Developing a Patient-Focused Study Design for Rare Disease Clinical Trials

by

Jian Yong

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Epidemiology

Department of Public Health Sciences
University of Alberta

Abstract

Randomized controlled trials for a rare disease face methodological difficulties in evaluating treatment effects due to characteristics of rare diseases such as a small patient population to recruit from, lack of knowledge about the disease itself (i.e., lack of clinically validated endpoints), and heterogeneity of patients.

The proposed trial design in this thesis for rare disease controlled trials aims not only efficiency in evaluating treatments compared to the standard parallel-group design of clinical trials, but also incorporates other important aspects in developing rare disease treatments such as providing more opportunities for rare disease patients to access new treatments, evaluating treatment effects where clinically validated endpoints are lacking, and identifying markers for treatment response for use in clinical practice. It is a patient-focused design consisting of two stages. Stage 1 provides all patients the opportunity to access the experimental treatment of the trial and identify patient characteristics that distinguish patients who respond to the experimental treatment and those who do not. Stage 2 is to evaluate treatment effects with a randomization of treatments to patients who had responded to the experimental treatment in Stage 1. To compare the effect of an experimental treatment to that of a standard treatment, Stage 2 uses cross-over design, series of n-of-1 trials design, or response-adaptive design. For both stages, patient-reported outcomes are collected to evaluate treatment effect on rare diseases where clinically validated endpoints may be lacking. Analysis methods and sample size calculations for the proposed two-stage design that uses cross-over design in Stage 2 are explained.

Table of Contents

List of Tables

List of Figures

Glossary of Terms

CNS	Central Nervous System
ERT	Enzyme Replacement Therapy
FGSR	Faculty of Graduate Studies and Research
FDA	Federal Drug Administration
LSD	Lysosomal Storage Disease
MLE	Maximum likelihood Estimation
RCT	Randomized Controlled Trials
RDT	Rare Disease Controlled Trials

Chapter 1: Introduction

1.1 Thesis Organization

This paper-based thesis was prepared in accordance to the Faculty of Graduate Studies and Research (FGSR) of the University of Alberta guidelines. The thesis is organized as follows:

Chapter 1 - Introduction

Chapter 2 - Background

Chapter 3 - A Two-Stage Patient-Focused Rare Disease Controlled Trials

Chapter 4 – Analysis methods and sample size calculations for the patient-focused two-stage study design

Chapter 5 - Discussion and Conclusion

1.2 Rationale

Rare disease is defined as, under the US Food and Drug Administration (FDA) guideline, a disease or a condition in which its prevalence is less than 1 in 200,000 in United States in any given year [1]. While the prevalence of a particular rare disease may be less than 1 in 200,000, it is estimated that collectively more than 7000 types of rare diseases have affected up to 8 percent of the world population [2]. Majority of these rare diseases do not have effective cure available and they are expensive to develop and to produce [3].

Using randomized controlled trials (RCTs) with parallel group design to evaluate treatments for rare diseases faces a number of methodological challenges [4-6]. These challenges include poorly understood etiology, natural history and epidemiology of rare diseases, small patient numbers, heterogeneous patient characteristics, lack of validated clinical endpoints, and scarcity of clinical experts [4-6]. Because of these challenges, rare disease controlled trials (RDTs) have limited capability to evaluate the effects of treatments on rare diseases compared to common diseases [7].

Alternatives to the standard parallel-group design of RCTs, other designs such as randomized withdrawal design and cross-over design have been suggested to overcome some of the methodological issues of evaluating treatment effects in RDTs [8-10]. For example, cross-over design, in which patients switch allocated treatments after a specific time and a washout period, requires less number of patients than the parallel-group design to achieve the equivalent statistical power: this may be applicable in certain rare disease scenarios where outcomes are reversible and can be evaluated under this design [9,10]. These alternative study designs also consider aspects other than study power that are important in evaluating treatment effects in RDTs [9,10]. For example, the use of randomized withdrawal design can reduce the chance/time of patients being exposed to placebo, and examine the average treatment effect of patients who respond to the treatment under study [9,11].

Even after rare disease treatments are regulatory approved, substantial challenges exist in administering the treatments. These challenges include high price for patients to access treatments and the fact that treatment may be effective in only a subgroup of patients with the rare disease of interest [12]. For example, in Gaucher disease, patients with the most severe Lysosomal Storage Disease (LSD) have both central nervous system (CNS) and systemic disease.

The less severe, most common phenotype (Gaucher disease type I) has no CNS involvement but does have a broad spectrum of systemic involvement. While Enzyme Replacement Therapy (ERT) is very effective for treating many of the systemic features of Gaucher disease, it is not effective in the treatment of the CNS disease [12]. Therefore, ERT is more suitable for Gaucher disease patients who only have systemic disease compared to Gaucher disease patients who have both CNS and systemic disease. Identifying such subgroup-specific treatment effects is challenging in a small population of a rare disease.

Health Canada indicated that more feasible and effective trial designs than the standard RCT designs are needed for RDTs in practice [14]. This is also an emphasized point in the US FDA's white paper entitled CPI Report to Congress: Improving the Prevention, Diagnosis and Treatment of Rare and Neglected Diseases released in March 2011 [13]. This thesis develops a study design with an analysis framework and a sample size calculation method to address some of the issues of RDTs.

1.3 Study Objectives

This thesis is aimed to propose a study design with analysis methods and sample size calculations for RDTs to evaluate treatment effect and to provide useful information for better clinical practice.

Following a background review in Chapter 2, Chapter 3 proposes a study design of RDT that aims to improve the use of RDT for treatment effect evaluation, better resource management of healthcare system and treatment access opportunities for patients during treatment effect

evaluation. Key considerations for the development of this study design will be given. Strengths and limitations of this study design will be discussed.

Chapter 4 proposes an analysis framework and a sample size calculation method for one of the proposed study designs in Chapter 3, specifically utilizing crossover design at Stage 2. Analysis methods for each stage of the proposed study design and a sample size calculation for the proposed study design are described. A rare disease scenario where this analysis framework is ideally suitable for and why it is suitable is illustrated.

Chapter 5 summarizes the results and conclusions of Chapters 2 to 4. Areas of interests for future work are also discussed.

1.4 References

1. U.S. Food and Drug Administration. Orphan drug act. Available: http://www.fda.gov/ForIndustry/DevelopingProductsforRareDiseasesConditions/default.htm.
2. Gatta G, Capocaccia R, Trama A, Martinez-Garcia C. The burden of rare cancers in Europe. Adv Exp Med Biol 2010;686:285-303.
3. Miles KA, Packer C, Stevens A. Quantifying emerging drugs for very rare conditions. QJM 2007;100:291-5.
4. Behera M, Kumar A, Soares HP, Sokol L, Djulbegovic B. Evidence-based medicine for rare diseases: implications for data interpretation and clinical trial design. Cancer Control 2007;14(2):160-6.
5. Dunoyer M. Accelerating access to treatments for rare diseases. Nat Rev Drug Discov 2011;10:475-6.
6. Coté RT, Xu K, Pariser RA. Accelerating orphan drug development. Nat Rev Drug Discov 2010; 9:901-2.
7. Tambuyzer E. Rare diseases, orphan drugs and their regulation: questions and misconceptions. Nat Rev Drug Discov 2010;9(12):921-9.
8. Gerss JW, Kopcke W. Clinical trials and rare diseases. Adv Exp Med Biol 2010;686:173-90.
9. Gupta S, Faughnan EM, Tomlinson AG, Bayoumi MA. A framework for applying unfamiliar trial designs in studies of rare diseases. J Clin Epidemiol 2011;64: 1085-94.
10. Institute of Medicine. Small clinical trials: issues and challenges. Washington:National Academy of Press; 2001.
11. Kianifard F. Islam MZ. A guide to the design and analysis of small clinical studies. Pharmaceutical Statistics 2011;10:363-8.

12. Cassino C, Orfali M, Charnigo RJ, Marsden DL. Rare and orphan diseases challenges: clinical development and clinical practice. Journal of Rare Disorders 2013;1(1):1-3.

13. U.S. Food and Drug Administration. Challenge and opportunity on the Critical Path to New Medical Products. Available: http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/default.htm.

14. Health Canada. An orphan drug framework for canada [updated 2012 October 3. About Health Canada. Available: http://www.hc-sc.gc.ca/ahc-asc/media/nr-cp/_2012/2012-147a-eng.php.

Chapter 2: Background

Rare diseases comprise a various group of over 7,000 disorders [1]. In United States, a rare disease is defined by prevalence of one out of 200,000 or less [2]. Even though the prevalence of all rare diseases combined is high, estimated to be 6-8% of the world population [1], treatments for these diseases is scarce [3]. Majority rare diseases have a genetic basis, 85% are serious or life threatening, and more than 50% affect children [4]. In this chapter, we describe: (1) issues in conducting standard randomized controlled trials to evaluate effects of treatments for rare diseases; (2) study designs that have been proposed for rare disease research; and (3) an analysis framework that has been suggested for rare disease research.

2.1 Rare disease treatment development issues

A randomized controlled trial has been considered as the gold standard in evidence-based medicine to evaluate treatment effects as it is well known for minimizing bias. Results of randomized controlled trials are often accepted as providing the strongest evidence in testing a hypothesis [5]. However, under the standard framework, conducting rare disease controlled trials to evaluate treatment effects for certain rare diseases might not only be technically challenging but also economically impractical.

2.1.1 Technical Issues

2.1.1.1 Patient recruitment and clinical endpoints

The prevalence of any rare diseases is 1 in 200,000 or less in United States [2]. As a result, for many rare diseases, there is a lack of clinically validated outcomes; hence recruiting hundreds of such scarce and difficult-to-be-identified patients into clinical trials may not be feasible [6-8]. Despite this, evaluation of rare diseases treatments requires similar standards of evidence for marketing authorization as treatments for common diseases [9]. Furthermore, the lack of clinically validated outcomes is not only causing challenges in identifying and recruiting patients, but also causing difficulties in the choice of primary endpoints.

2.1.1.2 Patient Heterogeneity

In addition to the small number of patients and lack of validated clinical endpoints available for clinical trials, many rare diseases are heterogeneous with respect to clinical phenotype and underlying pathophysiology [10]. It is therefore challenging for clinical trials to evaluate the effect of a treatment on a rare disease due to the large variability of treatment response (among a small number of patients). This is because randomized controlled trials usually evaluate average treatment effects on the underlying study population; an average might not be a close approximation to individuals of a rare disease population due to the wide range of responses to treatments

2.1.2 Economical Issues

The small number of patients is not only an issue in treatment effect evaluation, but also presents financial issues for the society. A common false impression is that rare disease

treatments are cheaper to develop than other drugs [6]. Researching, developing, manufacturing and authorizing any treatment to market is a long, complex process. Previous studies show that approximately 30% of all treatments fail in Phase III trials for common diseases and 37.1% of all treatments fail in Phase III trials for rare diseases [11, 12]. Since industries need to invest in the treatment development from biological laboratories to treatment effect evaluations in clinical trials, the higher proportion of failure to market authorization implies more testable treatments must be developed for one to be approved. The development of treatment for rare diseases is not cheaper than common diseases.

Since the cost of developing a rare disease treatment is higher than common diseases and the size of rare disease population is small, rare disease treatment is generally expensive [13]. For example, the price per patient of an enzyme replacement therapy can be over 400,000 euros per year [13]. As a result, it is financially challenging for individual patients to access rare disease treatments and reimbursement is often needed.

Rare disease patients have limited access to treatments because the cost is high and the reimbursement of government and insurance companies for rare disease treatments is not always available. Some may consider that, although rare disease treatments are very expensive, the prevalence is small and hence the financial impact of reimbursement is not substantially different from common diseases. Many rare disease treatments, however, do not provide cure but may improve patients' conditions or slow down the progression of the disease [14]. Therefore, for many rare diseases, patients need continuous reimbursement to receive treatments [14]. Thus, the financial impact of a rare disease can be substantial to the society.

