

© 2010

Tzu-Lin Hsu

ALL RIGHTS RESERVED

**COMPARATIVE POISSON TRIALS FOR COMPARING MULTIPLE NEW
TREATMENTS TO THE CONTROL**

By

TZU-LIN HSU

A Dissertation submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Statistics

written under the direction of

Donald R. Hoover

and approved by

New Brunswick, New Jersey

October, 2010

ABSTRACT OF THE DISSERTATION
COMPARATIVE POISSON TRIALS FOR COMPARING MULTIPLE NEW
TREATMENTS TO CONTROL

By TZU-LIN HSU

Dissertation Director:
Donald R. Hoover

Comparative Poisson Trials often test interventions to prevent rare adverse binomial outcomes. We extend Gail's "Design A" approach to continues the trial until a predetermined total number of disease cases, D , occur into comparing $K > 1$, treatments to one control. Controlling overall type I error and a post-hoc procedures to identify which treatments are better are addressed.

With the Poisson as the underlying distribution, conditioning on D disease cases total, the number in each group is multinomial distributed with parameters that depend on the incidence ratios of treatment to the control arms. Rejection regions based on the 1) numbers of cases that occur in control and/or 2) minimum number of cases among treatment groups are considered to test the global null hypothesis that no treatment is superior to the control. A tool known as the stochastic matrix simplifies size and power computations. Decision rules which are robust to some treatments being inferior to the control are discussed. There is no uniformly most powerful test against all alternatives, but rejection regions should have the Lower Left Quadrant Rule property. The discreteness of multinomial complicates derivation of theoretical results. Still, some identities are proven for comparing $K=2$ treatments to the control that we believe will extend to $K \geq 3$.

For $K=2$, the post-hoc procedure that applies standard binomial tests to each individual treatment vs. control hypothesis when the global hypothesis is rejected is superior to the Bonferroni adjustment; reducing by 7 % to 18 % the follow up disease cases required for the range of settings we studied. We

considered unbalanced allocation of follow up time to treatment and control groups. While discreteness of the multinomial distribution prevents analytic solution, a systematic point by point search that computes powers for a range of treatment / control allocation ratios with small increments is applied to find the optimum allocation ratio. In most cases the optimum allocation ratios do not perform substantially better than equal allocation in terms of minimization of the D or expected subject time needed to obtain D for given Type-1 error or power.

Acknowledgements

I am heartily express my gratitude to my dissertation advisor, Professor Donald R. Hoover, whose invaluable and constant encouragement, guidance and support from the beginning to the final enabled me to complete of this dissertation. I also extend my thanks to the graduate director, Dr. John Kolassa, for his advice on study, research, and his support since I entered the program. I also want to thanks to my dissertation committee members - Dr. Minge Xie, Dr. Ying Hung, Dr. Lee Dicker, and Dr. Xiaonan Xue for their time to review my thesis and their advice.

I am grateful to all the faculty members in the department of Statistics for their wonderful lectures and to the other students in the department whom have helped me enjoy learning and doing research in Rutgers University. My special thanks go to Women's Interagency HIV Study, U01-A1-34993, especially to Dr. Kaythyrn Anastos for their support and motivations in this study.

Finally, my deepest gratitude goes to my family, grandmother, aunts, and my parents, for endless love, support, and encouragements throughout these years.

Dedication

This dissertation is dedicated to my grandmother and my parents.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iv
Dedication.....	v
List of Tables.....	ix
List of Figures.....	x
List of Illustrations.....	xi
Notations.....	xiv
Chapter 1 INTRODUCTION AND OBJECTIVES.....	1
1.1 Motivation-Comparative Poisson Trials with Multiple New Treatments and One Control.....	1
1.2 Objectives.....	2
Chapter 2 BACKGROUND.....	3
2.1 Background on Comparative Poisson Trials.....	3
2.2 Background on Comparisons of Multiple Treatments to One Control.....	5
2.3 Summary and Goals of this Research.....	6
Chapter 3 METHODS TO TEST IF ANY TREATMENT IS BETTER THAN THE CONTROL.....	7
3.1 Multiple Comparative Poisson Trial Design and Hypothesis	7
3.1.1 Underlying Distribution for the Comparative Trial Design.....	7
3.1.2 Global Hypothesis Testing for the Comparative Poisson Design.....	8
3.2 Test Statistics for the Global Hypothesis Testing.....	10
3.3 Probabilities for Each Combination of Test Statistics Under Different Hypotheses.....	11
3.3.1 Representation of Table with the Probabilities Under Different Hypotheses.....	12
3.3.2 Computing Exact Joint Distribution for the Minimum of a Multinomial Distribution for $K \geq 2$	15
3.4 Criteria to Obtain the Rejection Region.....	18
3.4.1 Use the Control.....	18
3.4.2 Use the Minimal New Treatment.....	20
3.4.3 Difference Between Control And The Minimal Treatment.....	22

3.4.4 Cumulative Binomial For Control And The Minimal Treatment.....	25
3.4.5 Optimal Power/Size Ratio Based on a Specified Alternative.....	28
3.5 Accumulated Rejection Region Based on LLQR.....	33
3.6 Robustness of the Different Rejection Regions to Pathological Null and Pathological Alternatives.....	37
3.6.1 Robustness of Rejection Regions to False Positive Findings in the Pathological Null Setting..	37
3.6.2 Robustness of the Rejection Region to Pathological Alternatives.....	41
3.7 Summary.....	45
Chapter 4 METHODS TO TEST WHICH TREATMENT(S) IS (ARE) BETTER THAN THE CONTROL.....	47
4.1 Fisher’s L.S.D Adapted Approach (FLSDA).....	47
4.2 Comparison of FLSDA to Bonferroni Adjustment.....	49
4.3 Summary.....	51
Chapter 5 SAMPLE SIZE CALCULATIONS; EXPECTED WAITING TIME; WITH AN EXAMPLE FOR $K=2$ AND EQUAL ALLOCATION DESIGN.....	53
5.1 Disease Cases D Needed for Given Power and Size for Finding New Treatment(s) is (Are) Better Than the Control Using FLSDA-CB.....	53
5.2 Expected Subject Time/Subject Years N Followed up for Given Disease Cases D	54
5.3 Example on Prevention of HIV Transmission.....	54
Chapter 6 OPTIMIZING ALLOCATION RATIO OF CONTROL TO TREATMENT.....	58
6.1 Optimum Allocation Ratio that Minimizing the Follow Up Cases Required.....	59
6.2 Optimum Allocation Ratio that Minimizing the Expected Follow Up Subjects / Time.....	62
6.3 Results: Optimum Allocation Ratio For Different Alternatives.....	64
6.3.1 Optimum Allocation Ratio that Minimizing the Follow Up Disease Cases for Different Alternatives.....	64
6.3.2 Optimum Allocation Ratio that Minimizing the Follow Up Subjects / Time for Different Alternatives.....	66
Chapter 7 CONCLUSIONS AND FUTURE WORK.....	67

Appendix A UNBIASED PROPERTY FOR REJECTION REGION SATISFYING LLQR FOR $K=2$	
.....	69
A.1 Unbiased Properties for the Rejection Region Using “Minimal Treatment Only”	69
A.2 Unbiased Properties for Any Rejection Regions Satisfying LLQR	74
A.3 Unbiased Properties for Rejection Region Based On LLQR For Unequally Allocation For $K=2$	
.....	76
Appendix B: Proof that the Cumulative Binomial Metric under Equal Allocation is LLQR	77
Appendix C: Power Computations under the alternative $(1;r,1)$	79
C.1 Rejection Region for FLSDA for Given Total Disease Cases D	80
C.2 Rejection Region for Bonferroni Adjustment for Given Total Disease Cases D	83
C.3 Power Computation after the Rejection Region Is Determined	84
Appendix D Introduction to Design C	86
Appendix E List of R Functions Implemented In This Thesis	88
References Cited	90
Curriculum Vita	91

Lists of Tables

Table 3.6.1 Probability to reject the pathological null (1:1,2) using the $\alpha=0.05$ level on a rejection region based on regular null hypothesis (1:1,1).....	39
Table 3.6.2 Size based on the rejection region using “minimal treatment only” for its corresponding follow up cases and the pathological null under nominal level α equals to 0.01 and 0.025.....	40
Table 3.6.3 Probabilities to be reject under the hypothesis of (1: $r_1 = 0.1$, $R = 2$) and under (1: $r_1 = 0.1, 1$) based on different types of rejection region for $D=15$ and $D=30$ for size level $\alpha = 0.05$	44
Table 3.6.4 Probability to reject under the alternative (1: $r = 0.1, 1$) and two pathological alternatives (1: $r_1 = 0.1, R = 2$) and (1: $r_1 = 0.1, R = 5$) based on the rejection region using “control only” and “control boundary augmented” for size level 0.05 and follow up cases 15, 30, 60, 90, or 120.....	44
Table 4.2.1 Sample size reduction from FLSDA-CB to Bonferroni adjustment on different size levels $\alpha = 0.025$ and 0.05 under the setting (1: $r_1 = r, 1$) for $r = 0.1, 0.2$, or 0.5 for at least 80% power.....	51
Table 5.1.1 Minimum numbers of follow up disease cases required for given power 80%, 85% or 90% and size level 0.025 or 0.05 with the alternative setting (1: $r_1 = r, 1$), where $r=0.1, 0.2, 0.3, 0.4, 0.5$, or 0.6.....	53
Table 6.3.1 Smoothed optimum allocation ratio to minimize number of cases required for 80% power under size level $\alpha = 0.05$ and 0.025 as alternative is $H_a: (\rho: r_1 = r, 1)$ (where $r = 0.1, 0.2, 0.3, 0.4$, and 0.5) based on LOWESS with parameter $f = 0.2$ or 0.3 and smoothing spline with degrees of freedom $df = 5$	65
Table 6.3.2 Smoothed optimum allocation ratio to shorten expected waiting time for D^o cases to accrue of $\alpha = 0.05$ and 0.025 with power = 80% under $H_a: (\rho: r_1 = r, 1)$ is true (where $r = 0.1, 0.2, 0.3, 0.4$, and 0.5) based on exact search, LOWESS with parameter $f = 0.2$ or 0.3, and smoothing spline with degrees of freedom $df = 5$	66

Lists of Figures

- Figure 4.2.1** Power Comparison based on different $\alpha = 0.05, 0.025$ levels under the setting (1: $r_1=0.1,1$), (1: $r_1=0.2,1$) and (1: $r_1=0.5,1$) under equal allocation.....50
- Figure 6.1.1** Minimum follow up number of disease cases required for size 0.05 with 80% power under null hypothesis when alternative hypothesis ($\rho: r = 0.1,1$). Each point is (ρ, D^ρ) where D^ρ is the number of follow up cases required to obtain 80% power for the setting $\alpha = 0.05, r = 0.1$, and the its corresponding ρ 61
- Figure 6.1.2** Minimum follow up disease cases required for size 0.05 with 80% power under null hypothesis with an alternative hypothesis ($\rho: r_1 = 0.1, 1$) and smoothing spline curves. Each point is (ρ, D^ρ) where D^ρ is the numbers of follow up cases required to obtain 80% power for the setting $\alpha = 0.05, r = 0.1$, and the its corresponding ρ . The smoothing curve is obtained through locally weighted regression by function LOWESS in R software with smoothing parameter $f=0.2$ (red) and 0.3 (green) and through smoothing spline by function `smooth.spline` in software R with degrees of freedom $df = 5$ (blue).....62
- Figure 6.2.1** Expected waiting time for $\alpha = 0.05$ with 80% power under alternative hypothesis for ($\rho: r = 0.1, 1$). Each point is (ρ, D^ρ) where D^ρ is the numbers of follow up cases required to obtain 80% power for the setting $\alpha = 0.05, r = 0.1$, and the its corresponding ρ . The smoothing curve is obtained through locally weighted regression by function LOWESS in R software with smoothing parameter $f=0.2$ (red) or 0.3 (green) or through smoothing spline by function `smooth.spline` in software R with degrees of freedom $df = 5$ (blue).....64

List of illustrations

Illustration 3.2.1 LOWER LEFT QUADRANT RULE (LLQR) if the combination ($D_C = d_C, D_{(1)} = d_{(1)}$) is rejected.....	11
Illustration 3.2.2 One possible rejection region based on LLQR.....	11
Illustration 3.3.1 Possible cells (above the bolded line) to define the rejection region with impossible cells which is being beyond the bolded line being removed; Probabilities of ($D_C, D_{(1)}$) could be filled in under null or specified alternatives.....	12
Illustration 3.3.2 Probabilities of possible combinations $\{D_C, D_{(1)}\}$ under null hypothesis (1:1,1) for equal allocation.....	13
Illustration 3.3.3 Probabilities of possible combinations $\{D_C, D_{(1)}\}$ under one possible alternative hypothesis (1: $r_1=0.2, 1$) for equal allocation.....	13
Illustration 3.3.4 Probabilities of possible combinations $\{D_C, D_{(1)}\}$ under other one possible alternative hypothesis (1: $r_1=0.2, r_2=0.2$) for equal allocation.....	14
Illustration 3.3.5 Probabilities of possible combinations $\{D_C, D_{(1)}\}$ under null hypothesis ($\rho = 1.5: 1, 1$) for unequal allocation ratio $\rho = 1.5$	14
Illustration 3.3.6 Possible probabilities with marginal probabilities added in under null hypothesis (1:1,1) for equal allocation.....	15
Illustration 3.4.1 Rejection Region through “Controls Only” or “Controls Boundary Augmented” for $K = 2, D = 10$, and $\alpha = 0.05$ for equal allocation.....	19
Illustration 3.4.2 Rejection Region through “Controls Only” or “Controls Boundary Augmented” for $K = 2, D = 10$, and $\alpha = 0.05$ for unequal allocation with $\rho = 1.5$	20
Illustration 3.4.3 Rejection Region through “Minimal New Treatment Only” or “Minimal New Treatment Boundary Augmented” for $K = 2, D = 10$, and $\alpha = 0.05$ for equal allocation.....	21
Illustration 3.4.4 Rejection Region through “Minimal New Treatment Boundary Augmented” for $K = 2, D = 10$, and $\alpha = 0.05$ for unequal allocation with $\rho = 1.5$	22
Illustration 3.4.5 Rejection Region through “Difference Only” or “Difference Boundary Augmented” for $K = 2, D = 10$, and $\alpha = 0.05$ for equal allocation.....	23
Illustration 3.4.6 Rejection Region through “Difference Only” or “Difference Boundary Augmented” for $K = 2, D = 10$, and $\alpha = 0.05$ for unequal allocation design as $\rho = 1.5$. The first row is the modified difference while the second row is the probability of the combination under null hypothesis.....	24
Illustration 3.4.7 Rejection Region through “Cumulative Binomial” (shaded) for $K = 2, D = 10$, and $\alpha = 0.05$. For each ($D_C = c, D_{(1)} = t$), the numbers followed by “CB” are the cumulative binomial calculated by $P(D_{(1)} \leq t D_C + D_{(1)} = c + t)$ where $D_{(1)} D_C + D_{(1)} = c + t \sim \text{Bin}(D_C + D_{(1)} = c + t, 1/2)$ and the number followed by “size” is the probability of ($D_C = c, D_{(1)} = t$) under the null hypothesis discussed in section 3.3. For example, In the cell ($D_C = 9, D_{(1)} = 0$), $\text{CB} = P(D_{(1)} \leq 0 D_C + D_{(1)} = 9) = 0.00195$ and size = 0.0003 is by multinomial distribution from section 3.3 under null hypothesis for equal allocation.....	26
Illustration 3.4.8 Rejection Region through “Cumulative Binomial” (shaded) when $K = 2, D = 10$, and $\alpha = 0.05$ for unequal allocation ratio $\rho = 1.5$. For each ($D_C = c, D_{(1)} = t$), the numbers followed by “CB” are the cumulative binomial calculated by $P(D_{(1)} \leq t D_C + D_{(1)} = c + t)$ where $D_{(1)} D_C + D_{(1)} = c + t \sim \text{Bin}(D_C + D_{(1)} = c + t, 1/2.5)$ and the number followed by “size” is the probability of ($D_C = c, D_{(1)} = t$) under null	

hypothesis discussed in section 3.3. For example, In the cell ($D_C = 9, D_{(1)} = 0$), $CB = P(D_{(1)} \leq 0 \mid D_C + D_{(1)} = 9) = 0.01008$ and size = 0.0028 is by multinomial distribution from section 3.3 under null hypothesis for unequal allocation.....27


