $\boldsymbol{\theta} = (34, 34, 24)^\top$ with $T = 0.2$, $R = 55$, $D = 96$ (namely under scenario IIc of Table I of Sverdlov et al[22]), the authors of the paper[22] obtained $\boldsymbol{\rho}^{NP1} = (0.3, 0.3, 0.4)^\top$ and $\boldsymbol{\rho}^{NP2} = (0.32, 0.32, 0.36)^\top$, where the higher proportion of subjects is treated with the worst treatment. For $\boldsymbol{\rho}^{A_\epsilon}$ similar considerations of Remark 2.3.2 hold: such allocation is ethical as long as the reference treatment is also the superior one; whereas if e.g., $\boldsymbol{\theta} = (25, 29, 30)^\top$ (with $R = 55$ and $D = 96$), then $\boldsymbol{\rho}^{A_\epsilon} = (0.367, 0.310, 0.323)^\top$.*

We now compare the constrained target in presence of censoring with those mentioned previously by considering $K = 3$, $R = 55$ and $D = 96$. Table 2.9 reports the theoretical values of $\boldsymbol{\rho}^{D_\epsilon}$ and $\boldsymbol{\rho}^{A_\epsilon}$, the smoothed constrained optimal target $\boldsymbol{\rho}^{C_\epsilon}$, $\boldsymbol{\rho}^{NP1}$ and $\boldsymbol{\rho}^{NP2}$ with $T = 0.2$. The efficiency criteria

described in Section 2.3.3 have been considered to measure the performance of the competing targets with respect to statistical and ethical considerations. In particular, the target whose power efficiency's values are closest to 1 is $\boldsymbol{\rho}^{C_\epsilon}$, except for $\boldsymbol{\theta} = (7, 5, 4)^\top$, in which the loss with respect to $\boldsymbol{\rho}^{NP1}$ and $\boldsymbol{\rho}^{NP2}$ is 3.1% and 1.5%, respectively. The constrained target outperforms the other allocations in terms of ethical efficiency (except in the first scenario in which $\boldsymbol{\rho}^{A_\epsilon}$ gives nearly the same value). As far as estimation efficiency is concerned, $\boldsymbol{\rho}^{C_\epsilon}$ shows very good performance in terms of A_A efficiency, always higher than 90.2%, while D_A efficiency is slightly lower. Note that, in all the experimental settings of Table 2.9 except the last one, $\boldsymbol{\rho}^{NP1}$ and $\boldsymbol{\rho}^{NP2}$ coincide by allocating the same proportion T of patients to the intermediate and the worst treatments. Finally, the proposed $\boldsymbol{\rho}^{C_\epsilon}$ is superior to the balanced design in terms of power and ethical efficiency.

Table 2.9: Comparisons of optimal targets in presence of censoring for $K = 3$ treatments in different experimental scenarios with respect to power efficiency $E_\phi(\boldsymbol{\rho})$, ethical efficiency $E_e(\boldsymbol{\rho})$, D_A efficiency $E_{D_A}(\boldsymbol{\rho})$, and A_A efficiency $E_{A_A}(\boldsymbol{\rho})$. Censoring scheme in equation (2.4.1) with $R = 55$ and $D = 96$.

$\boldsymbol{\theta}$	$\boldsymbol{\rho}$	$E_\phi(\boldsymbol{\rho})$	$E_e(\boldsymbol{\rho})$	$E_{D_A}(\boldsymbol{\rho})$	$E_{A_A}(\boldsymbol{\rho})$
$(30, 20, 8)^\top$	$\boldsymbol{\rho}^{A_\epsilon} = (0.625, 0.274, 0.101)^\top$	0.787	0.834	0.922	1
	$\boldsymbol{\rho}^{D_\epsilon} = (0.450, 0.389, 0.161)^\top$	0.798	0.752	1	0.891
	$\boldsymbol{\rho}^{C_\epsilon} = (0.684, 0.158, 0.158)^\top$	0.915	0.832	0.818	0.902
	$\boldsymbol{\rho}^{B}$	0.762	0.644	0.888	0.702
	$\boldsymbol{\rho}^{NP1} = \boldsymbol{\rho}^{NP2} = (0.6, 0.2, 0.2)^\top$	0.900	0.787	0.884	0.929
$(30, 10, 8)^\top$	$\boldsymbol{\rho}^{A_\epsilon} = (0.731, 0.150, 0.119)^\top$	0.875	0.813	0.841	1
	$\boldsymbol{\rho}^{D_\epsilon} = (0.472, 0.293, 0.235)^\top$	0.646	0.632	1	0.788
	$\boldsymbol{\rho}^{C_\epsilon} = (0.792, 0.104, 0.104)^\top$	0.904	0.854	0.744	0.973
	$\boldsymbol{\rho}^{B}$	0.485	0.533	0.950	0.584
	$\boldsymbol{\rho}^{NP1} = \boldsymbol{\rho}^{NP2} = (0.6, 0.2, 0.2)^\top$	0.788	0.720	0.958	0.931
$(12, 5, 4)^\top$	$\boldsymbol{\rho}^{A_\epsilon} = (0.663, 0.188, 0.149)^\top$	0.840	0.791	0.892	1
	$\boldsymbol{\rho}^{D_\epsilon} = (0.452, 0.302, 0.246)^\top$	0.665	0.661	1	0.849
	$\boldsymbol{\rho}^{C_\epsilon} = (0.734, 0.133, 0.133)^\top$	0.873	0.833	0.799	0.969
	$\boldsymbol{\rho}^{B}$	0.528	0.583	0.959	0.670
	$\boldsymbol{\rho}^{NP1} = \boldsymbol{\rho}^{NP2} = (0.6, 0.2, 0.2)^\top$	0.817	0.750	0.943	0.980
$(7, 5, 4)^\top$	$\boldsymbol{\rho}^{A_\epsilon} = (0.527, 0.263, 0.210)^\top$	0.727	0.834	0.953	1
	$\boldsymbol{\rho}^{D_\epsilon} = (0.397, 0.329, 0.274)^\top$	0.665	0.789	1	0.935
	$\boldsymbol{\rho}^{C_\epsilon} = (0.586, 0.207, 0.207)^\top$	0.770	0.853	0.903	0.981
	$\boldsymbol{\rho}^{B}$	0.628	0.762	0.984	0.853
	$\boldsymbol{\rho}^{NP1} = (0.519, 0.200, 0.281)^\top$	0.801	0.823	0.931	0.964
	$\boldsymbol{\rho}^{NP2} = (0.576, 0.200, 0.224)^\top$	0.785	0.847	0.907	0.976

$\boldsymbol{\rho}^{A_\epsilon}$, A_A optimal design; $\boldsymbol{\rho}^{D_\epsilon}$, D_A optimal design; $\boldsymbol{\rho}^{C_\epsilon}$, smoothed version of the optimal constrained target; $\boldsymbol{\rho}^{B}$, balanced design; $\boldsymbol{\rho}^{NP1}$ and $\boldsymbol{\rho}^{NP2}$, Sverdlov et al[22] optimal targets (with minimum allocation proportion for each arm T equal to 0.2).

2.4.2 RAR implementation under censoring

As in Section 2.3.4, we wish to implement DBCD to target the optimal constrained allocation by taking into account delayed responses and staggered entries. Indeed, in trials with time-to-event outcomes the response of patients to a given treatment is often unavailable before the randomization of the next subject. However, it has been shown[32] that DBCD is quite insensitive to delayed responses under widely satisfied conditions that hold for exponential responses.[20]

We run 10000 trials to evaluate the operating characteristics of the optimal constrained design for $K = 3$ in several illustrative examples. In this case, initial data for the RAR procedure were collected by allocating subjects with equal probabilities until two events were observed in each treatment arm. After that, at each step, subject j is randomized to treatment i with probability Ψ_{ji} (as in Section 2.3.4). We studied the convergence of the allocation proportion to the desired target and we estimated the power of the Wald and log-rank (LR) tests. The proportion of responses observed during the recruitment, i.e. the observations used in RAR ($\%obs$) and the percentage of patients allocated with RAR ($\%RAR$) will be reported in the Tables.

Table 2.10 shows the results under the experimental scenarios of Table 2.8a for different choices of the sample size. In these scenarios $\%obs$=86% and the proportion of patients allocated with RAR varies from 88% for $n = 150$ to 92-93% for $n = 300$. As n grows, the convergence to $\boldsymbol{\rho}^{C_\epsilon}$ improves and the standard deviations become smaller. For $n \geq 150$ at least 83.7% of power is guaranteed for both tests in all the scenarios, while for $n \geq 250$ the simulated average power is always higher than 0.977.

The skewness in favour of the best performing treatment is achieved by DBCD, but, as is well-known (see Hu et al[32] and reference therein), the convergence to the desired target is affected by delayed responses. Indeed, in many circumstances, patients' outcomes may not be available before the randomization of the next patient and this delay could affect the estimators of the unknown parameters and consequently influence other properties of the design.

To further investigate this aspect, we implement DBCD under experimental scenarios of Table 2.8b, for $n = 100$ and $n = 250$. We fixed $\theta_2 = 8$ and $\theta_3 = 4$ and the duration of the trial (namely $D = 96$) and we choose two different lengths of the recruitment period R; results are summarized in Table 2.11. Since ϵ_1 is decreasing in θ_1 and in R (i.e. the censoring in the data is higher for longer survival times and clearly longer recruitment period), it stands out how the effect on the convergence of an increasing sample size is a minor matter with respect to the length of the recruitment, whose effect is as important as θ_1 increases. If the recruitment is longer, it is more likely that the number of responses observed, and used to allocate the next patient, is higher, resulting in a better convergence rate. From $R = 55$ to $R = 75$, the gain in terms of $\%obs$ is between 3% and 5%. An additional issue is related to $\%RAR$, which is smaller for longer survival times. Because of the combination of these factors better convergence is achieved when $\theta_1 = 10$ whereas, as for increasing θ_1 the target is more unbalanced in favour of treatment 1, DBCD procedure is slower in approaching $\boldsymbol{\rho}^{C_\epsilon}$. On the other hand, standard deviations are smaller for higher n and shorter R. With regard to power of the tests all these scenarios present values $> 90\%$. Furthermore, note that even if the convergence to the desired target is affected by delayed responses, when the superiority of the best treatment is pronounced (e.g., $\theta_1 = 15$ or 20), the simulated allocation proportion of subjects assigned to it is always more than twice with respect to that of the other treatments.

Table 2.10: Simulation results for different sample sizes and scenarios of Table 2.8a. Estimates of model parameters $\hat{\boldsymbol{\theta}}_n$, simulated allocation proportions $\boldsymbol{\pi}_n$ (standard deviation in square brackets) and average power of Wald's test (W) and log-rank (LR) test, $R = 55$, $D = 96$ and 10000 iterations. In all these settings $\%obs = 86\%$

			$\boldsymbol{\theta} = (10, 9, 5)^\top$			
n	$\%RAR$	$\hat{\boldsymbol{\theta}}_n^\top$	$\boldsymbol{\rho}^{C_\epsilon}$	$\boldsymbol{\pi}_n^\top$	W	LR
150	88%	$(9.8, 8.9, 5.0)$		$(0.40, 0.33, 0.27)$	0.885	0.875
				$[.110, .088, .051]$		
200	90%	$(9.9, 8.9, 5.0)$		$(0.40, 0.32, 0.28)$	0.962	0.875
				$[.100, .079, .041]$		
250	91%	$(9.9, 8.9, 5.0)$	$(0.444, 0.278, 0.278)^\top$	$(0.40, 0.32, 0.28)$	0.987	0.986
				$[.092, .073, .036]$		
300	92%	$(9, 9, 8.9, 5.0)$		$(0.40, 0.32, 0.28)$	0.995	0.996
				$[.088, .069, .034]$		
			$\boldsymbol{\theta} = (10, 7, 5)^\top$			
n	$\%RAR$	$\hat{\boldsymbol{\theta}}_n^\top$	$\boldsymbol{\rho}^{C_\epsilon}$	$\boldsymbol{\pi}_n^\top$	W	LR
150	88%	$(9.9, 6.9, 5.0)$		$(0.48, 0.27, 0.25)$	0.837	0.842
				$[.107, .069, .052]$		
200	90%	$(9.9, 6.9, 5.0)$		$(0.48, 0.27, 0.25)$	0.938	0.931
				$[.095, .0.6, .044]$		
250	91%	$(9.9, 6.9, 5.0)$	$(0.594, 0.203, 0.203)^\top$	$(0.49, 0.26, 0.25)$	0.977	0.975
				$[.088, .054, .041]$		
300	92%	$(9.9, 6.9, 5.0)$		$(0.49, 0.26, 0.25)$	0.992	0.991
				$[.084, .050, .039]$		
			$\boldsymbol{\theta} = (10, 5, 5)^\top$			
n	$\%RAR$	$\hat{\boldsymbol{\theta}}_n^\top$	$\boldsymbol{\rho}^{C_\epsilon}$	$\boldsymbol{\pi}_n^\top$	W	LR
150	89%	$(9.9, 5.0, 5.0)$		$(0.55, 0.23, 0.22)$	0.944	0.947
				$[.089, .054, .046]$		
200	90%	$(10.0, 5.0, 5.0)$		$(0.56, 0.22, 0.22)$	0.985	0.985
				$[.075, .044, .037]$		
250	92%	$(10.0, 5.0, 5.0)$	$(0.672, 0.164, 0.164)^\top$	$(0.57, 0.22, 0.21)$	0.996	0.996
				$[.065, .038, .032]$		
300	93%	$(10.0, 5.0, 5.0)$		$(0.58, 0.21, 0.21)$	0.999	0.999
				$[.058, .033, .029]$		

Table 2.11: Simulation results for $n = 100, 250$ and 10000 iterations. Estimates of model parameters $\hat{\boldsymbol{\theta}}_n$, simulated allocation proportions $\boldsymbol{\pi}_n$ (standard deviation in square brackets) and average power of Wald's (W) and log-rank (LR) tests, different R and $D = 96$ for $\theta_2 = 8$ and $\theta_3 = 4$.

				$\boldsymbol{\theta} = (10, 8, 4)^\top$				
R	n	$\%obs$	$\%RAR$	$\hat{\boldsymbol{\theta}}_n^\top$	$\boldsymbol{\rho}^{C_\epsilon}$	$\boldsymbol{\pi}_n^\top$	W	LR
55	100	86%	85%	$(9.8, 7.8, 3.9)$	$(0.550, 0.225, 0.225)^\top$	$(0.45, 0.30, 0.25)$ $[.131, .102, .065]$	0.910	0.904
	250	86%	91%	$(9.9, 7.9, 4.0)$		$(0.45, 0.29, 0.26)$ $[.096, .070, .038]$	1.000	1.000
75	100	89%	87%	$(9.8, 7.7, 3.9)$	$(0.552, 0.224, 0.224)^\top$	$(0.47, 0.30, 0.24)$ $[.142, .111, .068]$	0.907	0.896
	250	90%	92%	$(9.9, 7.9, 4.0)$		$(0.47, 0.28, 0.25)$ $[.102, .074, .039]$	0.999	1.000
				$\boldsymbol{\theta} = (15, 8, 4)^\top$				
R	n	$\%obs$	$\%RAR$	$\hat{\boldsymbol{\theta}}_n^\top$	$\boldsymbol{\rho}^{C_\epsilon}$	$\boldsymbol{\pi}_n^\top$	W	LR
55	100	81%	84%	$(14.9, 7.8, 3.9)$	$(0.714, 0.143, 0.143)^\top$	$(0.55, 0.24, 0.21)$ $[.117, .081, .059]$	0.996	0.995
	250	81%	90%	$(15.0, 7.9, 4.0)$		$(0.57, 0.22, 0.21)$ $[.076, .046, .036]$	1.000	1.000
75	100	85%	85%	$(14.9, 7.7, 3.9)$	$(0.716, 0.142, 0.142)^\top$	$(0.59, 0.21, 0.20)$ $[.119, .081, .059]$	0.996	0.994
	250	86%	92%	$(15.0, 7.9, 4.0)$		$(0.61, 0.20, 0.19)$ $[.073, .044, .034]$	1.000	1.000
				$\boldsymbol{\theta} = (20, 8, 4)^\top$				
R	n	$\%obs$	$\%RAR$	$\hat{\boldsymbol{\theta}}_n^\top$	$\boldsymbol{\rho}^{C_\epsilon}$	$\boldsymbol{\pi}_n^\top$	W	LR
55	100	76%	82%	$(20.1, 7.8, 3.9)$	$(0.782, 0.109, 0.109)^\top$	$(0.60, 0.20, 0.20)$ $[.106, .068, .060]$	0.999	0.999
	250	77%	90%	$(20.0, 7.9, 4.0)$		$(0.62, 0.19, 0.19)$ $[.065, .038, .034]$	1.000	0.999
75	100	81%	84%	$(20.1, 7.7, 3.9)$	$(0.786, 0.107, 0.107)^\top$	$(0.64, 0.19, 0.17)$ $[.100, .067, .055]$	1.000	0.999
	250	82%	91%	$(20.0, 7.9, 4.0)$		$(0.66, 0.17, 0.17)$ $[.060, .035, .031]$	1.000	0.999

The impact of the magnitude of the treatment effects on the convergence is additionally supported by the results in Table 2.12 which correspond to a set-up of a very rapid fatal disease. In this case the estimated allocation proportions are really close to the target, ensuring that 60% of patients or more receive the best treatment. These results show that, for small values of $\boldsymbol{\theta}$, our proposal has better performance even for small-moderate sample sizes, particularly common in clinical trials for rare diseases.

Finally Table 2.13 shows the simulate average type I errors for the Wald and log-rank (LR_α) tests. Both of them are very close to the significance level, that was set to 0.05. Better convergence rate is achieved by the second scenario $\boldsymbol{\theta} = (4, 4, 4)^\top$ as in this case the proportion of data used in the RAR procedure is greater than 90%, while in the first scenario it is not higher than 81%.

Table 2.12: Simulation results for different sample sizes, 10000 iterations and $\boldsymbol{\rho}^{C_\epsilon} = (0.634, 0.183, 0.183)^\top$. Estimates of model parameters $\hat{\boldsymbol{\theta}}_n$, simulated allocation proportions $\boldsymbol{\pi}_n$ (standard deviation in square brackets) and average power of the Wald (W) and log-rank (LR) tests, $R = 55$, $D = 96$. In all these settings $\%obs = 96\%$.

		$\boldsymbol{\theta} = (3, 2, 1)^\top$				
n	$\%RAR$	$\hat{\boldsymbol{\theta}}_n^\top$	$\boldsymbol{\pi}_n^\top$		W	LR
150	93%	(3.0, 1.9, 1.0)	(0.61, 0.20, 0.19)	[.114, .074, .053]	0.998	0.999
200	94%	(3.0, 2.0, 1.0)	(0.61, 0.20, 0.19)	[.093, .058, .042]	1.000	1.000
250	95%	(3.0, 2.0, 1.0)	(0.62, 0.19, 0.19)	[.080, .048, .037]	1.000	1.000
300	95%	(3.0, 2.0, 1.0)	(0.62, 0.19, 0.19)	[.072, .042, .034]	1.000	1.000

Table 2.13: Simulation results of type I error for the Wald (W_α) and log-rank (LR_α) tests ($\boldsymbol{\rho}^{C_\epsilon} = \boldsymbol{\rho}^B$). Estimates of model parameters $\hat{\boldsymbol{\theta}}_n$ and simulated allocation proportions $\boldsymbol{\pi}_n$ (standard deviation in square brackets), with $R = 55$, $D = 96$ and 10000 iterations.

			$\boldsymbol{\theta} = (12, 12, 12)^\top$				
n	$\%obs$	$\%RAR$	$\hat{\boldsymbol{\theta}}_n^\top$	$\boldsymbol{\pi}_n^\top$		W_α	LR_α
150	80%	86%	(11.8, 12.0, 12.0)	(0.32, 0.35, 0.33)	[.085, .075, .079]	0.048	0.061
200	81%	88%	(11.9, 12.0, 12.0)	(0.32, 0.35, 0.33)	[.080, .070, .071]	0.051	0.057
250	81%	89%	(11.9, 12.0, 12.0)	(0.32, 0.35, 0.33)	[.075, .066, .066]	0.047	0.054
300	81%	90%	(11.9, 12.0, 12.0)	(0.32, 0.35, 0.33)	[.072, .064, .064]	0.047	0.053
			$\boldsymbol{\theta} = (4, 4, 4)^\top$				
n	$\%obs$	$\%RAR$	$\hat{\boldsymbol{\theta}}_n^\top$	$\boldsymbol{\pi}_n^\top$		W_α	LR_α
150	92%	91%	(3.9, 4.0, 4.0)	(0.33, 0.34, 0.33)	[.095, .077, .083]	0.048	0.056
200	92%	92%	(3.9, 4.0, 4.0)	(0.33, 0.34, 0.33)	[.089, .071, .072]	0.048	0.055
250	93%	93%	(3.9, 4.0, 4.0)	(0.33, 0.34, 0.33)	[.084, .068, .068]	0.047	0.055
300	93%	94%	(3.9, 4.0, 4.0)	(0.33, 0.33, 0.34)	[.081, .066, .066]	0.046	0.054

2.4.3 Robustness of our methodology to model misspecifications and sensitivity analysis

In this section we investigate the performance of the optimal constrained target when the distribution of the survival times is non exponential. To take into account different shapes of the hazard function, assuming $\log Y_{ij} = \log \theta_j + s \cdot e_i$ and standard extreme value, standard logistic or standard normal distribution for e_i, then Y_{ij} follows Weibull (Weib), log-logistic (LL) or log-normal (LN) model. In particular we consider Weib with parameters $s = 0.8$ and 1.25 (monotone increasing and decreasing hazard, respectively), LL distribution with $s = 0.5$ and the LN distribution with $s = 0.8$ (non-monotone hazards).

As it is well-known, in both conventional and adaptive designs, wrong parametric assumptions often lead to power loss and type I error inflation of parametric tests.[33] In our robustness studies, besides the log-rank test, we consider the performance of the Wald test based on the correctly specified distribution of the survival times (as also suggested by Sverdlov et al[33]). More specifically, we compute the test statistic by analysing the final dataset with the distribution according to which the survival times were generated and we report the average power/type I error of Wald's test ($\tilde{W}$ and $\tilde{W}_\alpha$). We summarize in

Table 2.14 the simulation results under different models for the survival times for the DBCD targeting $\boldsymbol{\rho}^{C_\epsilon}$ and the CRD intending to target $\boldsymbol{\rho}^{B}$, taking into account $n = 250$ patients. These results show that our proposal is - in general - slightly more powerful and it results in a fewer average number of deaths than the balanced design even when the event time distribution is non exponential. The maximum loss of power with respect to the one under the exponential model for $\boldsymbol{\rho}^{C_\epsilon}$ occurs under a monotone increasing hazard function (-12% and -10% for $\tilde{\text{W}}$ and LR respectively) which is, at the same time, associated to a smaller average number of events in the study (-3 deaths). Conversely, when survival times follow a Weib with $s = 1.25$, higher values of power are present. In the case of non-monotone hazard functions an improvement in terms of power is combined with a substantial reduction in the number of deaths in the trial. The type I error is around the nominal level in most of the scenarios considered; a slight inflation is observed in the case of non-monotone hazards. We conclude that our procedure is quite robust to model misspecification in all the considered experimental set-ups for both the log-rank test and the parametric test $\tilde{\text{W}}$.

Table 2.14: Simulated average power/type I error of the Wald test ($\tilde{\text{W}}$ and $\tilde{\text{W}}_\alpha$) based on the correctly specified model and the log-rank test (LR and LR_α), average number of deaths (Deaths; with their standard deviations in brackets) for the DBCD targeting $\boldsymbol{\rho}^{C_\epsilon}$ and the CRD (intended to target $\boldsymbol{\rho}^{B}$) for $n = 250$ and 10000 iterations.

	$\boldsymbol{\theta} = (10, 7, 5)^\top$					
Model	$\boldsymbol{\rho}^{C_\epsilon}$			$\boldsymbol{\rho}^{B}$		
	$\tilde{\text{W}}$	LR	Deaths	$\tilde{\text{W}}$	LR	Deaths
Exp	0.977	0.975	229 (4)	0.970	0.972	231 (4)
Weib, $s = 0.8$	0.859	0.870	226 (5)	0.842	0.870	228 (4)
Weib, $s = 1.25$	0.999	0.998	230 (4)	0.999	0.998	232 (4)
LL, $s = 0.5$	0.996	0.985	218 (5)	0.996	0.983	221 (5)
LN, $s = 0.8$	0.999	0.996	221 (5)	0.999	0.995	223 (5)
	$\boldsymbol{\theta} = (12, 12, 12)^\top$					
Model	$\boldsymbol{\rho}^{C_\epsilon}$			$\boldsymbol{\rho}^{B}$		
	$\tilde{\text{W}}_\alpha$	LR_α	Deaths	$\tilde{\text{W}}_\alpha$	LR_α	Deaths
Exp	0.047	0.054	218 (5)	0.044	0.051	218 (5)
Weib, $s = 0.8$	0.050	0.052	213 (5)	0.044	0.050	213 (5)
Weib, $s = 1.25$	0.049	0.054	221 (5)	0.048	0.053	221 (4)
LL, $s = 0.5$	0.061	0.055	203 (6)	0.060	0.051	203 (6)
LN, $s = 0.8$	0.062	0.057	206 (6)	0.067	0.055	206 (6)

Another form of model misspecification may occur when the patients' accrual rate is non-uniform; to take into account several recruitment patterns we consider the Beta distribution.[33] The cases Beta $(1, 5)$ and Beta $(5, 1)$ (right/left skewed) encompass situations in which the accrual rate decreases/increases over time respectively. On the other hand, Beta $(1/5, 1/5)$ represents scenarios with accelerated recruitment at the beginning and at the end of the trial and Beta $(5, 5)$ refers to studies in which the highest recruitment rate occurs in the middle of the trial. Table 2.15 shows the simulating operating characteristics of the DBCD targeting $\boldsymbol{\rho}^{C_\epsilon}$ and the CRD targeting $\boldsymbol{\rho}^{B}$. We report the average proportion of patients whose outcomes are observed during the recruitment, the simulated average power of the Wald test/type I error and the total average number of events for $n = 250$ patients.

Table 2.15: Simulated average power/type I error of the Wald test (W and W_α), average percentage of observation during the recruitment period ($\%obs$; only for the DBCD procedure), average number of deaths (Deaths; with standard deviations in brackets) for the DBCD targeting $\boldsymbol{\rho}^{C_\epsilon}$ and the CRD (intended to target $\boldsymbol{\rho}^B$) under different recruitment patterns, $n = 250$ patients and 10000 iterations.

	$\boldsymbol{\theta} = (10, 7, 5)^\top$					$\boldsymbol{\theta} = (12, 12, 12)^\top$				
Recruitment	$\boldsymbol{\rho}^{C_\epsilon}$			$\boldsymbol{\rho}^B$		$\boldsymbol{\rho}^{C_\epsilon}$			$\boldsymbol{\rho}^B$	
	$\%obs$	W	Deaths	W	Deaths	$\%obs$	W_α	Deaths	W_α	Deaths
Uniform	86%	0.977	229 (5)	0.970	231 (4)	81%	0.047	218 (5)	0.045	218 (5)
Beta $(1, 5)$	95%	0.974	230 (4)	0.974	231 (4)	91%	0.048	219 (5)	0.045	219 (5)
Beta $(5, 1)$	59%	0.965	230 (5)	0.972	231 (4)	48%	0.052	217 (5)	0.045	217 (5)
Beta $(1/5, 1/5)$	67%	0.980	229 (5)	0.975	231 (4)	64%	0.054	218 (5)	0.048	218 (5)
Beta $(5, 5)$	89%	0.972	230 (4)	0.972	231 (4)	81%	0.052	219 (5)	0.045	219 (5)

The power of the Wald test is similar across the five recruitment patterns with a maximum loss of 1.2% corresponding to the left-skewed Beta distribution. On the other hand, under Beta$(1/5, 1/5)$ a slight increase in power is observed. The type I error rate is close to 0.05 showing that our proposal is also robust to misspecification of the recruitment pattern.

2.4.4 Redesign of KEYNOTE-010 clinical trial

In this section we illustrate the application of the constrained optimal target by redesigning the KEYNOTE-010 clinical trial[8] (registered at ClinicalTrials.gov, number NCT01905657). The aim of this Phase II/III study was to compare two doses of pembrolizumab (MK-3475) versus docetaxel in patients with non-small cell lung cancer and whose tumors were assessed as being PD-L1 positive. Between Aug 2013 and Feb 2015, 1034 participants were enrolled and were randomly allocate with a 1:1:1 ratio, with a central interactive voice-response system, to receive treatment A (pembrolizumab 2 mg/kg), treatment B (pembrolizumab 10 mg/kg) or treatment C (docetaxel). Among primary endpoints (overall survival and progression-free survival) we were interested in the overall survival in the intent-to-treat population. At the cut-off date, after 23 months, median overall survival was 10.4 months in group A, 12.7 months in group B, and 8.5 months in group C. In designing this study the authors assumed exponential distribution for overall survival, so that we run 10000 trials with $n = 1034$ patients with $\boldsymbol{\theta} = (\theta_A, \theta_B, \theta_C)^\top = (15, 18, 12)^\top$, adopting the above-mentioned censoring scheme with $R = 18$ and $D = 23$. The DBCD procedure has been implemented to target the optimal allocations $\boldsymbol{\rho}^{C_\epsilon}$, ρ^{A_ϵ}, ρ^{D_ϵ} and ρ^{NP1} (with $T = 0.2$) and we included the CRD to target $\boldsymbol{\rho}^B$, as sequential analogue to the equal allocation adopted in the original trial. Results are presented in Table 2.16 in which, for sake of comparison, we have also reported the simulated average efficiencies (see Section 2.3.3), number of deaths, total observed survival time, number of patients assigned to the inferior and to the superior treatment and the total expected hazard in the study (H=$\sum_{i=1}^K N_{in}/\theta_i$). In all the procedures $\%obs$ and $\%RAR$ were 60% and 90-91% respectively. The optimal constrained design presents the highest degree of skewness in favour of the best performing treatment arm (B in this case) with respect to to the other designs and the ethical property of Remark 2.3.1 does not hold for $NP1$ and the A_A allocations (see also Remark 3). All the procedures present similar statistical properties of power and estimation efficiency. Nevertheless, our proposal is superior in terms of ethical characteristics, resulting in 4-8 fewer deaths, longer total survival time, lower average hazard and higher ethical efficiency compared to the other designs. In addition, nInf and nSup highlight the potential advantages in adopting the optimal constrained design in which from 20

up to 74 fewer patients are assigned to the inferior treatment, whereas 428 patients receive the superior treatment, so that 24-83 more patients are injected with treatment B with the maximum gain achieved with respect to the balanced design.

Note that this result provide further significance to the ethical definition of Remark 2.3.1. Given the gravity of the outcome, a response-adaptive procedure targeting $\boldsymbol{\rho}^{C_\epsilon}$ would provide a better trade-off between statistical power and patients' benefits.

Table 2.16: Redesign of KEYNOTE-010[8] clinical trial.

	Statistical properties					*Ethical characteristics*					
	W	LR	E_ϕ	E_{D_A}	E_{A_A}	Deaths	TS	nInf	nSup	H	E_e
$\boldsymbol{\pi}_n^{A_\epsilon} = (0.30, 0.34, 0.36)^\top$ $[.028, .031, .029]$	0.89	0.90	0.74	0.97	0.98	445	6533	365	354	70	0.83
$\boldsymbol{\pi}_n^{D_\epsilon} = (0.34, 0.36, 0.30)^\top$ $[.020, .019, .019]$	0.88	0.89	0.73	0.99	0.91	441	6569	311	376	71	0.84
$\boldsymbol{\pi}_n^{C_\epsilon} = (0.31, 0.41, 0.28)^\top$ $[.053, .079, .035]$	0.88	0.88	0.76	0.98	0.94	437	6595	291	428	69	0.86
$\boldsymbol{\pi}_n^{B} = (0.33, 0.34, 0.33)^\top$ $[.007, .007, .007]$	0.88	0.89	0.72	0.98	0.87	444	6546	345	345	71	0.83
$\boldsymbol{\pi}_n^{NP1} = (0.29, 0.39, 0.32)^\top$ $[.057, .065, .047]$	0.89	0.89	0.77	0.96	0.91	441	6570	333	404	70	0.85

2.5 Conclusions and future research

The design of multi-arm clinical experiments is complex, especially when different objectives are involved. The proposed optimal constrained target guarantees a valid trade-off between the inferential goal of maximizing the power of Wald's test of homogeneity and the ethical demand of preserving subjects' care, whereas some targets for time-to-event outcomes could lead to undesirable allocation proportions, as we pointed out in this paper.

In addition, we implement our proposal via DBCD methodology. To assess its simulated operating characteristics we considered several experimental settings. In the uncensored model, we assumed that responses are available immediately and the convergence rate to the target is excellent. In the presence of censored observations we have also included delayed responses and staggered entries, which could seriously slowed down the convergence of the design. To assess such an impact, we ran simulation studies directed to illustrate how the convergence is related to the complex interplay between sample size, survival times, length of recruitment and duration of the trial. Overall, the optimal constrained design reaches a good rate of convergence, provided that a sufficient amount of responses are observed throughout the recruitment phase to let the adaptive procedure work. Moreover, even in experimental scenarios with slower convergence, a fair degree of skewness in favour of the most promising treatment is achieved. The practical applicability of our proposal has been also highlighted by performing robustness studies to model misspecification and by redesigning a real lung cancer trial. Authors are working on developing a user-friendly interface (R-based Shiny Web application) to implement the procedure, which will be available in the future.

The problem of testing the null-hypothesis of equality among treatment effects considered in this paper is useful in many applications. In multi-arm trials a global test comparing all treatments can be carried out prior to making individual pairwise comparisons.[13,21,22] Indeed, the overall null-hypothesis

is the first stage of multiple comparison methodologies for several stepwise procedures.[24,25] One of the most powerful is the Fisher's least significant difference method which is a two-step test for pairwise comparisons; in the first stage the overall null-hypothesis of homogeneity is tested at level α and then, in case of rejection, the pairs of interest are tested for equality at the same level of significance. As it is well-known, the design should be tailored on trial objectives: a single global test may be of interest, for instance, in Phase II trials.[34] Especially in anticancer research, due to the dramatic increase of new potential drugs under development, one of the primary Phase II objectives is to evaluate the effect of new treatments and to identify the one(s) that most warrants additional evaluation in a larger Phase III study.[35,36] Others interesting set-ups concern trials in which the choice of the control treatment is not unique: for instance, studies in which the new drug has to be compared to placebo, current commercial products, competitors products (see the comments by Owen and by Bechhofer and Tamhane in the discussion of Hedayat et al[37]). Finally there are situations such that no treatments with demonstrated efficacy exist and so no standard of care is available to be set as an appropriate/fixed control. This is particularly related to rapidly emerging novel infectious diseases such as Ebola[38] or COVID-19. In these cases clinical studies must start quickly and the overall null-hypothesis, as a first step, allows to evaluate several candidate treatments at once.

A further significant aspect of this work is to provide support to the arguments in favour of unbalanced allocation designs.[2,3] Unequal randomization is often more appropriate for patients' health and more powerful than the balanced design, especially for heteroscedastic treatment groups, and we showed that the proposed constrained target shares this property. Hence, the indiscriminate use of the popular equal allocation design in clinical trials should be reconsidered.

The promising performance of this optimization approach leads to further methodological developments to extend its applicability. One of the main directions of future research is to adopt this framework to derive optimal constrained targets for widely used heteroscedastic models like, e.g., for binary trials. Another interesting issue consists in including covariates/prognostic factors, to take also into account patients' heterogeneity.

2.6 Proofs

2.6.1 Proof of Theorem 1

The following Lemma is preliminary to the proof.