2.2 Efforts towards Development of Rare Disease Treatments

Many countries including the United States and European countries have implemented legislative actions to provide financial incentives for industries to develop rare disease treatments [15, 16]. Health Canada is developing a framework for the designation, authorization, and monitoring of treatments that will substantially improve the outcome of Canadians with rare diseases, and encourage innovation in methods to improve the technical issues of clinical trials to evaluate rare disease treatment effects [1].

In addition to legislative actions being taken in these countries to provide financial incentives for industries to develop rare disease treatments, alternative study designs and analysis frameworks have been proposed to overcome technical issues of conducting rare disease controlled trials [17-31]. Section 2.2.1 describes study designs that are proposed for rare disease controlled trials. Section 2.2.2 describes analysis methods that are proposed in the literature for rare disease controlled trials.

2.2.1 Study designs

As illustrated in Section 2.1.1, the technical issues of conducting parallel-group randomized controlled trials to evaluate treatment effects for rare diseases include small patient number, lack of validated outcomes and patient heterogeneity. A number of trial designs have been proposed in the literature as alternatives when the typical parallel-group design is not appropriate or feasible [18, 19, 24, 29,34-36]. These alternative designs include randomized withdrawal design, crossover design, various types of adaptive designs, and series of n-of-1 trials. These methods attempt to evaluate treatment effects satisfying the evidence requirement for regulatory approval in smaller and shorter studies compared to the parallel-group design [29].

These methods in certain scenarios are more efficient than the parallel-group design; however, there are not without drawbacks and require investigators to make a number of assumptions, some of which might be challenging to verify. For example, crossover design trials in which patients act as their own controls result in a smaller number of patients than the parallel-group design; however, assumptions regarding carryover effect, period effect, treatment by period interaction, and the characteristic of outcomes need to be considered and met [29].

2.2.2 Alternative analysis framework

Randomized controlled trials have become the standard method to assess the safety and effect of medical treatments. This establishment is partly due to the feature of its design such as randomization that can minimize bias and also its methodological sound analysis framework. The current widely-accepted and applied analysis framework in clinical trials, the frequentist framework, is an approach to statistical inference that focuses only on observed data in the trial [37]. To make conclusion on treatment effect and safety based solely on observed data in an experiment, a considerable number of patients are usually required. As described in Section 2.2.1, limited numbers of patients are available for recruitment for rare disease trials; hence, under this analysis framework, conducting randomised controlled trials is a challenge [23]. Researchers have been considering alternative analysis framework to collect evidence on treatment effects that meets the scientific standard of clinical research.

Bayesian analysis framework has been proposed as an alternative to frequentists' for rare disease clinical trials to evaluate treatment effect. Details of methodologies of Bayesian have been described and illustrated with examples (See Appendix 2) [37-41]. Briefly, unlike the

frequentist approach, Bayesian analysis can combine information from the trial itself and information from other sources (e.g., expert opinions and previous studies) to make statistical inference on the effect of treatment for rare diseases [23, 29, 37]. Furthermore, unlike the frequentist approach that depends on trial design to control type I and type II error for statistical inference, Bayesian can make statistical inference as information accrues within a trial or aggregate information across trials relatively more flexibly than the frequentist approach [39]. As a result, rare disease clinical trials with Bayesian approaches may require smaller sample sizes and, therefore, may be more feasible [41].

2.3 References

1. Eurodis. Rare disease: Understanding this public health priority 2005. http://www.eurodis.org/IMG/pdf/princeps_document-EN.pdf.
2. U.S. Food and Drug Administration. Orphan drug act. Available: http://www.fda.gov/ForIndustry/DevelopingProductsforRareDiseasesConditions/default.htm
3. Schieppati A, Henter JI, Daina E, Aperia A. Why rare diseases are an important medical and social issue. Lancet 2008; 371(9629):239-41.
4. Wellman-Labadie O, Zhou Y. The US Orphan Drug Act: rare disease research stimulator or commercial opportunity? Health Policy 2010; 95:216-8.
5. Emaneul EJ, Miller FG, The ethics of placebo-controlled trials: a middle ground, N.Eng. J. Med. 2001; 345: 915-9.
6. Tambuyzer E. Rare diseases, orphan drugs, and their regulation: questions and misconceptions. Nat Rev Drug Discov 2010; 12: 921-9.
7. MacLeod S. Optimal therapy for rare disorders and genetic diseases: ethical and political challenges. Proc West Pharmacol Soc 2007; 50:21-3.
8. BEhera M, Kumar A, Soares HP, Sokol L, Djulbegovic B. Evidence-based medicine for rare diseases: implications for data interpretation and clinical trial design. Cancer Control 2007;14 (2):160-6.
9. European Medicines Agency. Guidelines on clinical trials in small populations. Doc. Ref. CHMP/EWP/ 83561/2005.
10. Dunoyer M. Accelerating access to treatments for rare diseases. Nat Rev Drug Discov 2011; (10): 475-6.

11. CRA International. The current state of innovation in the pharmaceutical industry. European Federation of Pharmaceutical Industries and Associations 2008. Available: http://www.efpia.org/content/default.asp?PageID=559&DocID=4896.

12. Joppi J, Bertele V, Garattini S. Orphan drug development is not taking off. Br. J. Clin. Pharm 2009; 67;494-502.

13. Tambuyzer E. Rare disease, orphan drugs and their regulation: questions and misconceptions. Nat Rev Drug Discov 2010; 9: 921-9.

14. Haffner ME, Whitley J, M, Moses M. Two decades of orphan product development. Nat Rev Drug Discov 2002; 1:821-5.

15. The European Parliament and the Council of the European Union. Regulation (EC) no. 141/2000 of the european parliament and of the council of 16 December 1999 on orphan medical products. Official J. Eur. Communities 2000; L8/1-L18/5.

16. US Food and Drug Administration. Developing products for rare diseases & conditions. US FDA website 2010. Available: http://wwww.fda.gov/ForIndustry/DevelopingProducts for Rare DiseasesConditions/default.htm.

17. Coté TR, Xu K, Pariser AR. Accelerating orphan drug development. Nat Rev Drug Discov 2010; 9:901-2.

18. Gupta S, Faughnan ME, Tomlinson GA, Bayoumi AM. A framework for applying unfamiliar trial designs in studies of rare diseases. J Clin Epidemiology 2011; 64: 1085-94.

19. Kianifard F, Islam MZ. A guide to the design and analysis of small clinical studies. Pharm Stat 2011; 10:363-8.

20. van der Lee JH, Wesseling J, Tanck MWT, Offringa M. Efficient ways exist to obtain the optimal sample size in clinical trials in rare diseases. J Clin Epidemiology 2009; 61:324-30.

21. Gerß JWO, Kopcke W. Clinical trials and rare diseases. In: Paz MP, Groft SC, Eds. Rare Diseases Epidemiology (Advances in Experimental Medicine and Biology), New York: Springer;2010:173-90.

22. Honkanen VEA, Siegel AF, Szalai JP, Berger V, Feldman BM, Jeffrey NS. A three-stage clinical trial design for rare disorders. Statist Med 2001; 20: 3009-21.

23. Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of conundrum. BMJ 1995; 311: 1621-5.

24. Cornu C, Kassai B, Fisch R, Chiron C, Alberti C, Guerri R, et. al. Experimental designs for small randomised clinical trials: an algorithm for choice. Orphanet J Rare Dis 2013;8: 48-60.

25. Buckley BM. Clinical trials of orphan medicines. Lancet 2008; 371: 2051-5.

26. Lagakos SW. Clinical trials and rare diseases. N Engl J Med 2003; 348:2455-66.

27. Chow SC, Chang M. Adaptive design methods in clinical trials – a review. Orphanet J Rare Dis 2008; 3:11-7.

28. Day S. Evidence-based medicine and rare diseases. Adv Exp Med Biol 2010; 686:41-53.

29. Institute of Medicine. Small clinical trials: issues and challenges. Washington DC: National Academy Press; 2011.

30. Slilker B. Guide to clinical trials. New York: Raven; 1991.

31. Chow S-C, Chang M. Adaptive design methods in clinical trials. Chapman & Hall/CRC, Boca Raton; 2006.

32. Bashaw ED, Huang SM, Coté TR, Pariser AR, Garnett CE, Burckart G, et.al. Clinical pharmacology as a corner-stone of orphan drug development Nat Rev Drug Discov 2011; 10:795-6.

33. Kawut SM, Bagiella E, Lederer DJ, Shimbo D, Horn EM, Roberts KE, et.al. Randomized clinical trial of aspoirin and simvastatin for pulmonary arterial hypertension:ASA-STAT. Circulation 2011; 123:2985-93.

34. Griggs RC, Batshaw M, Dunkle M, Gopal-Srivastava R, Kaye E, Krischer J, et.al. For the rare diseases clinical research network. Clinical research for rare disease: Opportunities, challenges, and solutions. Molecular Genetics and Metabolism 2009; 96:20-6.

35. Honkanen VEA, Siegel AF, Szalai JP, Berger V, Feldman BM, Siegel JN. A three-stage clinical trial design for rare disorders. Statistics In Medicine 2001;20:3009-21.

36. Feldman B, Wang E, Willan A, Szalai JP. The randomized placebo-phase design for clinical trials. J Clin Epidemiology 2001; 54:550-7.

37. Berry DA. Bayesian clinical trials. Nat Rev Drug Discov 2006;5: 27-36.

38. Etzioni RD, Kadane JB. Bayesian statistical methods in public health and medicine. Annual Review of Public Health 1995; 16:23-41.

39. Fayers PM, Ashby D, Parmar MK. Tutorial in biostatistics Bayesian data monitoring in clinical trials. Statistics in Medicine 1997; 16:1413-30.

40. Berry DA, Stangl DK. Bayesian methods in health-related research. In: Bayesian Statistics. Berry DA and Stangl DK, eds. New York: Marcel-Dekker; 1996. p.13-66.

41. Spiegelhalter D, Freedman LS, Parmar MKB. Bayesian approaches to randomised trials. Journal of the Royal Statistical Society 1994; 157:357-416.

Chapter 3: A Two-Stage Patient-Focused Study Design for Rare Disease Controlled Trials

3.1 Introduction

A disease is defined as rare if it affects less than 1 in 200,000 people in the population [1]. With approximately 6-8% of people worldwide afflicted [2], rare diseases present a substantial burden to healthcare systems. For many of the over 6,000 diseases that fall into this characterization [1-2], there are little or no effective treatments available and the diseases themselves are poorly understood.

For the purpose of evaluating treatment effect, there are many challenges with using the standard randomized controlled trial (RCT) framework. These challenges include small patient numbers, lack of validated clinical endpoints, and heterogeneity in patient characteristics, all due to the rarity of diseases [3-7]. In addition, the rarity of disease, lack of information/knowledge on it, and high manufacturing costs for a relatively small patient population are some of the challenges faced during the development of orphan drugs, making them both scarce and expensive [3-7]. This in turn limits the accessibility and availability of treatments to patients [3-7].

Therefore, in addition to small sample size issues, rare disease controlled trials (RDT) have several other relevant considerations [2]. These considerations include increasing treatment availability and accessibility for rare disease patients and optimizing healthcare resource utilization for future treatment allocation, development and prioritization.

We propose an RDT study design that evaluates treatment effect comparatively under the standard trial requirements (i.e., randomization, blinding, allocation concealment) and measures treatment response where clinically validated outcomes may be lacking. Under this design, all enrolled patients have access to the experimental treatment and patient characteristics that are found to be associated with treatment response can be useful for tailoring treatments based on patient characteristics in clinical practice.

3.2 Methods

We systematically searched for English-language publications using multiple combinations of search terms (Table 1) in PubMed and included relevant literature up to January 2014. We selected literature that placed an emphasis on one of the following: 1) challenges with conducting RDTs; 2) review of previously proposed RDT study designs; or 3) application of RDT study designs.

Based on the literature identified as above, we created a list of important considerations that we deemed would aid investigators in selecting the most appropriate RDT study design given their disease of interest. Following discussions with two experts (Dr. Robin Casey, a physician involved in providing care for rare disease patients, and Dr. Robyn Lim, a senior scientific advisor at the Office of Legislative and Regulatory Modernization at Health Canada) we finalized a list of four key considerations: 1) patients' opportunity to access the new treatment; 2) assessment of outcomes where clinically validated outcomes may be lacking; 3) patient heterogeneity; and 4) duration of the study and number of patients required. With the inclusion of these four key considerations, we developed a new study design for RDTs.