Illustration 3.4.9 Rejection Region through “PowRatio” (shaded ) for $K = 2, D = 10, \alpha = 0.05$ and under specified alternative ($1:r_1 = 0.2, 1$). The power, size, and its ratio are calculated and located as top, middle, and third in each cell.....31


Illustration 3.4.10 Rejection Region through “PowRatio” (shaded ) for $K = 2, D = 10, \alpha = 0.05$ and under specified alternative ($1:r_1 = 0.2, r_2 = 0.2$). The power, size, and its ratio are calculated and located as top, middle, and third in each cell.....32


Illustration 3.4.11 Rejection Region through “PowRatio” (shaded ) for $K = 2, D = 10, \alpha = 0.05$ and under specified alternative ($1:r_1 = 0.2, r_2 = 0.5$). The power, size, and its ratio are calculated and located as top, middle, and third in each cell.....33

Illustration 3.5.1 Possible Candidate Cells Given Current Rejection Region; gray colored cells are current rejection region, and the green colored cells are the candidates to be accumulated to the new rejection region.....35

Illustration 3.5.2 Candidate cells (bolded) for the next rejection region given current rejection region (gray colored).....36

Illustration 3.6.1 Probability mass for $D = 15$. The top probability P_1 in the cell is the probability mass under regular null hypothesis ($1:1,1$) while the second probability P_2 in the cell is the probability mass under pathological hypothesis ($1:1,2$). 60% of the probability mass for each hypothesis is marked as bolded (red colored is under regular null and the black colored is under pathological null).....38

Illustration 3.6.2 Probability mass for $D = 15$. The top probability P_1 in the cell is the probability mass under regular alternative hypothesis ($1: r_1 = 0.1, 1$) while the second probability P_2 in the cell is the probability mass under pathological alternative ($1: r_1 = 0.1, R = 2$). 60% of the probability mass for each hypothesis is marked as bolded (red colored is under regular alternative and the black colored is under pathological alternative).....42

Illustration 4.1.1 Rejection Region (shaded in gray) in the step one of FLSDA with $\alpha=0.05, K=2, D=15$...49

Illustration 5.3.1 Rejection Region (shaded) for global hypothesis based on conditional binomial, $R_{\alpha-CB}$, for $K=2, D=19, \alpha=0.05$56

Illustration 5.3.2 Rejection region for each individual hypothesis for $K=2, D=19, \alpha=0.05$ and $k=1,2$. As $D_C + D_k = 19$, we will reject the null of H_{0k} as $D_k < 6$ and conclude that the new treatment k is better than the control.....56

Illustration C-1 Probabilities of all possible combinations of (D_C, D_1) given $D = 10$ under the setting ($1: r_1 = 0.5, 1$).....79

Illustration C-2 Rejection region using $(D_C, D_{(1)}), R_{\alpha}^{(i)}$ (shaded) for $K=2, D=10$ and size $\alpha = 0.05$ for the global hypothesis, the cell ($D_C=6, D_{(1)}=0$) which used in example is marked in red and the cell ($D_C=6, D_{(1)}=1$) is marked in blue.....81

Illustration C-3 Rejection region using $(D_C, D_1), R_{\alpha}^1$ (shaded) for $K=2, D=10$, and size $\alpha = 0.05$ for the global hypothesis, corresponding to $R_{\alpha}^{(i)}$ in Illustration C-2. For example, the cells ($D_C = 6, D_1 = 0$) and ($D_C = 6, D_1 = 4$) marked in red are rejected in R_{α}^1 since ($D_C=6, D_{(1)}=0$) is rejected in $R_{\alpha}^{(i)}$ and the cells ($D_C = 6, D_1 = 1$) or ($D_C = 6, D_1 = 3$) marked in blue is not rejected in R_{α}^1 as ($D_C=6, D_{(1)}=1$) is not rejected in $R_{\alpha}^{(i)}$ 81

Illustration C-4 Conditional p-values for the individual hypothesis of Treatment 1 versus the control conditioning on d_C+d_1 ; filled in on the global rejection region $R_\alpha^{(1)}$ that rejects in the global hypothesis for the minimum treatment being less than the control for $D=10$, $K=2$, and $\alpha=0.05$. The bolded enclosed region covers the conditional p-values $\leq \alpha$83

Illustration C-5 Conditional p-values to reject New Treatment 1 being equal to the control conditioning on d_C+d_1 for $D=10$, $K=2$. The bolded region covers all conditional p-values ≤ 0.025 and is the rejection region based on Bonferroni adjustment for the case $D=10$, $K=2$, and $\alpha=0.05$84

Illustration C-6 Multinomial probability of combinations of (D_C, D_1) that are in the given rejection region for given $D=10$ under the setting $(1: r_1 = 0.01, 1)$85

Notation

Symbol	Description
K	Numbers of new treatment groups (fixed for the design, $K=k$)
D	Total number of disease cases (this is fixed for design A, $D=d$)
D_C	Numbers of disease cases in control group, the control group could be placebo or current treatment (random or fixed depends on the design; random for design A)
D_k	Numbers of disease cases in k th new treatment group for $k=1, \dots, K$ (random)
$D_{(1)}$	Minimum number of disease cases observed in a new treatment group among the new treatment groups (random)
D_T	Total numbers of disease cases in all new treatment groups = $D - D_C$
N	Total population (or follow up) size in control and new treatment groups (random for design A, fixed, $N=n$, for design B)
N_C	Population (or follow up) size in control group (random for design A)
N_T	Population (or follow up) size in each new treatment group (random)
$\rho_C = N_C/N$	Portion of N in control group (fixed for the design)
ρ_k	Portion of N in the k th new treatment group, ($\rho_1 = \rho_2 = \dots = \rho_K$ in the design)
$\rho_T = N_T/N$	Portion of N in each new treatment group, $\rho_1 = \rho_2 = \dots = \rho_K = \rho_T$ (fixed for the design) Note: $\rho_C + K\rho_T = 1$. For equal allocation among control and new treatment arms, $\rho_C = \rho_k = 1 / K+1$ for $k = 1, 2, \dots, K$.
$\rho = \rho_C / \rho_T$	Ratio of the portion of total population in the control group to the portion in each new treatment group (fixed for the design)
D^p	Minimum numbers of disease cases required for given size level and power for design A for unequal allocation ratio ρ
i_C	Incident rate per subject year/time in the control group (fixed, but typically unknown)
i_k	Incident rate per subject year/time in the k th new treatment group (fixed, but unknown) for $k = 1, 2, \dots, K$
$r_k = i_k / \rho i_C$	the true but unknown relative risk of the new treatment group k to the control group for $k=1, 2, \dots, K$ $0 \leq r_k$ and assume $r_k \leq 1$ unless otherwise stated
λ_C	Poisson intensity rate for number of events in control subjects for a given N_C , $\lambda_C = i_C N_C$
$\lambda_k = r_k \lambda_C$	Poisson intensity rate for all subjects in new treatment group k for a given N_T .
R_α	The rejection region defined on $(D_C, D_{(1)})$ for the global null hypothesis that all Treatment groups are equal to the control Group. This region is denoted to have an overall Type-1 Error of α when none of the treatment groups exceed the control

Chapter 1 INTRODUCTION AND OBJECTIVES

1.1 Motivation – Comparative Poisson Trials with Multiple New Treatments and One Control

Studies are often conducted to compare outcome incidence rates in different groups when these outcomes either have a Poisson distribution or are binomial with a very low probability that can be approximated by a Poisson distribution. This problem is well studied for one new treatment compared to one control and the Comparative Poisson method for statistical inference was proposed and designed by, among others, by Gail (1974), Brown and Green (1982).

However, there are many settings where multiple treatments are compared to one control. For example, when K normally distributed new treatment outcomes are each compared to one normally distributed control with all having the same unknown variance, Dunnet (1955) reduced the numerical dimensionality of the distribution of the test statistics from $K+2$ to 3 by an integration that conditions on the distribution of the control mean and the independence of the treatment means. While Dunnet's method is widely used and has been expanded to many related settings (Hochberg, Tamhane, 1987), no analogous approach has yet been proposed for comparative Poisson trials.

However, settings with more than one treatment group compared to 1 control group with rare binary outcomes exist. For example, a randomized HIV trial in Malawi, N.I. Kumwenda et. al. (2008) compared $K=2$ new treatments, i) extended 14 week nevirapine with 1 week zidovudine and ii) extended 14 week nevirapine and zidovudine to the control 1 week zidovudine / nevirapine and found that the two extended prophylaxis reduced breast-milk HIV transmission from mother to non-infected infant more than control group by using Kaplan-Meier analyses. For a multiple vaccine study in an experiment on animals (chickens), Allen, Danforth, and Vinyard (2004) compared the protective index of $K=5$ new vaccines to a control. The design was divided into four subgroups with each vaccine group to challenge with one of the three *Eimeria maxima* isolated strains or water (unchallenged control). The comparison outcome of that study was a continuous outcome protective index which could be derived through other measurements and is done in each subgroup against the challenge strains. However, in other settings, the outcome could be binary such as shedding virus yes/no and comparative Poisson methods might be applicable.

We consider here, in particular, a clinical trial setting where K different new treatments or prophylactic interventions to prevent an uncommon disease are each compared to one control intervention.

For example, K new experimental vaccines for influenza (or Human papillomavirus) are each given to a large number n of subjects and compared to a control vaccine which is also given to say the same number n subjects; or $(K+1)n$ total subjects vaccinated. The objective is to simultaneously compare these treatments to the control in a way that preserves the overall type one error at a given level α .

The test statistic will be based on the number of disease cases that occurs in each new treatment group D_k (for $k = 1, \dots, K$) and the number of disease cases in the control group D_C at least one of the new treatments k will be deemed better than the Control by a decision rule designed to maintain this type 1 error and have power to identify important differences when they exist first. We discuss procedures for statistically finding which new treatment(s) is/are better than the control .

1.2 Objectives

The purpose of this thesis is to propose and evaluate procedures to do exact comparison for comparative Poisson trials comparing K treatment groups to one control group while maintaining an specific overall Type-1 error to falsely find at least one treatment is better than the control when in truth none are superior to the control. We begin by assuming equal allocation of subject time or subject years into each treatment and control group; $N_T=N_C$. Chapter 2 provides more background information on comparative Poisson studies and multiple comparisons and further develops the study notation. Chapter 3 proposes a test statistic approach that maintains the overall Type-I error for falsely finding at least one of K new treatments to be better than the control to be less or equal to α for the global hypothesis testing with emphasis on $K=2$. Chapter 4 presents a post-hoc approach for finding which treatment is superior to the control while the global hypothesis is rejected. Chapter 5 provides sample size calculation to find superiority when in fact a treatment is better for the $K=2$ setting. Chapter 6 explores study design strategies for this procedure for the unequal allocation that might minimum the follow up cases required for given power for the purposed procedures. Chapter 7 concludes with a conclusion and suggests future work. The Appendix contains some proofs for properties and power computation illustration and R software package to implement this design is included.

Chapter 2 BACKGROUND

2.1 Background on Comparative Poisson Trials

In designing studies to comparing incidence of a rare disease, it is often appropriate to assume that cases of disease occurred follow a Poisson distribution. This includes settings where the disease has a binomial distribution $\text{Bin}(n, p)$ approximated to $\text{Poisson}(\lambda = np)$ with sample size n is sufficient large and the rate p is very small (Ross, 2000), with $\lambda = np$ fixed. Since the binomial probability of the number of disease cases d is

$$P(X = d) = \frac{n!}{(n-d)!d!} p^d (1-p)^{n-d} = \frac{n(n-1)\dots(n-d+1)}{n^d} \frac{\lambda^d}{d!} \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^d}$$

For n is large, $(1-\frac{\lambda}{n})^n \approx e^{-\lambda}$, $\frac{n(n-1)\dots(n-d+1)}{n^d} \approx 1$, $(1-\frac{\lambda}{n})^d \approx 1$, (Eq. 2.1.1)

Thus $P(X = d) \approx e^{-\lambda} \frac{\lambda^d}{d!}$, which is the distribution function of $\text{Poisson}(\lambda)$.

One important issue with comparative Poisson trials is when to stop the study for analysis. With Comparative Poisson Trials comparing one new treatment to the control, Gail proposed two approaches which he referred to as “Design A” and “Design B.” These designs readily extended to studies of $K = k$ new treatments vs 1 control. Design A has the study continue until a fixed number of disease cases, D , occurs in all groups. For example, continue the study until $D = 50$ cases of influenza occur in all groups and then make the comparison. Design B has the study continue until the predetermined number of subjects (or subject time) occurs. For example, with a vaccine study, continue the study until 50,000 subject years have been observed and then make the comparison.

Both of these designs can be problematic if the disease incidence is lower than expected. For example with Design A, if the incidence was only 0.0001 per untreated subject year and the follow up was planned until $D = 50$, this might require $N = 500,000$ person years (or longer if the treatment reduces disease) which could be too costly. But if Design B was used for this study with the strategy that stops after 50,000 person years, then only approximately $D = 5$ cases (or less if the treatment works) would have occurred giving very small power to compare. Sometimes “hybrid” designs are used. For example, a study could be planned for $D=50$ cases with a contingency that it would be stopped at 200,000 subject years to limit costs if 50

cases had not occurred yet. Since, as we will see, power of comparative Poisson tests depend on D , not on N , Design A which fixes D is used most often and we will focus on Design A on this design study.