Lemma 4. *$\phi(\cdot)$ is a concave function of the vector of allocation proportions.*

Proof of Lemma 4. Let $\omega_i = (\rho_i\theta_i^{-2})/(\sum_{k=1}^K \rho_k\theta_k^{-2})$, then $\omega_i \geq 0$ for $i = 1, \ldots, K$ and $\sum_{i=1}^K \omega_i = 1$, namely $\omega = (\omega_1, \ldots, \omega_K)^\top$ could be regarded as a pdf of a (non-negative) discrete r.v. θ with K support points $\theta_1 \geq \ldots \geq \theta_K > 0$. So, letting

$$\bar{\theta}_\omega = \sum_{i=1}^K \theta_i\, \omega_i = \frac{\sum_{i=1}^K \rho_i\theta_i^{-1}}{\sum_{i=1}^K \rho_i\theta_i^{-2}} \tag{2.6.1}$$

and $M_\omega(\theta^2) = \sum_{i=1}^K \theta_i^2\omega_i = \left(\sum_{i=1}^K \rho_i\theta_i^{-2}\right)^{-1}$, then by (2.3.1), it is possible to show that

$$\phi(\boldsymbol{\rho}) = \frac{M_\omega(\theta^2) - \bar{\theta}_\omega^2}{M_\omega(\theta^2)} = \frac{V_\omega(\theta)}{M_\omega(\theta^2)}, \tag{2.6.2}$$

where $V_\omega(\theta)$ is the variance of θ evaluated w.r.t. ω. Moreover,

$$\phi(\boldsymbol{\rho}) = \frac{1}{M_\omega(\theta^2)} \sum_{i=1}^{K} (\theta_i - \bar{\theta}_\omega)^2 \omega_i = \sum_{i=1}^{K} a_i \rho_i, \tag{2.6.3}$$

where $a_i = \left(1 - \bar{\theta}_\omega/\theta_i\right)^2$, for $i = 1, \ldots, K$. Note that, for $i = 1, \ldots, K$, $\dfrac{\partial \phi(\boldsymbol{\rho})}{\partial \rho_i} = \sum_{j=1}^{K} \dfrac{\partial a_j}{\partial \rho_i} \rho_j + a_i$ where

$$\frac{\partial a_j}{\partial \rho_i} = 2\left(1 - \frac{\bar{\theta}_\omega}{\theta_j}\right)\left(-\frac{1}{\theta_j}\right)\frac{\partial \bar{\theta}_\omega}{\partial \rho_i} \quad \text{and} \quad \frac{\partial \bar{\theta}_\omega}{\partial \rho_i} = \frac{\theta_i^{-1}}{\sum_{i=1}^{K} \rho_i \theta_i^{-2}}\left(1 - \frac{\bar{\theta}_\omega}{\theta_i}\right).$$

Thus,

$$\frac{\partial \phi(\boldsymbol{\rho})}{\partial \rho_i} = a_i - \frac{2}{\theta_i}\left(1 - \frac{\bar{\theta}_\omega}{\theta_i}\right) \sum_{j=1}^{K} \left\{ \frac{\rho_j \theta_j^{-1}}{\sum_{i=1}^{K} \rho_i \theta_i^{-2}} \left(1 - \frac{\bar{\theta}_\omega}{\theta_j}\right) \right\} = a_i,$$

and

$$\frac{\partial^2 \phi(\boldsymbol{\rho})}{\partial \rho_i^2} = -\frac{2M_\omega(\theta^2)}{\theta_i^2}\left(1 - \frac{\bar{\theta}_\omega}{\theta_i}\right)^2, \quad \frac{\partial^2 \phi(\boldsymbol{\rho})}{\partial \rho_i \partial \rho_j} = -\frac{2M_\omega(\theta^2)}{\theta_i \theta_j}\left(1 - \frac{\bar{\theta}_\omega}{\theta_i}\right)\left(1 - \frac{\bar{\theta}_\omega}{\theta_j}\right).$$

Thus, the Hessian matrix is given by

$$H_\phi(\boldsymbol{\rho}) = -2M_\omega(\theta^2) \left[\left(1 - \frac{\bar{\theta}_\omega}{\theta_1}\right), \ldots, \left(1 - \frac{\bar{\theta}_\omega}{\theta_K}\right)\right]^\top \left[\left(1 - \frac{\bar{\theta}_\omega}{\theta_1}\right), \ldots, \left(1 - \frac{\bar{\theta}_\omega}{\theta_K}\right)\right], \tag{2.6.4}$$

having one eigenvalue equal to 0 (with multiplicity $K - 1$) and a non-null eigenvalue given by its trace, i.e. $tr\{H_\phi(\boldsymbol{\rho})\} = -2M_\omega(\theta^2) \sum_{i=1}^{K} \left(1 - \frac{\bar{\theta}_\omega}{\theta_i}\right)^2 < 0$, which implies the concavity of NCP.

Proof of Theorem 1

Let $CV_\omega(\theta)$ be the coefficient of variation of θ evaluated with respect to ω, by (2.6.2) the NCP can be rewritten as $\phi(\boldsymbol{\rho}) = V_\omega(\theta)\left(V_\omega(\theta) + \bar{\theta}_\omega^2\right)^{-1} = \left(1 + \frac{1}{CV_\omega^2(\theta)}\right)^{-1}$.

Within the class of pdfs with a given mean $\bar{\theta}_\omega$, the variance $V_\omega(\theta)$ is maximized by the one that assigns all the mass of probability to the extremes, in this case θ_1 and θ_K. So let $P(\theta = \theta_K) = \omega^* = 1 - P(\theta = \theta_1)$, then $\bar{\theta}_\omega = \theta_1(1 - \omega^*) + \theta_K \omega^*$ and $V_\omega(\theta) = (\theta_1 - \theta_K)^2 \omega^*(1 - \omega^*)$. Thus, $CV_\omega^2(\theta) = [\theta_1(1 - \omega^*) + \theta_K \omega^*]^{-2}(\theta_1 - \theta_K)^2 \omega^*(1 - \omega^*)$, where $\frac{\partial CV_\omega}{\partial \omega^*} = 0 \Leftrightarrow \omega^* = \theta_1/(\theta_1 + \theta_K)$. In such a case, $\omega^* = \frac{\rho_K}{\theta_K^2}\left(\frac{\rho_1}{\theta_1^2} + \frac{\rho_K}{\theta_K^2}\right)^{-1}$, thus $\rho_K = \theta_K/(\theta_1 + \theta_K) = 1 - \rho_1$ and, from (2.6.2), $\phi(\tilde{\boldsymbol{\rho}}) = \left(\frac{\theta_1 - \theta_K}{\theta_1 + \theta_K}\right)^2$.

As regards statement i), every allocation $\tilde{\boldsymbol{\rho}}$ such that $\sum_{i=1}^{j} \tilde{\rho}_i = \frac{\theta_1}{\theta_1 + \theta_K} = 1 - \sum_{i=h}^{K} \tilde{\rho}_i$ is optimal. Indeed, in this case $\bar{\theta}_\omega = \left(\sum_{i=1}^{K} \frac{\tilde{\rho}_i}{\theta_i^2}\right)^{-1} \sum_{i=1}^{K} \frac{\tilde{\rho}_i}{\theta_i} = \frac{2\theta_1\theta_K}{\theta_1 + \theta_K}$ and thus, from (2.6.3), $\phi(\tilde{\boldsymbol{\rho}}) = \left(1 - \frac{\bar{\theta}_\omega}{\theta_1}\right)^2 \frac{\theta_1}{\theta_1 + \theta_K} + \left(1 - \frac{\bar{\theta}_\omega}{\theta_K}\right)^2 \frac{\theta_K}{\theta_1 + \theta_K} = \left(\frac{\theta_1 - \theta_K}{\theta_1 + \theta_K}\right)^2$.

The proof of (ii) is straightforward. Moreover, since the Hessian matrix in (2.6.4) is negative semi-definite we have to check the stationary points. By setting the partial derivatives of the Lagrangian

$L(\boldsymbol{\rho}, \lambda) = \phi(\boldsymbol{\rho}) - \lambda\left(\sum_{i=1}^{K} \rho_i - 1\right)$ equal to zero, we obtain a system of K equations $a_i = \lambda$ for $i = 1, \ldots, K$, which admits solutions if and only if $a_1 = \cdots = a_K$. Notice that, for every $i = 1, \ldots, K-1$,

$$a_i \geq a_{i+1} \Leftrightarrow \left(\frac{\bar{\theta}_\omega}{\theta_i} + \frac{\bar{\theta}_\omega}{\theta_{i+1}}\right) \leq 2 \quad \text{or} \quad \theta_i = \theta_{i+1} \tag{2.6.5}$$

and, since $\theta_K < \bar{\theta}_\omega < \theta_1$, from $a_1 = a_K$ we obtain

$$\left(1 - \frac{\bar{\theta}_\omega}{\theta_1}\right)^2 = \left(1 - \frac{\bar{\theta}_\omega}{\theta_K}\right)^2 \Leftrightarrow \bar{\theta}_\omega = 2\left(\frac{1}{\theta_K} + \frac{1}{\theta_1}\right)^{-1} = \frac{2\theta_1\theta_K}{\theta_1 + \theta_K}. \tag{2.6.6}$$

Clearly, from (2.6.5), if $\theta_i = \theta_{i+1}$ then $a_i = a_{i+1}$. Furthermore, if $\exists\ j \in \{1, \ldots, K-1\}$ such that $\theta_1 = \ldots = \theta_j > \theta_{j+1} \geq \ldots \geq \theta_K$ then $a_1 = a_2 = \ldots = a_j$; since $\theta_i \neq \theta_1$ for $i = j+1, \ldots, K$, from (2.6.5) follows that $a_i = a_1 \ \Leftrightarrow \bar{\theta}_\omega\left(\frac{1}{\theta_1} + \frac{1}{\theta_i}\right) = 2 \ \ \forall\, i = j+1, \ldots, K$. By substituting $\bar{\theta}_\omega$ in (2.6.6), then $\theta_i = \theta_K$, for every $i = j+1, \ldots, K$.
On the other hand, if $\exists\ h \in \{2, \ldots, K\}$ such that $\theta_1 \geq \ldots \geq \theta_{h-1} > \theta_h = \ldots = \theta_K$, then $a_h = \ldots = a_K$. Clearly, $\theta_i \neq \theta_K$ for $i = 1, \ldots, h-1$ and therefore, from (2.6.5), $a_i = a_K \ \Leftrightarrow \bar{\theta}_\omega\left(\frac{1}{\theta_i} + \frac{1}{\theta_K}\right) = 2$. Thus, by substituting $\bar{\theta}_\omega$, it follows that $\theta_1 = \theta_i$ for every $i = 1, \ldots, h-1$.

2.6.2 Proof of Theorem 2

The following Lemma is preliminary to the proof.

Lemma 5. *Given a non-degenerate target $\boldsymbol{\rho}$ (i.e. such that $\rho_j > 0$ for every j), we have*

(i) If $\bar{\theta}_\omega \in \left[\frac{2\theta_K\theta_{K-1}}{\theta_K+\theta_{K-1}}, \frac{2\theta_1\theta_2}{\theta_1+\theta_2}\right]$, then $\exists\ \tilde{i} \in \{2, \ldots, K-1\}$ such that

$$a_1 \geq a_2 \geq \cdots \geq a_{\tilde{i}} \leq a_{\tilde{i}+1} \leq \cdots \leq a_K \tag{2.6.7}$$

In particular, if $\bar{\theta}_\omega \in \left[\theta_2, \frac{2\theta_1\theta_2}{\theta_1+\theta_2}\right]$, then $\tilde{i} = 2$; whereas, if $\bar{\theta}_\omega \in \left[\frac{2\theta_K\theta_{K-1}}{\theta_K+\theta_{K-1}}, \theta_{K-1}\right]$, then $\tilde{i} = K-1$.

(ii) If $\bar{\theta}_\omega \in \left[\frac{2\theta_1\theta_2}{\theta_1+\theta_2}, \theta_1\right)$ then $\tilde{i} = 1$, i.e.

$$a_1 \leq a_2 \leq \cdots \leq a_K. \tag{2.6.8}$$

(iii) If $\bar{\theta}_\omega \in \left(\theta_K, \frac{2\theta_K\theta_{K-1}}{\theta_K+\theta_{K-1}}\right]$, then $\tilde{i} = K$, i.e

$$a_1 \geq a_2 \geq \cdots \geq a_K. \tag{2.6.9}$$

Notice that $\frac{2\theta_K\theta_{K-1}}{\theta_K+\theta_{K-1}} \leq \theta_{K-1}$ and $\frac{2\theta_1\theta_2}{\theta_1+\theta_2} \geq \theta_2$.
Moreover,

(iv) if $\bar{a} = a_1$, then (2.6.8) and (2.6.9) are impossible and, therefore, (2.6.7) holds with $a_K > a_1$;

(v) if $\bar{a} > a_1$ then (2.6.9) is impossible;

(vi) if $\bar{a} > a_1 \geq a_2$, then (2.6.8) and (2.6.9) are impossible and, therefore, (2.6.7) holds with $a_K > a_1$.

Proof of Lemma 5. Clearly from (2.6.5), $a_i \geq a_{i+1} \iff \left(\frac{\bar{\theta}_\omega}{\theta_i} + \frac{\bar{\theta}_\omega}{\theta_{i+1}}\right) \leq 2$. Thus, if $\bar{\theta}_\omega \in [\theta_{K-1}, \theta_2]$, then $\exists\, \tilde{i} \in \{2, \ldots, K-1\}$ such that $a_1 \geq a_2 \geq \cdots \geq a_{\tilde{i}} \leq a_{\tilde{i}+1} \leq \cdots \leq a_K$.
Moreover, if $\bar{\theta}_\omega \in (\theta_2, \theta_1)$ and $\left(\frac{\bar{\theta}_\omega}{\theta_1} + \frac{\bar{\theta}_\omega}{\theta_2}\right) \leq 2$ (i.e. $\theta_2 < \bar{\theta}_\omega \leq \frac{2\theta_1\theta_2}{\theta_1+\theta_2}$), then $a_1 \geq a_2 \leq a_3 \leq \ldots \leq a_K$, (namely (2.6.7) holds with $\tilde{i} = 2$), while if $\bar{\theta}_\omega \in (\theta_2, \theta_1)$ and $\left(\frac{\bar{\theta}_\omega}{\theta_1} + \frac{\bar{\theta}_\omega}{\theta_2}\right) > 2$ (i.e. $\frac{2\theta_1\theta_2}{\theta_1+\theta_2} < \bar{\theta}_\omega < \theta_1$), then $a_1 < a_2 \leq a_3 \leq \ldots \leq a_K$, namely (2.6.7) holds with $\tilde{i} = 1$. Furthermore, if $\bar{\theta}_\omega \in (\theta_K, \theta_{K-1})$ and $\left(\frac{\bar{\theta}_\omega}{\theta_K} + \frac{\bar{\theta}_\omega}{\theta_{K-1}}\right) \leq 2$ (i.e. $\theta_K < \bar{\theta}_\omega \leq \frac{2\theta_K\theta_{K-1}}{\theta_K+\theta_{K-1}}$), then $a_1 \geq a_2 \geq \ldots \geq a_{K-1} \geq a_K$,(namely (2.6.7) holds with $\tilde{i} = K$), whereas if if $\bar{\theta}_\omega \in (\theta_K, \theta_{K-1})$ and $\left(\frac{\bar{\theta}_\omega}{\theta_K} + \frac{\bar{\theta}_\omega}{\theta_{K-1}}\right) > 2$ (i.e. $\frac{2\theta_K\theta_{K-1}}{\theta_K+\theta_{K-1}} < \bar{\theta}_\omega < \theta_{K-1}$), then $a_1 \geq a_2 \geq \ldots \geq a_{K-1} < a_K$, namely (2.6.7) holds with $\tilde{i} = K-1$. The rest of the proof is straightforward.
The behaviour of the sequence $\{a_i, i = 1, \ldots, K\}$ will be used for the proof of Theorem 2.

Proof of Theorem 2
Maximization problem (2.3.3) can be address via Lagrange multipliers with $L(\boldsymbol{\rho}, \lambda_1, \ldots, \lambda_K) = \phi(\boldsymbol{\rho}) - \sum_{i=1}^{K-1} \lambda_i(\rho_{i+1} - \rho_i) - \lambda_K \left(\sum_{i=1}^{K} \rho_i - 1\right)$. By setting $\partial L(\boldsymbol{\rho}, \lambda_1, \ldots, \lambda_K)/\partial\rho_i = 0$ for $i = 1, \ldots, K$ we obtain

$$\begin{cases} a_1 + \lambda_1 = \lambda_K \\ a_i - \lambda_{i-1} + \lambda_i = \lambda_K, \quad i = 2, \ldots, K-1, \\ a_K - \lambda_{K-1} = \lambda_K \end{cases}$$

namely, by summing all the equations, $\lambda_i = \sum_{j=1}^{i}(\bar{a} - a_j)$ $(i = 1, \ldots, K-1)$ and $\lambda_K = \bar{a} = \frac{1}{K}\sum_{i=1}^{K}\left(1 - \frac{\bar{\theta}_\omega}{\theta_i}\right)^2 > 0$.
Case 1 $\lambda_i > 0 \quad \forall i = 1, \ldots, K-1$.
In this case $\rho_1 = \ldots = \rho_K$, namely the corresponding target is $\boldsymbol{\rho}^B$, so that

$$\bar{\theta}_\omega = \left(\sum_{i=1}^{K} \frac{1}{\theta_i^2}\right)^{-1} \sum_{i=1}^{K} \frac{1}{\theta_i} \tag{2.6.10}$$

and, from (2.6.5), $a_1 \geq a_2$ since $\left(\frac{1}{\theta_2} - \frac{1}{\theta_1}\right)^2 + \sum_{i=3}^{K} \frac{1}{\theta_i}\left\{\left(\frac{1}{\theta_i} - \frac{1}{\theta_1}\right) + \left(\frac{1}{\theta_i} - \frac{1}{\theta_2}\right)\right\} \geq 0$.
Condition $\lambda_1 > 0$, i.e., $\bar{a} > a_1$, corresponds to $\left(\sum_{i=1}^{K} \frac{1}{\theta_i^2}\right)^{-1} \sum_{i=1}^{K} \frac{1}{\theta_i} > 2\left(\sum_{i=1}^{K} \frac{1}{\theta_i} - \frac{K}{\theta_1}\right)\left(\sum_{i=1}^{K} \frac{1}{\theta_i^2} - \frac{K}{\theta_1^2}\right)^{-1} \Leftrightarrow$ $x > K^{-1}$. Then, from (vi) of Lemma 5, the sequence $\{a_i, i = 1, \ldots, K\}$ behaves as in (2.6.7), with at least $a_K > a_1$. Therefore $i\bar{a} > \sum_{j=1}^{i} a_j$ for $i = 2, \ldots, \tilde{i}$, while for $i > \tilde{i}$, the sequence becomes increasing. Since $K - 1 > (a_1 + \ldots + a_{K-1})/a_K$ (we recall that $a_i < a_K$ for $i = 1, \ldots, K-1$), then $(K-1)\bar{a} > \sum_{j=1}^{K-1} a_j$ and therefore $\lambda_i > 0$ for $i = \tilde{i} + 1, \ldots, K-1$.
Case 2 $\lambda_1 = 0$ and $\lambda_i > 0 \quad \forall i = 2, \ldots, K-1$.
In this case $\rho_2 = \cdots = \rho_K = \zeta$, so the ensuing optimal target is $(1 - [K-1]\zeta, \zeta, \ldots, \zeta)$ (where, clearly, $\zeta \in [0, K^{-1}]$), which is admissible iff $\bar{a} = a_1$ and

$$i\bar{a} > \sum_{j=1}^{i} a_j \qquad i = 2, \ldots, K-1. \tag{2.6.11}$$

Now, $\bar{a} = a_1$ corresponds to

$$\bar{\theta}_\omega = \frac{2\left(\sum_{i=1}^K \frac{1}{\theta_i} - \frac{K}{\theta_1}\right)}{\sum_{i=1}^K \frac{1}{\theta_i^2} - \frac{K}{\theta_1^2}}. \tag{2.6.12}$$

Moreover, $\bar{\theta}_\omega = \dfrac{\frac{1}{\theta_1} + \zeta\left(\sum_{i=1}^K \frac{1}{\theta_i} - \frac{K}{\theta_1}\right)}{\frac{1}{\theta_1^2} + \zeta\left(\sum_{i=1}^K \frac{1}{\theta_i^2} - \frac{K}{\theta_1^2}\right)}$, so that $\zeta = x$ and the ensuing target $\boldsymbol{\rho}^C$ is admissible provided that $x \leq K^{-1}$. From (iv) in Lemma 5, the sequence of $\{a_i, i = 1, \ldots, K\}$ behaves as in (2.6.7), with at least $a_K > a_1$, and we need to check that $\lambda_i > 0$ for every $i = 2, \ldots, K-1$.

- If $\theta_1 = \theta_2$, then $a_1 = a_2$ which contradicts (2.6.11) (clearly, the same reasoning holds in the more general case of a cluster of several superior treatments).
- If $\theta_1 > \theta_2 = \cdots = \theta_K$, then $a_2 = \cdots = a_K$ and, together with $\bar{a} = a_1$, implies that $a_1 = \cdots = a_K$ which contradicts (2.6.11).
- If $\theta_1 > \theta_2 \geq \cdots \geq \theta_K$ with $\theta_2 > \theta_K$, from (3.6.6) after tedious algebra it follows that $\left(\frac{\bar{\theta}_\omega}{\theta_1} + \frac{\bar{\theta}_\omega}{\theta_2}\right) < 2$, i.e., $a_1 > a_2$ and, combined with $\bar{a} = a_1$, it ensures that $\lambda_i > 0$ for $i = 2, \ldots, \tilde{i}$. Moreover for $i > \tilde{i}$, the sequence $\{a_i\}$ becomes increasing and, therefore, if $(K-1)\bar{a} > \sum_{i=1}^{K-1} a_i$ then $\lambda_i > 0$, for $i = \tilde{i}+1, \ldots, K-1$. This condition is trivially satisfied, since $\bar{a} = a_1$ and

$$(K-1)\bar{a} + a_K > \sum_{i=1}^{K-1} a_i + a_K \Longleftrightarrow (K-1)\bar{a} + a_K > K\bar{a} \Longleftrightarrow a_K > \bar{a} = a_1. \tag{2.6.13}$$

Case 3 $\lambda_i = 0 \quad \forall i = 1, \ldots, K-1$
In such a case, $\bar{a} = a_1 = \cdots = a_{K-1}$, which clearly implies that $a_K = \bar{a}$. As shown in 2.6.1, this implies that $\theta_1 = \cdots = \theta_j > \theta_{j+1} = \cdots = \theta_K$ (i.e., there are two clusters of treatments). Thus, every $\boldsymbol{\rho}^{\tilde{C}} = (\rho_1^{\tilde{C}}, \ldots, \rho_K^{\tilde{C}})^\top$ such that $\sum_{i=1}^j \rho_i^{\tilde{C}} = \frac{\theta_1}{\theta_1+\theta_K}$ and $\sum_{i=j+1}^K \rho_i^{\tilde{C}} = \frac{\theta_K}{\theta_1+\theta_K}$, is optimal. Indeed, in such a case $\bar{\theta}_\omega = 2\theta_1\theta_K/(\theta_1+\theta_K)$ and $\phi(\boldsymbol{\rho}^{\tilde{C}}) = \bar{a} = (\theta_1-\theta_K)^2/(\theta_1+\theta_K)^2$, that coincides with the maximum of ϕ in the unconstrained optimization. Moreover, adopting $\boldsymbol{\rho}^{\tilde{C}}$, $x = \frac{\theta_K}{(K-j)(\theta_1+\theta_K)}$ and therefore, $\sum_{i=1}^j \rho_i^{\tilde{C}} = 1 - x(K-j)$=1 $- \sum_{i=j+1}^K \rho_i^{\tilde{C}}$. Since the components of $\boldsymbol{\rho}^{\tilde{C}}$ are ordered according to the magnitude of the treatment effects, then $\rho_j^{\tilde{C}} \leq \rho_k^{\tilde{C}}$ (for $k = 1, \ldots, j-1$) and $\rho_{j+1}^{\tilde{C}} \geq \rho_k^{\tilde{C}}$ (for $k = j+1, \ldots, K$). Thus, $\rho_j^{\tilde{C}} \leq \frac{1-x(K-j)}{j}$, $\rho_{j+1}^{\tilde{C}} \geq \frac{x(K-j)}{K-j} = x$ and, clearly, $j^{-1}[1-x(K-j)] \geq x$, i.e., $x \leq K^{-1}$.
Case 4 $\lambda_1 > 0$ and at least one $\lambda_i = 0$ for $i \in [2; K-1]$.
Under this scenario, $\bar{a} > a_1$ and thus (2.6.9) of Lemma 5 is impossible. Therefore, the sequence $\{a_i, i = 1, \ldots, K\}$ behaves as in (2.6.7) or (2.6.8). Moreover, $\lambda_i = 0 \Leftrightarrow i = \sum_{j=1}^i a_j/\bar{a}$ (with $2 \leq i \leq K-1$). Since $\frac{a_1}{\bar{a}} < 1$, then it exists at least one a_l (with $2 \leq l \leq i$) such that $\frac{a_l}{\bar{a}} > 1$, i.e. $\bar{a} < a_l$, and clearly $\bar{a} < a_l \leq a_i \leq a_{i+1}$. However, this is impossible since

- if $\lambda_{i+1} = 0$, then $(i+1)\bar{a} = \sum_{j=1}^i a_j + a_{i+1} \Leftrightarrow \bar{a} = a_{i+1}$ which contradicts $\bar{a} < a_{i+1}$;
- if $\lambda_{i+1} > 0$, then $(i+1)\bar{a} > \sum_{j=1}^i a_j + a_{i+1} \Leftrightarrow \bar{a} > a_{i+1}$ but this is impossible since $\bar{a} < a_{i+1}$.

Case 5 $\lambda_1 = 0$ and at least one $\lambda_i > 0$ and $\lambda_{i+1} = 0$ with $i \in \{2, \dots, K-2\}$
Under this setting, $\bar{a} = a_1$ and from Lemma 5 the sequence behaves as in (2.6.7). Since $\lambda_{i+1} = 0$, then

$$i = \frac{a_2}{\bar{a}} + \cdots + \frac{a_{i+1}}{\bar{a}}. \tag{2.6.14}$$

From (2.6.14) and $\lambda_i > 0$ (i.e. $i > \frac{a_1}{\bar{a}} + \frac{a_2}{\bar{a}} + \cdots + \frac{a_i}{\bar{a}}$) it follows that $\bar{a} = a_1 < a_{i+1}$ and, given the behaviour of $\{a_i, i = 1, \dots, K\}$, also $a_{i+1} \leq a_{i+2}$. This scenario is impossible since

- if $\lambda_{i+2} = 0$, then $(i+2)\bar{a} = a_1 + \cdots + a_{i+1} + a_{i+2}$, which combined with (2.6.14), gives $\bar{a} = a_{i+2}$ contradicting that $\bar{a} < a_{i+1} \leq a_{i+2}$;

- if $\lambda_{i+2} > 0$, then $(i+2)\bar{a} > a_1 + \cdots + a_{i+1} + a_{i+2}$, which combined with (2.6.14), gives $\bar{a} > a_{i+2}$ but this is impossible.

Case 6 $\lambda_1 = \lambda_2 = \cdots = \lambda_j = 0$ and $\lambda_i > 0 \; \forall i = j+1, \dots, K-1$
Under this setting, the ensuing optimal target is $\boldsymbol{\rho}^{\breve{C}} = (\rho_1^{\breve{C}}, \dots, \rho_j^{\breve{C}}, \nu, \dots, \nu)^\top$ with $\rho_i^{\breve{C}} \geq \rho_{i+1}^{\breve{C}} \geq \nu$ for $i = 1, \dots, j-1$ and $\sum_{i=1}^{j} \rho_i^{\breve{C}} = 1 - (K-j)\nu$ (where, clearly, $\nu \leq K^{-1}$). This target is admissible iff i) $\bar{a} = a_1 = \ldots = a_j$ and ii) $i\bar{a} > \sum_{k=1}^{i} a_k$ for $i = j+1, \ldots K-1$.
Condition $\lambda_1 = 0$ implies $\bar{a} = a_1$, so that (3.6.6) holds and, from (iv) of Lemma 5, the sequence $\{a_i, i = 1, \dots, K\}$ behaves as in (2.6.7) with $a_K > a_1$.

- If $\theta_1 > \theta_2$ then by (2.6.5), $a_1 = a_2 \Leftrightarrow \left(\frac{\bar{\theta}_\omega}{\theta_1} + \frac{\bar{\theta}_\omega}{\theta_2}\right) = 2$. Since $\theta_1 > \bar{\theta}_\omega$ it has to be $\theta_2 < \bar{\theta}_\omega$, namely $a_1 = a_2 \leq a_3 \leq \ldots \leq a_K$, which is impossible given that $\bar{a} = a_1$. Similar reasoning applies for all the pairs $\theta_i > \theta_{i+1}$ for $i = 2, \dots, j-1$.

- If $\theta_1 = \cdots = \theta_j > \theta_{j+1} \geq \ldots \geq \theta_K$, with $j \in \{2, \dots, K-1\}$, then $a_1 = \cdots = a_j$ and

$$\bar{\theta}_\omega = \frac{[1-(K-j)\nu]\theta_1^{-1} + \nu \sum_{i=j+1}^{K} \theta_i^{-1}}{[1-(K-j)\nu]\theta_1^{-2} + \nu \sum_{i=j+1}^{K} \theta_i^{-2}}$$

  so that, from (3.6.6), after tedious algebra it follows that $\nu = x$ (clearly, $x \leq K^{-1}$ in order to be admissible). Finally, we need to verify ii). These conditions are trivially satisfied following the same arguments of Case 2 (see (2.6.13)).

Since the components of $\boldsymbol{\rho}^{\breve{C}}$ are ordered according to the magnitude of the treatment effects, then $\rho_j^{\breve{C}} \leq \rho_k^{\breve{C}}$ (for $k = 1, \dots, j-1$) and, clearly, $\rho_j^{\breve{C}} \leq \frac{1-x(K-j)}{j}$ (recalling that $\sum_{i=1}^{j} \rho_i^{\breve{C}} = 1 - (K-j)x$). Moreover, given the above-mentioned ordering, $j^{-1}[1 - x(K-j)] \geq x$, i.e., $x \leq K^{-1}$. When $\theta_{j+1} = \ldots = \theta_K$, then $x = \frac{\theta_K}{(K-j)(\theta_1+\theta_K)}$ and this is a special case of Case 3. In addition, note that under $\boldsymbol{\rho}^{\breve{C}}$,

$$\phi(\boldsymbol{\rho}^{\breve{C}}) = a_1[1-(K-j)x] + x \sum_{j=i+1}^{K} a_i = \bar{a} - K\bar{a}x + K\bar{a}x = \bar{a}, \tag{2.6.15}$$

2.6.3 Proof of Theorem 3

Clearly, when $x > K^{-1}$, $\boldsymbol{\rho}^C = \boldsymbol{\rho}^B$ and thus, we will consider only the case $x \leq K^{-1}$. Denote by $\bar{\theta}^B_\omega = \sum_{i=1}^K \frac{1}{\theta_i} / \sum_{i=1}^K \frac{1}{\theta_i^2}$ the expression in (2.6.10). Then, from (2.6.3), $\phi(\boldsymbol{\rho}^B) = \frac{1}{K}\sum_{i=1}^K \left(1 - \frac{\bar{\theta}^B_\omega}{\theta_i}\right)^2 = 1 - \bar{\theta}^B_\omega \frac{1}{K}\sum_{i=1}^K \frac{1}{\theta_i}$.

- Under statement (i) of Theorem 2, $\boldsymbol{\rho}^C = (\rho_1^C, \ldots, \rho_j^C, x, \ldots, x)^\top$, and letting $\bar{\theta}^C_\omega$ be (2.6.1) under $\boldsymbol{\rho}^C$, then from (2.6.15) follows that $\phi(\boldsymbol{\rho}^C) = \frac{1}{K}\sum_{i=1}^K \left(1 - \frac{\bar{\theta}^C_\omega}{\theta_i}\right)^2 = 1 - 2\bar{\theta}^C_\omega \frac{1}{K}\sum_{i=1}^K \frac{1}{\theta_i} + \left(\bar{\theta}^C_\omega\right)^2 \frac{1}{K}\sum_{i=1}^K \frac{1}{\theta_i^2}$.

  Thus, $\phi(\boldsymbol{\rho}^B) \leq \phi(\boldsymbol{\rho}^C)$ since $-\bar{\theta}^B_\omega \frac{1}{K}\sum_{i=1}^K \frac{1}{\theta_i} \leq -2\bar{\theta}^C_\omega \frac{1}{K}\sum_{i=1}^K \frac{1}{\theta_i} + \left(\bar{\theta}^C_\omega\right)^2 \frac{1}{K}\sum_{i=1}^K \frac{1}{\theta_i^2} \Leftrightarrow (\bar{\theta}^B_\omega - \bar{\theta}^C_\omega)^2 \geq 0$.

- Under statement (ii) of Theorem 2, then $\phi(\boldsymbol{\rho}^C) = \phi(\tilde{\boldsymbol{\rho}}) = \left(\frac{\theta_1 - \theta_K}{\theta_1 + \theta_K}\right)^2$ and, clearly, $E_\phi(\boldsymbol{\rho}^C) = 1$.

Taking now into account the ethical efficiency, denoting by $\bar{\theta} = K^{-1}\sum_{i=1}^K \theta_i$, then $E_e(\boldsymbol{\rho}^B) = \frac{\bar{\theta}}{\theta_1}$. Under statement (i) of Theorem 2 we obtain $E_e(\boldsymbol{\rho}^C) = \theta_1^{-1}\left\{\theta_1[1-(K-j)]x + x\sum_{i=j+1}^K \theta_i\right\} = \theta_1^{-1}\left[\theta_1(1-Kx) + Kx\bar{\theta}\right]$. Thus, $E_e(\boldsymbol{\rho}^C) \geq E_e(\boldsymbol{\rho}^B)$ since $\theta_1(1-Kx) \geq \bar{\theta}(1-Kx)$ (recalling that $\theta_1 > \bar{\theta}$). The case under statement (ii) follows easily by similar arguments.

Bibliography

[1] Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med* 1987; 317(3): 141–145.

[2] Peckham E, Brabyn S, Cook L, et al. The use of unequal randomisation in clinical trials: An update. *Contemp Clin Trials* 2015; 45: 113–122. 10th Anniversary Special Issue.

[3] Sverdlov O, Ryeznik Y. Implementing unequal randomization in clinical trials with heterogeneous treatment costs. *Stat Med* 2019; 38(16): 2905–2927.

[4] Avins AL. Can unequal be more fair? Ethics, subject allocation, and randomised clinical trials. *J Med Ethics* 1998; 24(6): 401–408.

[5] Berry DA. Adaptive Clinical Trials: The promise and the caution. *J Clin Oncol* 2011; 29(6): 606–609.

[6] Zelen M. Application of exponential models to problems in cancer research. *J R Stat Soc Series A Stat Soc* 1966; 129(3): 368–398.

[7] Shulz D, Chernichovsky D, Allweis C. A novel method for quantifying passive-avoidance behavior based on the exponential distribution of step-through latencies. *Pharmacol Biochem Behav* 1986; 25(5): 979–983.

[8] Herbst RS, Baas P, Kim D, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet* 2016; 387(10027): 1540–1550.

[9] Fountzilas G, Ciuleanu E, Dafni U, et al. Concomitant radiochemotherapy vs radiotherapy alone in patients with head and neck cancer. *Med Oncol* 2004; 21(2): 95–107.

[10] Cook RD, Wong WK. On the equivalence of constrained and compound optimal designs. *J Am Stat Assoc* 1994; 89(426): 687–692.

