TWO-STAGE DESIGN FOR PHASE II CANCER CLINICAL TRIALS WITH MULTIPLE ENDPOINTS

By

HUI GU

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Public Health

Written under the direction of

Professor Yong Lin

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2017

ABSTRACT OF THE DISSERTATION

Two-Stage Design for Phase II Cancer Clinical Trials with Multiple Endpoints

by Hui Gu

Dissertation Directors: Professor Yong Lin

The main purpose of a single-arm phase II cancer trial of a new regimen is to determine whether it has sufficient anti-tumor activity against a specific type of tumor to warrant its further clinical development. Such a research question can be answered under the framework of hypothesis testing. With the advent of targeted therapies that prolong disease stabilization, cancer patients typically experience stable disease (SD) rather than tumor shrinkage. It has been shown that patients with SD also achieve clinical benefits. Therefore, when evaluating the anti-tumor activity of a new treatment, clinicians are interested not only in overall response rate (complete or partial response(s)), but also in other types of measurements indicating clinical benefit. Taking two primary efficacy endpoints as an example, if the new treatment can improve on either endpoint(s), it may be promising for further evaluation. Therefore, "OR" logical relationship between the two primary efficacy endpoints is used when specifying the alternative hypothesis. In phase II cancer clinical trials, two-stage designs rather than single-stage ones are widely used for its possibility of early termination for futility to protect cancer patients. Motivated by two real cancer clinical trials, we propose a single-arm two-stage phase II cancer clinical trial design with two dichotomous alternative primary efficacy endpoints. Because of unknown correlation between two endpoints at the design stage, minimax rule is used to determine the optimal

design, which minimizes the maximum of the expected sample size among all possible correlations, subject to the type I and II error constraints. Optimal designs for a variety of design parameters as well as the corresponding operating characteristics are provided. In addition, the statistical inferences of the design are studied. The MLE point estimators as well as confidence regions for the true event rates for the two efficacy endpoints are derived. Three types of confidence regions are obtained by inverting likelihood based test statistics: Wald, Score, and Likelihood ratio statistics. Among the three, the likelihood ratio-type confidence region performs the best in terms of good coverage probability and comparable expected area, and thus is recommended for this two-endpoint two-stage design.

Acknowledgements

I would like to express my sincere appreciation and many thanks to my advisor Dr. Yong Lin, who has been a tremendous mentor for me. His enthusiasm and immense knowledge guided me through all these years of research and writing of this dissertation. I am grateful for his patience and encouragement which motivates me to strive towards my goal and have made this journey much more enjoyable and easier.

I would especially like to thank the rest of my committee members, Dr. Weichung Joe Shih, Dr. Yaqun Wang, and Dr. Kejian Liu for their insightful comments and suggestions, perfecting the ideas and writing of this dissertation. I would also like to thank other faculty members in the Biostatistics department for their encouragement and support during the whole Ph.D. program.

In this journey, I am grateful that I found my Savior and Lord Jesus Christ. God has put so many kind and nice brothers and sisters around, especially in the RBSG cell group, to help me in all aspects of my life, in my difficulties.

Last but not least, I would like to thank my mother and father for their enduring love and support to make this sustained effort possible.

Dedication

To my parents

Aizhong Gu and Qingrong Dou

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Purpose of phase II trials

Phase I trials provide information about the maximum tolerated dose(s) (MTDs) of the treatment and most cancer treatments must be delivered at the MTD for maximum effect. Phase I trials generally treat only 3 to 6 patients per dose level, however, and the patients are diverse with regard to their cancer diagnosis. Consequently such trials provide little or no information about antitumor activity (Simon, 1989). The purpose of a phase IIA trial of a new anticancer drug is to determine whether the drug has sufficient activity against a specific type of tumor to warrant its further clinical development (phases IIB and phase III) (Simon, 1989). The research question of an initial rough estimate of the degree of antitumor activity of the treatment or drug can be answered under the framework of hypothesis testing.

In contrast with phase II designs in other medical fields, these phase IIA cancer trials are usually not performed in a controlled design but as single-arm studies. In early years of cancer treatment when chemotherapy is the main option, the objective response rate (ORR), defined as the proportion of patients whose tumors shrink by at least 50%, was chosen as the primary efficacy endpoint.

Typically, due to ethical considerations, phase II trials in oncology are performed with planned interim analyses to allow early termination for futility to protect patients. In early 1960s, when anticancer agents had low activity, Gehan (1961) proposed a two-stage trial design with a futility rule, which rejected a drug early if there were no observed responses in the first stage. A more general design that allows multistage testing was provided by Fleming (1982). Then Simon (1989) proposed a two-stage design which had either optimal or minimax properties. Due to its easy usage, Simon's two-stage design is widely adopted. There are also a lot of extension work based on Simon's design. For example, Chen's optimal

three-stage designs (Chen, 1997), Jung's admissible designs that balance the optimization criteria of expected sample size and maximum sample size (Jung et al., 2004), Lin and Shih's adaptive version of Simon's two-stage design considering two alternative response rates (either an optimistic or a skeptic target response rate) (Lin and Shih, 2004), Banerjee and Tsiatis' adaptive two-stage designs allowing different stage II sample sizes depending on different stage I results (Banerjee and Tsiatis, 2006), and Englert and Kieser's flexible phase II designs based on the conditional error function principle (Englert and Kieser, 2012). All of these designs are only for a single primary endpoint.

1.2 Concepts of multiple co-primary endpoints vs. multiple primary endpoints (or called 'alternative primary endpoints')

Later on, several authors proposed methods utilizing more than one primary endpoint. Bryant and Day (1995), Thall and Cheng (2001), and Conaway and Petroni (1995) take both efficacy and safety into consideration. The logical relationship between efficacy and safety endpoints is 'AND', which means recommending a new regimen only if it improves on both efficacy and safety endpoints. With the advent of targeted therapies that prolong disease stabilization, there are increasing needs of using multiple efficacy endpoints as alternative primary endpoints of cancer clinical studies. Motivated by two recent cancer clinical studies, in this dissertation we consider multiple measures on efficacy and the logical relationship among the multiple efficacy endpoints is 'OR'.

Most human diseases are characterized by multiple measures, including signs, symptoms, quantitative measurements, and patient-reported outcomes (Offen et al., 2007). And the effects of interventions are also multi-dimensional. In clinical trials, adoption of more than one primary endpoint offers an attractive design feature to capture a more comprehensive characterization of the intervention effects and provide more informative intervention comparisons. For these reasons, nowadays, using more than one primary endpoint has become a common design feature in clinical trials for disease areas such as oncology, infectious disease, and cardiovascular disease (Sozu et al., 2015).

When evaluating an intervention's efficacy, it is customary to classify efficacy endpoints

as primary or secondary. For clarity, two types of multiplicity of the primary endpoints need to be differentiated, that is, "alternative primary endpoints" and "multiple co-primary endpoints". The first type is when an intervention is deemed efficacious if it improves on at least one of the multiple primary endpoints. There may be no consensus on the single most appropriate measure of therapeutic benefit, or it may not be practical to evaluate the therapeutic effect on a unified scale. Take the following as an example of "alternative primary endpoints". A clinical trial to assess the therapeutic benefit in unstable angina patients could include efficacy endpoints such as mortality, myocardial infarction, urgent or emergency coronary revascularization, etc., all of which are of primary clinical interest (Sankoh et al., 1997).

The second type is when an intervention is deemed efficacious only if it improves on all of the multiple primary endpoints. The clinical rationale for adopting "multiple co-primary endpoints" should be clear and should not be due to experts' inability to choose among several endpoints (Offen et al., 2007). Offen et al. (2007) gives a list of disorders (such as Migraine and Alzheimer's disease) known to them for which regulatory agencies have required multiple co-primary endpoints when assessing the effect of an intervention in their Table 1 (Figure 1.1 is an extract of their Table 1).

List of Diseases (Along With Endpoints and Correlation Estimates) for Which Regulatory Agencies Have Required Two or More Co-primary Endpoints — TABLE 1

Disease	Endpoints	Correlation*
Migraine	A. Pain free at 2 hours B. Nausea at 2 hours C. Photosensitivity at 2 hours D. Phonosensitivity at 2 hours	B C D A M M M B L L C M
Acute pain (single-day multiple doses)	A. Patient global assessment B. Sum of pain intensity change over 24 hours C. Sum of pain relief over 24 hours	B C A M H B H
Alzheimer's disease	A. Alzheimer's Disease Assessment Scale–Cognitive (ADAS-Cog) B. Clinician Interview-Based Impression of Change (CIBIC)	B A L
Fibromyalgia	A. Pain reduction B. Patient Global Improvement C. Health Outcome Measure	B C A M M B M
Low back pain	A. Pain intensity (VAS) B. Functional status C. Patient global assessment	B C A M H B M
Osteoarthritis	A. Pain scale B. Patient global assessment C. Function (eg, HRQOL)	B C A H H B H

Figure 1.1: Part of the Table 1 of Offen et al. 2007

1.3 A Motivating Example

In recent years, our paradigm for understanding and treating cancer is changing. Cancers once viewed as relatively homogeneous in terms of organ location and treatment strategy ("one drug fits all") are now better understood to be increasingly heterogeneous across biomarker and genetically defined patient subpopulations (Renfro and Sargent, 2017). Thus, the landscape of cancer treatment is evolving from chemotherapy (the major treatment method when Simon's two-stage design was first proposed) to targeted therapies (including immunotherapies). Standard non-specific chemotherapy agents are cytotoxic (that is, they kill all kinds of cells, both normal cells and tumor cells), while targeted therapies are often cytostatic (that is, they block tumor cell proliferation). In terms of efficacy endpoints, historically, tumor response is an accepted endpoint to assess clinical benefit in phase II trials. Then with the advent of targeted therapies that prolong disease stabilization, patients typically experience stable disease (SD) rather than tumor shrinkage (Mandrekar et al., 2010). It has been shown that patients with SD also achieve clinical benefit (Shepherd et al., 2005), and hence it is not appropriate to ignore SD when assessing treatment efficacy. Another example is the development of the anticancer drug Sorafenib. Sorafenib (NEXAVAR) is a kinase inhibitor indicated for the treatment of unresectable hepatocellular carcinoma and advanced renal cell carcinoma. Clinical studies show that Sorafenib extends progression-free survival (PFS) but the response rate is only 2% (Highlights of prescribing information for Bayer drug NEXAVAR at the url: `https://www.accessdata.fda.gov/drugsatfda_docs/label/2010/021923s008s009lbl.pdf` and Llovet et al. (2008)). So the response rate (CR+PR) as used in Simon's two-stage design may not be appropriate to assess the anti-tumor activity of cytostatic drug such as Sorafenib. Progression-free survival (PFS) rate has now become an accepted alternate endpoint in assessing treatment efficacy as it includes a patient who achieves SD for an extended period of time as a success, in addition to those who achieve complete or partial response (Mandrekar et al., 2010). PFS is defined as the time from registration or randomization to the earlier of disease progression or death from any cause. An ongoing phase II, single arm study assessing the safety and efficacy of single agent CC-486 (oral Azacitidine) in previously treated subjects with locally advanced or

metastatic Nasopharyngeal carcinoma sponsored by Celgene (ClinicalTrials.gov identifier: NCT02269943) is also using two alternative primary efficacy endpoints: overall response rate (ORR) and 6-month PFS rate.

So when evaluating the anti-tumor activity of a new treatment/regimen to a specific type of tumor, clinicians are interested not only in CR+PR (in terms of tumor shrinkage), but also in other types of measurements indicating clinical benefit (for instance, progression-free survival). If the new treatment/regimen can improve on either type of endpoints, it may be promising for further evaluation.

1.4 Research questions and Objectives

The advancement of medicine has made the usage of more than one primary endpoint a common design feature in clinical trials for disease areas such as oncology, infectious disease, and cardiovascular disease. Previously, several authors have proposed two-stage phase II clinical trial designs considering both efficacy and toxicity, which belongs to the category of "multiple co-primary endpoints". Very few studies if any have discussed phase II trial designs with "alternative primary endpoints".

Our research question can be stated specifically as "Does the new treatment have sufficient anti-tumor activity when there are multiple alternative primary endpoints?" Hypothesis testing is one method capable of answering this research question. Recall the process of hypothesis testing and there is one important step of getting the critical values/region of the decision rule given a significance level. These critical values are part of the design parameters for a phase II cancer trial design.

In this dissertation we will develop a single-arm two-stage design for phase II cancer clinical trials with two dichotomous alternative primary endpoints of efficacy. (The two binary efficacy endpoints may come from two different pathways, or from different mechanisms.) The design is capable of detecting activity on either endpoint measure with high probability when the drug or regimen is truly active on one or both measures, and meanwhile be capable of rejecting the drug or regimen (i.e. accepting the null hypothesis) with high probability when the underlying truth is that there is little activity on both measures.

The objectives of this dissertation are:

1. Develop a single-arm two-stage design for phase II cancer clinical trials with two (or more) dichotomous alternative primary endpoints of efficacy.

2. Develop an efficient algorithm to find the optimal designs.

3. Develop methods of joint statistical inferences on the response rates for the two alternative primary efficacy endpoints.

The rest of this dissertation is organized as follows. In Chapter 2, we will review several previous studies that have provided some foundations for this dissertation. The proposed two-endpoint two-stage design for a Phase II cancer trial will be presented in Chapter 3. Statistical inference for the proposed design can be found in Chapter 4. Conclusions and possible future work are summarized in Chapter 5.

Chapter 2

Literature Review

There are quite a lot of papers related to phase II trials, either on one primary endpoint or on bivariate endpoints considering efficacy and safety simultaneously. Bayesian approaches have also been proposed. In the following, three papers on the design perspective for Phase II trials and two papers on inference perspective will be reviewed in details.

2.1 Simon(1989)'s Optimal Two-Stage Designs for Phase II Clinical Trials

In many studies, ethical concerns have led to the development of sequential and group sequential designs for Phase II studies. For instance, Simon (1989) proposed two-stage group sequential designs to determine whether a new anticancer drug has sufficient activity against a specified type of tumor to warrant its further development. Those designs are based on testing a null hypothesis $H_{0:}\ \pi \leq p^{(0)}$ versus an alternative hypothesis $H_{A:}\ \pi \geq p^{(A)}$. A cancer clinical trial following Simon's study designs can be carried out as follows.

- In the first stage of the trial, n_1 patients are enrolled, treated and observed for clinical response. The trial will be terminated at the end of the first stage and the drug will be rejected if r_1 or fewer responses are observed. This occurs with the probability of early termination $PET = B(r_1; n_1, \pi)$ where B denotes the cumulative binomial distribution and π denotes the true probability of response.

- Otherwise, an additional n_2 patients are accrued. The drug will be rejected at the end of the second stage if r or fewer total responses are observed.

After the investigator specifies the uninteresting level of response probability $p^{(0)}$ and the desirable target level $p^{(A)}$ together with type I and II error bounds α and β, the design parameters (n_1, n_2, r_1, r) for the "optimal" two-stage design can be found so as to minimize

the expected sample size under H_0 while satisfying type I/II error constraints. The optimization is taken over all values of n_1 and n_2 as well as r_1 and r. That is, find the design parameters $Q = (n_1, n_2, r_1, r)$ which

$$\text{minimizes } E(N|H_{0:}\ \pi = p^{(0)})$$

subject to

$$Pr(Y_1 > r_1 \text{ and } Y > r|\ \pi = p^{(0)}) \leq \alpha$$
$$Pr(Y_1 > r_1 \text{ and } Y > r|\ \pi = p^{(A)}) \geq 1 - \beta$$

where Y_1 and Y are total number of responses at the end of stage 1 and stage 2, respectively.

For Simon's "minimax" design, the design parameters can be found to minimize the maximum sample size n (where $n = n_1 + n_2$) while satisfying type I/II error constraints.

2.2 Design from Conaway and Petroni(1995)

In Phase II studies, information on treatment safety for the new therapy under investigation is also needed in addition to the preliminary information on treatment's anti-tumor activity. Previous proposed designs have the common feature that the hypothesis testing procedure is based on the response rate of single endpoint. Even though trials based on these designs implicitly consider safety information, sample size determination and stopping rules are based on the single endpoint of most interest, which calls researchers' attention to consider multiple endpoints in group sequential designs of phase II trials. However, many existing articles on multiple endpoints rely on large sample theory to derive their test statistics and stopping rules. Results based on these large sample test statistics may not be applicable to phase II trials with small sample size. Conaway and Petroni (1995) proposed a method for designing group sequential Phase II trials with two dependent binary endpoints, "response" and "toxicity" based on enumerating the exact distribution for the bivariate binary endpoints. They set up the problem to test the null hypothesis that "the new treatment is not sufficiently safe or effective" against the alternative that "the new treatment is sufficiently safe and effective".

Binary variables representing response and toxicity are observed in each of n patients. The data can be summarized in a 2×2 table where X_{ij} is the number of patients with response classification i and toxicity classification j (Table 2.1).

Table 2.1: 2×2 table summarizing data containing two endpoints

		Toxicity		
		Yes	No	
Response	Yes	X_{11}	X_{12}	X_r
	No	X_{21}	X_{22}	$X_{\overline{r}}$
		X_t	$X_{\overline{t}}$	n

Assume that $\mathbf{X} = (X_{11}, X_{12}, X_{21}, X_{22})$ has a multinomial distribution with underlying probabilities $\mathbf{P} = (p_{11},\ p_{12},\ p_{21},\ p_{22})$. The probability of a response is $p_r\ = p_{11} + p_{12}$, and the probability of a toxic event is $p_t\ = p_{11} + p_{21}$. The research hypotheses can be written as

$$H_{0:}\{(p_r, p_t) | p_r \leq p_{r_0} \text{ or } p_t \geq p_{t_0}\}$$

vs.

$$H_{A:}\{(p_r, p_t) | p_r > p_{r_0} \text{ and } p_t < p_{t_0}\}$$

where p_{r_0} is the response rate for the standard treatment, and p_{t_0} is the toxicity rate for the standard treatment. Their method for determining sample size is based on choosing a test statistic, T, and desired levels of type I error and type II error at a particular point in the alternative space, $p_r = p_{r_a}$ and $p_t = p_{t_a}$, under an assumed association between response and toxicity. The bivariate test they use is based on the joint distribution of the random variables (X_r, X_t), where $X_r = X_{11} + \ X_{12}$ and $X_t = X_{11} + \ X_{21}$. Odds ratio θ was chosen to describe the association between response and toxicity because it is a natural measure of association in 2×2 tables, and it does not depend on the marginal probabilities.

The critical values defining the reject region, and the sample size, n, are chosen to satisfy the following error requirements:

$$P[\text{reject } H_0 | (p_{r_0}, p_{t_0}), \theta] \leqslant \alpha,$$

$$\sup_{p_r \leqslant p_{r_0} \ or \ p_t \geqslant p_{t_0}} P[\text{reject } H_0 | (p_r, p_t), \theta] \leqslant \gamma,$$

$$P[\text{reject } H_0 | (p_{r_a}, p_{t_a}), \theta] \geqslant 1 - \beta.$$

Both α and γ represent type I error rates:

- α provides a bound on the type I error at a particular point (p_{r_0}, p_{t_0}), for a given θ; while
- γ bounds the maximum type I error over the entire null hypothesis region.

Typically $\alpha = 0.05, \gamma = 0.30$, and $\beta = 0.10$ in their proposed designs. For a fixed sample size n (single-stage design), the event "reject H_0" $=\{ X_r \geqslant c_r, X_t \leqslant c_t \}$ and the critical values (c_r, c_t) are found by enumerating the joint distribution of (X_r, X_t). Take for example a proposed phase II trial of high dose chemotherapy for patients with non-Hodgkin's lymphoma. Earlier studies results for this patient population have indicated that standard therapy results in an estimated response rate of 50% with approximate 30% of the patients experiencing life-threatening toxicities. In addition, previous results indicated that approximate 35-40% of the patients who experienced a complete response also experienced life-threatening toxicities. So (p_{r_0}, p_{t_0}) is assumed to be (0.5, 0.3) and the odds ration to be 2. With the desired power of 90% and a particular point in the alternative space $(p_{r_a}, p_{t_a}) = (0.75, 0.15)$, by enumerating the distribution of $(X_r, X_t), n = 30$ was found to be the smallest sample size that yielded a rule satisfying all the error requirements. The corresponding rule has (c_r, c_t)=(7,18) and rejection probabilities

$$P[X_r \geqslant c_r, X_t \leqslant c_t \ | (p_{r_0}, p_{t_0}), \theta] = 0.0377,$$

$$P[X_r \geqslant c_r, X_t \leqslant c_t \ | (p_{r_a}, p_{t_a}), \theta] = 0.9095,$$

$$\sup_{p_r \leqslant p_{r_0} \ or \ p_t \geqslant p_{t_0}} P[X_r \geqslant c_r, X_t \leqslant c_t \ | (p_r, p_t), \theta] = 0.2814.$$

They also extended the procedure to allow for early termination of a study if early results indicate that the treatment is not sufficiently effective or is too toxic (multiple-stage design). Examples were given to illustrate the two- and three-stage designs in their paper. This design focused on derivation of stopping rules for phase II trials with two dependent

binary endpoints, where a new treatment is recommended when it is both sufficiently safe and effective. The logical relationship between the two dependent endpoints in the alternative hypothesis is "AND". When searching for an optimal two-stage design, they used a particular value of odds ratio to describe the association between the two dependent endpoints. They have used two types of type I error rates in defining the critical values: α, a bound on type I error at a particular value of (p_r, p_t) corresponding to the current standard therapy and γ, the global bound on the maximum type I error over the entire null hypothesis region.

2.3 Design from John Bryant and Roger Day (1995)

Bryant and Day (1995) proposed a modified two-stage design incorporating toxicity considerations into a Simon design by requiring that the trial be terminated after the initial stage if **either** the number of observed responses is inadequate **or** the number of observed toxicities is excessive. Otherwise, the treatment under investigation is recommended following the second stage only if there are both a sufficient number of responses and an acceptably small number of toxicities in total. Let Y_{R1} and Y_{R2} denote the total number of clinical responses at the end of stage 1 and stage 2, respectively; Y_{T1} and Y_{T2} the total number of patients who **do not** experience toxicity at the end of stage 1 and stage 2, respectively; and C_{R1} and C_{R2} the critical boundaries for response while C_{T1} and C_{T2} the critical boundaries for **nontoxicity**. The trial proceeds as follows:

- Accrue N_1 patients in the first stage of the trial. Only if $Y_{R1} > C_{R1}$ and $Y_{T1} > C_{T1}$, continue to the second stage. Otherwise, terminate the trial early at the end of stage 1.

- In stage 2, accrue $(N_2 - N_1)$ additional patients. At the end of stage 2, recommend the treatment for further consideration only if we observe $Y_{R2} > C_{R2}$ and $Y_{T2} > C_{T2}$.

The directions in the decision rule for each endpoint are the same under this setting. The authors considered the following four possible states of nature hypotheses $H_{ij} : P_R = P_{Ri},\ P_T = P_{Tj},\ i = 0, 1;\ j = 0, 1$, where P_R, P_{R0}, P_{R1} denote the true response rate, the

investigator-specified "unacceptable" response rate, and the "acceptable" response rate ($0 < P_{R0} < P_{R1} < 1$), respectively; while P_T, P_{T0}, P_{T1} denote the true rate of nontoxicity, the investigator-specified "unacceptable" rate of nontoxicity, and the "acceptable" nontoxicity rate, respectively:

$$H_{00} : P_R = P_{R0},\ P_T = P_{T0}$$

$$H_{01} : P_R = P_{R0},\ P_T = P_{T1}$$

$$H_{10} : P_R = P_{R1},\ P_T = P_{T0}$$

$$H_{11} : P_R = P_{R1},\ P_T = P_{T1}$$

The design is specified by a vector $\mathbf{Q} = (N_1, N_2, C_{R1}, C_{R2}, C_{T1}, C_{T2})$. To determine the design parameter vector $\mathbf{Q}$ based on these hypotheses, the test statistics and the distribution of the test statistics are needed. To fully specify the joint distribution of response and toxicity under any of the above states of nature hypotheses, an additional parameter describing the association between response and toxicity is required in addition to the parameters P_R and P_T. This association between response and toxicity can be parameterized by the odds ratio $\phi = p_{00}p_{11}/p_{01}p_{10}$, where p_{00} is the proportion of patients who do not respond and who do become toxic, p_{01} is the proportion who do not respond and do not experience toxicity, p_{10} is the proportion who respond and who do become toxic, and p_{11} is the proportion who respond without toxicity.

For any study design $\mathbf{Q}$, denote by $E_{ij} = E_{ij}(\mathbf{Q}, \phi)$ the expected number of patients accrued, given state of nature H_{ij}: $i = 0, 1$; $j = 0, 1$. Let $\alpha_R > 0, \alpha_T > 0$, and $\beta > 0$ be error bounds specified by the investigators: α_R is an upper bound on the probability of erroneously recommending a treatment whose response rate is inadequate; α_T bounds the probability of erroneously recommending a treatment that is unacceptably toxic; and β is a bound on the probability of failing to recommend a treatment that is acceptable with respect to both response and toxicity. Define

$$\alpha_{ij}(\mathbf{Q}, \phi) = Pr\{\text{Recommend Treatment}|H_{ij}, \mathbf{Q}, \phi\},$$

and

$$\psi(\mathbf{Q}, \phi) = \max\{E_{01}(\mathbf{Q}, \phi), E_{10}(\mathbf{Q}, \phi)\}.$$

If nothing is assumed a priori concerning the value of ϕ, the design parameters $\mathbf{Q}$ may be determined by solving the mathematical program

$$\min_{\mathbf{Q}} \max_{\phi>0} \psi(\mathbf{Q}, \phi), \tag{2.1}$$

subject to

$$\max_{\phi>0} \alpha_{01}(\mathbf{Q}, \phi) \leq \alpha_R, \tag{2.2a}$$

$$\max_{\phi>0} \alpha_{10}(\mathbf{Q}, \phi) \leq \alpha_T, \tag{2.2b}$$

$$\min_{\phi>0} \alpha_{11}(\mathbf{Q}, \phi) \geq 1 - \beta. \tag{2.2c}$$

E_{00} does not appear in the object function ψ, because it is less than either E_{01} or E_{10}; and likewise the error rate α_{00} does not appear in the constraints (2.2a - 2.2c), because it is less than either α_{01} or α_{10}.

The functions E_{ij} and α_{ij} may be evaluated using a certain bivariate binomial distribution $\Pr\{Y_R = y_R, Y_T = y_T\}$ under an assumption of independent and identically distributed accruals. Consider an independent sequence of n patients, each having probability P_R for clinical response and P_T for nontoxicity. Let $P_{R|1} = p_{11}/P_T$, $P_{R|0} = p_{10}/(1 - P_T)$. Let Y_R and Y_T denote the total number of responses and nontoxicities, respectively. Denote by $b(\cdot\,; n, P)$ the binomial probability function with parameters n and P, and define

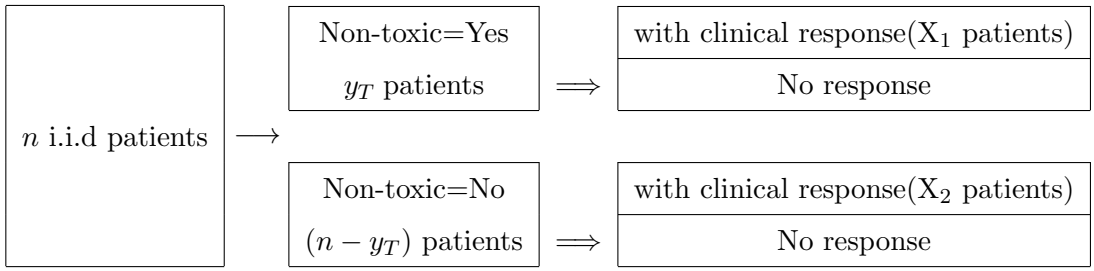
$$\overline{B}(y; n, P) = \sum_{j>y} b(j; n, P) = \Pr(Y > y | Y \sim \mathit{binom}(n, P)).$$

Denote the joint probability function $\Pr\{Y_R = y_R, Y_T = y_T\}$ by $d(y_R, y_T; n, P_R, P_T, \phi)$. A straightforward conditioning argument shows that

$$d(y_R, y_T; n, P_R, P_T, \phi) = \Pr\ \{Y_R = y_R, Y_T = y_T\}$$

$$= Pr(Y_T = y_T)Pr(Y_R = y_R|Y_T = y_T) \tag{2.3}$$

The idea of this conditioning can be sketched as:



where $Y_R = X_1 + X_2 = y_R$, $X_1 \sim Bin(y_T, P_{R|1})$, and $X_2 \sim Bin(n - y_T, P_{R|0})$.

Based on the above sketch, the last conditional probability in (2.3) can be expanded as follows:

$$\begin{aligned}
&Pr(Y_R = y_R|Y_T = y_T) \\
=&Pr(X_1 + X_2 = y_R|Y_T = y_T) \\
=&\sum_{\max(0,y_R+y_T-n)\leqslant j\leq \min(y_R,y_T)} Pr(X_1 = j|Y_T = y_T)Pr(X_2 = y_R - j|Y_T = y_T) \\
=&\sum_{\max(0,y_R+y_T-n)\leqslant j\leq \min(y_R,y_T)} b(j; y_T, P_{R|1})b(y_R - j; n - y_T, P_{R|0}).
\end{aligned}$$

So the above equation (2.3) can be further written as follows:

$$\begin{aligned}
&d(y_R, y_T; n, P_R, P_T, \phi) = \Pr\{Y_R = y_R, Y_T = y_T\} = Pr(Y_T = y_T)Pr(Y_R = y_R|Y_T = y_T) \\
&= b(y_T; n, P_T) \sum_{\max(0,y_R+y_T-n)\leq j\leq \min(y_R,y_T)} b(j; y_T, P_{R|1})b(y_R - j; n - y_T, P_{R|0}).
\end{aligned} \tag{2.4}$$

Now define

$$\overline{D}(y_R, y_T; n, P_R, P_T, \phi) = Pr(Y_R > y_R, Y_T > y_T) = \sum_{u>y_R} \sum_{v>y_T} d(u, v; n, P_R, P_T, \phi).$$

$$E_{ij}(\mathbf{Q}, \phi) = N_1 + (N_2 - N_1)\overline{D}(C_{R1}, C_{T1}; N_1, P_{Ri}, P_{Tj}, \phi) = N_1 PET + N_2(1 - PET).$$

Similarly, conditioning on the number of responses and nontoxicities observed during stage

1 leads to

$$\begin{aligned}
\alpha_{ij}(\mathbf{Q},\phi) =& Pr\{\text{Recommend Treatment}|H_{ij},\mathbf{Q},\phi\} \\
=& Pr(Y_{R1}>C_{R1},Y_{T1}>C_{T1},Y_{R2}>C_{R2},Y_{T2}>C_{T2}) \\
=& \sum_{y_{R1}>C_{R1}}\sum_{y_{T1}>C_{T1}}\{Pr(Y_{R2}-Y_{R1}>C_{R2}-y_{R1},Y_{T2}-Y_{T1}>C_{T2}-y_{T1}|Y_{R1}=y_{R1}, \\
& Y_{T1}=y_{T1})\times Pr(Y_{R1}=y_{R1},Y_{T1}=y_{T1})\} \\
=& \sum_{y_{R1}>C_{R1}}\sum_{y_{T1}>C_{T1}}\overline{D}(C_{R2}-y_{R1},C_{T2}-y_{T1};N_2-N_1,P_{Ri},P_{Tj},\phi) \\
& \times d(y_{R1},y_{T1};N_1,P_{Ri},P_{Tj},\phi) \qquad (2.5)
\end{aligned}$$

The mathematical programs (2.1 - 2.2c) will be considered in greater details based on the following theorems.

In the case where response and toxicity are assumed to be independent, the following Theorem shows that the computations required to evaluate E_{ij} and α_{ij} factor in a convenient way.

Theorem 2.3.1 *Let* $\boldsymbol{N}=(N_1,N_2)$, $\mathbf{C}_R=(C_{R1},C_{R2})$, *and* $\mathbf{C}_T=(C_{T1},C_{T2})$. *Let* C_1 *and* C_2 *be any integers, and denote* $\mathbf{C}=(C_1,C_2)$. *Define*

$$\alpha^*(\mathbf{N},\mathbf{C},P)=\sum_{y>C_1}b(y;N_1,P)\overline{B}(C_2-y;N_2-N_1,P).$$

Then in the special case of $\phi=1$,

$$E_{ij}(\mathbf{Q},1)=N_1+(N_2-N_1)\overline{B}(C_{R1};N_1,P_{Ri})\overline{B}(C_{T1};N_1,P_{Tj}),$$

and

$$\alpha_{ij}(\mathbf{Q},1)=\alpha^*(\mathbf{N},\mathbf{C}_R,P_{Ri})\alpha^*(\mathbf{N},\mathbf{C}_T,P_{Tj}),i=0,1;j=0,1.$$

The above Theorem 2.3.1 indicates that in the special case of independence ($\phi=1$), the calculations of expected sample sizes and error rates are much simplified and can be factored into product of simple binomial distributions.