For comparing two Poisson trial incident rates with fixed duration, Przyborowski and Wilenski (1940) presented to do the test conditional on the sum of the two variables $D_C + D_T = D$ where D_C is the number of cases in control and D_T is the number of cases in the single treatment group. Let two independent Poisson series be D_C and D_T with known rates λ_C and λ_T , the joint distribution of $D_C = d_C$ and $D_T = d_T$ can be expressed as Eq. 2.1.2.

$$P(D_C = d_C, D_T = d_T | \lambda_C, \lambda_T) = \frac{\lambda_C^{d_C} \lambda_T^{d_T} e^{-(\lambda_C + \lambda_T)}}{d_C! d_T!} \quad (\text{Eq. 2.1.2})$$

or if we let $d = d_T + d_C$

$$P(D_C = d_C, D_T = d_T | \frac{\lambda_C}{\lambda_C + \lambda_T}, \lambda_C + \lambda_T) = \frac{(\lambda_C + \lambda_T)^d e^{-(\lambda_C + \lambda_T)}}{d!} \frac{d!}{d_T! (d - d_T)!} \left(\frac{\lambda_T}{\lambda_C + \lambda_T}\right)^{d_T} \left(1 - \frac{\lambda_T}{\lambda_C + \lambda_T}\right)^{d - d_T} \quad (\text{Eq. 2.1.3})$$

The form of Eq. 2.1.3 shows that the joint distribution of D_C and D_T can be decomposed into two product terms:

i. $P(D = d | \lambda_C + \lambda_T) = \frac{(\lambda_C + \lambda_T)^d e^{-(\lambda_C + \lambda_T)}}{d!}$ where D is only depends on sum of the two rates $\lambda_C + \lambda_T$ and

thus contains no information on the ratio λ_T/λ_C .

ii. $P(D_T = d_T | D = d, \frac{\lambda_T}{\lambda_C + \lambda_T}) = \frac{d!}{d_T! (d - d_T)!} \left(\frac{\lambda_T}{\lambda_C + \lambda_T}\right)^{d_T} \left(1 - \frac{\lambda_T}{\lambda_C + \lambda_T}\right)^{d - d_T}$ which is Binomial with

parameters $n = d$ and $p = \frac{\lambda_T}{\lambda_C + \lambda_T}$ and thus can be used to test the ratio $\frac{\lambda_T}{\lambda_C + \lambda_T}$ or equivalently to test

λ_T/λ_C .

From ii above, the conditional distribution D_C and D_T of two Poisson distributions with rate λ_C and λ_T given

$D = d$ cases is Binomial with two parameters $n = d$ and $p = \frac{\lambda_T}{\lambda_C + \lambda_T}$. Through this approach, conditional

binomial rejection regions can define a UMPU (uniformly most powerful unbiased) test on whether the two rates, λ_C and λ_T are equal versus not equal (Lehmann, 2005).

2.2 Background on Comparisons of Multiple Treatments to One Control

For the normal outcome setting $X_1 \sim N(\mu_1, \sigma^2), \dots, X_k \sim N(\mu_k, \sigma^2), X_C \sim N(\mu_c, \sigma^2)$. Without loss of generality, assume that a smaller mean, μ , is better, then the general testing problem is $H_0: \mu_k \geq \mu_c$ for $k=1, \dots, K$ vs. $H_a: \mu_k < \mu_c$ for at least one $k=1, 2, \dots, K$. In other words, the control treatment will continue to be used unless at least one experimental treatment is found to be better.

This overall testing problem could be expressed as k separate hypothesis tests as follows:

$$\begin{aligned} H_{01}: \mu_1 - \mu_c \geq 0 \text{ vs } H_{a1}: \mu_1 - \mu_c < 0, \\ H_{02}: \mu_2 - \mu_c \geq 0 \text{ vs } H_{a2}: \mu_2 - \mu_c < 0, \\ \vdots \qquad \qquad \qquad \vdots \\ H_{0K}: \mu_K - \mu_c \geq 0 \text{ vs } H_{aK}: \mu_K - \mu_c < 0, \end{aligned} \tag{Eq. 2.1.4}$$

the original null hypothesis, $H_0: \mu_k \geq \mu_c$ for $k=1, \dots, K$, is the intersection of the each individual null hypothesis and the general alternative, $H_a: \mu_k < \mu_c$ for at least one $k = 1, 2, \dots, K$ with inequality holds, is the union of all the H_{ak} for $k=1, 2, \dots, K$.

Shaffer (1995) discussed the multiple ways to treat the multiple testing problems with tolerance of a particular error rate. One acceptable error rate is called family wise error rate (FWER) in strong control which is the probability of at least rejecting one null hypothesis H_{0k} wrong under any combination of true or false null hypothesis H_{0k} for $k = 1, 2, \dots, K$ is less than or equals to α . As the error rate only be controlled under all the hypotheses is true, i.e. probability to reject at least one hypothesis as all null hypotheses H_{01}, \dots, H_{0K} are true $\leq \alpha$, we called it weak control.

In the absence of a better approach for comparing each of K treatments to one control, the Bonferroni (1936) method which makes each comparison with a Type I error of α/K to falsely reject each of $H_{01}, H_{02}, \dots, H_{0K}$ will have an overall Type-I error of $\leq \alpha = \sum \alpha/K$ to falsely find any new treatment superior to the control when none are of $\leq \alpha$. However, this method is often too conservative and sacrifices power as the overall type 1 error is often substantially less than α . Holm (1979) proposed a step-down method that compares ordered p-values from the smallest to the adjusted size levels $(\alpha/K, \alpha/(K-1), \dots, \alpha)$. We will stop the procedure once the new treatment is not rejecting the null and conclude that all the corresponding rejected new treatments are better than the control. Holm's step-down procedure is not as conservative as Bonferroni and has higher power under normally distributed settings.

Dunnett(1955) developed an approach for comparing multiple treatments with one control group for normally distributed outcomes that had an overall Type I error of α and performed substantially better (in terms of size and power) than Bonferroni's method. Also, Paulson (1952) compared the K treatments to one control through comparing the best observation in the treatment groups to the control for normal variables and for Binomial variables using the inverse sine transformation approximation to normality. His procedure reduces K comparisons to 1 comparison,

$$H_0^* : \Delta^* = 0 \text{ versus } H_a^* : \Delta^* < 0 \text{ where } \Delta^* = \min\{ \mu_1 - \mu_C, \mu_2 - \mu_C, \dots, \mu_K - \mu_C \}.$$

2.3 Summary and Goals of this Research

In chapter 3, we extend Prozyborowski and Wilenski's conditional distribution of two Poisson random variables given sum of the random variables to derive a multinomial distribution the $K + 1$ setting in Section 3.1. In Section 3.2, we will reduce the dimensionality of the problem from $K+1$ down to 2 for using the test statistics $D_C, D_{(1)}$ which is much as Dunnet reduced the dimensionality for the continuous normal setting. Then in Section 3.3, we will extend Corrado's stochastic matrix approach to derive the exact distribution of $D_C, D_{(1)}$ for $K \geq 2$. In the rest of chapter 3, we will introduce the ways to construction the rejection regions for the global hypothesis testing problem and discuss the non-robustness of some rejection methods under some particular pathological conditions.

In chapter 4, we will purpose a test method for finding the new treatment which is better than the control once the global hypothesis is rejected which is similar to the concept of Fisher Least Significant Difference Test in section 4.1 and then in section 4.2, we compare the results to the test based on Bonferroni correction method in terms of power.

In chapter 5, a table of minimum sample size needed for different settings is provided in section 5.1. The expected follow up time which follows gamma distribution is discussed in section 5.2. The procedure is illustrated with an example in section 5.3.

In chapter 6, we discussed the strategy for the unequal allocation design that minimized total follow up disease cases for our procedure in section 6.1. Followed by section 6.2, we discussed the study strategy for unequal allocation design that shortened the expected follow up time. And then, we provide optimum allocation design for some possible settings in section 6.3.

Chapter 3 METHODS TO TEST IF ANY TREATMENT IS BETTER THAN THE CONTROL

3.1 Multiple Comparative Poisson Trial Design and Hypothesis

3.1.1 Underlying Distribution for the Comparative Trial Design

Consider a Comparative Poisson trial with one control and $K \geq 2$ new treatment groups. The extended Gail's Design A approach is used and the study continues until D cases are seen in all $K+1$ groups according to the following person allocation portions ρ_C into the control group and ρ_T into each experimental treatment group, with $\rho_C + K\rho_T = 1$. (Note that $\rho_C = \rho_T = \dots = \rho_K = 1/(K+1)$ for equal proportions of people recruited in each group.) A total of N subject-years-follow-up will need to be recruited to all K groups in order to obtain these D cases where N is random with $N_C = \rho_C N$ subject years seen in the control group and $N_T = \rho_T N$ subject years seen in each treatment group, also being random.

Let i_C and i_k be the true unit incident rates of disease per person follow up time in the control group and each respective new treatment group for $k=1, \dots, K$. Note, that in the extended Gail's Design B approach, N would be predetermined and the number of events in the control group is Poisson distributed with parameter $\lambda_C = i_C N_C$ and the number of events in each treatment group is Poisson distributed with parameter $\lambda_k = i_k N_T$ in k th treatment group for $k=1, \dots, K$.

For design A, D_C, D_1, \dots, D_K does not directly follow the unconditional Poisson setting when the total subject years or subject time of follow up is unknown for this Design. However, we will show that the conditional joint distribution of D_C, D_1, \dots, D_K given $D = d$ events does not depend on the follow-up-subject-years N . Let $r_k = i_k / \rho_T i_C$ be the ratio of per subject follow up time unit ratio for treatment group i_k to control group i_C . Since D_C and D_k are from independent Poisson with known rates $\lambda_C = i_C \rho_C N$ and $\lambda_k = i_k \rho_T N = r_k \lambda_C$ when we known the subject years N . We begin by assuming N is known (similar to Design B). By extension of (Eq. 2.1.3) to $K+1$ groups, for $D = D_C + \sum D_k$,

$$P(D_C, D_1, \dots, D_K | \lambda_C, \lambda_1, \dots, \lambda_K, N) =$$

$$\frac{(\lambda_C + \sum \lambda_k)^D e^{-(\lambda_C + \sum \lambda_k)}}{D!} \times \binom{D}{D_C \quad D_1 \quad \dots \quad D_K} \left(\frac{1}{1 + \sum r_k} \right)^{D_C} \prod \left(\frac{r_k}{1 + \sum r_k} \right)^{D_k}. \quad (\text{Eq. 3.1.1})$$

where the terms $1/(1 + \sum r_k)$ and $r_k/(1 + \sum r_k)$ do not depend on the follow up subject years N . So the conditional distribution of D_C, D_1, \dots, D_K given D is

$$\frac{(\lambda_c + \sum \lambda_k)^D e^{-(\lambda_c + \sum \lambda_k)}}{D!} \times \left(\begin{matrix} D \\ D_c \quad D_1 \quad \dots \quad D_K \end{matrix} \right) \left(\frac{1}{1 + \sum r_k} \right)^{D_c} \prod \left(\frac{r_k}{1 + \sum r_k} \right)^{D_k} = \frac{(\lambda_c + \sum \lambda_k)^D e^{-(\lambda_c + \sum \lambda_k)}}{D!}$$

$$\left(\begin{matrix} D \\ D_c \quad D_1 \quad \dots \quad D_K \end{matrix} \right) \left(\frac{1}{1 + \sum r_k} \right)^{D_c} \prod \left(\frac{r_k}{1 + \sum r_k} \right)^{D_k} \text{ or the standard multinomial with parameters } D, 1/(1 + \sum r_k)$$

and $r_k/(1 + \sum r_k)$, $k=1,2,\dots,K$.

Now, for Design A, we predetermined the total number of events D , so the first part of (Eq. 3.1.1) is no longer random. However, since for any given N ,

$$P(D_c, D_1, \dots, D_K | \lambda_c, \lambda_1, \dots, \lambda_K, N, D) = P(D_c, D_1, \dots, D_K | \lambda_c, \lambda_1, \dots, \lambda_K, N | D)$$

which as we showed was Multinomial distribution with parameters D , $1/(1 + \sum r_k)$ and $r_k/(1 + \sum r_k)$, for $k = 1, 2, \dots, K$, this distribution does not depend on the subject years N . Hence, we could treat N as a random variable without affecting the conditional distribution of D_c, D_1, \dots, D_K given D .

The comparison of the rates of D_c, D_1, \dots, D_K given a fixed D thus reduces to the comparison of the different portions in the multinomial distribution $1/(1 + \sum r_k)$ and $r_k/(1 + \sum r_k)$, $k=1,2,\dots,K$. Since the denominators of all terms are the same; $1 + \sum r_k$, these comparisons are most directly driven by the magnitude of differences in 1 and r_k , where $r_k = i_k / \rho_{i_c}$, for $k=1,\dots,K$. Through this setting, we also know that the conditional distribution of D given N person years is Poisson with a weighted averaged rate $= i_c \rho_{i_c} + \sum i_k \rho_{i_k}$ per unit subject years. Since D given N is a Poisson with waiting rate iN , the waiting time for one case to occur is exponential distributed with parameters $1/i$. If the trial ends after D cases occur (Design A) and the occurrence of each case is independent, the distribution of N at which the study ends for design A is Gamma with parameters D and $1/i$. Therefore, the expected duration for design A trial is the expectation of $\Gamma(D, 1/i)$ which is D/i . For example, if the incidence was 0.0001 per subject year in control subjects and none of the treatments changed this, and the design was follow up till $D=50$ cases. Then, the expected duration for this design is $N=D/i=50/0.0001=500,000$ persons would need to be recruited.

3.1.2 Global Hypothesis Testing for the Comparative Poisson Design

For comparative Poisson design problem, we are interested in seeing if the new treatment(s) is (are) more effective than the control. If the new treatment is more effective than the control, the incidence

with rate ρ for control group to each new treatment group, we use similar terminology and express $(\rho:1,\dots,1)$ under the null, $(\rho, r_1=r, \dots, 1)$, $(\rho: r_1=r, \dots, r_K=r)$ or $(\rho: r_1=c_1, \dots, r_K=c_K)$ under the specified alternative we mentioned above (i), (ii), and (iii).

There are too many possible combinations of the portion rate for the alternative to comprehensively cover. However, heuristically the conservative alternative is where only one treatment among the treatments works better than the control at a minimally specified level r that must be detected with a specified power $1-\beta$. We thus often focus on the setting with two new treatments with one control ($K=2$) of $H_0: (1:1,1)$ v.s. $H_a: (1:r_1=r,1)$.

3.2 Test Statistics for the Global Hypothesis Testing

We focus on comparing the best performing new treatments $D_{(1)}$ (or in the case of ties at $D_{(1)}$) to the controls. It is intuitive to assume that 1) the treatment(s) having the minimum disease cases ($D_{(1)}$) has the most statistical evidence to be the best of the new treatments and thus 2) comparing $D_{(1)}$ to the number of disease cases in the control group (D_C), most directly addresses whether the treatment that is most likely to be the best is better than the control, and 3) Controlling the type-I error for this comparison of the control to the treatments with the fewer cases as a first step to be $\leq \alpha$ makes the overall Type-I error to falsely reject any of the H_{0k} when all nulls are true to be $\leq \alpha$.

To illustrate this graphically, all possible combinations of D_C and $D_{(1)}$ could be represented as the table shown in Illustration 3.2.1 with rows representing D_C and columns representing $D_{(1)}$. Note that, mathematically, $D_{(1)}$ must be the integer that is less than or equal to D/K as $D_C = 0$. We define the rejection region R_α of D_C and $D_{(1)}$ as some collection of $\{D_C \text{ and } D_{(1)} \text{ pair}\}$ which cumulatively have less than or equal to α probability to obtain under the null hypothesis, but would be much more likely (i.e., $> \alpha$) to obtain under alternative hypotheses in which one or more treatments are better than the control. There are several ways to construct R_α and as we will show, there is no single R_α that is UMP (uniformly most powerful) for all possible alternatives simultaneously.