[11] Jennison C, Turnbull BW. *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton, FL: Chapman & Hall/CRC Press; 1999.

[12] Baldi Antognini A, Rosenberger WF, Wang Y, et al. Exact optimum coin bias in Efron's randomization procedure. *Stat Med* 2015; 34(28): 3760–3768.

[13] Tymofyeyev Y, Rosenberger WF, Hu F. Implementing optimal allocation in sequential binary response experiments. *J Am Stat Assoc* 2007; 102(477): 224–234.

[14] Biswas A, Mandal S, Bhattacharya R. Multi-treatment optimal response-adaptive designs for phase III clinical trials. *J Korean Stat Soc* 2011; 40(1): 33–44.

[15] Baldi Antognini A, Novelli M, Zagoraiou M. Optimal designs for testing hypothesis in multiarm clinical trials. *Stat Methods Med Res* 2019; 28(10-11): 3242–3259.

[16] Sverdlov O, Ryeznik Y, Wong WK. On optimal designs for clinical trials: An updated review. *J Stat Theory Pract* 2020; 14(10).

[17] Rosenberger WF, Lachin JM. *Randomization in clinical trials: theory and practice*. New York, NY: Wiley; 2002.

[18] Baldi Antognini A, Giovagnoli A. *Adaptive Designs for Sequential Treatment Allocation*. Boca Raton: Chapman & Hall/CRC Press; 2015.

[19] US Food and Drug Administration. Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics. `https://www.fda.gov/media/78495/download`. *Published December 2019. Accessed January 2020.*

[20] Zhang L, Rosenberger WF. Response-adaptive randomization for survival trials: the parametric approach. *J R Stat Soc Series C Appl Stat* 2007; 56(2): 153–165.

[21] Zhu H, Hu F. Implementing optimal allocation for sequential continuous responses with multiple treatments. *J Stat Plan Inference* 2009; 139(7): 2420–2430.

[22] Sverdlov O, Tymofyeyev Y, Wong WK. Optimal response-adaptive randomized designs for multi-armed survival trials. *Stat Med* 2011; 30(24): 2890–2910.

[23] Hu F, Zhang LX. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *Ann Stat* 2004; 32(1): 268–301.

[24] Hochberg Y, Tamhane A. *Multiple Comparison Procedures*. New York, NY: Wiley; 1987.

[25] Christensen R. *Plane Answers to Complex Questions*. New York, NY: Springer; 2002.

[26] Sverdlov O, Rosenberger WF. On recent advances in optimal allocation designs in clinical trials. *J Stat Theory Pract* 2013; 7: 753–773.

[27] Wong WK, Zhu W. Optimum treatment allocation rules under a variance heterogeneity model. *Stat Med* 2008; 27(22): 4581–4595.

[28] Rosenberger WF, Hu F. Maximizing power and minimizing treatment failures in clinical trials. *Clin Trials* 2004; 1(2): 141–147.

[29] Lawless J. *Statistical Models and Methods for Lifetime Data*. 2nd ed. New York, NY: Wiley; 2011.

[30] Latta RB. A Monte Carlo study of some two-sample rank tests with censored data. *J Am Stat Assoc* 1981; 76(375): 713–719.

[31] Rosenberger WF, Seshaiyer P. Adaptive survival trials. *J Biopharm Stat* 1997; 7(4): 617–624.

[32] Hu F, Zhang LX, Cheung SH, et al. Doubly adaptive biased coin designs with delayed responses. *Can J Stat* 2008; 36(4): 541–559.

[33] Sverdlov O, Rosenberger WF, Ryeznik Y. Utility of Covariate-Adjusted Response-Adaptive Randomization in survival trials. *Stat Biopharm Res* 2013; 5(1): 38–53.

[34] Jung S, George S. Between-arm comparisons in randomized Phase II trials. *J Biopharm Stat* 2009; 19(3): 456–468.

[35] Simon R, Wittes R, Ellenberg S. Randomized phase II clinical trials. *Cancer Treat Rep* 1985; 69(12): 1375–1381.

[36] Lee J, Feng L. Randomized phase II designs in cancer clinical trials: current status and future directions. *J Clin Oncol* 2005; 23(19): 4450–4457.

[37] Hedayat AS, Jacroux M, Majumdar D. Optimal designs for comparing test treatments with controls (with discussion). *Statist Sci* 1988; 3(4): 462–476.

[38] Magaret A, Angus D, Adhikari N, et al. Design of a multi-arm randomized clinical trial with no control arm. *Contemp Clin Trials* 2016; 46: 12–17.

Chapter 3

Paper B

Optimal designs for testing the efficacy of heterogeneous experimental groups

Alessandro Baldi Antognini[1], *Rosamarie Frieri*[1], *Marco Novelli*[1] *and Maroussa Zagoraiou*[1]

[1] Department of Statistical Sciences, University of Bologna

Abstract

This paper develops optimal design theory for testing the efficacy of several competing treatments. Adopting the general framework of heteroscedastic treatment groups, we derive the design maximizing the power of the multivariate test of homogeneity. In general, this optimal design is a generalized Neyman allocation involving only two experimental groups. In order to account for the ordering among the treatments, particularly relevant in the clinical context for ethical reasons, we provide the optimal design for testing under constraints reflecting their effectiveness. The advantages of the suggested allocations are illustrated both theoretically and through several numerical examples, also compared with other designs proposed in the literature, showing a substantial gain in terms of both power and ethics.

3.1 Introduction

This paper addresses the issue of designing experiments for comparing several treatments when the principal inferential aim is testing the homogeneity of the treatment effects. Starting from the classical one-way ANOVA of Sir. R.A. Fisher, the problem of comparing the equality of several means has a long history in the statistical literature and covers all the applied fields. Over the past 50's, there has been a growing stream of papers about the design of experiments for treatment comparisons; however, they are almost exclusively focused on estimation precision. In particular, having in mind the linear homoscedastic model set-up, balancing the allocations among treatments is often considered as desirable, since this strategy optimizes the usual alphabetical criteria for the estimation of the treatment effects. However, balance could be highly inefficient in the case of heteroscedasticity, or when inference is focused on the treatment contrasts; moreover, it could be strongly inappropriate for clinical trials, since the demand of individual care often induces to skew the allocations to the best performing drugs.
Although the ability of detecting a significant treatment difference is a fundamental issue for statistical inference, in the design of experiment literature very little attention has been devoted to hypothesis testing, also due to the underlined complex mathematical structure. Only recently there has been a growing interest on this topic, in particular in the clinical/pharmaceutical research, also due to the encouragement of Health Authorities[3]. In the context of binary trials, Tymofyeyev et al[9] were among the first authors to derive the design maximizing the power of the test of homogeneity. In general, this design is a degenerate allocation involving only the best and the worst treatments, with no observations on the intermediate ones: for this reason, a lower bound of each treatment allocation proportion is superimposed and the related constrained optimal design is derived. By applying the same methodology to exponential outcomes, Zhu and Hu[11] derived the corresponding optimal allocations, while Sverdlov et al[8] extended their results in the presence of censoring. In general, the ensuing designs are discontinuous (non-degenerate) functions of the unknown model parameters (i.e., they are locally optimal) and, by suitable smoothing transformations, they could be implemented in a sequential fashion via response adaptive randomization procedures, namely sequential rules that change the treatment allocation probabilities to approximate the chosen target (for a review see Baldi Antognini and Giovagnoli[1] and Rosenberger and Lachin[6]).

Recently, Baldi Antognini et al[2] derived the design maximizing the power of the test of homogeneity for normal homoscedastic data, which is a balanced allocation involving only the best and the worst treatments; moreover, by imposing the ethical constraint that the treatment allocation proportions should reflect the a-priori unknown ordering among their effects, they also derived a non-degenerate optimal target implementable via response-adaptive randomization. Under the same framework, assuming that the treatment ordering is known, Singh and Davidov[7] discussed the optimal designs for restricted and unrestricted statistical inference (which are equivalent for large samples) by adopting a maxi-min approach, in order to overcome local optimality problems.

The aim of the present paper is to present a unified framework for deriving optimal designs for hypothesis testing in the presence of several experimental groups, also encompassing the general ANOVA set-up with heteroscedastic errors. In particular, the optimal designs are generalized Neyman allocations involving only two treatments, not necessarily the best and the worst ones. In order to account for the ordering among treatments (which could be particularly relevant in the clinical context, for ethical reasons), we derive constrained optimal designs, where the allocation proportions are themselves ordered as the treatment efficacies. Since the ordering among the effects is generally a-priori unknown, the ensuing allocations are locally optimal designs that can be approached by response-adaptive randomization

procedures after suitable smoothing techniques. Several illustrative examples are provided for normal, binary, Poisson and exponential data (with and without censoring), also amending some previously obtained results (like e.g. Sverdlov et al[8]). The properties of these designs are described both theoretically and through numerical examples, showing a substantial gain in terms of power and ethics as well.

The paper is structured as follows. Section 3.2 deals with optimal designs for hypothesis testing, taking into account both unconstrained and constrained optimization. Section 3.3 discusses the performance of the proposed allocations both analytically and through numerical examples, also compared with other designs suggested in the literature. Section 3.4 deals with a general discussion about our results, including also their implementation via response-adaptive randomization and Section 3.5 concludes the paper. The mathematical details are available in Section 3.6.

3.2 Main results

3.2.1 Preliminaries

Suppose we have $K \geq 2$ competing treatments and let $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{iK})^\top$ be the indicator managing the allocation of the ith subject, namely $\delta_{ik} = 1$ if he/she is assigned to treatment k ($k = 1, \ldots, K$) and 0 otherwise. Given the assignments, the observations Y_is are assumed to be iid belonging to the exponential family parametrised in such a way that $\theta_k \in \Theta \subseteq \mathbb{R}$ denotes the mean effect of treatment k, while $v_k = v(\theta_k) \in \mathbb{R}^+$ represents the corresponding variance ($k = 1, \ldots, K$) and we set $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^\top$ and $\boldsymbol{v} = (v_1, \ldots, v_K)^\top$. Special cases of practical relevance are binary $B(\theta_k)$ ($\theta_k \in (0;1)$, $v(\theta_k) = \theta_k(1-\theta_k)$) and Poisson $P(\theta_k)$ ($\theta_k \in \mathbb{R}^+$, $v(\theta_k) = \theta_k$) trials for dichotomous and count data, respectively, while normal model $N(\theta_k; v_k)$ (with $\theta_k \in \mathbb{R}$ and $v(\theta_k) = v_k$ independent from θ_k) is also encompassed for continuous responses as well as the exponential one $exp(\theta_k)$ ($\theta_k \in \mathbb{R}^+$, $v(\theta_k) = \theta_k^2$) for survival outcomes.

After n allocations, let $\boldsymbol{N}_n = \sum_{i=1}^n \boldsymbol{\delta}_i$, where $N_{nk} = \sum_{i=1}^n \delta_{ik}$ denotes the number of assignments to treatment k and, clearly, $\boldsymbol{N}_n^\top \mathbf{1}_K = n$ (here $\mathbf{1}_K$ is the K-dim vector of ones); while $\boldsymbol{\rho} = n^{-1}\boldsymbol{N}_n$ is the vector of the treatment allocation proportions, where $\rho_k = n^{-1}N_{nk} \geq 0$ for $k = 1, \ldots, K$ and $\boldsymbol{\rho}^\top \mathbf{1}_K = 1$ for every n. Let $\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_{n1}, \ldots, \hat{\theta}_{nK})^\top$ be the MLEs of the treatment effects $\boldsymbol{\theta}$ (i.e., the sample means $\hat{\theta}_{nk} = N_{nk}^{-1} \sum_{i=1}^n \delta_{ik} Y_i$), under well-known regularity conditions $\hat{\boldsymbol{\theta}}_n$ is strongly consistent and asymptotically normal with $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathrm{N}(\mathbf{0}_K, \mathbb{M}^{-1})$, where $\mathbf{0}_K$ is the K-dim vector of zeros and $\mathbb{M} = \mathbb{M}(\boldsymbol{\rho}) = \mathrm{diag}\,(\rho_k/v_k)_{k=1,\ldots,K}$ is the Fisher information associated with $\boldsymbol{\theta}$.

In this setting the inferential focus is on the contrasts $\mathbb{A}^\top\boldsymbol{\theta}$ where, considering without loss of generality (wlog) the first treatment as the reference one, $\mathbb{A}^\top = [\mathbf{1}_{K-1}| - \mathbb{I}_{K-1}]$ and $\mathbb{I}_{K-1}$ is the $(K-1)$-dim identity matrix. The MLE $\mathbb{A}^\top\hat{\boldsymbol{\theta}}_n$ is strongly consistent and asymptotically normal with $\sqrt{n}\mathbb{A}^\top(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathrm{N}(\mathbf{0}_{K-1}, \mathbb{A}^\top\mathbb{M}^{-1}\mathbb{A})$. Let $\hat{v}_{kn}$s be consistent estimators of the treatment variances, then $\hat{\mathbb{M}}_n = \mathrm{diag}\,(\rho_k/\hat{v}_{kn})_{k=1,\ldots,K}$ and Wald statistic $W_n = n\hat{\boldsymbol{\theta}}_n^\top\mathbb{A}[\mathbb{A}^\top\hat{\mathbb{M}}_n^{-1}\mathbb{A}]^{-1}\mathbb{A}^\top\hat{\boldsymbol{\theta}}_n$ is usually employed for testing the hypothesis of homogeneity of the effects $\mathrm{H}_0 : \mathbb{A}^\top\boldsymbol{\theta} = \mathbf{0}_{K-1}$ vs $\mathrm{H}_1 : \mathbb{A}^\top\boldsymbol{\theta} \neq \mathbf{0}_{K-1}$. Under H_0, $W_n \xrightarrow{d} \chi^2_{K-1}$, namely it converges to a (central) χ^2 with $K-1$ degrees of freedom (df) (provided that $\rho_k > 0$ for every $k = 1, \ldots, K$), while under the alternative W_n converges to a non-central χ^2 distribution with $K-1$ df and non-centrality parameter (NCP) $n\phi$, denoted by $\chi^2_{K-1}(n\phi)$, where $\phi = \phi(\boldsymbol{\rho}) = \boldsymbol{\theta}^\top\mathbb{A}[\mathbb{A}^\top\mathbb{M}^{-1}\mathbb{A}]^{-1}\mathbb{A}^\top\boldsymbol{\theta}$. Thus, the power of the α-level test could be approximated by $\Pr\left(\chi^2_{K-1}(n\phi(\boldsymbol{\rho})) > q_{K-1,\alpha}\right)$, where $q_{c,\alpha}$ is the $(1-\alpha)$-percentile of a χ^2_c. For fixed df the non-central χ^2 distribution is stochastically increasing in the NCP so that the power is an increasing function of

$\phi(\boldsymbol{\rho})$. Thus, the problem is finding the design $\tilde{\boldsymbol{\rho}}$, defined on the simplex $\tilde{\rho}_k \geq 0$ for every $k = 1, \ldots, K$ with $\tilde{\boldsymbol{\rho}}^\top \mathbf{1}_K = 1$, maximizing $\phi(\cdot)$. As is well-known, for $K = 2$ the NCP is maximized by the Neyman allocation $\tilde{\rho}_1 = (1 + \sqrt{v_2/v_1})^{-1} = 1 - \tilde{\rho}_2$.

Notice that both the likelihood ratio and score tests are asymptotically equivalent to the Wald test and so are their power functions. Moreover, for normal outcomes $N(\theta_k; v_k)$, $W_n \sim \chi^2_{K-1}(n\phi(\boldsymbol{\rho}))$ for any sample size n and, in the case of homoscedasticity ($v_k = v$, for $k = 1, \ldots, K$), likelihood ratio and Wald's tests coincide.

3.2.2 Unconstrained optimal design for testing

In this section we derive the optimal design maximizing the NCP. Assuming without loss of generality that high responses are preferable, the treatment outcomes are ordered on the basis of their effects and, for ease of notation (without loss of generality), we assume that $\theta_1 \geq \ldots \geq \theta_K$ (i.e., the best treatment will be labelled as the first one, while the Kth treatment as the worst, admitting also clusters with the same efficacy), with at least one strict inequality. We wish to stress that this is a simple label-coding intended to avoid more complex notation like $\theta_{max} \geq \cdots \geq \theta_{min}$; clearly, the treatment ranking is a priori unknown, since the treatment effects are a priori unknown too (we do not know which treatment is associated to label 1 and which one to label K before the experiment). The treatment ordering can be then estimated sequentially, as we will discuss in Section 3.4.

Before the experiment, the treatment ordering is unknown, since the treatment effects are a priori unknown too (we do not know which treatment is associated to label 1 and which one to label K): the coding $\theta_1 \geq \cdots \geq \theta_K$ has been adopted in order to avoid a more complex notation like $\theta_{max} \geq \cdots \geq \theta_{min}$.

For Bernoulli, Poisson, exponential and normal homoscedastic models, this ordering corresponds to the classical stochastic order, while it does not imply a specific ordering for normal heteroscedastic data.

For a given design $\boldsymbol{\rho}$, let $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)^\top$ with $\pi_k = \pi_k(\boldsymbol{\rho}) = \rho_k v_k^{-1} / (\sum_{i=1}^K \rho_i v_i^{-1}) \geq 0$ ($k = 1, \ldots, K$) then $\boldsymbol{\pi}^\top \mathbf{1}_K = 1$ and, after straightforward calculation,

$$\phi(\boldsymbol{\rho}) = \left(\sum_{k=1}^K \rho_k v_k^{-1} \right) V_{\boldsymbol{\pi}}(\theta), \tag{3.2.1}$$

where $V_{\boldsymbol{\pi}}(\theta) = \sum_{k=1}^K (\theta_k - \bar{\theta}_{\boldsymbol{\pi}})^2 \pi_k$ and $\bar{\theta}_{\boldsymbol{\pi}} = \sum_{k=1}^K \theta_k \pi_k$ are the variance and the mean of a discrete r.v. θ (with K possibly different ordered support points $\theta_1 \geq \ldots \geq \theta_K$), evaluated with respect to the pdf $\boldsymbol{\pi}$. Clearly, for homoscedastic treatment groups $v_k = v$ ($k = 1, \ldots, K$) and $\boldsymbol{\pi} = \boldsymbol{\rho}$.

Remark 3.2.1. *This representation is quite general and covers the case of exponential outcomes subject to an independent right censoring scheme (which is a common feature of survival trials). Indeed, let $\epsilon_k = \epsilon(\theta_k) : \mathbb{R}^+ \to (0; 1)$ be the probability that a failure/death occurs before censoring in the kth group ($k = 1 \ldots, K$), that are assumed to be constant for every subject in each group, then ϵ is a decreasing function depending on the particular censoring scheme adopted in the trial (one of the most general is described in[10]). In such a case, $\phi(\boldsymbol{\rho})$ in (3.2.1) should be simply re-parametrized by substituting each treatment variance $v_k = \theta_k^2$ with $\dot{v}_k = \theta_k^2 / \epsilon(\theta_k)$ ($k = 1, \ldots, K$).*

The next Lemma shows some general properties of the function $\phi(\boldsymbol{\rho})$.

Lemma 3.2.1. *Let* $\boldsymbol{a} = (a_1, \dots, a_K)^\top$, *with* $a_k = v_k^{-1}\left(\theta_k - \bar{\theta}_{\boldsymbol{\pi}}\right)^2 \geq 0$ *for* $k = 1, \dots, K$. *Then,* $\phi(\boldsymbol{\rho}) = \boldsymbol{a}^\top \boldsymbol{\rho}$ *is a concave super-harmonic function with non-negative gradient* $\nabla\phi = \boldsymbol{a}$ *and Laplacian* $\nabla^2\phi < 0$.

Proof. See Appendix 3.6.1. □

From now on, we denote by $\boldsymbol{e}_1, \dots, \boldsymbol{e}_K$ the canonical base of $\mathbb{R}^K$ and we set $\varsigma_{ik} = \left(1 + \sqrt{v_k/v_i}\right)^{-1} = 1 - \varsigma_{ki}$, for any $i, k \in \{1, \dots, K\}$.

Theorem 3.2.1. *For every sample size* n, *the unconstrained optimal design* $\tilde{\boldsymbol{\rho}}$ *maximizing the NCP of Wald test is such that*

$$\phi(\tilde{\boldsymbol{\rho}}) = \max_{i,k \in \{1,\dots,K\}} \left(\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}}\right)^2. \tag{3.2.2}$$

Thus, if the pair of treatments $\{\tilde{i}, \tilde{k}\}$ *maximizing the of (3.2.2) is unique, then*

$$\tilde{\boldsymbol{\rho}} = \tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} = \boldsymbol{e}_{\tilde{i}}\varsigma_{\tilde{i}\tilde{k}} + \boldsymbol{e}_{\tilde{k}}\varsigma_{\tilde{k}\tilde{i}} = \left(0, \dots, 0, \frac{\sqrt{v_{\tilde{i}}}}{\sqrt{v_{\tilde{i}}} + \sqrt{v_{\tilde{k}}}}, 0, \dots, 0, \frac{\sqrt{v_{\tilde{k}}}}{\sqrt{v_{\tilde{i}}} + \sqrt{v_{\tilde{k}}}}, 0, \dots, 0\right)^\top,$$

i.e., $\tilde{\boldsymbol{\rho}}$ *corresponds to a Neyman allocation involving only this pair of treatments.*
Moreover, if the pair $\{\tilde{i}, \tilde{k}\}$ *maximizing the right-hand side (RHS) of (3.2.2) is not unique, but there exists one (or more) other pair(s) of indexes* $\{\tilde{i}', \tilde{k}'\}$ *such that*

$$\left(\frac{\theta_{\tilde{i}} - \theta_{\tilde{k}}}{\sqrt{v_{\tilde{i}}} + \sqrt{v_{\tilde{k}}}}\right)^2 = \left(\frac{\theta_{\tilde{i}'} - \theta_{\tilde{k}'}}{\sqrt{v_{\tilde{i}'}} + \sqrt{v_{\tilde{k}'}}}\right)^2 = \max_{i,k \in \{1,\dots,K\}} \left(\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}}\right)^2, \tag{3.2.3}$$

then $\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}$ *and* $\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}$ *are both optimal as well as every mixture, namely*

$$\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \phi\left(\omega\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} + (1-\omega)\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}\right) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}), \quad \forall\omega \in [0;1].$$

Example 3.2.1. *Let* $K = 3$ *be the treatments to be compared, with* $\boldsymbol{\theta} = (3, 2, 1)^\top$ *and* $\boldsymbol{v} = (1, 4, 9)^\top$. *From (3.2.2), the maximum of* ϕ *is attained at the pair* $\{1, 3\}$, *so that the optimal design is* $\tilde{\boldsymbol{\rho}}_{13} = (1/4, 0, 3/4)$, *with* $\phi\left(\tilde{\boldsymbol{\rho}}_{13}\right) = 1/4$. *Under the same setting, if* $v_3 = 25$ *(instead of* $v_3 = 9$*) then* $\tilde{\boldsymbol{\rho}}_{12}$ *and* $\tilde{\boldsymbol{\rho}}_{13}$ *are both optimal designs with* $\phi(\tilde{\boldsymbol{\rho}}_{12}) = \phi(\tilde{\boldsymbol{\rho}}_{13}) = 1/9$; *moreover, every convex combination of* $\tilde{\boldsymbol{\rho}}_{12}$ *and* $\tilde{\boldsymbol{\rho}}_{13}$ *is still optimal, i.e., every design* $\tilde{\boldsymbol{\rho}} = (\tilde{\rho}_1, 4\tilde{\rho}_1 - 2/3, 5/3 - 5\tilde{\rho}_1)^\top$ *with* $\tilde{\rho}_1 \in [1/3; 1/6]$ *is such that* $\phi(\tilde{\boldsymbol{\rho}}) = 1/9$.

The presence of clusters of treatments is also accounted for as a special case, as the following Corollary shows.

Corollary 3.2.1. *If* $\{\tilde{i}, \tilde{k}\}$ *is the pair of treatments maximizing the RHS of (3.2.2) and there exists one (or more) other treatment(s)* $\tilde{i}'$ *such that* $\theta_{\tilde{i}} = \theta_{\tilde{i}'}$ *and* $v_{\tilde{i}} = v_{\tilde{i}'}$ *(i.e., there exists a cluster of equal treatments involved in the maximization), then also* $\{\tilde{i}', \tilde{k}\}$ *maximizes the RHS of (3.2.2) and*

$$\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \phi(\omega\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} + (1-\omega)\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}}) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}}), \quad \forall\omega \in [0;1],$$

namely every design $\tilde{\boldsymbol{\rho}}$ *s.t.* $\tilde{\rho}_{\tilde{i}} + \tilde{\rho}_{\tilde{i}'} = \varsigma_{\tilde{i}\tilde{k}} = \varsigma_{\tilde{i}'\tilde{k}}$ *and* $\tilde{\rho}_{\tilde{k}} = \varsigma_{\tilde{k}\tilde{i}}$ *is optimal.*

If the treatments are grouped in two clusters, namely $\exists h \in \{1, \dots, K-1\}$ *s.t.* $(\theta_1, v_1) = (\theta_i, v_i)$ *for* $i = 1, \dots, h$ *and* $(\theta_K, v_K) = (\theta_i, v_i)$ *for* $i = h+1, \dots, K$, *with* $\theta_1 > \theta_K$, *then every design* $\tilde{\boldsymbol{\rho}}$ *s.t.* $\sum_{i=1}^{h} \tilde{\rho}_i = \varsigma_{1K}$ *and* $\sum_{i=h+1}^{K} \tilde{\rho}_i = \varsigma_{K1}$ *is optimal (namely, the Neyman allocation is spanned over the two clusters).*

Proof. See Section 3.6.3. □

Remark 3.2.2. *In general, the unconstrained optimal design $\tilde{\boldsymbol{\rho}}$ is a (generalized) Neyman allocation involving just two treatments (or mixture of Neyman targets in some special cases). However, adopting $\tilde{\boldsymbol{\rho}}$, the Wald test converges to a χ^2_1-distribution, so that power becomes $Pr\left(\chi^2_1(n\phi(\tilde{\boldsymbol{\rho}})) > q_{1,\alpha}\right)$. As correctly stated by Singh and Davidov[7], although $\tilde{\boldsymbol{\rho}}$ is a degenerate target, it is still the design maximizing the power, since $Pr\left(\chi^2_1(n\phi(\tilde{\boldsymbol{\rho}})) > q_{1,\alpha}\right) \geq Pr\left(\chi^2_b(n\phi(\tilde{\boldsymbol{\rho}})) > q_{b,\alpha}\right)$, for $b \geq 1$ (recalling that, for every fixed df b, χ^2-distributions are stochastically increasing in the NCP).*

As discussed in Remark 3.2.1, for exponential outcomes with censoring the treatment variance should be re-scaled. Thus, from Theorem 3.2.1, the optimal design maximizing the NCP is the generalized Neyman allocation $\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} = \boldsymbol{e}_{\tilde{i}}\varsigma_{\tilde{i}\tilde{k}} + \boldsymbol{e}_{\tilde{k}}\varsigma_{\tilde{k}\tilde{i}}$, where now $\varsigma_{\tilde{i}\tilde{k}} = \left(1 + \theta_{\tilde{k}}\sqrt{\epsilon(\theta_{\tilde{i}})}/\theta_{\tilde{i}}\sqrt{\epsilon(\theta_{\tilde{k}})}\right)^{-1}$, on the pair $\{\tilde{i}, \tilde{k}\}$ such that

$$\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \max_{i,k\in\{1,\ldots,K\}} \left(\frac{\theta_i - \theta_k}{\frac{\theta_i}{\sqrt{\epsilon(\theta_i)}} + \frac{\theta_k}{\sqrt{\epsilon(\theta_k)}}} \right)^2. \tag{3.2.4}$$

The maximization of the RHS in (3.2.4) depends on the specific form of the adopted censoring through $\epsilon(\cdot)$, and it could involve each pair of treatments, not necessarily $\{1, K\}$ (i.e., the one with the best and the worst treatments). Thus, this result conflicts with the optimal design obtained in Sverdlov et al[8] where, setting the minimum treatment allocation proportion equal to 0 leads to $\tilde{\boldsymbol{\rho}}_{1K}$, as the following example shows.

Example 3.2.2. *As in Sverdlov et al[8], we take into account the censoring scheme suggested in Zhang and Rosenberger[10] with duration $D = 96$ and recruitment period $R = 55$, for $K = 3$ treatments with $\boldsymbol{\theta}^\top = (150, 5, 1)$ we obtain $\epsilon(\theta_1) = 0.239$, $\epsilon(\theta_2) = 0.948$ and $\epsilon(\theta_3) = 0.990$. Thus, the maximum of the RHS in (3.2.4) is given by the pair $\{2, 3\}$ and therefore the optimal design is $\tilde{\boldsymbol{\rho}}_{23} = (0, 0.836, 0.164)^\top$ with $\phi(\tilde{\boldsymbol{\rho}}_{23}) = 0.424$, instead of $\tilde{\boldsymbol{\rho}}_{13} = (0.997, 0, 0.003)^\top$ for which $\phi(\tilde{\boldsymbol{\rho}}_{13}) = 0.234$.*

While the general cases of normal heteroscedastic outcomes and exponential responses with censoring should be analysed by Theorem 3.2.1 and Corollary 3.2.1, the next Corollary provides some useful simplifications for the most common models, where the Neyman target $\tilde{\boldsymbol{\rho}}_{1K}$ involving the best and the worst treatments is optimal.

Corollary 3.2.2. *Let $\theta_1 = \ldots = \theta_h \geq \theta_{h+1} \geq \ldots \geq \theta_{K-s} > \theta_{K-s+1} = \ldots = \theta_K$, where h, s are positive integers with $h + s \leq K$, then*

- *for binary trials, every $\tilde{\boldsymbol{\rho}}$ s.t.*

$$\sum_{i=1}^{h} \tilde{\rho}_i = \frac{\sqrt{\theta_1(1-\theta_1)}}{\sqrt{\theta_1(1-\theta_1)} + \sqrt{\theta_K(1-\theta_K)}} = \varsigma_{1K} = 1 - \sum_{i=K-s+1}^{K} \tilde{\rho}_i$$

is optimal with

$$\phi(\tilde{\boldsymbol{\rho}}) = \left(\frac{\theta_1 - \theta_K}{\sqrt{\theta_1(1-\theta_1)} + \sqrt{\theta_K(1-\theta_K)}} \right)^2;$$

- *under Poisson outcomes, every* $\tilde{\boldsymbol{\rho}}$ *s.t.* $\sum_{i=1}^{h}\tilde{\rho}_i = \sqrt{\theta_1}/(\sqrt{\theta_1}+\sqrt{\theta_K}) = \varsigma_{1K} = 1-\sum_{i=K-s+1}^{K}\tilde{\rho}_i$ *is optimal with* $\phi(\tilde{\boldsymbol{\rho}}) = \left\{(\theta_1-\theta_K)/(\sqrt{\theta_1}+\sqrt{\theta_K})\right\}^2$*;*
- *under exponential responses in the absence of censoring, every design* $\tilde{\boldsymbol{\rho}}$ *s.t.* $\sum_{i=1}^{h}\tilde{\rho}_i = \theta_1/(\theta_1+\theta_K) = \varsigma_{1K} = 1-\sum_{i=K-s+1}^{K}\tilde{\rho}_i$ *is optimal with* $\phi(\tilde{\boldsymbol{\rho}}) = \{(\theta_1-\theta_K)/(\theta_1+\theta_K)\}^2$*;*
- *under* $N(\theta_k; v)$ *for* $k = 1,\ldots,K$*, every* $\tilde{\boldsymbol{\rho}}$ *s.t.* $\sum_{i=1}^{h}\tilde{\rho}_i = 1/2 = \varsigma_{1K} = \sum_{i=K-s+1}^{K}\tilde{\rho}_i$ *is optimal with* $\phi(\tilde{\boldsymbol{\rho}}) = (\theta_1-\theta_K)^2/4v$*.*

Moreover, when $h = s = 1$ *(i.e., in the absence of clusters of better and worst treatments) the optimal design* $\tilde{\boldsymbol{\rho}} = \tilde{\boldsymbol{\rho}}_{1K} = \boldsymbol{e}_1\varsigma_{1K} + \boldsymbol{e}_K(1-\varsigma_{1K})$ *is unique.*

Proof. See Section 3.6.4. □

Remark 3.2.3. *Theorem 3.2.1 and Corollary 3.2.2 complement the results in Tymofyeyev et al*[9] *and Zhu and Hu*[11]*, by covering every possible scenario of clusters of treatments. Indeed, in both papers* h *and* s *are assumed to be positive integers with* $h+s<K$*, therefore the special case when all treatments are grouped into two clusters is excluded (e.g., for* $K=3$*,* $h=2$ *and* $s=1$ *or* $h=1$ *and* $s=2$*), also showing that every design spanning the Neyman target over the two clusters is optimal (instead of assuming as unique solution* $\tilde{\boldsymbol{\rho}} = h^{-1}\varsigma_{1K}\sum_{i=1}^{h}\boldsymbol{e}_i + s^{-1}\varsigma_{K1}\sum_{i=K-s+1}^{K}\boldsymbol{e}_i$ *as previously stated in the literature*[8,9,11]*).*

3.2.3 Constrained optimal designs for testing

This section deals with the problem of finding the design $\boldsymbol{\rho}^* = (\rho_1^*,\ldots,\rho_K^*)^\top$ maximizing the NCP of the multivariate Wald test of homogeneity under the (ethical) constraints $\rho_1^* \geq \ldots \geq \rho_K^*$, reflecting the treatment ordering $\theta_1 \geq \ldots \geq \theta_K$. Due to the complexity induced by this general framework, we need to introduce the following notation. Let

$$T = \frac{\left[\sum_{k=1}^{K} v_k^{-1}\right]\left[\sum_{k=1}^{K}(\theta_1-\theta_k)^2 v_k^{-1}\right]}{\left[\sum_{k=1}^{K}(\theta_1-\theta_k)v_k^{-1}\right]^2} \tag{3.2.5}$$

and, for $i = 1,\ldots,K-1$,

$$\sigma_i = \frac{\frac{1}{i}\sum_{k=1}^{i} v_k^{-1}}{\frac{1}{K}\sum_{k=1}^{K} v_k^{-1}},\ \beta_i = \frac{\frac{1}{i}\sum_{k=1}^{i}(\theta_1-\theta_k)v_k^{-1}}{\frac{1}{K}\sum_{k=1}^{K}(\theta_1-\theta_k)v_k^{-1}},\ \gamma_i = \frac{\frac{1}{i}\sum_{k=1}^{i}(\theta_1-\theta_k)^2 v_k^{-1}}{\frac{1}{K}\sum_{k=1}^{K}(\theta_1-\theta_k)^2 v_k^{-1}}$$

where, clearly, $\beta_1 = \gamma_1 = 0$.