3.3 Results

We first describe our study design which applies specifically to RDT scenarios where the outcome of the disease is reversible. In section 3.3.4, possible modifications that can be made to the proposed study design for non-reversible outcomes are described. Details on how this study design satisfies the four key considerations listed in the methods section will be given in Discussion.

3.3.1 Overview of Proposed Study Design

Our proposed study design has two stages. Stage 1 is an enrichment stage which distinguishes patients who respond to the treatment ("responders") from those who do not respond to the treatment ("non-responders") after assigning them all to the experimental treatment. Stage 2 evaluates the treatment effect comparatively among only those patients characterized as responders in stage 1 (Figure 1).

3.3.2 Proposed Stage 1 Framework

At the onset of stage 1, detailed baseline characteristics of all enrolled study patients, which include demographic information (e.g., age and sex), specific clinical/biological information such as characterizations of the disease, symptoms, known biological markers for the disease, and other physiological measurements, are collected. All enrolled patients are then given access to the experimental treatment with modifiable patient characteristics monitored throughout stage 1. Monitoring modifiable patient characteristics might provide useful

information about how certain patient characteristics change in response to the experimental treatment.

At the end of stage 1, patients are characterized as either responders or non-responders. Responsiveness to experimental treatment is based on pre-determined criteria on outcome improvement defined by patient self-reports and/or clinical evaluation. Patient self-reported outcomes include whether or not patients found the treatment to be beneficial or harmful. Based on this inference, if a small number of, or no, patients are characterized as responsive at the end of stage 1, the trial will be stopped with the conclusion that the treatment is overall ineffective. If sufficient improvement is seen, responders will proceed to stage 2 while non-responders will be withdrawn from the trial. Investigators should specify, in the study protocol, the minimum proportion of enrolled patients with clinically important absolute improvement that is need to be observed at the end of stage 1 to justify proceeding to stage 2.

Information about patient characteristics collected at baseline and throughout stage 1 will be analyzed to determine the association between patient characteristics and treatment response. This analysis can be done using standard statistical methods of binary outcomes (response vs. non response) such as logistic regression. Using the results from this analysis, patient subgroups can be created based on specific sets of characteristics to determine those most likely to get benefit and/or harm from the experimental treatment. This is useful information that can aid future treatment allocation, development, and prioritization since treatment accessibility and availability in rare disease populations are usually limited.

3.3.3 Proposed Stage 2 Framework

Following a washout period, patients who were characterized as responders at stage 1

proceed to stage 2 to evaluate the experimental treatment effect with a control group. Depending on the characteristic of the rare disease's outcome under study, an appropriate study design for evaluating the treatment effect in stage 2 will be used.

Some study designs for evaluating rare disease treatment effects comparatively have been suggested [8,9,13,15]. Of the previously proposed 14 study designs, we shortlist three for use in stage 2: cross-over design; series of n-of-1 trials design; and response-adaptive design (Table 2). These study designs were selected because they include randomization, allocation concealment, blinding of treatment allocations, and allow for interim analysis. Also, all three study designs provide enhanced opportunities for enrolled patients to access the experimental treatment.

The use of interim analysis can potentially make treatment effect evaluation at stage 2 more efficient. During stage 2 interim analysis, if patients' outcomes have already been found to have substantially improved, then fewer patients than anticipated may be required for stage 2.

For specific RDT scenarios, one of these study designs may be more suitable than others. A summary of these three designs' features are given here:

1) Cross-over design

- Definition: patients are randomly assigned to receive two or more treatments sequentially with wash-out periods between consecutive treatments and each patient acts as his/her own control [9].
- Assumptions: statistically-valid comparison of two or more treatments is possible where there is no treatment-period interaction and negligible carryover effect [9].
- Advantages: since the comparison of treatment effect is within the same patient, this study design can more precisely estimate treatment effect than

adaptive design when patient characteristics are heterogeneous [10,11,13,16,17].

- Limitations: 1) longer study duration is required than adaptive design because each patient receive more than one treatment; 2) sufficient washout period is required between each treatment received; and 3) consequences of dropouts are greater than adaptive design because more information is lost per dropout [9,11,13,16,17].
- Analysis: generalized linear model can be used in which the outcome can be binary or continuous. The effect of treatments, carry over, and periods are captured appropriately and evaluated in the model [18].

2) Series of n-of-1 trials design

- Definition: two or more treatments are consecutively and repeatedly given in a random order to each patient who contributes information to one n-of-1 trial. A series of n-of-1 trials are analyzed jointly for treatment effect evaluation [9,13,19].
- Assumptions: comparing multiple treatments under this design is statistically valid where there is no treatment-period interaction and no carry over effect [9,13,19].
- Advantages: since this design is capable of assessing the effect of the treatment on each patient, it can estimate treatment effect more precisely when the disease is extremely rare and when patient characteristics are highly heterogeneous [9, 13,19].
- Limitations: 1) longer study duration is required due to more number of treatments received and/or more repetition of same treatments received than

cross-over design and adaptive design; 2) similar to cross-over design, sufficient washout period is required between treatments received; and 3) the consequences of dropouts are greater than adaptive design because more information is lost per dropout.

- Analysis: meta-analysis can be performed for this study design by aggregating the data of each n-of-1 trial to obtain the average treatment effect for the population and individual patient, as well as estimates of between-patient variation [20, 21].

3) Response-adaptive study design

- Definition: a study design that allows modification of randomization schemes during the trial based on interim trial results. This is done by varying the probabilities of treatment assignment to increase the likelihood of patients being assigned to the superior treatment and minimizing the number of patients exposed to the inferior treatment [9,13,22-28].
- Advantages: 1) carry over effect is not an issue because patients receive only one treatment; 2) the study duration is generally shorter than cross over trials and series of n-of-1 trials since patients only receive one treatment; 3) response-adaptive design is a parallel group design if no adaptation is made during the trial; and 4) the outcome response of previous patients can guide trials to assign the better treatment with high probabilities to newly recruited patients.
- Limitations: 1) not suitable for heterogeneous populations; and 2) complex statistical analysis is usually required under frequentist statistical inference [9,13,22-28].

- Analysis: maximum likelihood estimation with consideration of correlation and permutation test are suggested as the analysis method for this study design [23, 26].

Stage 2 analysis, as in stage 1, includes the perspective of enrolled patients for treatment effect evaluation since patient feedback is important information in determining the experimental treatment's effectiveness: it describes aspects of patients' experience with the treatment that can be only obtainable from its actual users (e.g. pain, dizziness).

3.3.4 Potential modification for non-reversible outcomes

Stage 2 of our proposed study design can be modified for situations where the outcome is non-reversible by recruiting new patients directly into stage 2 instead of using stage 1 patients. Patient characteristics collected in stage 1 will be used to examine what subgroups of patients did not respond or had adverse response to the treatment. New recruitment will have the same inclusion criteria as stage 1 but will exclude patients who possess characteristics that were found to be associated with treatment non-response in stage 1. By excluding patients who are unlikely to benefit from the experimental treatment, stage 2 will focus the treatment effect evaluation on patients who are more likely to respond.

3.4 Discussion

It is estimated that up to 8% of the world has a rare medical condition [2]. Since rare disease treatments often cost much more than what many patients can afford, funding for the treatment access of a small population could significantly impact the budget of healthcare

systems [2]. Furthermore, since there is limited information related to the epidemiology of many rare diseases and rare disease treatment effects, treatment allocation in clinical practice might often be done without evidence from trials, which may result in poor utilization of healthcare resources. Thus, a study design that can aid prioritization of resource utilization in treatment development and treatment allocation in clinical practice and has provision of treatment access to patients, as well as evaluation of treatment effects, would make effective treatments more available for rare disease patients and would better utilize the limited healthcare resources. In sections 3.4.1-3.4.4, we provide a rationale for why each key consideration is important in these regards and then describe how each stage of our study design satisfies the consideration (summarized in Table 3).

3.4.1 Consideration 1 (Patient Opportunity to Receive the Experimental Treatment)

For many rare diseases, there is a lack of pharmaceutical industry's incentive to develop treatments for two reasons: 1) there is not a sufficiently large market to support treatment development; and 2) rare disease etiology is often poorly understood making treatment development less certain due to factors such as a lack of pharmacological targets for intervention. For these reasons, treatments for rare diseases are often lacking developments or unavailable. As an ethical consideration, it would greatly aid rare disease patients if RDTs made treatment access a priority given the scarcity of such opportunities in the rare disease community.

In order to enhance treatment access for patients in the trial, stage 1 of our proposed study design is an "enrichment" stage where all patients enrolled in the trial receive the experimental treatment. In addition, Stage 2 for treatment effect evaluation employs study

designs that give patients more opportunity to access the experimental treatment compared to a parallel group design. Specifically, n-of-1 and cross-over study designs ensure that all patients receive the experimental treatment during the treatment effect evaluation in Stage 2. For response-adaptive designs, the probability of receiving a superior treatment increases as the trial progresses, so patients have a greater chance of being selected into the superior treatment arm.

3.4.2 Consideration 2 (Assessment of outcomes where clinically validated outcomes may be lacking)

The effectiveness of a treatment in clinical trials is usually based on well-characterized clinical outcomes. If validated clinical outcomes are lacking, then it is challenging to justify the treatment effect on patients' health conditions. Therefore, it would aid treatment effect evaluation if RDTs can evaluate treatment effects in scenarios where no clinically validated endpoints for a disease are available.

Our study design takes this into consideration in both stages. In stage 1, we measure a range of modifiable patient characteristics before and after the treatment is administered and monitor the changes. This information can then be incorporated in stage 2 to define endpoints. In both stages 1 and 2, patient-feedback is used to evaluate treatment response. The measurements of modifiable patient characteristics and patient-feedback provide information about outcomes/endpoints relevant to evaluate treatment effect: they may also provide information about how the administration of the experimental treatment can be developed to be more patient-friendly.

3.4.3 Criteria 3 (Patient Heterogeneity)

Many rare diseases populations are known to be heterogeneous [2]. Patients with certain characteristics might respond to the experimental treatment whereas others might not. Since rare disease treatments are often expensive and scarce, the association between patient characteristics and treatment response is an important consideration for future treatment development and using treatments to patients who are found to respond in the clinical practice. This could provide future direction for treatment development by providing information regarding plausible biological mechanisms of treatments being more effective on certain patients than others.

To account for patient heterogeneity, in stage 1, we screen for responders and explore the patient characteristics associated with treatment response. This information is then used to create patient subgroups based on patient characteristics. For example, if a common biological marker exists among all responders but not among non-responders, this can inform new areas for pharmaceutical intervention and can also prioritize treatment allocation in the clinical setting. In stage 2, only patients who responded to the treatment in stage 1 are assessed, and hence the evaluated treatment effect is more specific to certain groups of the patient population that respond to the treatment.

3.4.4 Criteria 4 (Duration of the study and recruitment of a sufficient number of patients)

A more efficient study design can get an effective intervention to more rare disease patients quicker. For RDTs, a long study period is often required to recruit a sufficient number of patients into the study: finding/identifying the specific rare disease patients who may be

geographically dispersed takes time. For these reasons, investigators should consider how the trial can be made more efficient without sacrificing patient safety and integrity of the trial.

In our study design, investigators are asked to specify, in the study protocol, what proportion of patients showing improvement in stage 1 is sufficient to justify proceeding to stage 2. If this proportion of improvement is not met in stage 1, the trial is stopped without proceeding to stage 2. For stage 2, interim analysis is used; again, if sufficient improvement is not seen, as specified in the study protocol, the trial is stopped. This improves efficiency and also helps with resource utilization by stopping the trial, should the treatment be found ineffective at an earlier point.

Additionally, in section 3.3.4, we describe an efficiency modification that can be made. If a clear pattern of patient characteristics are found to be associated with treatment response in stage 1, the study design can be modified by having newly recruited patients enroll directly into stage 2. One limitation of this modification is that if a response-adaptive design is used in stage 2, some patients may not get the opportunity to access the experimental treatment. The impact of such a loss of opportunity should be considered prior to making such a study design modification.