When response and toxicity are assumed independent, Theorem 2.3.1 allows the reduction of the mathematical programs (2.1-2.2c) to:

$$\min_{\mathbf{Q}} \psi(\mathbf{Q}, 1), \tag{2.6}$$

subject to

$$\alpha_{01}(\mathbf{Q}, 1) = \alpha^*(\mathbf{N}, \mathbf{C}_R, P_{R0})\alpha^*(\mathbf{N}, \mathbf{C}_T, P_{T1}) \leq \alpha_R, \tag{2.7a}$$

$$\alpha_{10}(\mathbf{Q}, 1) = \alpha^*(\mathbf{N}, \mathbf{C}_R, P_{R1})\alpha^*(\mathbf{N}, \mathbf{C}_T, P_{T0}) \leq \alpha_T, \tag{2.7b}$$

$$\alpha_{11}(\mathbf{Q}, 1) = \alpha^*(\mathbf{N}, \mathbf{C}_R, P_{R1})\alpha^*(\mathbf{N}, \mathbf{C}_T, P_{T1}) \geq 1 - \beta. \tag{2.7c}$$

In the case where the odds ratio ϕ is unspecified, optimal design parameters are found by solving the mathematical programs (2.1-2.2c). The required computations are simplified by the fact that E_{ij} and α_{ij} are monotonic functions of ϕ:

Theorem 2.3.2 *For $i = 0, 1$; $j = 0, 1$ and for fixed $\boldsymbol{Q}$, $\alpha_{ij}(\mathbf{Q}, \phi)$ and $E_{ij}(\mathbf{Q}, \phi)$ are nondecreasing continuous functions of ϕ.*

Theorem 2.3.2 allows the reduction of mathematical programs (2.1-2.2c) to:

$$\min_{\mathbf{Q}} \psi(\mathbf{Q}, \infty), \tag{2.8}$$

subject to

$$\alpha_{01}(\mathbf{Q}, \infty) \leq \alpha_R, \tag{2.9a}$$

$$\alpha_{10}(\mathbf{Q}, \infty) \leq \alpha_T, \tag{2.9b}$$

$$\alpha_{11}(\mathbf{Q}, \ 0) \geq 1 - \beta. \tag{2.9c}$$

For either $\phi = 0$ or $\phi = \infty$, the calculation of the joint probability function $d(\cdot, \cdot; n, P_R, P_T, \phi)$ in (2.3) reduces to a product of binomial terms, thus simplify the computations required for mathematical programs (2.1-2.2c).

As long as β is moderately small, designs obtained by solving the mathematical programs

(2.8-2.9c) will have operating characteristics that differ negligibly from those obtained under the independence assumption $\phi = 1$ (mathematical programs (2.6-2.7c)). The following Theorem 2.3.3 will show that any $\mathbf{Q}$ feasible under the assumption that $\phi = 1$ will also be feasible, or nearly so, for all $\phi > 0$. Theorem 2.3.4 will show that the optimality criteria $\psi(\mathbf{Q}, 1)$ and $\underset{\phi>0}{max}\psi(\mathbf{Q}, \phi) = \psi(\mathbf{Q}, \infty)$ are nearly equal for designs which are feasible for $\phi = 1$.

Theorem 2.3.3 *If $\boldsymbol{Q}$ satisfies the constraints (2.7a-2.7c), then*

$$\alpha_{01}(\mathbf{Q}, \infty) = \max_{\phi>0} \alpha_{01}(\mathbf{Q}, \phi) \leq \alpha_R/(1-\beta), \tag{2.10a}$$

$$\alpha_{10}(\mathbf{Q}, \infty) = \max_{\phi>0} \alpha_{10}(\mathbf{Q}, \phi) \leq \alpha_T/(1-\beta), \tag{2.10b}$$

$$\alpha_{00}(\mathbf{Q}, \infty) = \max_{\phi>0} \alpha_{00}(\mathbf{Q}, \phi) \leq \min(\alpha_R, \alpha_T)/(1-\beta), \tag{2.10c}$$

$$\alpha_{11}(\mathbf{Q}, 0) = \min_{\phi>0} \alpha_{11}(\mathbf{Q}, \phi) \geq 2(1-\beta)^{1/2} - 1. \tag{2.10d}$$

Theorem 2.3.4 *Suppose that $\boldsymbol{Q}$ satisfies the constraints (2.7a-2.7c). Define*

$$R(\mathbf{Q}) = \{\underset{\phi>0}{max}\ \psi(\mathbf{Q}, \phi) - N_1\}/\{\psi(\mathbf{Q}, 1) - N_1\},$$

$$U_1(\mathbf{Q}) = 1/\min\{\overline{B}(C_{R1}; N_1, P_{R1}), \overline{B}(C_{T1}; N_1, P_{T1})\}$$

and

$$U_2(\mathbf{Q}) = 1/\Pr\{\textit{Continue to stage 2}|H_{11}, \mathbf{Q}, \phi = 1\}.$$

Then,

$$1 \leq R(\mathbf{Q}) \leq U_1(\mathbf{Q}) \leq U_2(\mathbf{Q}) \leq 1/(1-\beta).$$

2.4 Point estimation of the binomial probability in clinical trials with a single-endpoint, two-stage design

Although the primary goal of Phase II trials is decision making (terminate the trial or continue further to Phase III) rather than inference, obtaining an estimate of the true response rate p is often of interest, particularly when the trial was deemed successful and the new drug accepted for further evaluation in Phase III trials (Porcher and Desseaux, 2012). The most common estimator of p is the sample response rate, which is also the Maximum Likelihood Estimator (MLE) in a single-endpoint, two-stage design. Take a two-stage design allowing early termination for futility only for example.

Notations ($k = 1, 2$):

- n_k: number of patients accrued during stage k;
- X_k: number of responders during stage k;
- S_k: cumulative number of responders by stage k, $S_k = \sum_{i=1}^{k} X_i$;
- a_k: lower stopping boundaries for stage k;
- M: stopping stage.

The MLE for a two-stage design is:

$$\begin{aligned}\widehat{p} &= \begin{cases} \frac{X_1}{n_1}, & \text{stop at stage 1 } (m = 1, 0 \le s \le a_1) \\ \frac{X_1+X_2}{n_1+n_2}, & \text{stop at stage 2 } (m = 2, a_1 + 1 \le s \le n_1 + n_2) \end{cases} \\ &= \frac{X_1}{n_1} I\left(X_1 \le a_1\right) + \frac{X_1 + X_2}{n_1 + n_2} I\left(X_1 > a_1\right).\end{aligned}$$

In single-stage trials, the sample proportion is still unbiased for p. However, due to the sequential nature of the two-stage trials, the MLE is biased. By definition, the bias of the MLE is:

$$bias(\widehat{p}|p) = E\left(\widehat{p}|p\right) - p.$$

There are several ways to derive the bias of the MLE. Two ways are presented here: one through the use of indication function, and the other one based on the joint probability

mass function of a complete and sufficient statistic for p.

Method 1 to derive the bias of MLE(analytically)

$$\begin{aligned}
\widehat{p} &= \frac{X_1}{n_1} I\left(X_1 \leq a_1\right) + \frac{X_1 + X_2}{n_1 + n_2} I\left(X_1 > a_1\right) \\
&= \frac{X_1}{n_1} + \left(\frac{X_1 + X_2}{n_1 + n_2} - \frac{X_1}{n_1}\right) I\left(X_1 > a_1\right) \\
&= \frac{X_1}{n_1} + \frac{n_1 X_1 + n_1 X_2 - \left(n_1 + n_2\right) X_1}{\left(n_1 + n_2\right) n_1} I\left(X_1 > a_1\right) \\
&= \frac{X_1}{n_1} + \frac{n_1 X_2 - n_2 X_1}{\left(n_1 + n_2\right) n_1} I\left(X_1 > a_1\right) \\
E\left(\widehat{p}|p\right) &= E\left(\frac{X_1}{n_1}\right) + E\left(\frac{n_1 X_2 - n_2 X_1}{\left(n_1 + n_2\right) n_1} I\left(X_1 > a_1\right)\right) \\
&= p + \frac{n_2 p}{n_1 + n_2} E\left(I\left(X_1 > a_1\right)\right) - \frac{n_2}{\left(n_1 + n_2\right) n_1} E\left(X_1 I\left(X_1 > a_1\right)\right) \\
&= p + \frac{n_2}{\left(n_1 + n_2\right) n_1} E\left(n_1 p I\left(X_1 > a_1\right)\right) - \frac{n_2}{\left(n_1 + n_2\right) n_1} E\left(X_1 I\left(X_1 > a_1\right)\right) \\
&= p - \frac{n_2}{\left(n_1 + n_2\right) n_1} E\left(\left(X_1 - n_1 p\right) I\left(X_1 > a_1\right)\right) \\
&= p - \frac{n_2}{\left(n_1 + n_2\right) n_1} \sum_{x_1 > a_1}^{n_1} \left(x_1 - n_1 p\right) \left(\begin{array}{c} n_1 \\ x_1 \end{array} \right) p^{x_1} (1-p)^{n_1 - x_1} \\
bias(\widehat{p}|p) &= -\frac{n_2}{\left(n_1 + n_2\right) n_1} \sum_{x_1 > a_1}^{n_1} \left(x_1 - n_1 p\right) \left(\begin{array}{c} n_1 \\ x_1 \end{array} \right) p^{x_1} (1-p)^{n_1 - x_1}
\end{aligned}$$

So the bias of the MLE is always negative.

Method 2 to derive the bias of MLE(suitable for numerical calculations)

Let M denote the stage at which a trial is terminated and S denote the number of responders at Stage M. Jung and Kim (2004) showed that (M, S) is a complete and sufficient statistic for p in multistage designs such as Simon's two stage design. They also derived the probability mass function of the random vector (M, S) as: $f\left(m, s|p\right) = \Pr(M = m, S = s|p)$. The details are as follows.

When m=1,

$$f\left(m,s|p\right)=\Pr(M=1,S=s|p)=\Pr(X_1=s|p)=\left(\begin{array}{c} n_1 \\ s \end{array}\right)p^s(1-p)^{n_1-s}$$

When m=2,

$$\begin{aligned}
f\left(m,s|p\right) &= \Pr(M=2,S=s|p) \\
&= \Pr(S_2=s, a_1+1\le S_1\le n_1) \\
&= \Pr(X_1+X_2=s, a_1+1\le X_1\le n_1) \\
&= \sum_{x_1=a_1+1}^{\min(n_1,s)} \Pr(X_1+X_2=s, X_1=x_1) \\
&= \sum_{x_1=a_1+1}^{\min(n_1,s)} \Pr(X_2=s-x_1)\Pr(X_1=x_1) \\
&= \sum_{x_1=a_1+1}^{\min(n_1,s)} \left(\begin{array}{c} n_2 \\ s-x_1 \end{array}\right)p^{s-x_1}(1-p)^{n_2-s+x_1}\left(\begin{array}{c} n_1 \\ x_1 \end{array}\right)p^{x_1}(1-p)^{n_1-x_1} \\
&= p^s(1-p)^{n_1+n_2-s}\sum_{x_1=\max(a_1+1,s-n_2)}^{\min(n_1,s)} \left(\begin{array}{c} n_2 \\ s-x_1 \end{array}\right)\left(\begin{array}{c} n_1 \\ x_1 \end{array}\right) \\
bias(\widehat{p}|p) &= E\left(\widehat{p}|p\right)-p \\
&= \sum_{\text{all possible (m, s)}} \widehat{p}(m,s)f(m,s|p)-p \\
&= \frac{1}{n_1}\sum_{s=0}^{a_1} sf(1,s|p)+\frac{1}{n_1+n_2}\sum_{s=a_1+1}^{n_1+n_2} sf(2,s|p)-p
\end{aligned}$$

Since unbiasedness is a desired feature, Jung and Kim (2004) derived the uniformly minimum variance unbiased estimator (UMVUE) of p in multistage clinical trials. Considering that the sample proportion after the first stage, $\widehat{p_1}=\frac{X_1}{n_1}$, is an unbiased estimator of p, and based on the Rao-Blackwell theorem, they obtained the UMVUE by taking the conditional expectation of an unbiased estimator (the first stage sample proportion $\widehat{p_1}=\frac{X_1}{n_1}$) given a complete and sufficient statistic $(M,S)=(m,s)$, where m and s denote specific observations

of random variables M and S, respectively. So the UMVUE of p is obtained by

$$\begin{aligned}\widetilde{p} &= E\{\widehat{p_1}|(m,s)\} = E\{X_1|(m,s)\}/n_1 \\ &= \frac{1}{n_1}\sum x_1 \Pr(X_1 = x_1|(m,s))\end{aligned}$$

When $M = 1$, $\widetilde{p} = \widehat{p_1}$.

When $M = 2$, $for\ a_1 + 1 \le x_1 \le n_1$,

$$\begin{aligned}&\Pr(X_1 = x_1|(m,s)) \\ &\quad= \frac{\Pr(X_1 = x_1, M = m, S = s|p)}{\Pr(M = m, S = s|p)} \\ &\quad= \frac{\Pr(M = m, S = s|X_1 = x_1)\Pr(X_1 = x_1)}{f(m,s|p)} \\ &\quad= \frac{\Pr(S_2 = X_1 + X_2 = s, X_1 > a_1|X_1 = x_1)\Pr(X_1 = x_1)}{p^s(1-p)^{n_1+n_2-s}\sum_{x_1=\max(a_1+1,s-n_2)}^{\min(n_1,s)} \begin{pmatrix} n_2 \\ s-x_1 \end{pmatrix}\begin{pmatrix} n_1 \\ x_1 \end{pmatrix}} \\ &\quad= \frac{\Pr(X_2 = s - x_1|X_1 = x_1)\Pr(X_1 = x_1)}{p^s(1-p)^{n_1+n_2-s}\sum_{x_1=\max(a_1+1,s-n_2)}^{\min(n_1,s)} \begin{pmatrix} n_2 \\ s-x_1 \end{pmatrix}\begin{pmatrix} n_1 \\ x_1 \end{pmatrix}} \\ &\quad= \frac{\begin{pmatrix} n_2 \\ s-x_1 \end{pmatrix} p^{s-x_1}(1-p)^{n_2-s+x_1}\begin{pmatrix} n_1 \\ x_1 \end{pmatrix} p^{x_1}(1-p)^{n_1-x_1}}{p^s(1-p)^{n_1+n_2-s}\sum_{x_1=\max(a_1+1,s-n_2)}^{\min(n_1,s)} \begin{pmatrix} n_2 \\ s-x_1 \end{pmatrix}\begin{pmatrix} n_1 \\ x_1 \end{pmatrix}} \\ &\quad= \frac{\begin{pmatrix} n_2 \\ s-x_1 \end{pmatrix}\begin{pmatrix} n_1 \\ x_1 \end{pmatrix}}{\sum_{x_1=\max(a_1+1,s-n_2)}^{\min(n_1,s)} \begin{pmatrix} n_2 \\ s-x_1 \end{pmatrix}\begin{pmatrix} n_1 \\ x_1 \end{pmatrix}}\end{aligned}$$

So when $M = 2$, the UMVUE

$$\begin{aligned}\widetilde{p} &= E\{\widehat{p_1}|(m,s)\} = E\{X_1|(m,s)\}/n_1 \\ &= \frac{1}{n_1}\sum x_1 \Pr(X_1 = x_1|(m,s))\end{aligned}$$

$$= \frac{\sum_{x_1=\max(a_1+1,s-n_2)}^{\min(n_1,s)} \binom{n_2}{s-x_1}\binom{n_1-1}{x_1-1}}{\sum_{x_1=\max(a_1+1,s-n_2)}^{\min(n_1,s)} \binom{n_2}{s-x_1}\binom{n_1}{x_1}}$$

Then the authors gave an example to present the UMVUE and the MLE for observations from a Simon's optimal two-stage design with design parameters $Q = (n, n_1, a_1, a_2) = (43, 13, 3, 12), (p^{(0)}, p^{(A)}, \alpha, \beta) = (0.2, 0.4, 0.05, 0.20)$ in Table I of Jung and Kim (2004). They also plotted the distributions of the UMVUE and the MLE for the same optimal two-stage design in their Figure 1.

Table I of Jung and Kim (2004)

				$f(m,s\|p)$ for p				
m	s	UMVUE	MLE	0.1	0.2	0.3	0.4	0.5
1	0	0.000	0.000	0.254	0.055	0.010	0.001	0.000
1	1	0.077	0.077	0.367	0.179	0.054	0.011	0.002
1	2	0.154	0.154	0.245	0.268	0.139	0.045	0.010
1	3	0.231	0.231	0.100	0.246	0.218	0.111	0.035
2	4	0.308	0.093	0.001	0.000	0.000	0.000	0.000
2	5	0.312	0.116	0.004	0.002	0.000	0.000	0.000
2	6	0.317	0.140	0.007	0.006	0.001	0.000	0.000
2	7	0.322	0.163	0.008	0.015	0.002	0.000	0.000
2	8	0.328	0.186	0.006	0.027	0.006	0.000	0.000
2	9	0.335	0.209	0.004	0.038	0.015	0.001	0.000
2	10	0.343	0.233	0.002	0.043	0.030	0.003	0.000
2	11	0.351	0.256	0.001	0.041	0.049	0.008	0.000
2	12	0.360	0.279	0.000	0.033	0.068	0.018	0.001
2	13	0.371	0.302	0.000	0.023	0.081	0.033	0.003
2	14	0.382	0.326	0.000	0.014	0.084	0.054	0.006
2	15	0.395	0.349	0.000	0.007	0.076	0.076	0.013
2	16	0.409	0.372	0.000	0.003	0.062	0.096	0.025
2	17	0.424	0.395	0.000	0.001	0.044	0.107	0.042
2	18	0.440	0.419	0.000	0.001	0.029	0.108	0.063
2	19	0.458	0.442	0.000	0.000	0.017	0.098	0.085
2	20	0.477	0.465	0.000	0.000	0.009	0.080	0.105
2	21	0.496	0.488	0.000	0.000	0.004	0.059	0.116
2	22	0.517	0.512	0.000	0.000	0.002	0.040	0.118
2	23	0.538	0.535	0.000	0.000	0.001	0.025	0.108
2	24	0.560	0.558	0.000	0.000	0.000	0.014	0.091
2	25	0.582	0.581	0.000	0.000	0.000	0.007	0.069
2	26	0.605	0.605	0.000	0.000	0.000	0.003	0.048
2	27	0.628	0.628	0.000	0.000	0.000	0.001	0.030
2	28	0.651	0.651	0.000	0.000	0.000	0.001	0.017
2	29	0.674	0.674	0.000	0.000	0.000	0.000	0.009
2	30	0.698	0.698	0.000	0.000	0.000	0.000	0.004
2	31	0.721	0.721	0.000	0.000	0.000	0.000	0.002
2	32	0.744	0.744	0.000	0.000	0.000	0.000	0.001
2	33	0.767	0.767	0.000	0.000	0.000	0.000	0.000
2	34	0.791	0.791	0.000	0.000	0.000	0.000	0.000
2	35	0.814	0.814	0.000	0.000	0.000	0.000	0.000
2	36	0.837	0.837	0.000	0.000	0.000	0.000	0.000
2	37	0.861	0.861	0.000	0.000	0.000	0.000	0.000
2	38	0.884	0.884	0.000	0.000	0.000	0.000	0.000
2	39	0.907	0.907	0.000	0.000	0.000	0.000	0.000
2	40	0.930	0.930	0.000	0.000	0.000	0.000	0.000
2	41	0.954	0.954	0.000	0.000	0.000	0.000	0.000
2	42	0.977	0.977	0.000	0.000	0.000	0.000	0.000
2	43	1.000	1.000	0.000	0.000	0.000	0.000	0.000

When $m = 1$, UMVUE and MLE are the same.

When $m = 2$, UMVUE and MLE are very similar for large s values, while MLE is much smaller than UMVUE for small s values. But the probabilities of observing those scenarios for which UMVUE and MLE are very different (esp. at $(m, s) = (2, a_1 + 1) = (2, 4)$) are

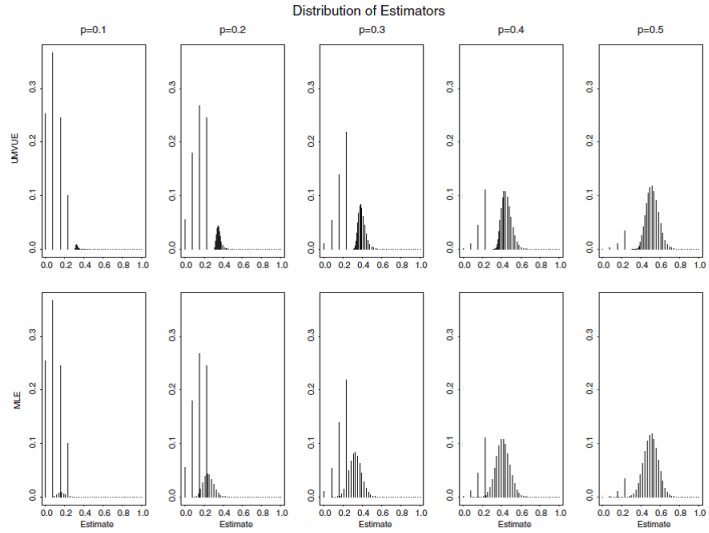


Figure 1. Distribution of the MLE and the UMVUE following two-stage optimal design with $\alpha = 0.05$ and $\beta = 0.2$ for $p_0 = 0.2$ and $p_1 = 0.4$.

very small based on the values of the p.m.f $f(m, s|p)$.

To understand the relative efficiency of the UMVUE as compared to the MLE defined as the ratio of the mean squared error(MSE) of the MLE to the variance of the UMVUE, Jung and Kim (2004) conducted numerical studies based on two-stage designs with lower stopping boundaries only. The results show that for all designs (optimal or minimax designs), the MLE has smaller MSE for smaller true response rates than UMVUE, but larger MSE for larger true response rates. It is also shown that there appears to be some efficiency loss with UMVUE as compared to MLE, particularly for optimal designs, a reasonable price for unbiasedness.

2.5 Interval estimation of binomial probability in clinical trials with a single-endpoint two-stage design

Besides point estimates, confidence intervals are often reported in Phase II trials. Since the interval estimation of response rate has been extensively developed in the literature for a binomial response in a single-stage design, Tsai et al. (2008) considered the interval estimation of the response probability when the second stage is allowed to continue in a two-stage design. This is the method of conditional inference. Two asymptotic interval estimators, Wald and score, as well as two exact interval estimators, Clopper-Pearson and

Sterne, are constructed according to the two binomial responses from this two-stage design, where the binomial response after the first stage follows a truncated binomial distribution. These interval estimators are based on the conditional MLE given the trial proceeds to the second stage.

Notations:

- $L(p)$: Likelihood function conditioning on the second stage is allowed to continue;
- $G(a_1, n_1, p) = Pr(a_1 < X_1 \leq n_1) = Pr(a_1 + 1 \leq X_1 \leq n_1)$;
- $G_1(a_1, n_1, p)$: first derivative of $G(a_1, n_1, p)$ with respect to p;
- $G_2(a_1, n_1, p)$: second derivative of $G(a_1, n_1, p)$ with respect to p.

When the trial is allowed to continue to the second stage, the probability distribution of X_1 is referred to as truncated binomial distribution with response probability p. Its p.m.f. is

$$P_{a_1}(X_1 = x_1) = \frac{\binom{n_1}{x_1} p^{x_1}(1-p)^{n_1-x_1}}{\sum_{k=a_1+1}^{n_1} \binom{n_1}{k} p^k(1-p)^{n_1-k}} = \frac{\binom{n_1}{x_1} p^{x_1}(1-p)^{n_1-x_1}}{G(a_1, n_1, p)}$$

The distribution of X_2 is simple binomial and its p.m.f. is

$$P(X_2 = x_2) = \binom{n_2}{x_2} p^{x_2}(1-p)^{n_2-x_2}$$

The likelihood function based jointly on $X_1 = x_1$ and $X_2 = x_2$ can be expressed as

$$L(p) = \frac{1}{G(a_1, n_1, p)} \binom{n_1}{x_1} \binom{n_2}{x_2} p^{x_1+x_2}(1-p)^{n_1+n_2-x_1-x_2}$$

The score function is

$$S(p) = -\frac{G_1(a_1, n_1, p)}{G(a_1, n_1, p)} + \frac{x_1 + x_2}{p} - \frac{n_1 + n_2 - x_1 - x_2}{1-p}.$$

The graph of $S(p)$ displays that a unique solution can be obtained for $a_1 + 1 < x_1 + x_2 < n_1 + n_2$. The equation $S(p) = 0$ does not have a closed-form solution. Hence, the Newton-Raphson (NR) algorithm is employed to obtain a numerical estimate of p, e.g., when $x_1 + x_2 = 25$, the NR algorithm found the solution as 0.580.

To apply the NR algorithm, the observed Fisher information must be derived and it is as follows:

$$I_n(X,p) = \frac{G(a_1,n_1,p)G_2(a_1,n_1,p) - G_1^2(a_1,n_1,p)}{G^2(a_1,n_1,p)} + \frac{x_1+x_2}{p^2} + \frac{n_1+n_2-x_1-x_2}{(1-p)^2}$$

The NR iterative algorithm starts with an initial guess of p, say 0.5; then update by the iteration equation:

$$p_{[j]} = p_{[j-1]} + S(p_{[j-1]})/I_n(X, p_{[j-1]})$$

In this manner, this algorithm generates a sequence of estimates and the conditional MLE, $\widehat{p}_c$, is the limit of $p_{[j]}$ as $j \to \infty$.

The asymptotic confidence intervals, say Wald and Score intervals, can be derived based on the asymptotic distribution of the conditional MLE $\widehat{p}_c$.

Wald Confidence Interval

Under regularity conditions, the asymptotic distribution of the conditional MLE

$$\widehat{p}_c \sim AN(p, I_T^{-1}(p))$$

where $I_T(p)$ is the Fisher information of $\widehat{p}_c$,

$$\begin{aligned} I_T(p) &= E[I_n(X,p)] \\ &= \frac{G(a_1,n_1,p)G_2(a_1,n_1,p) - G_1^2(a_1,n_1,p)}{G^2(a_1,n_1,p)} + \frac{\mu(p)}{p^2} + \frac{n_1+n_2-\mu(p)}{(1-p)^2} \end{aligned}$$

where

$$\mu(p) = E(X_1 + X_2) = n_1 p G(a_1 - 1, n_1 - 1, p)/G(a_1, n_1, p) + n_2 p$$

The Fisher information $I_T(p)$ can be consistently estimated by replacing p by $\widehat{p}_c$ in it.

For true p, the Wald test statistic

$$\frac{\widehat{p}_c - p}{I_T^{-1/2}(\widehat{p}_c)}$$

has an asymptotic standard normal distribution. The confidence intervals can be constructed by inverting the hypothesis test based on the acceptance region at level α

$$-z_{1-\frac{\alpha}{2}} \leq \frac{\widehat{p}_c - p}{I_T^{-1/2}(\widehat{p}_c)} \leq z_{1-\frac{\alpha}{2}}.$$

The resulting $(1-\alpha)\times 100\%$ Wald confidence interval is

$$[\widehat{p}_c - z_{1-\frac{\alpha}{2}} I_T^{-1/2}(\widehat{p}_c), \widehat{p}_c + z_{1-\frac{\alpha}{2}} I_T^{-1/2}(\widehat{p}_c)]$$

except that the lower limit is 0 if this is negative and the upper limit is 1 if this is greater than 1. Since for $x_1 + x_2 = a$, the MLE of p is 0 and for $x_1 + x_2 = n$, the MLE is 1, their interval estimates are 0 and 1, respectively.

Because a discrete random variable can take on only specified values, the correction for continuity adjustment is employed. $1/(2(n-a_1-1))$ is used as the factor for continuity correction(CC) for a two-stage design. Thus, the Wald interval with CC (*Wald_c*) in Simon's two-stage design is:

$$[\widehat{p}_c - (z_{1-\frac{\alpha}{2}} I_T^{-1/2}(\widehat{p}_c) + 1/(2(n-a_1-1))),\ \widehat{p}_c + (z_{1-\frac{\alpha}{2}} I_T^{-1/2}(\widehat{p}_c) + 1/(2(n-a_1-1)))].$$

As investigated by many researchers, it is well known that the coverage probability of the Wald interval is usually below the nominal confidence level for a single-stage design. Thus the authors also derived the Score confidence interval for p.

Score Confidence Interval

Following Wilson's concept, the lower and upper limits of the Score confidence interval can be solved by replacing $\widehat{p}_c$ with the actual success probability p in the denominator of

the Wald statistic. That is, the Score statistic is

$$\frac{\widehat{p}_c - p}{I_T^{-1/2}(p)}$$

For a given $\widehat{p}_c$, the lower interval limit is the solution to the equation

$$h_L(p) = (\widehat{p}_c - p) - z_{1-\frac{\alpha}{2}} I_T^{-1/2}(p) \ = 0$$

and the upper interval limit is the solution to the equation

$$h_U(p) = (\widehat{p}_c - p) + z_{1-\frac{\alpha}{2}} I_T^{-1/2}(p) \ = 0.$$

The graph of $h_L(p)$ for various x shows that a unique solution of p can be obtained except for $x = a_1 + 1$, and the graph of $h_U(p)$ for various x shows that a unique solution of p can be obtained except for $x = n_1 + n_2$. Thus, when $x = a_1 + 1$, the lower limit is 0 and the upper limit of the score interval is the solution to the equation $h_U(p) = 0$. When $x = n_1 + n_2$, the upper limit is 1 and the lower limit of the score interval is the solution to the equation $h_L(p) = 0$.

The bisection numerical algorithm can be employed to obtain solutions to $h_L(p) = 0$ and $h_U(p) = 0$ more easily than the NR algorithm. The resulting Score CI is denoted by $[p_L^{Score}, p_U^{Score}]$. The lower limit of the score interval with CC is the solution to the equation

$$(\widehat{p}_c - p) - 1/(2(n - a_1 - 1)) - z_{1-\frac{\alpha}{2}} I_T^{-1/2}(p) \ = 0$$

and the upper limit of the score interval with CC is the solution to the equation

$$(\widehat{p}_c - p) + 1/(2(n - a_1 - 1)) + z_{1-\frac{\alpha}{2}} I_T^{-1/2}(p) \ = 0.$$

The resulting CI ($Score\_c$) is denoted by $[p_L^{Score\_c}, p_U^{Score\_c}]$.

Coverage probability and the expected width were used to evaluate the performance of

interval estimators. Coverage probability of an interval estimator is defined as:

$$CP(p) = \sum_{x=0}^{n} I(x,p)P(X=x)$$

where x is the possible values of the Binomial random variable X and $I(x,p)$ is an indicator function indicating whether the interval based on $X = x$ contains p or not.

Expected width of an interval estimator is defined as

$$EW(p) = \sum_{x=0}^{n} (\widehat{p}_U - \widehat{p}_L)P(X=x).$$

Note that both coverage probability and expected width are functions of p.

Regarding unbiasedness, the (conditional) MLE, $\widehat{p}_c$, is slightly underestimated, and the sample proportion $\widehat{p}$ is overestimated for Simon's designs when only cumulative data from second stage is used. Note that the reason why Tsai et al. (2008) claimed sample proportion overestimated in terms of unbiasedness is that Tsai et al. (2008) has used only the distribution of the second stage data. In contrast, Jung and Kim (2004) claimed the overall bias of sample proportion, which considers data from both stage 1 and stage 2, is always negative (that is, sample proportion underestimates).

In general, the bias of $\widehat{p}_c$ is much smaller than that of sample proportion $\widehat{p}$ (conditional likelihood vs. full likelihood, and the former utilized one more information that knowing M=2), and hence $\widehat{p}_c$ is recommended for use in practice in Tsai et al. (2008).

With conditional inference, the coverage probability of the Wald intervals without CC are below the nominal confidence level (95%) for most of Simon's optimal designs. The Wald interval with CC seems slightly better than that without CC. The coverage probability of the Score intervals with and without CC are the same for most of Simon's optimal designs, except for two designs. Except for the Wald interval without CC, the score interval has the smallest interval width as compared to the other interval estimators examined here.

In summary, Tsai et al. (2008) recommends the Score interval for both Simon's optimal and minimax designs using conditional inference.

Chapter 3

Two-stage Design for Phase II Cancer Clinical Trials With Two Endpoints

Recall that our research question is "Does the new treatment have sufficient anti-tumor activity when there are multiple alternative primary endpoints of efficacy?" It can be written mathematically as:

$$H_0 : \pi_1 \leq p_1^{(0)}, \pi_2 \leq p_2^{(0)}, \cdots, \pi_k \leq p_k^{(0)}$$
$$H_A : \pi_1 \geq p_1^{(A)} \text{ or } \pi_2 \geq p_2^{(A)} \cdots \text{ or } \pi_k \geq p_k^{(A)}$$

Let's narrow down the research question a little bit and start with $k = 2$. Let probabilities of event/success for each of the binary efficacy endpoints be designated as π_1 and π_2, respectively. Then the research hypotheses can be formulated as follows:

$$H_0 : \pi_1 \leq p_1^{(0)} \text{ and } \pi_2 \leq p_2^{(0)}$$
$$H_A : \pi_1 \geq p_1^{(A)} \text{ or } \pi_2 \geq p_2^{(A)}$$

where $p_1^{(0)}$ and $p_2^{(0)}$ are specified values that are believed to be uninteresting or comparable to the current standard of care, $p_1^{(A)}$ and $p_2^{(A)}$ are the targeted response rates with $p_1^{(0)} < p_1^{(A)}$ and $p_2^{(0)} < p_2^{(A)}$.