Lemma 3.2.2. *For every* $i = 1,\ldots,K-1$*,* $\sigma_i \geq \beta_i \geq \gamma_i \geq 0$ *and, if* $\gamma_i = 1$*, then* $\beta_i > 1$*. Moreover,* $T > 1$*.*

Proof. See Section 3.6.5. □

Theorem 3.2.2. *If there exists* $\breve{k} \in \{1,\ldots,K-1\}$ *such that:*

P1a: $\beta_{\breve{k}}\left(2\sigma_{\breve{k}} - \beta_{\breve{k}} - \beta_{\breve{k}}\sigma_{\breve{k}}\right)/\sigma_{\breve{k}}^2 < T(1-\gamma_{\breve{k}}) \leq 1 + \sigma_{\breve{k}} - 2\beta_{\breve{k}}$*,*

P1b: $A_{\breve{k}}^2(1-\sigma_i) - 2A_{\breve{k}}(1-\beta_i) + T(1-\gamma_i) > 0$, *for every* $i \neq \breve{k}$,

where

$$A_{\breve{k}} = \begin{cases} \frac{T(1-\gamma_{\breve{k}})}{1-\beta_{\breve{k}}+\left[(1-\beta_{\breve{k}})^2-T(1-\sigma_{\breve{k}})(1-\gamma_{\breve{k}})\right]^{1/2}}, & \text{if } \gamma_{\breve{k}} \neq 1 \\ 2(\beta_{\breve{k}}-1)/(\sigma_{\breve{k}}-1), & \text{if } \gamma_{\breve{k}} = 1, \end{cases} \tag{3.2.6}$$

then

$$\boldsymbol{\rho}^* = \boldsymbol{\rho}^*_{\breve{k}} = \left(\frac{1-\tau(K-\breve{k})}{\breve{k}}\right) \sum_{i=1}^{\breve{k}} \boldsymbol{e}_i + \tau \sum_{i=\breve{k}+1}^{K} \boldsymbol{e}_i, \tag{3.2.7}$$

with

$$\tau = \frac{(\sigma_{\breve{k}} A_{\breve{k}} - \beta_{\breve{k}})}{\left\{K[1-\beta_{\breve{k}} - A_{\breve{k}}(1-\sigma_{\breve{k}})]\right\}}$$

and

$$\phi(\boldsymbol{\rho}^*_{\breve{k}}) = \left(\frac{A_{\breve{k}}^2 - 2A_{\breve{k}} + T}{TK}\right) \sum_{k=2}^{K} (\theta_1 - \theta_k)^2 v_k^{-1}. \tag{3.2.8}$$

If $\breve{k}$ *is not unique but there exists (one or more)* $\breve{i}$ *satisfying* $P1$ *(namely* $P1a$ *and* $P1b$*, where now* $P1b$ *should hold for every* $i \neq \{\breve{k}, \breve{i}\}$*), then* $A_{\breve{k}} = A_{\breve{i}}$ *and* $\boldsymbol{\rho}^*$ *can be obtained by any convex combination of the corresponding constrained optimal designs* $\boldsymbol{\rho}^*_{\breve{k}}$ *and* $\boldsymbol{\rho}^*_{\breve{i}}$ *in (3.2.7), so that NCP in (3.2.8) still holds.*

Whereas, if

P2: $T(1-\gamma_i) > 1 + \sigma_i - 2\beta_i$, *for* $i = 1, \ldots, K-1$,

the constrained optimal design is balanced, namely $\boldsymbol{\rho}^* = \boldsymbol{\rho}_B = K^{-1}\mathbf{1}_K$*, with*

$$\phi(\boldsymbol{\rho}_B) = \left(\frac{T-1}{TK}\right) \sum_{k=2}^{K} (\theta_1 - \theta_k)^2 v_k^{-1}. \tag{3.2.9}$$

In all the other scenarios $\boldsymbol{\rho}^{*t} = \left(\boldsymbol{\rho}^{*t}_{[K-\dot{c}]}, \mathbf{0}^\top_{\dot{c}}\right)$*, where* $\dot{c} \in \{1, \ldots, K-2\}$ *is the minimum number of inferior treatments that should be omitted in order to satisfy* $P1$ *or* $P2$ *and* $\boldsymbol{\rho}^*_{[K-\dot{c}]}$ *is the previously defined constrained optimal design evaluated by taking into account the remaining* $K - \dot{c}$ *(superior) treatments.*

Proof. See Section 3.6.6. □

Remark 3.2.4. *If* $\dot{c} = K-2$*, then* $P1b$ *vanishes, while* $P1a$ *and* $P2$ *identify two exhaustive and disjoint sets, namely* $T \leq 1 + \sigma_1$ *(i.e.,* $v_1 \geq v_2$*) and* $T > 1 + \sigma_1$*, respectively. In such a case,*

$$\boldsymbol{\rho}^{*t} = \begin{cases} \left(\varsigma_{12}; \varsigma_{21}, \mathbf{0}^\top_{K-2}\right), & \text{if } v_1 \geq v_2, \\ \left(2^{-1}\mathbf{1}^\top_2, \mathbf{0}^\top_{K-2}\right), & \text{if } v_1 < v_2 \end{cases}$$

and therefore the constrained optimal design is the Neyman or a balanced allocation involving just the two superior treatments.

Example 3.2.3. *Consider now the case of $K = 3$ treatments. When $\boldsymbol{\theta} = (23, 22.5, 22)^\top$ and $\boldsymbol{v} = (100, 10, 11)^\top$, condition P2 holds and therefore the constrained optimal design is the balanced one, i.e. $\boldsymbol{\rho}^* = 3^{-1}\mathbf{1}_3$, with $\phi(\boldsymbol{\rho}^*) = 0.0057$. Under the same setting, if $\boldsymbol{v} = (65, 10, 3.1)^\top$ then there exists a unique $\breve{k} \in \{1, 2\}$ satisfying $P1a$-$P1b$, namely $\breve{k} = 1$, so that $\boldsymbol{\rho}^* = \boldsymbol{\rho}_1^* = (0.508, 0.246, 0.246)^\top$ with $\phi(\boldsymbol{\rho}_1^*) = 0.0104$, whereas if $\boldsymbol{v} = (80, 10, 3.1)^\top$ then $\breve{k} = 2$ and therefore $\boldsymbol{\rho}^* = \boldsymbol{\rho}_2^* = (0.361, 0.361, 0.278)^\top$ with $\phi(\boldsymbol{\rho}_2^*) = 0.0096$. When $\boldsymbol{v} = (65.37, 10, 3.1)^\top$, then both $\breve{k} = 1$ and $\breve{k} = 2$ satisfy $P1a$-$P1b$ and $A_1 = A_2 = 0.965$; thus, every combination $\boldsymbol{\rho}^* = \omega\boldsymbol{\rho}_1^* + (1 - \omega)\boldsymbol{\rho}_2^*$ - where $\omega \in [0; 1]$, $\boldsymbol{\rho}_1^* = (0.504, 0.248, 0.248)^\top$ and $\boldsymbol{\rho}_2^* = (0.360, 0.360, 0.280)^\top$ - is optimal with $\phi(\boldsymbol{\rho}^*) = \phi(\boldsymbol{\rho}_1^*) = \phi(\boldsymbol{\rho}_2^*) = 0.0103$. If $\boldsymbol{v} = (5, 1, 65)^\top$ instead, neither $P1a$ nor $P2$ are satisfied and $\dot{c} = 1$; since $v_1 > v_2$, then $\boldsymbol{\rho}^{*t} = (\varsigma_{12}, \varsigma_{21}, 0) = (0.691, 0.309, 0)$ with $\phi(\boldsymbol{\rho}^*) = 0.0239$; analogously, $\dot{c} = 1$ also when $\boldsymbol{v} = (1, 5, 65)^\top$, but now $\boldsymbol{\rho}^* = (0.5, 0.5, 0)^\top$ with $\phi(\boldsymbol{\rho}^*) = 0.0208$, since $v_1 \leq v_2$.*

Despite the generality of Theorem 3.2.2, the next Corollary shows that in many practical situations the constrained optimal design $\boldsymbol{\rho}^*$ is a non-degenerate target having a simple functional form.

Corollary 3.2.3. *Assume that the variance $v(\cdot)$ is non-decreasing and*

$$\sigma_1(1 - \gamma_i) \geq \sigma_i + \gamma_i - 2\beta_i, \qquad \text{for } i = 2, \ldots, K - 1. \tag{3.2.10}$$

Thus, when $\theta_1 > \theta_2$, the constrained optimal design is

$$\boldsymbol{\rho}^* = \begin{cases} \boldsymbol{\rho}_1^* & \text{if } T \leq 1 + \sigma_1, \\ \boldsymbol{\rho}_B & \text{if } T > 1 + \sigma_1. \end{cases} \tag{3.2.11}$$

Whereas, in the presence of a cluster of superior treatments $\theta_1 = \cdots = \theta_j > \theta_{j+1} \geq \cdots \geq \theta_K$ ($j = 2, \ldots, K - 1$), then $\boldsymbol{\rho}^* = \boldsymbol{\rho}^B$ when $T > 1 + \sigma_1$, while for $T \leq 1 + \sigma_1$ every convex combination of $\boldsymbol{\rho}_1^*, \ldots, \boldsymbol{\rho}_j^*$ is optimal.

In particular, normal homoscedastic, Poisson and exponential models satisfy condition (3.2.10) and

$$\tau = \begin{cases} \frac{\sum_{k=2}^K (\theta_1 - \theta_k)^2}{2\left[\sum_{k=2}^K (\theta_1 - \theta_k)\right]^2}, & \text{for } N(\theta_k; v); \\ \frac{\theta_1 \left[\sum_{k=2}^K \left(\frac{1}{\theta_k} - \frac{1}{\theta_1}\right)\right]^{1/2} - \left[\sum_{k=2}^K (\theta_1 - \theta_k)\right]^{1/2}}{\left[\sum_{k=1}^K \frac{\theta_1}{\theta_k} - K\right]\left[\sum_{k=2}^K (\theta_1 - \theta_k)\right]^{1/2}}, & \text{for } P(\theta_k); \\ \frac{\frac{1}{\theta_1}\sum_{k=2}^K \left(\frac{1}{\theta_k} - \frac{1}{\theta_1}\right)^2}{\left[\sum_{k=2}^K \left(\frac{1}{\theta_k} - \frac{1}{\theta_1}\right)\right]\left[\sum_{k=2}^K \left(\frac{1}{\theta_k^2} - \frac{1}{\theta_1^2}\right)\right]}, & \text{for } exp(\theta_k). \end{cases}$$

Proof. See Section 3.6.7. □

Although the hypothesis of Corollary 3.2.3 do not hold for binary outcomes, the constrained optimal design as an analogous form, as the following proposition shows.

Proposition 3.2.1. *Under the binary model, when $\theta_1 > \theta_2$ the constrained optimal design is $\boldsymbol{\rho}^*$ in (3.2.11), where*

$$\tau = \frac{R^{-1}\left[\sum_{k=1}^K \frac{(\theta_1 - \theta_k)}{(1-\theta_k)\theta_k}\right] - 1}{\sum_{k=1}^K \frac{(1-\theta_1)\theta_1}{(1-\theta_k)\theta_k} - K} \text{ and } R = \sqrt{\left[\left(\sum_{k=1}^K \frac{\theta_1 - \theta_k}{(1-\theta_1)(1-\theta_k)}\right)\left(\sum_{k=1}^K \frac{\theta_1 - \theta_k}{\theta_1\theta_k}\right)\right]}.$$

Whereas, if $\theta_1 = \cdots = \theta_j$ *(*$j = 2, \ldots, K-1$*), then* $\boldsymbol{\rho}^* = \boldsymbol{\rho}_B$ *when* $T > 1+\sigma_1$*, while for* $T \leq 1+\sigma_1$ *every convex combination of* $\boldsymbol{\rho}_1^*, \ldots, \boldsymbol{\rho}_j^*$ *is optimal.*

Proof. See Section 3.6.8 □

3.3 Analytical and numerical comparisons

This section is dedicated to the performance assessment of the newly introduced optimal designs. Starting with the normal model, $\boldsymbol{\rho}^*$ and $\tilde{\boldsymbol{\rho}}$ will be compared with the balanced allocation $\boldsymbol{\rho}_B$ and the design $\boldsymbol{\rho}_M = (\boldsymbol{e}_1 + \boldsymbol{e}_K)/2$ proposed by Baldi Antognini et al[2] and Singh and Davidov[7], which is the optimal design for normal homoscedastic data (i.e., $\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}_M$) and it is also the target maximizing the minimum power for both restricted and unrestricted likelihood ratio tests under the simple order restriction $\theta_1 \geq \ldots \geq \theta_K$. Whereas for binary and exponential responses, we compare our proposals with $\boldsymbol{\rho}_B$ and the design $\boldsymbol{\rho}_H$ proposed by Tymofyeyev et al[9] and Zhu and Hu[11], that maximizes the power under the constraint of a minimum prefixed threshold of allocation to each treatment.

In what follows, both inferential and ethical criteria will be assessed. For every design $\boldsymbol{\rho}$, besides the statistical power $\mathcal{P}_n(\boldsymbol{\rho}) = Pr\left(\chi_b^2(n\phi(\boldsymbol{\rho})) > q_{b,0.05}\right)$ (where b is the appropriate df), a first approximated measure of efficiency is simply provided by $\Lambda(\boldsymbol{\rho}) = \phi(\boldsymbol{\rho})/\phi(\tilde{\boldsymbol{\rho}})$, namely the ratio between the NCP induced by $\boldsymbol{\rho}$ over its optimal value (i.e., the one corresponding to the unconstrained optimal design $\tilde{\boldsymbol{\rho}}$). Taking into account $\boldsymbol{\rho}_M$, from (3.2.1) and (3.2.2) it follows that

$$\Lambda(\boldsymbol{\rho}_M) = \frac{\frac{(\theta_1-\theta_K)^2}{2(v_1+v_K)}}{\max\limits_{i,k\in\{1,\ldots,K\}}\left(\frac{\theta_i-\theta_k}{\sqrt{v_i}+\sqrt{v_k}}\right)^2}. \tag{3.3.1}$$

From Theorem 3.2.1, when $\{1; K\}$ is the pair of treatments maximizing the side of (3.2.2), then $\Lambda(\boldsymbol{\rho}_M) = \left\{2[\varsigma_{1K}^2 + (1-\varsigma_{1K})^2]\right\}^{-1} \geq 1/2$, guaranteeing good performance even in the presence of strong heteroscedasticity between the two extreme treatment groups (namely when $\varsigma_{1K} \to \{0; 1\}$). Whereas, if $\{\tilde{i}, \tilde{k}\} \neq \{1; K\}$, then $\Lambda(\boldsymbol{\rho}_M)$ tends to vanish; indeed, assuming for example the normal heteroscedastic model with $v_1 = v_K$ and $v_{\tilde{i}} = v_{\tilde{k}}$ (where, clearly, $v_1 > v_{\tilde{i}}$), then $\Lambda(\boldsymbol{\rho}_M) \to 0$ as $v_{\tilde{i}}/v_1 \to 0$ (e.g., by letting $\boldsymbol{\theta} = (15, 14, 13, 10, 9)^\top$ and $\boldsymbol{v} = (40, 1, 35, 1, 40)^\top$, then $\Lambda(\boldsymbol{\rho}_M) = 0.056$). Adopting $\boldsymbol{\rho}_B$ instead, from (3.2.1) and (3.2.9)

$$\Lambda(\boldsymbol{\rho}_B) = \frac{\left(\frac{T-1}{T}\right)\frac{1}{K}\sum_{k=1}^K \frac{(\theta_1-\theta_k)^2}{v_k}}{\max\limits_{i,k\in\{1,\ldots,K\}}\left(\frac{\theta_i-\theta_k}{\sqrt{v_i}+\sqrt{v_k}}\right)^2}. \tag{3.3.2}$$

For normal homoscedastic data, $\Lambda(\boldsymbol{\rho}_B) = 4\,(\theta_1-\theta_K)^{-2}K^{-1}\sum_{i=1}^K(\theta_i-\bar{\theta})^2$, where $\bar{\theta} = K^{-1}\sum_{k=1}^K\theta_k$. By applying the Von Szokefalvi-Nagy inequality, $\Lambda(\boldsymbol{\rho}_B) \geq 2K^{-1}$, namely $\boldsymbol{\rho}_B$ tends to exhibit poor performances as the number of treatment groups grows. In this regard, Figure 3.1 displays the behaviour of $\Lambda(\boldsymbol{\rho}_B)$ for different models with $K = 3$ and 5 treatments as θ_2 varies. In particular, for the normal homoscedastic (with $v = 1$), exponential and Poisson models we consider $\boldsymbol{\theta}^\top = (11, \theta_2, 1)$ for $K = 3$ and $\boldsymbol{\theta}^\top = (11, \theta_2, 2, 1.5, 1)$ for $K = 5$, where in both scenarios θ_2 varies between 2 and 10; whereas for the binary model we set $\boldsymbol{\theta}^\top = (0.9, \theta_2, 0.1)$ for $K = 3$ and $\boldsymbol{\theta}^\top = (0.9, \theta_2, 0.2, 0.15, 0.1)$ for $K = 5$,

with θ_2 varying between 0.2 and 0.8. For all the considered models, the performance of $\boldsymbol{\rho}_B$ considerably deteriorates for $K = 5$, especially for binary, exponential and Poisson responses, whose efficiencies are always lower than 61%. For normal and binary outcomes, $\Lambda(\boldsymbol{\rho}_B)$ tents to grow for values of θ_2 close to θ_3, with maximum efficiency around 80%. As regards Poisson and exponential models the performances improve as $\theta_2 \to \theta_1$, whereas for θ_2 close to θ_3 the efficiency becomes even lower than 40%.

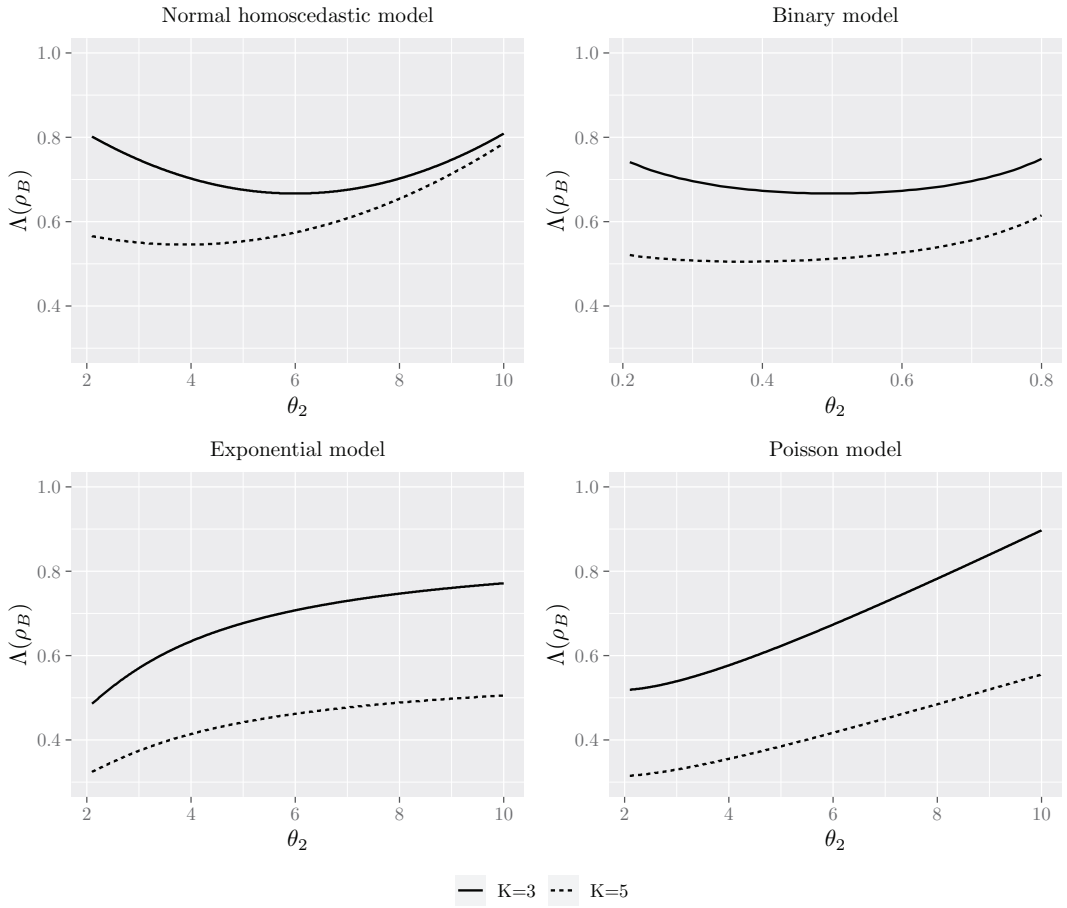


Figure 3.1: Behaviour of $\Lambda(\boldsymbol{\rho}_B)$ for $K = 3$ and 5 treatments as θ_2 varies. For the normal homoscedastic, exponential and Poisson models $\boldsymbol{\theta}^\top = (11, \theta_2, 1)$ and $\boldsymbol{\theta}^\top = (11, \theta_2, 2, 1.5, 1)$, while for binary data $\boldsymbol{\theta}^\top = (0.9, \theta_2, 0.1)$ and $\boldsymbol{\theta}^\top = (0.9, \theta_2, 0.2, 0.15, 0.1)$.

As regards ethics, in the multi-treatment context several ethical measures could be adopted, some of them are only model-specific. In our general set-up, as a measure of ethics we take into account the total expected outcome $\mathcal{E}_n(\boldsymbol{\rho}) = n\boldsymbol{\theta}^\top\boldsymbol{\rho}$; the corresponding ethical efficiency $(\boldsymbol{\theta}^\top\boldsymbol{\rho} - \theta_K)/(\theta_1 - \theta_K) \in [0; 1]$ will be provided within brackets. Since $\mathcal{E}_{2n}(\boldsymbol{\rho}) = 2\mathcal{E}_n(\boldsymbol{\rho})$, in the following tables the ethical criterion will be provided only for $n = 100$.

As previously showed, the unconstrained optimal design $\tilde{\boldsymbol{\rho}}$ maximizes the NCP and the power as well; therefore, it does not exists a target with better performances in terms of both ethics and power with respect to $\tilde{\boldsymbol{\rho}}$, simultaneously. Whereas, in some circumstances, $\tilde{\boldsymbol{\rho}}$ dominates $\boldsymbol{\rho}_B$ and $\boldsymbol{\rho}_M$, as discussed in

the following proposition.

Proposition 3.3.1. *Let $\{\tilde{i}, \tilde{k}\}$ be the pair of treatments maximizing the RHS of (3.2.2) with $\tilde{i} < \tilde{k}$. If $\varsigma_{\tilde{i}\tilde{k}} \geq (\bar{\theta} - \theta_{\tilde{k}})/(\theta_{\tilde{i}} - \theta_{\tilde{k}})$ then $\tilde{\boldsymbol{\rho}}$ dominates $\boldsymbol{\rho}_B$, namely $\mathcal{E}_n(\tilde{\boldsymbol{\rho}}) \geq \mathcal{E}_n(\boldsymbol{\rho}_B)$ and $\mathcal{P}_n(\tilde{\boldsymbol{\rho}}) \geq \mathcal{P}_n(\boldsymbol{\rho}_B)$, simultaneously, for every sample size n. Analogously, if $\varsigma_{\tilde{i}\tilde{k}} \geq (\theta_1 + \theta_K - 2\theta_{\tilde{k}})/[2(\theta_{\tilde{i}} - \theta_{\tilde{k}})]$, then $\tilde{\boldsymbol{\rho}}$ dominates $\boldsymbol{\rho}_M$. Moreover, if the variance $v(\cdot)$ of a given statistical model is non-decreasing and $\theta_{\tilde{i}} + \theta_{\tilde{k}} \geq \max\{2\bar{\theta}; \theta_1 + \theta_K\}$, then $\tilde{\boldsymbol{\rho}}$ dominates both $\boldsymbol{\rho}_M$ and $\boldsymbol{\rho}_B$.*

Proof. The proof follows easily from (3.2.1) after some algebra, by observing that $\mathcal{E}_n(\boldsymbol{\rho}_B) = n\bar{\theta}$, $\mathcal{E}_n(\boldsymbol{\rho}_M) = n(\theta_1 + \theta_K)/2$, while $\mathcal{E}_n(\tilde{\boldsymbol{\rho}}) = n\left[\theta_{\tilde{i}}\varsigma_{\tilde{i}\tilde{k}} + \theta_{\tilde{k}}(1 - \varsigma_{\tilde{i}\tilde{k}})\right]$. □

Starting from the case of normal responses, Tables 3.1, 3.2 and 3.3 summarize the performances in terms of power $\mathcal{P}_n$ with $n = 50$ and 100 (within brackets the corresponding efficiency evaluated with respect to $\tilde{\boldsymbol{\rho}}$), and ethical criterion $\mathcal{E}_n$ (with $n = 100$) of the considered allocations for $K = 3$, 4 and 5 treatments, as $\boldsymbol{\theta}$ and $\boldsymbol{v}$ vary. Whereas, Tables 3.4 and 3.5 show the results in the case of binary and exponential outcomes, respectively, where the minimum proportion of subjects assigned to each treatment group for $\boldsymbol{\rho}_H$ is set to 0.2 ($K = 3, 4$) and 0.15 ($K = 5$).

Let us first consider the results for normal response trials in Tables 3.1, 3.2 and 3.3. Clearly, the unconstrained optimal design $\tilde{\boldsymbol{\rho}}$ exhibits the highest power, with an ethical efficiency varying between 42.9% and 72.2%. Whereas, the constrained optimal target $\boldsymbol{\rho}^*$ substantially shows the highest ethical efficiency (between 51.8% and 70.2%) also guaranteeing valid performances in terms of power (for $n = 100$, its efficiency is always greater than 65.6%). Excluding the homoscedastic scenario, the power of $\boldsymbol{\rho}_M$ is strictly related to the treatment variances: for unordered variances $\boldsymbol{\rho}_M$ exhibits poor performances (showing in some cases an extremely low efficiency, equal to 39.3% for $K = 4$ and $n = 50$), while when the variances are ordered as the treatment effects its power tends to increase. Moreover, the ethical efficiency of $\boldsymbol{\rho}_M$ is always equal to 0.5 and in several scenarios $\boldsymbol{\rho}_M$ is dominated by $\boldsymbol{\rho}^*$, especially when the number of the treatment groups increases. The balanced design $\boldsymbol{\rho}_B$ shows the worst performances in terms of both power and ethical gain and it is always dominated by $\boldsymbol{\rho}^*$. The power provided by the unconstrained optimal design $\tilde{\boldsymbol{\rho}}$ with $n = 50$ observations tends to be quite similar to the one of $\boldsymbol{\rho}_B$ with $n = 100$ subjects.

As regards binary trials in Table 3.4, the constrained optimal design $\boldsymbol{\rho}^*$ and $\boldsymbol{\rho}_H$ tend to perform quite similarly in terms of power, with an efficiency always higher than 68%, whereas $\boldsymbol{\rho}_B$ shows the lowest statistical power with a maximum loss up to 44%. Taking into account the ethical criterion, that in this case corresponds to the total expected successes, $\boldsymbol{\rho}^*$ and $\tilde{\boldsymbol{\rho}}$ guarantee the highest ethical efficiency (with only one exception, where $\mathcal{E}_n(\boldsymbol{\rho}_B)$ is slightly bigger with respect to $\mathcal{E}_n(\tilde{\boldsymbol{\rho}})$), with an ethical gain up to 8 and 12 successes with respect to $\boldsymbol{\rho}_H$ and $\boldsymbol{\rho}_B$, respectively. Similar considerations still hold for exponential responses reported in Table 3.5, where the minimum power efficiency becomes 73% for $\boldsymbol{\rho}^*$, 67% for $\boldsymbol{\rho}_H$ and 58% for the balanced allocation. The ethical gain (i.e., the additional total expected survival time in this context) induced by $\boldsymbol{\rho}^*$ with respect to $\boldsymbol{\rho}_H$ ranges from 36 to 90 (corresponding to a gain in efficiency between $5\% - 18\%$) and it is even more evident for $\boldsymbol{\rho}_B$ (the additional expected survival is up to 250, with an ethical gain up to 33% in terms of efficiency). The unconstrained optimal design $\tilde{\boldsymbol{\rho}}$ exhibits the highest power and, at the same time, the greatest ethical gain, and therefore it dominates all the other designs. Similarly to the case of normal responses, the power induced by $\tilde{\boldsymbol{\rho}}$ with $n = 50$ observations tends to be quite similar to those of $\boldsymbol{\rho}_B$ and $\boldsymbol{\rho}_H$ with $n = 100$ subjects. In general, for all the considered models, both $\tilde{\boldsymbol{\rho}}$ and $\boldsymbol{\rho}^*$ present high values of ethical efficiency. Moreover, for normal homoscedastic, binary and exponential responses, it is evident that $\boldsymbol{\rho}_B$ is dominated by $\boldsymbol{\rho}^*$

and also by $\tilde{\boldsymbol{\rho}}$ (with the exception of the second last scenario of Table 3.4). As further comparisons (omitted here for brevity) showed, the results for Poisson outcomes are substantially the same of those of the exponential model. The case of exponential responses with censoring tends to be similar to that of Normal heteroscedastic data with variances ordered as the treatment effects, and it is strongly affected by the chosen censoring scheme.

Table 3.1: Power $\mathcal{P}_n$ and total expected outcomes $\mathcal{E}_n$ (efficiencies within brackets) for normal responses with $K = 3$ treatments, as $\boldsymbol{\theta}$ and $\boldsymbol{v}$ vary.

$\boldsymbol{\theta} = (1.5, 1.1, 1)^\top$	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
$\boldsymbol{v} = (1, 1, 1)^\top$	$\boldsymbol{\rho}^* = (.494, .253, .253)^\top$	.283 (.667)	.519 (.736)	127 (.544)
	$\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}_M$	.424 (1)	.705 (1)	125 (.500)
	$\boldsymbol{\rho}_B$	.257 (.606)	.475 (.674)	120 (.400)
$\boldsymbol{v} = (1, 2, 6)^\top$	$\boldsymbol{\rho}^* = (.5, .5, 0)^\top$	.211 (.977)	.372 (.976)	130 (.600)
	$\tilde{\boldsymbol{\rho}} = (.414, .586, 0)^\top$	.216 (1)	.381 (1)	127 (.531)
	$\boldsymbol{\rho}_M$	.157 (.727)	.267 (.701)	125 (.500)
	$\boldsymbol{\rho}_B$	.153 (.708)	.269 (.706)	120 (.400)
$\boldsymbol{v} = (6, 2, 1)^\top$	$\boldsymbol{\rho}^* = (.668, .166, .166)^\top$	.121 (.688)	.200 (.656)	135 (.702)
	$\tilde{\boldsymbol{\rho}} = (.71, 0, .29)^\top$	.176 (1)	.305 (1)	136 (.710)
	$\boldsymbol{\rho}_M$	.157 (.892)	.267 (.875)	125 (.500)
	$\boldsymbol{\rho}_B$	.098 (.557)	.151 (.495)	120 (.400)
$\boldsymbol{v} = (2, 1, 6)^\top$	$\boldsymbol{\rho}^* = \tilde{\boldsymbol{\rho}} = (.586, .414, 0)^\top$	.216 (1)	.381 (1)	133 (.669)
	$\boldsymbol{\rho}_M$	.143 (.662)	.240 (.630)	125 (.500)
	$\boldsymbol{\rho}_B$	.135 (.625)	.230 (.604)	120 (.400)

Table 3.2: Power $\mathcal{P}_n$ and total expected outcomes $\mathcal{E}_n$ (efficiencies within brackets) for normal responses with $K = 4$ treatments, as $\boldsymbol{\theta}$ and $\boldsymbol{v}$ vary.

$\boldsymbol{\theta} = (2, 1.8, 1.1, 1)^\top$	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
$\boldsymbol{v} = (1, 1, 1, 1)^\top$	$\boldsymbol{\rho}^* = (.37, .21, .21, .21)^\top$	.747 (.793)	.971 (.972)	156 (.560)
	$\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}_M$	.942 (1)	.999 (1)	150 (.500)
	$\boldsymbol{\rho}_B$	.729 (.774)	.965 (.966)	148 (.475)
$\boldsymbol{v} = (1, 1.5, 2, 7)^\top$	$\boldsymbol{\rho}^* = (1/3, 1/3, 1/3, 0)^\top$	.467 (.622)	.778 (.810)	163 (.633)
	$\tilde{\boldsymbol{\rho}} = (.414, 0, .586, 0)^\top$	.751 (1)	.961 (1)	147 (.473)
	$\boldsymbol{\rho}_M$	.424 (.565)	.705 (.734)	150 (.500)
	$\boldsymbol{\rho}_B$	.386 (.514)	.692 (.720)	148 (.475)
$\boldsymbol{v} = (7, 2, 1.5, 1)^\top$	$\boldsymbol{\rho}^* = (.309, .309, .191, .191)^\top$	.364 (.561)	.662 (.726)	158 (.575)
	$\tilde{\boldsymbol{\rho}} = (0, .586, 0, .414)^\top$	.649 (1)	.912 (1)	147 (.469)
	$\boldsymbol{\rho}_M$	.424 (.653)	.705 (.773)	150 (.500)
	$\boldsymbol{\rho}_B$	.347 (.535)	.637 (.698)	148 (.475)
$\boldsymbol{v} = (12, 1.5, 9, 1)^\top$	$\boldsymbol{\rho}^* = (.275, .275, .225, .225)^\top$	.340 (.472)	.627 (.661)	152 (.518)
	$\tilde{\boldsymbol{\rho}} = (0, .55, 0, .45)^\top$	.720 (1)	.949 (1)	144 (.440)
	$\boldsymbol{\rho}_M$	.284 (.393)	.501 (.528)	150 (.500)
	$\boldsymbol{\rho}_B$	.337 (.468)	.622 (.655)	148 (.475)

Table 3.3: Power $\mathcal{P}_n$ and total expected outcomes $\mathcal{E}_n$ (efficiencies within brackets) for normal responses with $K = 5$ treatments, as $\boldsymbol{\theta}$ and $\boldsymbol{v}$ vary.

$\boldsymbol{\theta} = (3, 2.7, 2, 1.2, 1)^\top$	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
$\boldsymbol{v} = (1, 1, 1, 1, 1)^\top$	$\boldsymbol{\rho}^* = (.36, .16, .16, .16, .16)^\top$	.999 (.999)	1 (1)	218 (.592)
	$\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}_M$	1 (1)	1 (1)	200 (.500)
	$\boldsymbol{\rho}_B$	.998 (.998)	1 (1)	198 (.490)
$\boldsymbol{v} = (1, 1.5, 2, 3, 15)^\top$	$\boldsymbol{\rho}^* = (.277, .241, .241, .241, 0)^\top$	.843 (.846)	.992 (.992)	225 (.626)
	$\tilde{\boldsymbol{\rho}} = (.366, 0, 0, .634, 0)^\top$	.997 (1)	1 (1)	186 (.429)
	$\boldsymbol{\rho}_M$	.705 (.707)	.942 (.942)	200 (.500)
	$\boldsymbol{\rho}_B$	.765 (.767)	.978 (.978)	198 (.490)
$\boldsymbol{v} = (12, 3, 2, 1.5, 1)^\top$	$\boldsymbol{\rho}^* = (.287, .287, .142, .142, .142)^\top$	.794 (.800)	.985 (.985)	223 (.616)
	$\tilde{\boldsymbol{\rho}} = (0, .634, 0, 0, .366)^\top$	.993 (1)	1 (1)	208 (.539)
	$\boldsymbol{\rho}_M$	.792 (.798)	.975 (.975)	200 (.500)
	$\boldsymbol{\rho}_B$	.762 (.767)	.978 (.978)	198 (.490)
$\boldsymbol{v} = (5, 3, 10, 1, 15)^\top$	$\boldsymbol{\rho}^* = (.4, .2, .2, .2, 0)^\top$	.836 (.857)	.991 (.991)	238 (.690)
	$\tilde{\boldsymbol{\rho}} = (.691, 0, 0, .309, 0)^\top$	.976 (1)	1 (1)	244 (.722)
	$\boldsymbol{\rho}_M$	.609 (.624)	.885 (.885)	200 (.500)
	$\boldsymbol{\rho}_B$	.696 (.713)	.957 (.957)	198 (.490)

Table 3.4: Power $\mathcal{P}_n$ and total expected outcomes $\mathcal{E}_n$ (efficiencies within brackets) for binary responses with $K = 3$, 4 and 5 treatments, as $\boldsymbol{\theta}$ varies.