3.4.5 Comparison with another enrichment design

Our study design shares some similarities with a previously proposed RDT study design, the randomized withdrawal design [28-36]. Specifically, both our proposed study design and the randomized withdrawal design have 2 phases/stages; there is an initial enrichment stage where

all recruited patients receive the experimental treatment and only patients who respond to the treatment continue to the second stage for treatment effect evaluation.

The salient features of our design are: 1) the association of patient characteristics and responsiveness to the experimental treatment are analyzed at the end of stage 1; and 2) there are three study design options to evaluate treatment effect in stage 2. Regarding 1), we gain knowledge on patient subgroups for which the treatment is effective, which has an important implications on treatment development, allocation, and prioritization. On the other hand, the enrichment design is often motivated to "enrich" the patient population and increase power in the second stage's efficacy evaluation [31]. With 2), more study design options in stage 2 makes it possible to evaluate treatment effect for more rare diseases with those proposed designs, such as cross-over design, series of n-of-1 trials, compare to randomized withdrawal design in which only the parallel group design is used in stage 2. Our design considers specific RDT issues that patients, healthcare providers, funders and treatment developers face as described above.

3.4.6 Limitations

This proposed RDT framework has some potential limitations which should be considered. In general, investigations of the association between patient characteristics and treatment response in a clinical trial are recommended to be exploratory: the association may be specific to some aspects of the trial (e.g., season). The results on the association from stage 1 should, therefore, cautiously be generalized to the target population. Our suggestion to alleviate this limitation is to ensure sufficient power for identifying the associations in stage 1 as well as soliciting from patients and care providers potential specific aspects of the trials that may potential bias the trial results.

It is also possible that the treatment is found to be effective following stage 2, but there is ambiguity about how the characteristics of responders and non-responders differ. In such a scenario, the trial has a limited capacity to prioritize treatment based on patient characteristics. However, the information about treatment effects is still valuable for stakeholders in determining the proportion of the target population that is expected to respond to the treatment. Given that rare disease treatments often are very expensive, such a consideration will help with resource utilization.

Since the development of this study design did not consider regulatory aspects, our study design might be modified after the inclusion of regulatory aspects.

3.5. Conclusion

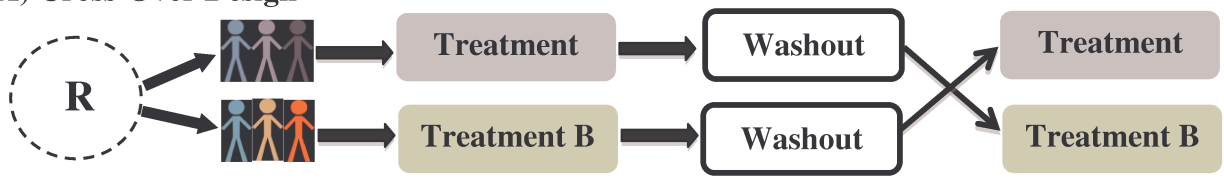
In addition to treatment effect evaluation, RDTs can potentially be of great benefit to rare disease patients and clinical practice by increasing opportunities to access experimental treatments and by providing relevant information that can be used for tailoring treatments to certain subgroups, aiding future research in treatment development, and improving healthcare resource utilization. Future work in applying our proposed design to rare disease clinical trials is needed in order to evaluate its robustness in practice. Although this design has some limitations and may not be suitable for all rare disease scenarios, it may serve as a basis for the future development of RDT study designs that better fulfill the needs of the rare disease community.

Table 3-1 Search terms for the identification of articles on rare disease clinical trial topics

• Rare disease	• Randomized withdrawal	• Crossover
• Research Design	• Adaptive design	• Series N-of-1
• Epidemiologic Methods	• Response adaptive design	• Small Clinical Trial
• Clinical trials	• Sequential design	• Orphan Drug
• Ranking and Selection	• Enrichment design	• Orphan
• N-of-1	• Clinical Trials as Topics	• Analysis

Table 3-2 Study design options for Stage 2 treatment effect evaluation

A) Cross-Over Design

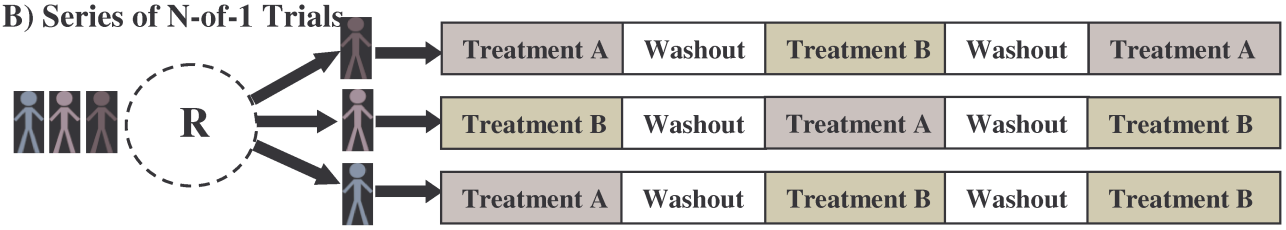


Definition:
Randomly assign sequences of two treatments to patient groups

Advantages:
- Comparison of two treatments
- No treatment-period interaction
- No carry over effect
- Heterogeneous populations

Limitations:
- Longer study duration
- Sufficient washout period needed
- Dropouts have greater influence

B) Series of N-of-1 Trials

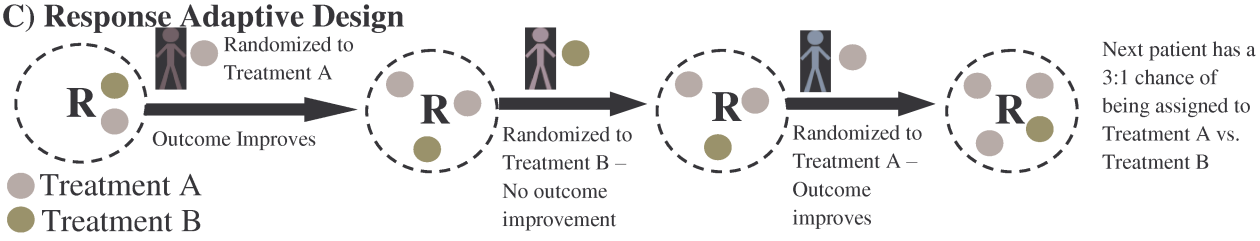


Definition
Randomly assign sequences of multiple treatments to only one patient in a trial for a series of trials

Advantages:
- Comparison of two or more treatments
- No treatment-period interaction
- No carry over effect
- Heterogenous populations
- Extremely rare disease

Limitations:
- Longer study duration
- Sufficient washout period needed
- Dropouts have greater influence

C) Response Adaptive Design



Definition:
Each treatment's chance of being assigned to patients is adapted based on accumulated patients' responses to treatments

Advantages:
- Evidence of one superior treatment
- Shorter study duration preferred

Limitations:
- Not suitable for heterogeneous populations
- Complex analysis

Note: "R" represents randomization. Treatment A and B refer to an experimental treatment and a standard treatment, respectively. For response-adaptive design, each patient receives Treatment A or B; the probability of receiving a specific treatment depends on the number of balls for that treatment in ths "ball-in-urn" design.

Table 3-2 Summary of how Stage 1 and 2 Satisfy the Four Key Considerations

Key Considerations	Rationale	Stage 1	Stage 2
1. Patient opportunity to access the new treatment	Due to a lack of accessible and effective treatments, treatment access is a high priority for rare disease patients	All patients receive the experimental treatment	Study designs give more opportunity to patients to receive treatment than parallel group design does
2. Assessment of outcomes where clinically validated outcomes may be lacking	The trial should evaluate treatment effect on all rare disease patients even without validated clinical endpoints so as to benefit clinical practice	- Patient-feedback on treatment effect	Patient feedback on treatment effect
3. Patient heterogeneity	Treatments should be prioritized to those patients most likely to benefit from them	- Patients are characterized into responders and non-responders - Measurement of patient characteristics and association with treatment response	- Only patients who responded to the treatment in Stage 1 are assessed - N-of-1 and cross-over study designs in stage 2 use patients as their own controls
4. Duration of the study and number of patients required	An efficient study design will get an effective intervention to patients sooner and will aid resource utilization	If sufficient stage 1 improvement is not seen, the trial is terminated	- Interim analysis is performed to justify continuing the trial - Only responders from stage 1 patients are evaluated for treatment effects

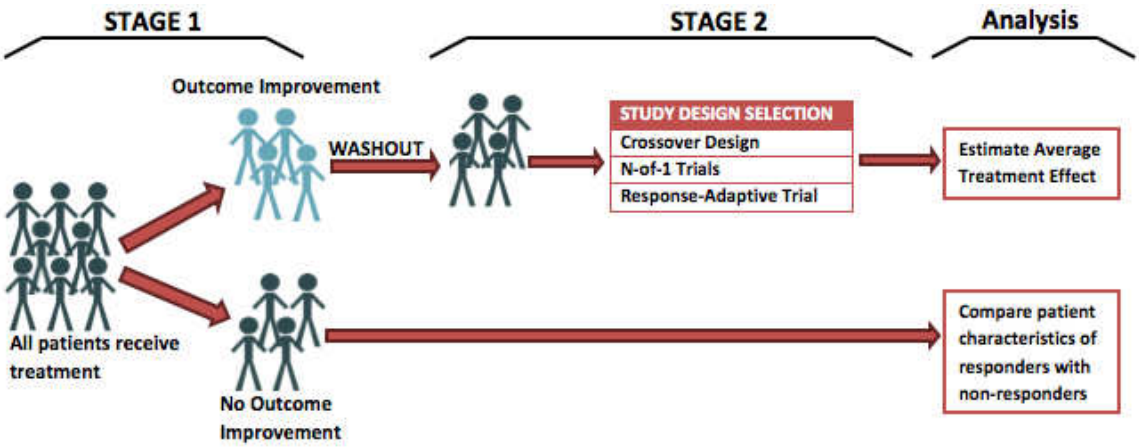


Figure 3-1 Overview of the proposed framework

During Stage 1, all patients receive the experimental treatment and are identified as "responders" or "non-responders" based on their outcome improvement. Following a washout period, responders proceed to Stage 2 for comparative evaluation of treatment effects among them. Analysis has two goals: 1) estimation of the average treatment effect on responsive patients; and 2) comparison of the characteristics of responsive and non-responsive patients.

3.6 References

1. Bashaw ED, Huang S, Cote TR, Pariser AR, Garnett CE, Burckart G, et.al. Clinical pharmacology as a corner-stone of orphan drug development. Nature Rev Drug Discov 2011:10:795-6.
2. Gatta G, Capocaccia R, Trama A, Martinez-Garcia C. The burden of rare cancers in Europe. Adv Exp Med Biol 2010;686:285-303.
3. Dunoyer M. Accelerating access to treatments for rare diseases. Nat Rev Drug Discov 2011;10:475-6.
4. Field MJ, Boat TF. Rare Diseases and Orphan Products: Accelerating Research and Development. Washington: National Academies Press;2010.
5. Haffner ME, Whitley J, Moses M. Two decades of orphan product development. Nat Rev Drug Discov 2002;1:821-5.
6. US Food and Drug Administration. Accelerating orphan drug development. Nat Rev Drug Discov 2010;9;901-2.
7. Abrahamyan L. Designing Randomized Clinical Trials for Rare Diseases [dissertation]. Toronto (ON) : University of Toronto; 2010.
8. Cornu C, Kassai B, Fisch R, Chiron C, Alberti C, Guerrini R, et.al. Experimental design of small clinical trials: an algorithm for choice. J Rare Diseases 2013;8(1):48-60.
9. Edward SJL, Lillford RJ, Braunholtz D, Jackson J. Why "underpowered" trials are not necessarily unethical. Lancet 1997;350:804-7.
10. Evans CH, Ildstad ST. Small Clinical Trials: Issues and Challenges. Washington: National Academy Press;2001.