3.1 Design Settings

Our methodology has been developed in the context of phase II cancer clinical trials. Suppose we are interested in two binary alternative primary endpoints of efficacy, either originally binary or being dichotomized from continuous efficacy variables. The probablities of

event/success for each of the binary efficacy endpoints are denoted as π_1 and π_2, respectively. The original data in a two-stage design looks like:

Table 3.1: The original data structure

	ID	*Endpoint* 1 ($Yes = 1$ *or* $No = 0$)	*Endpoint* 2 ($Yes = 1$ *or* $No = 0$)
Stage 1	1	X_{11}	X_{12}
	2	X_{21}	X_{22}
	·	·	·
	·	·	·
	n_1	$X_{n_1 1}$	$X_{n_1 2}$
Total at the end of stage 1		$X_1 = \sum_{i=1}^{n_1} X_{i1}$	$Y_1 = \sum_{i=1}^{n_1} X_{i2}$
Stage 2	$n_1 + 1$	$X_{(n_1+1)1}$	$X_{(n_1+1)2}$
	·	·	·
	·	·	·
	$n_1 + n_2 (= n)$	X_{n1}	X_{n2}
Total at the end of stage 2		$X = \sum_{i=1}^{n} X_{i1}$	$Y = \sum_{i=1}^{n} X_{i2}$

where

- $X_{11}, X_{21}, ..., X_{n_1 1}, X_{(n_1+1)1}, ..., X_{n1} \overset{i.i.d}{\sim} Bernoulli(\pi_1)$. Let X_1 and X denote the total number of responses for endpoint 1 at the end of stage 1 and of stage 2 if any, respectively, that is, $X_1 = \sum_{i=1}^{n_1} X_{i1}$ and $X = \sum_{i=1}^{n} X_{i1}$. Then $X_1 \sim Binomial(n_1, \pi_1)$ and $X \sim Binomial(n, \pi_1)$; and

- $X_{12}, X_{22}, ..., X_{n_1 2}, X_{(n_1+1)2}, ..., X_{n2} \overset{i.i.d}{\sim} Bernoulli(\pi_2)$. Let Y_1 and Y denote the total number of responses for endpoint 2 at the end of stage 1 and of stage 2 if any, respectively, that is, $Y_1 = \sum_{i=1}^{n_1} X_{i2}$ and $Y = \sum_{i=1}^{n} X_{i2}$. Then $Y_1 \sim Binomial(n_1, \pi_2)$ and $Y \sim Binomial(n, \pi_2)$;

Within the same individual i, there are some correlation between X_{i1} and X_{i2}; and people are independent of each other.

A two-stage design allowing for early termination for futility only can be specified by a vector of six parameters $\mathbf{Q} = (n_1, n, s_1, t_1, s, t)$:

- n_1: the number of patients enrolled in stage 1;

- n: total number of enrolled patients in stage 1 and stage 2;

- (s_1, s): critical values associated with the occurrence of endpoint 1 at the end of stage 1 and of stage 2, respectively;
- (t_1, t): critical values associated with the occurence of endpoint 2 at the end of stage 1 and of stage 2, respectively.

The trial proceeds as follows:

- Accrue n_1 patients in stage 1. If $X_1 \leq s_1$ and $Y_1 \leq t_1$, terminate the trial due to futility.
- Otherwise, accrue additional $(n - n_1)$ patients into the 2nd stage. Recommend the treatment only if $\{(X_1 > s_1$ or $Y_1 > t_1)$ and $(X > s$ or $Y > t)\}$.

Note that in two-stage designs, the sample size is a random variable. When the numbers of responses at the end of stage 1 pass the interim critical boundaries, the trial continues to the second stage. In such cases, we only make decisions (reject the null hypothesis or accept the null hypothesis) at the end of the study, rather than at the end of stage 1. For instance, if a clinical trial indeed successfully enters the second stage, and is with the responses $\{(X_1 > s_1, Y_1 \leq t_1)$ and $(X \leq s, Y > t)\}$, we recommend the treatment at the conclusion of the study and the efficacy claim is based on both efficacy endpoints at the second stage. This is an example in which the evidence of efficacy for both endpoints is not accumulated sufficiently at the end of stage 1 due to limited sample size even though the response of the either endpoint *appears* to be effective at the end of stage 1, we need to use additional information from the second stage to confirm our conclusion.

One of the goals of this dissertation is, given 8 parameters (α, β_1, β_2, β, $p_1^{(0)}$, $p_2^{(0)}$, $p_1^{(A)}$, $p_2^{(A)}$), to search for feasible solutions of $\mathbf{Q}= (n_1, n, s_1, t_1, s, t)$ that satisfy type I/II error constraints, and then use the optimality criteria to find the "best" solution.

3.2 Derivation of the joint probability function of two correlated binary variables

When doing hypothesis testing, we need to find out appropriate test statistics. In a single-stage design with n patients, let the collected summary data be denoted as (X, Y), where

X and Y are total number of responses for endpoint 1 and for endpoint 2, respectively. X and Y are convenient test statistics for π_1 and π_2, respectively. The details of deriving the joint probability function of (X, Y) are as follows.

The original individual data in a single-stage design with n patients can be collapsed into 4 response patterns, which can be represented by a 2×2 table as follows:

Table 3.2: Response Pattern

Pattern	*Endpoint* 1	*Endpoint* 2	*total obs.*
1	Yes	Yes	n_{11}
2	Yes	No	n_{10}
3	No	Yes	n_{01}
4	No	No	n_{00}

Table 3.3: Observed Counts

		Endpoint 1		
		Yes	No	
Endpoint 2	Yes	$C_{11} = n_{11}$	$C_{01} = n_{01}$	$Y = n_{+1}$
	No	$C_{10} = n_{10}$	$C_{00} = n_{00}$	
		$X = n_{1+}$		n

Table 3.4: Probability

		Endpoint 1		
		Yes	No	
Endpoint 2	Yes	π_{11}	π_{01}	π_2
	No	π_{10}	π_{00}	
		π_1		

The random quantities in each cell of the 2×2 table (Table 3.3), $(C_{11}, C_{10}, C_{01}, C_{00})$, are distributed as:

$$(C_{11,}\ C_{10,}\ C_{01,}\ C_{00}) \sim Multinomial(n, (\pi_{11}, \pi_{10}, \pi_{01}, \pi_{00}))$$

Recall that the probability mass function of a multinomial distribution is:

$$\begin{aligned} f(x_1, x_2, ..., x_k; n, p_1, p_2, ..., p_k) &= Pr(X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k) \\ &= \binom{n}{x_1, x_2, \cdots, x_k} p_1^{x_1} {p_2}^{x_2} ... p_k^{x_k} \\ &= \frac{n!}{x_1!\ x_2! ... x_k!} p_1^{x_1} {p_2}^{x_2} \cdots p_k^{x_k}. \end{aligned}$$

Table 3.5: Sketch of the joint p.m.f of (X, Y)

		X					
		0	1	2	...	...	n
	0						
	1						
Y	2						
	...						
	...						
	n						

The joint distribution of (X, Y) is sketched in Table 3.5 (given n and π_{11}) and the joint probability mass function (p.m.f) of (X, Y) can be derived:

$$\begin{aligned} & p(x, y; n, \pi_1, \pi_2, \pi_{11}) = P(X = x, Y = y) \\ &= \sum_{n_{11}} \Pr(C_{11} = n_{11}, C_{10} = x - n_{11}, C_{01} = y - n_{11}, C_{00} = n - x - y + n_{11}) \\ &= \sum_{\max(0,\ x+y-n) \le n_{11} \le \min(x,\ y)} \binom{n}{n_{11}, x - n_{11}, y - n_{11}, n - x - y + n_{11}} \times \pi_{11}^{n_{11}} \times \pi_{10}^{x - n_{11}} \\ &\qquad \times \pi_{01}^{y - n_{11}} \times \pi_{00}^{n - x - y + n_{11}} \\ &= \sum_{\max(0,\ x+y-n) \le n_{11} \le \min(x,\ y)} \binom{n}{n_{11}, x - n_{11}, y - n_{11}, n - x - y + n_{11}} \times \pi_{11}^{n_{11}} \\ &\qquad \times (\pi_1 - \pi_{11})^{x - n_{11}} \times (\pi_2 - \pi_{11})^{y - n_{11}} \times (1 - \pi_1 - \pi_2 + \pi_{11})^{n - x - y + n_{11}} \end{aligned} \tag{3.1}$$

where

$$\max(0, \pi_1 + \pi_2 - 1) \le \pi_{11} \le \min(\pi_1, \pi_2).$$

Regarding the correlation/association between the two endpoints X_{i1} and X_{i2} within

the same individual i , it can be expressed through any convenient measure such as the correlation coefficient or the odds ratio. Let such an association between X_{i1} and X_{i2} be represented by $Pr[X_{i1} = 1, X_{i2} = 1] = \pi_{11}$, rather than the complicated correlation coefficient.

The following brief derivations will show that:

- π_{11} can be representative of the correlation between X_{i1} and X_{i2};
- The correlation between X_{i1} and X_{i2} within the same individual i is equal to the correlation between X and Y;
- Thus, π_{11} can be representative of the correlation/association between X and Y.

$$
\begin{aligned}
&Pr[X_{i1} = 1, X_{i2} = 1] = \pi_{11},\\
&Pr[X_{i1} = 1, X_{i2} = 0] = \pi_{10} = \pi_1 - \pi_{11},\\
&Pr[X_{i1} = 0, X_{i2} = 1] = \pi_{01} = \pi_2 - \pi_{11},\\
&Pr[X_{i1} = 0, X_{i2} = 0] = \pi_{00} = 1 - \pi_1 - \pi_2 + \pi_{11}.
\end{aligned}
$$

X_{i1}	1	1	0	0
X_{i2}	1	0	1	0
probability	π_{11}	$\pi_1 - \pi_{11}$	$\pi_2 - \pi_{11}$	$1- \pi_1 - \pi_2 + \pi_{11}$
count	n_{11}	$x - n_{11}$	$y - n_{11}$	$n - x - y + n_{11}$
$E[X_{i1}X_{i2}\ ] = 1 \times 1 \times \pi_{11} + 1 \times 0 \times (\pi_1 - \pi_{11}) + 0 + 0 = \pi_{11} = Pr[X_{i1} = 1, X_{i2} = 1]$				

$$
\begin{aligned}
corr(X_{i1}, X_{i2}) &= \frac{cov(X_{i1}, X_{i2})}{\sqrt{Var(X_{i1})} \times \sqrt{Var(X_{i2})}}\\
&= \frac{E[X_{i1} \times X_{i2}] \ - \ E[X_{i1}] \ \times \ E[X_{i2}]}{\sqrt{Var(X_{i1})} \times \sqrt{Var(X_{i2})}}\\
&= \frac{\pi_{11} \ - \ \pi_1 \ \times \ \pi_2}{\sqrt{\pi_1 \times \ (1 \ - \ \pi_1)} \times \sqrt{\pi_2 \times \ (1 \ - \ \pi_2)}}
\end{aligned}
$$

Since $X = \sum_{i=1}^{n} X_{i1}, Y = \sum_{i=1}^{n} X_{i2}$,

$$
cov(X, \ Y) = cov\left(\sum_{i=1}^{n} X_{i1}, \sum_{i=1}^{n} X_{i2}\right) = ncov(X_{11}, X_{12}),
$$

so $corr(X, Y) = corr(X_{i1}, X_{i2})$, for all i.

3.3 Derivation of the Power Function

In order to get the sample sizes and the critical values in the design parameters $\mathbf{Q}$, we need first to define the power function as follows:

$$G(\mathbf{Q}, \pi_{11}|H) = Pr(\text{Recommend Treatment}|\mathbf{Q}, \pi_{11}, H)$$

3.3.1 Single-Stage Designs

For single-stage designs, recommend the treatment if $X > s$ or $Y > t$. The power function for Single-Stage designs can be written as:

$$\begin{aligned}
&G_s(\mathbf{Q}, \pi_{11}|H) = Pr(\text{Recommend Treatment}|\mathbf{Q}, \pi_{11}, H) \\
=&Pr(X > s \text{ or } Y > t \mid \mathbf{Q}, \pi_{11}, H) \\
=&1 - Pr(X \leq s, Y \leq t \mid \mathbf{Q}, \pi_{11}, H) \\
=&1 - \sum_{x=0}^{s} \sum_{y=0}^{t} p(x, y; n, \pi_1, \pi_2, \pi_{11})
\end{aligned}$$

Let

$$\begin{aligned}
D(s, t; n, \pi_1, \pi_2, \pi_{11}) &= Pr(X \leq s, Y \leq t \mid X \sim Bin(n, \pi_1), Y \sim Bin(n, \pi_2), \pi_{11}) \\
&= \sum_{x=0}^{s} \sum_{y=0}^{t} p(x, y; n, \pi_1, \pi_2, \pi_{11}),
\end{aligned}$$

then the power function for Single-Stage designs can be denoted as:

$$\begin{aligned}
&G_s(\mathbf{Q}, \pi_{11}|H) = G_s(n_1, n, s_1, t_1, s, t, \pi_{11}, \pi_1, \pi_2) \\
=&1 - \sum_{x=0}^{s} \sum_{y=0}^{t} p(x, y; n, \pi_1, \pi_2, \pi_{11}) = 1 - D(s, t; n, \pi_1, \pi_2, \pi_{11})
\end{aligned}$$

3.3.2 Two-Stage Designs

For two-stage designs, recommend the treatment if $\{(X_1 > s_1 \text{ or } Y_1 > t_1)$ and $(X > s \text{ or } Y > t)\}$. The power function for two-stage designs can be written as:

$$
\begin{aligned}
&G_t(\mathbf{Q}, \pi_{11}|H) = Pr(\text{Recommend Treatment}|\mathbf{Q}, \pi_{11}, H) \\
=&G_t(n_1, n, s_1, t_1, s, t, \pi_{11}, \pi_1, \pi_2) \\
=&Pr(X_1 > s_1 \text{ or } Y_1 > t_1), (X > s \text{ or } Y > t)) \\
=&Pr((X > s \text{ or } Y > t)) - Pr(X_1 \le s_1, Y_1 \le t_1, (X > s \text{ or } Y > t)) \\
=&1 - Pr(X \le s, Y \le t) - Pr((X > s \text{ or } Y > t)|X_1 \le s_1, Y_1 \le t_1)Pr(X_1 \le s_1, Y_1 \le t_1) \\
=&1 - Pr(X \le s, Y \le t) - (1 - Pr(X \le s, Y \le t|X_1 \le s_1, Y_1 \le t_1))Pr(X_1 \le s_1, Y_1 \le t_1) \\
=&1 - Pr(X \le s, Y \le t) - Pr(X_1 \le s_1, Y_1 \le t_1) \\
&\quad + Pr(X \le s, Y \le t|X_1 \le s_1, Y_1 \le t_1)Pr(X_1 \le s_1, Y_1 \le t_1) \\
=&1 - Pr(X \le s, Y \le t) - Pr(X_1 \le s_1, Y_1 \le t_1) \\
&\quad + \underline{Pr(X \le s, Y \le t, X_1 \le s_1, Y_1 \le t_1)} \\
=&1 - D(s, t; n, \pi_1, \pi_2, \pi_{11}) - D(s_1, t_1; n_1, \pi_1, \pi_2, \pi_{11}) \\
&\quad + \sum_{i=0}^{s_1}\sum_{j=0}^{t_1}\{D(s-i, t-j; n_2, \pi_1, \pi_2, \pi_{11}) \times p(i, j; n_1, \pi_1, \pi_2, \pi_{11})\}
\end{aligned}
$$

The detailed derivation of calculation of the underlined term of probability is as follows:

$$
\begin{aligned}
&Pr(X \le s, Y \le t, X_1 \le s_1, Y_1 \le t_1) \\
=&Pr(X_1 + X_2 \le s, Y_1 + Y_2 \le t, X_1 \le s_1, Y_1 \le t_1) \\
=&\sum_{i=0}^{s_1}\sum_{j=0}^{t_1} Pr(X_1 + X_2 \le s, Y_1 + Y_2 \le t, X_1 = i, Y_1 = j) \\
=&\sum_{i=0}^{s_1}\sum_{j=0}^{t_1} Pr(X_2 \le s - i, Y_2 \le t - j, X_1 = i, Y_1 = j) \\
=&\sum_{i=0}^{s_1}\sum_{j=0}^{t_1} Pr(X_2 \le s - i, Y_2 \le t - j)Pr(X_1 = i, Y_1 = j).
\end{aligned}
$$

3.4 Properties of the power function for two-stage designs with bivariate endpoints.

In this section, we will study the relationship between the power function and each of the design parameters π_1, π_2, π_{11}, t and s.

3.4.1 With respect to π_1 or π_2

Previous studies (Simon, 1989; Banerjee and Tsiatis, 2006) did not rigidly prove that the power function is a monotone function of the marginal probability such as π_1. Instead, some has used numerical calculation results to project the monotonicity. For our study, in the beginning, we also use this empirical strategy to present the monotonicity.

The following 3D plots (given a fixed $(Q,\ \pi_{11})$) show that the power function is a non-decreasing continuous function of π_1, and of π_2 as well. The values of the power function in the space defined by the composite null hypothesis $H_0 : \pi_1 \leqslant p_1^{(0)}$ and $\pi_2 \leqslant p_2^{(0)}$, which are essentially the values of the type I error rate, have its maximum value at $\pi_1 = p_1^{(0)}$ and $\pi_2 = p_2^{(0)}$.

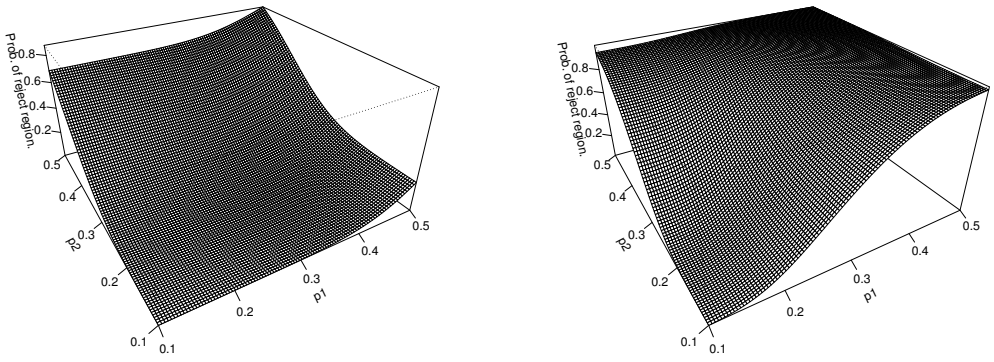


Figure 3.1: Relationship between the Power Function and (π_1, π_2).
(left): Given $(Q,\ \pi_{11})$=$(n_1,\ n,\ s_1,\ t_1,\ s,\ t,\ \pi_{11})$ = (15, 29, 8, 6, 15, 10, 0.1).
(right): Given $(Q,\ \pi_{11})$=$(n_1,\ n,\ s_1,\ t_1,\ s,\ t,\ \pi_{11})$ = (20, 50, 5, 6, 10, 15, 0.1)

Using the method of mathematical induction (MI), we can theoretically prove that the power function is a non-decreasing function of π_1, and of π_2.

Theorem 3.4.1 *The power function is a non-decreasing function of* π_1 *and a non-decreasing function of* π_2 *as well.*

Proof: We will use the method of mathematical induction(MI) to prove the theorem. Recall the steps to implement MI:

- Base case: prove the given statement for the 1st natural number;
- Inductive step: prove that, if the statement holds for some natural number n, then the statement holds for $n+1$.

Briefly, the proof for **Theorem** 3.4.1 consists of three parts:

- Part 1: Prove $D(s,t;n,\pi_1,\pi_2,\pi_{11})$ is monotone in π_1;
- Part 2: Prove the power function for single-stage designs $G_s(\mathbf{Q},\pi_{11}|H)$ is monotone in π_1;
- Part 3: Prove the power function for two-stage designs $G_t(\mathbf{Q},\pi_{11}|H)$ is monotone in π_1.

Part 1. Prove $D(s,t;n,\pi_1,\pi_2,\pi_{11})$ **is monotone in** π_1

Step 1: For $n=1$, we have

$$D(s,t;n,\pi_1,\pi_2,\pi_{11}) = Pr(X \le s, Y \le t | X \sim Ber(\pi_1), Y \sim Ber(\pi_2), \pi_{11}).$$

If $(s,t)=(0,0)$,

$$\begin{aligned} D(s,t;n,\pi_1,\pi_2,\pi_{11}) &= Pr(X=0,Y=0) = 1-\pi_1-\pi_2+\pi_{11}, \\ \frac{\partial}{\partial \pi_1} D(s,t;n,\pi_1,\pi_2,\pi_{11}) &= -1; \end{aligned}$$

if $(s,t)=(1,0)$,

$$\begin{aligned} D(s,t;n,\pi_1,\pi_2,\pi_{11}) &= Pr(X \le 1, Y=0) = (1-\pi_1-\pi_2+\pi_{11}) + (\pi_1-\pi_{11}) \\ &= 1-\pi_2, \\ \frac{\partial}{\partial \pi_1} D(s,t;n,\pi_1,\pi_2,\pi_{11}) &= 0; \end{aligned}$$

if $(s,t)=(0,1)$,

$$D(s,t;n,\pi_1,\pi_2,\pi_{11}) = Pr(X=0, Y\leq 1) = 1-\pi_1,$$
$$\frac{\partial}{\partial \pi_1} D(s,t;n,\pi_1,\pi_2,\pi_{11}) = -1;$$

if $(s,t)=(1,1)$,

$$D(s,t;n,\pi_1,\pi_2,\pi_{11}) = Pr(X\leq 1, Y\leq 1) = 1,$$
$$\frac{\partial}{\partial \pi_1} D(s,t;n,\pi_1,\pi_2,\pi_{11}) = 0.$$

Therefore, when $n=1$, $D(s,t;n,\pi_1,\pi_2,\pi_{11})$ is a non-increasing function of π_1.

Step 2:

Assume that for $n=k\geq 1$, $D(s,t;k,\pi_1,\pi_2,\pi_{11})$ is a non-increasing function of π_1, that is,

$$\frac{\partial}{\partial \pi_1} D(s,t;k,\pi_1,\pi_2,\pi_{11}) \leq 0.$$

Therefore when $n=k+1$, consider an independent sequence of $k+1$ patients. Let P_{hm} denote the probability of the response pattern for the first patient,

$$P_{hm} = Pr(X_{11}=h, X_{12}=m) = \begin{cases} \pi_{11}, & \text{when } h=m=1 \\ \pi_1-\pi_{11}, & \text{when } h=1, m=0 \\ \pi_2-\pi_{11}, & \text{when } h=0, m=1 \\ 1-\pi_1-\pi_2+\pi_{11}, & \text{when } h=m=0, \end{cases}$$

and $X_{i1}\sim Ber(\pi_1), X_{i2}\sim Ber(\pi_2)$ are the endpoint 1 response status and endpoint 2 response status from the i^{th} patient for $i=1,\cdots,k+1$, respectively. By conditioning on the number of responses for Endpoint 1 and Endpoint 2 for the first patient, we have

$$\begin{aligned} &D(s,t;\underline{k+1},\pi_1,\pi_2,\pi_{11}) \\ &= Pr(X\leq s, Y\leq t \mid X\sim Bin(\underline{k+1},\pi_1), Y\sim Bin(\underline{k+1},\pi_2),\pi_{11}) \\ &= Pr(X_{11}+X_{21}+\cdots+X_{(k+1),1}\leq s, X_{12}+X_{22}+\cdots+X_{(k+1),2}\leq t) \end{aligned}$$

$$= \sum_{h=0,1} \sum_{m=0,1} Pr(X_{21} + \cdots + X_{(k+1),1} \leq s-h, X_{22} + \cdots + X_{(k+1),2} \leq t-m | X_{11} = h,$$
$$X_{12} = m) P_{hm}$$
$$= \sum_{h=0,1} \sum_{m=0,1} Pr(X_{21} + \cdots + X_{(k+1),1} \leq s-h, X_{22} + \cdots + X_{(k+1),2} \leq t-m\ ) P_{hm}$$
$$= \sum_{h=0,1} \sum_{m=0,1} D(s-h,\ t-m; \underline{k}, \pi_1, \pi_2, \pi_{11}) \times P_{hm}$$

Denote by D' the partial derivative of D with respect to π_1, we get

$$\begin{aligned} & D'(s,t;k+1,\pi_1,\pi_2,\pi_{11}) \\ & = \sum_{h=0,1} \sum_{m=0,1} \frac{\partial P_{hm}}{\partial \pi_1} D(s-h,\ t-m; k, \pi_1, \pi_2, \pi_{11}) \\ & \quad + \sum_{h=0,1} \sum_{m=0,1} D'(s-h,\ t-m; k, \pi_1, \pi_2, \pi_{11}) P_{hm}. \end{aligned}$$

By the previous assumption for $n = k$, the second part is non-positive. Expanding the first part of the above equation,

$$\begin{aligned} & \sum_{h=0,1} \sum_{m=0,1} \frac{\partial P_{hm}}{\partial \pi_1} D(s-h,\ t-m; k, \pi_1, \pi_2, \pi_{11}) \\ & = D(s-1,\ t; k, \pi_1, \pi_2, \pi_{11}) - D(s,\ t; k, \pi_1, \pi_2, \pi_{11}) \\ & = Pr(X \leq s-1, Y \leq t \mid X \sim Bin(k, \pi_1), Y \sim Bin(k, \pi_2), \pi_{11}) \\ & \qquad\qquad - Pr(X \leq s, Y \leq t \mid X \sim Bin(k, \pi_1), Y \sim Bin(k, \pi_2), \pi_{11}) \\ & = -\ Pr(X = s, Y \leq t \mid X \sim Bin(k, \pi_1), Y \sim Bin(k, \pi_2), \pi_{11}) \\ & \leq 0. \end{aligned}$$

Therefore,

$$D'(s,t;k+1,\pi_1,\pi_2,\pi_{11}) \leq 0.$$

Combining the above step 1 and step 2, and by mathematical induction, $D(s,t;n,\pi_1,\pi_2,\pi_{11})$ is non-increasing in π_1 for any $n \geq 1$.

Part 2. Prove the power function for single-stage designs $G_s(\mathbf{Q}, \pi_{11}|H)$ is monotone in π_1

For single-stage designs, the power function is

$$\begin{aligned} &G_s(\mathbf{Q}, \pi_{11}|H) = G_s(n_1, n, s_1, t_1, s, t, \pi_{11}, \pi_1, \pi_2) \\ =&1 - D(s, t; n, \pi_1, \pi_2, \pi_{11}). \end{aligned}$$

Since $D(s, t; n, \pi_1, \pi_2, \pi_{11})$ is non-increasing in π_1, therefore the power function for single-stage designs is non-decreasing in π_1.

Part 3. Prove the power function for two-stage designs $G_t(\mathbf{Q}, \pi_{11}|H)$ is monotone in π_1

The power function for two-stage designs is:

$$\begin{aligned} &G_t(\mathbf{Q}, \pi_{11}|H) = G_t(n_1, n, s_1, t_1, s, t, \pi_{11}, \pi_1, \pi_2) \\ =&1 - D(s_1, t_1; n_1, \pi_1, \pi_2, \pi_{11}) - D(s, t; n, \pi_1, \pi_2, \pi_{11}) \\ &+ \sum_{i=0}^{s_1} \sum_{j=0}^{t_1} \{D(s - i, t - j; n_2, \pi_1, \pi_2, \pi_{11}) \times p(i, j; n_1, \pi_1, \pi_2, \pi_{11})\}. \end{aligned}$$

Step 1: when $n_1 = 0$, the design is reduced to the single stage design and we have proved that the power function is non-decreasing in π_1.

Step 2: Assume that for $n_1 = k \geq 0$ and any $n \geq k + 1, s_1, t_1, s, t$, the power function $G_t(n_1 = k, n, s_1, t_1, s, t, \pi_{11}, \pi_1, \pi_2)$ is non-decreasing in π_1,

$$G_t'(k, n, s_1, t_1, s, t, \pi_{11}, \pi_1, \pi_2) = \frac{\partial}{\partial \pi_1} G_t(k, n, s_1, t_1, s, t, \pi_{11}, \pi_1, \pi_2) \geq 0,$$

then for $n_1 = k + 1$ and any $n \geq k + 2, s_1, t_1, s, t$, the power function can be expressed by conditioning on the number of responses for Endpoint 1 and Endpoint 2 for the first patient, who is surely in the first stage,

$$\begin{aligned} &G_t(k + 1, n, s_1, t_1, s, t, \pi_{11}, \pi_1, \pi_2) \\ &= 1 - P\left(X \leq s, Y \leq t\right) - P\left(X_1 \leq s_1, Y_1 \leq t_1 | X_1 \sim Bin(k + 1, \pi_1), Y_1 \sim Bin(k + 1, \pi_2), \pi_{11}\right) \end{aligned}$$

$$
\begin{aligned}
&\quad + P(X \le s, Y \le t, X_1 \le s_1, Y_1 \le t_1) \\
&= 1 - P(\underline{X_{11}} + X_{21} + \cdots + X_{n1} \le s, \underline{X_{12}} + X_{22} + \cdots + X_{n2} \le t) \\
&\quad - P(\underline{X_{11}} + X_{21} + \cdots + X_{(k+1),1} \le s_1, \underline{X_{12}} + X_{22} + \cdots + X_{(k+1),2} \le t_1) \\
&\quad + P(\underline{X_{11}} + X_{21} + \cdots + X_{n1} \le s, \underline{X_{12}} + X_{22} + \cdots + X_{n2} \le t, \\
&\qquad \underline{X_{11}} + X_{21} + \cdots + X_{(k+1),1} \le s_1, \underline{X_{12}} + X_{22} + \cdots + X_{(k+1),2} \le t_1) \\
&= 1 - \sum_{h=0,1} \sum_{m=0,1} \{P(X_{21} + \cdots + X_{n1} \le s-h, X_{22} + \cdots + X_{n2} \le t-m) P_{hm}\} \\
&\quad - \sum_{h=0,1} \sum_{m=0,1} \{P(X_{21} + \cdots + X_{(k+1),1} \le s_1-h, X_{22} + \cdots + X_{(k+1),2} \le t_1-m) P_{hm}\} \\
&\quad + \sum_{h=0,1} \sum_{m=0,1} \{P(X_{21} + \cdots + X_{n1} \le s-h, X_{22} + \cdots + X_{n2} \le t-m, \\
&\qquad X_{21} + \cdots + X_{(k+1),1} \le s_1-h, X_{22} + \cdots + X_{(k+1),2} \le t_1-m) P_{hm}\} \\
&= 1 - \sum_{h=0,1} \sum_{m=0,1} \{P_{hm}[1 - G_t(k, n-1, s_1-h, t_1-m, s-h, t-m, \pi_{11}, \pi_1, \pi_2)]\} \\
&= \sum_{h=0,1} \sum_{m=0,1} \{P_{hm}\, G_t(k, n-1, s_1-h, t_1-m, s-h, t-m, \pi_{11}, \pi_1, \pi_2)\}.
\end{aligned}
$$

Take partial derivative with respect to π_1 and we get

$$
\begin{aligned}
&G_t'(k+1, n, s_1, t_1, s, t, \pi_{11}, \pi_1, \pi_2) \\
&= \sum_{h=0,1} \sum_{m=0,1} \frac{\partial P_{hm}}{\partial \pi_1}\, G_t(k, n-1, s_1-h,\ t_1-m, s-h, t-m, \pi_{11}, \pi_1, \pi_2) \\
&\qquad + \sum_{h=0,1} \sum_{m=0,1} G_t'(k, n-1, s_1-h, t_1-m, s-h, t-m, \pi_{11}, \pi_1, \pi_2) P_{hm}
\end{aligned}
$$

By previous assumption that for $n_1 = k$, the second part is non-negative. Expanding the first part of the above equation,

$$
\begin{aligned}
&\sum_{h=0,1} \sum_{m=0,1} \frac{\partial P_{hm}}{\partial \pi_1}\, G_t(k, n-1, s_1-h,\ t_1-m, s-h, t-m, \pi_{11}, \pi_1, \pi_2) \\
&= G_t(k, n-1, s_1-1,\ t_1, s-1, t, \pi_{11}, \pi_1, \pi_2) - G_t(k, n-1, s_1, t_1, s, t, \pi_{11}, \pi_1, \pi_2) \\
&= \{1 - P(X \le s-1, Y \le t) - P(X_1 \le s_1-1, Y_1 \le t_1) \\
&\qquad + P(X \le s-1, Y \le t, X_1 \le s_1-1, Y_1 \le t_1)\} \\
&\quad - \{1 - P(X \le s, Y \le t) - P(X_1 \le s_1, Y_1 \le t_1) + P(X \le s, Y \le t, X_1 \le s_1, Y_1 \le t_1)\}
\end{aligned}
$$

$$= P(X = s, Y \leq t) + P(X_1 = s_1, Y_1 \leq t_1)$$
$$- \{P(X = s, Y \leq t, X_1 \leq s_1 - 1, Y_1 \leq t_1) + P(X \leq s, Y \leq t, X_1 = s_1, Y_1 \leq t_1)\}$$
$$\geq 0.$$

Hence, $G'_t(k+1, n, s_1, t_1, s, t, \pi_{11}, \pi_1, \pi_2) \geq 0$.

Combining the above step 1 and step 2, and by mathematical induction, the power function for two-stage designs $G_t(n_1, n, s_1, t_1, s, t, \pi_{11}, \pi_1, \pi_2)$ is non-decreasing in π_1.

Combining the above Parts 1-3, the power function, no matter for one-stage designs or for two-stage designs, is non-decreasing in π_1.

By symmetry, the power function is also non-decreasing in π_2.

□

Theorem 3.4.1 shows that the possible values of the power function under H_0, (that is, the possible values of the joint type I error), in the space constructed by the composite null hypothesis $H_{0:}\ \pi_1 \leqslant p_1^{(0)}$ and $\pi_2 \leqslant p_2^{(0)}$ have its maximum value at the pair point constructed by the simple null hypothesis $H_{0:}\ \pi_1 = p_1^{(0)}$ and $\pi_2 = p_2^{(0)}$. To satisfy the type I error constraint, we only need to consider designs for the simple null hypothesis $H_{0:}\ \pi_1 = p_1^{(0)}$ and $\pi_2 = p_2^{(0)}$.