However, there are some important rules that can be used in the construction of optimal R_α . Most notably rule is referred to “**LOWER LEFT QUADRANT RULE**” (LLQR). Once, we decide to reject a given cell ($D_C = d_C, D_{(1)} = d_{(1)}$), then all other cells for which ($D_C \geq d_C, D_{(1)} \leq d_{(1)}$) from its left and down

should be rejected as well since there is more evidence to reject null hypothesis when D_C is larger or $D_{(1)}$ is smaller while the other one is fixed. This means that the left-lower quadrant of the table with $(D_C = d_C, D_{(1)} = d_{(1)})$ as the upper-right vertex should be included in R_α (hence the LLQR).

$D_C \backslash D_{(1)}$	0	1	$d_{(1)}$	$\lfloor D/K \rfloor$
0						
1						
...						
d_C				reject		
...						
...						
D						

Illustration 3.2.1 LOWER LEFT QUADRANT RULE (LLQR) if the combination $(D_C = d_C, D_{(1)} = d_{(1)})$ is rejected

The rejection region based on this rule will be like combination of rectangles and it could be one rectangle or stair down shape with non-constant step height and depth as following (Illustration 3.2.2).

$D_C \backslash D_{(1)}$	0	1	$\lfloor D/K \rfloor$
0						
1						
...		reject				
...				reject		
...						
...					reject	
D						

Illustration 3.2.2 One possible rejection region based on LLQR

3.3 Probabilities for Each Combination of Test Statistics Under Different Hypotheses

We are now going to look at strategies to identify a rejection region R_α which optimizes power for given size level α and fixed D . As a first step, an approach to calculate probabilities to be in the different cells of the $D_C, D_{(1)}$ table must be developed. To do this, we first note that for any combination $D_C, D_{(1)}$ that

$P\{D_C, D_{(1)} | D\} = P\{D_C | D\} \times P\{D_{(1)} | D_C, D\}$. Since the conditional distribution of D_C is $\text{Bin}(D, \rho/(\rho+K))$, i.e.

$$P\{D_C | D\} = \binom{D}{D_C} \left(\frac{\rho}{\rho+K}\right)^{D_C} \left(\frac{K}{\rho+K}\right)^{D-D_C}.$$

Now, for any given $D_C, D_{(1)}$, the distribution of D_1, \dots, D_K is multinomial(S, π_1, \dots, π_K), where $S = D - D_C$ and $\pi_k = r_k / \sum r_i$, for $k=1, \dots, K$. We now present an approach to compute the probability for a minimum of a multinomial distribution.

3.3.1 Representation of Table with the Probabilities Under Different Hypotheses

First we restrict our setting to $K=2$. For the null hypothesis against any specified alternative hypothesis of any incidence rate we interested in testing, we could fill in the probabilities for each cell of the all possible combinations of $\{D_C, D_{(1)}\}$ with its corresponding D_C and $D_{(1)}$. For equal allocation ($\rho=1$) and follow up till 10 cases ($D=10$), the table would be as follows and the values of cells could be filled for different probability or under different hypothesis. The impossible combination of $(D_C, D_{(1)})$ are removed from the table as shown in Illustration 3.3.1.

$D_C \backslash D_{(1)}$	0	1	2	3	4	5
0						
1						
2	Would be filled in based on different hypothesis/ probability					
3						
4						
5						
6						
7						
8						
9						
10						

Here are the impossible cells $(D_C, D_{(1)})$ which are removed

Illustration 3.3.1 Possible cells (above the bolded line) to define the rejection region with impossible cells which is being beyond the bolded line being removed; Probabilities of $(D_C, D_{(1)})$ could be filled in under null or specified alternatives

The filled in probabilities for $(D_C = d_C, D_{(1)} = d_{(1)})$ could be obtain through the sum of all possible probabilities $(D_C = d_C, D_1 = d_1, D_2 = d_2)$ where minimum of the d_1, d_2 is $d_{(1)}$. Since the probability of (D_C, D_1, D_2) is followed by multinomial distribution, the filled in value could be calculate easily for $K=2$. For the same case of $D=10$ and $K=2$, under the null hypothesis $H_0: (1:1,1)$, the table with filled in probabilities is as Illustration 3.3.2. Since the distribution of (D_C, D_1, D_2) under the null hypothesis is followed by Multi(10,

1/3, 1/3, 1/3). The value 0.0122 is the probability of $(D_C = 2, D_{(1)} = 1)$ which would be calculate as $P(D_C=2, D_{(1)} = 1) = P(\{D_C = 2, D_1 = 1, D_2 = 7\} \cup \{D_C = 2, D_1 = 7, D_2 = 1\}) = P(D_C = 2, D_1 = 1, D_2 = 7) + P(D_C=2, D_1=7, D_2 = 1)$. And in the case of $(D_C = 8, D_{(1)} = 1)$, there is only one possible combination $(D_C = 8, D_1=1, D_2=1)$. Hence, $P(D_C = 8, D_{(1)} = 1) = P(D_C = 8, D_1 = 1, D_2 = 1) = 0.0015$.

$D_C \backslash D_{(1)}$	0	1	2	3	4	$\lfloor 10/2 \rfloor = 5$
0	<0.0001	0.0003	0.0015	0.0041	0.0071	0.0043
1	0.0003	0.0031	0.0122	0.0285	0.0427	
2	0.0015	0.0122	0.0427	0.0854	0.0534	
3	0.0041	0.0285	0.0854	0.1423		
4	0.0071	0.0427	0.1067	0.0711		
5	0.0085	0.0427	0.0854			
6	0.0071	0.0285	0.0213			
7	0.0041	0.0122				
8	0.0015	0.0015				
9	0.0003					
10	<0.0001					

Illustration 3.3.2 Probabilities of possible combinations $\{D_C, D_{(1)}\}$ under null hypothesis (1:1,1) for equal allocation

As mentioned earlier, we can fill in the probabilities under any specific alternative hypothesis. For example, the following tables (Illustration 3.3.3, Illustration 3.3.4) are filled in the probabilities for all possible cells under the specific possible alternatives $(1: r_1 = 0.2, 1)$ or $(1: r_1 = 0.2, r_2 = 0.2)$, respectively for $K=2, D=10$. Those probabilities can be applied to calculate the power once we have decided the rejection region.

$D_C \backslash D_{(1)}$	0	1	2	3	4	$\lfloor 10/2 \rfloor = 5$
0	0.0004	0.0008	0.0007	0.0004	0.0001	0.0001
1	0.0038	0.0068	0.0054	0.0026	0.0009	
2	0.0169	0.0271	0.0190	0.0079	0.0019	
3	0.0452	0.0633	0.0383	0.0152		
4	0.0791	0.0950	0.0493	0.0127		
5	0.0949	0.0956	0.0455			
6	0.0792	0.0658	0.0190			
7	0.0455	0.0325				
8	0.0176	0.0068				
9	0.0045					
10	0.0004					

Illustration 3.3.3 Probabilities of possible combinations $\{D_C, D_{(1)}\}$ under one possible alternative hypothesis $(1: r_1 = 0.2, 1)$ for equal allocation

$D_C \backslash D_{(1)}$	0	1	2	3	4	$\lfloor 10/2 \rfloor = 5$
0	<0.00001	<0.00001	<0.00001	0.000001	0.000001	<0.00001
1	<0.00001	<0.00001	0.00001	0.00003	0.00004	
2	0.00001	0.00006	0.00022	0.00045	0.00028	
3	0.00011	0.00074	0.00223	0.00372		
4	0.00093	0.00558	0.01394	0.00929		
5	0.00558	0.02788	0.05576			
6	0.02323	0.09293	0.06970			
7	0.06638	0.19913				
8	0.12446	0.12446				
9	0.13829					
10	0.03457					

Illustration 3.3.4 Probabilities of possible combinations $\{D_C, D_{(1)}\}$ under other one possible alternative hypothesis (1: $r_1=0.2, r_2=0.2$) for equal allocation

Similarly, we could filled in the probabilities for unequal allocation with allocation ratio ρ under null hypothesis ($\rho: 1, 1$) and alternative ($\rho: r_1 = r, 1$) or ($\rho: r_1 = r, r_2 = r$). For example with $\rho=1.5$ and $D=10$, the probabilities of possible combination $\{D_C, D_{(1)}\}$ under null hypothesis ($\rho:1,1$) could be seen as following Illustration 3.3.5. Under allocation ratio $\rho=1.5$ and null hypothesis that the incident rate for each new treatment group is the same as the control group, the conditional distribution of D_C, D_I, D_2 given D is multinomial($D, \rho/\rho+2, \rho/\rho+2, \rho/\rho+2$). The probability that $(D_C = 10, D_{(1)} = 0)$ as follow up till 10 disease occur is $P(D_C = 10, D_{(1)} = 0 | D) = P(D_C = 10, D_I = 0, D_2 = 0 | D = 10) = (1.5/3.5)^{10} (1/3.5)^0 = 0.0002$.

$D_C \backslash D_{(1)}$	0	1	2	3	4	$\lfloor 10/2 \rfloor = 5$
0	<0.0001	0.0001	0.0003	0.0009	0.0015	0.0009
1	0.0001	0.0010	0.0039	0.0091	0.0137	
2	0.0007	0.0059	0.0206	0.0411	0.0257	
3	0.0029	0.0206	0.0617	0.1028		
4	0.0077	0.0462	0.1156	0.0771		
5	0.0139	0.0694	0.1387			
6	0.0173	0.0694	0.0520			
7	0.0149	0.0446				
8	0.0084	0.0084				
9	0.0028					
10	0.0002					

Illustration 3.3.5 Probabilities of possible combinations $\{D_C, D_{(1)}\}$ under null hypothesis ($\rho = 1.5: 1, 1$) for unequal allocation ratio $\rho = 1.5$

For some reasons, it might be good to include the information of its marginal probabilities for given D_C or $D_{(1)}$. The table under H_0 could be looked like as the following. The second row of the table is the marginal probability given $D_{(1)} = t$ which is the sum of all D_C for fixed $D_{(1)}$, i.e. $\sum_{D_C} P(D_C, D_{(1)} = t)$.

The second column of the table is the marginal probability given $D_C = c$ which is the probability sum over all $D_{(1)}$ for fixed $D_C = c$, i.e. $\sum_{D_{(1)}} P(D_C = c, D_{(1)})$. The marginal value 0.1716 is the marginal probability of $P(D_{(1)}=1)$ and which could be calculated as $\sum_{d_c=0}^8 P(D_C = d_c, D_{(1)} = 1)$ and the value of 0.0867 is the marginal probability of $P(D_C = 1) = \sum_{d_{(1)}=0}^4 P(D_C = 1, D_{(1)} = d_{(1)})$ or simply compute $P(D_C = 1)$ where $D_C | D \sim \text{Binomial}(10, 1/3)$.

$$\sum_{D_C} P(D_C, D_{(1)} = t)$$

$D_C \backslash D_{(1)}$	Marginal	0	1	2	3	4	$\lfloor 10/2 \rfloor = 5$
Marginal	1	0.0347	0.1716	0.3551	0.3313	0.1031	0.0043
0	0.0173	<0.0001	0.0003	0.0015	0.0041	0.0071	0.0043
1	0.0867	0.0003	0.0031	0.0122	0.0285	0.0427	
2	0.1951	0.0015	0.0122	0.0427	0.0854	0.0534	
3	0.2601	0.0041	0.0285	0.0854	0.1423		
4	0.2276	0.0071	0.0427	0.1067	0.0711		
5	0.1366	0.0085	0.0427	0.0854			
6	0.0569	0.0071	0.0285	0.0213			
7	0.0163	0.0041	0.0122				
8	0.0030	0.0015	0.0015				
9	0.0003	0.0003					
10	<0.0001	<0.0001					

$$\sum_{D_{(1)}} P(D_C = c, D_{(1)})$$

Illustration 3.3.6 Possible probabilities with marginal probabilities added in under null hypothesis (1:1,1) for equal allocation

For $K > 2$, we could obtain similar table and the filled in probabilities, however, it is not as simple to calculate as the case $K=2$. Next section 3.3.2, we will apply stochastic matrix method introduced by K. C. J. CORRADO to compute the minimum of the multinomial distribution.

3.3.2 Computing Exact Joint Distribution for the Minimum of a Multinomial Distribution for $K \geq 2$

This section contains some notation unique to find probabilities of the minimum of the multinomial that will only be used here. As stated above, given D_C and D , the conditional distribution of D_1, D_2, \dots, D_K is multinomial (S, π_1, \dots, π_K) , where $S = D - D_C$ and $\pi_k = r_k / \sum r_i$, for $k=1, \dots, K$. C. J. CORRADO (2007) presented the stochastic matrix method to compute the 1 minus the cumulative density of the minimum of multinomial distribution; $P(D_{(1)} \geq c | S, \pi_1, \dots, \pi_K)$. Once this probability, $P(D_{(1)} \geq c | S, \pi_1, \dots, \pi_K)$, is calculated, we can obtain $P(D_{(1)} = c | S, \pi_1, \dots, \pi_K)$ through $P(D_{(1)} \geq c | S, \pi_1, \dots, \pi_K) - P(D_{(1)} \geq c+1 | S, \pi_1, \dots,$

π_K). In this method, let $S_I = \sum_{k=1}^I dT_k$ be the total of disease cases of the first I treatment groups. The idea of the stochastic matrix approach is to define the transition matrices Q_I where each element in the matrices is the conditional probability of S_I given S_{I-1} , which could be express as a Binomial distribution with parameter $S-S_{I-1}$ and $\pi_I / \sum_{j=1}^K \pi_j$, assume $S_0=0$,

$$P(S_I = s_I | S_{I-1} = s_{I-1}, \pi_I^*) = \begin{cases} \binom{s - s_{I-1}}{s_I - s_{I-1}} (\pi_I^*)^{s_I - s_{I-1}} (1 - \pi_I^*)^{s - s_I} & \text{for } s_I \geq s_{I-1} \\ 0 & \text{otherwise} \end{cases} \quad (\text{Eq. 3.3.1})$$

and $\pi_I^* = \pi_I / \sum_{j=1}^K \pi_j$, since we could think that during the rest of the disease cases $s-s_{I-1}$, there are $D_I = s - s_{I-1}$ cases in the I -th treatment group with probability $\pi_I / \sum_{j=1}^K \pi_j$ as we know that there are s_{I-1} disease cases in the first $I-1$ treatment groups. Hence, we could define the matrices as follows:

$$Q_1 = [P(0|0, \pi_1), P(1|0, \pi_1), \dots, P(s|0, \pi_1)], \quad Q_j = \begin{bmatrix} P(0|0, \pi_j^*) & P(1|0, \pi_j^*) & \dots & P(s|0, \pi_j^*) \\ 0 & P(1|1, \pi_j^*) & \dots & P(s|1, \pi_j^*) \\ \vdots & \dots & \dots & \vdots \\ 0 & 0 & \dots & P(s|s, \pi_j^*) = 1 \end{bmatrix} \text{ for}$$

$j=2, \dots, K-1$, and $Q_K = [1 \ 1 \ 1 \ \dots \ 1]$. The product of $Q_1 \times Q_2 \times \dots \times Q_K$ represents the sum of all the probabilities of the trace of $D_1 = d_1 \rightarrow D_2 = d_2 \rightarrow \dots \rightarrow D_K = d_K$ through the trace of

$$S_0 = 0 \xrightarrow{d_1} S_1 = s_1 = d_1 \xrightarrow{d_2} S_2 = s_2 = d_1 + d_2 \xrightarrow{d_3} \dots \xrightarrow{d_K} S_K$$

where $S_K = \sum d_k$ is known. The value of the products $Q_1 \times Q_2 \times \dots \times Q_K$ should be the same as the probability of all possible combinations which equals to 1.