11. Gerß JWO, Kopcke W. Clinical trials and rare diseases. In: Paz MP, Groft SC, Eds. Rare Diseases Epidemiology (Advances in Experimental Medicine and Biology), New York: Springer;2010:173-90.

12. Gupta S, Faughan ME, Tomlinson GA, Bayoumi AM. A framework for applying unfamiliar trial designs in studies of rare disease. Journal of Clinical Epidemiology 2011;64:1085-94.

13. Johnson SR, Feldman BM, Pope J, Tomlinson GA. Shifting our thinking about uncommon disease trials: The case of methotrexate in scleroderma. The Journal of Rheumatology 2009;36:323-9.

14. Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: A way out of a conundrum. British Medical Journal 1995;311(7020):1621-5.

15. Piantadosi S. Clinical Trials: A Methodological Perspective. Hoboken; John Wiley & Sons, Inc;2005.

16. Baiardi P, Giaquinto C, Girotto S, Manfredi C, Ceci A. Innovative study design for paediatric clinical trials. European Journal of Clinical Pharmacology 2011;67:S109-15.

17. Grieve A, Senn S. Estimating treatment effects in clinical crossover trials. Journal of Biopharmaceutical Statistics 1998;8(2):191-233.

18. Guyatt GH, Heyting A, Jaeschke R, Keller J, Adachi J, Roberts RB. N of 1 randomized trials for investigating new drugs. Controlled Clinical Trials 1990;11:88-100.

19. Piantadosi S. Clinical Trials: A Methodological Perspective. Hoboken; John Wiley & Sons, Inc;2005.

20. Zucker DR, Schmid CH, McIntosh MW, D'Agostino RB, Selker HP, Lau J. Combining single patient (n-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. Journal of clinical epidemiology 1997;50(4):401-10.

21. Benda N, Brannath W, Bretz F, Burger H, Friede T, Maurer W, Wang S. Perspectives on the use of adaptive designs in clinical trials. Part II. Panel discussion. Journal of Biopharmaceutical Statistics 2010;20:1098-112.

22. Berry SM, Carlin BP, Lee JJ, Muler P. Bayesian adaptive methods for clinical trials. New York: CRC Press; 2011.

23. Berry DA, Eick SG. Adaptive assignment versus balanced randomization in clinical trials: A decision analysis. Statistics in medicine 1995;14:231-46.

24. Chow S, Corey R. Benefits, challenges and obstacles of adaptive clinical trial designs. Orphanet Journal of Rare Diseases 2011;6(79):8-10.

25. Chow S, Chang M. Adaptive Design Methods in Clinical Trials. Boca Raton: CRC Press;2011.

26. Chow S, Chang M. Adaptive design methods in clinical trials – a review. Orphanet Journal of Rare Diseases 2008;3(11) :8-10.

27. Wang S. Perspectives on the use of adaptive designs in clinical trials. Part I. Statistical considerations and issues. Journal of Biopharmaceutical Statistics 2010;10:1090-7.

28. Amery W, Dony J. A clinical trial design avoiding undue placebo treatment. Journal of Clinical Pharmacology 1975;15:674-9.

29. Freidlin B, Simon R. Evaluation of randomized discontinuation design. Journal of Clinical Oncology 2005;23:5094-8.

30. Katz N. Enriched enrollment randomized withdrawal trial designs of analgesics. The Clinical Journal of Pain 2009;25:797-807.

31. Kopec JA, Abrahamowicz M, Esdaile JM. Randomized discontinuation trials: utility and efficiency. Journal of Clinical Epidemiology 1993;46(9):959-71.

32. Rosner GL, Stadler W, Ratain MJ. Randomized discontinuation design: application to cytostatic antineoplastic agents. Journal of Clinical Oncology 2002;20(22):4478-84.

33. Schwartz S, Etropolski M, Shapiro DY, Okamoto A, Lange R, Haeussler J, et.al. Safety and efficacy of tapentadol ER in patients with painful diabetic peripheral neuropathy: results of a randomized-withdrawal, placebo-controlled trial. Current Medical Research & Opinion 2011;27(1):151-62.

34. Toth C, Mawani S, Brady S, Chan C, Liu C, Mehina E, Garven A, et.al. An enriched-enrolment, randomized withdrawal, flexible-dose, double-blind, placebo-controlled, parallel assignment efficacy study of nabilone as adjuvant in the treatment of diabetic peripheral neuropathic pain. Pain 2012;153:2073-82.

35. Temple RJ. Enrichment designs: Efficiency in development of cancer treatments. Journal of Clinical Oncology 2005;23(22):4838-9.

36. Temple R. Enrichment of clinical study populations. Clinical Pharmacology and Therapeutics 2010;88(6):774-8.

Chapter 4: Analysis methods and sample size calculations for the patient-focused two-stage study design

4.1 Introduction

An appropriate statistical analysis is a necessary component of a clinical trial [1-2]. Statistical analysis is used to test how likely observed differences between two or more interventions in a clinical trial are due to chance [1-2]. In the previous chapter, we discussed issues and challenges of using rare disease clinical trials to evaluate treatment effect and focused on trial design. In this chapter, we propose an analysis framework as well as a sample size calculation method for the evaluation of treatment effect for a specific study design, in which cross-over design is used in Stage2, proposed in Chapter 3.

As described in Chapter 3, the proposed study design has two stages. In Stage 1, the outcome of interest and patient characteristics are collected at baseline, i.e., before patients receive the experimental treatment, and the outcome is collected again after patients receive the treatment. Patients are considered to be responding to the experimental treatment if their changes in outcome(s) are greater than or equal to a pre-defined threshold δ. Patients who respond to the experimental treatment will continue to the second stage, whereas patients who do not respond to the experimental treatment will not. The second stage is a treatment effect evaluation stage. At this stage, the experimental treatment effect on responding patients is evaluated by comparing it to another treatment (e.g., placebo, standard care). The study design at Stage 2 is proposed to be one of cross-over design, series of n-of-1 trials, and response-adaptive design.

There is a primary analysis and two secondary analyses associated with the proposed study design. The primary analysis is to evaluate the treatment effect of the experimental treatment on responding patients. Since the number of patients who continue to Stage 2 depends on the underlying proportion p in the patient population who respond to the experimental treatment in Stage 1, one of the secondary analyses is to evaluate whether p is at least an expected proportion p_0 in the study population who respond to the experimental treatment. The value of p_0 is specified in the study protocol which may depends on prevalence, social values, and severity of the disease. The other secondary analysis is to perform an exploratory analysis on the association of patients' response to the experimental treatment and their characteristic. In this chapter, we describe an analysis framework and a sample size calculation method for the primary and secondary analyses. For the purpose of illustration, we show the analysis method and the sample size calculation when cross-over design is being used to evaluate treatment effect in Stage 2.

The organization of the chapter is as follows: in section 4.2 we describe analysis methods for the primary analysis and secondary analyses; in section 4.3 we present the sample size calculations for the primary and secondary analyses; section 4.4 discusses a rare disease scenario where the proposed study design and the analysis framework are applicable.

4.2 Proposed Statistical Analysis Framework

Analysis methods presented in this chapter illustrates analysis methods for the primary analysis and two secondary analyses. The organization of this section is as follows: in section 4.2.1 we describe an analysis method for treatment effect evaluation in Stage 2 (primary

analysis); in section 4.2.2 we describe an analysis method for the investigation of the proportion of patients who respond to the experimental treatment in Stage 1 and continue to Stage 2, and an analysis method for an exploratory analysis of the association between patients' response to the experimental treatment and their characteristics (secondary analyses).

4.2.1 Primary Analysis

This section describes an analysis method for the treatment effect evaluation with cross-over design in Stage 2 [1-2].

A crossover design is a repeated measurements design in which each participant receives different treatments during the different time periods of a trial, i.e., participants cross over from one treatment to another during the study [1-2]. As described in the previous chapter, patients in Stage 2 are randomized into two groups. Both groups receive the experimental and standard treatment but in different sequences. There are two periods in Stage 2 when there are an experimental treatment and a standard treatment. At this stage, patients receive their respective first treatment in Period 1 and, followed by a wash out period, they receive their second treatment in Period 2. Patients who are randomized to Sequence 1 receive the experimental treatment in Period 1 and the standard treatment in Period 2: patients who are randomized to Sequence 2 receive the standard treatment in Period 1 and the experimental treatment in Period 2.

Let D^*_{jkl} denote the change in the outcome of l^{th} patient (l=1,2, .., n^*_j, n^*_j is the number of patients in j^{th} sequence) during k^{th} period (k=1,2) at j^{th} sequence (j=1,2). Define $\mu_{experimental}$ and $\mu_{standard}$ be the mean effect of the experimental treatment and that of the standard

treatment, respectively, in the patient population: their difference, θ_d $= \mu_{experimental} - \mu_{standard}$, is the parameter of interest in the primary analysis. Under the following two assumptions, namely, (1) the carry-over effects of the experimental and standard treatment are the same (if $n_1^* = n_2^*$) or the carry-over effects are negligible, and (2) there is no treatment-by-period interaction, θ_d can be estimated by

$$\hat{\theta}_d = \hat{\mu}_{experimental} - \hat{\mu}_{standard}$$

where

$$\hat{\mu}_{experimental} = \left(\sum_{l=1}^{n_1^*} D_{11l}^* + \sum_{l=1}^{n_2^*} D_{22l}^* \right) \Big/ (n_1^* + n_2^*)$$

$$\hat{\mu}_{standard} = \left(\sum_{l=1}^{n_1^*} D_{12l}^* + \sum_{l=1}^{n_2^*} D_{21l}^* \right) \Big/ (n_1^* + n_2^*).$$

[1-2]. The notation τ_d^2, denotes the population variance of the effect size differences between experimental and standard treatments in responding patients [1-2]. ; $\hat{\tau}_d^2$ is estimated by

$$\hat{\tau}_d^2 = \frac{1}{\sum_{j=1}^{2} n_j^* - 1} \sum_{j=1}^{2} \sum_{l=1}^{n_j^*} \left\{ (-1)^j (D_{j2l}^* - D_{j1l}^*) - \hat{\theta}_d \right\}^2$$

[1-2]. The null hypothesis of the stage 2 analysis, H_{02}, is that there is no difference in the mean effects of the two treatments, i.e. H_{02}: $\theta_d = 0$.

A test for the hypothesis H_{02}: $\theta_d = 0$ can be obtained using a test statistic in the form of the t-test statistic as follows:

$$T_d = \frac{\hat{\theta}_d - 0}{\hat{\tau}_d \sqrt{\frac{1}{\sum_{j=1}^{2} n_j^*}}}$$

[1-2]. Under the null hypothesis H_{02}: $\theta_d = 0$, T_d approximately follows a normal distribution with a sufficiently large sample size [1-2]. We reject the null hypothesis H_{02}: $\theta_d = 0$ if $|T_d| > Z_{1-\frac{\alpha_2}{2}}$ where α_2 is the type I error probability of testing the primary hypothesis in Stage 2 [1-2].

4.2.2 Secondary Analyses

This section illustrates an analysis for each of the following analyses: (1) to determine the association between patients' response and patients' characteristic using data collected in Stage 1; and (2) to determine the proportion of patients who respond to the experimental treatment in Stage 1.

4.2.2.1 Evaluation of the Association between Patients' Response and Patients' Characteristic at Stage 1

The proposed study design collects baseline patient characteristics before they receive the experimental treatment in Stage 1 [3]. Without loss of generality, suppose there is only one patient characteristic of interest denoted by X. Let the change in i^{th} patient's outcome after s/he receives the experimental treatment in Stage 1 be D_i. Patient i is considered to be a respondent if D_i is greater than or equal to a threshold δ. We would be interested in assessing if the proportion of patients who respond to the experimental treatment Pr $(D_i \geq \delta | X_i)$ depends on X. This can be tested using a logistic regression model:

$$\log\left(\frac{\Pr\,(D_i \geq \delta | X_i)}{1-\Pr\,(D_i \geq \delta | X_i)}\right) = \gamma_0 + \gamma_1 X_i \qquad \text{for i = 1, 2, …, } n_1$$

where n_1 is the total number of patients enrolled in the trial (Stage 1) [3]. The binary random variable $D_i \geq \delta$ is assumed to follow the Bernoulli distribution with the success probability of $\Pr\,(D_i \geq \delta)|X_i)$. The parameter γ_1 represents the log odds ratio of the outcome change being greater than or equal to δ of those patients who have the characteristic X=1 compared to patients who have the characteristic X=0: γ_0 is the log odds of the outcome change being greater than or equal to δ of those patients who have the characteristic X=0.