Previous composite alternative hypothesis $H_{A:}\ \pi_1 \geqslant p_1^{(A)}$ or $\pi_2 \geqslant p_2^{(A)}$ can be divided into 3 separate spaces:

$$\begin{aligned} H_{A1:} &\quad \pi_1 \geqslant p_1^{(A)} \quad \text{and } 0 \leqslant \pi_2 < p_2^{(A)}, \\ H_{A2:}\ &0 \leqslant \pi_1 < p_1^{(A)} \text{ and } \quad \pi_2 \geqslant p_2^{(A)}, \\ H_{A3:} &\quad \pi_1 \geqslant p_1^{(A)} \quad \text{and} \quad \pi_2 \geqslant p_2^{(A)}. \end{aligned}$$

The possible values of the power function in the space defined by $H_{A1:}\ \pi_1 \geqslant p_1^{(A)}$ and $0 \leqslant \pi_2 < p_2^{(A)}$ have its minimum value at $\pi_1 = p_1^{(A)}$ and $\pi_2 = 0$ mathematically, but we use the value of the power function at $\pi_1 = p_1^{(A)}$ and $\pi_2 = p_2^{(0)}$ instead since the latter value makes more sense for clinicians. The latter value is the statistical power when the drug or regimen is with desired clinical response rate in efficacy Endpoint 1 and with non-zero uninteresting

response rate in efficacy Endpoint 2. Similarly, the minimum power in the space defined by $H_{A2:}\ 0 \leqslant \pi_1 < p_1^{(A)}$ and $\pi_2 \geqslant p_2^{(A)}$ is at $\pi_1 = p_1^{(0)}$ and $\pi_2 = p_2^{(A)}$ and the minimum power in the space defined by $H_{A3:}\ \pi_1 \geqslant p_1^{(A)}$ and $\pi_2 \geqslant p_2^{(A)}$ is at $\pi_1 = p_1^{(A)}$ and $\pi_2 = p_2^{(A)}$.

That is the rationale we, from here on, will only consider the error constraints for the simple null hypothesis:

$$H_{0:}\ \ \pi_1 = p_1^{(0)} \text{ and } \pi_2 = p_2^{(0)}$$

versus the simple alternative hypotheses:

$$\begin{aligned} H_{A1:}&\ \ \pi_1 = p_1^{(A)} \text{ and } \pi_2 = p_2^{(0)}, \\ H_{A2:}&\ \ \pi_1 = p_1^{(0)} \text{ and } \pi_2 = p_2^{(A)}, \\ H_{A3:}&\ \ \pi_1 = p_1^{(A)} \text{ and } \pi_2 = p_2^{(A)}. \end{aligned}$$

Note: given $\mathbf{Q}$ and π_{11},

$$\begin{aligned} &G(\mathbf{Q}, \pi_{11}|H_{A1}) < G(\mathbf{Q}, \pi_{11}|H_{A3}),\ \text{that is, } power(H_{A1}) < power(H_{A3}), \\ &G(\mathbf{Q}, \pi_{11}|H_{A2}) < G(\mathbf{Q}, \pi_{11}|H_{A3}),\ \text{that is, } power(H_{A2}) < power(H_{A3}). \end{aligned}$$

3.4.2 With respect to π_{11}

In general case with π_{11} unspecified, it seems that the joint type I error for two-stage designs (the value of the power function under H_0) is not monotone in π_{11} (see Figure 3.2 for example), though the joint type I error for one-stage designs is still monotone in π_{11}).

Note: the reason why ours is not monotone in π_{11} while the power function of two-stage designs in Bryant and Day (1995) is monotone in π_{11} is as follows. The power function in Bryant and Day(1995) is essentially:

$$\begin{aligned} &G(\mathbf{Q}, \pi_{11}|H_{ij}) = Pr\{\text{Recommend Treatment}|H_{ij}, \mathbf{Q}, \pi_{11}\} \\ =&Pr(X_1 > s_1, Y_1 > t_1, X > s, Y > t) \end{aligned}$$

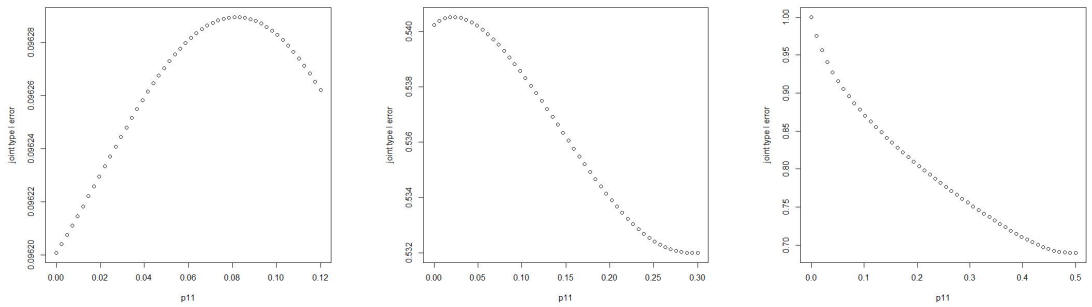


Figure 3.2: Relationship between the Joint Type I error rate (y axis) for two-stage designs and π_{11}(x axis).
(left): Given $(Q, \pi_1, \pi_2) = (n_1, n, s_1, t_1, s, t, \pi_1, \pi_2)$ =(15, 29, 8, 6, 15, 10, 0.2, 0.3);
(mid): Given $(Q, \pi_1, \pi_2) = (n_1, n, s_1, t_1, s, t, \pi_1, \pi_2) = (15, 29, 8, 6, 15, 10, 0.3, 0.45)$;
(right): Given $(Q, \pi_1, \pi_2) = (n_1, n, s_1, t_1, s, t, \pi_1, \pi_2) = (15, 29, 8, 6, 15, 10, 0.5, 0.5)$.

while the power function for two-stage designs in our study is more complicated, and consists of three parts:

$$
\begin{aligned}
&G(\mathbf{Q}, \pi_{11}|H) = Pr(\text{Recommend Treatment}|\mathbf{Q}, \pi_{11}, H) \\
=&Pr(X_1 > s_1 \text{ or } Y_1 > t_1), (X > s \text{ or } Y > t)) \\
=&1 - Pr(X \leq s, Y \leq t) - Pr(X_1 \leq s_1, Y_1 \leq t_1) \\
&+ \underline{Pr(X \leq s, Y \leq t, X_1 \leq s_1, Y_1 \leq t_1)}
\end{aligned}
$$

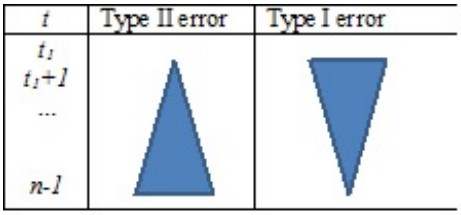
3.4.3 With respect to t or s

Theorem 3.4.2 *Given fixed π_1, π_2 and π_{11}, the power function is a non-increasing function of t given (n, n_1, s_1, t_1, s).*

It is straightforward to prove.

Application of **Theorem 3.4.2**: In searching algorithm, we search for optimal value of t in descending order and when this parameter constellation $\mathbf{Q} = (n, n_1, s_1, t_1, s, t)$ with this maximum value of t satisfying type II error constraints is found, this parameter constellation $\mathbf{Q}$ will be further judged for other conditions.

I used this property in the searching process—searching for optimal t value in descending order of t and find the maximum value of t which satisfies the type II error constraint. Since s is in similar position of t, I search for optimal s value in descending order of s as well.



The same applies for t_1 and s_1. That is, type II error is a non-decreasing function of t_1 given (n, n_1, s_1, s, t), and the type II error is a non-decreasing function of s_1 given (n, n_1, t_1, s, t).

3.5 Properties of the expected sample size under H_0 for the two stage designs with bivariate endpoints

There may be many feasible designs that satisfy type I/II error constraints, so additional optimality criteria are needed to select one of these feasible designs. e.g. :

$$\text{minimizing } E(N|H_0, \mathbf{Q}, \pi_{11}).$$

Let's look at the properties of $E(N|H, \mathbf{Q}, \pi_{11})$ for two-stage designs first.

$$E(N|H, \mathbf{Q}, \pi_{11}) = n_1 PET_H + n(1 - PET_H) = n - PET_H(n - n_1)$$

where

$$PET_H = Pr(X_1 \leq s_1, Y_1 \leq t_1|H, \mathbf{Q}, \pi_{11}) = D(s_1, t_1; n_1, \pi_1, \pi_2, \pi_{11})$$

Theorem 3.5.1 *For two-stage designs, given* $\boldsymbol{Q}$ *and* π_{11}, $E(N|H, \boldsymbol{Q}, \pi_{11})$ *is a non-decreasing function of* π_1 *and of* π_2 *as well.*

Proof: Recall that in the proof of Part 1 of Theorem 3.4.1, we have already proved that $D(s, t; n, \pi_1, \pi_2, \pi_{11})$ is non-increasing in π_1 for any $n \geq 1$, so $E(N|H, \mathbf{Q}, \pi_{11})$ is non-decreasing in π_1.

The same applies to the relationship between $E(N|H, \mathbf{Q}, \pi_{11})$ and π_2.

Theorem 3.5.2 *For two-stage designs, given* $\boldsymbol{Q}$ *and fixed* π_1 *and of* π_2, $E(N|H, \boldsymbol{Q}, \pi_{11})$ *is a non-increasing function of* π_{11}.

Proof: The method of Mathematical Induction (MI) will be used to prove that $D(s,t;n,\pi_1,\pi_2,\pi_{11})$ is non-decreasing in π_{11} for any $n \geq 1$. Hence this Theorem follows.

Steps to prove the monotonicity of $D(s,t;n,\pi_1,\pi_2,\pi_{11})$ in π_{11}:

Step 1: when $n = 1$,

$$D(s,t;n,\pi_1,\pi_2,\pi_{11}) = Pr(X \leq s, Y \leq t | X \sim Ber(\pi_1), Y \sim Ber(\pi_2), \pi_{11}).$$

If $(s,t) = (0,0)$,

$$\begin{aligned} D(s,t;n,\pi_1,\pi_2,\pi_{11}) &= Pr(X = 0, Y = 0) = 1 - \pi_1 - \pi_2 + \pi_{11}, \\ \frac{\partial}{\partial \pi_{11}} D(s,t;n,\pi_1,\pi_2,\pi_{11}) &= 1; \end{aligned}$$

If $(s,t) = (1,0)$,

$$\begin{aligned} D(s,t;n,\pi_1,\pi_2,\pi_{11}) &= Pr(X \leq 1, Y = 0) \\ &= (1 - \pi_1 - \pi_2 + \pi_{11}) + (\pi_1 - \pi_{11}) = 1 - \pi_2, \\ \frac{\partial}{\partial \pi_{11}} D(s,t;n,\pi_1,\pi_2,\pi_{11}) &= 0; \end{aligned}$$

If $(s,t) = (0,1)$,

$$\begin{aligned} D(s,t;n,\pi_1,\pi_2,\pi_{11}) &= Pr(X = 0, Y \leq 1) = 1 - \pi_1, \\ \frac{\partial}{\partial \pi_{11}} D(s,t;n,\pi_1,\pi_2,\pi_{11}) &= 0; \end{aligned}$$

If $(s,t) = (1,1)$,

$$\begin{aligned} D(s,t;n,\pi_1,\pi_2,\pi_{11}) &= Pr(X \leq 1, Y \leq 1) = 1, \\ \frac{\partial}{\partial \pi_{11}} D(s,t;n,\pi_1,\pi_2,\pi_{11}) &= 0. \end{aligned}$$

In summary, when $n = 1$, $D(s, t; n, \pi_1, \pi_2, \pi_{11})$ is a non-decreasing function of π_{11}.

Step 2:

Assume that for $n = k \geq 1$, $D(s, t; k, \pi_1, \pi_2, \pi_{11})$ is a non-decreasing function of π_{11}, that is (denote by $D'_{\pi_{11}}$ the partial derivative of D with respect to π_{11},),

$$D'_{\pi_{11}}(s, t; k, \pi_1, \pi_2, \pi_{11}) = \frac{\partial}{\partial \pi_{11}} D(s, t; k, \pi_1, \pi_2, \pi_{11}) \geq 0.$$

Then, when $n = k+1$, consider an independent sequence of $k+1$ patients. Let P_{hm} denote the probability of the response pattern for the first patient,

$$P_{hm} = Pr(X_{11} = h, X_{12} = m) = \begin{cases} \pi_{11}, & \text{when } h = m = 1 \\ \pi_1 - \pi_{11}, & \text{when } h = 1, m = 0 \\ \pi_2 - \pi_{11}, & \text{when } h = 0, m = 1 \\ 1 - \pi_1 - \pi_2 + \pi_{11}, & \text{when } h = m = 0, \end{cases}$$

and $X_{i1} \sim Ber(\pi_1), X_{i2} \sim Ber(\pi_2)$ are the endpoint 1 response status and endpoint 2 response status from the i^{th} patient for $i = 1, \cdots, k+1$, respectively. By conditioning on the number of responses for Endpoint 1 and Endpoint 2 for the first patient, we have

$$\begin{aligned} & D(s, t; \underline{k+1}, \pi_1, \pi_2, \pi_{11}) \\ & = Pr(X \leq s, Y \leq t \mid X \sim Bin(\underline{k+1}, \pi_1), Y \sim Bin(\underline{k+1}, \pi_2), \pi_{11}) \\ & = Pr(X_{11} + X_{21} + \cdots + X_{(k+1),1} \leq s, X_{12} + X_{22} + \cdots + X_{(k+1),2} \leq t) \\ & = \sum_{h=0,1} \sum_{m=0,1} Pr(X_{21} + \cdots + X_{(k+1),1} \leq s - h, \\ & \qquad\qquad X_{22} + \cdots + X_{(k+1),2} \leq t - m | X_{11} = h, X_{12} = m) P_{hm} \\ & = \sum_{h=0,1} \sum_{m=0,1} D(s - h,\ t - m; \underline{k}, \pi_1, \pi_2, \pi_{11}) \times P_{hm} \end{aligned}$$

Take partial derivative of D with respect to π_{11}, we get

$$\begin{aligned} & D'_{\pi_{11}}(s, t; k+1, \pi_1, \pi_2, \pi_{11}) \\ & = \sum_{h=0,1} \sum_{m=0,1} \frac{\partial P_{hm}}{\partial \pi_{11}} D(s - h,\ t - m; k, \pi_1, \pi_2, \pi_{11}) \end{aligned}$$

$$+ \sum_{h=0,1} \sum_{m=0,1} D'_{\pi_{11}}(s-h,\ t-m; k, \pi_1, \pi_2, \pi_{11}) P_{hm}$$

$$= \sum_{h=0,1} \sum_{m=0,1} (-1)^{h+m} D(s-h,\ t-m; k, \pi_1, \pi_2, \pi_{11})$$

$$+ \sum_{h=0,1} \sum_{m=0,1} D'_{\pi_{11}}(s-h,\ t-m; k, \pi_1, \pi_2, \pi_{11}) P_{hm}$$

By the previous assumption for $n = k$, $D'_{\pi_{11}}(s-h,\ t-m; k, \pi_1, \pi_2, \pi_{11}) \geq 0$, so the second part of the above equation is non-negative. And the first part

$$\sum_{h=0,1} \sum_{m=0,1} (-1)^{h+m} D(s-h,\ t-m; k, \pi_1, \pi_2, \pi_{11})$$

$$= D(s,t; k, \pi_1, \pi_2, \pi_{11}) - D(s-1,t; k, \pi_1, \pi_2, \pi_{11}) - D(s,t-1; k, \pi_1, \pi_2, \pi_{11})$$

$$+ D(s-1,t-1; k, \pi_1, \pi_2, \pi_{11})$$

$$= Pr(X = s, Y \leq t \mid X \sim Bin(k, \pi_1), Y \sim Bin(k, \pi_2), \pi_{11})$$

$$- Pr(X = s, Y \leq t-1 \mid X \sim Bin(k, \pi_1), Y \sim Bin(k, \pi_2), \pi_{11})$$

$$= Pr(X = s, Y = t | X \sim Bin(k, \pi_1), Y \sim Bin(k, \pi_2), \pi_{11})$$

$$\geq 0,$$

$D'_{\pi_{11}}(s,t; k+1, \pi_1, \pi_2, \pi_{11}) \geq 0$.

In summary, when $n = k+1$, $D(s,t; k+1, \pi_1, \pi_2, \pi_{11})$ is a non-decreasing function of π_{11}.

Combining the above step 1 and step 2, and by mathematical induction, $D(s,t; n, \pi_1, \pi_2, \pi_{11})$ is a non-decreasing function of π_{11} for any $n \geq 1$. Since

$$E(N|H_0, \mathbf{Q}, \pi_{11}) = n - n_2 D(s_1, t_1; n_1, \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(0)}, \pi_{11}),$$

so $E(N|H_0, \mathbf{Q}, \pi_{11})$ is non-increasing in π_{11}. □

Theorem 3.5.3 *For two-stage designs, given fixed π_1, π_2 and π_{11}, $E(N|H, \boldsymbol{Q}, \pi_{11})$ is a non-increasing function of s_1 and of t_1 as well.*

Proof is Straightforward.

3.6 Determination of design parameters based on marginal and overall Type I and Type II error constrains

In trial designs for multiple endpoints, taking bivariate endpoints as an example here, we need to consider not only the overall type I error constraint based on the joint distribution of the bivariate endpoints, but also the marginal type I error constraint for each endpoint.

Let α, β_1, β_2, β be error bounds specified by the investigators: α bounds the probability of erroneously recommending a treatment whose two primary alternative efficacy endpoints both have unacceptable response rate; β_1 bounds the probability of failing to recommend a treatment whose first of the two primary alternative efficacy endpoints has adequate response rate, second inadequate; β_2 bounds the probability of failing to recommend a treatment whose first of the two primary alternative efficacy endpoints has inadequate response rate, second adequate; β bounds the probability of failing to recommend a treatment when both of the two primary alternative efficacy endpoints have adequate response rates. Let $G(\mathbf{Q}, \pi_{11}|H)$ denote the power function of the design under H. The design parameters $\mathbf{Q} = (n_1, n, s_1, t_1, s, t)$ may be determined by solving the following mathematical program

$$\min_{Q} \max_{\pi_{11}} E(N|Q, \pi_{11}, H_0 : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(0)}), \tag{3.2}$$

subject to

$$\begin{aligned}
&\max_{\pi_{11}} G(\mathbf{Q}, \pi_{11}|H_0 : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(0)}) \\
&= \max_{\pi_{11}} P\{(X_1 > s_1 \text{ or } Y_1 > t_1), (X > s \text{ or } Y > t)|H_0 : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(0)}\} \leq \alpha \qquad (3.3a) \\
&P(X_1 > s_1 \text{ and } X > s|H_{0_1} : \pi_1 = p_1^{(0)}) \leq \alpha \text{ and} \\
&P(Y_1 > t_1 \text{ and } Y > t|H_{0_2} : \pi_2 = p_2^{(0)}) \leq \alpha \qquad (3.3b) \\
&\min_{\pi_{11}} G(Q, \pi_{11}|H_{A_1} : \pi_1 = p_1^{(A)}, \pi_2 = p_2^{(0)}) \geqslant 1 - \beta_1, \qquad (3.3c) \\
&\min_{\pi_{11}} G(Q, \pi_{11}|H_{A_2} : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(A)}) \geqslant 1 - \beta_2 \qquad (3.3d) \\
&\min_{\pi_{11}} G(Q, \pi_{11}|H_{A_3} : \pi_1 = p_1^{(A)}, \pi_2 = p_2^{(A)}) \geqslant 1 - \beta \qquad (3.3e)
\end{aligned}$$

The following brief proof would suggest that the above constraints in (3.3) can be further

simplified based on the relationship between joint and marginal probabilities.

In single-stage designs, under the simple null hypothesis,

$$\begin{aligned}
\text{Joint type I error} &= G_s(\mathbf{Q}, \pi_{11}|H_0) \\
&= Pr(X > s \text{ or } Y > t|H_0 : \pi_1 = p_1^{(0)} \text{ and } \pi_2 = p_2^{(0)}) \\
&= 1 - Pr(X \leq s \text{ and } Y \leq t|H_0 : \pi_1 = p_1^{(0)} \text{ and } \pi_2 = p_2^{(0)}) \\
\text{Marginal type I error for endpoint 1} &= Pr(X > s|H_0 : \pi_1 = p_1^{(0)}) \\
&= 1 - Pr(X \leq s|H_0 : \pi_1 = p_1^{(0)})
\end{aligned}$$

It is straightforward that the marginal type I error for either endpoint is less than or equal to the joint type I error in single-stage designs based on the sketch of the joint p.m.f of (X, Y) in Table 3.5. If we put the error constraint on the joint type I error, it will guarantee that the marginal type I error is also constrained. That is, marginal type I error for either endpoint $\leq$ joint type I error $\leq \alpha$.

For two-stage designs,

$$\text{Joint type I error} = G_t(\mathbf{Q}, \pi_{11}|H_0 : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(0)})$$

$$= \underline{P\{(X_1 > s_1 \textit{ or } Y_1 > t_1), (X > s \textit{ or } Y > t)|H_0 : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(0)}\}} \tag{3.4}$$

$$\text{Marginal type I error for endpoint 1} = P(X_1 > s_1 \textit{ and } X > s|H_{0_1} : \pi_1 = p_1^{(0)}) \tag{3.5}$$

$$\text{Marginal type I error for endpoint 2} = P(Y_1 > t_1 \textit{ and } Y > t|H_{0_2} : \pi_2 = p_2^{(0)}) \tag{3.6}$$

We would intuitively prove that the underlined probability term in (3.4) is always greater than or equal to each of the two probabilities in (3.5) and (3.6). Let

event A1=$(X_1 > s_1)$, event B1=$(Y_1 > t_1)$, event A=$(X > s)$, event B=$(Y > t)$,

then

$$\{(A1 \cup B1) \cap (A \cup B)\} \supset (A1 \cap A),$$

$$\{(A1 \cup B1) \cap (A \cup B)\} \supset (B1 \cap B).$$

Therefore,

$$P\{(X_1 > s_1 \text{ or } Y_1 > t_1), (X > s \text{ or } Y > t) | H_0 : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(0)}\}$$
$$> P(X_1 > s_1 \text{ and } X > s | H_{0_1} : \pi_1 = p_1^{(0)}),$$
$$P\{(X_1 > s_1 \text{ or } Y_1 > t_1), (X > s \text{ or } Y > t) | H_0 : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(0)}\}$$
$$> P(Y_1 > t_1 \text{ and } Y > t | H_{0_2} : \pi_2 = p_2^{(0)}).$$

So if we put the error constraint on joint type I error in two-stage designs as stated in (3.4), the error constraints on marginal type I errors are guaranteed to be satisfied.

By combining the above brief proof on the relationship between marginal and joint type I error rates in both single-stage designs and two-stage designs, the mathematical program in (3.2) and (3.3) can be simplified as the following:

$$\min_{Q} \max_{\pi_{11}} E(N | Q, \pi_{11}, H_0 : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(0)}), \tag{3.7}$$

subject to

$$\max_{\pi_{11}} G(Q, \pi_{11} | H_0 : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(0)}) \leq \alpha, \tag{3.8a}$$
$$\min_{\pi_{11}} G(Q, \pi_{11} | H_{A_1} : \pi_1 = p_1^{(A)}, \pi_2 = p_2^{(0)}) \geqslant 1 - \beta_1, \tag{3.8b}$$
$$\min_{\pi_{11}} G(Q, \pi_{11} | H_{A_2} : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(A)}) \geqslant 1 - \beta_2, \tag{3.8c}$$
$$\min_{\pi_{11}} G(Q, \pi_{11} | H_{A_3} : \pi_1 = p_1^{(A)}, \pi_2 = p_2^{(A)}) \geqslant 1 - \beta. \tag{3.8d}$$

If $\beta_1 = \beta_2 = \beta$, then (3.8d) is included in (3.8b) or (3.8c).

3.7 Methods and algorithms

For specified values of α, β, β_1, β_2, $p_1^{(0)}$, $p_2^{(0)}$, $p_1^{(A)}$, and $p_2^{(A)}$, we have determined optimal designs by enumeration using exact bivariate binomial probabilities. For each value of total sample size n and each value of n_1 in the range $[1, n-1]$, we will find the feasible solutions

$Q = (n_1, n, s_1, t_1, s, t)$ which satisfies the error constraints (3.8a) – (3.8d) and then use the optimality criteria, that is, minimizing the expected sample size under the null hypothesis as shown in (3.7), to find the "best" solution.

Technically, this is found by searching over the ranges of s_1 in $[0, n_1 - 1]$ and t_1 in $[0, n_1 - 1]$. For each value of s_1, t_1, and s, we determine the maximum value of t that satisfies the type II error constraints since the type II error is a monotone function of t given (n_1, n, s_1, t_1, s). We then examine whether that set of parameters $Q = (n_1, n, s_1, t_1, s, t)$ satisfies the type I error constraint. If it does, save it to the set of feasible solutions. Then among all the feasible solutions, locate the minimum expected sample size of n under the null hypothesis and its corresponding Q. That is the globally optimal design.

For this two-stage design for cancer clinical trials with two alternative primary endpoints, how to determine the searching range of n and how to clearly specify the optimality criteria require more attention.

3.7.1 Determination of the searching range of n

For "alternative primary endpoints", the central issue is to control the false-positive rate (type I error) at the study level since there are many chances to declare efficacy. It is a traditional multiplicity problem. A common approach to handle this traditional multiplicity problem is to adjust the significance level downward for individual testings so that the overall false-positive rate can be maintained at the desirable level (Sankoh et al., 1997).

We calculate sample sizes for endpoint 1 and endpoint 2, respectively, using Dr. Richard Simon's formula (as shown in the following) under the unadjusted power and adjusted (Bonferroni adjustment) nominal type I error rate:

$$n = \overline{p}(1 - \overline{p}) \left[\frac{z_{1-\alpha} + z_{1-\beta}}{p^{(A)} - p^{(0)}} \right]^2 \tag{3.9}$$

where $\overline{p} = (p^{(0)} + p^{(A)})/2$, $p^{(0)}$ and $p^{(A)}$ are probabilities of event under the null and alternative hypotheses for each endpoint, respectively. The α level here in this formula is the adjusted nominal type I error rate, say, 0.025 if the total nominal type I error rate for the entire study with two alternative primary endpoints is 0.05. β here is the unadjusted

nominal type II error rate for the study. The smaller of the two sample sizes (which corresponds to larger 'effect size') is multiplied by 0.85 and 1.5 to serve as the lower and upper limits of the searching range for n respectively.

3.7.2 Optimality Criteria

We found that more than one, actually many, feasible solutions share the same minimum expected sample size of N under the null hypothesis because of discreteness of the underlying bivariate binomial distribution and the small difference in the value of $\mathrm{E}(N|\ H_0)$ between feasible solutions sharing the same (n, n_1, s_1, t_1). So the optimality criteria for the optimal design now is:

1. Minimum $E(N|H_0)$;
2. Maximum type I error(closer to nominal level) since there are three type II errors and the directions of the magnitude of them are not the same in most time.

3.7.3 Naive exhaustive searching algorithm

Given investigator-specified values of (α, β_1, β_2, β, $p_1^{(0)}$, $p_2^{(0)}$, $p_1^{(A)}$, $p_2^{(A)}$), use the *exact bivariate binomial probability* to exhaustively search among loops (from outer loop to inner loops: $n \to n_1 \to s_1 \to t_1 \to s \to t$) to find feasible solutions to satisfy error constraints.

searching range for n : follow the strategy mentioned in section 3.7.1 .

$$\text{for } n_1 : [\frac{n}{4}, \frac{3n}{4}];$$
$$\text{for } s_1 : [n_1 p_1^{(0)}, n_1 p_1^{(A)}];$$
$$\text{for } t_1 : [n_1 p_2^{(0)}, n_1 p_2^{(A)}];$$
$$\text{for } s : [\max(s_1 + 1, np_1^{(0)}), np_1^{(A)}];$$
$$\text{for } t : [\max(t_1 + 1, np_2^{(0)}), np_2^{(A)}].$$

Then among those feasible solutions, apply the *optimality criteria*(as mentioned in section 3.7.2) to locate the optimal design.

3.7.4 Refined searching algorithm starting from independence assumption

Initially we tried to do exhaustive searching directly (just using the naive exhaustive searching algorithm as mentioned in section 3.7.3). However, due to the introduction of the correlation parameter π_{11} into the bivariate joint distribution, the time cost in optimization and exhaustive searching has increased dramatically, especially when $n > 30$.

Theorems 2.3.3 and 2.3.4 from Bryant and Day (1995) have inspired us to adopt a pre-screening strategy via starting searching assuming the two alternative primary efficacy endpoints are independent. Find those feasible solutions satisfying type I/II error constraints under independence assumption, and sort them by the optimality criteria under independence assumption. Next, among the top 5% of the sorted feasible solutions, relax the independence assumption, do computation-intensive calculations of real maximized type I/II error rates allowing π_{11} to assume any values in its defined range, and search and locate the optimal design after applying the optimality criteria.

3.8 Results

The following tables 3.6 to 3.9 show optimal two-stage designs for cancer clinical trials with two alternative primary endpoints for a variety of design parameters with $p_1^{(0)} \leq p_2^{(0)}$. The following notations are used in these tables

$$
\begin{aligned}
G_0(Q, \pi_{11}) &= G(Q, \pi_{11} | H_0 : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(0)}), \\
G_1(Q, \pi_{11}) &= G(Q, \pi_{11} | H_{A_1} : \pi_1 = p_1^{(A)}, \pi_2 = p_2^{(0)}), \\
G_2(Q, \pi_{11}) &= G(Q, \pi_{11} | H_{A_2} : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(A)}), \\
G_3(Q, \pi_{11}) &= G(Q, \pi_{11} | H_{A_3} : \pi_1 = p_1^{(A)}, \pi_2 = p_2^{(A)}).
\end{aligned}
$$

Table 3.6 is for cancer trials with $\delta = 0.20 = p_1^{(A)} - p_1^{(0)} = p_2^{(A)} - p_2^{(0)}$ and Table 3.7 is for trials with $\delta = 0.15 = p_1^{(A)} - p_1^{(0)} = p_2^{(A)} - p_2^{(0)}$. Table 3.8 is for trials with $\delta_1 = 0.15 = p_1^{(A)} - p_1^{(0)}, \delta_2 = 0.20 = p_2^{(A)} - p_2^{(0)}$, and Table 3.9 is for trials with $\delta_1 = 0.20 = p_1^{(A)} - p_1^{(0)}, \delta_2 = 0.15 = p_2^{(A)} - p_2^{(0)}$. The operating characteristics for each of the optimal two-stage sequential design (including the maximized type I error rate, minimized

powers and the minimized value of maximum possible expected sample size under the null hypothesis in the defined range of π_{11}) are presented as well.

Table 3.6: Given $(\alpha, \beta_1, \beta_2) = (0.05, 0.20, 0.20)$. $\delta = 0.20$. Reject H_0 if $(X_1 > s_1$ or $Y_1 > t_1)$ and $(X > s$ or $Y > t)$.

$p_1^{(0)}$	$p_1^{(A)}$	$p_2^{(0)}$	$p_2^{(A)}$	(n, n_1, s_1, t_1, s, t)	$\max_{\pi_{11}} G_0(Q, \pi_{11})$	$\min_{\pi_{11}} G_1(Q, \pi_{11})$	$\min_{\pi_{11}} G_2(Q, \pi_{11})$	$\min_{\pi_{11}} G_3(Q, \pi_{11})$	$\max_{\pi_{11}} E(N\|H_0, Q, \pi_{11})$
0.05	0.25	0.05	0.25	(25, 12, 1, 1, 3, 3)	0.0497	0.8061	0.8061	0.8062	15.0
		0.1	0.3	(27, 15, 1, 2, 4, 5)	0.0491	0.8161	0.8101	0.8688	19.0
		0.2	0.4	(37, 17, 2, 4, 5, 11)	0.0490	0.8130	0.8040	0.9010	22.6
		0.3	0.5	(39, 19, 2, 6, 6, 16)	0.0483	0.8308	0.8041	0.9236	26.8
		0.4	0.6	(53, 21, 3, 10, 7, 26)	0.0470	0.8045	0.8010	0.9223	27.1
		0.5	0.7	(39, 26, 3, 15, 6, 24)	0.0496	0.8623	0.8018	0.9491	28.6
		0.6	0.8	(38, 18, 2, 12, 6, 27)	0.0477	0.8066	0.8119	0.9455	23.2
		0.7	0.9	(28, 13, 1, 10, 4, 23)	0.0494	0.8082	0.8045	0.9513	17.9
0.1	0.3	0.1	0.3	(38, 14, 2, 2, 7, 7)	0.0460	0.8053	0.8053	0.8054	21.2
		0.2	0.4	(41, 22, 4, 6, 8, 12)	0.0499	0.8053	0.8042	0.8760	25.6
		0.3	0.5	(47, 22, 4, 8, 9, 19)	0.0462	0.8142	0.8013	0.9061	28.0
		0.4	0.6	(45, 22, 4, 10, 9, 23)	0.0491	0.8030	0.8035	0.9195	28.7
		0.5	0.7	(49, 18, 3, 10, 9, 30)	0.0482	0.8227	0.8004	0.9360	28.2
		0.6	0.8	(43, 18, 3, 12, 8, 31)	0.0456	0.8175	0.8005	0.9437	25.5
		0.7	0.9	(38, 14, 2, 11, 7, 31)	0.0489	0.8115	0.8215	0.9515	21.7
0.2	0.4	0.2	0.4	(50, 22, 6, 6, 15, 15)	0.0446	0.8025	0.8025	0.8027	29.3
		0.3	0.5	(54, 22, 6, 8, 16, 22)	0.0480	0.8110	0.8038	0.8777	31.9
		0.4	0.6	(55, 25, 7, 12, 17, 28)	0.0418	0.8008	0.8011	0.8969	32.8
		0.5	0.7	(59, 22, 6, 13, 18, 35)	0.0498	0.8035	0.8021	0.9157	32.4

Table 3.6 (Cont.). Given $(\alpha, \beta_1, \beta_2) = (0.05, 0.20, 0.20)$. $\delta = 0.20$. Reject H_0 if $(X_1 > s_1$ or $Y_1 > t_1)$ and $(X > s$ or $Y > t)$.