$\boldsymbol{\theta}$	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
$(.4, .1, .05)^\top$	$\boldsymbol{\rho}^* = (.658, .171, .171)^\top$	.827 (.882)	.987 (.988)	29 (.682)
	$\tilde{\boldsymbol{\rho}} = (.692, 0, .308)^\top$	.938 (1)	.999 (1)	29 (.692)
	$\boldsymbol{\rho}_H = (.593, .2, .207)^\top$	.821 (.875)	.986 (.987)	27 (.624)
	$\boldsymbol{\rho}_B$	.663 (.707)	.932 (.933)	18 (.381)
$(.6, .4, .25)^\top$	$\boldsymbol{\rho}^* = (.480, .260, .260)^\top$	.516 (.675)	.827 (.855)	46 (.591)
	$\tilde{\boldsymbol{\rho}} = (.531, 0, .469)^\top$	.765 (1)	.967 (1)	44 (.531)
	$\boldsymbol{\rho}_H = (.432, .2, .368)^\top$	.566 (.74)	.869 (.899)	43 (.518)
	$\boldsymbol{\rho}_B$	.485 (.634)	.796 (.823)	42 (.476)
$(.4, .3, .1, .05)^\top$	$\boldsymbol{\rho}^* = (.562, .146, .146, .146)^\top$	.725 (.773)	.964 (.965)	29 (.688)
	$\tilde{\boldsymbol{\rho}} = (.692, 0, 0, .308)^\top$	.938 (1)	.999 (1)	29 (.692)
	$\boldsymbol{\rho}_H = (.4, .2, .2, .2)^\top$	.693 (.739)	.952 (.953)	25 (.571)
	$\boldsymbol{\rho}_B$	.611 (.651)	.910 (.911)	21 (.464)
$(.5, .2, .15, .1)^\top$	$\boldsymbol{\rho}^* = (.583, .139, .139, .139)^\top$	.729 (.774)	.965 (.966)	35 (.635)
	$\tilde{\boldsymbol{\rho}} = (.625, 0, 0, .375)^\top$	.942 (1)	.999 (1)	35 (.625)
	$\boldsymbol{\rho}_H = (.4, .2, .2, .2)^\top$	.670 (.711)	.942 (.943)	29 (.475)
	$\boldsymbol{\rho}_B$	.525 (.557)	.846 (.847)	24 (.344)
$(.8, .7, .6, .5, .1)^\top$	$\boldsymbol{\rho}^* = (.316, .171, .171, .171, .171)^\top$	.992 (.992)	1 (1)	58 (.683)
	$\tilde{\boldsymbol{\rho}} = (.571, 0, 0, 0, .429)^\top$	1 (1)	1 (1)	50 (.571)
	$\boldsymbol{\rho}_H = (.246, .15, .15, .15, .304)^\top$	.998 (.998)	1 (1)	50 (.567)
	$\boldsymbol{\rho}_B$	.990 (.990)	1 (1)	54 (.629)
$(.55, .4, .3, .1, .05)^\top$	$\boldsymbol{\rho}^* = (.544, .114, .114, .114, .114)^\top$	.924 (.925)	.999 (.999)	40 (.692)
	$\tilde{\boldsymbol{\rho}} = (.695, 0, 0, 0, .305)^\top$	.999 (1)	1 (1)	40 (.695)
	$\boldsymbol{\rho}_H = (.378, .15, .15, .15, .172)^\top$	.909 (.910)	.998 (.998)	34 (.573)
	$\boldsymbol{\rho}_B$	.817 (.818)	.989 (.989)	28 (.460)

Table 3.5: Power $\mathcal{P}_n$ and total expected outcomes $\mathcal{E}_n$ (efficiencies within brackets) for exponential responses with $K = 3$, 4 and 5 treatments, as $\boldsymbol{\theta}$ varies.

$\boldsymbol{\theta}$	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
$(4,2,1)^\top$	$\boldsymbol{\rho}^* = (.722,.139,.139)^\top$	.950 (.961)	.999 (.999)	331 (.769)
	$\tilde{\boldsymbol{\rho}} = (.8,0,.2)^\top$	.989 (1)	1 (1)	340 (.800)
	$\boldsymbol{\rho}_H = (.6,.2,.2)^\top$	.941 (.951)	.999 (.999)	300 (.667)
	$\boldsymbol{\rho}_B$	.856 (.866)	.992 (.992)	233 (.444)
$(10,7,3)^\top$	$\boldsymbol{\rho}^* = (.634,.183,.183)^\top$	.902 (.932)	.997 (.997)	817 (.739)
	$\tilde{\boldsymbol{\rho}} = (.769,0,.231)^\top$	.968 (1)	1 (1)	838 (.769)
	$\boldsymbol{\rho}_H = (.574,.2,.226)^\top$	.902 (.932)	.997 (.997)	782 (.688)
	$\boldsymbol{\rho}_B$	.849 (.877)	.991 (.991)	667 (.524)
$(11,9,5,3)^\top$	$\boldsymbol{\rho}^* = (.625,.125,.125,.125)^\top$	.875 (.892)	.995 (.995)	899(.749)
	$\tilde{\boldsymbol{\rho}} = (.786,0,0,.214)^\top$	.981 (1)	1 (1)	929 (.786)
	$\boldsymbol{\rho}_H = (.4,.2,.2,.2)^\top$	.843 (.859)	.992 (.992)	780 (.600)
	$\boldsymbol{\rho}_B$	.789 (.804)	.982 (.982)	700 (.500)
$(14,10,7,5)^\top$	$\boldsymbol{\rho}^* = (.619,.127,.127,.127)^\top$	.672 (.732)	.943 (.946)	1145 (.717)
	$\tilde{\boldsymbol{\rho}} = (.737,0,0,.263)^\top$	.918 (1)	.997 (1)	1163 (.737)
	$\boldsymbol{\rho}_H = (.4,.2,.2,.2)^\top$	.619 (.674)	.915 (.918)	1000 (.556)
	$\boldsymbol{\rho}_B$	.534 (.582)	.854 (.857)	900 (.444)
$(7,5,4,3,2)^\top$	$\boldsymbol{\rho}^* = (.624,.094,.094,.094,.094)^\top$	.759 (.778)	.977 (.977)	569 (.737)
	$\tilde{\boldsymbol{\rho}} = (.778,0,0,0,.222)^\top$	.975 (1)	1 (1)	589 (.778)
	$\boldsymbol{\rho}_H = (.378,.15,.15,.15,.172)^\top$	.713 (.731)	.963 (.963)	479 (.558)
	$\boldsymbol{\rho}_B$	.603 (.618)	.911 (.911)	420 (.440)
$(14,13,10,5,4)^\top$	$\boldsymbol{\rho}^* = (.58,.105,.105,.105,.105)^\top$	.825 (.846)	.990 (.990)	1148 (747)
	$\tilde{\boldsymbol{\rho}} = (.778,0,0,0,.222)^\top$	.975 (1)	1 (1)	1178 (.778)
	$\boldsymbol{\rho}_H = (.4,.15,.15,.15,.15)^\top$	.804 (.825)	.987 (.987)	1040 (.640)
	$\boldsymbol{\rho}_B$	.741 (.760)	.972 (.972)	920 (.520)

3.4 Implementation via response-adaptive randomization and discussion

The unconstrained optimal design $\tilde{\boldsymbol{\rho}}$ in Theorem 3.2.1 and the constrained one $\boldsymbol{\rho}^*$ in Theorem 3.2.2 depend on the unknown model parameters and therefore they are a-priori unknown (i.e., *locally optimal*). The dependence on the model parameters acts in terms of both i) the a-priori unknown treatment ordering and ii) the functional form of the optimal design itself, which is often a degenerate allocation with no assignments to some treatment groups.

Instead of other alternative approaches suggested in the literature, which are intended to mediate the design criterion onto the entire parameter space to obtain good overall performances (such as, e.g.,

Bayesian or maxi-min approaches), in this paper we consider response-adaptive randomization as the natural solution to this local optimality problem. Under this framework, the exact optimal designs are sequentially estimated step-by-step: on the basis of earlier responses and past assignments, the unknown parameters are estimated along with the treatment ordering (which could change as the trial progresses) and thus, the next assignment is randomly forced to progressively approach the optimal target (for instance, by applying the Doubly Adaptive Biased Coin Design of Hu and Zhang[5]).

Even if response-adaptive randomization methodology seems a natural choice in order to implement the optimal proposed allocations, we wish to stress that the optimal designs $\tilde{\boldsymbol{\rho}}$ and $\boldsymbol{\rho}^*$ cannot be targeted directly, due to the fact that i) their functional forms are locally discontinuous around the subset of Θ^K under which the treatment ordering changes (namely, where θ_i tends to coincide with one or more θ_ks) and ii) in several scenarios these optimal designs lay on the boundary. To overcome these drawbacks, that prevent the applicability of standard response-adaptive randomization methodology, a smoothing transformation (e.g., via a Gaussian kernel) can be applied to obtain a continuous and non-degenerate version of these targets (see, e.g., Tymofyeyev et al[9]). In particular, for the mono-parametric exponential family we take into account the convolution of $\tilde{\boldsymbol{\rho}} = \tilde{\boldsymbol{\rho}}(\boldsymbol{\theta}) = (\tilde{\rho}_1(\boldsymbol{\theta}), \ldots, \tilde{\rho}_K(\boldsymbol{\theta}))^\top$ (or, analogously, $\boldsymbol{\rho}^*$), with a K-dim Gaussian kernel

$$G(\boldsymbol{\theta}) = (2\pi\sigma^2)^{-K/2} \exp\left(-\frac{\theta_1^2 + \cdots + \theta_K^2}{2\sigma^2}\right),$$

($\sigma^2 > 0$ controls the degree of smoothing), namely we define the smoothed version

$$\tilde{\boldsymbol{\rho}}^S(\boldsymbol{\theta}) = (\tilde{\rho}_1^S(\boldsymbol{\theta}), \ldots, \tilde{\rho}_K^S(\boldsymbol{\theta}))^\top$$

of $\tilde{\boldsymbol{\rho}}$, by letting

$$\tilde{\rho}_k^S(\boldsymbol{\theta}) = (\tilde{\rho}_k * G)(\boldsymbol{\theta}) = \int_{\Theta^K} \tilde{\rho}_k(\boldsymbol{x}) G(\boldsymbol{\theta} - \boldsymbol{x}) d\boldsymbol{x}, \quad k = 1, \ldots, K \tag{3.4.1}$$

(which could be naturally extended to the case of heteroscedastic normal model where $\tilde{\boldsymbol{\rho}} = \tilde{\boldsymbol{\rho}}(\boldsymbol{\theta}; \boldsymbol{v}) : \Theta^K \times \mathbb{R}^{+K} \to [0; 1]$).

This smoothing transformation essentially impacts on the points of discontinuity of $\tilde{\boldsymbol{\rho}}$ and on its boundary, so that the smoothed optimal design $\tilde{\boldsymbol{\rho}}^S$ obey the classical regularity conditions of continuity, non-degeneracy and differentiability (see, e.g., Baldi Antognini and Giovagnoli[1] or Hu and Rosenberger[4]), that allows for the standard asymptotic inference for response-adaptive randomization procedures (therefore, all the asymptotics in Section 3.2.1 are still valid). For instance, taking into account binary trials with $K = 3$ treatments, Figure 3.2 illustrates the behaviour of the first component of $\tilde{\boldsymbol{\rho}}$ and the one of its smoothed version $\tilde{\boldsymbol{\rho}}^S$, with $\boldsymbol{\theta} = (\theta_1, \theta_2, 0.15)^\top$ as θ_1 and θ_2 vary in $[0; 1]$ (where now the treatment ordering is free to change on the basis of the values of θ_1 and θ_2).

In order to show how the performances of $\tilde{\boldsymbol{\rho}}$ and $\tilde{\boldsymbol{\rho}}^S$ (as well as those of $\boldsymbol{\rho}^*$ and $\boldsymbol{\rho}^{*S}$) are quite similar, Table 3.6 presents the statistical power $\mathcal{P}_n$ and total expected outcomes $\mathcal{E}_n$ under the same scenarios of Table 3.4 with $K = 3$ treatments.

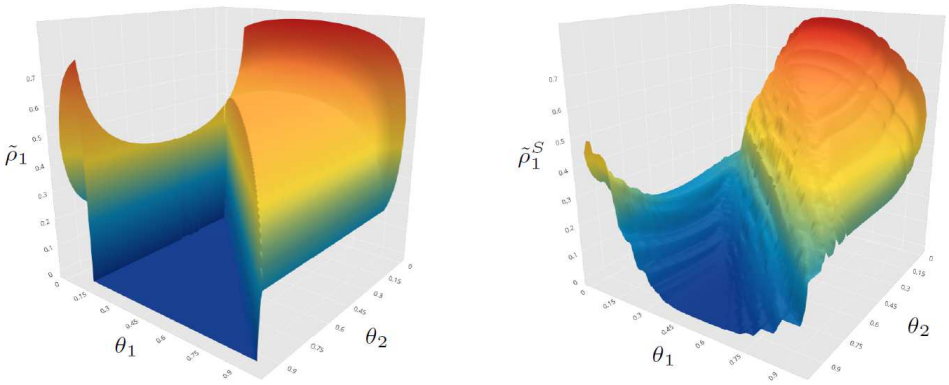


Figure 3.2: Comparison of $\tilde{\rho}_1(\boldsymbol{\theta})$ and $\tilde{\rho}_1^S(\boldsymbol{\theta})$ as $\boldsymbol{\theta} = (\theta_1, \theta_2, 0.15)^\top$ vary.

Table 3.6: Power $\mathcal{P}_n$ and total expected outcomes $\mathcal{E}_n$ (efficiencies within brackets) for binary responses with $K = 3$ treatments, as $\boldsymbol{\theta}$ varies.

$\boldsymbol{\theta}$	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
$(.4, .1, .05)^\top$	$\boldsymbol{\rho}^* = (.658, .171, .171)^\top$	.827 (.882)	.987 (.988)	29 (.682)
	$\boldsymbol{\rho}^{*S} = (.660, .170, .170)^\top$	.827 (.882)	.987 (.988)	29 (.684)
	$\tilde{\boldsymbol{\rho}} = (.692, 0, .308)^\top$	.938 (1)	.999 (1)	29 (.692)
	$\tilde{\boldsymbol{\rho}}^S = (.742, .031, .227)^\top$	.872 (.930)	.994 (.995)	31 (.747)
$(.6, .4, .25)^\top$	$\boldsymbol{\rho}^* = (.480, .260, .260)^\top$	.516 (.675)	.827 (.855)	46 (.589)
	$\boldsymbol{\rho}^{*S} = (.476, .262, .262)^\top$	.516 (.675)	.827 (.855)	46 (.591)
	$\tilde{\boldsymbol{\rho}} = (.531, 0, .469)^\top$	.765 (1)	.967 (1)	44 (.531)
	$\tilde{\boldsymbol{\rho}}^S = (.549, .041, .409)^\top$	.645 (.843)	.923 (.954)	45 (.567)

3.5 Conclusions

This paper discusses optimal designs for hypothesis testing in the presence of heterogeneous experimental groups, also encompassing the general one-way ANOVA with heteroscedastic errors. In particular, we derive the allocation maximizing the NCP of the classical Wald test of homogeneity about the treatment contrasts; this optimal design is a generalized Neyman allocation involving only two treatments, not necessarily the best and the worst ones. Moreover, to account for the ordering among treatments, we derive the optimal design maximizing the NCP of the homogeneity test, subject to an ethical constraint reflecting the efficacy of the competing treatments. Due to the dependence on the unknown model parameters, these allocations are locally optimal and therefore a-priori unknown. Moreover, these designs are degenerate allocations with possible local discontinuities. To avoid these drawbacks, a smoothing transformation via Gaussian kernel implemented via response-adaptive randomization has been proposed.

The suggested convolution avoids i) degeneracies by assigning a non-null mass on each treatment and ii) potential discontinuities. Thus, these smoothed optimal designs could be approached via standard response-adaptive randomization procedures that, by estimating at each step the unknown parameters as well as the treatment ordering, change sequentially the probabilities of treatment assignments in order to converge to the desired target.

As showed in Section 3.3, the balanced allocation is strongly inappropriate under heteroscedasticity. Moreover, any deviation from the assumption of homoscedasticity could also affect the inferential performances of $\boldsymbol{\rho}_M$, which tents to exhibit very low statistical power, especially in the case of unordered variances. In such scenarios, the new unconstrained optimal design $\tilde{\boldsymbol{\rho}}^S$, combined with response-adaptive randomization methodology, provides a remarkable gain in terms of power with respect to all the other designs, guaranteeing also high ethical standards. Nevertheless, whenever the ethical dimension plays a crucial role, $\boldsymbol{\rho}^{*S}$ could be preferable since, although it places more emphasis on the ethical aspects, it still guarantees good performances in terms of statistical power, also with respect to $\boldsymbol{\rho}_H$.

3.6 Proofs

3.6.1 Proof of Lemma 3.2.1

From (3.2.1), $\phi(\boldsymbol{\rho}) = \left(\sum_{k=1}^K \rho_k v_k^{-1}\right) V_{\boldsymbol{\pi}}(\theta) = \boldsymbol{a}^\top \boldsymbol{\rho}$ where, for every $i = 1, \dots, K$,

$$\frac{\partial V_{\boldsymbol{\pi}}(\theta)}{\partial \rho_i} = \left(v_i \sum_{k=1}^K \frac{\rho_k}{v_k}\right)^{-1} [(\theta_i - \bar{\theta}_{\boldsymbol{\pi}})^2 - V_{\boldsymbol{\pi}}(\theta)]$$

and therefore

$$\frac{\partial \phi(\boldsymbol{\rho})}{\partial \rho_i} = \frac{V_{\boldsymbol{\pi}}(\theta)}{v_i} - \frac{V_{\boldsymbol{\pi}}(\theta) - (\theta_i - \bar{\theta}_{\boldsymbol{\pi}})^2}{v_i} = a_i.$$

Furthermore,

$$\frac{\partial^2 \phi(\boldsymbol{\rho})}{\partial \rho_i \partial \rho_j} = -2 \left(\sum_{k=1}^K \frac{\rho_k}{v_k}\right)^{-1} \frac{(\theta_i - \bar{\theta}_{\boldsymbol{\pi}})}{v_i} \frac{(\theta_j - \bar{\theta}_{\boldsymbol{\pi}})}{v_j},$$

so the Hessian matrix

$$H(\phi) = -2 \left(\sum_{k=1}^K \frac{\rho_k}{v_k}\right)^{-1} \begin{bmatrix} \frac{(\theta_1 - \bar{\theta}_{\boldsymbol{\pi}})}{v_1} \\ \vdots \\ \frac{(\theta_K - \bar{\theta}_{\boldsymbol{\pi}})}{v_K} \end{bmatrix} \cdot \begin{bmatrix} \frac{(\theta_1 - \bar{\theta}_{\boldsymbol{\pi}})}{v_1} & \dots & \frac{(\theta_K - \bar{\theta}_{\boldsymbol{\pi}})}{v_K} \end{bmatrix}$$

is negative semi-definite, since it has $K - 1$ null eigenvalues and one eigenvalue (that coincides with the Laplacian)

$$\text{tr}\,(H(\phi)) = \nabla^2 \phi = -2 \left(\sum_{k=1}^K \frac{\rho_k}{v_k}\right)^{-1} \sum_{k=1}^K \frac{a_k}{v_k} < 0,$$

provided that $\mathbb{A}^\top \boldsymbol{\theta} \neq \mathbf{0}_{K-1}$.

3.6.2 Proof of Theorem 3.2.1

Firstly, we will show that the unconstrained optimal design $\tilde{\boldsymbol{\rho}}$ maximising the NCP of Wald test is of the form of (3.2.2). Let $\boldsymbol{\rho}^{(l)}$ denote a generic target with l components different from 0, then $\boldsymbol{\rho}^{(1)}$ corresponds to one of the standard basis $\boldsymbol{e}_k$ ($k = 1, \ldots, K$) and clearly $\phi(\boldsymbol{\rho}^{(1)}) = 0$, attaining the minimum of the NCP. Taking now into account $\boldsymbol{\rho}^{(2)} = \sum_{k\in\{i,j\}} \boldsymbol{e}_k\rho_k$, where $\rho_j = 1 - \rho_i \in (0; 1)$, from (3.2.1) we have

$$\phi\left(\boldsymbol{\rho}^{(2)}\right) = \left(\sum_{k\in\{i,j\}} \frac{\rho_k}{v_k}\right)^{-1} (\theta_i - \theta_j)^2 \left\{\frac{\rho_i}{v_i}\frac{1-\rho_i}{v_j}\right\},$$

and

$$\phi\left(\boldsymbol{\rho}^{(2)}\right) \leq \max_{i,j\in\{1,2,\ldots,K\}} (\theta_i - \theta_j)^2 \left\{\max_{\rho_i\in[0,1]} \left(\sum_{k\in\{i,j\}} \frac{\rho_k}{v_k}\right)^{-1} \frac{\rho_i}{v_i}\frac{1-\rho_i}{v_j}\right\}. \tag{3.6.1}$$

First, observe that

$$\max_{\rho_i\in[0,1]} \left(\sum_{k\in\{i,j\}} \frac{\rho_k}{v_k}\right)^{-1} \frac{\rho_i(1-\rho_i)}{v_iv_j} = \max_{\rho_i\in[0,1]} \left\{\frac{v_i}{\rho_i} + \frac{v_j}{1-\rho_i}\right\}^{-1} = \min_{\rho_i\in[0,1]} \frac{v_i}{\rho_i} + \frac{v_j}{1-\rho_i},$$

which is clearly minimized by the Neyman allocation $\tilde{\rho}_i = \varsigma_{ij}$. Thus, from (3.6.1),

$$\phi(\tilde{\boldsymbol{\rho}}) = \max_{i,j\in\{1,2,\ldots,K\}} (\theta_i - \theta_j)^2 \left\{\frac{\varsigma_{ij}(1-\varsigma_{ij})}{\varsigma_{ij}(v_j - v_i) + v_i}\right\} = \max_{i,j\in\{1,2,\ldots,K\}} \left(\frac{\theta_j - \theta_i}{\sqrt{v_i} + \sqrt{v_j}}\right)^2,$$

so that if the pair of treatments $\{\tilde{i}, \tilde{k}\}$ maximising the RHS of the previous equation is unique, then $\tilde{\boldsymbol{\rho}} = \tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} = \boldsymbol{e}_{\tilde{i}}\varsigma_{\tilde{i}\tilde{k}} + \boldsymbol{e}_{\tilde{k}}\varsigma_{\tilde{k}\tilde{i}}$. When the pairs $\{\tilde{i}, \tilde{k}\}$ and $\{\tilde{i}', \tilde{k}'\}$ satisfy (3.2.3), then both $\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}$ and $\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}$ are optimal designs, namely

$$\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}) = \max_{i,k\in\{1,\ldots,K\}} \left(\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}}\right)^2. \tag{3.6.2}$$

Moreover, every mixture of $\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}$ and $\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}$ is still optimal, since for every $\omega \in [0; 1]$

$$\phi\left(\omega\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} + (1-\omega)\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}\right) \geq \omega\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) + (1-\omega)\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}),$$

due to the concavity of $\phi(\cdot)$ (see Lemma 3.2.1), but from (3.6.2), $\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}) \geq \phi\left(\omega\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} + (1-\omega)\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}\right)$. Thus, $\phi\left(\omega\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} + (1-\omega)\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}\right) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'})$.

In order to deal with other scenarios, namely $\phi(\boldsymbol{\rho}^{(l)})$ with $l > 2$, assume wlog $i < j < t$. For the case $\boldsymbol{\rho}^{(3)} = \sum_{k\in\{i,j,t\}} \boldsymbol{e}_k\rho_k$, with $\rho_t = 1 - \rho_i - \rho_j$, note that $\nabla\phi(\boldsymbol{\rho}^{(3)})$ can be expressed as

$$\frac{\partial\phi(\boldsymbol{\rho}^{(3)})}{\partial\rho_k} = \frac{(\theta_k - \bar{\theta}_{\boldsymbol{\pi}})^2}{v_k} - \frac{(\theta_t - \bar{\theta}_{\boldsymbol{\pi}})^2}{v_t} = a_k - a_t, \quad k = i, j.$$

If $\exists\, k$ s.t. $a_k \neq a_t$, then $a_k > a_t$ implies that $\phi(\boldsymbol{\rho}^{(3)})$ is increasing in ρ_k, namely the NCP is maximised for $\tilde{\rho}_t = 0$, and conversely, $a_k < a_t$ implies that $\phi(\boldsymbol{\rho}^{(3)})$ is decreasing in ρ_k that is, it is maximised for

$\tilde{\rho}_k = 0$, so that in both cases $\phi(\boldsymbol{\rho}^{(3)}) \leq \phi(\tilde{\boldsymbol{\rho}})$. If instead $\nabla\phi(\boldsymbol{\rho}^{(3)}) = \mathbf{0}_K$, namely $a_i = a_j = a_t$, from $a_i = a_t$ we get $\bar{\theta}_{\boldsymbol{\pi}} = (\sqrt{v_t}\theta_i + \sqrt{v_i}\theta_t)/(\sqrt{v_t} + \sqrt{v_i})$ and, from (3.2.1),

$$\phi(\boldsymbol{\rho}^{(3)}) = a_t = v_t^{-1}\left(\theta_t - \frac{\sqrt{v_t}\theta_i + \sqrt{v_i}\theta_t}{\sqrt{v_t} + \sqrt{v_i}}\right)^2 = \left(\frac{\theta_i - \theta_t}{\sqrt{v_t} + \sqrt{v_i}}\right)^2 \leq \phi(\tilde{\boldsymbol{\rho}}).$$

By applying the same reasoning, it follows that $\phi(\boldsymbol{\rho}^{(l)}) \leq \phi(\tilde{\boldsymbol{\rho}})$ for $l = 4, \ldots, K$. Indeed, taking into account $\phi(\boldsymbol{\rho}^{(K)})$, if $\nabla\phi(\boldsymbol{\rho}^{(K)}) = \mathbf{0}_K$ then $a_1 = \ldots = a_K$; from $a_1 = a_K$ we get $\bar{\theta}_{\boldsymbol{\pi}} = (\sqrt{v_K}\theta_1 + \sqrt{v_1}\theta_K)/(\sqrt{v_K} + \sqrt{v_1})$, so that

$$\phi(\tilde{\boldsymbol{\rho}}^{(K)}) = \left(\frac{\theta_1 - \theta_K}{\sqrt{v_1} + \sqrt{v_K}}\right)^2 \leq \phi(\tilde{\boldsymbol{\rho}});$$

while if $\nabla\phi(\boldsymbol{\rho}^{(K)}) \neq \mathbf{0}_K$, then $\phi(\boldsymbol{\rho}^{(K)})$ could be maximised by some $\boldsymbol{\rho}^{(K-1)}$ and, recursively, it follows that $\phi(\boldsymbol{\rho}^{(K)}) \leq \phi(\tilde{\boldsymbol{\rho}})$, which concludes the proof.

3.6.3 Proof of Corollary 3.2.1

Since $\theta_{\tilde{i}} = \theta_{\tilde{i}'}$ and $v_{\tilde{i}} = v_{\tilde{i}'}$, the pair $\{\tilde{i}, \tilde{k}\}$ maximises the RHS of (3.2.2) if and only if $\{\tilde{i}', \tilde{k}\}$ maximises the RHS of (3.2.2). Therefore, from Theorem 3.2.1, both $\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}$ and $\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}}$ are optimal designs as well as every mixture of them, where clearly $\varsigma_{\tilde{i}\tilde{k}} = \varsigma_{\tilde{i}'\tilde{k}}$.

The case of two clusters of treatments follows easily from the previous result by noticing that

$$\max_{i,k\in\{1,\ldots,K\}}\left(\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}}\right)^2 = \left(\frac{\theta_1 - \theta_K}{\sqrt{v_1} + \sqrt{v_K}}\right)^2,$$

namely the pair $\{1, K\}$ surely maximises the RHS of (3.2.2).

3.6.4 Proof of Corollary 3.2.2

Letting $i < k$, then $\theta_i \geq \theta_k$ and therefore the RHS of (3.2.2) becomes

$$\max_{i<k}\left(\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}}\right)^2 = \left(\max_{i<k}\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}}\right)^2. \tag{3.6.3}$$

For binary trials, by letting

$$\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}} = \frac{\theta_i - \theta_k}{\sqrt{\theta_i(1-\theta_i)} + \sqrt{\theta_k(1-\theta_k)}} = F(\theta_i, \theta_k), \tag{3.6.4}$$

then $F(\cdot)$ attains its minimum when $\theta_i = \theta_k$; moreover, $F(\theta_i, \theta_k)$ is increasing in $\theta_i \in (\theta_k; \theta_1]$, since

$$\frac{\partial F(\theta_i, \theta_k)}{\partial\theta_i} = \frac{\theta_i(1-\theta_k) + \theta_k(1-\theta_i) + 2\sqrt{\theta_i\theta_k(1-\theta_i)(1-\theta_k)}}{2\sqrt{(1-\theta_i)\theta_i}\left(\sqrt{(1-\theta_i)\theta_i} + \sqrt{(1-\theta_k)\theta_k}\right)^2} > 0,$$

and decreasing in $\theta_k \in [\theta_K; \theta_i)$, because

$$\frac{\partial F(\theta_i, \theta_k)}{\partial\theta_k} = \frac{-\theta_i(1-\theta_k) - \theta_k(1-\theta_i) - 2\sqrt{\theta_i\theta_k(1-\theta_i)(1-\theta_k)}}{2\sqrt{(1-\theta_k)\theta_k}\left(\sqrt{(1-\theta_i)\theta_i} + \sqrt{(1-\theta_k)\theta_k}\right)^2} < 0.$$

Thus,

$$\max_{i<k}\left(\frac{\theta_i-\theta_k}{\sqrt{\theta_i(1-\theta_i)}+\sqrt{\theta_k(1-\theta_k)}}\right)=\frac{\theta_1-\theta_K}{\sqrt{\theta_1(1-\theta_1)}+\sqrt{\theta_K(1-\theta_K)}},$$

namely, the optimal design is $\tilde{\boldsymbol{\rho}}_{1K}$ with $\varsigma_{1K}=\sqrt{\theta_1(1-\theta_1)}/(\sqrt{\theta_1(1-\theta_1)}+\sqrt{\theta_K(1-\theta_K)})$ and $\phi(\tilde{\boldsymbol{\rho}}_{1K})=\left\{(\theta_1-\theta_K)/(\sqrt{\theta_1(1-\theta_1)}+\sqrt{\theta_K(1-\theta_K)})\right\}^2$. Taking into account Poisson outcomes,

$$\max_{i<k}\left(\frac{\theta_i-\theta_k}{\sqrt{\theta_i}+\sqrt{\theta_k}}\right)=\max_{i<k}\left(\sqrt{\theta_i}-\sqrt{\theta_k}\right)=\sqrt{\theta_1}-\sqrt{\theta_K}$$

and therefore $\tilde{\boldsymbol{\rho}}_{1K}$ is the optimal design with $\varsigma_{1K}=\sqrt{\theta_1}/(\sqrt{\theta_1}+\sqrt{\theta_K})$ and $\phi(\tilde{\boldsymbol{\rho}}_{1K})=\left\{(\theta_1-\theta_K)/(\sqrt{\theta_1}+\sqrt{\theta_K})\right\}^2$. For exponential responses,

$$\max_{i<k}\left(\frac{\theta_i-\theta_k}{\theta_i+\theta_k}\right)=\frac{\theta_1-\theta_K}{\theta_1+\theta_K}=\max_{x>1}\left(\frac{x-1}{x+1}\right),$$

where $x=\theta_i/\theta_k\in[1;\theta_1/\theta_K]$; since the RHS is increasing in x, the optimal design is $\tilde{\boldsymbol{\rho}}_{1K}$ with $\varsigma_{1K}=\theta_1/(\theta_1+\theta_K)$ and $\phi(\tilde{\boldsymbol{\rho}}_{1K})=\{(\theta_1-\theta_K)/(\theta_1+\theta_K)\}^2$. For normal homoscedastic outcomes, from (3.6.3),

$$\max_{i<k}\left(\frac{\theta_i-\theta_k}{2\sqrt{v}}\right)^2=\frac{(\theta_1-\theta_K)^2}{4v}$$

is attained at the optimal design $\tilde{\boldsymbol{\rho}}_{1K}$ with $\varsigma_{1K}=1/2$. In the presence of clusters of best(worst) treatments, ς_{1K} (ς_{K1}) should be spanned over the corresponding group of treatments.

3.6.5 Proof of Lemma 3.2.2

From now on let $\vartheta_i=\theta_1-\theta_i$ for $i=1,\ldots,K$, then $0=\vartheta_1\leq\ldots\leq\vartheta_K$, with at least one strict inequality holds. For every $i=1,\ldots,K-1$, $\sigma_j\geq\beta_j$; indeed,

$$\left(\sum_{i=1}^{K}\frac{\vartheta_i}{v_i}\right)\left(\sum_{i=1}^{j}\frac{1}{v_i}\right)\geq\left(\sum_{i=1}^{j}\frac{\vartheta_i}{v_i}\right)\left(\sum_{i=1}^{K}\frac{1}{v_i}\right)\iff$$
$$\left(\sum_{i=j+1}^{K}\frac{\vartheta_i}{v_i}\right)\left(\sum_{i=1}^{j}\frac{1}{v_i}\right)\geq\left(\sum_{i=1}^{j}\frac{\vartheta_i}{v_i}\right)\left(\sum_{i=j+1}^{K}\frac{1}{v_i}\right)$$

and therefore, by rearranging the terms in the summations,

$$\sigma_j-\beta_j=\sum_{k=1}^{j}\frac{1}{v_k}\left\{\sum_{i=j+1}^{K}\frac{\vartheta_i-\vartheta_j}{v_i}\right\}\geq 0,$$

since $\vartheta_i\geq\vartheta_j$ for every $i>j$. Moreover, note that $\sigma_j=\beta_j$ if and only if $\vartheta_j=\ldots=\vartheta_K$. Analogously, $\beta_j\geq\gamma_j$, since

$$\beta_j-\gamma_j=\sum_{k=1}^{j}\frac{\vartheta_k}{v_k}\left\{\sum_{i=j+1}^{K}\frac{\vartheta_i(\vartheta_i-\vartheta_j)}{v_i}\right\}\geq 0,$$

and the equality holds if i) $0 = \vartheta_1 = \ldots = \vartheta_j$ or ii) $\vartheta_j = \ldots = \vartheta_K$. Case i) corresponds to $\beta_1 = \ldots = \beta_j = \gamma_1 = \ldots = \gamma_j = 0$, while under ii) $\sigma_j = \beta_j = \gamma_j$. If $\gamma_j = 1$, then $\beta_j = \gamma_j$ if and only if $0 = \vartheta_1 = \ldots = \vartheta_K$; indeed, under ii), $\sigma_j = \beta_j = \gamma_j = 1$, that implies $(K-j)\sum_{i=1}^{j} v_i^{-1} = j\sum_{i=j+1}^{K} v_i^{-1}$ and

$$(K-j)\sum_{i=1}^{j}\frac{\vartheta_i}{v_i} = j\vartheta_j\sum_{i=j+1}^{K}\frac{1}{v_i} = \vartheta_j(K-j)\sum_{i=1}^{j}\frac{1}{v_i},$$

namely $\sum_{i=1}^{j}(\vartheta_i - \vartheta_j)/v_i = 0$, i.e. $0 = \vartheta_1 = \ldots = \vartheta_j$, which, combined with ii), corresponds to $\vartheta_1 = \ldots = \vartheta_K = 0$ (i.e., $\theta_1 = \ldots = \theta_K$).

Furthermore, from (3.2.5), due to the Cauchy-Schwarz inequality

$$T = \left[\sum_{k=1}^{K}\left(v_k^{-1/2}\right)^2\right]\left[\sum_{k=1}^{K}\left(\vartheta_k v_k^{-1/2}\right)^2\right]\left[\sum_{k=1}^{K}\vartheta_k v_k^{-1}\right]^{-2} \geq 1,$$

where the equality holds if and only if $\vartheta_1 = \ldots = \vartheta_K = 0$.