A likelihood ratio test can be used to test the null hypothesis that the log odds ratio of the outcome greater than or equal to δ is zero, i.e., $\Pr\,(D_i \geq \delta)|X_i)$ is not associated with patient characteristic $X1$, H_{03}: $\gamma_1 = 0$ [3]. The test statistics follows approximately a chi-square distribution with 1 degree of freedom when n_1 is sufficiently large [3]. The null hypothesis is rejected if $D > \mathrm{X}^2_{1-\alpha_3}$ where α_3 is the type I error probability of this test [3].

4.2.2.2 Evaluation of the Proportion of Patients Responding to Experimental Treatment at Stage 1

This section describes an analysis method to evaluate whether the proportion, p, of patients responding to the experimental treatment in Stage 1 is at least as large as a threshold, p_0^*, pre-specified prior to the study.

As described in the introduction, all n_1 patients at Stage 1 receive the experimental treatment. Patients' outcome is collected before and after they receive the experimental treatment. The change, D_i, in the outcome of i^{th} patient in Stage 1 is measured by the difference of the i^{th} patient's outcome before and after she/he receives the experimental treatment. The

i^{th} patientis considered respondent to treatment if his/her change (D_i) is greater than or equal to a pre-specified threshold, δ.

A one–sample proportion test is used to test the null hypothesis H_{01} that the proportion of patients with greater than δ outcome changes in the population is at most p_0^*, $H_{01}: p \leq p_0^*$, where p can be estimated by

$$\hat{p} = \frac{1}{n_1}\sum_{i=1}^{n_1} I(D_i \geq \delta) \qquad \text{with } I(D_i \geq \delta) = \begin{cases} 1, & D_i \geq \delta \\ 0, & D_i < \delta \end{cases}$$

[1-2] and the variance of $\hat{p}$ can be estimated by $\frac{1}{n_1}\hat{p}(1-\hat{p})$. The test statistics

$$Z = \frac{\hat{p} - p_0^*}{\sqrt{\frac{1}{n_1}p_0^*(1-p_0^*)}}$$

follows approximately the standard normal distribution if $p = p_0^*$ and the sample size n_1 is sufficiently large [1-2]. The null hypothesis is rejected if $Z > Z_{1-\alpha_1}$ where α_1 is the type I error probability of the test [1-2].

4.3 Sample Size Considerations

This section describes a sample size calculation method for rare disease clinical trials that uses the proposed study design.

4.3.1 Primary Analysis

Let β_2 be the type II error and α_2 be the type I error probabilities of the evaluation of the experimental treatment effect in Stage 2 and $n_1^{(1)}$be the number of patients at Stage 1. We need

to ensure a sufficient number of patients in Stage 2 to have $1 - \beta_2$ power. With underlying proportion, p_0, in the patient population who would respond to the experimental treatment in Stage 1, the number of patients, $n_1^{(1)}$, in Stage 1 needs to satisfy the following condition to have $1 - \beta_2$ power in Stage 2:

$$\sum_{x=1}^{n_1^{(1)}} b\left(x; p_0, n_1^{(1)}\right)\left\{1 - \Phi\left(-\theta_d\sqrt{\frac{x}{\tau_d^2}} + Z_{1-\frac{\alpha_2}{2}}\right)\right\} \geq 1 - \beta_2 \qquad (1)$$

where $b\left(x; p_0, n_1^{(1)}\right) = \binom{n_1^{(1)}}{x} {p_0}^x (1 - p_0)^{(n_1^{(1)} - x)}$ denotes the binomial probability mass function of observing x "successes" out of $n_1^{(1)}$ independent trials with a common success probability of p_0.

To calculate the minimum sample size $n_1^{(1)}$ that satisfies this condition, we employ simulation. First input an arbitrary number n_2 for $n_1^{(1)}$and with an assumed p_0 simulate x based on the binomial model. With the simulated x, calculate the power for Stage 2 which is $\left\{1 - \Phi\left(-\theta_d\sqrt{\frac{x}{\tau_d^2}} + Z_{1-\frac{\alpha_2}{2}}\right)\right\}$. If the power is less than $1 - \beta_2$, then increase n_2 and calculate the power with the new value of n_2. Repeat this process again until the calculated power by the above formula is greater than or equal to $1 - \beta_2$ power.

4.3.2 Secondary Analyses

4.3.2.1 Evaluation of the Association between Patients' Response and Patients' Characteristic at Stage 1

Let α_3 be the type I error probability and β_3 be the type II error probability for evaluating the association of treatment response with patient characteristics X in Stage 1. To simplify, we consider the binary case of X. Let $n_{X=1}$ be the number of patients whom their characteristic $X = 1$ in Stage 1; $n_{X=0}$ be the number of patients whom their characteristic $X = 0$ in Stage 1; $n_1^{(3)}$ be the number of patients needed in Stage 1; $p_{X=1}$ be the proportion of patients with $X = 1$ who are responders; $p_{X=0}$ be the proportion of patients with $X = 0$ who are responders; $\bar{p}_1$ be the proportion of responders across all categories of patient characteristic X; and $\bar{p}_2$ be the proportion of non-responders across all categories of patient characteristic X (i.e., $\bar{p}_1 = \frac{p_{X=1}+p_{X=0}}{2}$ and $\bar{p}_2 = 1 - \bar{p}_1$). With a given ratio of $n_{X=1}$ / $n_{X=0}$ as w prior to the study (i.e., $n_{X=1} = \frac{w}{w+1} n_1^{(3)}$ and $n_{X=0} = \frac{1}{w+1} n_1^{(3)}$), the number of patients needed in Stage 1, $n_1^{(3)}$, such that there is sufficient power to reject the null hypothesis $H_0: \gamma_1 = 1.0$ is calculated by

$$n_1^{(3)} = \frac{3(w+1)^2 [Z_{1-\frac{\alpha_3}{2}} + Z_{1-\beta}]^2}{w\gamma_1^2 [1-\bar{p}_1^3-\bar{p}_2^3]}$$

[5].

4.3.2.2 Evaluation of the Proportion of Patients Responding at Stage 1

Let β_1 be the type II error probability and α_1 be the type I error probability of the evaluation of proportion of patents responding at Stage 1. The sample size needed at Stage 1, $n_1^{(2)}$, such that there is sufficient power to reject the null hypothesis $H_{01}: p \leq p_0^*$ where the null hypothesis specifies the value of p_0^*. We assume that the true value of p, i.e., the underlying proportion in the patient population who would respond to the experimental treatment, is p_0. Then the minimal sample size $n_1^{(2)}$ must satisfy:

$$n_1^{(2)} = \frac{[Z_{1-\alpha_1}\sqrt{p_0^*(1-p_0^*)}+Z_{1-\beta_1}\sqrt{p_0(1-p_0)}]}{(p_0-p_0^*)^2}$$

[1,2,4].

4.3.3 Sample size for the proposed study design

There are three analyses in this proposed study design. Therefore, there are three type I errors with probabilities α_1, α_2 and α_3 and three type II errors with probabilities β_1, β_2 and β_3 in this proposed study design. $n_1^{(1)}$ number of patients is needed for the primary analysis. To ensure there is sufficient number of patients for all three analyses of the proposed study design, the minimal sample size needed is $n = \max(n_1^{(1)}, n_1^{(2)}, n_1^{(3)})$.

4.4 Discussion

The purpose of this chapter is to illustrate an analysis framework of the proposed study design described in Chapter 3. The use of a two-stage patient-focused design for RDT can be used not only for treatment effect evaluation, but also exploring the association of patient characteristics and patient's response to treatment. In this chapter, we proposed analysis methods that measure the association between response to treatment and patient characteristics, examine whether the proportion of patients whose outcome changes is at leastp_0^*, and the effect size difference between experimental treatment and the other treatment on responders. We further proposed a sample size calculation method for the primary analysis to have $1-\beta_2$ power and for the two secondary analyses which are described in Section 4.2.2.1 and Section 4.2.2.2 to have $1-\beta_1$ power and $1-\beta_3$ power, respectively.

Since limited rare disease patients are usually available for recruitment into clinical trials, collecting sufficient patients needed for the power and type I error requirement of our proposed study design might only be suitable in certain scenarios. As illustrated in Table 1, an ideal rare disease scenario where this analysis framework would be suitable is (1) all treatments have equal or no carryover effects; (2) outcome is reversible and stable; (3) experimental treatment effect is large compare to standard treatment; (4) patient characteristics are highly associated with being respondent or not; (5) small variability of outcome response to treatments between patients with same known patient characteristics such that variance of outcome is much smaller than the difference in effect between treatments.

For convenience sake, the illustration of the analysis framework is based on the assumptions that the carryover effect of treatments are considered to be the same, no patients dropout, only a patient characteristic is examined at Stage 1, and there is no treatment by period interaction. These assumptions might not hold in practice. This study design is not recommended if treatments' carryover effects are substantially different. Larger sample size might be needed if there is treatment by period interaction and substantial amount of patient dropout.

The proposed study design in the previous chapter with this analysis framework needs to be tested in suitable rare disease treatment evaluation scenarios. It should be noted that, for illustration purposes, our proposed analysis framework uses typical analysis methods and is only for crossover design. Alternative analysis methods have been suggested elsewhere [6-8]. Future work is needed on the development of analysis framework for two other proposed designs at Stage 2.

Table 4-1 Ideal Scenario where the proposed analysis framework is applicable

Ideal Scenario	Reasons
All treatments have equal or no carryover effects	the measurement of treatment effect is not bias by carryover effects
Outcome is reversible	each patient in the study is in similar disease condition when they receive both treatments and hence both treatments' effect on outcomes is comparable
Experimental treatment effect is large compared to standard treatment	less number of patients are required than when the experimental treatment is small to evaluate experimental treatment effect at Stage 2
Large difference in proportion of outcome response to experimental treatment between patients with different known patient characteristics	enrichment at Stage 1 requires small number of patients to examine the association between patient characteristics and the proportion of patients responding to treatment effect
Small variability of outcome response to treatments between patients with same known patient characteristics such that variance of outcome is much smaller than the difference in effect between treatments	only small number of patients are required to evaluate treatment effect at Stage 2

4.5 References

1. Chow SC, Liu JP. Design and analysis of clinical trials: concepts and methodologies. 2nd ed. New Jersey: Wiley; 2004.
2. The Pennsylvania State University. c2014. Stat 509-design and analysis of clinical trials. Available from: https://onlinecourses.science.psu.edu/stat509/node/123.
3. Kleinbaum DG, Mitchel K. Logistic regression:A self-learning text. 3rd ed. New Jersey: Springer; 2012.
4. Julious SA. Sample sizes for clinical trials. New York: CRC Press; 2009.
5. Whitehead J. Sample size calculations for ordered categorical data. Statistics In Medicine 1993; 12:2257-71.
6. Zucker DR, Schmid CH, McIntosh MW, D'Agostino RB, Selker HP, Lau J. Combining single patient (n-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. Journal of clinical epidemiology 1997;50(4):401-10.
7. Benda N, Brannath W, Bretz F, Burger H, Friede T, Maurer W, Wang S. Perspectives on the use of adaptive designs in clinical trials. Part II. Panel discussion. Journal of Biopharmaceutical Statistics 2010;20:1098-112.
8. Chow S, Chang M. Adaptive design methods in clinical trials – a review. Orphanet Journal of Rare Diseases 2008;3(11) :8-10.

Chapter 5: Discussion and Conclusion

Conducting clinical trials to evaluation the effect of treatments for rare diseases can be challenging. As indicated by the US Food & Drug Administration [1] and Health Canada [2], the development of new research methods for rare disease clinical trials is considered one of the key factors towards the improvement of management and treatment of rare diseases. This thesis project proposed a study design and illustrated the study design's analysis framework with sample size calculation methods.