		0.6	0.8	(47, 22, 6, 15, 14, 34)	0.0494	0.8039	0.8074	0.9319	29.2
		0.7	0.9	(44, 22, 6, 18, 13, 36)	0.0471	0.8040	0.8192	0.9451	26.4
0.3	0.5	0.3	0.5	(59, 24, 9, 9, 24, 24)	0.0456	0.8031	0.8031	0.8036	34.5
		0.4	0.6	(55, 25, 9, 12, 23, 28)	0.0478	0.8057	0.8010	0.8765	35.2
		0.5	0.7	(62, 24, 9, 14, 25, 38)	0.0460	0.8031	0.8048	0.9063	35.8
		0.6	0.8	(54, 22, 8, 15, 22, 39)	0.0437	0.8034	0.8188	0.9243	32.9
		0.7	0.9	(47, 22, 8, 18, 19, 39)	0.0451	0.8008	0.8078	0.9366	28.2
0.4	0.6	0.4	0.6	(63, 25, 12, 12, 32, 32)	0.0492	0.8005	0.8005	0.8329	36.8
		0.5	0.7	(59, 27, 13, 16, 30, 36)	0.0496	0.8137	0.8100	0.8819	35.6
		0.6	0.8	(60, 23, 11, 16, 30, 43)	0.0464	0.8062	0.8118	0.9077	33.7
		0.7	0.9	(47, 22, 10, 18, 24, 39)	0.0479	0.8068	0.8083	0.9322	29.5
0.5	0.7	0.5	0.7	(57, 24, 14, 14, 35, 35)	0.0471	0.8014	0.8014	0.8017	34.1
		0.6	0.8	(50, 26, 15, 18, 31, 36)	0.0489	0.8047	0.8216	0.8890	32.7
		0.7	0.9	(48, 21, 12, 17, 29, 40)	0.0499	0.8086	0.8110	0.9199	28.6
0.6	0.8	0.6	0.8	(50, 22, 15, 15, 36, 36)	0.0488	0.8009	0.8009	0.8400	31.0
		0.7	0.9	(46, 21, 14, 17, 33, 38)	0.0459	0.8187	0.8125	0.9143	28.0
0.7	0.9	0.7	0.9	(39, 14, 11, 11, 32, 32)	0.0430	0.8024	0.8024	0.8028	21.8

Table 3.7: Given $(\alpha, \beta_1, \beta_2) = (0.05, 0.20, 0.20)$. $\delta = 0.15$. Reject H_0 if $(X_1 > s_1$ or $Y_1 > t_1)$ and $(X > s$ or $Y > t)$.

$p_1^{(0)}$	$p_1^{(A)}$	$p_2^{(0)}$	$p_2^{(A)}$	(n, n_1, s_1, t_1, s, t)	$\max_{\pi_{11}} G_0(Q, \pi_{11})$	$\min_{\pi_{11}} G_1(Q, \pi_{11})$	$\min_{\pi_{11}} G_2(Q, \pi_{11})$	$\min_{\pi_{11}} G_3(Q, \pi_{11})$	$\max_{\pi_{11}} E(N\|H_0, Q, \pi_{11})$
0.05	0.20	0.05	0.20	(36, 22, 2, 2, 4, 4)	0.0496	0.8011	0.8011	0.8013	24.5
		0.1	0.25	(48, 23, 2, 3, 6, 8)	0.0445	0.8027	0.8073	0.8632	30.0
		0.2	0.35	(57, 32, 3, 7, 8, 16)	0.0485	0.8015	0.8075	0.8967	41.0
		0.3	0.45	(74, 35, 4,12, 10, 28)	0.0468	0.8131	0.8011	0.9185	44.8
		0.4	0.55	(79, 38, 5, 17, 9, 38)	0.0487	0.8014	0.8015	0.9325	47.4
		0.5	0.65	(81, 33, 4, 18, 10, 47)	0.0474	0.8134	0.8019	0.9393	45.5
		0.6	0.75	(70, 27, 3, 17, 9, 48)	0.0477	0.8039	0.8076	0.9459	42.0
		0.7	0.85	(54, 26, 3, 19, 6, 43)	0.0486	0.8056	0.8004	0.9584	35.2
		0.8	0.95	(43, 16, 1, 14, 5, 38)	0.0488	0.8077	0.8149	0.9587	24.9
0.1	0.25	0.1	0.25	(51, 24, 3, 3, 9, 9)	0.0483	0.8017	0.8017	0.8020	34.5
		0.2	0.35	(74, 32, 5, 8, 13, 20)	0.0485	0.8115	0.8042	0.8829	42.9
		0.3	0.45	(76, 33, 5, 11, 14, 29)	0.0477	0.8100	0.8025	0.9069	48.5
		0.4	0.55	(76, 37, 6, 16, 14, 37)	0.0495	0.8051	0.8008	0.9211	50.6
		0.5	0.65	(77, 35, 5, 19, 15, 45)	0.0494	0.8000	0.8034	0.9296	50.5
		0.6	0.75	(77, 31, 5, 20, 13, 53)	0.0482	0.8141	0.8024	0.9436	46.2
		0.7	0.85	(77, 26, 4, 20, 13, 60)	0.0432	0.8024	0.8021	0.9417	39.8
		0.8	0.95	(53, 18, 2, 16, 9, 47)	0.0482	0.8055	0.8042	0.9540	30.9
0.2	0.35	0.2	0.35	(88, 38, 10, 10, 24, 24)	0.0495	0.8011	0.8011	0.8014	50.0
		0.3	0.45	(87, 37, 9, 13, 25, 33)	0.0499	0.8007	0.8117	0.8828	55.5

Table 3.7 (Cont.). Given $(\alpha, \beta_1, \beta_2) = (0.05, 0.20, 0.20)$. $\delta = 0.15$. Reject H_0 if $(X_1 > s_1$ or $Y_1 > t_1)$ and $(X > s$ or $Y > t)$.

		0.4	0.55	(93, 42, 11, 19, 26, 45)	0.0499	0.8115	0.8090	0.9053	57.8
		0.5	0.65	(96, 39, 10, 22, 27, 56)	0.0491	0.8066	0.8054	0.9165	56.7
		0.6	0.75	(90, 36, 9, 24, 25, 62)	0.0495	0.8160	0.8094	0.9309	53.8
		0.7	0.85	(87, 32, 8, 25, 24, 68)	0.0499	0.8008	0.8223	0.9393	48.0
		0.8	0.95	(56, 35, 7, 31, 16, 51)	0.0492	0.8017	0.8409	0.9570	44.7
0.3	0.45	0.3	0.45	(100, 44, 16, 16, 38, 38)	0.0491	0.8022	0.8022	0.8026	59.7
		0.4	0.55	(92, 48, 17, 22, 36, 45)	0.0476	0.8005	0.8002	0.8711	62.5
		0.5	0.65	(93, 45, 16, 25, 36, 55)	0.0488	0.8033	0.8043	0.8997	61.8
		0.6	0.75	(88, 38, 13, 25, 34, 61)	0.0500	0.8033	0.8022	0.9176	58.5
		0.7	0.85	(90, 37, 13, 29, 34, 71)	0.0481	0.8053	0.8049	0.9329	52.4
		0.8	0.95	(77, 30, 10, 27, 29, 70)	0.0480	0.8027	0.8063	0.9478	44.6
0.4	0.55	0.4	0.55	(96, 48, 22, 22, 47, 47)	0.0494	0.8011	0.8011	0.8015	63.8
		0.5	0.65	(98, 48, 22, 27, 48, 58)	0.0457	0.8017	0.8008	0.8729	64.0
		0.6	0.75	(95, 38, 17, 25, 46, 66)	0.0492	0.8043	0.8002	0.9025	61.4
		0.7	0.85	(86, 37, 16, 29, 42, 68)	0.0491	0.8054	0.8035	0.9292	55.6
		0.8	0.95	(79, 38, 17, 34, 38, 72)	0.0478	0.8014	0.8446	0.9532	48.7
0.5	0.65	0.5	0.65	(100, 43, 24, 24, 59, 59)	0.0459	0.8008	0.8008	0.8012	63.5
		0.6	0.75	(86, 47, 26, 31, 51, 60)	0.0499	0.8017	0.8048	0.8777	60.6
		0.7	0.85	(85, 40, 22, 31, 50, 68)	0.0454	0.8014	0.8059	0.9107	54.4
		0.8	0.95	(75, 29, 15, 26, 44, 68)	0.0490	0.8032	0.8229	0.9458	47.4

Table 3.7 (Cont.). Given $(\alpha, \beta_1, \beta_2) = (0.05, 0.20, 0.20)$. $\delta = 0.15$. Reject H_0 if $(X_1 > s_1$ or $Y_1 > t_1)$ and $(X > s$ or $Y > t)$.

0.6	0.75	0.6	0.75	(95, 43, 29, 29, 65, 65)	0.0498	0.8011	0.8011	0.8014	56.0
		0.7	0.85	(83, 35, 23, 27, 57, 66)	0.0465	0.8019	0.8066	0.8886	50.4
		0.8	0.95	(70, 31, 20, 28, 48, 63)	0.0493	0.8002	0.8021	0.9313	41.9
0.7	0.85	0.7	0.85	(76, 30, 23, 23, 60, 60)	0.0453	0.8036	0.8036	0.8040	44.3
		0.8	0.95	(57, 23, 17, 20, 45, 52)	0.0468	0.8032	0.8059	0.9119	35.9
0.8	0.95	0.8	0.95	(46, 14, 12, 12, 41, 41)	0.0497	0.8106	0.8106	0.8110	25.8

Table 3.8: Given $(\alpha, \beta_1, \beta_2) = (0.05, 0.20, 0.20)$. $\delta_1 = 0.15, \delta_2 = 0.20$. Reject H_0 if $(X_1 > s_1$ or $Y_1 > t_1)$ and $(X > s$ or $Y > t)$.

$p_1^{(0)}$	$p_1^{(A)}$	$p_2^{(0)}$	$p_2^{(A)}$	(n, n_1, s_1, t_1, s, t)	$\max_{\pi_{11}} G_0(Q, \pi_{11})$	$\min_{\pi_{11}} G_1(Q, \pi_{11})$	$\min_{\pi_{11}} G_2(Q, \pi_{11})$	$\min_{\pi_{11}} G_3(Q, \pi_{11})$	$\max_{\pi_{11}} E(N\|H_0, Q, \pi_{11})$
0.05	0.20	0.05	0.25	(37, 16, 1, 2, 4, 4)	0.0458	0.8055	0.8034	0.8776	20.8
		0.1	0.3	(40, 15, 1, 2, 4, 8)	0.0474	0.8026	0.8141	0.8882	23.2
		0.2	0.4	(44, 22, 2, 6, 5, 13)	0.0474	0.8066	0.8055	0.9078	26.8
		0.3	0.5	(50, 22, 2, 8, 6, 20)	0.0451	0.8017	0.8141	0.9238	29.6
		0.4	0.6	(52, 23, 2, 11, 6, 26)	0.0497	0.8322	0.8058	0.9394	30.5
		0.5	0.7	(53, 21, 2, 12, 6, 32)	0.0467	0.8021	0.8188	0.9413	29.6
		0.6	0.8	(43, 22, 2, 15, 5, 31)	0.0435	0.8050	0.8081	0.9472	27.2
		0.7	0.9	(44, 16, 1, 13, 5, 35)	0.0476	0.8118	0.8119	0.9555	24.1
0.1	0.25	0.1	0.3	(48, 23, 3, 4, 8, 10)	0.0419	0.8069	0.8164	0.8644	29.5
		0.2	0.4	(50, 25, 3, 7, 9, 15)	0.0450	0.8003	0.8157	0.9030	33.2
		0.3	0.5	(64, 26, 4, 10, 10, 26)	0.0488	0.8008	0.8055	0.9108	34.7
		0.4	0.6	(59, 23, 3, 11, 10, 30)	0.0495	0.8154	0.8032	0.9273	35.2
		0.5	0.7	(59, 27, 4, 16, 10, 36)	0.0460	0.8044	0.8157	0.9363	35.0
		0.6	0.8	(53, 23, 3, 16, 9, 38)	0.0484	0.8089	0.8160	0.9459	32.3
		0.7	0.9	(47, 19, 2, 15, 8, 39)	0.0466	0.8099	0.8553	0.9628	30.8
0.2	0.35	0.2	0.4	(64, 31, 7, 9, 18, 20)	0.0453	0.8057	0.8185	0.8704	42.2
		0.3	0.5	(69, 33, 8, 13, 19, 29)	0.0472	0.8106	0.8040	0.8915	43.1
		0.4	0.6	(72, 30, 7, 15, 20, 37)	0.0465	0.8098	0.8005	0.9117	43.9
		0.5	0.7	(69, 33, 8, 20, 19, 43)	0.0486	0.8107	0.8052	0.9250	43.1

Table 3.8 (Cont.). Given $(\alpha, \beta_1, \beta_2) = (0.05, 0.20, 0.20)$. $\delta_1 = 0.15, \delta_2 = 0.20$. Reject H_0 if $(X_1 > s_1$ or $Y_1 > t_1)$ and $(X > s$ or $Y > t)$.

		0.6	0.8	(64, 31, 7, 22, 18, 47)	0.0438	0.8059	0.8076	0.9367	42.2
		0.7	0.9	(66, 23, 5, 19, 18, 55)	0.0476	0.8042	0.8195	0.9485	38.5
0.3	0.45	0.3	0.5	(73, 36, 12, 14, 28, 31)	0.0493	0.8036	0.8182	0.8610	49.0
		0.4	0.6	(76, 38, 13, 19, 29, 40)	0.0488	0.8066	0.8117	0.8922	49.4
		0.5	0.7	(73, 41, 14, 25, 28, 46)	0.0496	0.8086	0.8153	0.9146	50.0
		0.6	0.8	(77, 35, 12, 25, 29, 57)	0.0487	0.8074	0.8051	0.9265	46.8
		0.7	0.9	(68, 36, 12, 30, 26, 58)	0.0496	0.8018	0.8078	0.9401	45.3
0.4	0.55	0.4	0.6	(76, 41, 18, 20, 37, 41)	0.0484	0.8039	0.8009	0.8523	53.2
		0.5	0.7	(73, 44, 18, 26, 36, 47)	0.0453	0.8000	0.8055	0.8875	57.6
		0.6	0.8	(79, 40, 18, 29, 38, 58)	0.0493	0.8037	0.8106	0.9160	49.4
		0.7	0.9	(77, 38, 17, 32, 37, 65)	0.0492	0.8010	0.8140	0.9360	47.4
0.5	0.65	0.5	0.7	(73, 44, 24, 26, 43, 47)	0.0500	0.8063	0.8043	0.8571	53.0
		0.6	0.8	(77, 35, 19, 25, 45, 57)	0.0482	0.8002	0.8005	0.8939	48.0
		0.7	0.9	(79, 30, 16, 25, 46, 67)	0.0470	0.8025	0.8020	0.9241	45.7
0.6	0.75	0.6	0.8	(73, 31, 20, 22, 50, 53)	0.0494	0.8051	0.8180	0.8720	44.4
		0.7	0.9	(67, 27, 17, 22, 46, 57)	0.0491	0.8003	0.8105	0.9101	41.4
0.7	0.85	0.7	0.9	(56, 29, 22, 24, 44, 47)	0.0492	0.8105	0.8062	0.8794	34.9

Table 3.9: Given $(\alpha, \beta_1, \beta_2) = (0.05, 0.20, 0.20)$. $\delta_1 = 0.2, \delta_2 = 0.15$. Reject H_0 if $(X_1 > s_1$ or $Y_1 > t_1)$ and $(X > s$ or $Y > t)$.

$p_1^{(0)}$	$p_1^{(A)}$	$p_2^{(0)}$	$p_2^{(A)}$	(n, n_1, s_1, t_1, s, t)	$\max_{\pi_{11}} G_0(Q, \pi_{11})$	$\min_{\pi_{11}} G_1(Q, \pi_{11})$	$\min_{\pi_{11}} G_2(Q, \pi_{11})$	$\min_{\pi_{11}} G_3(Q, \pi_{11})$	$\max_{\pi_{11}} E(N\|H_0, Q, \pi_{11})$
0.05	0.25	0.05	0.20	(37, 16, 2, 1, 4, 4)	0.0458	0.8034	0.8055	0.8776	20.8
		0.1	0.25	(40, 22, 3, 2, 6, 7)	0.0424	0.8117	0.8032	0.8775	29.2
		0.2	0.35	(58, 27, 4, 6, 9, 16)	0.0500	0.8231	0.8007	0.8968	36.3
		0.3	0.45	(65, 48, 7, 17, 12, 25)	0.0495	0.8497	0.8010	0.9179	50.9
		0.4	0.55	(77, 38, 6, 17, 14, 37)	0.0484	0.8256	0.8010	0.9255	46.6
		0.5	0.65	(79, 30, 5, 16, 12, 46)	0.0463	0.8054	0.8026	0.9376	44.7
		0.6	0.75	(65, 32, 5, 20, 11, 45)	0.0442	0.8291	0.8013	0.9508	42.9
		0.7	0.85	(52, 23, 3, 17, 9, 41)	0.0485	0.8072	0.8002	0.9473	31.4
		0.8	0.95	(36, 12, 1, 10, 5, 32)	0.0492	0.8222	0.8330	0.9639	21.4
0.1	0.3	0.1	0.25	(48, 23, 4, 3, 10, 8)	0.0419	0.8164	0.8069	0.8644	29.5
		0.2	0.35	(62, 30, 6, 7, 13, 17)	0.0487	0.8152	0.8083	0.8691	38.3
		0.3	0.45	(68, 36, 7, 12, 16, 26)	0.0490	0.8042	0.8018	0.8898	45.0
		0.4	0.55	(84, 30, 6, 13, 19, 40)	0.0491	0.8032	0.8085	0.9130	46.7
		0.5	0.65	(69, 39, 9, 19, 13, 41)	0.0498	0.8067	0.8000	0.9693	54.0
		0.6	0.75	(67, 27, 5, 17, 15, 46)	0.0485	0.8164	0.8012	0.9384	41.0
		0.7	0.85	(63, 25, 5, 19, 12, 49)	0.0496	0.8095	0.8076	0.9436	33.7
		0.8	0.95	(45, 13, 2, 11, 8, 40)	0.0498	0.8024	0.8391	0.9572	24.8
0.2	0.4	0.2	0.35	(64, 31, 9, 7, 20, 18)	0.0453	0.8185	0.8057	0.8704	42.2
		0.3	0.45	(76, 33, 10, 11, 24, 29)	0.0450	0.8006	0.8016	0.8600	46.3

Table 3.9 (Cont.). Given $(\alpha, \beta_1, \beta_2) = (0.05, 0.20, 0.20)$. $\delta_1 = 0.2, \delta_2 = 0.15$. Reject H_0 if $(X_1 > s_1$ or $Y_1 > t_1)$ and $(X > s$ or $Y > t)$.

		0.4	0.55	(79, 38, 11, 17, 26, 38)	0.0483	0.8258	0.8014	0.8965	49.5
		0.5	0.65	(77, 35, 10, 19, 25, 45)	0.0483	0.8306	0.8008	0.9177	48.4
		0.6	0.75	(71, 30, 9,19, 22, 49)	0.0460	0.8027	0.8020	0.9302	44.3
		0.7	0.85	(64, 25, 7, 19, 20, 50)	0.0488	0.8073	0.8059	0.9379	36.7
		0.8	0.95	(44, 22, 6, 19, 13, 40)	0.0461	0.8082	0.8060	0.9554	28.4
0.3	0.5	0.3	0.45	(73, 36, 14, 12, 31, 28)	0.0493	0.8182	0.8036	0.8610	49.0
		0.4	0.55	(73, 44, 18, 18, 31, 36)	0.0476	0.8094	0.8000	0.8807	56.5
		0.5	0.65	(79, 38, 15, 21, 34, 46)	0.0496	0.8061	0.8016	0.8890	49.9
		0.6	0.75	(73, 38, 15, 25, 31, 50)	0.0493	0.8156	0.8106	0.9167	47.1
		0.7	0.85	(62, 29, 11, 22, 26, 49)	0.0468	0.8059	0.8072	0.9305	39.6
		0.8	0.95	(50, 26, 10, 23, 20, 45)	0.0476	0.8025	0.8288	0.9508	31.0
0.4	0.6	0.4	0.55	(76, 41, 20, 18, 41, 37)	0.0484	0.8009	0.8039	0.8523	53.2
		0.5	0.65	(76, 38, 19, 20, 40, 45)	0.0469	0.8100	0.8028	0.8676	52.8
		0.6	0.75	(76, 31, 15, 20, 40, 52)	0.0495	0.8100	0.8103	0.8975	47.7
		0.7	0.85	(71, 30, 15, 23, 36, 56)	0.0431	0.8005	0.8040	0.9175	40.7
		0.8	0.95	(49, 27, 13, 24, 25, 44)	0.0495	0.8024	0.8253	0.9453	31.8
0.5	0.7	0.5	0.65	(73, 44, 26, 24, 47, 43)	0.0500	0.8043	0.8063	0.8571	53.0
		0.6	0.75	(73, 38, 23, 25, 46, 50)	0.0498	0.8095	0.8099	0.8630	47.1
		0.7	0.85	(70, 30, 18, 23, 43, 55)	0.0443	0.8108	0.8089	0.9050	40.4
		0.8	0.95	(51, 24, 14, 21, 31, 46)	0.0469	0.8062	0.8360	0.9433	31.0

Table 3.9 (Cont.). Given $(\alpha, \beta_1, \beta_2) = (0.05, 0.20, 0.20)$. $\delta_1 = 0.2, \delta_2 = 0.15$. Reject H_0 if $(X_1 > s_1$ or $Y_1 > t_1)$ and $(X > s$ or $Y > t)$.

0.6	0.8	0.6	0.75	(73, 31, 22, 20, 53, 50)	0.0494	0.8180	0.8051	0.8720	44.4
		0.7	0.85	(63, 26, 18, 19, 46, 50)	0.0462	0.8168	0.8073	0.8832	41.2
		0.8	0.95	(55, 15, 10, 13, 39, 49)	0.0469	0.8018	0.8054	0.9263	29.8
0.7	0.9	0.7	0.85	(56, 29, 24, 22, 47, 44)	0.0492	0.8062	0.8105	0.8794	34.9
		0.8	0.95	(45, 14, 11, 12, 37, 40)	0.0433	0.8071	0.8167	0.8933	24.5

Some discussions:

1. In the study design stage, we may not have too much information about the nuisance correlation parameter π_{11} and we want to be conservative. So we searched thoroughly in the defined range of π_{11}. The searching results show that π_{11} may assume different values to achieve the maximized type I error rate, minimized powers, the minimum of the maximized value among all possible expected sample sizes under the null hypothesis.

2. Due to the time cost of thorough searching in the defined range of π_{11}, we only did the computation-intensive calculations of real maximized type I and II error rates among the top 5% of sorted feasible solutions from independence assumption, so the resulting designs we got may not be global optimal, but close to as shown in Bryant and Day (1995).

Chapter 4

Statistical Inference For Proposed Two-stage Design For Phase II Cancer Clinical Trials with Two Endpoints

In this chapter, we will focus on methodology for making statistical inference based on the proposed two-stage design for phase II cancer clinical trials with two endpoints.

4.1 Review of the MLE for π and the total Fisher information in a single-endpoint two-stage design

When a clinical trial with a single-endpoint two-stage design completes and we get all the data at hand, that is, given (n, n_1, s_1, s), we have observed the individual data $(M, X_{11}, X_{21}, \cdots, X_{n_1 1}, X_{(n_1+1)1}, \cdots, X_{n1})$ using the notations in Chapter 3. The individual data can be summarized by (X_1, X, M), where

- X_1: total number of responses at the end of stage 1;
- X: total number of responses at the end of stage 2;
- M: which stage the trial stops at.

Consider the likelihood function in this single-endpoint two-stage design:

1. If $X_1 \leq s_1$, the trial stops early at the end of stage $M = 1$, the likelihood function is:

$$\begin{aligned} L(\pi|X_1, M=1) &= \Pr(X_{11} = x_{11}, X_{21} = x_{21}, ..., X_{n_1 1} = x_{n_1 1}) \\ &= \prod_{i=1}^{n_1} \pi^{x_{i1}} (1-\pi)^{1-x_{i1}} = \pi^{\sum_{i=1}^{n_1} x_{i1}} (1-\pi)^{n_1 - \sum_{i=1}^{n_1} x_{i1}} \\ &= \pi^{x_1} (1-\pi)^{n_1 - x_1}. \end{aligned}$$

2. If $X_1 > s_1$, enroll additional n_2 patients into the trial after stage 1, and record the number of responses during stage 2 as X_2, thus $X_2 = X - X_1$. The likelihood function is:

$$L(\pi|X_1, X, M = 2) = \pi^{x_1+x_2}(1-\pi)^{n_1+n_2-x_1-x_2} = \pi^x(1-\pi)^{n-x}.$$

Summarizing the above 2 scenarios, the likelihood function for a single-endpoint two-stage design can be written as:

$$L(\pi|X_1, X, M) = \pi^{x_1+x_2(M-1)}(1-\pi)^{(n_1-x_1)+(n_2-x_2)(M-1)},$$

and the log-likelihood

$$l(\pi) = \log L(\pi) = [x_1 + x_2(M-1)]\log\pi + [(n_1 - x_1) + (n_2 - x_2)(M-1)]\log(1-\pi).$$

We have

$$l'(\pi) = \frac{\partial l(\pi)}{\partial \pi} = \frac{x_1 + x_2(M-1)}{\pi} - \frac{(n_1 - x_1) + (n_2 - x_2)(M-1)}{1-\pi}$$

Let the above item $l'(\pi)$ be equal to 0, then we can get the maximum likelihood estimator (MLE) for π in a single-endpoint two-stage design:

$$\begin{aligned}\widehat{\pi}_{MLE} &= \frac{x_1 + x_2(M-1)}{n_1 + n_2(M-1)} \\ &= \begin{cases} \frac{x_1}{n_1}, & \text{if } M = 1; \\ \frac{x_1+x_2}{n_1+n_2}, & \text{if } M = 2. \end{cases}\end{aligned}$$

The MLE for π in a single-endpoint two-stage design is sample proportion in essence. Since

$$\frac{\partial^2 \log L(\pi)}{\partial \pi^2} = -\frac{x_1 + x_2(M-1)}{\pi^2} - \frac{(n_1 - x_1) + (n_2 - x_2)(M-1)}{(1-\pi)^2},$$

the total Fisher information is:

$$I(\pi) = E\left[-\frac{\partial^2 \log L(\pi)}{\partial \pi^2}\right]$$

$$= E\left[\frac{X_1 + X_2(M-1)}{\pi^2} + \frac{(n_1 - X_1) + (n_2 - X_2)(M-1)}{(1-\pi)^2}\right].$$

Given that M is a random variable dependent on X_1 ($M = 1$ if $X_1 \leq s_1$, $M = 2$ otherwise), M is a function of X_1. Since X_1 and X_2 are independent, M and X_2 are independent. Since

$$\begin{aligned} E(M) &= 1 \times \Pr(X_1 \leq s_1) + 2 \times [1 - \Pr(X_1 \leq s_1)] \\ &= 2 - \Pr(X_1 \leq s_1) \\ &= 2 - \sum_{k=0}^{s_1} \left(\begin{array}{c} n_1 \\ k \end{array} \right) \pi^k (1-\pi)^{n_1 - k}, \end{aligned}$$

the total Fisher information for a single-endpoint two-stage design is:

$$\begin{aligned} I(\pi) &= E\left[-\frac{\partial^2 \log L(\pi)}{\partial \pi^2}\right] \\ &= E\left[\frac{X_1 + X_2(M-1)}{\pi^2} + \frac{(n_1 - X_1) + (n_2 - X_2)(M-1)}{(1-\pi)^2}\right] \\ &= \frac{n_1\pi + n_2\pi\left[1 - \sum_{k=0}^{s_1}\binom{n_1}{k}\pi^k(1-\pi)^{n_1-k}\right]}{\pi^2} \\ &\quad + \frac{n_1 - n_1\pi + n_2(1-\pi)\left[1 - \sum_{k=0}^{s_1}\binom{n_1}{k}\pi^k(1-\pi)^{n_1-k}\right]}{(1-\pi)^2} \\ &= \frac{n_1 + n_2\left[1 - \sum_{k=0}^{s_1}\binom{n_1}{k}\pi^k(1-\pi)^{n_1-k}\right]}{(1-\pi)\pi}. \end{aligned}$$

4.2 Maximum Likelihood Estimators for (π_1, π_2) in a two-endpoint two-stage design

We will use the following notations for a two-endpoint two-stage design:

- n_1, n_2: number of patients enrolled in stage 1, stage 2, respectively; these two quantities are constants by design;
- $n_{11}^{(1)}$, $n_{11}^{(2)}$: number of patients with (endpoint1 response = Yes, endpoint 2 response = Yes) = $(1,1)$ response pattern at stage 1 and stage 2, respectively;
  note that we have $n_{11}^{(1)} \sim Bin(n_1, \pi_{11})$, $n_{11}^{(2)} \sim Bin(n_2, \pi_{11})$, with π_{11} defined in Table 3.4;

- X_1, X_2: total number of respondents for endpoint 1 at stage 1 and stage 2, respectively; $X_1 \sim Bin(n_1, \pi_1)$, $X_2 \sim Bin(n_2, \pi_1)$;

- Y_1, Y_2: total number of respondents for endpoint 2 at stage 1 and stage 2, respectively, and $Y_1 \sim Bin(n_1, \pi_2)$, $Y_2 \sim Bin(n_2, \pi_2)$;

- M: a random variable describing at which stage the clinical trial stops,

$$M = \begin{cases} 1 & \text{if } X_1 \leq s_1 \text{ and } Y_1 \leq t_1 \\ 2 & \text{if } X_1 > s_1 \text{ or } Y_1 > t_1 \end{cases}$$

We have

$$\begin{aligned} \Pr(M = 1) &= \Pr(X_1 \leq s_1 \text{ and } Y_1 \leq t_1) \\ &= \sum_{x=0}^{s_1} \sum_{y=0}^{t_1} p(x, y; n_1, \pi_1, \pi_2, \pi_{11}) = PET; \\ E(M) &= 1 \times \Pr(M = 1) + 2 \times [1 - \Pr(M = 1)] = 2 - \Pr(M = 1) \end{aligned}$$

Consider the likelihood function for a cancer clinical trial with the two-endpoint two-stage design:

1. If $X_1 \leq s_1$ and $Y_1 \leq t_1$, the trial stops early at the end of stage 1. By using the individual data, the likelihood function can be written as:

$$\begin{aligned} & L(\pi_1, \pi_2, \pi_{11} | n_1, n_{11}^{(1)}, x_1, y_1, M = 1) \\ &= \Pr[(X_{11}, X_{12}) = (x_{11}, x_{12}), ..., (X_{n_1 1}, X_{n_1 2}) = (x_{n_1 1}, x_{n_1 2})] \\ &= \prod_{i=1}^{n_1} \pi_{11}^{x_{i1}x_{i2}} (\pi_1 - \pi_{11})^{x_{i1}(1-x_{i2})} (\pi_2 - \pi_{11})^{(1-x_{i1})x_{i2}} (1 - \pi_1 - \pi_2 + \pi_{11})^{(1-x_{i1})(1-x_{i2})} \\ &= \pi_{11}^{\sum_{i=1}^{n_1} x_{i1}x_{i2}} (\pi_1 - \pi_{11})^{\sum_{i=1}^{n_1} x_{i1}(1-x_{i2})} (\pi_2 - \pi_{11})^{\sum_{i=1}^{n_1} (1-x_{i1})x_{i2}} (1 - \pi_1 - \pi_2 + \pi_{11})^{\sum_{i=1}^{n_1} (1-x_{i1})(1-x_{i2})} \\ &= \pi_{11}^{n_{11}^{(1)}} (\pi_1 - \pi_{11})^{x_1 - n_{11}^{(1)}} (\pi_2 - \pi_{11})^{y_1 - n_{11}^{(1)}} (1 - \pi_1 - \pi_2 + \pi_{11})^{n_1 - x_1 - y_1 + n_{11}^{(1)}} \end{aligned}$$

2. If $X_1 > s_1$ or $Y_1 > t_1$, then continue to the 2nd stage and enroll additional n_2 patients.

The likelihood function is

$$\begin{aligned}
&L(\pi_1, \pi_2, \pi_{11} | n_1, n_2, n_{11}^{(1)}, n_{11}^{(2)}, x_1, x_2, y_1, y_2, M = 2) \\
&= \pi_{11}^{n_{11}^{(1)} + n_{11}^{(2)}} (\pi_1 - \pi_{11})^{x_1 - n_{11}^{(1)} + x_2 - n_{11}^{(2)}} (\pi_2 - \pi_{11})^{y_1 - n_{11}^{(1)} + y_2 - n_{11}^{(2)}} \\
&\quad \times (1 - \pi_1 - \pi_2 + \pi_{11})^{n_1 - x_1 - y_1 + n_{11}^{(1)} + n_2 - x_2 - y_2 + n_{11}^{(2)}}
\end{aligned}$$