3.6.6 Proof of Theorem 3.2.2

To avoid a cumbersome notation, we let $\bar{\vartheta}_{\boldsymbol{\pi}} = \sum_{k=1}^{K}\vartheta_k\pi_k$ and $\bar{\vartheta}_{\boldsymbol{\pi}_B} = (\sum_{k=1}^{K}\vartheta_k v_k^{-1})/(\sum_{k=1}^{K} v_k^{-1})$ (namely assuming $\boldsymbol{\pi}_B = \boldsymbol{\pi}(\boldsymbol{\rho}_B)$). Recalling that $\phi(\boldsymbol{\rho}) = \boldsymbol{a}^\top\boldsymbol{\rho}$, where clearly $a_i = v_i^{-1}(\theta_i - \bar{\theta}_{\boldsymbol{\pi}})^2 = v_i^{-1}(\vartheta_i - \bar{\vartheta}_{\boldsymbol{\pi}})^2$. Moreover, from now on we let $\bar{a}_k = k^{-1}\sum_{i=1}^{k} a_i$ for $k = 1, \ldots, K-1$, while $\bar{a} = K^{-1}\sum_{i=1}^{K} a_i$.

The constrained maximisation problem of Theorem 3.2.2 can be address via Lagrange multipliers, where $L(\boldsymbol{\rho}, \lambda_1, \ldots, \lambda_K) = \phi(\boldsymbol{\rho}) - \sum_{i=1}^{K-1}\lambda_i(\rho_{i+1} - \rho_i) - \lambda_K\left(\sum_{i=1}^{K}\rho_i - 1\right)$.
Setting $\partial L(\boldsymbol{\rho}, \lambda_1, \ldots, \lambda_K)/\partial\rho_i = 0$ for $i = 1, \ldots, K$, then

$$\begin{cases} a_1 + \lambda_1 = \lambda_K \\ a_i - \lambda_{i-1} + \lambda_i = \lambda_K, \quad i = 2, \ldots, K-1, \\ a_K - \lambda_{K-1} = \lambda_K \end{cases}$$

namely, by summing all the equations, $\lambda_K = \bar{a} > 0$ (since $\vartheta_K > 0$) and $\lambda_i = i(\bar{a} - \bar{a}_i)$ for $i = 1, \ldots, K-1$.

Case 1: $\lambda_i = 0$ and $\lambda_j > 0$ for every $j \neq i$. Under this scenario, the corresponding target is $\boldsymbol{\rho}_i^* = \xi\sum_{k=1}^{i}\boldsymbol{e}_k + \tau\sum_{k=i+1}^{K}\boldsymbol{e}_k$, where $\tau \in (0, K^{-1}]$ and $\xi = [1 - \tau(K-i)]i^{-1} \geq \tau$, since $\mathbf{1}_K^\top\boldsymbol{\rho}_i^* = 1$. Under $\boldsymbol{\rho}_i^*$,

$$\bar{\vartheta}_{\boldsymbol{\pi}} = \bar{\vartheta}_{\boldsymbol{\pi}_B}\cdot\left\{\frac{\tau K + (1-\tau K)\beta_i}{\tau K + (1-\tau K)\sigma_i}\right\}, \tag{3.6.5}$$

where $[\tau K + (1-\tau K)\beta_i]/[\tau K + (1-\tau K)\sigma_i] \in (\beta_i/\sigma_i, 1]$. Therefore, $\boldsymbol{\rho}_i^*$ is optimal iff $\bar{a} = \bar{a}_i$ and $\bar{a} > \bar{a}_j$ for every $j \neq i$. Condition $\bar{a} = \bar{a}_i$ can be restated as

$$\bar{\vartheta}_{\boldsymbol{\pi}}^2\left(\frac{1}{K}\sum_{k=1}^{K}\frac{1}{v_k}\right)[1-\sigma_i] - 2\bar{\vartheta}_{\boldsymbol{\pi}}\left(\frac{1}{K}\sum_{k=1}^{K}\frac{\vartheta_k}{v_k}\right)[1-\beta_i] + \left(\frac{1}{K}\sum_{k=1}^{K}\frac{\vartheta_k^2}{v_k}\right)[1-\gamma_i] = 0.$$

Clearly, if $\sigma_i = 1$ then $\bar{\vartheta}_{\boldsymbol{\pi}} = \bar{\vartheta}_{\boldsymbol{\pi}_B}\{T(1-\gamma_i)/[2(1-\beta_i)]\}$. If $\sigma_i \neq 1$ instead, this is a quadratic function with respect to $\bar{\vartheta}_{\boldsymbol{\pi}}$ and the corresponding roots are $\bar{\vartheta}_{\boldsymbol{\pi}} = \bar{\vartheta}_{\boldsymbol{\pi}_B} \cdot A_i^{\pm}$, where $A_i^{\pm} = [1-\beta_i \pm \sqrt{R_i}]/(1-\sigma_i)$ and $R_i = (1-\beta_i)^2 - T(1-\sigma_i)(1-\gamma_i)$. Thus, by (3.6.5), $A_i^{\pm} = [\tau K + (1-\tau K)\beta_i]/[\tau K + (1-\tau K)\sigma_i] \in (\beta_i/\sigma_i, 1]$. However, A_i^+ is not an admissible solution since (unless the degenerate case $\theta_1 = \ldots = \theta_K$, which is clearly excluded):

- if $\sigma_i < 1$, $A_i^+ \leq 1 \iff \sqrt{R_i} \leq \beta_i - \sigma_i < 0$;
- if $\sigma_i > 1$, $A_i^+ > \beta_i/\sigma_i \iff \sigma_i\sqrt{R_i} < \beta_i - \sigma_i < 0$.

Taking into account A_i^-, when $\gamma_i \neq 1$ it can be rewritten as $A_i^- = T(1-\gamma_i)/[1-\beta_i+\sqrt{R_i}]$ (which also encompasses the case $\sigma_i = 1$), and therefore the definition of A_i in (3.2.6) follows immediately. Now, $A_i \in (\beta_i/\sigma_i, 1]$ when $P1a$ holds. Indeed,

- when $\gamma_i < 1$, $A_i \leq 1 \iff T(1-\gamma_i) \leq 1 + \sigma_i - 2\beta_i$, which also guarantees that $R_i \geq 0$ (since $(1-\beta_i)^2 \geq T(1-\sigma_i)(1-\gamma_i)$ when $\sigma_i \geq 1$, while for $\sigma_i < 1$ then $(1-\beta_i)^2/(1-\sigma_i) \geq 1+\sigma_i-2\beta_i$). In addition, $A_i > \beta_i/\sigma_i$ when $T(1-\gamma_i) > \beta_i(2\sigma_i - \beta_i - \beta_i\sigma_i)/\sigma_i^2$ (which also implies that $T(1-\gamma_i) > (1-\beta_i)\beta_i/\sigma_i$, since $\beta_i(1-\beta_i)/\sigma_i < \beta_i(2\sigma_i - \beta_i - \beta_i\sigma_i)/\sigma_i^2$);
- when $\gamma_i > 1$, $A_i > \beta_i/\sigma_i \iff T(1-\gamma_i) > \beta_i(2\sigma_i - \beta_i - \beta_i\sigma_i)/\sigma_i^2$, also ensuring that $R_i \geq 0$ (since $(1-\beta_i)^2/(1-\sigma_i) \leq \beta_i(2\sigma_i - \beta_i - \beta_i\sigma_i)/\sigma_i^2$). Moreover $A_i \leq 1 \iff 1-\beta_i \leq T(1-\gamma_i) \leq 1 + \sigma_i - 2\beta_i$ (where $T(1-\gamma_i) \geq 1-\beta_i$ is trivially satisfied when $A_i > \beta_i/\sigma_i$);
- if $\gamma_i = 1$, condition $P1a$ reduces to $2\sigma_i - \beta_i - \beta_i\sigma_i < 0 \leq 1 + \sigma_i - 2\beta_i$.

Condition $\bar{a} > \bar{a}_j$ for every $j \neq i$ corresponds to $P1b$; indeed, by combining

$$\bar{\vartheta}_{\boldsymbol{\pi}}^2 \left(\frac{1}{K}\sum_{k=1}^K \frac{1}{v_k}\right)[1-\sigma_j] - 2\bar{\vartheta}_{\boldsymbol{\pi}} \left(\frac{1}{K}\sum_{k=1}^K \frac{\vartheta_k}{v_k}\right)[1-\beta_j] + \left(\frac{1}{K}\sum_{k=1}^K \frac{\vartheta_k^2}{v_k}\right)[1-\gamma_j] > 0$$

($j = 1, \ldots, K-1$, $j \neq i$) with $\bar{\vartheta}_{\boldsymbol{\pi}} = \bar{\vartheta}_{\boldsymbol{\pi}_B} \cdot A_i$, then

$$\frac{\left(\sum_{k=1}^K \vartheta_k v_k^{-1}\right)^2 \left[A_i^2(1-\sigma_j) - 2A_i(1-\beta_j)\right]}{\sum_{k=1}^K v_k^{-1}} + \left(\sum_{k=1}^K \frac{\vartheta_k^2}{v_k}\right)(1-\gamma_j) > 0,$$

namely $T^{-1}\left(\sum_{k=1}^K \vartheta_k^2 v_k^{-1}\right)\left[A_i^2(1-\sigma_j) - 2A_i(1-\beta_j) + T(1-\gamma_j)\right] > 0$, i.e., $f_j(A_i) > 0$, where $f_j(z) = z^2(1-\sigma_j) - 2z(1-\beta_j) + T(1-\gamma_j)$. Thus, if there exists a treatment $i \in \{1, \ldots, K-1\}$ s.t. $P1$ holds, then $\boldsymbol{\rho}_i^*$ is optimal with $\tau = (\sigma_i A_i - \beta_i)/\{K[1-\beta_i - A_i(1-\sigma_i)]\}$. Moreover under $\boldsymbol{\rho}_i^*$, since $\bar{a} = \bar{a}_i$,

$$\phi(\boldsymbol{\rho}_i^*) = \boldsymbol{a}^\top \boldsymbol{\rho}_i^* = i\bar{a}[1-\tau(K-i)]i^{-1} + \tau\bar{a}(K-i) = \bar{a}$$

and therefore (3.2.8) follows directly.

Case 2: $\lambda_i = \lambda_j = 0$ and $\lambda_k > 0$ for every $k \neq \{i, j\}$. Assuming (wlog) $1 \leq i < j \leq K-1$, the corresponding target has the form

$$\boldsymbol{\rho}^* = \left(\frac{1-(j-i)\eta - (K-j)\nu}{i}\right)\sum_{k=1}^i \boldsymbol{e}_k + \eta\sum_{k=i+1}^j \boldsymbol{e}_k + \nu\sum_{k=j+1}^K \boldsymbol{e}_k, \tag{3.6.6}$$

where $\eta \geq \nu \in (0, K^{-1}]$ and $j(\eta-\nu) \leq 1-K\nu \leq j(\eta-\nu)+i(1-\eta)$. Under this scenario, $\bar{a} = \bar{a}_i = \bar{a}_j$ and $\bar{a} > \bar{a}_k$ for $k \neq \{i, j\}$. Thus, if $P1$ holds for both i and j, then $A_i = A_j$ (namely, $\bar{\vartheta}_{\boldsymbol{\pi}} = \bar{\vartheta}_{\boldsymbol{\pi}_B} \cdot A_i = \bar{\vartheta}_{\boldsymbol{\pi}_B} \cdot A_j$) and $P1b$ should be satisfied for every $k \neq \{i, j\}$. Indeed, if $A_i \neq A_j$, from $P1b$ it follows that $f_j(A_i) > 0$ but, at the same time, $f_i(A_j) > 0$, which is impossible since $f_j(A_i) > 0 \iff f_i(A_j) < 0$. Adopting $\boldsymbol{\rho}^*$ in (3.6.6), since $\bar{a} = \bar{a}_i = \bar{a}_j$,

$$\begin{aligned}\phi(\boldsymbol{\rho}^*) = \boldsymbol{a}^\top \boldsymbol{\rho}^* &= \left(\frac{1-(j-i)\eta-(K-j)\nu}{i}\right)\sum_{k=1}^{i} a_k + \eta\sum_{k=i+1}^{j} a_k + \nu\sum_{k=j+1}^{K} a_k \\ &= \left(\frac{1-(j-i)\eta-(K-j)\nu}{i}\right) i\bar{a} + \eta\bar{a}(j-i) + \nu\bar{a}(K-j) = \bar{a},\end{aligned}$$

so that, combined with (3.2.8), $\phi(\boldsymbol{\rho}^*) = \phi(\boldsymbol{\rho}_i^*) = \phi(\boldsymbol{\rho}_j^*)$. Moreover, every mixture of $\boldsymbol{\rho}_i^*$ and $\boldsymbol{\rho}_j^*$ is optimal since, due to the concavity of $\phi(\cdot)$,

$$\phi(\boldsymbol{\rho}_i^*) = \phi(\boldsymbol{\rho}_j^*) \geq \phi\left(\omega\boldsymbol{\rho}_i^* + (1-\omega)\boldsymbol{\rho}_j^*\right) \geq \omega\phi(\boldsymbol{\rho}_i^*) + (1-\omega)\phi(\boldsymbol{\rho}_j^*) = \phi(\boldsymbol{\rho}_i^*) = \phi(\boldsymbol{\rho}_j^*).$$

Case 3: $\lambda_i > 0$ for every $i = 1, \ldots, K-1$. Under this setting, $\boldsymbol{\rho}^* = \boldsymbol{\rho}_B$ and $\bar{\vartheta}_{\boldsymbol{\pi}} = \bar{\vartheta}_{\boldsymbol{\pi}_B}$. Balance is optimal provided that $\bar{a} > \bar{a}_i$ for $i = 1, \ldots, K-1$, i.e.

$$\bar{\vartheta}^2_{\boldsymbol{\pi}_B}\left(\frac{1}{K}\sum_{k=1}^{K}\frac{1}{v_k}\right)[1-\sigma_i] - 2\bar{\vartheta}_{\boldsymbol{\pi}_B}\left(\frac{1}{K}\sum_{k=1}^{K}\frac{\vartheta_k}{v_k}\right)[1-\beta_i] + \left(\frac{1}{K}\sum_{k=1}^{K}\frac{\vartheta_k^2}{v_k}\right)[1-\gamma_i] > 0,$$

namely $f_i(1) > 0$ for every $i = 1, \ldots, K-1$ (condition $P2$) and

$$\phi(\boldsymbol{\rho}_B) = \bar{a} = \frac{1}{K}\left[\bar{\vartheta}^2_{\boldsymbol{\pi}_B}\sum_{k=1}^{K}\frac{1}{v_k} - 2\bar{\vartheta}_{\boldsymbol{\pi}_B}\sum_{k=1}^{K}\frac{\vartheta_k}{v_k} + \sum_{k=1}^{K}\frac{\vartheta_k^2}{v_k}\right] = \left(\frac{T-1}{TK}\right)\sum_{k=1}^{K}\frac{\vartheta_k^2}{v_k}.$$

Case 4: in all the other scenarios, namely when $P2$ does not hold and $\nexists i$ satisfying $P1$, then the Lagrangian does not admit critical points, so that only boundary solutions could be optimal. Due to the constraints $\rho_1^* \geq \ldots \geq \rho_K^*$, then $\rho_i^* = 0$ implies that $\rho_j^* = 0$ for every $j > i$; moreover, $\rho_2^* = 0$ should be excluded since it corresponds to the minimum of the NCP. Starting from the case with $\rho_K^* = 0$ and using the same notation of Appendix 3.6.2, notice that $\phi\left(\boldsymbol{\rho}^{(K-1)^\top}; 0\right) = \phi\left(\boldsymbol{\rho}^{(K-1)}\right)$; thus, if $P1$ or $P2$ are satisfied, then $\dot{c} = 1$ and therefore $\phi\left(\boldsymbol{\rho}^{(K-1)}\right) \leq \phi\left(\boldsymbol{\rho}^*_{[K-1]}\right)$; otherwise $\rho^*_{K-1} = \rho^*_K = 0$ and, iteratively, the same reasoning should be applied at most for $\dot{c} = K-2$, for which results of Remark 3.2.4 hold.

3.6.7 Proof of Corollary 3.2.3

If $v(\theta)$ is non-decreasing in θ, then $\{v_k^{-1}; k = 1, \ldots, K\}$, $\{\vartheta_k v_k^{-1}; k = 1, \ldots, K\}$ and $\{\vartheta_k^2 v_k^{-1}; k = 1, \ldots, K\}$ are non decreasing in k and therefore $\{(\sigma_k, \beta_k, \gamma_k); k = 1, \ldots, K-1\}$ is still non-decreasing with $1 \geq \sigma_k \geq \beta_k \geq \gamma_k \geq 0$ for every $k = 1, \ldots, K-1$. Under condition (3.2.10),

- $T > 1 + \sigma_1$ implies that $P2$ holds, since $(1+\sigma_1)(1-\gamma_i) > 1 + \sigma_i - 2\beta_i$ for $i = 2, \ldots, K-1$;

- while if $T \leq 1 + \sigma_1$, then $A_1 \in (0; 1]$ and $P1$ is satisfied, since $f_i(A_1) > 0$ for every $i = 2, \ldots, K-1$; indeed, it corresponds to

$$T\left[(1-\sigma_i) - (1-\gamma_i)(1-\sigma_1)\right] + 2(\beta_i - \gamma_i)\left[1 + \sqrt{1 - T(1-\sigma_1)}\right] > 0, \tag{3.6.7}$$

  which trivially holds if $1 - \sigma_i \geq (1-\gamma_i)(1-\sigma_1)$, while for $1 - \sigma_i < (1-\gamma_i)(1-\sigma_1)$ the LHS in (3.6.7) is monotonically decreasing in T and (3.6.7) is satisfied for $T = 1 + \sigma_1$.

For the normal homoscedastic model the proof is straightforward, since $\sigma_i = 1$ for $i = 1, \ldots, K-1$ and condition (3.2.10) is trivially satisfied because $\beta_i \geq \gamma_i$ for $i = 2, \ldots, K-1$. Under Poisson data, let $\psi_k = \theta_1/\theta_k$ for $k = 1, \ldots, K$ (where $1 = \psi_1 \leq \ldots \leq \psi_K$) and $\bar{\psi} = K^{-1}\sum_{i=1}^{K}\psi_i$ (while $\bar{\psi}_k = k^{-1}\sum_{i=1}^{k}\psi_i$). Thus, $K^{-1}\sum_{k=1}^{K} v_k^{-1} = \bar{\psi}/\theta_1$ and $K^{-1}\sum_{k=1}^{K}\vartheta_k v_k^{-1} = \bar{\psi} - 1$, so that (3.2.10) becomes

$$\frac{1-\gamma_j}{\bar{\psi}} \geq \frac{\bar{\psi}_j}{\bar{\psi}} + \gamma_j - 2\left(\frac{\bar{\psi}_j - 1}{\bar{\psi} - 1}\right),$$

namely

$$(\bar{\psi}_j - 1)\left[\frac{2}{\bar{\psi} - 1} - \frac{1}{\bar{\psi}}\right] \geq \left(\frac{\bar{\psi} + 1}{\bar{\psi}}\right)\gamma_j \iff \frac{\bar{\psi}_j - 1}{\bar{\psi} - 1} = \beta_j \geq \gamma_j.$$

For exponential data, by letting $\bar{\psi^2} = K^{-1}\sum_{i=1}^{K}\psi_i^2$ and $\bar{\psi}_k^2 = k^{-1}\sum_{i=1}^{k}\psi_i^2$, then $K^{-1}\sum_{k=1}^{K} v_k^{-1} = \bar{\psi^2}/\theta_1^2$ and $K^{-1}\sum_{k=1}^{K}\vartheta_k v_k^{-1} = (\bar{\psi^2} - \bar{\psi})/\theta_1$, while $K^{-1}\sum_{k=1}^{K}\vartheta_k^2 v_k^{-1} = \bar{\psi^2} - 2\bar{\psi} + 1$. Hence, after some algebra, (3.2.10) becomes

$$\bar{\psi}_k^2 - 1 \geq \frac{\bar{\psi}_k^2 - 2\bar{\psi}_k + 1}{\bar{\psi^2} - 2\bar{\psi} + 1}(\bar{\psi^2} - 1) \iff \frac{\bar{\psi}_k - 1}{\bar{\psi} - 1} \geq \frac{\bar{\psi}_k^2 - 1}{\bar{\psi^2} - 1}$$

and, by rearranging the terms in the summation, we obtain

$$\begin{aligned}
&(K-k)\left[\sum_{i=1}^{k}\psi_i(\psi_i - 1)\right] - \sum_{h=1}^{k}\sum_{i=k+1}^{K}\psi_h\psi_i(\psi_h - \psi_i) \geq k\sum_{i=k+1}^{K}\psi_i(\psi_i - 1) \iff \\
&\sum_{h=1}^{k}\sum_{i=k+1}^{K}\psi_h(\psi_i - 1)(\psi_i - \psi_h + 1) \geq \sum_{h=1}^{k}\sum_{i=k+1}^{K}\psi_i(\psi_i - 1) \iff \\
&\sum_{h=1}^{k}\sum_{i=k+1}^{K}(\psi_i - 1)(\psi_h - 1)(\psi_i - \psi_h) \geq 0,
\end{aligned}$$

which is satisfied since $\psi_i \geq 1$ for every $i = 1, \ldots, K$ and $\psi_i \geq \psi_h$ (since $h < i$).

3.6.8 Proof of Proposition 3.2.1

For binary trials, $\exists\,!\,\tilde{i} \in \{1, \ldots, K\}$ s.t. $a_1 \geq \cdots \geq a_{\tilde{i}} \leq a_{\tilde{i}+1} \leq \cdots \leq a_K$, namely $\{a_i, i = 1, \ldots, K\}$ is decreasing for $k \leq \tilde{i}$ and increasing for $k > \tilde{i}$. Indeed, by equating $a_i = a_{i+1}$ we get

$$\bar{\vartheta}_{\boldsymbol{\pi}}^2(v_{i+1} - v_i) - 2\bar{\vartheta}_{\boldsymbol{\pi}}(\vartheta_i v_{i+1} - \vartheta_{i+1}v_i) + \vartheta_i^2 v_{i+1} - \vartheta_{i+1}^2 v_i = 0,$$

where clearly if $v_i = v_{i+1}$, then $\bar{\vartheta}_{\boldsymbol{\pi}} = (\vartheta_i + \vartheta_{i+1})/2$, while if $v_i \neq v_{i+1}$ this is a quadratic function with respect to $\bar{\vartheta}_{\boldsymbol{\pi}}$ whose corresponding roots are $\bar{\vartheta}_{\boldsymbol{\pi}} = Q_i^{\pm}$, with $Q_i^{\pm} = [\vartheta_i\sqrt{v_{i+1}} \pm \vartheta_{i+1}\sqrt{v_i}]/[\sqrt{v_{i+1}} \pm \sqrt{v_i}]$. However, since $0 = \vartheta_1 < \bar{\vartheta}_{\boldsymbol{\pi}} < \vartheta_K \leq 1$, then Q_i^- is not a feasible solution; indeed, if $v_{i+1} > v_i$, then $Q_i^- < 0$, while if $v_{i+1} < v_i$ then $Q_i^- > 1$. Thus, the only feasible solution is $Q_i^+ = \varsigma_{i+1,i}\vartheta_i + \varsigma_{i,i+1}\vartheta_{i+1} \in [\vartheta_i; \vartheta_{i+1}]$ (which also encompasses the case $v_i = v_{i+1}$) and

$$a_i \geq a_{i+1} \Longleftrightarrow (\bar{\vartheta}_{\boldsymbol{\pi}} \geq Q_i^+) \cup (\vartheta_i = \vartheta_{i+1}). \tag{3.6.8}$$

Since the sequence $\{Q_i^+, i = 1, \ldots, K-1\}$ is increasing in i, combined with (3.6.8) it guarantees that $\exists ! \, \tilde{i} \in \{1, \ldots, K\}$, such that $a_1 \geq \cdots \geq a_{\tilde{i}} \leq a_{\tilde{i}+1} \leq \cdots \leq a_K$. Moreover, if $\bar{a} \geq a_1$, then $\bar{a} > \bar{a}_{K-1}$ (i.e., $\bar{a} < a_K$), since $(K-1)\bar{a} + a_K > \sum_{k=1}^{K-1} a_k + a_K$ is always satisfied due to the fact that $\bar{a} < \max_{i=1,\ldots,K} a_i$.

Case 1: $\lambda_1 = 0$ and $\lambda_i > 0$ for $i = 2, \ldots, K-1$. Under this setting the optimal target is $\boldsymbol{\rho}_1^* = [1 - (K-1)\tau]\boldsymbol{e}_1 + \tau \sum_{k=2}^{K} \boldsymbol{e}_k$, where τ follows easily from Theorem 3.2.2 after simple algebra. Condition $\bar{a} = a_1$, namely $A_1 \in (0, 1]$, holds iff $T \leq 1 + \sigma_1$. In this case also $P1b$ is satisfied, i.e. $\bar{a} > \bar{a}_i$ for $i = 2, \ldots, K-1$, since $\bar{a} > \bar{a}_{K-1}$ due to the behaviour of $\{a_i, i = 1, \ldots, K\}$.

Case 2 $\lambda_1 = \cdots = \lambda_j = 0$ (with $j = 2, \ldots, K-2$) and $\lambda_i > 0$ for $i > j$. Under this setting $\bar{a} = \bar{a}_i$ for $i = 1, \ldots, j$ and $\bar{a} > \bar{a}_i$ for $i = j+1, \ldots, K-1$, then $\theta_1 = \cdots = \theta_j$ (which implies $A_1 = \cdots = A_j$) and $P1a$ holds for $i = 1, \ldots, j$ iff $T \leq 1 + \sigma_1$. Following the same reasoning of Case 1, $P1b$ for $i = j+1, \ldots, K-1$ also holds, so that $\boldsymbol{\rho}_1^*, \boldsymbol{\rho}_2^*, \ldots, \boldsymbol{\rho}_j^*$ are optimal designs. Moreover, due to the concavity of $\phi(\cdot)$, every convex combination of $\boldsymbol{\rho}_1^*, \boldsymbol{\rho}_2^*, \ldots, \boldsymbol{\rho}_j^*$ is still optimal.

Case 3: $\lambda_i > 0$ for every $i = 1, \ldots, K-1$. Under this scenario $\boldsymbol{\rho}^* = \boldsymbol{\rho}_B$ and when $T > 1 + \sigma_1$ then $\bar{a} > a_1$, which also guarantees $\bar{a} > \bar{a}_i \, \forall i = 2, \ldots, K-1$ (namely $P2$ holds), due to the behaviour of $\{a_i, i = 1, \ldots, K\}$.

Every other scenario is impossible since:

- $\lambda_1 > 0$ and at least one $\lambda_i = 0$ with $i \in \{2, \ldots, K-1\}$, namely $\bar{a} > a_1$ and $\bar{a} = \bar{a}_i$ where at least $a_i > \bar{a}$, clearly implies that $\bar{a} < a_i < a_{i+1}$. However, if $\lambda_{i+1} > 0$, then $\bar{a} = \bar{a}_i > \bar{a}_{i+1} \Longleftrightarrow \bar{a} > a_{i+1}$, while if $\lambda_{i+1} = 0$, then $\bar{a} = \bar{a}_i = \bar{a}_{i+1} \Longleftrightarrow \bar{a} = a_{i+1}$;
- $\lambda_1 = 0$ and at least one $\lambda_i > 0$ and $\lambda_{i+1} = 0$ with $i \in \{2, \ldots, K-2\}$, namely $\bar{a} = a_1$ and $\bar{a} = \bar{a}_{i+1}$, so that $a_2 < \bar{a} < a_{i+1} < a_{i+2}$, but if $\lambda_{i+2} = 0$, then $\bar{a} = a_{i+2}$, while if $\lambda_{i+2} > 0$, then $\bar{a} > a_{i+2}$.

Bibliography

[1] Baldi Antognini A, Giovagnoli A. *Adaptive Designs for Sequential Treatment Allocation*. Chapman & Hall/CRC Biostatistics; 2015.

[2] Baldi Antognini A, Novelli M, Zagoraiou M. Optimal designs for testing hypothesis in multiarm clinical trials. *Stat Methods Med Res* 2019; 28: 3242–3259.

[3] US Food and Drug Administration. Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics (draft document). *Available online.* 2019.

[4] Hu F, Rosenberger WF. *The Theory of Response-Adaptive Randomization in Clinical Trials.* New York: John Wiley & Sons; 2006.

[5] Hu F, Zhang LX. Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. *Ann Stat* 2004; 32: 268–301.

[6] Rosenberger WF, Lachin JM. *Randomization in clinical trials: theory and practice*. John Wiley & Sons; 2015.

[7] Singh SP, Davidov O. On the design of experiments with ordered treatments. *J R Stat Soc Series B Stat Methodol* 2019; 81: 881–900.

[8] Sverdlov O, Tymofyeyev Y, Wong WK. Optimal response-adaptive randomized designs for multi-armed survival trials. *Stat Med* 2011; 30: 2890–2910.

[9] Tymofyeyev Y, Rosenberger WF, Hu F. Implementing optimal allocation in sequential binary response experiments. *J Am Stat Assoc* 2007; 102: 224–234.

[10] Zhang LX, Rosenberger WF. Response-adaptive randomization for survival trials: the parametric approach. *Appl Statist* 2007; 56: 153–165.

[11] Zhu H, Hu F. Implementing optimal allocation for sequential continuous responses with multiple treatments. *J Stat Plan Inference* 2009; 139: 2420–2430.

Part II

Chapter 4

Paper C

Design of experiments and manufacturing design space for multi-step processes

Rosamarie Frieri[1,2] *, Marco Mariti* [2] *and Marilena Paludi* [2]

[1] Department of Statistical Sciences, University of Bologna

[2] GSK, Technical R&D, Siena, Italy

Abstract

Most industrial processes are composed of multiple subsequent steps. In this paper we provide a statistical approach to design experiments and to define the manufacturing design space of multi-step processes by taking into account the complex system of interactions among steps. We consider each intermediate outcome as an additional input factor in the next step and we plan experiments following a particular sequential structure. To encompass the potential deviations from the target levels of such input factors, designs are selected according to the D-optimality in average criterion and, in order to assess their prediction capabilities, a suitable extension of the fraction of design space technique has been proposed. The manufacturing design space of the process is then defined by combining the interconnected manufacturing design spaces of the process steps and by deriving the linear combination of the process inputs that ensures the required quality standard for the final outcome. Appealing properties of this approach are also shown by the application to a three steps biochemical process of expression and purification of a recombinant protein in which ten input factors are included in the design.

4.1 Introduction

A multi-step process is a system composed of multiple subsequent stages: the quality of the final product is the result of the interactions among the steps. More specifically, the quality of the product at one stage is not only affected by the operations performed at that stage, but also by those of previous stages.[1] Dealing with multi-step processes is common in many industrial fields, but their complex structure still presents significant challenges for researchers.

In particular in pharmaceutical industry, the introduction of the Quality by Design (QbD)[2] paradigm has increased the demand of systematic and science-based approaches to support pharmaceutical development and manufacturing activities. QbD principles have been developed to guarantee high level of product and process understanding and, hence, high level of quality for patients.[3,4] The QbD implementation includes the identification of the Critical Quality Attributes (CQAs) of pharmaceutical products, defined as those attributes which impact the clinical performance. Then, to ensure product quality, the manufacturing process should be designed to obtain these attributes consistently at the desired values. The multidimensional combination and interaction of inputs and process parameters such that the final CQAs meet the desired ranges is defined as the *manufacturing design space**, which is one of the fundamental concept in the QbD paradigm.[2]

Clearly, the establishment of the manufacturing design space in a multi-step context should take into account that i) the quality of the final product is the result of the combination of operations and resources employed in the subsequent steps, which are, in turn, controlled by multiple input factors/process parameters and ii) input factors of different steps interact to determine the intermediate and the final outcome/CQA. Such structure and the lack of first principle models pose technical issues in the design, analysis and optimization of multi-step processes.

Usually, given the complexity of the problem, the process is investigated one step at-a-time.[5] This strategy fails to detect potentially critical interactions between steps and requires a remarkable amount of experimental resources. On the other hand, by approaching the multi-step process as a big single-stage with traditional experimental designs would involve a large study, with potentially around 10 or more interacting factors, usually not affordable in practice (for biochemical processes the number of factors typically investigated in a single experiment is below 10). Indeed, a full description and understanding of the system should explore all the potential interactions between all the input factors of the process, requiring a high experimental effort. Therefore a suitable strategy to design experimental studies for multi-step processes is fundamental for scientists, especially in pharmaceutical industry in which the compliance with regulatory guidelines is often matched with the need to speed-up the development process.

Literature on multi-step processes has mainly addressed process robustness, control and optimization and discussion of case studies.[6–8] To the best of our knowledge, research on design of experiment in a multi-stage framework has been focused on split-plot designs and its variants (see the review of Yuangyai and Lin[9]). Multi-stage experimentation can be performed in a split-plot fashion, but the complexity ramp up as the number of steps increases.[5,10] In general, the proposals are not flexible in terms of number of steps involved, number of factors, number of levels for each factor and type of experimental design used.[9]

In this paper, we introduce an efficient approach to design an experimental study for a multi-step

*Note that in the ICH guideline Q8 on pharmaceutical development[2] this CQAs-consistent production space is named *design space*. However, this terminology may be ambiguous in the context of this manuscript since the term design space is commonly used in the design of experiment literature to refer to the set of all the possible design points of the experiment.

process. Motivated by QbD principles, the final objective of the design strategy is the determination of the manufacturing design space of the whole process. The procedure consists in designing a set of experiments to fit an empirical model for the final outcome. We present a new framework in which the output of each step is included as an input factor in the subsequent step and experiments are planned following a particular sequential structure. The setting of the additional input factor is subject to error since it can be set only by changing the input factors of the previous step, according to a model. Then, to fit a tentative model for each step, the D-optimal design in average [11,12] has been adopted as a design criterion. Since the error in setting input factors affects the prediction properties of the design, we propose an extension of the fraction of design space technique, which includes this effect. Once the model for the final outcome has been estimated, starting from the quality requirements on the final product, the multi-step manufacturing design space is derived by combining the interconnected manufacturing design spaces of the steps. As a result, a set of operating ranges for all the directly controlled input factors of the process is obtained. The application of the procedure is particularly relevant for - but not limited to - biochemical processes within pharmaceutical industry, for which we report an illustrative case-study example.

The paper is organized as follows: in Section 4.2 we formalize the structure of multi-step processes and we introduce the notation and in Section 4.3 we present the sequential strategy to design experiments. In Section 4.4 we report the method to define the multi-step manufacturing design space, while in Section 4.5, the complete procedure is implemented for a three step process of expression and purification of a recombinant protein. In Section 4.6 we conclude with a discussion and future developments.

4.2 Multi-step processes and notation

Let us consider a process made of V steps, $\mathcal{S}_1, \dots, \mathcal{S}_V$, and let us denote with $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots)'$ the set of controllable input factors of $\mathcal{S}_i$, for $i = 1, \dots, V$, and with $y^{(V)}$ the output of the process (Figure 4.1). The main concern of multi-step systems is that the behaviour of the final outcome depends both on the effect of process inputs and on the interactions among steps, namely $y^{(V)}$ depends on $\mathbf{x}^{(i)}, \forall i = 1, \dots, V$ and their interactions.
In such complex systems, mechanistic models relating inputs and outputs are not generally available so that researchers have to rely on Design of Experiment (DoE) techniques. [13] Hence, empirical models are devoted to the identification of the manufacturing design space, which then consists in learning how to set $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(V)}$ in such a way that $y^{(V)}$ has the desired characteristics (e.g. quality, safety).