However, we are interested in obtaining the $P(D_{(1)} \geq c | S, \pi_1, \dots, \pi_K)$ in the current stage, the conditional probability of the minimum number of disease cases observed in a treatment group is more or equal to c given total sum of numbers of disease cases in all treatment groups and their incidence rates. In this case, we know that the number of disease cases must exceed or equal c in all the treatment groups which is equivalent to saying that the difference of $D_j = s_j - s_{j-1}$ can't be less than c for each $j=1, \dots, K$. As the result, we set $P(S_I = s_I | S_{I-1} = s_{I-1}, \pi_I^*) = 0$ whenever $s_I - s_{I-1} < c$ for $I = 1, 2, \dots, K$, in each element of the transition matrices and call the new stochastic matrices Q_I^* for $I = 1, 2, \dots, K$ after replacing

$$P(S_I = s_I | S_{I-1} = s_{I-1}, \pi_I^*) = 0 \text{ as } s_I - s_{I-1} < c \text{ for each } I = 1, 2, \dots, K..$$

Then the product of matrices, $Q_1^* \times \dots \times Q_K^*$ gives $P(D_{(1)} \geq c | S, \pi_1, \dots, \pi_K)$ exactly since we have zeroed out all the possibilities of $D_j = s_j - s_{j-1} < c$ for $j = 1, \dots, K$, and thus summed over all the possible traces D_1 to D_K under the constraint $D_j = s_j - s_{j-1} \geq c, \forall j = 1, \dots, K$. Then the mass probability of the minimum of the multinomial distribution would be obtained through (Eq. 3.3.2) and have numerical results through computing.

$$P(D_{(1)} = c | S, \pi_1, \dots, \pi_K) = P(D_{(1)} \geq c | S, \pi_1, \dots, \pi_K) - P(D_{(1)} \geq c + 1 | S, \pi_1, \dots, \pi_K) \quad (\text{Eq. 3.3.2})$$

For example, if there are three treatment groups ($K=3$) and one control group with equal allocation ($\rho = 1$) and the equal incidence rate in each groups ($r_k = 1$ for $k=1, \dots, K$), and the trials were followed until 10 disease cases accrued, ($D = 10$). We could obtain the conditional joint distribution of D_C and $D_{(1)}$ given $D = 10$ disease cases through the equation:

$$P(D_C = d_C, D_{(1)} = d_{(1)} | D = 10) = P(D_{(1)} = d_{(1)} | D = 10, D_C = d_C)P(D_C = d_C | D = 10).$$

Since it's the product of two distributions and $P(D_{(1)} | D = 10, D_C = d_C) = P(D_{(1)} | D_T = 10 - d_C)$ is the minimum of the multinomial distribution for known d_C , we could obtain $P(D_{(1)} \geq c | D = 10, D_C = d_C)$ and c through the products of stochastic matrices. Assume $D_C = 7$ and $c=1$, $P(D_{(1)} \geq 1 | D = 10, D_C = 7) = P(D_{(1)} \geq 1 | \sum d_k = 3)$ would be given through $Q_1^* \times Q_2^* \times Q_3^*$, where

$$Q_1^* = [\text{Pr}(S_1 = 0 | S_0 = 0, \pi_1) = 0 \quad \text{Pr}(S_1 = 1 | S_0 = 0, \pi_1) \quad \text{Pr}(S_1 = 2 | S_0 = 0, \pi_1) \quad \text{Pr}(S_1 = 3 | S_0 = 0, \pi_1)]$$

$$Q_2^* = \begin{bmatrix} \text{Pr}(S_2 = 0 | S_1 = 0, \pi_2^*) = 0 & \text{Pr}(S_2 = 1 | S_1 = 0, \pi_2^*) & \text{Pr}(S_2 = 2 | S_1 = 0, \pi_2^*) & \text{Pr}(S_2 = 3 | S_1 = 0, \pi_2^*) \\ 0 & \text{Pr}(S_2 = 1 | S_1 = 1, \pi_2^*) = 0 & \text{Pr}(S_2 = 2 | S_1 = 1, \pi_2^*) & \text{Pr}(S_2 = 3 | S_1 = 1, \pi_2^*) \\ 0 & 0 & \text{Pr}(S_2 = 2 | S_1 = 2, \pi_2^*) = 0 & \text{Pr}(S_2 = 3 | S_1 = 2, \pi_2^*) \\ 0 & 0 & 0 & \text{Pr}(S_2 = 3 | S_1 = 3, \pi_2^*) = 0 \end{bmatrix}$$

$Q_3^* = [1 \ 1 \ 1 \ 0]^T$. In those new matrices, we reset some elements in the matrices to be 0 to eliminate settings where the minimal number of disease cases observed in the treatment groups is less than 1 ($D_{(1)} \geq 1$).

Based on the (Eq. 3.3.1), we could calculate each element in the matrices, Q_1^* and Q_2^* .

3.4 Criteria to Obtain the Rejection Region

Due to the wide range of potential alternatives to the global null as well as anomalies from discreteness of the multinomial distribution, (as we will show later) no rejection region is uniformly most powerful against all alternatives. Therefore, anchoring on the Lower Left Quadrant Rule (LLQR), we try to define rejection regions based on different heuristic criteria that will perform well from a power for size standpoint against reasonable and likely alternatives and evaluate their performances against these alternatives.

3.4.1 Use the Control

Use $D_C > c$ to define rejection region, R_α and the smallest value of c for which $P_{H_0}(D_C > c) \leq \alpha$.

We denote this method “control only”, or abbreviate this size α rejection region as “ $R_{\alpha-c}$ ”. While larger D_C provides evidence that the control group is worse than the treatments collectively, it ignores information contained in $D_{(1)}$ that could reflect one treatment being very good. Since the rejection region is only based on D_C , the conditional distribution of D_C given D is $\text{Bin}(D, 1/(1+K))$ under null hypothesis of $(1:1, \dots, 1)$ for equal allocation or $\text{Bin}(D, \rho/(\rho+K))$ under null hypothesis and ρ is the allocation ratio for the control to each new treatments. The probability of falsely reject the null hypothesis would be calculated based on this Binomial distribution. However, due to discreteness of the binomial, in many settings, the Type I error could be much smaller than the nominal level α for $D_C > c$, but the type I error exceeds the requested level α for the rejection region $D_C > c-1$. This sacrifice in actual size under H_0 will also be associated with a loss power (compared to a rejection rule with actual size closer to the nominal α when a given H_a is true). One approach to lessen this loss would be to incorporate information from a small $D_{(1)}$ when the number in the control group falls on the boundary $D_C = c$. The smaller $D_{(1)}$ is for given $D_C = c$, the more evidence there is that the incidence rate for the best treatment is smaller than the control. The rejection region is then $\{D_C > c \text{ and } D_{(1)} < t \text{ when } D_C = c\}$ where c is the smallest value for which $P_{H_0}(D_C > c | D) \leq \alpha$ and t is the largest value for which

$$P_{H_0}(D_C > c \cup \{D_{(1)} < t \cap D_C = c\} | D) \leq \alpha.$$

This method is called “Control Boundary Augmented”, or abbreviate this augmented size α rejection region as “ $R_{\alpha-CBA}$ ”.

In the following example, we are going to demonstrate how the rejection region is constructed by “controls only” and “controls boundary augmented” or other methods which we will purposed and illustrate this example again later for given size level α . For comparing two new treatment groups ($K=2$) to one control group with equal allocation ($\rho=1$) and the trial is followed up until 10 disease cases ($D=10$). For testing $(1: r_1 = r, 1)$ or $(1: r_1 = r, r_2 = r)$ as alternative (in a scenario where one of the two new treatments fully works while the other is ineffective) and the required type I error of 0.05. By section 3.3.1, for each different value of D_C , we could compute the probability of $D_C = d_c$ from Bin(10, 1/3) under H_0 . Also, for any values of $D_{(1)} | D_C = d_c$, we could compute its probability on the boundary under H_0 and the probabilities are shown in Illustration 3.4.1. For the “control only” method, if we define the rejection region to be $D_C > 6$ the type I error would be 0.0197 but at or $D_C > 5$, the type-1 error is 0.0766. The later one exceeds required level 0.05. As the result, the rejection region should then be $\{(D_C, D_{(1)}) | D_C > 6\}$ for a level 5% test. For the “control boundary augmented” method, the rejection region is $\{(D_C, D_{(1)}) | D_C > 6, (D_{(1)} = 0 \cap D_C = 6)\}$ with a type I error 0.0268 as including $\{(D_{(1)} = 1 \cap D_C = 6)\}$ will increase the overall Type 1 error to 0.0553. The rejection regions are shaded in the following table.

$D_C \backslash D_{(1)}$	Marginal	0	1	2	3	4	$\lfloor 10/2 \rfloor = 5$
Marginal	1	0.0347	0.1716	0.3551	0.3313	0.1031	0.0043
0	0.0173	<0.0001	0.0003	0.0015	0.0041	0.0071	0.0043
1	0.0867	0.0003	0.0031	0.0122	0.0285	0.0427	
2	0.1951	0.0015	0.0122	0.0427	0.0854	0.0534	
3	0.2601	0.0041	0.0285	0.0854	0.1423		
4	0.2276	0.0071	0.0427	0.1067	0.0711		
5	0.1366	0.0085	0.0427	0.0854			
6	0.0569	0.0071	0.0285	0.0213			
7	0.0163	0.0041	0.0122				
8	0.0030	0.0015	0.0015				
9	0.0003	0.0003					
10	<0.0001	<0.0001					

■ : Rejection region by “controls only”

■ U ■ : Rejection region by “controls boundary augmented”

Illustration 3.4.1 Rejection Region through “Controls Only” or “Controls Boundary Augmented” for $K = 2$, $D = 10$, and $\alpha = 0.05$ for equal allocation

For unequal allocation case and let $\rho = 1.5$ be the allocation ratio for the control to each new treatment, the probabilities for each combination of D_C and $D_{(1)}$ and the marginal probabilities under null hypothesis for comparing 2 new treatments ($K=2$) to one control and $D = 10$ are in the following table.

For level $\alpha = 0.05$, the rejection region based on “Control Only” is $\{D_C > 7\}$ since the size of the rejection region $\{D_C > 6\}$ is about 0.0595 which exceeds 0.05. For the rejection region based on “Control Boundary Augmented”, the rejection region is $\{(D_C, D_{(1)}) \mid D_C > 7 \text{ or } (D_C = 7, D_{(1)} = 0)\}$ with size 0.0316.

$D_C \backslash D_{(1)}$	Marginal	0	1	2	3	4	$\lfloor 10/2 \rfloor = 5$
Marginal	1	0.0689	0.2654	0.3928	0.2310	0.0409	0.0009
0	0.0037	<0.0001	0.0001	0.0003	0.0009	0.0015	0.0009
1	0.0278	0.0001	0.0010	0.0039	0.0091	0.0137	
2	0.0940	0.0007	0.0059	0.0206	0.0411	0.0257	
3	0.1879	0.0029	0.0206	0.0617	0.1028		
4	0.2467	0.0077	0.0462	0.1156	0.0771		
5	0.2220	0.0139	0.0694	0.1387			
6	0.1387	0.0173	0.0694	0.0520			
7	0.0595	0.0149	0.0446				
8	0.0167	0.0084	0.0084				
9	0.0028	0.0028					
10	0.0002	0.0002					

: Rejection region by “controls only”
 U : Rejection region by “controls boundary augmented”

Illustration 3.4.2 Rejection Region through “Controls Only” or “Controls Boundary Augmented” for $K = 2$, $D = 10$, and $\alpha = 0.05$ for unequal allocation with $\rho = 1.5$

3.4.2 Use the Minimal New Treatment

Conversely to use only the “controls” we could only use the “minimal new treatment” $D_{(1)} < t$, to define R_α and it called “Minimal New Treatment Only”, or abbreviate this size α rejection region as “ $R_{\alpha-T}$ ”. A smaller $D_{(1)}$ could reflect either only one treatment being very good or even all treatments being better than the control. This, however ignores information from D_C that could comparatively reflect the control being worse (or for that matter if something has gone horribly wrong, better) that all treatments as a group. However, due to discreteness of the multinomial, the type I error could again be much smaller than the nominal level α for $D_{(1)} < t$ for a given t with the type I error exceeding the requested level α for the rejection region expanded to $D_{(1)} < t+1$. This sacrifice in actual size under H_0 will also be associated with a loss power (compared to a rejection rule with actual size closer to the nominal α when a given H_a is true). One approach to lessen this impact would be to incorporate information from a large D_C when $D_{(1)} = t$ on the boundary, i.e. the rejection region is $\{D_{(1)} < t \text{ and } D_C > c \mid D_{(1)} = t\}$ where t is the largest value for which $P(D_{(1)} < t) \leq \alpha$ as H_0 is true and c is the smallest value for which

$$P_{H_0} \left(D_{(1)} < t \cup \{D_{(1)} = t \cap D_C > c\} \right) \leq \alpha.$$

The method which accumulates extra cells on the boundary based on above as the rejection region is called “Minimal new Treatment Boundary Augmented”, or abbreviate this rejection region as “ $R_{\alpha-TBA}$ ”. Continuing with the previous example, with $K=2$ and equal allocation ($\rho = 1$) followed up until 10 disease cases ($D=10$), if we define the rejection region from treatment, the type I errors are 0.0347 and 0.206 for the rejection region $\{(D_C, D_{(1)}) | D_{(1)} < 1\}$ and $\{(D_C, D_{(1)}) | D_{(1)} < 2\}$ respectively. For a level 5% test, the rejection region is $\{(D_C, D_{(1)}) | D_{(1)} < 1\}$ based on “Minimal New Treatment Only”. However, the type I error for this rejection region is far below the nominal 5% level. The type I error of the rejection region based on “treatment boundary augmented” $\{(D_C, D_{(1)}) | D_{(1)} < 1 \text{ or } (D_C \geq 6, D_{(1)} = 1)\}$ is 0.0696 (which exceeds 0.05) but for $\{(D_C, D_{(1)}) | D_{(1)} < 1 \text{ or } (D_C \geq 7, D_{(1)} = 1)\}$ is 0.0484 which is not greater than level 5% and should have greater power than the rejection region based on $\{(D_C, D_{(1)}) | D_{(1)} < 1\}$. The rejection region based on “minimal new treatment only” and “minimal new treatment boundary augmented” are illustrated below.