Summarizing the above 2 scenarios, the likelihood function for a clinical trial with a two-endpoint two-stage design is

$$\begin{aligned}
&L(\pi_1, \pi_2, \pi_{11} | n_1, n_2, n_{11}^{(1)}, n_{11}^{(2)}, x_1, x_2, y_1, y_2, M) \\
&= \pi_{11}^{n_{11}^{(1)} + n_{11}^{(2)}(M-1)} (\pi_1 - \pi_{11})^{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)} (\pi_2 - \pi_{11})^{(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)} \\
&\quad \times (1 - \pi_1 - \pi_2 + \pi_{11})^{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)},
\end{aligned}$$

and the log-likelihood is

$$\begin{aligned}
&\log L(\pi_1, \pi_2, \pi_{11}) \\
&= [n_{11}^{(1)} + n_{11}^{(2)}(M-1)] \log \pi_{11} + [(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)] \log(\pi_1 - \pi_{11}) \\
&\quad + [(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)] \log(\pi_2 - \pi_{11}) \\
&\quad + [(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)] \log(1 - \pi_1 - \pi_2 + \pi_{11}).
\end{aligned}$$

The first partial derivatives of the log-likelihood function are

$$\begin{aligned}
\frac{\partial \log L(\pi_1, \pi_2, \pi_{11})}{\partial \pi_1} &= \frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{\pi_1 - \pi_{11}} \\
&\quad - \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{1 - \pi_1 - \pi_2 + \pi_{11}}, \\
\frac{\partial \log L(\pi_1, \pi_2, \pi_{11})}{\partial \pi_2} &= \frac{(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)}{\pi_2 - \pi_{11}} \\
&\quad - \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{1 - \pi_1 - \pi_2 + \pi_{11}}, \\
\frac{\partial \log L(\pi_1, \pi_2, \pi_{11})}{\partial \pi_{11}} &= \frac{n_{11}^{(1)} + n_{11}^{(2)}(M-1)}{\pi_{11}} - \frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{\pi_1 - \pi_{11}}
\end{aligned}$$

$$- \frac{(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)}{\pi_2 - \pi_{11}}$$
$$+ \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{1 - \pi_1 - \pi_2 + \pi_{11}}.$$

Let each of the above three equations be zero, then

$$\frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{\pi_1 - \pi_{11}} - \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{1 - \pi_1 - \pi_2 + \pi_{11}} = 0,$$
$$\frac{(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)}{\pi_2 - \pi_{11}} - \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{1 - \pi_1 - \pi_2 + \pi_{11}} = 0,$$
$$\frac{n_{11}^{(1)} + n_{11}^{(2)}(M-1)}{\pi_{11}} - \frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{\pi_1 - \pi_{11}} - \frac{(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)}{\pi_2 - \pi_{11}} + \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{1 - \pi_1 - \pi_2 + \pi_{11}} = 0.$$

This implies

$$\frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{\pi_1 - \pi_{11}} = \frac{(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)}{\pi_2 - \pi_{11}} = \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{1 - \pi_1 - \pi_2 + \pi_{11}}$$
$$\frac{n_{11}^{(1)} + n_{11}^{(2)}(M-1)}{\pi_{11}} = \frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{\pi_1 - \pi_{11}}.$$

Hence

$$\frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{\pi_1 - \pi_{11}} = \frac{(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)}{\pi_2 - \pi_{11}}$$
$$\frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{\pi_1 - \pi_{11}} = \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{1 - \pi_1 - \pi_2 + \pi_{11}}$$

$$\frac{\pi_1 - \pi_{11}}{\pi_{11}} = \frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{n_{11}^{(1)} + n_{11}^{(2)}(M-1)}$$
$$= \frac{x_1 + x_2(M-1) - [n_{11}^{(1)} + n_{11}^{(2)}(M-1)]}{n_{11}^{(1)} + n_{11}^{(2)}(M-1)}.$$

This implies

$$\pi_2 = \frac{(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)}{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}(\pi_1 - \pi_{11}) + \pi_{11},$$
$$\pi_1 = \frac{x_1 + x_2(M-1)}{n_{11}^{(1)} + n_{11}^{(2)}(M-1)}\pi_{11},$$

and

$$\frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{\pi_1 - \pi_{11}}$$
$$= \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{1 - \pi_1 - \pi_2 + \pi_{11}}, \qquad (4.1)$$

thus we can express π_1 and π_2 as functions of π_{11}

$$\pi_1 = \frac{x_1 + x_2(M-1)}{n_{11}^{(1)} + n_{11}^{(2)}(M-1)}\pi_{11},$$
$$\pi_2 = \frac{y_1 + y_2(M-1)}{n_{11}^{(1)} + n_{11}^{(2)}(M-1)}\pi_{11}.$$

Plugging these two equations for π_1, π_2 as functions of π_{11} back to Equation (4.1) and keeping only π_{11}, we get

$$\widehat{\pi}_{11,MLE} = \frac{n_{11}^{(1)} + n_{11}^{(2)}(M-1)}{n_1 + n_2(M-1)}$$
$$= \begin{cases} \frac{n_{11}^{(1)}}{n_1}, & \text{if } M = 1, \\ \frac{n_{11}^{(1)} + n_{11}^{(2)}}{n_1 + n_2}, & \text{if } M = 2, \end{cases}$$
$$\widehat{\pi}_{1,MLE} = \frac{x_1 + x_2(M-1)}{n_{11}^{(1)} + n_{11}^{(2)}(M-1)}\pi_{11} = \frac{x_1 + x_2(M-1)}{n_1 + n_2(M-1)}$$
$$= \begin{cases} \frac{x_1}{n_1}, & \text{if } M = 1, \\ \frac{x_1 + x_2}{n_1 + n_2}, & \text{if } M = 2. \end{cases}$$

$$\widehat{\pi}_{2,MLE} = \frac{y_1 + y_2\ (M-1)}{n_{11}^{(1)} + n_{11}^{(2)}(M-1)}\pi_{11} = \frac{y_1 + y_2(M-1)}{n_1 + n_2\ (M-1)}$$
$$= \begin{cases} \frac{y_1}{n_1}, & \text{if } M = 1, \\ \frac{y_1+y_2}{n_1+n_2}, & \text{if } M = 2. \end{cases}$$

4.3 Estimation of Confidence region for (π_1, π_2) in a two-endpoint two-stage design

Based on Chapter 9 of the textbook "Statistical Inference" by George Casella, there is a strong correspondence between acceptance regions of hypothesis tests and confidence sets/regions. So we can obtain a confidence region by inverting a hypothesis test. In this section, we will propose three types of confidence regions based on the inverses of three types of likelihood based test statistics − Wald, Score and Likelihood Ratio statistics.

4.3.1 Wald-type confidence region for (π_1, π_2) in a two-endpoint two-stage design

The first type of confidence region we considered is created by inverting a Wald statistic. Under regularity conditions, $\widehat{\theta}_{MLE}$ is AN(θ, $\{\mathrm{I}_T(\theta)\}^{-1}$), i.e.,

$$\begin{pmatrix} \widehat{\pi}_{1,MLE} \\ \widehat{\pi}_{2,MLE} \\ \widehat{\pi}_{11,MLE} \end{pmatrix} \sim AN\left(\begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_{11} \end{pmatrix}, \sum\right)$$

where $\sum = \{I_T(\theta)\}^{-1}$ is a 3×3 variance-covariance matrix.

Throughout this chapter, we will partition the parameter vector θ as follows:

$$\underset{3\times1}{\theta} = \begin{pmatrix} \underset{2\times1}{\theta_1} \\ \underset{1\times1}{\theta_2} \end{pmatrix}$$

where $\theta_2 = \pi_{11}$ is treated as a nuisance parameter, and our parameters of interest is

$$\theta_1 = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}.$$

With this partitioning $H_0 : \theta_1^T = (\pi_1, \pi_2)^T = (\pi_{1,0}, \pi_{2,0})^T$, the Wald statistic is:

$$T_w = (\widehat{\theta}_1 - \theta_{1,0})^T \{[I_T(\widehat{\theta})]^{-1}_{(1,1)}\}^{-1} (\widehat{\theta}_1 - \theta_{1,0}).$$

In order to derive the Wald statistic, we first need to find the Fisher Information matrix for the design.

Let $\theta^T = [\pi_1, \pi_2, \pi_{11}]$, the second derivatives of the log-likelihood function are:

$$\frac{\partial^2 \log L(\pi_1, \pi_2, \pi_{11})}{\partial\theta\partial\theta^T} = \begin{bmatrix} l''_{11} & l''_{12} & l''_{13} \\ l''_{21} & l''_{22} & l''_{23} \\ l''_{31} & l''_{32} & l''_{33} \end{bmatrix}$$

where

$$\begin{aligned}
l''_{11} &= -\frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{(\pi_1 - \pi_{11})^2} \\
&\quad - \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{(1 - \pi_1 - \pi_2 + \pi_{11})^2} \\
l''_{12} = l''_{21} &= -\frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{(1 - \pi_1 - \pi_2 + \pi_{11})^2} \\
l''_{13} = l''_{31} &= \frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{(\pi_1 - \pi_{11})^2} \\
&\quad + \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{(1 - \pi_1 - \pi_2 + \pi_{11})^2} \\
l''_{22} &= -\frac{(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)}{(\pi_2 - \pi_{11})^2} \\
&\quad - \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{(1 - \pi_1 - \pi_2 + \pi_{11})^2} \\
l''_{23} = l''_{32} &= \frac{(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)}{(\pi_2 - \pi_{11})^2} \\
&\quad + \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{(1 - \pi_1 - \pi_2 + \pi_{11})^2} \\
l''_{33} &= -\frac{n_{11}^{(1)} + n_{11}^{(2)}(M-1)}{\pi_{11}^2} - \frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{(\pi_1 - \pi_{11})^2} \\
&\quad - \frac{(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)}{(\pi_2 - \pi_{11})^2}
\end{aligned}$$

$$- \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{(1 - \pi_1 - \pi_2 + \pi_{11})^2}.$$

Fisher information matrix is

$$\begin{aligned} I(\theta) &= E\left[-\frac{\partial^2 \log L(\pi_1, \pi_2, \pi_{11})}{\partial\theta\partial\theta^T}\right] \\ &= E\begin{bmatrix} -l''_{11} & -l''_{12} & -l''_{13} \\ -l''_{21} & -l''_{22} & -l''_{23} \\ -l''_{31} & -l''_{32} & -l''_{33} \end{bmatrix} = \begin{bmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{bmatrix}. \end{aligned}$$

Let

$$\eta = E(M-1) = \Pr(M=2) = 1 - \Pr(M=1) = 1 - PET,$$

then

$$E(N) = n_1 PET + (n_1 + n_2)(1 - PET) = n_1(1-\eta) + (n_1+n_2)\eta = n_1 + n_2\eta,$$

and we have

$$\begin{aligned} I_{11} &= E\left[\frac{(X_1 - N_{11}^{(1)}) + (X_2 - N_{11}^{(2)})(M-1)}{(\pi_1 - \pi_{11})^2}\right. \\ &\qquad \left. + \frac{(n_1 - X_1 - Y_1 + N_{11}^{(1)}) + (n_2 - X_2 - Y_2 + N_{11}^{(2)})(M-1)}{(1 - \pi_1 - \pi_2 + \pi_{11})^2}\right] \\ &= \frac{E(X_1 - N_{11}^{(1)}) + E(X_2 - N_{11}^{(2)})E(M-1)}{(\pi_1 - \pi_{11})^2} \\ &\qquad + \frac{E(n_1 - X_1 - Y_1 + N_{11}^{(1)}) + E(n_2 - X_2 - Y_2 + N_{11}^{(2)})E(M-1)}{(1 - \pi_1 - \pi_2 + \pi_{11})^2} \\ &= \frac{(n_1\pi_1 - n_1\pi_{11}) + (n_2\pi_1 - n_2\pi_{11})\eta}{(\pi_1 - \pi_{11})^2} \\ &\qquad + \frac{(n_1 - n_1\pi_1 - n_1\pi_2 + n_1\pi_{11}) + (n_2 - n_2\pi_1 - n_2\pi_2 + n_2\pi_{11})\eta}{(1 - \pi_1 - \pi_2 + \pi_{11})^2} \\ &= \frac{(n_1 + n_2\eta)(1 - \pi_2)}{(\pi_1 - \pi_{11})(1 - \pi_1 - \pi_2 + \pi_{11})} \\ I_{12} &= I_{21} = E\left[\frac{(n_1 - X_1 - Y_1 + N_{11}^{(1)}) + (n_2 - X_2 - Y_2 + N_{11}^{(2)})(M-1)}{(1 - \pi_1 - \pi_2 + \pi_{11})^2}\right] \\ &= \frac{(n_1 - n_1\pi_1 - n_1\pi_2 + n_1\pi_{11}) + (n_2 - n_2\pi_1 - n_2\pi_2 + n_2\pi_{11})\eta}{(1 - \pi_1 - \pi_2 + \pi_{11})^2} \end{aligned}$$

$$= \frac{n_1 + n_2\eta}{1 - \pi_1 - \pi_2 + \pi_{11}}$$

$$I_{13} = I_{31} = E\left[-\frac{(X_1 - N_{11}^{(1)}) + (X_2 - N_{11}^{(2)})(M-1)}{(\pi_1 - \pi_{11})^2} - \frac{(n_1 - X_1 - Y_1 + N_{11}^{(1)}) + (n_2 - X_2 - Y_2 + N_{11}^{(2)})(M-1)}{(1 - \pi_1 - \pi_2 + \pi_{11})^2}\right]$$

$$= -\frac{(n_1 + n_2\eta)(1 - \pi_2)}{(\pi_1 - \pi_{11})(1 - \pi_1 - \pi_2 + \pi_{11})} = -I_{11}$$

$$I_{22} = E\left[\frac{(Y_1 - N_{11}^{(1)}) + (Y_2 - N_{11}^{(2)})(M-1)}{(\pi_2 - \pi_{11})^2} + \frac{(n_1 - X_1 - Y_1 + N_{11}^{(1)}) + (n_2 - X_2 - Y_2 + N_{11}^{(2)})(M-1)}{(1 - \pi_1 - \pi_2 + \pi_{11})^2}\right]$$

$$= \frac{(n_1\pi_2 - n_1\pi_{11}) + (n_2\pi_2 - n_2\pi_{11})\eta}{(\pi_2 - \pi_{11})^2} + \frac{(n_1 - n_1\pi_1 - n_1\pi_2 + n_1\pi_{11}) + (n_2 - n_2\pi_1 - n_2\pi_2 + n_2\pi_{11})\eta}{(1 - \pi_1 - \pi_2 + \pi_{11})^2}$$

$$= \frac{n_1 + n_2\eta}{\pi_2 - \pi_{11}} + \frac{n_1 + n_2\eta}{1 - \pi_1 - \pi_2 + \pi_{11}}$$

$$= \frac{(n_1 + n_2\eta)(1 - \pi_1)}{(\pi_2 - \pi_{11})(1 - \pi_1 - \pi_2 + \pi_{11})}$$

$$I_{23} = I_{32} = E\left[-\frac{(Y_1 - N_{11}^{(1)}) + (Y_2 - N_{11}^{(2)})(M-1)}{(\pi_2 - \pi_{11})^2} - \frac{(n_1 - X_1 - Y_1 + N_{11}^{(1)}) + (n_2 - X_2 - Y_2 + N_{11}^{(2)})(M-1)}{(1 - \pi_1 - \pi_2 + \pi_{11})^2}\right]$$

$$= -\frac{(n_1 + n_2\eta)(1 - \pi_1)}{(\pi_2 - \pi_{11})(1 - \pi_1 - \pi_2 + \pi_{11})}$$

$$= -I_{22}$$

$$I_{33} = E\left[\frac{N_{11}^{(1)} + N_{11}^{(2)}(M-1)}{\pi_{11}^2} + \frac{(X_1 - N_{11}^{(1)}) + (X_2 - N_{11}^{(2)})(M-1)}{(\pi_1 - \pi_{11})^2}\right] + E\left[\frac{(Y_1 - N_{11}^{(1)}) + (Y_2 - N_{11}^{(2)})(M-1)}{(\pi_2 - \pi_{11})^2} + \frac{(n_1 - X_1 - Y_1 + N_{11}^{(1)}) + (n_2 - X_2 - Y_2 + N_{11}^{(2)})(M-1)}{(1 - \pi_1 - \pi_2 + \pi_{11})^2}\right]$$

$$= \frac{n_1\pi_{11} + n_2\pi_{11}\eta}{\pi_{11}^2} + \frac{(n_1\pi_1 - n_1\pi_{11}) + (n_2\pi_1 - n_2\pi_{11})\eta}{(\pi_1 - \pi_{11})^2} + \frac{(n_1\pi_2 - n_1\pi_{11}) + (n_2\pi_2 - n_2\pi_{11})\eta}{(\pi_2 - \pi_{11})^2} + \frac{(n_1 - n_1\pi_1 - n_1\pi_2 + n_1\pi_{11}) + (n_2 - n_2\pi_1 - n_2\pi_2 + n_2\pi_{11})\eta}{(1 - \pi_1 - \pi_2 + \pi_{11})^2}$$

$$= (n_1 + n_2\eta)(\frac{1}{\pi_{11}} + \frac{1}{\pi_1 - \pi_{11}} + \frac{1}{\pi_2 - \pi_{11}} + \frac{1}{1 - \pi_1 - \pi_2 + \pi_{11}})$$
$$= \frac{(n_1 + n_2\eta)(\pi_1\pi_2 + 2\pi_1\pi_2\pi_{11} - \pi_{11}^2 - \pi_1\pi_2^2 - \pi_1^2\pi_2)}{\pi_{11}(\pi_1 - \pi_{11})(\pi_2 - \pi_{11})(1 - \pi_1 - \pi_2 + \pi_{11})}$$

In summary, the Fisher information matrix is

$$I = (n_1 + n_2\eta)$$
$$\begin{bmatrix} \frac{1}{\pi_1-\pi_{11}} + \frac{1}{1-\pi_1-\pi_2+\pi_{11}} & \frac{1}{1-\pi_1-\pi_2+\pi_{11}} & -\frac{1}{\pi_1-\pi_{11}} - \frac{1}{1-\pi_1-\pi_2+\pi_{11}} \\ \frac{1}{1-\pi_1-\pi_2+\pi_{11}} & \frac{1}{\pi_2-\pi_{11}} + \frac{1}{1-\pi_1-\pi_2+\pi_{11}} & -\frac{1}{\pi_2-p_{11}} - \frac{1}{1-\pi_1-\pi_2+\pi_{11}} \\ -\frac{1}{\pi_1-\pi_{11}} - \frac{1}{1-\pi_1-\pi_2+\pi_{11}} & -\frac{1}{\pi_2-\pi_{11}} - \frac{1}{1-\pi_1-\pi_2+\pi_{11}} & \frac{1}{\pi_{11}} + \frac{1}{\pi_1-\pi_{11}} + \frac{1}{\pi_2-\pi_{11}} + \frac{1}{1-\pi_1-\pi_2+\pi_{11}} \end{bmatrix}$$
$$= (n_1 + n_2\eta) \begin{bmatrix} a+b & b & -a-b \\ b & c+b & -c-b \\ -a-b & -c-b & d+a+c+b \end{bmatrix}$$

where

$$a = \frac{1}{\pi_1 - \pi_{11}},$$
$$b = \frac{1}{1 - \pi_1 - \pi_2 + \pi_{11}},$$
$$c = \frac{1}{\pi_2 - \pi_{11}},$$
$$d = \frac{1}{\pi_{11}}.$$

We have

$$I^{-1} = \frac{1}{n_1 + n_2\eta} \begin{bmatrix} \frac{ab+ac+bd+cd}{abc+abd+acd+bcd} & \frac{ac-bd}{abc+abd+acd+bcd} & \frac{ab+ac}{abc+abd+acd+bcd} \\ \frac{ac-bd}{abc+abd+acd+bcd} & \frac{ac+ad+bc+bd}{abc+abd+acd+bcd} & \frac{ac+bc}{abc+abd+acd+bcd} \\ \frac{ab+ac}{abc+abd+acd+bcd} & \frac{ac+bc}{abc+abd+acd+bcd} & \frac{ab+ac+bc}{abc+abd+acd+bcd} \end{bmatrix}.$$

Therefore

$$I^{-1}_{(1,1)} = \frac{1}{n_1 + n_2\eta} \frac{ab + ac + bd + cd}{abc + abd + acd + bcd}$$
$$= \frac{1}{n_1 + n_2\eta}$$
$$\times \frac{\frac{1}{\pi_1-\pi_{11}}(\frac{1}{1-\pi_1-\pi_2+\pi_{11}} + \frac{1}{\pi_2-\pi_{11}}) + \frac{1}{\pi_{11}}(\frac{1}{1-\pi_1-\pi_2+\pi_{11}} + \frac{1}{\pi_2-\pi_{11}})}{\frac{1}{\pi_1-\pi_{11}}\frac{1}{1-\pi_1-\pi_2+\pi_{11}}(\frac{1}{\pi_2-\pi_{11}} + \frac{1}{\pi_{11}}) + \frac{1}{\pi_2-\pi_{11}}\frac{1}{\pi_{11}}(\frac{1}{\pi_1-\pi_{11}} + \frac{1}{1-\pi_1-\pi_2+\pi_{11}})}$$
$$= \frac{1}{n_1 + n_2\eta}$$
$$\times \frac{(\frac{1}{1-\pi_1-\pi_2+\pi_{11}} + \frac{1}{\pi_2-\pi_{11}})(\frac{1}{\pi_1-\pi_{11}} + \frac{1}{\pi_{11}})}{\frac{1}{\pi_{11}(\pi_1-\pi_{11})(\pi_2-\pi_{11})(1-\pi_1-\pi_2+\pi_{11})}(\pi_{11} + (\pi_2 - \pi_{11}) + (1 - \pi_1 - \pi_2 + \pi_{_{11}}) + (\pi_1 - \pi_{11}))}$$

$$
\begin{aligned}
&= \frac{1}{n_1+n_2\eta}\frac{\frac{(1-\pi_1)}{(\pi_2-\pi_{11})(1-\pi_1-\pi_2+\pi_{11})}\frac{\pi_1}{\pi_{11}(\pi_1-\pi_{11})}}{\frac{1}{\pi_{11}(\pi_1-\pi_{11})(\pi_2-\pi_{11})(1-\pi_1-\pi_2+\pi_{11})}} \\
&= \frac{\pi_1(1-\pi_1)}{n_1+n_2\eta} \\
I^{-1}_{(2,2)} &= \frac{1}{n_1+n_2\eta}\frac{ac+ad+bc+bd}{abc+abd+acd+bcd} \\
&= \frac{1}{n_1+n_2\eta}\frac{\frac{1}{\pi_1-\pi_{11}}(\frac{1}{\pi_2-\pi_{11}}+\frac{1}{\pi_{11}})+\frac{1}{1-\pi_1-\pi_2+\pi_{11}}(\frac{1}{\pi_2-\pi_{11}}+\frac{1}{\pi_{11}})}{\frac{1}{\pi_1-\pi_{11}}\frac{1}{1-\pi_1-\pi_2+\pi_{11}}(\frac{1}{\pi_2-\pi_{11}}+\frac{1}{\pi_{11}})+\frac{1}{\pi_2-\pi_{11}}\frac{1}{\pi_{11}}(\frac{1}{\pi_1-\pi_{11}}+\frac{1}{1-\pi_1-\pi_2+\pi_{11}})} \\
&= \frac{1}{n_1+n_2\eta}\frac{(\frac{1}{\pi_2-\pi_{11}}+\frac{1}{\pi_{11}})(\frac{1}{\pi_1-\pi_{11}}+\frac{1}{1-\pi_1-\pi_2+\pi_{11}})}{\frac{1}{\pi_1-\pi_{11}}\frac{1}{1-\pi_1-\pi_2+\pi_{11}}(\frac{1}{\pi_2-\pi_{11}}+\frac{1}{\pi_{11}})+\frac{1}{\pi_2-\pi_{11}}\frac{1}{\pi_{11}}(\frac{1}{\pi_1-\pi_{11}}+\frac{1}{1-\pi_1-\pi_2+\pi_{11}})} \\
&= \frac{1}{n_1+n_2\eta} \\
&\times \frac{\frac{1}{\pi_{11}(\pi_2-\pi_{11})}(\pi_{11}+(\pi_2-\pi_{11}))\frac{1}{(\pi_1-\pi_{11})(1-\pi_1-\pi_2+\pi_{11})}((1-\pi_1-\pi_2+\pi_{11})+(\pi_1-\pi_{11}))}{\frac{1}{\pi_{11}(\pi_1-\pi_{11})(\pi_2-\pi_{11})(1-\pi_1-\pi_2+\pi_{11})}(\pi_{11}+(\pi_2-\pi_{11})+(1-\pi_1-\pi_2+\pi_{11})+(\pi_1-\pi_{11}))} \\
&= \frac{1}{n_1+n_2\eta}\frac{\pi_2(1-\pi_2)}{1} \\
&= \frac{\pi_2(1-\pi_2)}{n_1+n_2\eta} \\
I^{-1}_{(1,2)} = I^{-1}_{(2,1)} &= \frac{1}{n_1+n_2\eta}\frac{ac-bd}{abc+abd+acd+bcd} \\
&= \frac{1}{n_1+n_2\eta}\frac{\frac{1}{\pi_1-\pi_{11}}\frac{1}{\pi_2-\pi_{11}}-\frac{1}{1-\pi_1-\pi_2+\pi_{11}}\frac{1}{\pi_{11}}}{\frac{1}{\pi_1-\pi_{11}}\frac{1}{1-\pi_1-\pi_2+\pi_{11}}(\frac{1}{\pi_2-\pi_{11}}+\frac{1}{\pi_{11}})+\frac{1}{\pi_2-\pi_{11}}\frac{1}{\pi_{11}}(\frac{1}{\pi_1-\pi_{11}}+\frac{1}{1-\pi_1-\pi_2+\pi_{11}})} \\
&= \frac{1}{n_1+n_2\eta}\frac{\frac{1}{\pi_{11}(\pi_1-\pi_{11})(\pi_2-\pi_{11})(1-\pi_1-\pi_2+\pi_{11})}(\pi_{11}\times(1-\pi_1-\pi_2+\pi_{11})-(\pi_1-\pi_{11})(\pi_2-\pi_{11}))}{\frac{1}{\pi_{11}(\pi_1-\pi_{11})(\pi_2-\pi_{11})(1-\pi_1-\pi_2+\pi_{11})}(\pi_{11}+(\pi_2-\pi_{11})+(1-\pi_1-\pi_2+\pi_{11})+(\pi_1-\pi_{11}))} \\
&= \frac{1}{n_1+n_2\eta}\frac{\pi_{11}-\pi_1\pi_{11}-\pi_2\pi_{11}+\pi_{11}^2-\pi_1\pi_2+\pi_1\pi_{11}+\pi_2\pi_{11}-\pi_{11}^2}{1} \\
&= \frac{\pi_{11}-\pi_1\pi_2}{n_1+n_2\eta}
\end{aligned}
$$

MLEs and the Fisher Information matrix can now be plugged in to construct the Wald test statistic under $H_0 : (\pi_1, \pi_2)^T = (\pi_{1,0}, \pi_{2,0})^T$:

$$
\begin{aligned}
T_w &= (\widehat{\theta}_1-\theta_{1,0})^T\{[I_T(\widehat{\theta})]^{-1}_{(1,1)}\}^{-1}(\widehat{\theta}_1-\theta_{1,0}) \qquad (4.2)\\
&= \begin{pmatrix} \widehat{\pi}_1-\pi_1, & \widehat{\pi}_2-\pi_2 \end{pmatrix}\begin{pmatrix} \frac{\widehat{\pi}_1(1-\widehat{\pi}_1)}{n_1+n_2\widehat{\eta}} & \frac{\widehat{\pi}_{11}-\widehat{\pi}_1\widehat{\pi}_2}{n_1+n_2\widehat{\eta}} \\ \frac{\widehat{\pi}_{11}-\widehat{\pi}_1\widehat{\pi}_2}{n_1+n_2\widehat{\eta}} & \frac{\widehat{\pi}_2(1-\widehat{\pi}_2)}{n_1+n_2\widehat{\eta}} \end{pmatrix}^{-1}\begin{pmatrix} \widehat{\pi}_1-\pi_1 \\ \widehat{\pi}_2-\pi_2 \end{pmatrix} \\
&= \frac{n_1+n_2\widehat{\eta}}{\widehat{\pi}_1\widehat{\pi}_2(1-\widehat{\pi}_1)(1-\widehat{\pi}_2)-(\widehat{\pi}_{11}-\widehat{\pi}_1\widehat{\pi}_2)^2}\begin{pmatrix} \widehat{\pi}_1-\pi_1, & \widehat{\pi}_2-\pi_2 \end{pmatrix} \\
&\quad\times\begin{pmatrix} \widehat{\pi}_2(1-\widehat{\pi}_2) & -(\widehat{\pi}_{11}-\widehat{\pi}_1\widehat{\pi}_2) \\ -(\widehat{\pi}_{11}-\widehat{\pi}_1\widehat{\pi}_2) & \widehat{\pi}_1(1-\widehat{\pi}_1) \end{pmatrix}\begin{pmatrix} \widehat{\pi}_1-\pi_1 \\ \widehat{\pi}_2-\pi_2 \end{pmatrix} \\
&= \frac{n_1+n_2\widehat{\eta}}{\widehat{\pi}_1\widehat{\pi}_2(1-\widehat{\pi}_1)(1-\widehat{\pi}_2)-(\widehat{\pi}_{11}-\widehat{\pi}_1\widehat{\pi}_2)^2}\{\widehat{\pi}_2(1-\widehat{\pi}_2)(\widehat{\pi}_1-\pi_1)^2
\end{aligned}
$$

$$-2(\widehat{\pi}_1-\pi_1)(\widehat{\pi}_2-\pi_2)(\widehat{\pi}_{11}-\widehat{\pi}_1\widehat{\pi}_2)+\widehat{\pi}_1(1-\widehat{\pi}_1)(\widehat{\pi}_2-\pi_2)^2\Big\}$$

where $\widehat{\eta}=1-\widehat{PET}=1-\sum_{x=0}^{s_1}\sum_{y=0}^{t_1}p(x,y;n_1,\widehat{\pi}_1,\widehat{\pi}_2,\widehat{\pi}_{11})$, and for simplicity of notations we dropped the subscript 'MLE' in the MLE estimators in T_w and also from now on, i.e., $\widehat{\pi}_i=\widehat{\pi}_{i,MLE}$ for $i=1,2$, likewise the convention will apply to other MLE estimators as well.

Under regularity conditions, we have

$$T_w \dot{\sim} \chi^2_2$$

The confidence region then can be constructed by inverting a hypothesis test based on the acceptance region at level α. Therefore, an approximate 100(1-α)% Wald-type confidence region for (π_1,π_2) is the ellipse determined by all (π_1,π_2) such that

$$T_w \le \chi^2_{2,1-\alpha}.$$

For instance, when $\alpha=0.05$, $\chi^2_{2,1-\alpha}=5.991$. The approximate 95% Wald-type confidence region for (π_1,π_2) is an ellipse consisting of all (π_1,π_2) satisfying

$$\begin{aligned}&\widehat{\pi}_2(1-\widehat{\pi}_2)(\widehat{\pi}_1-\pi_1)^2-2(\widehat{\pi}_1-\pi_1)(\widehat{\pi}_2-\pi_2)(\widehat{\pi}_{11}-\widehat{\pi}_1\widehat{\pi}_2)+\widehat{\pi}_1(1-\widehat{\pi}_1)(\widehat{\pi}_2-\pi_2)^2\\&\le\frac{5.991*\widehat{\pi}_1\widehat{\pi}_2(1-\widehat{\pi}_1)(1-\widehat{\pi}_2)-(\widehat{\pi}_{11}-\widehat{\pi}_1\widehat{\pi}_2)^2}{n_1+n_2\widehat{\eta}}.\end{aligned}$$

When we get the data and plug in the MLEs, the above inequality represents an ellipse. But when $\widehat{\pi}_2=0$ or $\widehat{\pi}_2=1$ or $\widehat{\pi}_1=0$ or $\widehat{\pi}_1=1$ or $\widehat{\pi}_2=\widehat{\pi}_1=\widehat{\pi}_{11}$, the above inequality no longer represents a whole ellipse.

An example of an approximate 95% Wald-type confidence region (CR) for a simulated data (assuming true $(\pi_1,\pi_2,\pi_{11})=(0.05,0.25,0.03)$): $(x_1,x,y_1,y,n_{11}^{(1)},n_{11}^{(2)})=(3,5,2,7,1,2)$ from the optimal design $(n,n_1,s_1,t_1,s,t)=(25,12,1,1,3,3)$, is shown in Figure 4.1.