On the one hand, if the process is intended as a big single stage (Figure 4.1), the behaviour of $y^{(V)}$ can be typically described by a regression model $y^{(V)} = f(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(V)}; \boldsymbol{\beta}) + \epsilon_V$ where $f(\cdot; \boldsymbol{\beta})$ is the model response, $\boldsymbol{\beta}$ is the unknown model parameters vector and ϵ_V is a random error term. Designing an experiment to fit such a complex model would require, in general, a high number of runs, due to the high number of parameters to estimate. Even if few steps are involved in the process, the planning of the experiment with traditional DoE techniques [13] is hardly affordable in practice. Indeed, let us consider an experimental scenario with $V = 3$, in which the experimenter would like to study three input factors per step. Assuming $f(\cdot, \boldsymbol{\beta})$ linear in $\boldsymbol{\beta}$ involving main, quadratic and two factor interactions effects, the number of parameters to be estimated in the model for $y^{(3)}$ would be 54 (plus the intercept). On the other hand, designing experiments by not taking into account the interactions among steps (e.g. by changing input factors one-step-at-a-time and observing the final outcome) would require prohibitive experimental resources giving only partial knowledge of the process. These drawbacks call for an *ad hoc*

design strategy to effectively plan experiments in a multi-step context.

Figure 4.1: Multi-step process with V steps.

Let us assume that the output of each intermediate step can be observed, so that a multi-step process can be represented as in Figure 4.2, where $y^{(i)}$ is the outcome of $\mathcal{S}_i$ for $i = 1, \ldots, V$. In contrast to the setting of Figure 4.1, we consider the structure of the process as made of multiple components. From now on, we assume that the experimenter is interested in a single characteristic (i.e. a single CQA in the context of pharmaceutical development) of each intermediate material. This assumption is relevant for the development to our proposal (for a thorough discussion see Section 4.6): only one output is carried out as an input in the next step.

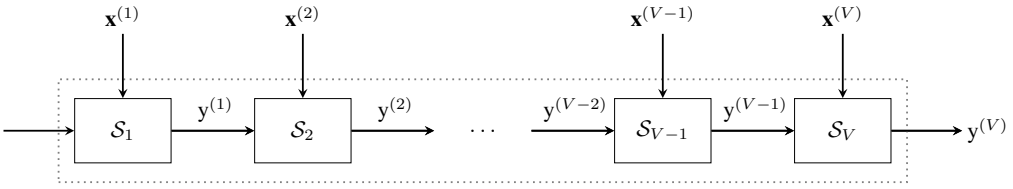


Figure 4.2: Multi-step process with V steps: new modelling.

Within this setting, the first step and so $y^{(1)}$, depends on $\mathbf{x}^{(1)}$ while each $y^{(i)}$ for $i = 2, \ldots, V$ is affected by the input factors proper of that step, $\mathbf{x}^{(i)}$, but also by the outcome of the previous one, $y^{(i-1)}$. We then introduce a mathematical formalization of a multi-step process in which each step is characterized by a model,

$$
\begin{aligned}
y^{(1)} &= f_1(\mathbf{x}^{(1)}; \boldsymbol{\beta}^{(1)}) + \epsilon^{(1)}, \\
&\vdots \\
y^{(i)} &= f_i(\mathbf{x}^{(i)}, y^{(i-1)}; \boldsymbol{\beta}^{(i)}) + \epsilon^{(i)}, \\
&\vdots \\
y^{(V)} &= f_V(\mathbf{x}^{(V)}, y^{(V-1)}; \boldsymbol{\beta}^{(V)}) + \epsilon^{(V)},
\end{aligned}
\tag{4.2.1}
$$

where for the i-th step, $f_i(\cdot)$ is the response model, $\mathbf{x}^{(i)}$ is the vector of input factors, $\boldsymbol{\beta}^{(i)}$ is the vector of the unknown model parameters and $\epsilon^{(i)}$ is a random error. We assume $f_i(\cdot)$ linear in the parameters so that the i-th equation of (4.2.1) can be restated as $y^{(i)} = \mathbf{z}_i^\top \boldsymbol{\beta}^{(i)} + \epsilon^{(i)}$, where $\mathbf{z}_i^\top$ is the vector (of size p_i) whose components are the term included in the model $f_i(\cdot)$. For instance, if the i-th step has two input factors and $f_i(\cdot)$ is a full quadratic model (without intercept) then $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)})$ and $\mathbf{z}_i^\top = (x_1^{(i)}, x_2^{(i)}, y^{(i-1)}, x_1^{(i)} \cdot x_2^{(i)}, x_1^{(i)} \cdot y^{(i-1)}, x_2^{(i)} \cdot y^{(i-1)}, [x_1^{(i)}]^2, [x_2^{(i)}]^2, [y^{(i-1)}]^2)$ with $p_i = 9$. In addition, by

letting $\mathbf{y}^{(i)}$ be the vector of the n_i observations from step i, then $\mathbf{y}^{(i)} = \mathbf{Z}_i\boldsymbol{\beta}^{(i)}+\boldsymbol{\epsilon}^{(i)}$ where $\mathbf{Z}_i$ is the design matrix expanded to the $f_i(\cdot)$ model form (i.e. it has size $n_i \times p_i$). For $\boldsymbol{\epsilon}^{(i)} = (\epsilon_1^{(i)}, \ldots, \epsilon_u^{(i)}, \ldots, \epsilon_{n_i}^{(i)})^\top$ we assume that $\epsilon_u^{(i)}$ has zero mean and variance $\sigma_i^2, \forall u = 1, \ldots, n_i$ (homoschedasticity) and that the $\{\epsilon_u^{(i)}\}$ are uncorrelated random variables. Finally, since we deal with situations in which the real experimental conditions may fluctuates around those specified in the design, i.e the actual design matrix could be different to the the planned design matrix, we denote with $\tilde{\mathbf{Z}}_i$ the observed design matrix expanded to the form of $f_i(\cdot)$.

4.3 Design of experiments for multi-step processes

In this section we present the design strategy to plan the experiments in multi-step processes. In general, the model for the final outcome is used to draw conclusions on the end-process material and so to derive the manufacturing design space. Following the set-up in (4.2.1), to fit an empirical model for $y^{(V)}$ we have to estimate V models. Thus, we need to generate V experimental plans: the procedure consists of a sequence of V phases where, in the generic *phase* i, the objective is to fit a model for the outcome of $\mathcal{S}_i$. On this purpose, *phase* i consists of (a) the design of the experiment for $\mathcal{S}_i$ (b) the implementation of the experiments to observe the $y^{(i)}$s and finally (c) the analysis of the results for fit a model for $y^{(i)}$. So in *phase* i (i.e. *phase* 1,..., *phase* $i-1$ have already been run), the experimental design for $\mathcal{S}_i$ is derived according to a suitable optimality criterion. Such experimental design provides target levels for $\mathbf{x}^{(i)}$ and $y^{(i-1)}$. However, since $y^{(i-1)}$ is the output of $\mathcal{S}_{i-1}$, its target levels cannot be set directly as for $\mathbf{x}^{(i)}$. The desired target values can be reached only after running the previous steps, $\mathcal{S}_1$, $\mathcal{S}_2$,...,$\mathcal{S}_{i-1}$, under proper settings of $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(i-1)}$. In particular, $y^{(i-1)}$ can be controlled only through the inputs of $\mathcal{S}_{i-1}$ and the model estimates $\hat{f}_{i-1}(\cdot)$ (estimator of $f_{i-1}(\cdot)$) computed in *phase* $i-1$, but due to the prediction error intrinsic in each model, $y^{(i-1)}$ cannot be set precisely to the desired levels. More specifically, while all the input factors $\mathbf{x}^{(i)}$ in step i are subjected to a negligible experimental error, $y^{(i-1)}$ is also subjected to the error of the model for $\mathcal{S}_{i-1}$. Therefore, due to the multi-step structure of the process, each experimental plan involves one factor whose levels are set with error: the experiments should be designed by taking into account the potential deviations from the target levels that may occur due to this error.

To the best of our knowledge, the first author interested in the effect of error in setting factor levels was Box[14] and this topic has been more recently addressed by Pronzato[11] and Donev[12] in the context of optimal designs. However, robustness to error in setting factor levels of optimal designs has not been much addressed in the literature.[15]

In order to clarify the experimental strategy proposed, Example 4.3.1 and Figure 4.3 illustrate the multi-step procedure in the case of three steps.

Example 4.3.1. *The multi-step experimental strategy is illustrated in Figure 4.3 in the case of a three-step process: the procedure consists of three phases. In phase 1 the experiments are performed on* $\mathcal{S}_1$ *to fit a model for* $y^{(1)}$*. In phase 2 the experiments are performed on* $\mathcal{S}_1$ *and* $\mathcal{S}_2$ *to fit a model for* $y^{(2)}$*. However, note the difference between the experiments on* $\mathcal{S}_1$ *in phase* 1 *and in phase* 2*. In phase* 1*, the experiments on* $\mathcal{S}_1$ *are performed under some settings of* $\mathbf{x}^{(1)}$ *that are optimal to fit a model for* $y^{(1)}$*. In phase* 2*, the experiments on* $\mathcal{S}_1$ *are run once again but under different set-ups of* $\mathbf{x}^{(1)}$*, that are the ones ensuring* $y^{(1)}$ *close to the target levels optimal to fit a model for* $y^{(2)}$*. The same reasoning holds for phase 3.*

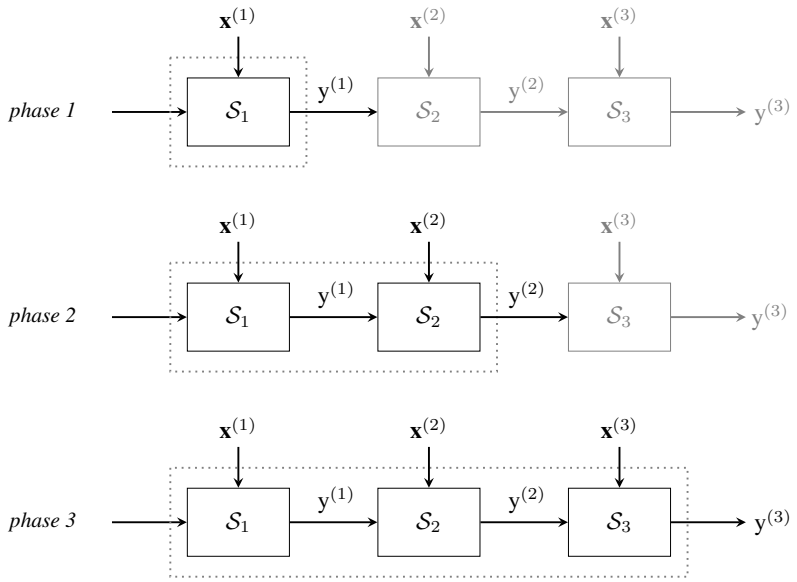


Figure 4.3: Multi-step experimental strategy for a three-step process.

Section 4.3.1 is focused on how to design the experiment for $\mathcal{S}_i$ at *phase* i and the complete and detailed procedure to implement the proposed experimental strategy from *phase* 1 up to *phase* V is reported in Section 4.3.2.

4.3.1 Design of experiment for step i in phase i

Let us consider that *phase* 1, *phase* 2, ..., *phase* $i-1$ have already been run, i.e. $\hat{f}_1(\cdot), \ldots, \hat{f}_{i-1}(\cdot)$ are available. For instance, *phase* $i-1$ is devoted to make experiments to fit a model for $y^{(i-1)}$, thus the experiments have been carried out from $\mathcal{S}_1$ up to $\mathcal{S}_{i-1}$ and the experimental conditions (values of $\mathbf{x}^{(i-1)}$ and $y^{(i-2)}$) that generated the observed $y^{(i-1)}$s are known so that $\tilde{\mathbf{Z}}_{i-1}$ is available and used to estimate the parameters $\hat{\boldsymbol{\beta}}^{(i-1)}$ via least squares.

We are now in *phase* i. First, we wish to design an experiment to fit an empirical model for $y^{(i)}$ (*phase* i (a)) Typically, input factors are set to target levels according to a given experimental plan. The plan is generated following optimality criteria, which are generally based on convex functions of the information matrix of the design[16] (e.g. the well-known D-optimality requires the maximization of its determinant or, equivalently, the minimization of the determinant of the inverse information matrix[17]).

However, for the i-th step,

(i) the levels of $\mathbf{x}^{(i)}$ are directly controllable and so adjustable to the target values required by the experimental plan,

(ii) the levels of $y^{(i-1)}$ can be controlled by running $\mathcal{S}_{i-1}$ and by changing $\mathbf{x}^{(i-1)}$ and $y^{(i-2)}$ according to the model $\hat{f}_{i-1}(\cdot)$.

In particular, the levels of $y^{(i-1)}$ are set through $\hat{f}_{i-1}(\mathbf{x}^{(i-1)}, y^{(i-2)}; \hat{\boldsymbol{\beta}}^{(i-1)}) = \tilde{\boldsymbol{z}}_{i-1}^{\top}\hat{\boldsymbol{\beta}}^{(i-1)}$. If following a design criterion τ_u is the target level for $y^{(i-1)}$ in the u-th experimental run and t_u is the design point

such that $\hat{f}_{i-1}(t_u) = \tau_u$, then the point estimator $\hat{f}_{i-1}(t_u) = t_u^\top \hat{\boldsymbol{\beta}}^{(i-1)}$ is unbiased with variance

$$\eta_u^2 = \sigma_{i-1}^2(1 + t_u^\top(\tilde{\mathbf{Z}}_{i-1}^\top \tilde{\mathbf{Z}}_{i-1})^{-1} t_u), \tag{4.3.1}$$

which can be estimated by replacing σ_{i-1}^2 with the sample variance $\hat{\sigma}_{i-1}^2$ (note that the choice of t_u may not be unique, see Remark 4.3.1). Then, experimental data in the u-th run will be generated by $Y_u^{(i-1)} = \tau_u + e_u$ rather than by τ_u, where e_u is the error in setting $y^{(i-1)}$ in the u-th observation. By assuming e_u normally distributed with zero mean and variance η_u^2 (and also $\mathbb{E}[e_u \cdot e_j] = 0$ for $u \neq j$), the level of $y^{(i-1)}$ becomes a random variable i.e. $Y_u^{(i-1)} \sim N(\tau_u, \eta_u^2)$. Thus, the planned $\mathbf{Z}_i$ is composed of deterministic elements, associated to the input factors $\mathbf{x}^{(i)}$, and random elements associated to $Y^{(i-1)}$, i.e. the generic row of $\mathbf{Z}_i$ is $\mathbf{z}_i^\top = (x_1^{(i)}, x_2^{(i)}, \ldots, Y^{(i-1)}, x_1^{(i)} \cdot x_2^{(i)}, \ldots, x_1^{(i)} \cdot Y^{(i-1)}, \ldots, [x_1^{(i)}]^2, \ldots, [Y^{(i-1)}]^2, \ldots)$.

Remark 4.3.1. *Consider the target level τ_u for $y^{(i-1)}$ in the u-th experimental run. There may be many possible choices of the design points t_u that satisfies $\hat{f}_{i-1}(t_u) = \tau_u$. From a statistical perspective, the choice of t_u can lead to different η_u^2 i.e. to different magnitudes of the prediction variance. A natural choice is then to take the t_u for which the corresponding η_u^2 is the lowest. The selection of the most appropriate t_u could be also driven by practical considerations like cost reasoning or operative purposes.*

Remark 4.3.2. *Note that in phase i (a), the planned $\mathbf{Z}_i$ has random components due to the still unobserved outcomes of $\mathcal{S}_{i-1}$. Since we work under the assumption that intermediate outcomes can be measured, when the experiment on $\mathcal{S}_{i-1}$ in phase i (b) takes place, the realizations of $Y^{(i-1)}$ can be observed. Thus, the experimental conditions that generate $\boldsymbol{y}^{(i)}$ are not random: $\tilde{\mathbf{Z}}_i$ is the matrix $\mathbf{Z}_i$ conditional on the realizations of the $Y^{(i-1)}$s, and $\tilde{\mathbf{Z}}_i$ is not random (see also Donev[12]). Therefore least squares estimates of $\boldsymbol{\beta}^{(i)}$ can can be obtained in phase i (c) by $\hat{\boldsymbol{\beta}}^{(i)} = (\tilde{\mathbf{Z}}_i^\top \tilde{\mathbf{Z}}_i)^{-1} \tilde{\mathbf{Z}}_i^\top \boldsymbol{y}^{(i)}$.*

In *phase i* (a) the design and so the choice of the target levels for both input factors set directly and with error of $\mathcal{S}_i$ should be made according to some optimality considerations. In the presence of error in setting the levels of some input factors, the average D-optimal criterion has been proposed.[11,12] Average D-optimal design is based on minimizing the expected value of the determinant of the random inverse information matrix, i.e.

$$\min \mathbb{E}[\det(\mathbf{Z}_i^\top \mathbf{Z}_i)^{-1}]. \tag{4.3.2}$$

The distribution of $\det(\mathbf{Z}_i^\top \mathbf{Z}_i)^{-1}$ is very complex and, in general, not available in closed-form, so that we compute $\mathbb{E}[\det(\mathbf{Z}_i^\top \mathbf{Z}_i)^{-1}]$ with Monte Carlo approximation. To implement the criterion in (4.3.2), we have extended the Fedorov's exchange algorithm.[17] Starting from an initial design of a given size, at each iteration the algorithm will select a point of the design to be removed and exchanged with a new point from a candidate set. The point is chosen as the one that gives the best improvement in terms of the criterion in (4.3.2). This procedure is iterated until no further exchanges are found to improve the criterion more than a given small threshold (e.g. 10^{-10}). The pseudo code is reported in the Appendix and the R code is available upon request to the first author.

The criterion in (4.3.2) provides a design which is optimal in average. The real experimental conditions recorded in $\tilde{\mathbf{Z}}_i$ are a single realization of $\mathbf{Z}_i$. However, in the next section we report a suitable extension of the fraction of design space (FDS) technique[18] to evaluate the impact of input factors set with error on the prediction capabilities of the design. In the exposition of Section 4.3.1 we refer to the i-th step of the process in *phase i*.

Prediction properties of designs with error in factor levels

One of the most used measures of the prediction properties of a response surface design are the scaled prediction variance (SPV) and the unscaled prediction variance (UPV), which, for a given extended design matrix $\tilde{\mathbf{Z}}_i$ and a point $\tilde{\mathbf{z}}_0$ (expanded to the model space), are given by $SPV(\tilde{\mathbf{z}}_0) = n_i \tilde{\mathbf{z}}_0^\top (\tilde{\mathbf{Z}}_i^\top \tilde{\mathbf{Z}}_i)^{-1} \tilde{\mathbf{z}}_0$ and $UPV(\tilde{\mathbf{z}}_0) = SPV(\tilde{\mathbf{z}}_0)/n_i$ respectively. Zahran et al[18] introduced the fraction of design space technique to both assess the prediction capability of a single design and make comparisons between competing designs. They present a graphical method, the FDS plot, to quantify the fraction of the design space with SPV less than or equal than any SPV values. We have extended such technique to evaluate designs in presence of error in factor levels. Let us assume to be in *phase* i (a). In this case, due to the random $\mathbf{Z}_i$, in any $\tilde{\mathbf{z}}_0$, instead of a single value we get a distribution of $UPV(\tilde{\mathbf{z}}_0)$. As analogous to the UPV, which is sorted and displayed in the FDS plot, we consider $QPV_{1-\alpha}$, defined, for each design point, as the $1-\alpha$ quantile of $UPV(\tilde{\mathbf{z}}_0)$. Hence, for a given fraction of the design space γ, the corresponding $QPV_{1-\alpha}$ value denotes that the $1-\alpha\%$ of the possible designs occurrence have the $\gamma\%$ of the design space with UPV at or below that value (a typical value used in experiments is $\gamma\% = 80\%$).

For numerical comparisons we consider the half-width of the confidence interval for the predicted mean (divided by the expected variability on the tentative model σ_i^2), computed with respect to the $1-\alpha$ quantile of the UPV distribution. In particular, by letting $d_\gamma = q_\gamma(UPV) \cdot t_{0.975,\, n_i - p_i}$, where q_γ is the γ-quantile of the UPV and $t_{0.975, n_i-p_i}$ is the 0.975 quantile of the student-t distribution with $n_i - p_i$ degrees of freedom, each d_γ value indicates that the $\gamma\%$ of the design space is precise enough to predict the mean within $\pm\hat{\sigma}_i d_\gamma$. To accommodate the error in factor levels, following similar reasoning, we define

$$d_{\gamma,\alpha} = q_\gamma(\sqrt{QPV_{1-\alpha}}) \cdot t_{1-0.05/2,\, n_i - p_i},$$

whose value denotes that the $1-\alpha\%$ of the realizations of the designs have the $\gamma\%$ of the design space precise enough to predict the mean within $\pm\hat{\sigma}_i d_{\gamma,\alpha}$.

Notice that the usefulness of $QPV_{1-\alpha}$ and $d_{\gamma,\alpha}$ is not limited to multi-step processes. Also in the case of a single-step process, in which for some reason one or more input factors are set with error, these tools may be adopted to evaluate the prediction capabilities of the experimental designs.

4.3.2 Experimental strategy for multi-step processes

In this section we report the complete procedure that starting from $\mathcal{S}_1$ provides a model for $\mathcal{S}_V$. According to our proposal, the experiments are performed following a particular sequential structure. The experimental strategy is articulated in a sequence of phases in which the same operations are substantially repeated by adding each time one more step as follows (see also Table 4.1).

phase 1
- *(a)* **Design the experiment for step 1.** The experimental plan can be derived in this case according to a classical designs (i.e. a design that does not consider input factors set with error, like D-optimal design, central composite design).
- *(b)* **Perform the experiment on step 1.** Perform the n_1 runs on $\mathcal{S}_1$.
- *(c)* **Analyse the results from step 1.** Estimate the parameters $\boldsymbol{\beta}^{(1)}$ of the response model $f_1(\cdot)$ and the variance of the observations σ_1^2.

phase 2
- *(a)* **Design the experiment for step 2.** Use informations from *phase* 1 *(c)* to find the D-optimal design in average for step 2 as follows. Fix a range $[y_L^{(1)}, y_U^{(1)}]$ of values to explore

in input for $\mathcal{S}_2$, choose target levels within this interval (e.g. low, mid and $high$ level) and compute the corresponding variance of the error in setting $y^{(1)}$ to the desired levels (e.g. $\eta^2_{low}, \eta^2_{mid}, \eta^2_{high}$). Implement D-optimality in average to find the optimal design for $\mathcal{S}_2$ (with n_2 runs); compute $\mathbf{Z}_2$.

(b) **Perform the experiments on step 1 and step 2.** Perform the n_2 runs on $\mathcal{S}_1$ and $\mathcal{S}_2$. The input factors levels of $\mathcal{S}_1$ are set to obtain the outputs of $\mathcal{S}_1$ close to the target levels computed as optimal for $\mathcal{S}_2$. At the end of $\mathcal{S}_1$ record the realizations of $Y_u^{(1)}$, $\forall u = 1, \ldots, n_2$ (which are different from those obtained at the end of the *phase* 1), so that the real experimental conditions (i.e. observed design matrix) that generate the $\boldsymbol{y}^{(2)}$ are known; compute $\tilde{\mathbf{Z}}_2$.

(c) **Analyse the results from step 2.** Estimate the parameters $\boldsymbol{\beta}^{(2)}$ of the response model $f_2(\cdot)$ and the variance of the observations σ_2^2.

⋮ ⋮

phase i *(a)* **Design the experiment for step i.** Use informations from *phase* $i-1$ *(c)* and implement D-optimality in average to find the optimal design for $\mathcal{S}_i$ (with n_i runs); compute $\mathbf{Z}_i$.

(b) **Perform the experiments on step 1, step 2, ... step i.** Perform n_i runs on $\mathcal{S}_1$, $\mathcal{S}_2$, ..., $\mathcal{S}_i$. For step $\mathcal{S}_{i-1}$ factors levels are set to obtain the output of $\mathcal{S}_{i-1}$ close to the target levels computed as optimal for $\mathcal{S}_i$ and factors' levels for $\mathcal{S}_{i-2}, \ldots, \mathcal{S}_1$ are then set accordingly. At the end of $\mathcal{S}_{i-1}$ record the realizations of $Y_u^{(i-1)}$, $\forall u = 1, \ldots, n_i$ (which are different from those obtained at the end of *phase* $i-1$) so that the real experimental conditions (i.e. observed design matrix) that generate the $\boldsymbol{y}^{(i)}$ are known; compute $\tilde{\mathbf{Z}}_i$.

(c) **Analyse the results from step i.** Estimate the parameters $\boldsymbol{\beta}^{(i)}$ of the response model $f_i(\cdot)$ and the variance of the observations σ_i^2.

⋮ ⋮

phase V *(a)* **Design the experiment for step V.** Use informations from *phase* $V-1$ *(c)* and implement D-optimality in average to find the optimal design for $\mathcal{S}_V$ (with n_V runs); compute $\mathbf{Z}_V$.

(b) **Perform the experiments on step 1, step 2, ..., step V.** Perform n_V runs on all the process steps and proceed similarly to *phase* i *(b)*; compute $\tilde{\mathbf{Z}}_V$.

(c) **Analyse the results from step V.** Estimate the parameters $\boldsymbol{\beta}^{(V)}$ of the response model $f_V(\cdot)$ and the variance of the observations σ_V^2.

The complete procedure requires $n_1 + n_2 + \cdots + n_V$ runs where the n_1 runs are performed only on step 1 (*phase* 1), the n_2 runs are performed on step 1 and step 2 (*phase* 2) and so on. At the end of *phase* V we obtain V fitted models describing each step of the process. An example for a case study is reported in Section 4.5 for $V = 3$.

In the next example we show the potential advantages of our proposal in terms of required experimental effort.

Example 4.3.2. *In the case of a three-step process with three input factors per step and a full tentative quadratic model with intercept for $y^{(3)}$, we report in Table 4.2 the number of parameters to be estimated in the case the experimenter considers the process as a single-stage (single-stage DoE) and in*

the proposed multi-step approach (multi-step DoE). We compared the two procedures as the number of controllable input factors per stage (F) increases.

Clearly a larger number of parameters to estimate requires a higher number of runs. However, note that the single-stage and the multi-step procedures cannot be directly compared in terms of the number of runs. This number depends on the design criterion adopted and on the required prediction properties of the experimental design. The main advantage of our set-up is related to the number of steps involved in each experiment: according to the single-stage DoE, each of the planned runs involve all the process' steps whereas adopting our proposal the number of runs carried out from the first up to the last step is limited, inducing, in general, a gain in terms of experimental resources. This example also highlights how the saving of experimental resources brought by the multi-step DoE increases as the number of controllable input factors per stage increases.

Table 4.1: Strategy of experimentation for V-step processes.

Ph.	Experiment	Input	How is set	Design criterion	Output	For next step
1	n_1 runs on $\mathcal{S}_1$	$\mathbf{x}^{(1)}$	directly	Classic DoE	observe $\mathbf{y}^{(1)}$, $\tilde{\mathbf{Z}}_1$, compute $\hat{f}_1(\cdot)$	$[\mathrm{y}_L^{(1)}, \mathrm{y}_U^{(1)}]$ η_u^2
2	n_2 runs on $\mathcal{S}_1$ and $\mathcal{S}_2$	$\mathbf{x}^{(2)}$ $\mathrm{y}^{(1)}$	directly by changing $\mathbf{x}^{(1)}$ through $\hat{f}_1(\cdot)$	Ave D-opt ($\mathbf{Z}_2$)	observe $\mathbf{y}^{(2)}$, $\tilde{\mathbf{Z}}_2$; compute $\hat{f}_2(\cdot)$	$[\mathrm{y}_L^{(2)}, \mathrm{y}_U^{(2)}]$ η_u^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	n_i runs on $\mathcal{S}_1, \ldots, \mathcal{S}_i$	$\mathbf{x}^{(i)}$ $\mathrm{y}^{(i-1)}$	directly by changing $\mathbf{x}^{(i-1)}$ and $\mathrm{y}^{(i-2)}$ through $\hat{f}_{i-1}(\cdot)$	Ave D-opt ($\mathbf{Z}_i$)	observe $\mathbf{y}^{(i)}$, $\tilde{\mathbf{Z}}_i$; compute $\hat{f}_i(\cdot)$	$[\mathrm{y}_L^{(i)}, \mathrm{y}_U^{(i)}]$ η_u^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
V	n_V runs on $\mathcal{S}_1, \ldots, \mathcal{S}_V$	$\mathbf{x}^{(V)}$ $\mathrm{y}^{(V-1)}$	directly by changing $\mathbf{x}^{(V-1)}$ and $\mathrm{y}^{(V-2)}$ through $\hat{f}_{V-1}(\cdot)$	Ave D-opt ($\mathbf{Z}_V$)	observe $\mathbf{y}^{(V)}$, $\tilde{\mathbf{Z}}_V$; compute $\hat{f}_V(\cdot)$	$[\mathrm{y}_L^{(V)}, \mathrm{y}_U^{(V)}]$ η_u^2

Table 4.2: Number of parameters to be estimated: single-stage DoE vs. multi-step DoE ($\mathcal{S}_2$ and $\mathcal{S}_3$ involve an additional input factor).

Controlled Inputs			Number of parameters	
$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	single-stage DoE	multi-step DoE
2	2	2	28	26
3	3	3	55	40
4	4	4	91	57
F	F	F	$(9F^2 + 9F + 2)/2$	$(3F^2 + 13F + 14)/2$

4.4 Manufacturing design space definition for multi-step processes

Once the procedure of Section 4.3.2 has been performed, V fitted models, one for each outcome, are available. These models are used to define a set of interconnected acceptable ranges expressed in terms of the directly controlled input factors of the process. We will refer to *multi-step* manufacturing design space (denoted by DS) to intend the manufacturing design space of the entire process and to *individual* manufacturing design space as this space for a single step (denoted by DS_i for $\mathcal{S}_i$).

Starting from the last step, let us assume that the quality target for the final outcome of the process requires $y^{(V)} \in R$. Then, from $\hat{f}_{V-1}(\cdot)$, the corresponding multidimensional combination of $\mathbf{x}^{(V)}$ and $y^{(V-1)}$ such that $y^{(V)} \in R$ is derived. This region is usually restricted by a confidence/prediction interval on the fitted response, giving an acceptable (more robust) smaller region for $\mathbf{x}^{(V)}$ and $y^{(V-1)}$. Then, the determined range for $y^{(V-1)}$ becomes the quality target used to derive the individual manufacturing design space for $\mathcal{S}_{V-1}$, so that DS_{V-1} is defined from $\hat{f}_{V-1}(\cdot)$ and is expressed in terms of $\mathbf{x}^{(V-1)}$ and $y^{(V-2)}$.
The same reasoning is iterated until the manufacturing design space for $y^{(1)}$, which will be expressed in terms of $\mathbf{x}^{(1)}$, is derived. The combination of the acceptable ranges of the directly controllable input factors gives the multi-step manufacturing design space

$$DS = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(V)} \text{ such that } y^{(V)} \in R\}.$$

A similar procedure has been adopted in the work of Eon-Duval et al[19] but the authors do not address two fundamental issue. First, they do not formally provide a procedure to model a multi-step process and second they do not take into account the multi-step structure of the process in the designing of experiments.

Remark 4.4.1. *Two potential presentations of the manufacturing design space are reported in the guidelines.[2] i) It can be defined by a non-linear combination of inputs' ranges that makes $y^{(V)} \in R$. In this case, the manufacturing design space is explained by mathematical equations describing relationships between inputs that lead to successful outputs. While this approach allows the maximum operative range to achieve the required quality standard, it makes the manufacturing design space be a complex set. Otherwise ii) the manufacturing design space can be defined as a smaller region, based on a linear combination of input factors. Even if this approach is more limiting, it is often preferred in the applications due to operational simplicity. For this reason, in this paper we adopt definition in ii). In principle the experimenter could select any sub-region based on a linear combination of inputs; often this choice is driven by scientific and practical considerations. For example in Figure 4.8, the manufacturing design space as in i) is the yellow region whereas the manufacturing design space as in ii) is the red-delimited rectangle.*

4.5 Case study: Manufacturing design space for a three step biochemical process

In this section we implement our proposal to an illustrative case-study. We considered a biochemical process commonly used in pharmaceutical industry to produce and purify recombinant proteins expressed by *E. Coli*. The process consists of three separate steps (Figure 4.4):

- **Fermentation**, where the *E. Coli* culture is grown and the recombinant protein is expressed in the bacterial cells;

- **Capture separation**, which is the first purification step where the target recombinant protein is captured in the column and a first portion of impurities is removed (DNA fragments, HCP, endotoxin etc);

- **Hydrophobic separation**, which is the last purification step, where the hydrophobic interactions are used to separate our target recombinant proteins from other residual proteins.

For the Fermentation step, four input factors were identified as potentially critical process parameters and included in the study: trace element concentration in the fermentation media (TE), optical density of induction (OD_{ind}), pH of the fermentation media (pH), which is maintained fixed during the whole fermentation and the duration of the expression phase (DE), i.e. the time between the induction point and the end of fermentation. Since the protein purity from the Fermentation is expected to impact the performance of the purification steps, the output selected to be included as input factor for the Capture separation step is the protein purity in the capture load material (denoted by $Pur1$). In the Capture step, instead, we considered three process parameters and $Pur1$ as input factors. The three capture process parameters are the following: pH level of the wash ($pH.W$), the molarity and the pH of the load material ($Mol.L$ and $pH.L$ respectively). Also in this case, the purity of the target protein ($Pur2$) has been identified as the output of the Capture to be included as input in the Hydrophobic separation. In this latter step, three process parameters have been identified as potentially critical for the final purity: the pH of the material loaded in the column ($pH.C$), the protein concentration and the conductivity of the loaded material ($dens$ and $cond$ respectively). Finally, we include the purity of the incoming material ($Pur2$) as additional input factor.

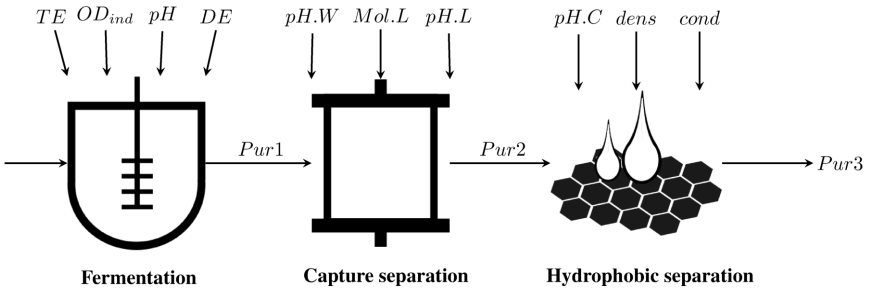


Figure 4.4: Three step process of expression and purification of a recombinant protein.

Other steps, like the final filtration or the centrifugation at the end of the fermentation steps were not considered in our study, however the approach can be easily extended to include them. The aim of the study is to determine the multi-step manufacturing design space corresponding, in this case, to the region of input factors/process parameters which can guarantee a final purity of the target protein ($Pur3$) above 88%. More specifically, in the following, the Protein purity is defined as the relative amount of the recombinant protein of interest with respect to the total amount of all the components present in the material, i.e. host cell proteins, fragmented recombinant protein, aggregated forms. The methods used to determine the protein purity are: SDS-PAGE after Fermentation and Reverse Phase HPLC after Capture separation and after Hydrophobic separation.

In Section 4.5.1 we report the design and analysis of the experiments for the three steps and in Section 4.5.2 we derive the multi-step manufacturing design space.

4.5.1 Design and Analysis of three step process

STEP 1: Fermentation For each input factor of the Fermentation five levels were identified within the ranges reported in Table 4.3a.

Table 4.3: Step 1: Fermentation.