$D_C \backslash D_{(1)}$	Marginal	0	1	2	3	4	$\lfloor 10/2 \rfloor = 5$
Marginal	1	0.0347	0.1716	0.3551	0.3313	0.1031	0.0043
0	0.0173	<0.0001	0.0003	0.0015	0.0041	0.0071	0.0043
1	0.0867	0.0003	0.0031	0.0122	0.0285	0.0427	
2	0.1951	0.0015	0.0122	0.0427	0.0854	0.0534	
3	0.2601	0.0041	0.0285	0.0854	0.1423		
4	0.2276	0.0071	0.0427	0.1067	0.0711		
5	0.1366	0.0085	0.0427	0.0854			
6	0.0569	0.0071	0.0285	0.0213			
7	0.0163	0.0041	0.0122				
8	0.0030	0.0015	0.0015				
9	0.0003	0.0003					
10	<0.0001	<0.0001					

: Rejection region by “Minimum New Treatment only”
 U : Rejection region by “Treatment boundary augmented”

Illustration 3.4.3 Rejection Region through “Minimal New Treatment Only” or “Minimal New Treatment Boundary Augmented” for $K = 2$, $D = 10$, and $\alpha = 0.05$ for equal allocation

For unequal allocation ratio with $\rho = 1.5$, there will be no rejection for the rejection region based on “new treatment only” since $P\{D_{(1)} = 0\}$ under null hypothesis is $0.0689 > 0.05$ while the rejection region based “treatment boundary augmented” is $(D_C > 5, D_{(1)} = 0)$ with size 0.0434.

$D_C \backslash D_{(1)}$	Marginal	0	1	2	3	4	$\lfloor 10/2 \rfloor = 5$
Marginal	1	0.0689	0.2654	0.3928	0.2310	0.0409	0.0009
0	0.0037	<0.0001	0.0001	0.0003	0.0009	0.0015	0.0009
1	0.0278	0.0001	0.0010	0.0039	0.0091	0.0137	
2	0.0940	0.0007	0.0059	0.0206	0.0411	0.0257	
3	0.1879	0.0029	0.0206	0.0617	0.1028		
4	0.2467	0.0077	0.0462	0.1156	0.0771		
5	0.2220	0.0139	0.0694	0.1387			
6	0.1387	0.0173	0.0694	0.0520			
7	0.0595	0.0149	0.0446				
8	0.0167	0.0084	0.0084				
9	0.0028	0.0028					
10	0.0002	0.0002					


 : Rejection region by
 “treatment boundary augmented”

Illustration 3.4.4 Rejection Region through “Minimal New Treatment Boundary Augmented” for $K=2$, $D=10$, and $\alpha=0.05$ for unequal allocation with $\rho=1.5$

3.4.3 Difference Between Control And The Minimal Treatment

For equal allocation $\rho=1$, use $D_C - D_{(1)} > d$ to define R_α . Unlike the previous two approaches, here we simultaneously take D_C and $D_{(1)}$ into account. A larger difference of $D_C - D_{(1)}$ suggests there is more evidence to show that at least one treatment is better than the control. We denote this method “Difference Only”, or abbreviate as “ $R_{\alpha-D}$ ”. Again, due to discreteness of the multinomial, the type I error could be much smaller than the required level α , that is $P\{D_C - D_{(1)} > d\} < \alpha$ while $P\{D_C - D_{(1)} > d+1\} > \alpha$. This sacrifice in actual size under H_0 will also be associated with a loss power (compared to a rejection rule with actual size closer to the nominal α when a given H_a is true). In an alternate approach targeting the alternative hypothesis of $(1: r_1 = r, 1)$, we could incorporate information from a small $D_{(1)}$ when $D_C - D_{(1)} = d$ to lessen this impact. If the alternative hypothesis is $(1: r_1 = r, r_2 = r)$, we could incorporate information from large D_C when $D_C - D_{(1)} = d$ to lessen the impact of loss power. Here, we emphasize the worst case scenario of only one treatment work better than the control, $(1: r_1 = r, 1)$, and use the augmented rejection region $\{D_C - D_{(1)} > d \cup (D_{(1)} < t \mid D_C - D_{(1)} = d)\}$ where d is the smallest value for which $P\{D_C - D_{(1)} > d\}$ under H_0 is $\leq \alpha$ and t is the largest value for which $P\{(D_C - D_{(1)} > d) \cup (D_{(1)} < t \cap D_C - D_{(1)} = d)\}$ under H_0 is $\leq \alpha$. This method is denoted “Difference Boundary Augmented”(DBA), or abbreviate this rejection region as “ $R_{\alpha-DBA}$ ”.

Continuing with the previous example for $K=2$, $\rho=1$ and $D=10$, the rejection region for size = $0.0268 \leq 5\%$ level obtained from method “difference only” is $D_C - D_{(1)} > 5$ since the size for $\{D_C - D_{(1)} > 4\}$

is 0.0638 which exceeds request 0.05. The size of the rejection region based on “difference boundary augmented” from worst case scenario, i.e. $\{(D_C, D_{(1)})|D_C - D_{(1)} > 5 \cup \{D_{(1)} = 0 \cap D_C - D_{(1)} = 5\}$ is 0.0353 which improves the power computation comparing to the rejection region based on “difference only” and adding $D_{(1)} = 1, D_C - D_{(1)} = 5$ as a sufficient condition increases the type 1 error to 0.0638. The top number in each cell of the Illustration 3.4.5 below shows the rejection regions we just described for difference of the control and the minimum treatment ($D_C - D_{(1)}$) and the second item of the cells has the size of corresponding combination of $D_C, D_{(1)}$ under the null hypothesis for the setting $\alpha=0.05, K=2$ and $D = 10$.

$D_C \backslash D_{(1)}$	0	1	2	3	4	$\lfloor 10/2 \rfloor = 5$
0	$D_C - D_{(1)} = 0$ Size<0.0001	$D_C - D_{(1)} = -1$ Size=0.0003	$D_C - D_{(1)} = -2$ Size=0.0015	$D_C - D_{(1)} = -3$ Size=0.0041	$D_C - D_{(1)} = -4$ Size=0.0071	$D_C - D_{(1)} = -5$ Size=0.0043
1	$D_C - D_{(1)} = 1$ Size=0.0003	$D_C - D_{(1)} = 0$ Size=0.0031	$D_C - D_{(1)} = -1$ Size=0.0122	$D_C - D_{(1)} = -2$ Size=0.0285	$D_C - D_{(1)} = -3$ Size=0.0427	
2	$D_C - D_{(1)} = 2$ Size=0.0015	$D_C - D_{(1)} = 1$ Size=0.0122	$D_C - D_{(1)} = 0$ Size=0.0427	$D_C - D_{(1)} = -1$ Size=0.0854	$D_C - D_{(1)} = -2$ Size=0.0534	
3	$D_C - D_{(1)} = 3$ Size=0.0041	$D_C - D_{(1)} = 2$ Size=0.0285	$D_C - D_{(1)} = 1$ Size=0.0854	$D_C - D_{(1)} = 0$ Size=0.1423		
4	$D_C - D_{(1)} = 4$ Size=0.0071	$D_C - D_{(1)} = 3$ Size=0.0427	$D_C - D_{(1)} = 2$ Size=0.1067	$D_C - D_{(1)} = 1$ Size=0.0711		
5	$D_C - D_{(1)} = 5$ Size=0.0085	$D_C - D_{(1)} = 4$ Size=0.0427	$D_C - D_{(1)} = 3$ Size=0.0854			
6	$D_C - D_{(1)} = 6$ Size=0.0071	$D_C - D_{(1)} = 5$ Size=0.0285	$D_C - D_{(1)} = 4$ Size=0.0213			
7	$D_C - D_{(1)} = 7$ Size=0.0041	$D_C - D_{(1)} = 6$ Size=0.0122				
8	$D_C - D_{(1)} = 8$ Size=0.0015	$D_C - D_{(1)} = 7$ Size=0.0015				
9	$D_C - D_{(1)} = 9$ Size=0.0003					
10	$D_C - D_{(1)} = 10$ Size<0.0001					

: Rejection region by “Difference only”

U : Rejection region by “Difference boundary augmented”

Illustration 3.4.5 Rejection Region through “Difference Only” or “Difference Boundary Augmented” for $K = 2, D = 10$, and $\alpha = 0.05$ for equal allocation

For the case of unequal allocation with allocation ratio ρ , there are different numbers of subject recruited in the control group and each new treatment group. Hence, the difference of D_C and $D_{(1)}, D_C - D_{(1)}$, is not the best way to compare the control to the new treatment group and some modification should be applied. It could be more reasonable to compare the difference of estimated incident rate for the control group to the minimum new treatment group. Since the difference of incident rate is $D_C / N_C - D_{(1)} / N_T = D_C / \rho_C N - D_{(1)} / \rho_T N = (D_C - \rho D_{(1)}) / \rho_C N$ where ρ, ρ_C and D are determined and fixed before the trial study and N is also fixed since it only depends on the ρ and total follow up cases D , the modified difference, defined as

$D_C - \rho D_{(1)}$, could be seen as a modified criteria to obtain the rejection region using “difference” for unequal allocation with ratio ρ . For example of $K=2, D=10$, and the unequal allocation that allocates 0.5 times more subjects to control group than the treatment group ($\rho=1.5$), we define the rejection region based on the modified difference. In Illustration 3.4.6, the first row of each cell of is the modified difference and the second row is the size of its corresponding combination $(D_C, D_{(1)})$ under the null hypothesis with allocation rate $\rho=1.5$ where the conditional distribution of D_C, D_1, D_2 given $D = 10$ is from Multinomial $(10, 1.5/3.5, 1/3.5, 1/3.5)$. For $(D_C=8, D_{(1)}=1)$, the first row is modified $D_C - \rho D_{(1)} = 8 - 1.5 \times 1 = 6.5$ and $P(D_C = 8, D_{(1)} = 1 | D = 10)$ under H_0 is $P(D_C = 8, D_1=1, D_2=1 | D = 10) = 90 \times (1.5/3.5)^8 \times (1/3.5)^1 \times (1/3.5)^1 = 0.0084$.

$D_C - D_{(1)}$	0	1	2	3	4	$\lfloor 10/2 \rfloor = 5$
0	$D_C - \rho D_{(1)} = 0$ size = 0.0001	$D_C - \rho D_{(1)} = -1.5$ size = 0.0001	$D_C - \rho D_{(1)} = -3$ size = 0.0003	$D_C - \rho D_{(1)} = -4.5$ Size = 0.0009	$D_C - \rho D_{(1)} = -6$ size = 0.0015	$D_C - \rho D_{(1)} = -7.5$ size = 0.0009
1	$D_C - \rho D_{(1)} = 1$ size = 0.0001	$D_C - \rho D_{(1)} = -0.5$ size = 0.0010	$D_C - \rho D_{(1)} = -2$ size = 0.0039	$D_C - \rho D_{(1)} = -3.5$ Size = 0.0091	$D_C - \rho D_{(1)} = -5$ size = 0.0137	
2	$D_C - \rho D_{(1)} = 2$ size = 0.0007	$D_C - \rho D_{(1)} = 0.5$ size = 0.0059	$D_C - \rho D_{(1)} = -1$ size = 0.0206	$D_C - \rho D_{(1)} = -2.5$ Size = 0.0411	$D_C - \rho D_{(1)} = -4$ size = 0.0257	
3	$D_C - \rho D_{(1)} = 3$ size = 0.0029	$D_C - \rho D_{(1)} = 1.5$ size = 0.0206	$D_C - \rho D_{(1)} = 0$ size = 0.0617	$D_C - \rho D_{(1)} = -1.5$ Size = 0.1028		
4	$D_C - \rho D_{(1)} = 4$ size = 0.0077	$D_C - \rho D_{(1)} = 2.5$ size = 0.0462	$D_C - \rho D_{(1)} = 1$ size = 0.1156	$D_C - \rho D_{(1)} = -0.5$ Size = 0.0771		
5	$D_C - \rho D_{(1)} = 5$ size = 0.0139	$D_C - \rho D_{(1)} = 3.5$ size = 0.0694	$D_C - \rho D_{(1)} = 2$ size = 0.1387			
6	$D_C - \rho D_{(1)} = 6$ size = 0.0173	$D_C - \rho D_{(1)} = 4.5$ size = 0.0694	$D_C - \rho D_{(1)} = 3$ size = 0.0520			
7	$D_C - \rho D_{(1)} = 7$ size = 0.0149	$D_C - \rho D_{(1)} = 5.5$ size = 0.0446				
8	$D_C - \rho D_{(1)} = 8$ size = 0.0084	$D_C - \rho D_{(1)} = 6.5$ size = 0.0084				
9	$D_C - \rho D_{(1)} = 9$ size = 0.0028					
10	$D_C - \rho D_{(1)} = 10$ size = 0.0002					

: Rejection region by “Difference only”

Illustration 3.4.6 Rejection Region through “Difference Only” or “Difference Boundary Augmented”

(shaded) for $K = 2, D = 10$, and $\alpha = 0.05$ for unequal allocation design as $\rho = 1.5$. The first row is the modified difference while the second row is the probability of the combination under null hypothesis.

For the rejection region using modified difference, we reject the null as $\{(D_C, D_{(1)}) | D_C - \rho D_{(1)} \geq d\}$ and determined d through $P(D_C - \rho D_{(1)} \geq d) \leq \alpha$ under null hypothesis. In this example, the rejection region use “difference only” could be determined as $\{(D_C, D_{(1)}) | D_C - 1.5 D_{(1)} \geq 6.5\}$ which is $\{(D_C \geq 7, D_{(1)} = 0), (D_C$

$\geq 8, D_{(1)} = 1\}$. Since there is no tie on modified difference, we do not consider the “Difference Boundary Augmented” in this case.

3.4.4 Cumulative Binomial Metric For Control And The Minimal Treatment

One could argue that the difference $D_C - D_{(1)}$ for equal allocation as a metric is problematic as it does not factor in the sum of $D_{(1)}$ and D_C . For example $D_C - D_{(1)} = 3$ might have stronger implications when this sum is 3; $D_C = 3, D_{(1)} = 0$, than it does when this sum is 7, $D_C = 5, D_{(1)} = 2$. In order to account for this possibility, one could instead use as a metric the cumulative probability to observe $D_{(1)}$ cases based on a binomial distribution $\text{Bin}(D_C + D_{(1)}, 0.5)$ for equal allocation i.e., with only one treatment ($K=1$) ($D_{(1)}$) with one control (D_C) or $\text{Bin}(D_C + D_{(1)}, 1/\rho+1)$ for unequal allocation with allocation ratio ρ for the control to each treatment group. While in taking the minimum of the $K > 1$ treatments, one would expect $D_{(1)}$ to be shifted to the left under H_0 conditional on $D_{(1)} + D_C$, than it would under the binomial conditional on $D_{(1)} + D_C$, as we have seen, there is no simple closed interpretable formula to quantify this. Thus we have used the $\text{Bin}(D_C + D_{(1)}, 0.5)$ for equal allocation or $\text{Bin}(D_C + D_{(1)}, 1/\rho+1)$ for unequal allocation with allocation ratio ρ as a metric for the likelihood of $D_{(1)}$ given $D_C + D_{(1)}$ as a convenient surrogate. Although the true conditional distribution of $D_{(1)}$ given $D_C + D_{(1)}$ is not Binomial distribution, a smaller cumulative probability on this metric indicates the less likely the chance to observe $D_{(1)}$ given $D_C + D_{(1)}$ if no treatment is better than the control. Hence, the metric would be determined from Binomial distribution: for each combination of $D_C = c$ and $D_{(1)} = t$, such that $c+t=Q$, we would calculate $\text{CB} = \text{P}\{D_{(1)} \leq t \mid D_C + D_{(1)} = c + t\}$. The smaller this CB is, the more evidence (among this cumulative binomial metric) the data provides to conclude that at least one rate of the treatment groups has a lower rate than the control group. Therefore, we could accumulate the rejection region using the cumulative binomial metric from the smallest as shown in the following section 3.4.5. The size level α rejection region based on this metric is called “cumulative binomial” or abbreviated as “ $R_{\alpha\text{-CB}}$ ”.