In simulation studies, some confidence regions are of a complete ellipse shape, just like Figure 4.1 shows. However, there are some special cases we only have a partial ellipse as shown in Figure 4.2. The occurrence of incomplete ellipse patterns shown in Figure 4.2 is because the approximate 95% Wald-type confidence ellipse is constrained by the fact that

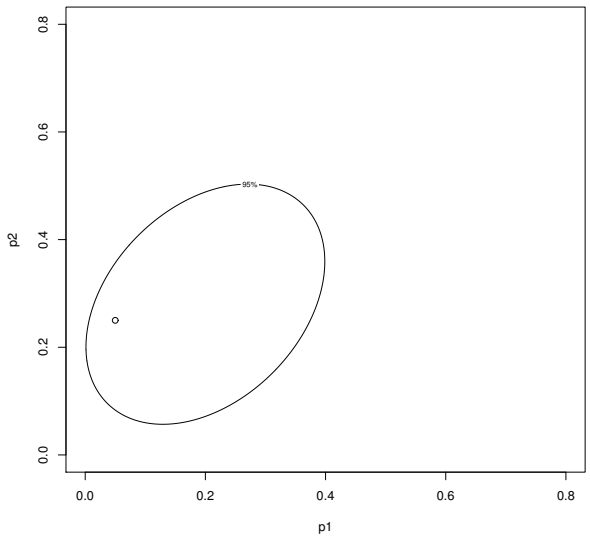


Figure 4.1: The approx. 95% Wald-type Confidence Region for (π_1, π_2) for a simulated data

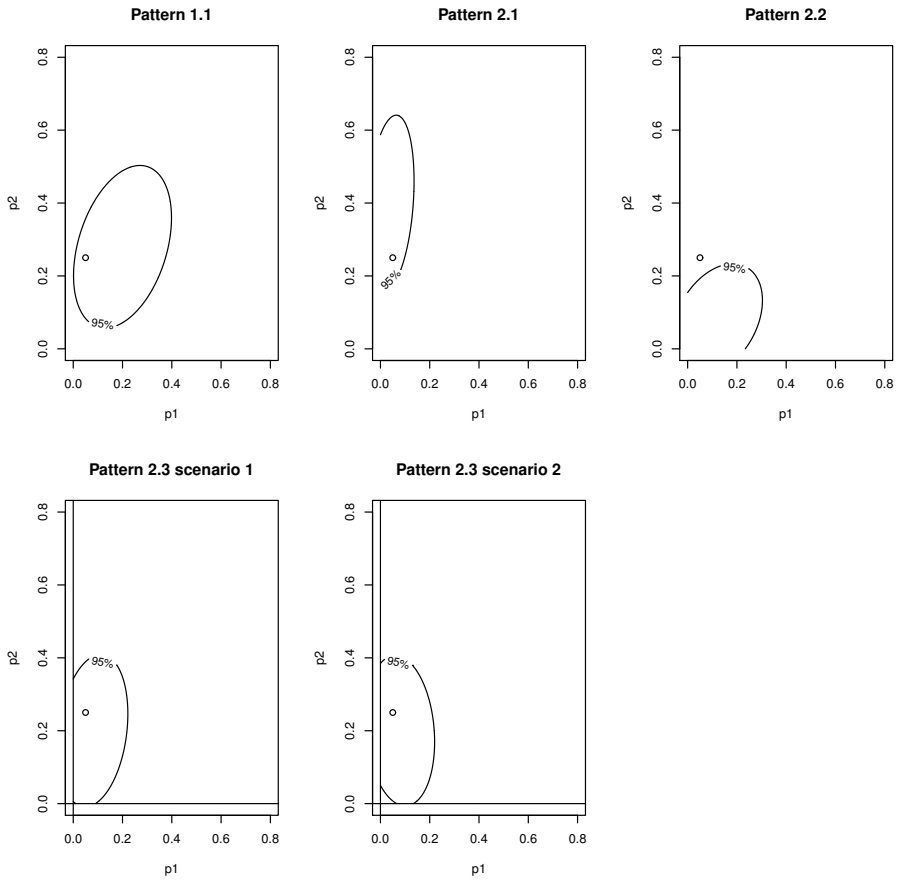


Figure 4.2: Some special examples for the approx. 95% Wald-type CRs from simulations
(upper): Pattern 1.1 Pattern 2.1 Pattern 2.2
(bottom): Pattern 2.3 (left: scenario 1; right: scenario 2)

the range for both π_1 and π_2 is $[0, 1]$. So in the event a Wald-type confidence region was partially outside of the parameter space, we truncated that region to only contain values in the unit square.

4.3.2 Score-type confidence region for (π_1, π_2) in a two-endpoint two-stage design

The second asymptotic confidence region we propose is created by inverting a Score test statistic. Recall that the Score test statistic is:

$$T_s = S_1(\widetilde{\theta})^T \{I_T(\widetilde{\theta})^{-1}\}_{(1,1)} S_1(\widetilde{\theta}),$$

where $\widetilde{\theta}$ is the MLE of θ under H_0, that is,

$$\widetilde{\theta} = \begin{pmatrix} \widetilde{\pi}_1 \\ \widetilde{\pi}_2 \\ \widetilde{\pi}_{11} \end{pmatrix} = \begin{pmatrix} \pi_{1,0} \\ \pi_{2,0} \\ \widetilde{\pi}_{11} \end{pmatrix},$$

and $\widetilde{\pi}_{11}$ is a value of π_{11} from the range $\max(0, \pi_{1,0} + \pi_{2,0} - 1) \leq \pi_{11} \leq \min(\pi_{1,0}, \pi_{2,0})$ such that it maximizes the log-likelihood function $logL(\pi_{1,0}, \pi_{2,0}, \pi_{11})$ under H_0. Under regularity conditions, we have

$$T_s \dot{\sim} \chi^2_2$$

The construction of a Score-type confidence region is shown in the following four steps:

Step 1: Solve for $\widetilde{\pi}_{11}$

There are two methods to get $\widetilde{\pi}_{11}$: method 1 is to find the solution that maximizes the likelihood, for example using "optimize" function in R, and method 2 is solving $\widetilde{\pi}_{11}$ from the cubic equation from the score equation. The details of method 2 are as follows.

The log-likelihood function is

$$\begin{aligned} &\log L(\pi_1, \pi_2, \pi_{11}) \\ &= [n_{11}^{(1)} + n_{11}^{(2)}(M-1)] \log \pi_{11} + [(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)] \log(\pi_1 - \pi_{11}) \end{aligned}$$

$$+ [(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)] \log(\pi_2 - \pi_{11})$$
$$+ [(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)] \log(1 - \pi_1 - \pi_2 + \pi_{11}).$$

Under H_0:

$$\begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} = \begin{pmatrix} \pi_{1,0} \\ \pi_{2,0} \end{pmatrix},$$

the MLE of the nuisance parameter π_{11}, $\widetilde{\pi}_{11}$, can be found by solving the corresponding score equation

$$\frac{\partial \log L(\pi_{1,0}, \pi_{2,0}, \pi_{11})}{\partial \pi_{11}} = 0.$$

That is

$$\begin{aligned}
0 = & \frac{n_{11}^{(1)} + n_{11}^{(2)}(M-1)}{\pi_{11}} - \frac{(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)}{\pi_{1,0} - \pi_{11}} \\
& - \frac{(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)}{\pi_{2,0} - \pi_{11}} \\
& + \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)}{(1 - \pi_{1,0} - \pi_{2,0} + \pi_{11})}, \\
0 = & \frac{1}{\pi_{11}(\pi_{1,0} - \pi_{11})(\pi_{2,0} - \pi_{11})(1 - \pi_{1,0} - \pi_{2,0} + \pi_{11})} \\
& \times \{[n_{11}^{(1)} + n_{11}^{(2)}(M-1)](\pi_{1,0} - \pi_{11})(\pi_{2,0} - \pi_{11})(1 - \pi_{1,0} - \pi_{2,0} + \pi_{11}) \\
& - [(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)]\pi_{11}(\pi_{2,0} - \pi_{11})(1 - \pi_{1,0} - \pi_{2,0} + \pi_{11}) \\
& - [(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)]\pi_{11}(\pi_{1,0} - \pi_{11})(1 - \pi_{1,0} - \pi_{2,0} + \pi_{11}) \\
& + [(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)] \\
& \quad \times \pi_{11}(\pi_{1,0} - \pi_{11})(\pi_{2,0} - \pi_{11})\}, \\
0 = & [n_{11}^{(1)} + n_{11}^{(2)}(M-1)](\pi_{1,0} - \pi_{11})(\pi_{2,0} - \pi_{11})(1 - \pi_{1,0} - \pi_{2,0} + \pi_{11}) \\
& - [(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)]\pi_{11}(\pi_{2,0} - \pi_{11})(1 - \pi_{1,0} - \pi_{2,0} + \pi_{11}) \\
& - [(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)]\pi_{11}(\pi_{1,0} - \pi_{11})(1 - \pi_{1,0} - \pi_{2,0} + \pi_{11}) \\
& + [(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)] \\
& \quad \times \pi_{11}(\pi_{1,0} - \pi_{11})(\pi_{2,0} - \pi_{11}).
\end{aligned}$$

Let

$$A = [n_{11}^{(1)} + n_{11}^{(2)}(M-1)], B = [(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)],$$
$$C = [(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)],$$
$$D = [(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)],$$

then the above equation can be represented as:

$$\begin{aligned}
0 &= A(\pi_{1,0} - \pi_{11})(\pi_{2,0} - \pi_{11})(1 - \pi_{1,0} - \pi_{2,0} + \pi_{11}) \\
&\quad - B\pi_{11}(\pi_{2,0} - \pi_{11})(1 - \pi_{1,0} - \pi_{2,0} + \pi_{11}) - C\pi_{11}(\pi_{1,0} - \pi_{11})(1 - \pi_{1,0} - \pi_{2,0} + \pi_{11}) \\
&\quad + D\pi_{11}(\pi_{1,0} - \pi_{11})(\pi_{2,0} - \pi_{11}) \\
&= (A + B + C + D)\pi_{11}^3 \\
&\quad + \{A(1 - \pi_{1,0} - \pi_{2,0} - \pi_{1,0} - \pi_{2,0}) + B(1 - \pi_{1,0} - \pi_{2,0} - \pi_{2,0}) \\
&\quad + C(1 - \pi_{1,0} - \pi_{2,0} - \pi_{1,0}) \quad - D(\pi_{1,0} + \pi_{2,0})\}\pi_{11}^2 \\
&+ \{A(\pi_{1,0}\pi_{2,0} - (\pi_{1,0} + \pi_{2,0})(1 - \pi_{1,0} - \pi_{2,0})) - B\pi_{2,0}(1 - \pi_{1,0} - \pi_{2,0}) \\
&\quad - C\pi_{1,0}(1 - \pi_{1,0} - \pi_{2,0}) + D\pi_{1,0}\pi_{2,0}\}\pi_{11} \\
&\quad + A\pi_{1,0}\pi_{2,0}(1 - \pi_{1,0} - \pi_{2,0}) \\
&= (A + B + C + D)\pi_{11}^3 \\
&\quad + \pi_{11}^2\{(A + B + C) - \pi_{1,0}(2A + B + 2C + D) - \pi_{2,0}(2A + 2B + C + D)\} \\
&\quad + \pi_{11}\{\pi_{1,0}^2(A + C) + \pi_{2,0}^2(A + B) + \pi_{1,0}\pi_{2,0}(3A + B + C + D) - \pi_{1,0}(A + C) \\
&\quad - \pi_{2,0}(A + B)\} + A\pi_{1,0}\pi_{2,0}(1 - \pi_{1,0} - \pi_{2,0}).
\end{aligned}$$

We have

$$\begin{aligned}
A + B + C + D &= n_{11}^{(1)} + n_{11}^{(2)}(M-1) + [(x_1 - n_{11}^{(1)}) + (x_2 - n_{11}^{(2)})(M-1)] \\
&\quad + [(y_1 - n_{11}^{(1)}) + (y_2 - n_{11}^{(2)})(M-1)] \\
&\quad + [(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (n_2 - x_2 - y_2 + n_{11}^{(2)})(M-1)] \\
&= n_1 + (M-1)n_2,
\end{aligned}$$

$$
\begin{aligned}
A+B+C &= (x_1+y_1-n_{11}^{(1)})+(M-1)(x_2+y_2-n_{11}^{(2)}),\\
2A+B+2C+D &= n_1+(M-1)n_2+n_{11}^{(1)}+n_{11}^{(2)}(M-1)+[(y_1-n_{11}^{(1)})\\
&\quad +(y_2-n_{11}^{(2)})(M-1)]\\
&= n_1+y_1+(M-1)(n_2+y_2),\\
2A+2B+C+D &= n_1+(M-1)n_2+n_{11}^{(1)}+n_{11}^{(2)}(M-1)+[(x_1-n_{11}^{(1)})\\
&\quad +(x_2-n_{11}^{(2)})(M-1)]\\
&= n_1+x_1+(M-1)(n_2+x_2),\\
A+C &= y_1+(M-1)y_2, A+B = x_1+(M-1)x_2,\\
3A+B+C+D &= n_1+(M-1)n_2+2[n_{11}^{(1)}+n_{11}^{(2)}(M-1)]\\
&= (n_1+2n_{11}^{(1)})+(M-1)(n_2+2n_{11}^{(2)}).
\end{aligned}
$$

After re-arrangement, the above equation can be expressed in the format of

$$a\pi_{11}^3+b\pi_{11}^2+c\pi_{11}+d=0$$

where

$$
\begin{aligned}
a &= n_1+(M-1)n_2,\\
b &= -\pi_{1,0}[(y_1+n_1)+(M-1)(y_2+n_2)]-\pi_{2,0}[(x_1+n_1)+(M-1)(x_2+n_2)]\\
&\quad +[(x_1+y_1-n_{11}^{(1)})+(M-1)(x_2+y_2-n_{11}^{(2)})],\\
c &= \pi_{1,0}\pi_{2,0}[(n_1+2n_{11}^{(1)})+(M-1)(n_2+2n_{11}^{(2)})]-\pi_{2,0}(1-\pi_{2,0})[x_1+(M-1)x_2]\\
&\quad -\pi_{1,0}(1-\pi_{1,0})[y_1+(M-1)y_2],\\
d &= \pi_{1,0}\pi_{2,0}(1-\pi_{1,0}-\pi_{2,0})[n_{11}^{(1)}+(M-1)n_{11}^{(2)}].
\end{aligned}
$$

For cubic equation $ax^3+bx^2+cx+d=0$, the solution is

$$x=\{q+[q^2+(r-p^2)^3]^{1/2}\}^{1/3}+\{q-[q^2+(r-p^2)^3]^{1/2}\}^{1/3}+p$$

where

$$p = -\frac{b}{3a}, q = p^3 + \frac{bc - 3ad}{6a^2}, r = \frac{c}{3a}$$

i.e.

$$x = \sqrt[3]{\left(-\frac{b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right) + \sqrt{\left(-\frac{b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right)^2 + \left(\frac{c}{3a} - \frac{b^2}{9a^2}\right)^3}} + \sqrt[3]{\left(-\frac{b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right) - \sqrt{\left(-\frac{b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right)^2 + \left(\frac{c}{3a} - \frac{b^2}{9a^2}\right)^3}} - \frac{b}{3a}$$

This is an explicit solution formula for $\widetilde{\pi}_{11}$ from a cubic equation. With the help of a computer, the complicated-looking roots, which may contain complex numbers, can be obtained. If the solution for $\widetilde{\pi}_{11}$ from the cubic equation does not fall in the range $\{\max(0, \pi_{1,0} + \pi_{2,0} - 1) \leq \pi_{11} \leq \min(\pi_{1,0}, \pi_{2,0})\}$, then check the boundaries of the range for π_{11}.

Step 2: Derive the Score statistic after getting $\widetilde{\pi}_{11}$

By inverting the Score-type test statistic, an approximate $100(1-\alpha)\,\%$ Score-type confidence region for (π_1, π_2) is the region determined by all (π_1, π_2) such that

$$S_1(\widetilde{\theta})^T \{I_T(\widetilde{\theta})^{-1}\}_{(1,1)} S_1(\widetilde{\theta}) \leq \chi^2_{2,1-\alpha},$$

that is,

$$\left\{(\pi_1, \pi_2) | T_s = S_1(\widetilde{\theta})^T \{I_T(\widetilde{\theta})^{-1}\}_{(1,1)} S_1(\widetilde{\theta}) \leq \chi^2_{2,1-\alpha}\right\}.$$

Let

$$A = \frac{(x_1 - n_{11}^{(1)}) + (M-1)(x_2 - n_{11}^{(2)})}{\pi_1 - \widetilde{\pi}_{11}}$$
$$B = \frac{(y_1 - n_{11}^{(1)}) + (M-1)(y_2 - n_{11}^{(2)})}{\pi_2 - \widetilde{\pi}_{11}}$$
$$C = \frac{(n_1 - x_1 - y_1 + n_{11}^{(1)}) + (M-1)(n_2 - x_2 - y_2 + n_{11}^{(2)})}{1 - \pi_1 - \pi_2 + \widetilde{\pi}_{11}}$$

then

$$
\begin{aligned}
T_s &= S_1(\widetilde{\theta})^T \{I_T(\widetilde{\theta})^{-1}\}_{(1,1)} S_1(\widetilde{\theta}) \\
&= \begin{pmatrix} \frac{(x_1-n_{11}^{(1)})+(M-1)(x_2-n_{11}^{(2)})}{\pi_1-\widetilde{\pi}_{11}} - \frac{(n_1-x_1-y_1+n_{11}^{(1)})+(M-1)(n_2-x_2-y_2+n_{11}^{(2)})}{1-\pi_1-\pi_2+\widetilde{\pi}_{11}} \\ \frac{(y_1-n_{11}^{(1)})+(M-1)(y_2-n_{11}^{(2)})}{\pi_2-\widetilde{\pi}_{11}} - \frac{(n_1-x_1-y_1+n_{11}^{(1)})+(M-1)(n_2-x_2-y_2+n_{11}^{(2)})}{1-\pi_1-\pi_2+\widetilde{\pi}_{11}} \end{pmatrix}^T \\
&\quad \times \begin{pmatrix} \frac{\pi_1(1-\pi_1)}{n_1+n_2\eta} & \frac{\widetilde{\pi}_{11}-\pi_1\pi_2}{n_1+n_2\eta} \\ \frac{\widetilde{\pi}_{11}-\pi_1\pi_2}{n_1+n_2\eta} & \frac{\pi_2(1-\pi_2)}{n_1+n_2\eta} \end{pmatrix} \\
&\quad \times \begin{pmatrix} \frac{(x_1-n_{11}^{(1)})+(M-1)(x_2-n_{11}^{(2)})}{\pi_1-\widetilde{\pi}_{11}} - \frac{(n_1-x_1-y_1+n_{11}^{(1)})+(M-1)(n_2-x_2-y_2+n_{11}^{(2)})}{1-\pi_1-\pi_2+\widetilde{\pi}_{11}} \\ \frac{(y_1-n_{11}^{(1)})+(M-1)(y_2-n_{11}^{(2)})}{\pi_2-\widetilde{\pi}_{11}} - \frac{(n_1-x_1-y_1+n_{11}^{(1)})+(M-1)(n_2-x_2-y_2+n_{11}^{(2)})}{1-\pi_1-\pi_2+\widetilde{\pi}_{11}} \end{pmatrix} \\
&= \frac{1}{n_1+n_2\eta} \begin{pmatrix} A-C & B-C \end{pmatrix} \begin{pmatrix} \pi_1(1-\pi_1) & \widetilde{\pi}_{11}-\pi_1\pi_2 \\ \widetilde{\pi}_{11}-\pi_1\pi_2 & \pi_2(1-\pi_2) \end{pmatrix} \begin{pmatrix} A-C \\ B-C \end{pmatrix} \\
&= \frac{1}{n_1+n_2\eta} \{(A-C)^2\pi_1(1-\pi_1) + 2(A-C)(B-C)(\widetilde{\pi}_{11}-\pi_1\pi_2) \\
&\qquad + (B-C)^2\pi_2(1-\pi_2))\}
\end{aligned}
$$

Since $\widetilde{\pi}_{11}$ does not have a simple form and is a function of $(\pi_{1,0}, \pi_{2,0})$, and also η is a function of $(\pi_{1,0}, \pi_{2,0}, \widetilde{\pi}_{11})$, there is no simple form for the Score statistic T_s. We can use the grid searching method to find the asymptotic Score-type Confidence Region numerically.

The process for locating the Score-type confidence region for (π_1, π_2) based on the observed data through grid searching is as follows.

1. For the observed data $(M, x_1, x, y_1, y, n_{11}^{(1)}, n_{11}^{(2)})$ with the design (n, n_1, s_1, t_1), divide the parameter space for (π_1, π_2) into small grids. At each grid point of (π_1, π_2), solve for $\widetilde{\pi}_{11}$, either through the cubic equation

$$a\pi_{11}^3 + b\pi_{11}^2 + c\pi_{11} + d = 0,$$

   or through the R function "optimize( )" (preferring the latter method).

2. Then η can be calculated since η is a function of $(\pi_1, \pi_2, \widetilde{\pi}_{11})$, i.e.,

$$\eta(\pi_1, \pi_2, \widetilde{\pi}_{11}) = 1 - PET = 1 - \sum_{x=0}^{s_1} \sum_{y=0}^{t_1} p(x, y; n_1, \pi_1, \pi_2, \widetilde{\pi}_{11}).$$

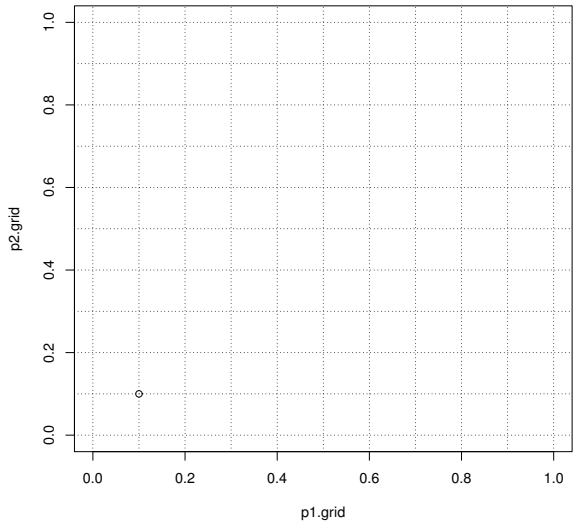


Figure 4.3: Sketch of the division of the parameter space for (π_1, π_2) into small grids

3. Calculate the Score test statistic T_s at this grid point. If $T_s \leq \chi^2_{2,1-\alpha}$, then the current grid point is within the Score-type confidence region. Otherwise, the current grid point is not part of the Score-type confidence region.

4. Repeat Step $(1) - (3)$ for each grid point of (π_1, π_2).

After exhaustive grid searching, and if the grid is sufficiently small, the Score-type confidence region will consist of those grid points satisfying $T_s \leq \chi^2_{2,1-\alpha}$.

The following Figure 4.4 exemplifies the located Score-type confidence region by grid searching for the simulated dataset $(M, x_1, x_2, y_1, y_2, n_{11}^{(1)}, n_{11}^{(2)}) = (2, 16, 15, 15, 20, 11, 11)$, which is generated from the design $(\pi_1, \pi_2, n, n_1, s_1, t_1) = (0.4, 0.6, 60, 23, 11, 16)$ assuming $\pi_{11} = 0.3$.

Since for each simulated dataset, each of the grid points in the parameter space for (π_1, π_2) will be checked to see whether the Score statistic at that point satisfying $T_s \leq \chi^2_{2,1-\alpha}$ and then decide whether that grid point is part of the Score-type CR, so the time cost for locating a Score-type CR is higher than that for locating a Wald-type CR.

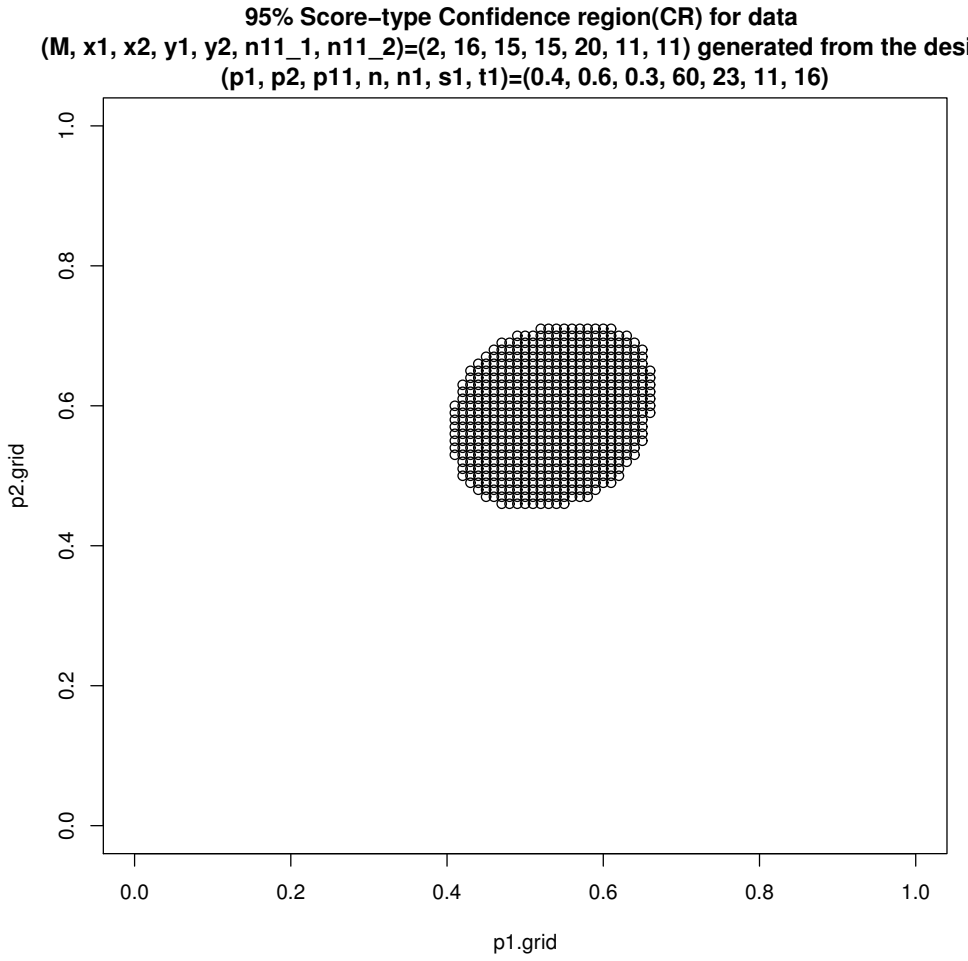


Figure 4.4: 95% Score-type Confidence Region for (π_1, π_2) for a simulated dataset

4.3.3 Likelihood ratio type confidence region for (π_1, π_2) in a two-endpoint two-stage design

The third asymptotic confidence region we propose is created by inverting a Likelihood ratio test statistic. The Likelihood ratio statistic is:

$$T_{LR} = -2\log\frac{\sup_{\theta\in H_0} L(\theta|Y)}{\sup_{\theta\in\Theta} L(\theta|Y)} = -2(l(\widetilde{\theta}|Y) - l(\widehat{\theta}|Y)),$$

where $\widetilde{\theta}$ is the MLE of θ under H_0, that is,

$$\widetilde{\theta} = \begin{pmatrix} \widetilde{\pi}_1 \\ \widetilde{\pi}_2 \\ \widetilde{\pi}_{11} \end{pmatrix} = \begin{pmatrix} \pi_{1,0} \\ \pi_{2,0} \\ \widetilde{\pi}_{11} \end{pmatrix},$$

and $\widetilde{\pi}_{11}$ maximize log-likelihood$(\pi_{1,0}, \pi_{2,0}, \pi_{11})$ under H_0 in the range of $\max(0, \pi_{1,0} + \pi_{2,0} - 1) \leq \pi_{11} \leq \min(\pi_{1,0}, \pi_{2,0})$. Under regularity conditions, we have

$$T_{LR} \dot{\sim} \chi^2_2.$$

When the data at hand implies the MLE estimates under the whole sample space such that

$$\widehat{\pi}_{11} = \widehat{\pi}_1 = \widehat{\pi}_2,$$

then the statistic T_{LR} is undefined. So in simulation studies later, when calculating coverage probability of LR-type CRs and the expected area of the LR-type CRs, such datasets will be excluded.

Since $\widetilde{\pi}_{11}$ is involved, we can also use the strategy of grid searching to locate the asymptotic Likelihood ratio-type Confidence Region numerically. The process is very similar to that described in the section of locating a Score-type confidence region.

The following Figure 4.5 exemplifies the located Likelihood ratio-type confidence region by grid searching for the simulated dataset

$$(M, x_1, x_2, y_1, y_2, n_{11}^{(1)}, n_{11}^{(2)}) = (2, 11, 12, 20, 24, 4, 6),$$

which is generated from the design $(\pi_1, \pi_2, n, n_1, s_1, t_1) = (0.35, 0.45, 87, 37, 9, 13)$ assuming $\pi_{11} = 0.18$.

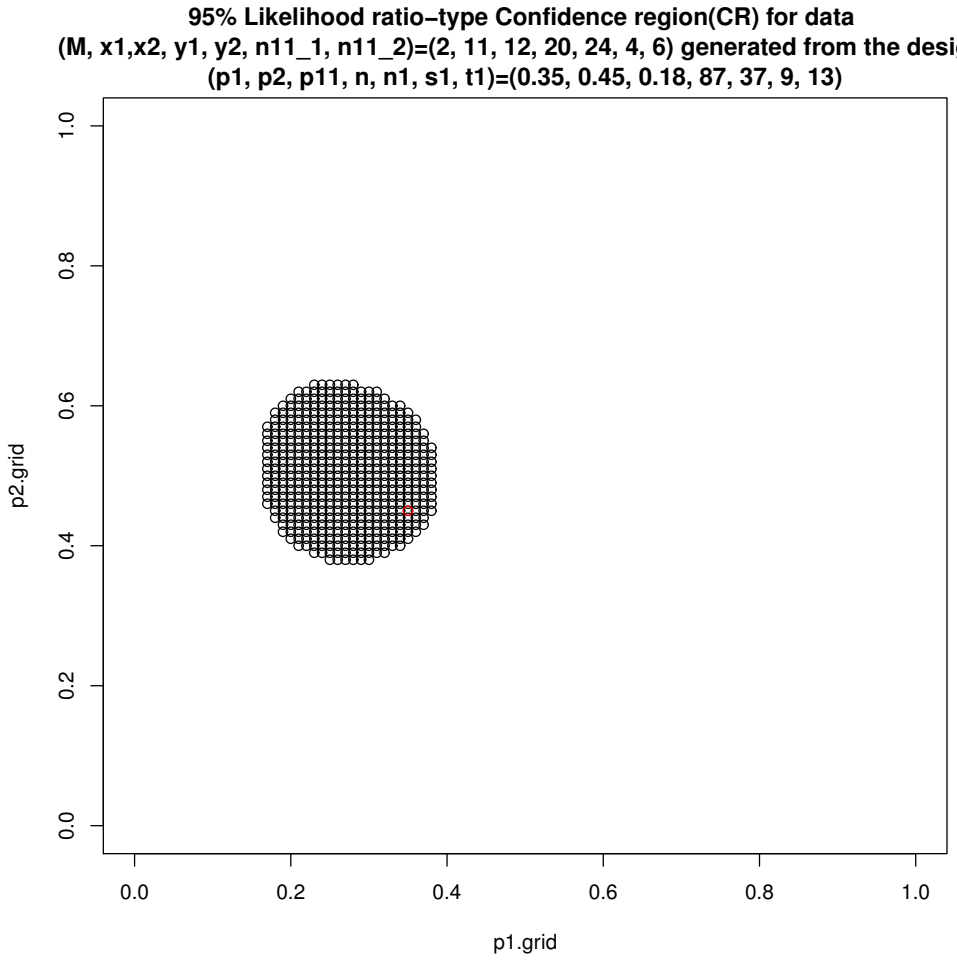


Figure 4.5: 95% Likelihood ratio-type Confidence Region for (π_1, π_2) for a simulated dataset

4.3.4 Comparison of Wald-type, Score-type and LR-type confidence regions

Coverage probability and expected area are used as evaluation criteria to compare the performance of Wald-type, Score-type and LR-type confidence region estimators.

Criterion I. Coverage probability

The region R $(\mathbf{X})$ is said to be a $100(1-\alpha)\%$ *confidence region* if,

$$P[R(\mathbf{X}) \text{ will cover the true } \theta] = 1 - \alpha.$$

This probability is calculated under the true, but unknown, value of θ.

The actual coverage probability (CP) of a confidence region can be calculated as:

$$CP = \sum_{all\ possible\ (m,x,y)} I_{CR}(\pi_1, \pi_2) f(m, x, y|\pi_1, \pi_2)$$

where $I_{CR}(\pi_1, \pi_2)$ is an indicator function that is equal to 1 if the true (π_1, π_2) lies in the region CR, and equal to 0 otherwise; and $f(m, x, y|\pi_1, \pi_2)$ is the probability mass function for the random vector (M, X, Y). Given the impracticality of summing over infinitely many values of (M, X, Y), we estimate the coverage probability of each confidence region using a Monte Carlo simulation, rather than exact theoretical calculation. The schema of calculating the coverage probability of each type of confidence region is shown in Figure 4.6.