(a) Input Factors of Step 1 and ranges.

$\mathbf{x}^{(1)}$	*Input*	*Range*
$x_1^{(1)}$	TE	$[1.3, 1.7]$
$x_2^{(1)}$	OD_{ind}	$[3.0, 7.0]$
$x_3^{(1)}$	pH	$[6.2, 6.8]$
$x_4^{(1)}$	DE	$[5.0, 7.0]$

(b) Output of Step 1, coded level in brackets.

Output	*Input*						
Level	$Pur1$	OD_{ind}		pH		DE		η_u^2
low	13.32	7.00	$(+1.00)$	6.80	$(+1.00)$	5.00	(-1.00)	0.71^2
medium	15.54	6.70	$(+0.85)$	6.41	(-0.30)	6.35	$(+0.35)$	0.68^2
high	17.75	4.52	(-0.24)	6.20	(-1.00)	7.00	$(+1.00)$	0.70^2

To determine the relationship between $Pur1$ and the four fermentation parameters, a modified face-centred central composite design with 34 runs was defined (3 extra runs were added as confirmation runs). From the results reported in Table 4.6, the fitted model for $Pur1$ (in terms of the coded unit) is

$$\hat{Pur1} = 16.32 - 0.52 \cdot OD_{ind} - 0.44 \cdot pH + 0.94 \cdot DE - 1.10 \cdot OD_{ind}^2, \tag{4.5.1}$$

with $\hat{\sigma}_1^2 = 0.64^2$ and $R^2 = 0.74$ (predictive $R^2 = 0.63$). We used the model in (4.5.1) to determine the combination of OD_{ind}, pH, and DE such that $Pur1$ reaches its minimum value, 13.32, (for $OD_{ind} = 1, pH = 1, DE = -1$) and maximum, 17.75 (for $OD_{ind} = -0.24, pH = -1, DE = 1$). Within this interval we selected three target levels for $Pur1$, low, medium and high, in which we compute η_u^2, as in (4.3.1). These informations are summarized in Table 4.3b.

STEP 2: Capture separation In order to fit a quadratic tentative model with two factor interactions effects, for each of the four factors of this step (see Table 4.4a), we consider three levels, $[-1, 0, 1]$. As far as $Pur1$ is concerned, it is set by changing Fermentations input factors through the model in (4.5.1). According to the criterion in (4.3.2), we generated D-optimal designs with different number of runs (from 29 to 33), with $\eta_{+1}^2 = 0.316^2$, $\eta_0^2 = 0.305^2$ and $\eta_{-1}^2 = 0.321^2$. In this case we select the design with $d_{0.8,0.95} = 1.49$ which corresponds to the design with 33 runs. In Figure 4.5 we report the corresponding FDS plot. The black curve is related to the UPV values obtained in the ideal case of no error in setting $Pur1$, while the grey curve refers $QPV_{0.95}$ (Section 4.3.1). For the chosen design, the UPV value for a fraction of the design space $\gamma = 0.8$ (grey dot) indicates that the 95% of the possible realizations of this design have the 80% of the design space with $UPV \leq 0.49$. Whereas, if ideally we could

Table 4.4: Step 2: Capture separation.

(a) Input factors of Step 2 and ranges.

	Input	*Range*	
$x_1^{(2)}$	$pH.W$	$[6.80, 7.20]$	set directly
$x_2^{(2)}$	$Mol.L$	$[70.00, 80.00]$	set directly
$x_3^{(2)}$	$pH.L$	$[6.80, 7.20]$	set directly
$y^{(1)}$	$Pur1$	$[13.32, 17.75]$	set indirectly

(b) Output of Step 2, coded levels in brackets.

Output		*Input*				
Level	$Pur2$	$Mol.L$		$Pur1$		η_u^2
low	72.90	80.00	$(+1.00)$	12.27	(-1.00)	3.30^2
medium	84.48	78.00	$(+0.06)$	15.14	$(+0.05)$	3.19^2
high	96.06	74.00	(-0.20)	18.31	$(+1.00)$	3.39^2

run the experiment by setting precisely all input factors, the 80% of the design space would have had $UPV \leq 0.42$ (black dot).

The generated design provides target levels for $pH.W$, $Mol.L$, $pH.L$ and $Pur1$. In order to proceed with the experiment, we rearrange the experimental plan in terms of the directly controllable input factors: target levels for $Pur1$ become target levels for OD_{ind}, pH and DE. The 33 runs are independently performed on the Fermentation and the Capture steps. Clearly, the $Pur1$ values achieved in the Fermentations will be different from the target values due to the model error (e.g. 13.36 is just prediction given by the model, experimental values are expected within $13.36 \pm e_u$). From the results (reported in Table 4.7), we estimate the following model (in terms of coded unit) for the $Pur2$ outcome:

$$\hat{Pur}2 = 87.80 - 1.91 \cdot Mol.L + 8.08 \cdot Pur1 - 4.91 \cdot Mol.L^2, \tag{4.5.2}$$

with $\hat{\sigma}_2^2 = 3.1^2$ and $R^2 = 0.80$ (predictive $R^2 = 0.75$). Following the same procedure of the previous step, by using the model in (4.5.2), we derived the combination of $Mol.L$ and $Pur1$ such that $Pur2$ reaches its minimum and maximum (see Table 4.4b). Within this range we selected three target level, low, medium and high for $Pur2$ and we compute the corresponding variance of the error for a future prediction by (4.3.1).

STEP 3: Hydrophobic separation The input factors of the Hydrophobic separation step are reported in Table 4.5 together with the selected ranges. As regards $Pur2$, it can be controlled by tuning the inputs of the Capture and Fermentation steps according to the models in (4.5.1) and (4.5.2).

Assuming a tentative model with quadratic terms and two factor interactions, we proceeded as for step 2: we rescaled the input factors to the range $[-1; 1]$ and consequently the variance of the error in setting $Pur2$ (obtaining $\eta_{+1}^2 = 0.29^2$, $\eta_0^2 = 0.27^2$ and $\eta_{-1}^2 = 0.28^2$). Among D-optimal designs in average, with 30-33 runs, by comparing FDS plots and the expected half-width values, we selected the experimental plan with 33 runs. For this plan, the expected half-width of the mean predicted from the model is below 1.54 in 80% of the design space. For the same proportion of the design space the 95%

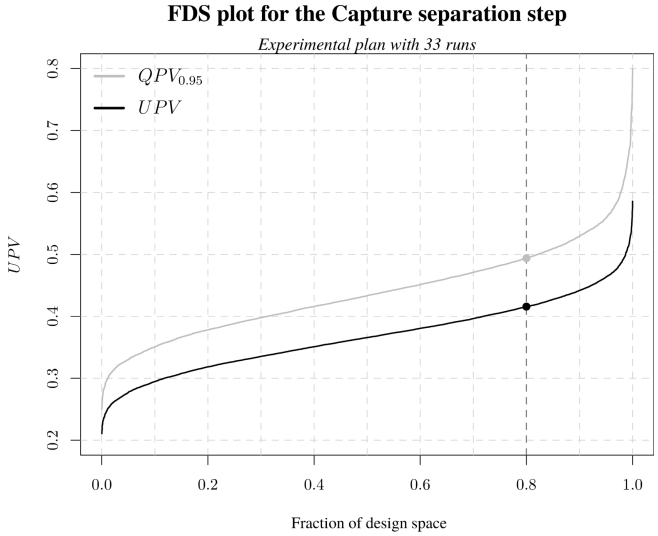


Figure 4.5: Fraction of Design Space plot for the 33 runs design of the Capture separation step.

of the experimental plan will have $UPV \leq 0.54$, whereas $UPV \leq 0.46$ if the experiment would have been run with no error (FDS plot in Figure 4.6).

The experimental plan is then rearranged in order to express the target values for $Pur2$ as target values for $Mol.L$ and $Pur1$. In turn, target levels for $Pur1$ are translated into target levels for OD_{ind}, pH and DE. The values obtained experimentally for $Pur1$ and $Pur2$ have been measured and recorded (data reported in Table 4.8) and have been used to fit the following model (in terms of coded units) for $Pur3$,

$$\hat{Pur3} = 88.81 + 0.77 \cdot pH.C + 1.58 \cdot dens + 4.38 \cdot Pur2 + 3.75 \cdot dens^2 + 1.76 \cdot pH.C \cdot Pur2, \quad (4.5.3)$$

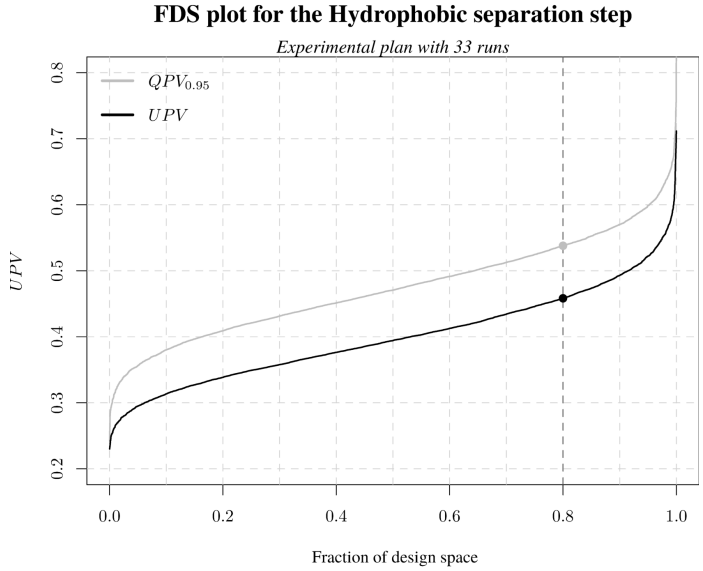
with $\hat{\sigma}_3^2 = 2.01^2$ and $R^2 = 0.80$ (predictive $R^2 = 0.69$).

Since $Pur3$ is a CQA of the final outcome we now proceed to the manufacturing design space definition. For this three-step process, three interconnected models are available, one for $Pur1$, one for $Pur2$ and one for $Pur3$.

Table 4.5: Step 3: Hydrophobic Separation. Input factors and ranges.

Input		*Range*	
$x_1^{(3)}$	$pH.C$	$[6.40, 7.00]$	set directly
$x_2^{(3)}$	$dens$	$[2.00, 8.00]$	set directly
$x_3^{(3)}$	$cond$	$[80.00, 110.00]$	set directly
$y^{(2)}$	$Pur2$	$[92.90, 96.06]$	set indirectly

Figure 4.6: Fraction of Design Space plot for the 33 runs design of the Hydrophobic separation step.



4.5.2 Multi-step manufacturing design space determination

In this study, the manufacturing design space is the region of the input factors' space which can consistently guarantee a final protein purity at or above 88% (often called specification limits in this kind of studies). Therefore, by the model in (4.5.3) and by graphical optimization we selected the region of $pH.C$, $dens$ and $Pur2$ such that $\hat{Pur}3 \geq 88\%$. To take into account the uncertainty on the model predictions, this region is typically reduced following operative considerations. More specifically we considered the region restricted by the 95% one-sided prediction interval (P.I.) on a single future observation of $Pur3$, as shown in Figure 4.7, but other intervals can be considered as well, like e.g. tolerance intervals[20]. In Figure 4.7 the solid black curve is the model prediction such that $\hat{Pur}3 = 88\%$, while the dashed black curve is the corresponding bound given by the prediction interval equal to 88% We highlight in yellow the region of the input factors such that specification limits are satisfied. In this case, within this region, we identified a suitable sub-region of operating conditions (see Remark 4.4.1) - the red-bordered rectangles - so that the manufacturing design space is defined by the linear combination of the significant input factors for the Hydrophobic step and the outcome of the Capture step as follows,

$$DS_3 = \{dens \in [6.87, 8.00], pH.C \in [6.82, 7.00]\, Pur2 \geq 84.64\%\}.$$

Now, $Pur2 \geq 84.64\%$ becomes the specification limit for the Capture step: the resulting manufacturing design space is identified by the suitable region delimited by the one-sided 95% prediction interval on $Pur2$, as shown in Figure 4.8 and it can be defined as

$$DS_2 = \{Mol.L \in [72.90, 75.10],\ Pur1 \geq 16.02\%\}.$$

The same procedure is repeated for the Fermentation step (see Figure 4.9), whose manufacturing design space is given by

$$DS_1 = \{OD \in [4.05, 5.05],\ pH \in [6.63, 7.00],\ DE \in [3.00, 6.05]\}.$$

The multi-step manufacturing design space is then derived by combining the individual ones of each process step, obtaining

$$DS = \{dens \in [6.87, 8.00], pH.C \in [6.82, 7.00], Mol \in [72.90, 75.10],$$
$$OD_{ind} \in [4.05, 5.05],\ pH \in [6.63, 7.00],\ DE \in [3.00, 6.05]\}.$$

Individual manufacturing design space for step 3

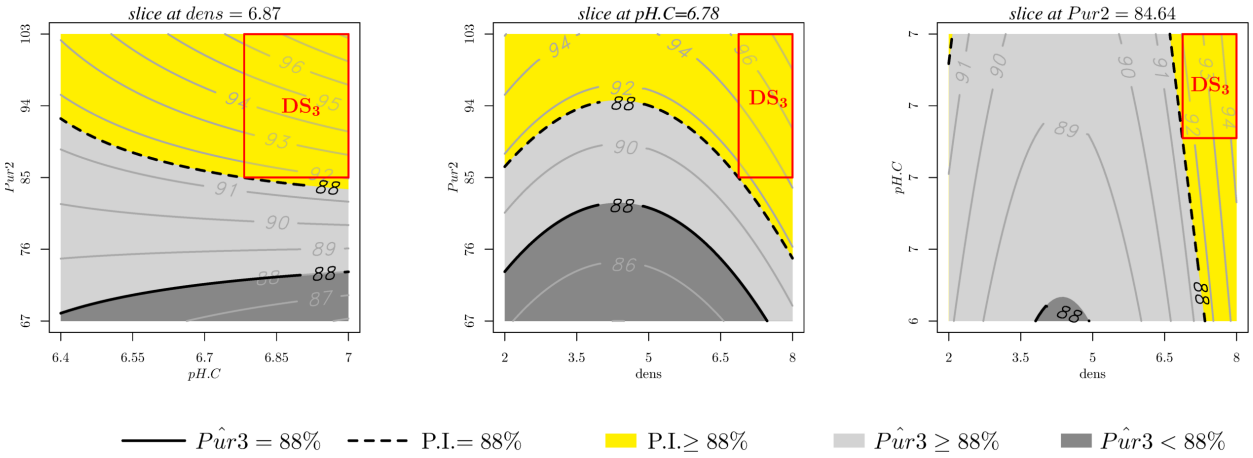


Figure 4.7: Manufacturing design space for the Hydrophobic separation step.

Individual manufacturing design space for step 2

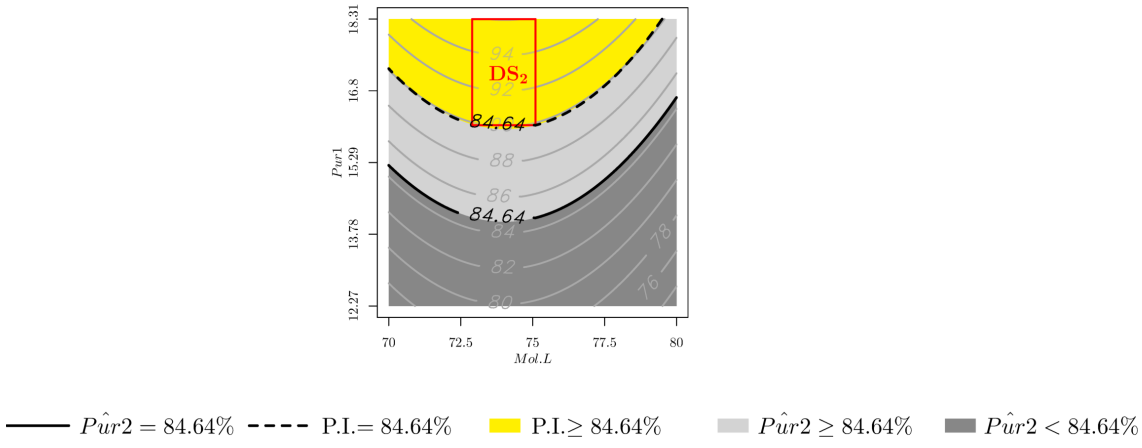


Figure 4.8: Manufacturing design space for the Capture separation step.

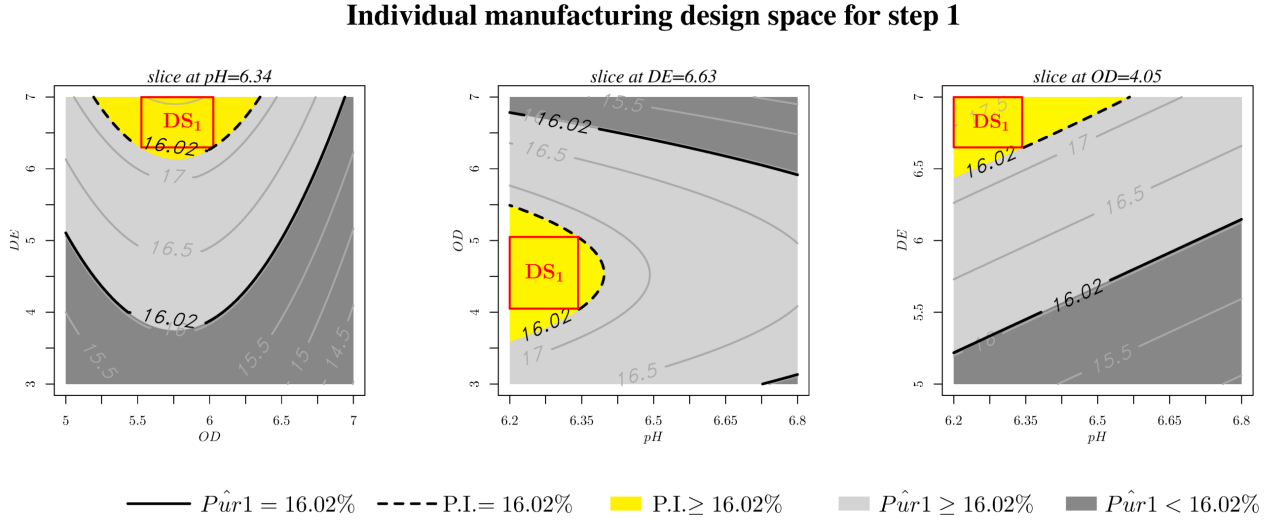


Figure 4.9: Manufacturing design space for the Fermentation step.

4.6 Discussion and conclusions

In this paper we provide a statistical approach to define the manufacturing design space of a process composed of multiple steps. Since the multi-step manufacturing design space does consider the interactions among subsequent steps, it leads to a very good process understanding, it guarantees quality and safety of products, faster and more consistent product development. In the context of pharmaceutical industry, these aspects consistently guarantee a drug product with the desired properties leading to benefits for patients.[2] Moreover, the multi-step manufacturing design space increases both manufacturing flexibility and process robustness, which are crucial for reducing costs and batch discarding. Indeed, in the first place, working within the manufacturing design space is not considered as a change, whereas movements outside would normally initiate a regulatory post approval. In the second place, it is well-known that a manufacturing design space that spans the entire process can increase the operational flexibility.[2] The experimental effort required by the procedure, although moderate, is still affordable by scientists, making our proposal a valid compromise between process knowledge and experimental resources. Despite our motivating set-up is related to pharmaceutical processes, our methodology is general and it can be considered for multi-step experimentation in various industrial fields.

In the three-steps Fermentation and Purification process considered, three experiments for three steps are required: overall, 10 input factors were involved in the study. If we would have approached this process as a big single-stage, the experiment would have had 10 input factors, resulting in 66 parameters to estimate (in the case of a full-quadratic model for the final outcome). In principle, the D-optimal design with 100 runs could have also been a possible solution to design the experiment. However, by considering the process as a big single-stage, each of the 100 runs would have involved all the three process' steps. According to our proposal only the last set of experiments, namely 33 runs, has to be performed on the whole process. Therefore, the multi-step design strategy requires a limited number of runs for the experiments carried out from the first up to the last step.

The multi-step approach relies on the assumption that the outcomes of the intermediate steps can be

measured to derive intermediate models. The quality of these models depends on many elements and, of course, on the features of the step itself. The impact of an input factor set with error on the intermediate models, and so on the model for the final outcome, strongly depends on how many input factors has the step in view of one factor set with error and on the effect (linear, quadratic, ...) that the input factor set with error is expected to have on the outcome i.e. on the tentative model.[12] Finally, the quality of an intermediate model is also affected by the observed realization of the planned design matrix. We would recommend a case-by-case simulation study to evaluate this impact. We would also advise to evaluate the prediction properties of the design by taking into account the possible deviations from the target levels that could occur in setting the input factor with error, as suggested in Section 4.3.1. This prevents an overestimation of the prediction properties of the design.

Notice that the design strategy proposed here can be also applied to contexts in which, at the end one step (or more) it is not possible to stop and observe the outcome. In these cases, such step can be simply merged with the subsequent one and considered as a single-stage in the design of experiment and in the manufacturing design space.

The design strategy presented in this paper is based on planning one experimental design for each step including, as additional input factor, the output of the previous one. This setting requires the selection of a single outcome/CQA of the previous step that the experimenter is interested to study, in interactions with the input factors of the current one. This assumption is essential in practice: an input - which is actually an output of the previous step, say y_A - is set by running the previous step and by adjusting its inputs to achieve the desired values for y_A given in the experimental plan. In the case the experimental design would include two set indirectly inputs - say y_A and y_B - as they are controlled by changing the same inputs of the previous step, they cannot be, in general, set independently to the desired level. Essentially, if the levels of y_A are appropriately changed in the experiment, the levels of y_B can only be observed. Thus, at the end of each step, multiple CQAs can be still measured and monitored, but the experimental plan will only be optimal for the selected one. In many industrial fields, scientific knowledge and discussion with process' experts should help to identify the appropriate CQA to be included as an input for the next step. Otherwise, the two outcomes could be treated separately into two different experiments but the procedure would employ a quite large amount of experimental resources. Further research is surely requested on this point.

In addition, the multi-step framework provides hints for future research in many directions. We focused on processes in which the behaviour of the outcome of each step can be well approximated by linear models in the parameters and we consider only the interactions among subsequent steps. This framework encompasses several practical situations, however, the procedure could be appropriately extended, with increasing complexity, to relax these assumptions. Moreover, since first principle models may eventually exist just for one of the steps that makes up the process, one of the main direction for future studies is the extension of our proposal to accommodate both mechanistic and empirical models. As regards the design criterion, in our framework it should take into account that one input factors is not set directly. As a starting point we adopted the average D-optimality proposed in the literature but this problem offers insights for further research to derive alternatives design criteria suitable for our set-up. Future developments will be also dedicated to the optimization of the total number of experimental runs to be performed on each step of the process.

Acknowledgements

We wish to thank Alessandro Pieri and Eva Grassi of GSK TRD Drug Substance team, who provided a very useful support to this work.

Disclosure statement

The authors have declared the following conflicts of interest: Rosamarie Frieri is a PhD student at the University of Bologna and participates in a post graduate studentship program at GSK, with Marilena Paludi as her supervisor. Marilena Paludi and Marco Mariti are employees of the GSK group of companies.

4.7 Appendix

4.7.1 Algorithm: D-optimal design in average

In exchange algorithms, starting from an initial design, each design point is considered for exchange with each of the point of a candidate list. The selected pair of points to exchange is the one which gives the best improvement in terms of the chosen design criterion (in this case is the pair which most decreases the expected determinant). This procedure is iterated until no further improvement in the criterion can be obtained by a pairwise exchange.[21] The algorithm finds local optimum so the procedure is usually repeated for multiple initial designs.

Notation:
design points: $\tilde{\mathbf{z}}_i$ for $i = 1, \ldots, n$
candidate points: $\tilde{\mathbf{z}}_j$ for $i = 1, \ldots, N$
$\delta(\tilde{\mathbf{z}}_i) = \tilde{\mathbf{z}}_i^\top (\tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}})^{-1} \tilde{\mathbf{z}}_i$
$\delta^2(\tilde{\mathbf{z}}_i, \tilde{\mathbf{z}}_j) = \mathbf{z}_i^\top (\tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}})^{-1} \tilde{\mathbf{z}}_j$
$\Delta_{ij} = \delta(\tilde{\mathbf{z}}_j) - \delta(\tilde{\mathbf{z}}_i) - \delta(\tilde{\mathbf{z}}_i)\delta(\tilde{\mathbf{z}}_j) + \delta^2(\tilde{\mathbf{z}}_i, \tilde{\mathbf{z}}_j)$
$D_{ij} = [\det(\tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}}) + (1 + \Delta_{ij})]^{-1}$, namely the change in $\det(\tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}})^{-1}$ which would be obtained by switching $\tilde{\mathbf{z}}_i$ with $\tilde{\mathbf{z}}_j$
t: small threshold (e.g. 10^{-10})
$nsim$: number of Monte Carlo simulations[1]. To select $nsim$ we run a series of preliminary simulation studies in which we check the distributions of D_{ij} and their expectations for increasing $nsim$. In our case study example $nsim$=1000 (as also in Donev[12]) was enough to provide reasonable stable results.

generate a (random) start design with n points
while $-(D_{new} - D_{old}) \leq t$ {
  for $1, \ldots, nsim$ {
    generate a realization* of $\mathbf{Z}_{old} \rightarrow \tilde{\mathbf{Z}}_{old}$
    generate a realization* of the N candidate points
    compute $D_{old} = \det(\tilde{\mathbf{Z}}_{old}^{\top}\tilde{\mathbf{Z}}_{old})^{-1}$
      for i in $1, \ldots, n$ {
        for j in $1, \ldots, N$ {
          compute Δ_{ij}
          compute D_{ij}
      } }
  }
  compute $\mathbb{E}[D_{ij}]$
  select i and j minimizing $\mathbb{E}[D_{ij}]$ (if more than one best exchange, select one randomly)
  exchange $\mathbf{z}_i$ with $\mathbf{z}_j$ so that $\mathbf{Z}_{old} \longrightarrow \mathbf{Z}_{new}$
  update the determinant $D_{new} = \det(\mathbf{Z}_{new}^{\top}\mathbf{Z}_{new})^{-1}$
}

*a random draw from normal distribution (with zero mean and variance in (4.3.1)) is added to each target level of the input factor set with error in the main effect column of $\mathbf{Z}_{old}$; quadratic and two factor interactions effects are then computed accordingly.

4.7.2 Experimental Data

Table 4.6: Experimental results from the Fermentation step.

run	$Pur1$	TE	OD_{ind}	pH	DE
1	16.20	1.00	-1.00	1.00	1.00
2	15.80	0.00	0.00	0.00	0.00
3	16.10	1.00	1.00	-1.00	1.00
4	16.80	0.00	0.00	0.00	0.00
5	14.80	1.00	-1.00	-1.00	-1.00
6	17.50	0.00	0.00	0.00	0.00
7	15.10	-0.50	-0.50	-0.67	-0.50
8	16.60	-1.00	-1.00	-1.00	1.00
9	12.55	1.00	1.00	1.00	-1.00
10	15.45	-1.00	1.00	1.00	1.00
11	13.10	-1.00	1.00	-1.00	-1.00
12	14.50	-1.00	-1.00	1.00	-1.00
13	16.10	-1.00	-1.00	-1.00	-1.00
14	16.70	0.00	0.00	0.00	0.00
15	15.00	1.00	1.00	1.00	1.00
16	15.30	1.00	1.00	1.00	1.00
17	16.03	0.50	0.50	0.67	0.50
18	16.70	0.00	0.00	0.00	0.00
19	17.80	1.00	-1.00	-1.00	1.00
20	15.30	1.00	1.00	-1.00	-1.00
21	13.70	-1.00	1.00	1.00	-1.00
22	15.40	-1.00	-1.00	1.00	1.00
23	14.90	1.00	-1.00	1.00	-1.00
24	17.20	-1.00	0.00	0.00	0.00
25	15.00	0.00	0.00	1.00	0.00
26	16.00	0.00	0.00	-1.00	0.00
27	16.50	0.00	0.00	0.00	0.00
28	16.10	0.00	0.00	0.00	0.00
29	17.60	0.00	0.00	0.00	1.00
30	16.20	-0.50	0.50	-1.00	0.50
31	16.50	1.00	0.00	0.00	0.00
32	14.90	0.00	1.00	0.00	0.00
33	14.80	0.00	0.00	0.00	-1.00
34	15.70	0.00	-1.00	0.00	0.00

Table 4.7: Experimental results from the Capture separation step.

run	$Pur2$	$pH.W$	$Mol.L$	$pH.L$	$Pur1$
1	74.86	-1.00	1.00	1.00	-0.35
2	89.50	-1.00	1.00	1.00	0.67
3	80.65	-1.00	0.00	-1.00	-0.86
4	90.51	1.00	0.00	-1.00	0.61
5	87.95	0.00	0.00	1.00	-0.13
6	96.20	0.00	-1.00	-1.00	0.87
7	80.42	1.00	-1.00	-1.00	-0.60
8	91.94	1.00	-1.00	0.00	0.62
9	89.72	-1.00	-1.00	-1.00	0.88
10	77.44	0.00	1.00	-1.00	-0.11
11	79.59	1.00	-1.00	1.00	-0.11
12	90.10	-1.00	-1.00	0.00	0.80
13	91.95	-1.00	1.00	-1.00	0.64
14	95.41	0.00	0.00	0.00	0.64
15	75.04	1.00	1.00	1.00	0.09
16	85.45	1.00	0.00	0.00	-0.79
17	85.11	-1.00	-1.00	1.00	-0.05
18	80.47	0.00	-1.00	1.00	-1.00
19	75.30	1.00	1.00	-1.00	-0.52
20	78.76	1.00	0.00	1.00	-0.81
21	88.63	1.00	1.00	1.00	0.66
22	82.42	1.00	-1.00	-1.00	-0.07
23	90.66	1.00	-1.00	1.00	0.93
24	88.27	1.00	1.00	-1.00	0.93
25	78.79	-1.00	1.00	1.00	-0.89
26	81.14	-1.00	1.00	0.00	-0.14
27	77.55	0.00	-1.00	0.00	-0.93
28	89.40	-1.00	0.00	1.00	0.54
29	80.41	-1.00	1.00	-1.00	-0.18
30	82.22	-1.00	-1.00	1.00	-0.46
31	91.46	0.00	-1.00	1.00	1.00
32	70.64	1.00	1.00	1.00	-0.77
33	79.55	-1.00	-1.00	-1.00	-0.61

Table 4.8: Experimental results from the Hydrophobic separation step.

run	$Pur3$	$pH.C$	$cond$	$dens$	$Pur2$
1	86.10	-1.00	1.00	-1.00	-0.80
2	91.23	1.00	1.00	1.00	-0.60
3	91.18	-1.00	-1.00	0.00	0.60
4	89.47	1.00	1.00	-1.00	-0.30
5	97.89	0.00	-1.00	1.00	0.70
6	91.98	-1.00	-1.00	1.00	-0.90
7	96.63	1.00	-1.00	-1.00	0.70
8	93.02	-1.00	-1.00	-1.00	1.00
9	98.47	1.00	1.00	1.00	0.40
10	91.18	0.00	1.00	1.00	-0.70
11	95.68	1.00	-1.00	-1.00	0.70
12	90.43	-1.00	1.00	1.00	-0.10
13	90.08	-1.00	-1.00	-1.00	-0.40
14	90.64	-1.00	0.00	-1.00	0.10
15	88.61	0.00	1.00	0.00	0.70
16	95.72	-1.00	1.00	-1.00	1.00
17	97.11	0.00	-1.00	1.00	0.50
18	86.20	0.00	0.00	0.00	-1.00
19	96.68	-1.00	1.00	1.00	0.30
20	85.62	1.00	1.00	-1.00	-0.90
21	85.80	-1.00	1.00	0.00	-0.60
22	84.39	1.00	0.00	0.00	-0.40
23	90.25	0.00	-1.00	-1.00	0.00
24	100.65	1.00	-1.00	1.00	0.40
25	90.95	-1.00	0.00	1.00	-0.90
26	88.49	-1.00	-1.00	0.00	-0.20
27	97.38	1.00	0.00	1.00	0.60
28	91.58	-1.00	0.00	1.00	0.90
29	95.61	1.00	1.00	0.00	0.60
30	90.14	1.00	-1.00	-1.00	-0.70
31	89.79	1.00	-1.00	1.00	-0.60
32	92.57	1.00	1.00	-1.00	0.60
33	91.64	-1.00	1.00	-1.00	0.70

Bibliography

[1] Shi J, Zhou S. Quality control and improvement for multistage systems: A survey. *IIE Trans* 2009; 41(9): 744-753.

[2] ICH guideline Q8 (R2) on pharmaceutical development. *U.S. Department of Health and Human Services, Food and Drug Administration.* November 2009.

[3] Yu L. Pharmaceutical Quality by Design: Product and Process Development, Understanding, and Control. *Pharm Res* 2008; 25: 781-791.

[4] Jiang C, Flansburg L, Ghose S, Jorjorian P, Shukla AA. Defining process design space for a hydrophobic interaction chromatography (HIC) purification step: Application of quality by design (QbD) principles. *Biotechnol Bioeng* 2010; 107(6): 985-997.

[5] Tyssedal J, Kulahci M. Experiments for Multi-Stage Processes. *Qual Technol Quant Manag* 2015; 12(1): 13-28.

[6] Bjørkestøl K, Sivertsen E, Næs T. Optimization of two-step batch processes and the method of compensation for random error. *J Chemom* 2012; 26(6): 311-321.

[7] Beshah S, Desta G. Multi-Stage and multi-response process optimization in Taguchi method. *JEEA* 2015; 33.

[8] Moslemi A, Seyyed-Esfahani M. Robustness Indices in Multistage Problems. *Qual and Reliab Eng In* 2017; 33.

[9] Yuangyai C, Lin DK. Understanding multistage experiments. *International IJEDPO* 2013; 3(4): 384-409.

[10] Kulahci M, Tyssedal J. Split-plot designs for multistage experimentation. *J Appl Stat* 2017; 44(3): 493-510.

[11] Pronzato L. Information matrices with random regressors. Application to experimental design. *J Stat Plan Inference* 2002; 108: 189-200.

[12] Donev A. Design of experiments in the presence of errors in factor levels. *J Stat Plan Inference* 2004; 126(2): 569–585.

[13] Montgomery D. *Design and Analysis of Experiments*. New York: John Wiley & Sons; 2008.

[14] Box GEP. The Effects of Errors in the Factor Levels and Experimental Design. *Technometrics* 1963; 5(2): 247-262.

[15] Jensen WA. Open problems and issues in optimal design. *Qual Eng* 2018; 30(4): 583-593.

[16] Silvey S. *Optimal design: an introduction to the theory for parameter estimation*. Springer Netherlands; 1980.

[17] Fedorov V. *Theory Of Optimal Experiments*. New York: Academic Press; 1972.

[18] Zahran A, Anderson-Cook CM, Myers RH. Fraction of Design Space to Assess Prediction Capability of Response Surface Designs. *J Qual Technol* 2003; 35(4): 377-386.

[19] Eon-Duval A, Valax P, Solacroup T, Broly H, Gleixner R, Strat CLE, Sutter J. Application of the quality by design approach to the drug substance manufacturing process of an Fc fusion protein: Towards a global multi-step design space. *J of Pharm Sci* 2012; 101(10): 3604 - 3618.

[20] Krishnamoorthy K, Mathew T. *Statistical tolerance regions: theory, applications, and computation*. New York: John Wiley & Sons, Wiley Series in Probability and Statistics; 2008.

[21] Miller AJ, Nguyen NK. Algorithm AS 295: A Fedorov Exchange Algorithm for D-Optimal Design. *J R Stat Soc Ser C Appl Stat* 1994; 43(4): 669–677.