Continuing with the example of equal allocation, with 2 treatment groups ($K=2$) and ($D=10$). We could obtain the cumulative binomial through $\text{Bin}(D_C + D_{(1)}, 1/2)$. If we take a look at Illustration 3.4.7, the top number in each cell is the cumulative binomial probability, $\text{CB} = \text{P}(D_{(1)} \leq t \mid D_C + D_{(1)} = c + t)$, of the corresponding ($D_C = c, D_{(1)} = t$) obtained from $D_{(1)} \mid D_C + D_{(1)} = c + t \sim \text{Bin}(D_C + D_{(1)} = c + t, 1/2)$ and the

second number in each cell is the probability of the combination $(D_C, D_{(1)})$ under null hypothesis. For example the cell $(D_C = 9, D_{(1)} = 0)$, in red for Illustration 3.4.7, shows that $CB = 0.00195$ (top number) which is calculated by $CB = P(D_{(1)} \leq 0 | D_C + D_{(1)} = c + t = 9)$ where $D_{(1)} | D_C + D_{(1)} = 9 \sim \text{Bin}(9, 1/2)$ and the size = 0.0003 (lower number) $P(D_C = 9, D_{(1)} = 0) = P(D_C = 9, D_1 = 0, D_2 = 1) + P(D_C = 9, D_1 = 1, D_2 = 0)$ under the distribution of D_C, D_1, D_2 given D is multinomial(10, 1/3, 1/3, 1/3) under null hypothesis.

$D_C \backslash D_{(1)}$	0	1	2	3	4	5
0	CB=1.00000 size <0.0001	CB=1.00000 size=0.0003	CB=1.00000 size =0.0015	CB=1.00000 size =0.0041	CB=1.00000 size =0.0071	CB=1.00000 size =0.0043
1	CB=0.50000 size =0.0003	CB=0.75000 size =0.0031	CB=0.87500 size =0.0122	CB=0.93750 size =0.0285	CB=0.96875 size =0.0427	CB=0.98438
2	CB=0.25000 size =0.0015	CB=0.50000 size =0.0122	CB=0.68750 size =0.0427	CB=0.81250 size =0.0854	CB=0.89063 size =0.0534	CB=0.93750
3	CB=0.12500 size =0.0041	CB=0.31250 size =0.0285	CB=0.50000 size =0.0854	CB=0.65625 size =0.1423	CB=0.77344	CB=0.85547
4	CB=0.06250 size =0.0071	CB=0.18750 size =0.0427	CB=0.34375 size =0.1067	CB=0.50000 size =0.0711	CB=0.63672	CB=0.74609
5	CB=0.03125 size =0.0085	CB=0.10938 size =0.0427	CB=0.22656 size =0.0854	CB=0.36328	CB=0.50000	CB=0.62305
6	CB=0.01563 size =0.0071	CB=0.06250 size =0.0285	CB=0.14453 size =0.0213	CB=0.25391	CB=0.37695	CB=0.50000
7	CB=0.00781 size =0.0041	CB=0.03516 size =0.0122	CB=0.08984	CB=0.17188	CB=0.27441	CB=0.38721
8	CB=0.00391 size =0.0015	CB=0.01953 size =0.0015	CB=0.05469	CB=0.11328	CB=0.19385	CB=0.29053
9	CB=0.00195 Size=0.0003	CB=0.01074	CB=0.03271	CB=0.07300	CB=0.13342	CB=0.21198
10	CB=0.00098 size <0.0001	CB=0.00586	CB=0.01929	CB=0.04614	CB=0.08978	CB=0.15088

Illustration 3.4.7 Rejection Region through “Cumulative Binomial” (shaded) for $K=2, D = 10$, and $\alpha = 0.05$. For each $(D_C = c, D_{(1)} = t)$, the numbers followed by “CB” are the cumulative binomial calculated by $P(D_{(1)} \leq t | D_C + D_{(1)} = c + t)$ where $D_{(1)} | D_C + D_{(1)} = c + t \sim \text{Bin}(D_C + D_{(1)} = c + t, 1/2)$ and the number followed by “size” is the probability of $(D_C = c, D_{(1)} = t)$ under the null hypothesis discussed in section 3.3. For example, In the cell $(D_C = 9, D_{(1)} = 0)$, $CB = P(D_{(1)} \leq 0 | D_C + D_{(1)} = 9) = 0.00195$ and size = 0.0003 is by multinomial distribution from section 3.3 under null hypothesis for equal allocation

Under the null hypothesis that the incident rate is the same in the control and each treatment group, based on the cumulative binomial metric (top number of the cell), as the arrows connecting the bolded top boxes on the figure show we would reject the combination $(D_C = 10, D_{(1)} = 0)$ first since 0.00098 is the lowest CB and followed by $(D_C = 9, D_{(1)} = 0)$ which is the next lowest CB and then so on (reject the combinations $(D_C = 8, D_{(1)} = 0)$, $(D_C = 7, D_{(1)} = 0)$, $(D_C = 6, D_{(1)} = 0)$, $(D_C = 8, D_{(1)} = 1)$, $(D_C = 5, D_{(1)} = 0)$, $(D_C = 7, D_{(1)} = 1)$, $(D_C = 4, D_{(1)} = 0)$ based on the next lowest CB). This path is followed until the cumulated binomial probability (top number in the cells) and preserving the LLQR property (as shown in the next

section) until the type I error reaches the maximum value that does not exceed 5% level of type I error.

Also, combinations such as $(D_C=10, D_{(1)}=1)$, $(D_C=9, D_{(1)}=1)$,... do not put into the rejection region since those combinations are impossible with $D = 10$. The shaded cells of the following table represent the rejection region obtained through this method. Note The cells beyond the bolded line are impossible combinations of $(D_C, D_{(1)})$ but are shown as we need the information in these cells to compute the cumulative binomial in the metric being used.

$D_C \backslash D_{(1)}$	0	1	2	3	4	5
0	CB=1.00000 size <0.0001	CB=1.00000 size =0.0001	CB=1.00000 Size=0.0003	CB=1.00000 size =0.0009	CB=1.00000 size =0.0015	CB=1.00000 size =0.0009
1	CB=0.60000 size =0.0001	CB=0.84000 size =0.0010	CB=0.93600 Size=0.0039	CB=0.97440 size =0.0091	CB=0.98976 size =0.0137	CB=0.99590
2	CB=0.36000 size =0.0007	CB=0.64800 size =0.0059	CB=0.82080 Size=0.0206	CB=0.91296 size =0.0411	CB=0.95904 size =0.0257	CB=0.98116
3	CB=0.21600 size =0.0029	CB=0.47520 size =0.0206	CB=0.68256 size =0.0617	CB=0.82080 size =0.1028	CB=0.90374	CB=0.95019
4	CB=0.12960 size =0.0077	CB=0.33696 size =0.0462	CB=0.54432 size =0.1156	CB=0.71021 size =0.0771	CB=0.82633	CB=0.90065
5	CB=0.07776 size =0.0139	CB=0.23328 size =0.0694	CB=0.41990 size =0.1387	CB=0.59409	CB=0.73343	CB=0.83376
6	CB=0.04666 size =0.0173	CB=0.15863 size =0.0694	CB=0.31539 size =0.0520	CB=0.48261	CB=0.63310	CB=0.75350
7	CB=0.02729 size =0.0149	CB=0.10638 size =0.0446	CB=0.23179	CB=0.38228	CB=0.53277	CB=0.66521
8	CB=0.01680 size =0.0084	CB=0.07054 size =0.0084	CB=0.16729	CB=0.29628	CB=0.43818	CB=0.57440
9	CB=0.01008 size=0.0028	CB=0.04636	CB=0.11892	CB=0.22534	CB=0.35304	CB=0.48585
10	CB=0.00605 size =0.0002	CB=0.03023	CB=0.08344	CB=0.16858	CB=0.27926	CB=0.40322

Illustration 3.4.8 Rejection Region through “Cumulative Binomial” (shaded) when $K=2$, $D = 10$, and $\alpha = 0.05$ for unequal allocation ratio $\rho = 1.5$. For each $(DC = c, D(1) = t)$, the numbers followed by “CB” are the cumulative binomial calculated by $P(D(1) \leq t | DC + D(1) = c + t)$ where $D(1) | DC + D(1) = c + t \sim \text{Bin}(DC + D(1) = c + t, 1/2.5)$ and the number followed by “size” is the probability of $(DC = c, D(1) = t)$ under null hypothesis discussed in section 3.3. For example, In the cell $(DC = 9, D(1) = 0)$, $CB=P(D(1) \leq 0 | DC + D(1) = 9) = 0.01008$ and size = 0.0028 is by multinomial distribution from section 3.3 under null hypothesis for unequal allocation.

From the example of unequal allocation with allocation ratio $\rho = 1.5$ with 2 treatment groups ($K=2$) and followed up until 10 disease cases occur ($D=10$), the surrogate cumulative binomial metric and the probabilities for all possible combinations under null hypothesis could be determined as following Illustration 3.4.8. Again, the second row of each cell is the probability for $D_{(1)}, D_C$ under the null hypothesis for the allocation ratio 1.5 which is discussed in Section 3.3 and the first row in each cell is the cumulative binomial which is calculated by $P(D_{(1)} \leq t | D_C + D_{(1)} = c + t)$ where $D_{(1)} | D_C + D_{(1)} = c + t \sim \text{Bin}(D_C + D_{(1)} =$

$c + t, 1/2.5)$. For the combination ($D_C = 9, D_{(1)} = 0$) (shown in red), the cumulative binomial could be calculated as $P(D_{(1)} \leq 0 | D_C + D_{(1)} = 9) = (1.5/2.5)^9 = 0.01008$.

3.4.5 Optimal Power/Size Ratio Based on a Specified Alternative

Optimal Power/Size Ratio is a metric and based on the value to define the rejection region. This method is denoted as ‘‘PowRatio’’. For each combination of D_C and $D_{(1)}$, we could have two probabilities, one is obtained under the specified alternative hypothesis and the other is the one under null hypothesis. If that probability under specified alternative is much larger than the probability under the null, or more precisely, the ratio of these two probabilities is larger since both values are less than one, we could consider that at least one treatment is better than the control. To illustrate this we will present a table which contains all the values of the ratio of the specified alternative / null probabilities for each combination of D_C and $D_{(1)}$. Hence, we look the ratios and see the combinations which might have bigger power with smaller size. Then, we accumulate the rejection region based on LLQR and this ratio from the largest to smaller.

For a specific alternative, the rejection region defined as power/ratio is larger than a number under size level α . However, the shape of the rejection based on this method could be very different as the alternative changes. Here we will discussed the shape of the rejection region under the alternative ($1:r_1 = r, 1$) or ($1: r_1 = r, r_2 = r$), $r < 1$, for the case of $K=2$. As the alternative is as ($1: r_1 = r, 1$) for $r < 1$, the power could be represent as the following formula:

$$\begin{aligned} P_{H_a}(D_C = c, D_{(1)} = t) &= P_{H_a}(D_C = c, D_{(1)} = t, D_1 \leq D_2) + P_{H_a}(D_C = c, D_{(1)} = t, D_1 > D_2) \\ &= P_{H_a}(D_C = c, D_1 = t, D_1 \leq D_2) + P_{H_a}(D_C = c, D_2 = t, D_1 > D_2) \\ &= p \binom{D}{c \quad t \quad D-c-t} \left(\frac{r}{2+r}\right)^t \left(\frac{1}{2+r}\right)^{D-t} + (1-p) \binom{D}{c \quad D-c-t \quad t} \left(\frac{r}{2+r}\right)^{D-c-t} \left(\frac{1}{2+r}\right)^{c+t}, \end{aligned}$$

where $t \leq D-c-t$ and $1 > p = P_{H_a}(D_1 \leq D_2) > 1-p > 0$ under alternative hypothesis as ($1: r_1 = r, 1$) for $r < 1$ and

which does not depend on t. Similarly, under null hypothesis, $r = 1$, and $p^* = P_{H_0}(D_1 \leq D_2)$,

$$P_{H_0}(D_C = c, D_{(1)} = t) = \left(\frac{1}{3}\right)^D \binom{D}{c \quad t \quad D-c-t}. \text{ The ratio of the Power/Size could be express as}$$

$$\frac{P_{H_a}(D_C = c, D_{(1)} = t)}{P_{H_0}(D_C = c, D_{(1)} = t)} \propto (\text{proportional to}) p \left(\frac{1}{2+r} \right)^D r^t + (1-p) \left(\frac{1}{2+r} \right)^D r^{D-c-t} \propto pr^t + (1-p)r^{D-c-t} \doteq f(t).$$

The derivative of $f(t)$ is $f'(t) = p \ln(r) r^t - (1-p) \ln(r) r^{D-c-t} < \ln(r) (r^t - r^{D-c-t})/2 < 0$ as $\ln(r) < 0$, $r^t - r^{D-c-t} \geq 0$

for $r < 1$ and $t \leq D-c-t$ and $f'(t)$ is decreasing on p and with maximum at $p = 1/2$ (in this cases, p is always larger than $1/2$). Therefore, the ratio of the Power/Size under $K=2$ has following properties under alternative $(1:r,1)$:

- (1) Power/Size ratio is increasing as t decreases for fixed numbers of control, c .
- (2) the ratio is increasing as the numbers of controls, c , increasing given fixed t .
- (3) the ratio is increasing as t decreases for fixed of total numbers of disease cases in the control group and the minimum treatment group $(c+t)$

Hence, for any given $(D_C = c, D_{(1)} = t)$ to be rejected, the lower and left cells which will have larger ratio and tend to have more evident to be rejected. This approach for $K=2$ and alternative is $(1: r_1 = r, 1)$, it satisfies the property LLQR directly. And also, the rejection region tends to accumulated from the small minimum treatment (small t) first and might be similar to the rejection region based on “minimum treatment” since the impact for power/size ratio is more for small minimum treatment than the large control by (1), (2), and (3). However, it is hard to show the rejection region exactly by mathematical formula since it is hard to compare the ratio which has different combinations of D_C and $D_{(1)}$ in some case. Hence, similar to the method of “cumulative binomial”, there is a metric filled with power/size ratio in the table of combinations $(D_C, D_{(1)})$ and the rejection region based on this condition could not be decided as directly as other methods. Instead, the rejection region could be obtained through this metric and LLQR which is shown in the following section. But through the computational results, we found that the rejection region based on this method is similar to the “treatment boundary augmented”.