5.1 Synthesis of Results

5.1.1 Chapter 3: A two-stage patient-focused study design for rare disease controlled trials

A two-stage patient-focused study design was proposed for rare disease controlled trials to evaluate treatment effect on reversible and non-reversible outcomes. In addition, the development of this study design also considered other aspects that are relevant to rare diseases, such as the lack of clinically validated outcomes, opportunities for patients to access treatments during treatment development, discovering patient characteristics associated with treatment response among heterogeneous patient population, and duration of the study and number of patients required.

At Stage 1 of the proposed design, all patients receive the experimental treatment. This procedure provides all patients who enroll into the trial to access the experimental treatment.

Since all patients receive the experimental treatment at Stage 1 and their patient characteristics are collected, an exploratory analysis of the association between patient characteristic and patients' outcome is possible. The information regarding the association between outcome response to treatment and patient characteristics could be useful to assign treatment to suitable patients in clinical practice. At the end of Stage 1, if the proportion of patients who respond to the experimental treatment is substantially less than the threshold specified in the study protocol p_0^*, then the trial is terminated. Under this rule, treatments that were not found to benefit clinically important number of patients in the trial will not be continued for treatment effect evaluation at Stage 2.

With a washout period preceding Stage 2, only those patients who respond to the experimental treatment at Stage 1 proceed to Stage 2 and contribute to evaluation of treatment effectiveness. This scheme evaluates whether the experimental treatment is substantially better than the standard treatment at improving the outcome. The suggested choices of designs for evaluating treatment effects on Stage 2 include crossover design, series of n-of-1 trials, and adaptive trials, and all of which provide further opportunities for the participants to receive the experimental treatment than the parallel-group design would. Since patients' feedback on the treatment and its effect is collected at both stages, an exploratory assessment of outcomes in the rare disease under study is possible when clinically validated outcomes are lacking.

5.1.2 Chapter 4: Analysis methods and sample size calculations for the patient-focused two-stage study design

Analysis methods and sample size estimation methods for our proposed study design in which the cross-over design is used at Stage 2 were illustrated. There is a primary analysis and two secondary analyses. The primary analysis of the proposed study design is treatment effect evaluation at Stage 2. Two secondary analyses of the proposed study design are estimation of the proportion of patients who respond positively to the experimental treatment and assessment of the association between patients' outcome response to the experimental treatment and their patient characteristics.

For the primary analyses at Stage 2, the mean difference change in the pre- vs. post-treatment outcome between the experimental treatment and standard treatment of the cross-over trial is estimated. The variance calculation for the mean difference takes into account potential positive correlation of repeated measurements and used in testing whether the null hypothesis that the effect of experimental treatment and standard treatment is the same holds or not: the null hypothesis is rejected if p-value of the test is smaller than or equal to the type I error probability of the test.

The two secondary analyses of Stage 1 are as follows:

1) A one-sample proportion test is used to assess whether the proportion who respond is less than or equal to a pre-specified threshold, p_0. The null hypothesis is rejected if the p-value is smaller than or equal to the type I error probability of this test.

2) A simple logistic regression can be used for assessing the association between patient characteristics and response to the experimental treatment. The null hypothesis that there is no

association between the patient characteristic of interest and response to the experimental treatment is rejected if the p-value of the likelihood ratio test is less than the type I error probability of this test.

A sample size calculation method was proposed for the proposed study design in which cross-over design is used at Stage 2 to have a given power for the primary analysis, along with a specified power for the analysis of the proportion of patients responding to the experimental treatment and another specified power for analysis of the association between the patient characteristic of interest and response to experimental treatment.

5.2 Discussion

This thesis project proposed a patient-focused two-stage study design and an analysis framework with a sample size calculation method. Chapter 3 proposed a study design for treatment effect evaluation for rare diseases where clinically validated outcomes may be lacking, opportunity to access treatment is limited, patients' response to treatment is heterogeneous, and the duration of the study and number of patients required are important considerations. With these features, our study design can be useful to address the following issues of developing rare disease treatments that are illustrated in Chapter 1 and Chapter 2:

(1) Limited opportunities for patients to access treatments: By increasing the opportunity of patients in the trial to receive the experimental treatment also provides more opportunities for patients of those communities to access experimental treatments if they enroll into the trial.

(2) Clinically validated outcomes may be lacking: By using patient reported outcomes (e.g., pain or ability to perform certain activities) in evaluating treatment effect provides investigators measures of patient's quality of life; hence provides information and ways to evaluate treatment effect in a meaningful way for patients.

(3) Patient heterogeneity: By evaluating the association between patient characteristics and their outcome respond to experimental treatment at Stage 1 provides information for prioritization in allocating treatments to patients based on their patient characteristics and future direction on treatment development.

(4) Small number of patients enroll into clinical trials: (i) By distinguish responders and non-responders at Stage 1 and evaluate treatment effect only on responders at Stage 2 might reduce the number of patients needed at Stage 2 compared to a one stage parallel group design; (ii) By analyzing the proportion of patients responding to the experimental treatment at the end of Stage 1 can terminate the trial early if the proportion of patients respond to the experimental treatment is much smaller than a prior anticipated proportion.

By addressing the above issues, this would not only solve technical issues in evaluating treatment effect, but also improve patients' outcomes and utilizing healthcare resources. This is because: (1) more patients having the opportunity to access treatment means more patients who will improve relevant outcomes; (2) our approach's capacity for evaluating treatment effects for rare diseases where clinically validated outcomes is lacking is important for regulatory approval and reimbursement, and enables clinicians to better understanding of treatment effect, which in turn will improve patients' outcome and better use of healthcare resources; (3) allocating treatments to patients who are more likely to respond reduces the risk of incorrectly not providing treatment that is useful to patients and the risk of incorrectly providing treatment that

is not useful to patients; and (4) more efficient study design requires less resources for rare disease treatment development and hence improves the utilization of healthcare resources.

However, to use this study design for rare disease controlled trials, one should be aware of the following limitations:

(1) The investigation of the association between patient characteristics and treatment response in a study is recommended to be exploratory since the association may be specific to some other aspects of the study (e.g., season). The results on the association from Stage 1 should, therefore, cautiously be generalized to the target population. Our suggestion to alleviate this limitation is to ensure sufficient power for identifying clinically important associations in Stage 1 as well as soliciting from patients and care providers potential specific aspects of the trials that may potentially bias the trial results.

(2) It is also possible that the treatment is found to be effective following Stage 2, but there is ambiguity about how the characteristics of responders and non-responders differ. In such a scenario, the trial has a limited capacity to prioritize treatment for future patients based on patient characteristics. However, the information about treatment effects is still valuable for stakeholders in determining the proportion of the target population that is expected to respond to the treatment.

(3) Since the development of this study design did not consider regulatory aspects, modification of the study design might be required to meet specific regulatory evidence requirement.

Chapter 4 proposed an analysis framework and the sample size calculation for the study design. Compare to series of n-of-1 trials and response-adaptive design, crossover design is a relatively common study design. The proposed analysis framework and simple size calculation

method for our proposed study design were illustrated only for crossover design. The secondary analyses methods and sample size calculation for the other two designs can follow the same principles as illustrated in Chapter 4. However, the primary analysis and its sample size calculation of each of those two designs need to be modified since they are different study designs compared to cross-over study design.

For some rare disease clinical trials, only small numbers of patients are available for recruitment; hence other source of information in addition to clinical trials might be needed to evaluate treatment effect. Our proposed analysis framework is a frequentist's approach, Bayesian analysis framework has been proposed to be a more suitable analysis method to evaluate treatment effect in Stage 2. We only highlight the arguments in the literature as to why Bayesian is preferred than the frequentist approach. The frequentist approach is the most common approach to analyze trial data and uses the trial data alone to evaluate treatment effect [3]. Since treatment effect evaluation is based solely on trial data, a substantial number of patients is often needed to evaluate treatment effect [3-5]. Furthermore, since the analysis method of the frequentist approach is closely related to the study design of trials, this in turns limits the ability of the frequentist approach to aggregate information across trials that use different study designs [7]. These features of the frequentist approach can sometimes limit the ability of clinical trials to evaluate treatment effects of certain rare diseases [4].

Unlike the frequentist approach, Bayesian approach uses probability as a measure of belief and a parameter has a probability distribution reflecting degrees of subjective belief over values the parameter might take [8]. Under Bayesian approach, before conducting a trial, a probability distribution is assigned for each parameter to represent the subjective belief on its values based on previous studies and/or expert's opinions [9]. Data from a research study

changes the subjective belief of what the true parameter values could be [9]. Since Bayesian approach can include other information in addition to trial data for making statistical inference on trial data, Bayesian statistical inference usually requires a smaller number of patients than the frequentist approach to evaluate treatment effect. Furthermore, unlike the frequentist approach that depends on trial design to control type I and type II error for statistical inference, Bayesian can make statistical inference as information accrues within a trial or aggregate information across trials relatively more flexibly than the frequentist approach [5]. As a result, under the Bayesian analysis framework, rare disease clinical trials are more feasible [5]. However, there are barriers in using Bayesian analysis for rare disease controlled trials:

(1) challenges in quantifying other sources of information that is qualitative such as expert opinions; and

(2) challenges in assigning weights to different sources of information and opinions of different experts.

5.3 Future research recommendations

Future research aims to extend the development of study designs and analysis methods for rare disease controlled trials to evaluate treatment effect. Specifically, some of the research goals that arose from the work presented in this thesis are:

- The proposed study design has not yet included the perspectives of regulatory agencies. Therefore, a future study on how the proposed design can be modified to meet the regulatory aspects of countries is recommended.

- As indicated in section 4.2, analysis methods and sample size calculation for the primary analysis of our proposed study design that uses series of n-of-1 trials or response-adaptive design at Stage 2.
- Alternative analysis framework for clinical trials such as those of Bayesians (see Appendix) have been suggested in the literature to be a more suitable option for rare diseases controlled trials. Therefore future research using Bayesian framework for our proposed study design is suggested.

5.4 References

1. U.S. Food and Drug Administration. Orphan drug act. Available: http://www.fda.gov/ForIndustry/DevelopingProductsforRareDiseasesConditions/default.htm.
2. Gatta G, Capocaccia R, Trama A, Martinez-Garcia C. The burden of rare cancers in Europe. Adv Exp Med Biol 2010;686:285-303.
3. Gerss JW, Kopcke W. Clinical trials and rare diseases. Adv Exp Med Biol 2010;686:173-90.
4. Gupta S, Faughnan EM, Tomlinson AG, Bayoumi MA. A framework for applying unfamiliar trial designs in studies of rare diseases. J Clin Epidemiol 2011;64: 1085-94.
5. Evans HC, Jr., Ildstad TS. Small clinical trials: issues and challenges. National Academy of Press 2001.
6. Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of conundrum. BMJ 1995; 311: 1621-5.
7. Berry SM, Carlin BP, Lee JJ, Muller. Bayesian adaptive methods for clinical trials. New York: CRC Press; 2011.
8. Clayton D, Hills M. Statistical Models in Epidemiology. New York: Oxford University Press; 1993.

Appendices

Appendix 1: Study Designs in Chapter 3

The following content describes the three study designs that are considered to be used at Stage 2 of the proposed study design, i.e. cross-over, series of n-of 1 trials, and response adaptive design; as well as a study design, randomized withdrawal design, that is compared to the proposed study design in the discussion section of Chapter 3.

A1.1 Cross-over design [1-3]

Description

Patients are sequentially recruited and randomized into two arms in which one arm has the same treatments as the other but in different order. Patients in the study receive a treatment, wash out, and then receive another treatment. During the study, outcome of the patients are collected several times throughout the study at different time point. The difference between the two treatments' outcomes is compared to estimate the overall effect difference between the two treatments.

Time: Since all patients receive multiple treatments and are washed out, this study takes much longer than studies in which each patient receives only a single treatment.

Internal Validity:

- Blinding due to randomization reduces placebo effect.

- Carryover effect might occur if wash out period between the treatment periods is not sufficient.
- Treatment by period effect if present cannot be easily detected due to limited power.
- Overall effect between two treatments estimated in the study is not confounded by prognostic factors.
- 100% receiving the new treatment attracts more patients to enter the study but longer trial duration increases the chance for patients to drop out of the study.