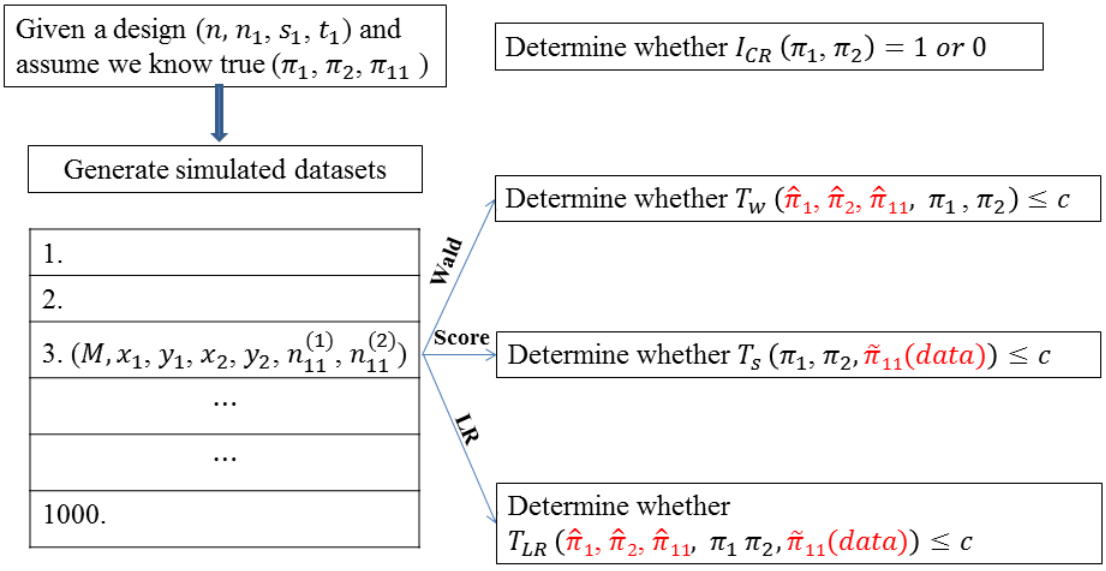


Figure 4.6: Schema of Calculating the Coverage Prob. of each type of CR

After choosing some representative optimal designs from Table 3.6 to 3.9, and considering

the scenarios of low/medium/high positive correlation between the two alternative binary efficacy endpoints, where the level of correlation can be described by π_{11} or the correlation coefficient ρ, the coverage probabilities of the Wald-type, Score-type and Likelihood ratio-type CRs were calculated and compared using simulation sample size = 1000. The results are in Table 4.1.

Because discrete random variables can only take on specified values, the correction for continuity adjustment is employed. Since only Wald test statistic has closed form, the continuity correction is only applied to Wald statistic by plugging in $(\widehat{\theta}_1 - \theta_{1,0} - cc)$ to the underlined parts: $T_w = \underline{(\widehat{\theta}_1 - \theta_{1,0})^T}\{[I_T(\widehat{\theta})]^{-1}_{(1,1)}\}^{-1}\underline{(\widehat{\theta}_1 - \theta_{1,0})}$. Four different factors for continuity correction were tried in the position of "cc": $cc1 = -ccf, cc2 = ccf, cc3 = 2 \times ccf, cc4 = -2 \times ccf$, where

$$ccf = \frac{1}{2 \times [n_1 + n_2(M-1)]}$$

based on Tsai et al. (2008).

Table 4.1: Table of Coverage Probabilities (samsize=1000)

(n, n_1, s_1, t_1, s, t)	(p_1, p_2)	π_{11}	ρ	wald	wald_cc1	wald_cc2	wald_cc3	wald_cc4	score	LR
(51, 24, 3, 3, 9, 9)	(0.1, 0.1)	0.02	0.11	0.875	0.874	0.805	0.611	0.869	0.905	0.956
		0.05	0.44	0.880	0.881	0.766	0.612	0.878	0.844	0.975
		0.08	0.78	0.923	0.922	0.889	0.770	0.918	0.592	0.953
	(0.25, 0.1)	0.04	0.12	0.882	0.900	0.804	0.794	0.897	0.940	0.956
		0.08	0.42	0.886	0.901	0.814	0.798	0.899	0.892	0.977
		0.09	0.50	0.867	0.881	0.807	0.792	0.882	0.793	0.975
	(0.1, 0.25)	0.04	0.12	0.874	0.903	0.778	0.773	0.902	0.951	0.968
		0.08	0.42	0.869	0.886	0.806	0.797	0.887	0.888	0.976
		0.09	0.50	0.866	0.874	0.802	0.784	0.869	0.792	0.959
	(0.25, 0.25)	0.08	0.09	0.931	0.924	0.897	0.891	0.923	0.947	0.952
		0.15	0.47	0.910	0.918	0.903	0.883	0.921	0.949	0.947
		0.20	0.73	0.917	0.930	0.912	0.891	0.940	0.958	0.981
(87, 37, 9, 13, 25, 33)	(0.2, 0.3)	0.08	0.11	0.907	0.923	0.889	0.836	0.937	0.918	0.957
		0.15	0.49	0.902	0.915	0.870	0.831	0.931	0.901	0.956
		0.19	0.71	0.872	0.894	0.852	0.837	0.899	0.629	0.960
	(0.35, 0.3)	0.13	0.11	0.915	0.930	0.893	0.862	0.934	0.955	0.953
		0.20	0.43	0.911	0.919	0.891	0.859	0.921	0.949	0.938
		0.27	0.75	0.897	0.905	0.878	0.851	0.922	0.902	0.947
	(0.2, 0.45)	0.11	0.10	0.916	0.919	0.898	0.875	0.921	0.948	0.948

Table 4.1 (Cont.). Table of Coverage Probabilities (samsize=1000)

		0.18	0.45	0.903	0.910	0.882	0.854	0.912	0.788	0.953
		0.19	0.50	0.890	0.900	0.871	0.853	0.905	0.571	0.957
	(0.35, 0.45)	0.18	0.09	0.925	0.923	0.923	0.920	0.914	0.954	0.942
		0.26	0.43	0.926	0.931	0.917	0.912	0.929	0.952	0.942
		0.34	0.77	0.898	0.910	0.892	0.866	0.911	0.551	0.947
(55, 25, 9, 12, 23, 28)	(0.3, 0.4)	0.14	0.09	0.926	0.938	0.894	0.855	0.949	0.920	0.964
		0.22	0.45	0.901	0.914	0.887	0.850	0.926	0.915	0.962
		0.29	0.76	0.903	0.903	0.899	0.854	0.925	0.619	0.956
	(0.5, 0.4)	0.23	0.12	0.913	0.925	0.888	0.868	0.923	0.961	0.951
		0.31	0.45	0.903	0.901	0.879	0.868	0.919	0.946	0.945
		0.39	0.78	0.898	0.908	0.873	0.864	0.919	0.503	0.937
	(0.3, 0.6)	0.21	0.13	0.918	0.927	0.904	0.875	0.919	0.953	0.958
		0.28	0.45	0.900	0.909	0.885	0.871	0.901	0.648	0.950
		0.29	0.49	0.918	0.923	0.888	0.880	0.916	0.419	0.951
	(0.5, 0.6)	0.33	0.12	0.918	0.919	0.917	0.915	0.916	0.945	0.938
		0.41	0.45	0.919	0.933	0.912	0.897	0.931	0.949	0.943
		0.48	0.73	0.906	0.905	0.898	0.886	0.905	0.726	0.946
(60, 23, 11, 16, 30, 43)	(0.4, 0.6)	0.30	0.25	0.926	0.930	0.890	0.867	0.933	0.909	0.957
		0.36	0.50	0.925	0.929	0.897	0.881	0.939	0.824	0.947
		0.39	0.63	0.905	0.905	0.893	0.866	0.918	0.553	0.929

Table 4.1 (Cont.). Table of Coverage Probabilities (samsize=1000)

	(0.6, 0.6)	0.39	0.13	0.911	0.918	0.906	0.862	0.922	0.952	0.953
		0.47	0.46	0.910	0.906	0.900	0.871	0.907	0.947	0.942
		0.54	0.75	0.914	0.915	0.890	0.853	0.923	0.941	0.952
	(0.4, 0.8)	0.34	0.10	0.907	0.916	0.899	0.873	0.897	0.931	0.946
		0.36	0.20	0.920	0.925	0.908	0.878	0.907	0.862	0.956
		0.39	0.36	0.914	0.916	0.903	0.878	0.903	0.433	0.969
	(0.6, 0.8)	0.50	0.10	0.910	0.911	0.911	0.902	0.899	0.951	0.941
		0.56	0.41	0.923	0.925	0.910	0.904	0.918	0.895	0.961
		0.59	0.56	0.925	0.925	0.906	0.892	0.916	0.510	0.964

Observation in Table 4.1: Similar to the phenomenon observed by many other researchers, the coverage probability of the Wald-type CR is below the nominal confidence level for this two-endpoint two-stage design. The calculation result shows that using $cc4 = -2 \times ccf$ has slightly better improvement on the coverage probability of Wald-type CR (Table 4.1). The factor "2" in $cc4$ may be due to that we have two endpoints here, and the minus sign in $cc4$ may be justified by the fact that the MLEs (sample proportion) are always negatively biased (Jung and Kim, 2004). It is unusual to see that the coverage probability of the Score-type CR is so unstable. In some cases, the coverage probability of the Score-type CR is much better than that of Wald-type CR; while in some other cases, the coverage probability of Score-type CR is really small. The overall coverage probability of the LR-type CR is good.

Criterion II. Expected Area

A. Calculation of the expected area of Wald-type CRs

The area of one Wald-type confidence region(an ellipse) for an observed dataset can be regarded as an area between two curves, which can be calculated as the integral of "top curve minus bottom curve". Suppose the upper and lower curves can respectively be expressed as:

$$\text{upper curve: } Y = F(x)$$
$$\text{lower curve: } Y = G(x)$$

and the two curves intersect at two end points $(a_{L,}F(a_L))$ and $(a_{R,}F(a_R))$(as shown in Figure 4.7),

then

$$\text{Area between two curves} = \int_{a_L}^{a_R} [F(x) - G(x)]dx,$$

where a_L and a_R are the solutions for $F(x) = G(x)$.

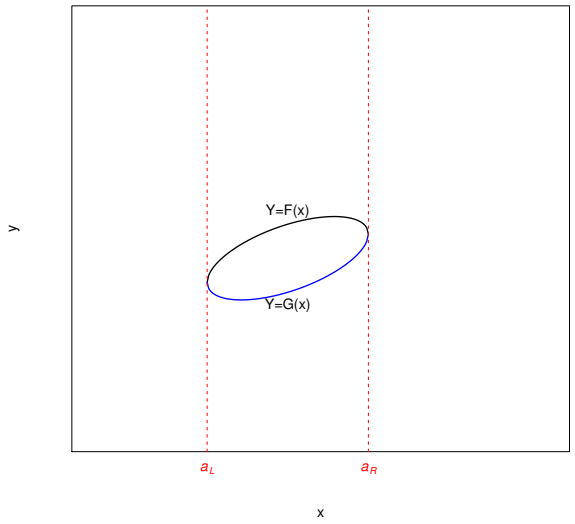


Figure 4.7: Calculation of the area between two curves

Suppose the equation outlining the Wald-type confidence region/ellipse is

$$ax^2 + bxy + cy^2 = d,$$

then we have

$$\begin{aligned} cy^2 + bxy + (ax^2 - d) &= 0, \\ y &= \frac{-bx \pm \sqrt{(bx)^2 - 4c(ax^2 - d)}}{2c}. \end{aligned} \tag{4.3}$$

This implies

$$\begin{aligned} F(x) &= \frac{-bx + \sqrt{(bx)^2 - 4c(ax^2 - d)}}{2c}, \\ G(x) &= \frac{-bx - \sqrt{(bx)^2 - 4c(ax^2 - d)}}{2c}. \end{aligned}$$

Therefore a_L and a_R are essentially the solutions of $(bx)^2 - 4c(ax^2 - d) = 0$.

$$\begin{aligned} a_L &= -\sqrt{\frac{4cd}{4ac - b^2}}, \\ a_R &= \sqrt{\frac{4cd}{4ac - b^2}}. \end{aligned}$$

Recall that the equation outlining the ellipse of a 95% Wald-type confidence region is

$$5.991 = \frac{n_1 + n_2\widehat{\eta}}{\widehat{\pi}_1\widehat{\pi}_2(1-\widehat{\pi}_1)(1-\widehat{\pi}_2) - (\widehat{\pi}_{11} - \widehat{\pi}_1\widehat{\pi}_2)^2} \left\{\widehat{\pi}_2(1-\widehat{\pi}_2)(\widehat{\pi}_1 - \pi_1)^2 \right.$$
$$\left. -2(\widehat{\pi}_1 - \pi_1)(\widehat{\pi}_2 - \pi_2)(\widehat{\pi}_{11} - \widehat{\pi}_1\widehat{\pi}_2) + \widehat{\pi}_1(1-\widehat{\pi}_1)(\widehat{\pi}_2 - \pi_2)^2\right\}.$$

After re-arrangement, it can be written as

$$\begin{aligned} 0 = {} & [\widehat{\pi}_1(1-\widehat{\pi}_1)]{\pi_2}^2 + 2\pi_2\{\widehat{\pi}_1\widehat{\pi}_{11} - \widehat{\pi}_1\widehat{\pi}_2 - \pi_1\widehat{\pi}_{11} + \pi_1\widehat{\pi}_1\widehat{\pi}_2\} \\ & - 5.991\frac{\widehat{\pi}_1\widehat{\pi}_2(1-\widehat{\pi}_1)(1-\widehat{\pi}_2) - (\widehat{\pi}_{11} - \widehat{\pi}_1\widehat{\pi}_2)^2}{n_1 + n_2\widehat{\eta}} \\ & + \widehat{\pi}_1\widehat{\pi}_2(\widehat{\pi}_1 - 2\widehat{\pi}_{11} + \widehat{\pi}_2) + 2\pi_1\widehat{\pi}_2(\widehat{\pi}_{11} - \widehat{\pi}_1) + \widehat{\pi}_2(1-\widehat{\pi}_2)\pi_1^2. \end{aligned}$$

If we use the notation in formula (4.3) and let $y = \pi_2$, then

$$\begin{aligned} c &= \widehat{\pi}_1(1-\widehat{\pi}_1) \\ bx &= 2\{\widehat{\pi}_1\widehat{\pi}_{11} - \widehat{\pi}_1\widehat{\pi}_2 - \pi_1\widehat{\pi}_{11} + \pi_1\widehat{\pi}_1\widehat{\pi}_2\} \\ ax^2 - d &= -5.991\frac{\widehat{\pi}_1\widehat{\pi}_2(1-\widehat{\pi}_1)(1-\widehat{\pi}_2) - (\widehat{\pi}_{11} - \widehat{\pi}_1\widehat{\pi}_2)^2}{n_1 + n_2\widehat{\eta}} \\ & + \widehat{\pi}_1\widehat{\pi}_2(\widehat{\pi}_1 - 2\widehat{\pi}_{11} + \widehat{\pi}_2) + 2\pi_1\widehat{\pi}_2(\widehat{\pi}_{11} - \widehat{\pi}_1) + \widehat{\pi}_2(1-\widehat{\pi}_2)\pi_1^2. \end{aligned}$$

Thus

$$\begin{aligned} F(\pi_1) &= \frac{-2\{\widehat{\pi}_1\widehat{\pi}_{11} - \widehat{\pi}_1\widehat{\pi}_2 - \pi_1\widehat{\pi}_{11} + \pi_1\widehat{\pi}_1\widehat{\pi}_2\} + \sqrt{\Delta(\pi_1)}}{2\widehat{\pi}_1(1-\widehat{\pi}_1)}, \\ G(\pi_1) &= \frac{-2\{\widehat{\pi}_1\widehat{\pi}_{11} - \widehat{\pi}_1\widehat{\pi}_2 - \pi_1\widehat{\pi}_{11} + \pi_1\widehat{\pi}_1\widehat{\pi}_2\} - \sqrt{\Delta(\pi_1)}}{2\widehat{\pi}_1(1-\widehat{\pi}_1)}, \end{aligned}$$

where

$$\begin{aligned} \Delta(\pi_1) = {} & 4\{\widehat{\pi}_1\widehat{\pi}_{11} - \widehat{\pi}_1\widehat{\pi}_2 - \pi_1\widehat{\pi}_{11} + \pi_1\widehat{\pi}_1\widehat{\pi}_2\}^2 \\ & - 4\widehat{\pi}_1(1-\widehat{\pi}_1)\{\widehat{\pi}_1\widehat{\pi}_2(\widehat{\pi}_1 - 2\widehat{\pi}_{11} + \widehat{\pi}_2) + 2\pi_1\widehat{\pi}_2(\widehat{\pi}_{11} - \widehat{\pi}_1) \\ & + \widehat{\pi}_2(1-\widehat{\pi}_2)\pi_1^2 - 5.991\frac{\widehat{\pi}_1\widehat{\pi}_2(1-\widehat{\pi}_1)(1-\widehat{\pi}_2) - (\widehat{\pi}_{11} - \widehat{\pi}_1\widehat{\pi}_2)^2}{n_1 + n_2\widehat{\eta}}\} \end{aligned}$$

Solutions of $\Delta(\pi_1) = 0$ are a_L and a_R, therefore

$$\text{Area of a Wald-type CR} = \int_{a_L}^{a_R} [F(\pi_1) - G(\pi_1)] d\pi_1 = \frac{1}{\widehat{\pi}_1(1-\widehat{\pi}_1)} \int_{a_L}^{a_R} \sqrt{\Delta(\pi_1)} d\pi_1.$$

When the approximate 95% Wald-type CR is a complete ellipse, locate the two intersection points of the two curves $(a_L, F(a_L))$ and $(a_R, F(a_R))$ and then use the above formula to calculate the area of the ellipse. This is straightforward.

However, when the Wald-type CR is close to the boundaries, the confidence region may be an incomplete ellipse, such as one of the patterns in Figure 4.2. In such cases, we need to identify the pattern first, locate the points intersecting with the x-axis and y-axis, and then use the above formula accordingly to calculate the area of the incomplete ellipse. Then the expected area (EA) of Wald-type CRs in one simulation with 1000 simulated datasets is the average of the areas of all Wald-type CRs, each CR from one simulated dataset.

B. Calculation of the expected area of Score-type and LR-type CRs.

For the calculation of the area of a Score-type (or LR-type) confidence region, employ the strategy of grid searching as mentioned previously , and then use the proportion of the total grid points within the Score-type (or LR-type) confidence region as an approximation to the area of the Score-type (or LR-type) CR, that is,

$$\text{Area of a Score-type (or LR-type) CR} \approx \frac{\text{number of grids satisfying } T \leq \chi^2_{2,1-\alpha}}{\text{total number of grid points}}.$$

Thus the expected area (EA) of Score-type (or LR-type) CRs is just the average of areas of Score-type (or LR-type) CRs, with each CR derived from one simulated dataset.

C. Comparison of the expected area among Wald-, Score-, and LR-type CRs

The following Figure 4.8 is a graphical comparison among the three types of confidence regions in two different scenarios under the same design: an early stop trial($M = 1$) and a trial continuing to the second stage ($M = 2$). A numerical comparison of the expected areas (EA) among Wald-, Score-, LR-type confidence regions is shown in Table 4.2. The corresponding coverage probability (CP) is shown in parentheses. We only chose some optimal designs with relatively good coverage probabilities to do EA comparison.

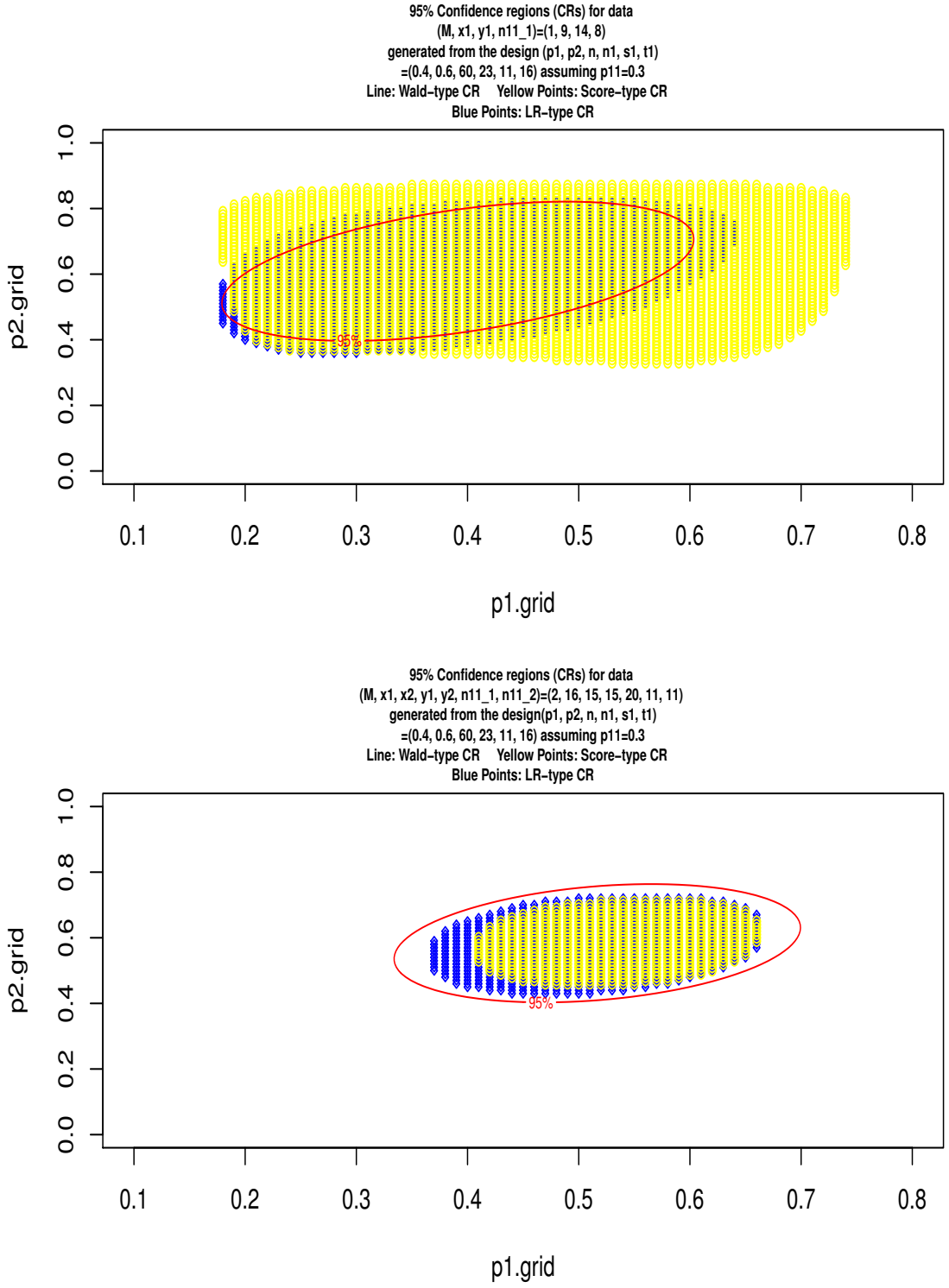


Figure 4.8: Comparison of 95% Score-type(yellow points), LR-type(blue points), and Wald-type (red line) Confidence Regions for (π_1, π_2) based on different simulated datasets, which were generated from the same design parameters

Table 4.2: Expected area comparison between Wald-type, Score-type and LR-type CRs (samsize=1000)

(n, n_1, s_1, t_1, s, t)	(p_1, p_2)	π_{11}	proportion of early-stop random datasets (M=1)	wald EA (CP)	score EA (CP)	LR-type CRs	
						EA (CP)	number of valid datasets to derive EA of LR-type CRs
(51, 24, 3, 3, 9, 9)	(0.25, 0.10)	0.08	0.106	0.046 (0.886)	0.051 (0.892)	0.048 (0.977)	600
(51, 24, 3, 3, 9, 9)	(0.25, 0.25)	0.15	0.027	0.066 (0.910)	0.064 (0.949)	0.061 (0.947)	976
(87, 37, 9, 13, 25, 33)	(0.20, 0.30)	0.08	0.672	0.058 (0.907)	0.126 (0.918)	0.069 (0.957)	962
(87, 37, 9, 13, 25, 33)	(0.35, 0.45)	0.18	0.022	0.051 (0.925)	0.051 (0.954)	0.051 (0.942)	999
(55, 25, 9, 12, 23, 28)	(0.30, 0.40)	0.14	0.705	0.111 (0.926)	0.194 (0.920)	0.127 (0.964)	982
(55, 25, 9, 12, 23, 28)	(0.50, 0.60)	0.33	0.023	0.083 (0.918)	0.080 (0.945)	0.082 (0.938)	1000
(60, 23, 11, 16, 30, 43)	(0.40, 0.60)	0.30	0.740	0.127 (0.926)	0.222 (0.909)	0.145 (0.957)	926
(60, 23, 11, 16, 30, 43)	(0.60, 0.80)	0.50	0.034	0.062 (0.910)	0.063 (0.951)	0.063 (0.941)	992

Observations in Figure 4.8 and Table 4.2

From Figure 4.8, it seems that when $M = 2$, the Score-type CR is the smallest, and that when $M = 1$ (the trial stops early at stage 1), the Wald-type CR is the smallest. Based on Table 4.2, when the proportion of the total randomly generated datasets with early stopping (i.e. $M = 1$) is high, the expected area of the Score-type CR will become much larger. This may be explained part by Figure 4.8. The expected areas between Wald-type and LR-type CRs are comparable in general.

In summary, when taking both coverage probability and expected area into account, the Likelihood ratio-type CR performs best: with good coverage probability and comparable expected area. The Likelihood ratio-type CR is a compromise between the Wald-type and Score-type CRs. In general, the Likelihood ratio-type CR is recommended.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

This dissertation has developed a two-stage optimal design for a single-arm phase II cancer clinical trial with two alternative binary primary efficacy endpoints under a variety of parameter settings. Since the two alternative primary efficacy endpoints within a patient are correlated, the inclusion of the nuisance correlation parameter has made the joint distribution and the power function more complicated for the study design in terms of controlling both Type I and II error constraints of the study. Sill et al. (2012) has mentioned the necessity of considering the case with two alternative primary efficacy endpoints although they used different terminology. They only considered three relatively extreme cases for the correlation parameter: independent, partially and fully dependent. This dissertation, however, has considered all possible values of this correlation parameter since we may not have much information about this nuisance correlation parameter at the design stage and we want to be conservative.

This thorough consideration on the correlation parameter has made whole searching process for optimal designs very computation-intensive and the time cost is very expensive. Due to the time cost of the searching process, the resulting designs we got may not be the globally optimal ones though it is sub-optimal. The resulting designs show that the correlation parameter may assume different values to achieve the maximized type I error rate, the minimized powers, the minimum of the maximized value among all possible expected sample sizes under H_0.

The optimal two-stage designs and the corresponding operating characteristics listed in Tables 3.6 to 3.9 can be referenced when planning a phase II cancer clinical trial with two binary alternative primary efficacy endpoints.

Although the primary goal of phase II cancer trials is testing whether the new treatment has sufficient anti-tumor activity for further development, obtaining estimates of the true event rates for the two binary alternative primary efficacy endpoints is also of interest, especially when the trial is deemed successful to go to Phase III. In this dissertation, we derived the most intuitive and commonly used point estimators, maximum likelihood estimators (MLE), for the true event rates for the two binary alternative primary efficacy endpoints. We also proposed three types of confidence regions for (π_1, π_2) based on the inverses of three types of likelihood based test statistics — Wald, Score and Likelihood Ratio statistics. The Wald-type confidence region for (π_1, π_2) is of a closed form, while Score-type and Likelihood ratio-type CRs are located using grid searching method.

The performance of the three types of confidence regions is compared using simulation data and the evaluation criteria include coverage probability and expected area. The coverage probability of the Wald-type confidence region is below the nominal confidence level for this two-endpoint two-stage design, which is similar to the phenomenon observed by many other researchers. The coverage probability of Score-type confidence region fluctuates: in some cases it is much better than that of Wald-type confidence region; while in some other cases, it is really small. The coverage probability of the likelihood ratio-type confidence region is good in general.

The expected areas between the Wald-type and the likelihood ratio-type confidence regions are comparable in general. If it is more likely for the trial to stop early, the expected area of the Wald-type confidence region is the smallest among the three, and the expected area of the Score-type confidence region is the largest (actually dramatically large). If it is more likely for the trial to continue to the second stage, the expected areas of the three types of confidence regions are similar.

All the three types of confidence regions are derived based on asymptotic distributions. It seems that the Score-type confidence region is more conservative and more sensitive to small sample size in Phase II cancer clinical trials.

In conclusion, when taking both coverage probability and expected area into account, the likelihood ratio-type confidence region performs best among the three: with good coverage probability and comparable expected area. The likelihood ratio-type confidence region

is a compromise between the Wald-type and Score-type confidence regions. In general, the likelihood ratio-type confidence region is recommended for this two-endpoint two-stage design.

5.2 Future Work

In addition to the optimal two-stage design proposed in this dissertation, minimax two-stage design can also be considered for the same research question by changing one of the optimality criteria of minimizing $E(N|H_0)$ into minimizing the maximum sample size. In some cases, the "minimax" design may be more attractive than the optimal design with minimum expected sample size. This will be the case when the difference in expected sample sizes is small and the patient accrual rate is low (Simon, 1989).

The sample size for stage 2 in the currently proposed design is pre-specified. An adaptive design, which allows the sample size of the second stage to depend on the results from the first stage $n_2(R)$, where R is the number of responses in the first stage, can be further considered for the same research question with two alternative primary efficacy endpoints. Banerjee and Tsiatis (2006) is a good reference for this.

The maximum likelihood estimators(MLE) are usually the first type of point estimators for consideration. There is no best estimator and each type of estimator has its own advantages and disadvantages. The MLE for (π_1, π_2) in the two-stage design here is negatively biased. Other point estimators such as bias reduced estimators and conditional MLEs can be explored and compared. Conditional MLEs are of special interest because conditional estimator may reduce bias and variance given

$$Var(X) = Var(E(X|Y)) + E(Var(X|Y)).$$

Tsai et al. (2008) can be a reference and a starting point.

The proposed confidence regions are now based on asymptotic distributions, and the exact confidence regions based on Clopper-Pearson method and Sterne method can be further considered.

The proposed design and inference methods can be generalized from currently two endpoints to more than two endpoints.

Bibliography

Richard Simon. Optimal two-stage designs for phase ii clinical trials. *Controlled Clinical Trials*, (10):1–10, 1989.

Edmund Gehan. The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *Journal of Chronic Diseases*, 13 (4):346–353, 1961.

Thomas R. Fleming. One-sample multiple testing procedure for phase ii clinical trials. *Biometrics*, 38(1):143–151, 1982.

TT Chen. Optimal three-stage designs for phase ii cancer clinical trials. *Statistics in Medicine*, 16(23):2701–2711, 1997.

SH Jung, T Lee, KM Kim, and SL George. Admissible two stage designs for phase ii cancer clinical trials. *Statistics in Medicine*, 23(4):561–569, 2004.

Yong Lin and Weichung J. Shih. Adaptive two-stage designs for single-arm phase iia cancer clinical trials. *Biometrics*, 60(2):482–490, 2004.

Anindita Banerjee and Anastasios A. Tsiatis. Adaptive two-stage designs in phase ii clinical trials. *Statistics in Medicine*, 25(19):3382–3395, 2006.

Stefan Englert and Meinhard Kieser. Improving the flexibility and efficiency of phase ii designs for oncology trials. *Biometrics*, 68:886–892, 2012.

John Bryant and Roger Day. Incorporating toxicity considerations into the design of two-stage phase ii clinical trials. *Biometrics*, 51(4):1372–1383, 1995.

Peter F. Thall and Su-Chun Cheng. Optimal two-stage designs for clinical trials based on safety and efficacy. *Statistics in Medicine*, 20(7):1023–1032, 2001.

Mark R. Conaway and Gina R. Petroni. Bivariate sequential designs for phase-ii trials. *Biometrics*, 51(2):656–664, 1995.

Walter Offen, Christy Chuang-Stein, Alex Dmitrienko, Gary Littman, Jeff Maca, Laura Meyerson, Robb Muirhead, Paul Stryszak, Alex Boddy, Kun Chen, Kati Copley-Merriman, Willard Dere, Sam Givens, David Hall, David Henry, Joseph D. Jackson, Alok Krishen, Thomas Liu, Steve Ryder, A. J. Sankoh, Julia Wang, and Chyon-Hwa Yeh. Multiple co-primary endpoints: Medical and statistical solutions - a report from the multiple endpoints expert team of the pharmaceutical research and manufacturers of america. *Drug Information Journal*, 41(1):31–46, 2007.

Takashi Sozu, Tomoyuki Sugimoto, Toshimitsu Hamasaki, and Scott R. Evans. *Sample size determination in clinical trials with multiple endpoints*. Springer, 2015.

A. J. Sankoh, M. F. Huque, and S. D. Dubey. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Statistics in Medicine*, 16(22):2529–2542, 1997.

L. A. Renfro and D. J. Sargent. Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocols: a review and examples. *Annals of Oncology*, 28:34–43, 2017.

S. J. Mandrekar, Yingwei Qi, S. L. Hillman, KL Allen Ziegler, NF Reuter, KM Rowland, SA Kuross, RS Marks, SE Schild, and AA Adjei. Endpoints in phase ii trials for advanced non-small cell lung cancer. *Journal of Thoracic Oncology*, 5(1):3–9, 2010.

FA Shepherd, PJ Rodrigues, T Ciuleanu, and et al. Erlotinib in previously treated non-small-cell lung cancer. *N Engl J Med*, 353:123–32, 2005.

Josep M. Llovet, et al., and the SHARP Investigators Study Group. Sorafenib in advanced hepatocellular carcinoma. *New England Journal of Medicine*, 359(4):378–90, 2008.

Raphael Porcher and Kristell Desseaux. What inference for two-stage phase ii trials? *BMC Medical Research Methodology*, 12:117, 2012.

Sin-Ho Jung and Kyung Mann Kim. On the estimation of the binomial probability in multistage clinical trials. *Statistics in Medicine*, 23:881–896, 2004.

Wei-Yann Tsai, Yunchan Chi, and Chia-Min Chen. Interval estimation of binomial proportion in clinical trials with a two-stage design. *Statistics in Medicine*, 27:15–35, 2008.

Michael W Sill, Larry Rubinstein, Samuel Litwin, and Greg Yothers. A method for utilizing co-primary efficacy outcome measures to screen regimens for activity in two-stage phase ii clinical trials. *Clinical Trials*, 9(4):385–395, 2012.