Opportunity to receive new treatment: 100% receiving the new treatment.

Sample size requirement: In this study design, patients are their own control. This allows the treatment difference to be more precisely estimated compare to parallel-groups. Due to the above reasons, the sample size requirement for this study design is less than parallel group design.

This design is appropriate in the scenario which must have:

- Stable disease
- None of the treatments changes the patient's outcome permanently.
- Confounders do not have to be well known or difficult to be controlled for in the study.

This design is recommended if the described scenario above also contains the following conditions:

- The response of patient to treatment is short.
- Patients of the disease have strong desire to receive new treatment.

A1.2 Series of N-of-1 Trials [1,4-6]

Description

Only a patient is required to run the trial. The patient receives at least a pair of treatment (new treatment and placebo or current treatment) multiple times at different time periods in a study. There is a wash out period between each pair of adjacent treatment periods. The trial is begun by randomizing the patient into treatment groups. After every second wash out period since randomization, the patient is randomized again into receiving either pair of treatment. Each result of an n-of-1 trial is only applicable to the particular patient. With a protocol defined on the eligibility criteria to enter an n-of-1 trial, a number of n-of-1 trials can be conducted to generalize the findings to the target population.

Time: Since all patients receive multiple treatments and washed out, this study takes much longer than studies in which each patient receives only a single treatment.

Internal Validity:

- N-of-1 is for one patient. In the end of the trial, it is helpful to identify what treatment is best for the particular patient. A series of n-of-1 trials is required to generalize study results to population.
- Overall effect between two treatments estimated in the study is not confounded by prognostic factors.
- 100% receiving the new treatment attracts more patients to enter the study. This increases internal validity. Longer trial duration increases the chance for patients to drop out of the study.
- Blinding due to randomization reduces placebo effect.

- Carryover effect might occur if wash out period between treatment periods is not sufficient.

Opportunity to receive new treatment: 100% receiving the new treatment.

Sample size requirement: For any trial, only a patient is required. A number of trials are required to generalize the results to the target population.

This design requires:

- Stable disease;
- None of the treatments changes the patient's outcome permanently; and

This design is recommended if the described scenario above also contains the following conditions:

- The response of patient to treatment is short.
- Patients of the disease have strong desire to receive new treatment.
- Confounders do not have to be well known or difficult to be controlled for in the study.
- Disease condition of patient population is sparse and heterogeneous

A1.3 Response Adaptive Randomization Design [1,7-9]

Description

All patients are sequentially recruited and randomized to two arms with probability in consideration of the response of the previous patients. Treatment that is observed to be more effective than the other is given a higher chance of being assigned as the treatment for the

upcoming patient. The intention of this method is to provide the observed better treatment during the study to patients such that more patients benefit from the better treatment in the study.

Time: Since all patients only receive a single treatment, this study takes much less time than studies in which each patient receives at least two treatments.

Internal Validity:

- For ethical reason, patients should be informed that the later they come into the study, the greater is their chance of being assigned to the superior treatment. For this reason, patients may prefer to wait. Patients with more severe conditions are usually enrolled into the study first because they could not wait. This might result early termination of the trial because the treatment effect might be more pronounced in patients with more severe condition.
- Selection bias might occur if investigators guess the assignment of the next patient based on knowing the treatments assigned to the past patients and their outcome.
- Higher opportunity to receive better treatment attracts patients into the study and hence increases internal validity of the study.
- If some of the important covariates in the study is time dependent, such covariates need to be collected and adjusted for in the analysis.

Sample size requirement: Because of sequential enrolment of patients into the trial, sufficient information might be obtained to reach conclusion before the apriori defined number of patients is reached. If the number of patients required for the study to obtain sufficient information can be perfectly anticipated, the sample size required for this design is larger than parallel group design.

This design is recommended if the scenario contains the following conditions:

- Not stable disease
- Treatments that may change the patient's outcome permanently.
- The response of patient to treatment is quick.
- Patients of the disease have strong desire or needs to receive new treatment.
- patients are similar with respect to the important prognostic factors
- negligible changes in the types of patients entering into a trial over time.

A1.4 Randomized Withdrawal Design [1,10-12]

Description

Patients are recruited based on a set of selection criterion. There are two phases. In the first phase, all patients are treated with the new drug. For those who responded to the new treatment are in the response group and the rest are in the non-response group. In the second phase, the response group is randomized into either the new treatment group or current treatment group. The overall treatment effect observed in the two randomized groups is compared in the end of the trial. For the non-response group, they are withdrawn from the study.

Time: All recruited patients can be quickly identified if they are responsive to the new treatment. Non-responsive patients can be quickly identified and be withdrawn from the study. For those that response to the new treatment, the time in the second stage it takes to distinguish the treatment effect from placebo effect is very similar to parallel group design. The duration of the study is longer than parallel group design because of two phases.

Internal Validity:

- The response of patients found in the first phase can be due to placebo effect.
- Carryover effect might occur if wash out period between first phase and second phase is not long enough.
- Treatment by period effect is a serious treat in both stages.
- Misclassification of response and non-response in the first phase affects observed outcome in the second phase.
- 100% receiving the new treatment attracts patients to enter the study.
- Effect found in the second phase can only be generalized to response group

Opportunity to receive new treatment: 100% receiving the new treatment.

Sample size requirement: The sample size requirement to detect the true clinical effect size applies to on the second phase. The patients in the second phase are expected to have baseline characteristics relatively more similar compare to patients in the first phase. Therefore, the estimate of overall treatment effect in the second phase is expected to be more precise. The true overall clinical effect size in the second phase is expected to be larger than regular parallel group design because it is not diluted by those that are non-response to the new treatment. Because of the increase in precision and expected treatment effect difference, if majority of the recruited patient responded to the new treatment, the sample size required for a study with this study design can be potentially less than the parallel group design.

This design is appropriate in the following scenario which must have:

- Stable disease
- None of the treatments completely change the patient's outcome permanently.

This design is recommended if the described scenario above also contains the following conditions:

- Interest in applying new treatment to a subset of patients who are likely to have a certain degree of response to the new treatment.

A1.5 References

1. Institute of Medicine. Small clinical trials: issues and challenges. Washington DC: National Academy Press; 2001.

2. Chow SC, Liu JP. Design and Analysis of Clinical Trials: Concepts and Methodologies. 2nd ed. New Jersey: Wiley; 2004.

3. The Pennsylvania State University. c2014. STAT 509-Design and Analysis of Clinical Trials. Available from: https://onlinecourses.science.psu.edu/stat509/node/123.

4. Gordon H, Guyatt MD, Heyting A, Jaeschke R, Keller J, Adachi J, Roberts RS. N of 1 Randomized Trials for Investing New Drugs. Controlled Clinical Trials 1990; 11:88-100.

5. Zucker DR, Schmid CH, McIntosh MW, Agostino RBD, Selker HP, Lau J. J Clin Epi 1997;50(4):401-10.

6. Zucker DR, Ruthazer R, Schmid CH. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: Methodologic considerations. J Clin Epi 2010; 63(12):1312-23.

7. Chow SC, Chang M. Adaptive Design Methods in Clinical Trials. 2nd Ed. New York: CRC Press; 2012.

8. Zhang L, Rosenberger W. Optimal Response Adaptive Randomization for Clinical Trials. In: Pong A, Chow SC, editors. Handbook of Adaptive Designs in Pharmaceutical and Clinical Development. New York: CRC Press; 2011. p. 15-30.

9. Berry SC, Carlin BP, Lee JJ, Muller P. Bayesian Adaptive Methods for Clinical Trials. New York: CRC Press; 2011. p. 155-167.

10. Kopec JA, Abrahamowicz M, Esdaile JM. Randomized Discontinuation Trials: Utility and Efficiency. J Clin Epidemiol 1993;46(9): 959-71.

11. Amery W, Dony J. A clinical trial design avoiding undue placebo treatment. Journal of clinical pharmacology 1975;15:674-9.

Freidlin B, Simon R. Evaluation of randomized discontinuation design. Journal of Clinical Oncology 2005;23:5094-8.

12. Katz N. Enriched enrollment randomized withdrawal trial designs of analgesics. The clinical Journal of Pain 2009;25:797-807.

Appendix 2: Bayesian Statistics

The following content briefly describes Bayesian Statistics and also compares Bayesian Statistics to Frequentist Statistics.

A2.1 Brief introduction to Bayesian Statistics

In Bayesian statistics, probability is used as a measure of degree of subjective belief [1, 2]. In this paradigm, a probability distribution is used to describe the current belief about a quantity of interest [1, 2]. As more information about the quantity accumulates, its probability distribution changes and eventually converges to a single value, which is the truth. For example, a specific quantity of interest, say the average height of School A students in year 2014. Prior to a study, some background knowledge such as the average height of School A students in the previous years is known. A probability distribution is formulated to describe the uncertainty of the quantity of interest. This probability distribution that is set prior to a study is called prior distribution [3]. After the study is conducted, additional information regarding the average height of School A students in 2014 updates what is known about the average height School A students in 2014. After the update of the prior distribution with additional information, a new probability distribution of the quantity of interest, called posterior probability, is derived to describe what is known about the quantity of interest [3]. Therefore, there is a probability distribution to describe the uncertainty of the quantity of interest before a study is conducted and another probability distribution to describe it after the study is conducted [3]. A prior distribution of uncertainty before studies are modified to posterior uncertainty after studies by new data from the study [3].

As more relevant information accumulates, certain values have higher probabilities than others indicating that these values are believed to be more likely to be the truth than other values.

Evidence from new data is expressed as a likelihood function for the quantity of interest, and the normalized product of the prior and the likelihood combines the subjective belief about parameter values before the study and information from the study. This is expressed as the posterior distribution (i.e., updated belief) of parameter values after studies [4,5]. As more information is accumulated, this posterior distribution will eventually converge to a probability mass of 1.0 on the true parameter value.

A2.2 Comparison of Bayesian and frequentist Approaches

Probabilities of parameters: In frequentist approaches, probabilities are defined only on the sample space. In Bayesian approaches, probabilities are defined on parameter space as well as the sample space [3].

The use of information for statistical inference: Under frequentist approaches, statistical inference regarding parameters of interest is made based only on data from experiments. In Bayesian approaches, statistical inference regarding parameters of interest includes information outside of the experiments and also the experiments [3].

Flexibility: In frequentist approaches, statistical inference is closely related to study design that defines sample space (i.e., data that could have been observed) whereas statistical inference under Bayesian approaches is uses the probabilities of what was observed only [3]. An example to illustrate this concept is shown below:

Suppose that an experimental treatment is being used in a disease for which the historical success rate for standard treatment was 35%. Consider the first of the above two designs, the one that calls for treating exactly 10 patients. An important value of p is less than or equal to 0.35, called the null hypothesis. So when p = 0.35, the probability of the actual observation (7 successes) is 0.021. The conventional frequentist approach is to add in the probabilities of more extreme results - in this case 8, 9, or 10 successes - giving 0.026. This sum is called the significance level, or more briefly, the P-value. In this case the P-value = 0.026 is less than a usual type I error probability 0.05; hence under a typical scenario, this results are called statistically significant.

Had the design of the trial been other than taking exactly 10 observations, then the P-value for these data would be different as well. For example, if the design was to continue the trial until obtaining the third failure, then the P-value would have been 0.004, an order of magnitude smaller than the first P-value. So the evidence is now stronger that the success rate on the experimental treatment is greater than the historical rate, even though the results of the experiment are identical. This close relation between trial design and consequent inferences describes the frequentist approach and showed its inflexibility [3].

A2.3 References

1. Berry DA. Bayesian clinical trials. Nat Rev Drug Discov 2006; 5: 27-36.

2. Rothman KJ. An Introduction to Epidemiology. 2nd ed. New York: Oxford University Press. 2012.

3. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. London: Chapman & Hall; 1995.

4. Berry SM, Carlin BP, Lee JJ, Muller P. Bayesian adaptive methods for clinical trials. New York: CRC Press; 2011.

5. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials. J. R. Statist. Soc. A 1994; 157(3): 357-416.