For special cases of $K=2$ and the alternative is as $(1: r_1 = r, r_2 = r)$ for $r < 1$, the rejection region based on “PowRatio” is exactly the same as the one based on “Use the controls”. The Power/Size ratio is

$$\frac{\binom{D}{D_C \quad D_1 \quad D_2} \left(\frac{1}{1+2r} \right)^{D_C} \left(\frac{r}{1+2r} \right)^{D-D_C}}{\binom{D}{D_C \quad D_1 \quad D_2} \left(\frac{1}{3} \right)^D} = \left(\frac{3}{1+2r} \right)^D \left(\frac{1}{r} \right)^{D_C},$$

and for given r , the ratio is only depends on D_C and is constant for the same D_C . Also, the ratio is larger as D_C increase since r is less than 1 which has the same metric as the one we discussed in “Use the Control”. Hence it is more evidence to get rejection for large D_C . Also, similar to the method of “control boundary augmented”, the accumulated extra cells from the smallest $D_{(1)}$ for given D_C would increase the power. The rejection region based on Power/Size ratio is the same as the one based on “control only” and “control boundary augmented”.

The power ratio could be seen as the likelihood ratio test for a specific alternative. Based on Neyman-Pearson Lemma, the most powerful test for the simple test $H_0: (1:1,1)$ v.s. $H_a:(1: r_1 = r, r_2 = r)$ is the rejection region based on “control” and the most powerful test for test $H_0: (1:1,1)$ v.s. $H_a:(1: r_1 = r, 1)$ is similar to the rejection region based on “minimum treatment”. Hence there is no UMP for the test $H_0: (1:1,1)$ v.s. composite alternative $H_a: (1:r_1, r_2)$ where $r_1, r_2 \leq 1$ with one strict inequality since the shape changes on the different alternative dramatically.

The rejection region for the power ratio approach depends on the alternative hypothesis too much and it is hard to decide which specific alternative hypothesis to pick to construct the rejection region. Hence, we will not consider this method for the rest of the study after this section.

Continuing the example with $(K=2)$ and $(D=10)$ occur under equal allocation $(\rho=1)$. For each possible cell, we show the Power as the top number, size as the middle number and the power/size ratio as the third number. First, for the rejection region under alternative $(1: r_1 = 0.2, 1)$, the ratio gets smaller for larger value of $D_{(1)}$ and smaller D_C . As the result, we could reject the $(D_C, D_{(1)})$ in the shaded area as shown in below (Illustration 3.4.9) which is the same region we had for the “treatment boundary augmented”.

For the alternative of interest is change to $(1: r_1 = 0.2, r_2 = 0.2)$ and other setting remain the same, we could find that the ratio value is the same for the same D_C and the value is larger for larger D_C and is the same metric as the one from “Use the control”. The rejection region based on “PowRatio” and LLQR (as in Illustration 3.4.10) is to reject $(D_C = 10, D_{(1)} = 0)$ first, and then reject the cells as the following order $(D_C = 9, D_{(1)} = 0)$, $(D_C = 8, \text{all } D_{(1)})$, $(D_C = 7, \text{all } D_{(1)})$, and collect the cell, $(D_C=6, D_{(1)} = 0)$, for increasing the power from the boundary. Hence the rejection region for this alternative is the same as the one based on “controls boundary augmented” as shown in the previous paragraph.

$D_c D_{(t)}$	0	1	2	3	4	5
0	Power:0.0004 Size:<0.0001 Ratio:11.1162	Power:0.0008 Size:0.0003 Ratio:2.2232	Power:0.0007 Size:0.0015 Ratio:0.4447	Power:0.0004 Size:0.0041 Ratio:0.0891	Power:0.0001 Size:0.0071 Ratio:0.0185	Power<0.0001 Size:0.0043 Ratio:0.0071
1	Power:0.0038 Size:0.0003 Ratio:11.1162	Power:0.0068 Size:0.0031 Ratio:2.2233	Power:0.0054 Size:0.0122 Ratio:0.4448	Power:0.0026 Size:0.0285 Ratio:0.0896	Power:0.0009 Size:0.0427 Ratio:0.0213	
2	Power:0.0169 Size:0.0015 Ratio:11.1162	Power:0.0271 Size:0.0122 Ratio:2.2234	Power:0.0190 Size:0.0427 Ratio:0.4454	Power:0.0079 Size:0.0854 Ratio:0.0925	Power:0.0019 Size:0.0534 Ratio:0.0356	
3	Power:0.0452 Size:0.0041 Ratio:11.1163	Power:0.0633 Size:0.0285 Ratio:2.2239	Power:0.0383 Size:0.0854 Ratio:0.4482	Power:0.0152 Size:0.1423 Ratio:0.1067		
4	Power:0.0791 Size:0.0071 Ratio:11.1169	Power:0.0950 Size:0.0427 Ratio:2.2268	Power:0.0493 Size:0.1067 Ratio:0.4624	Power:0.0127 Size:0.0711 Ratio:0.1779		
5	Power:0.0949 Size:0.0085 Ratio:11.1197	Power:0.0956 Size:0.0427 Ratio:2.2410	Power:0.0455 Size:0.0854 Ratio:0.5336			
6	Power:0.0792 Size:0.0071 Ratio:11.1340	Power:0.0658 Size:0.0285 Ratio:2.3122	Power:0.0190 Size:0.0213 Ratio:0.8893			
7	Power:0.0455 Size:0.0041 Ratio:11.2051	Power:0.0325 Size:0.0122 Ratio:2.6679				
8	Power:0.0176 Size:0.0015 Ratio:11.5608	Power:0.0068 Size:0.0015 Ratio:4.4465				
9	Power:0.0045 Size:0.0003 Ratio:13.3394					
10	Power:0.0004 Size:<0.0001 Ratio:22.2324					



 : Rejection region based on “PowRatio”
 In each possible cell, the power of corresponding combinations will be located on the top, size will be in the middle, and the Power/Size ratio will be at the third.

Illustration 3.4.9 Rejection Region through “PowRatio” (shaded ) for $K=2$, $D=10$, $\alpha=0.05$ and under specified alternative ($1:r_1=0.2,1$). The power, size, and its ratio are calculated and located as top, middle, and third in each cell.

$D_c \backslash D_{(+)}$	0	1	2	3	4	5
0	Power<0.0001 Size<0.0001 Ratio:0.0002	Power<0.0001 Size:0.0003 Ratio:0.0002	Power<0.0001 Size:0.0015 Ratio:0.0002	Power<0.0001 Size:0.0041 Ratio:0.0002	Power<0.0001 Size:0.0071 Ratio:0.0002	Power<0.0001 Size:0.0043 Ratio:0.0002
1	Power<0.0001 Size:0.0003 Ratio:0.0010	Power<0.0001 Size:0.0031 Ratio:0.0010	Power<0.0001 Size:0.0122 Ratio:0.0010	Power<0.0001 Size:0.0285 Ratio:0.0010	Power<0.0001 Size:0.0427 Ratio:0.0010	
2	Power<0.0001 Size:0.0015 Ratio:0.0052	Power:0.0001 Size:0.0122 Ratio:0.0052	Power:0.0002 Size:0.0427 Ratio:0.0052	Power:0.0004 Size:0.0854 Ratio:0.0052	Power:0.0003 Size:0.0534 Ratio:0.0052	
3	Power:0.0001 Size:0.0041 Ratio:0.0261	Power:0.0007 Size:0.0285 Ratio:0.0261	Power:0.0022 Size:0.0854 Ratio:0.0261	Power:0.0037 Size:0.1423 Ratio:0.0261		
4	Power:0.0009 Size:0.0071 Ratio:0.1307	Power:0.0056 Size:0.0427 Ratio:0.1307	Power:0.0139 Size:0.1067 Ratio:0.1307	Power:0.0093 Size:0.0711 Ratio:0.1307		
5	Power:0.0056 Size:0.0085 Ratio:0.6533	Power:0.0279 Size:0.0427 Ratio:0.6533	Power:0.0558 Size:0.0854 Ratio:0.6533			
6	Power:0.0232 Size:0.0071 Ratio:3	Power:0.0929 Size:0.0285 Ratio:3	Power:0.0697 Size:0.0213 Ratio:3			
7	Power:0.0664 Size:0.0041 Ratio:16	Power:0.1991 Size:0.0122 Ratio:16				
8	Power:0.1245 Size:0.0015 Ratio:82	Power:0.1245 Size:0.0015 Ratio:82				
9	Power:0.1383 Size:0.0003 Ratio:408					
10	Power:0.0346 Size<0.0001 Ratio:2041					



 : Rejection region based on “PowRatio”
 In each possible cell, the power of corresponding combinations will be located on the top, size will be in the middle, and the Power/Size ratio will be at the third.

Illustration 3.4.10 Rejection Region through “PowRatio” (shaded ) for $K=2$, $D=10$, $\alpha=0.05$ and under specified alternative ($1:r_1=0.2, r_2=0.2$). The power, size, and its ratio are calculated and located as top, middle, and third in each cell.

Other example for the alternatives such as ($1:r_1=0.2, r_2=0.5$) and others conditions remain the same would be obtained through Power/Size ratio metric and LLQR. As the table below (Illustration 3.4.11) shows, in this case ($1:r_1=0.2, r_2=0.5$), the rejection region is no longer like “controls boundary augmented” or “treatment boundary augmented” as previous examples.

Again, since the shape of the rejection region based on “PowRatio” depends on the alternative hypothesis and could be quite different; there is no uniformly most powerful test based these two statistics for un-specific alternative for our study. Also, we are looking for a rejection region method that is good for different alternatives. From now on, we only study the rejection regions that do not depend on alternative settings which are discussed in section 3.4.1 to section 3.4.4.

$D_c \backslash D_{+}$	0	1	2	3	4	5
0	Power:<0.0001 Size:<0.0001 Ratio:0.1430	Power:<0.0001 Size:0.0003 Ratio:0.0572	Power:<0.0001 Size:0.0015 Ratio:0.0230	Power:<0.0001 Size:0.0041 Ratio:0.0094	Power:<0.0001 Size:0.0071 Ratio:0.0042	Power:<0.0001 Size:0.0043 Ratio:0.0029
1	Power:0.0001 Size:0.0003 Ratio:0.2861	Power:0.0003 Size:0.0031 Ratio:0.1146	Power:0.0006 Size:0.0122 Ratio:0.0462	Power:0.0006 Size:0.0285 Ratio:0.0195	Power:0.0004 Size:0.0427 Ratio:0.0103	
2	Power:0.0009 Size:0.0015 Ratio:0.5725	Power:0.0028 Size:0.0122 Ratio:0.2298	Power:0.0040 Size:0.0427 Ratio:0.0939	Power:0.0036 Size:0.0854 Ratio:0.0425	Power:0.0016 Size:0.0534 Ratio:0.0293	
3	Power:0.0047 Size:0.0041 Ratio:1.1460	Power:0.0132 Size:0.0285 Ratio:0.4623	Power:0.0166 Size:0.0854 Ratio:0.1948	Power:0.0146 Size:0.1423 Ratio:0.1025		
4	Power:0.0163 Size:0.0071 Ratio:2.2977	Power:0.0401 Size:0.0427 Ratio:0.9388	Power:0.0453 Size:0.1067 Ratio:0.4247	Power:0.0208 Size:0.0711 Ratio:0.2929		
5	Power:0.0395 Size:0.0085 Ratio:4.6235	Power:0.0831 Size:0.0427 Ratio:1.9478	Power:0.0875 Size:0.0854 Ratio:1.0252			
6	Power:0.0668 Size:0.0071 Ratio:9.3875	Power:0.1208 Size:0.0285 Ratio:4.2471	Power:0.0625 Size:0.0213 Ratio:2.9290			
7	Power:0.0792 Size:0.0041 Ratio:19.4780	Power:0.1250 Size:0.0122 Ratio:10.2516				
8	Power:0.0647 Size:0.0015 Ratio:42.4709	Power:0.0446 Size:0.0015 Ratio:29.2903				
9	Power:0.0347 Size:0.0003 Ratio:102.516					
10	Power:0.0050 Size:<0.0001 Ratio:292.903					



 : Rejection region based on “PowRatio”
 In each possible cell, the power of corresponding combinations will be located on the top, size will be in the middle, and the Power/Size ratio will be at the third.

Illustration 3.4.11 Rejection Region through “PowRatio” (shaded ) for $K=2$, $D=10$, $\alpha=0.05$ and under specified alternative ($r_1=0.2, r_2=0.5$). The power, size, and its ratio are calculated and located as top, middle, and third in each cell.

3.5 Accumulated Rejection Region Based on LLQR and its Representation

As discussed in section 3.4, the rejection region using “cumulative binomial” metric can not be obtained from simple way. Here, we introduce a method to accumulate the cells until the size of the region is not exceeded level α based on LLQR. The rejection region is obtained through comparison of its ordering criteria among the candidate cells where are the cells might be sequentially rejected and accumulated in the rejection region based on the current rejection region under LLQR given the cumulative type I error remains less or equal to α . We accumulate the rejection region starting at the cell which is located in the last row and first column.

In this process, the cell $(D_C = D, D_{(1)} = 0)$ where all cases are controls would start (be the first included in) the reject region which is the current rejection region at start. The rejection region would be then grow with new cells added as we continue to add by the following 2 steps until the type I error exceeds α .

1. Given the current rejection region and the criteria metric, we first search for the set of candidate cells which may be accumulated to new reject region in next stage followed by LLQR and then compare the criteria among candidate cells. For the cases of tie, we might pick either from the smallest $D_{(1)}$ or largest D_C depends on the alternative of interests.
2. Computed the size (under H_0) of the new rejection region which combine the new cell chosen in Step 1 and the existing rejection region.
 - a. If the size of new rejection region is less than the target α level we are pursuing, we assume the new rejection region as the current rejection region and go back to Step 1 to add a new cell
 - b. If the size is equal or more than the target α level, then do not add the cell and keep the current rejection region as the final result.

In step 1, in order to maintain the LLQR property; the candidate cells should have two properties: A. the candidate cells to be added must be immediately to the right or immediately above the current rejection region. B. furthermore, the candidate cells cannot have any cells in the same row to the left or any cells in the same column below them which are not in the rejection region. For the case as shown in the Illustration 3.5.1, suppose we have current rejection region shaded as gray, we could find that $\{(D_C = 1, D_{(1)} = 0), (D_C = 3, D_{(1)} = 1), (D_C = 3, D_{(1)} = 2), (D_C = 2, D_{(1)} = 3)\}$ are the cells immediately right or above the current rejection region which are possible cells to be the candidate cells. However, when we take a closer look, if we pick the cell $(D_C = 3, D_{(1)} = 2)$, the rejection region is not followed by LLQR since a cell that is not in the rejection region $(D_C = 3, D_{(1)} = 1)$ is in the same row to the left of $(D_C = 3, D_{(1)} = 2)$. Similarly, $(D_C = 2, D_{(1)} = 1)$ could not be the candidate cell since a cell in the same column below it $(D_C = 3, D_{(1)} = 1)$ should be collected in the reject first. Hence, the candidate cells for increasing the rejection region at this stage are $\{(D_C, D_{(1)}) | (D_C = 1, D_{(1)} = 0), (D_C = 3, D_{(1)} = 1)\}$.

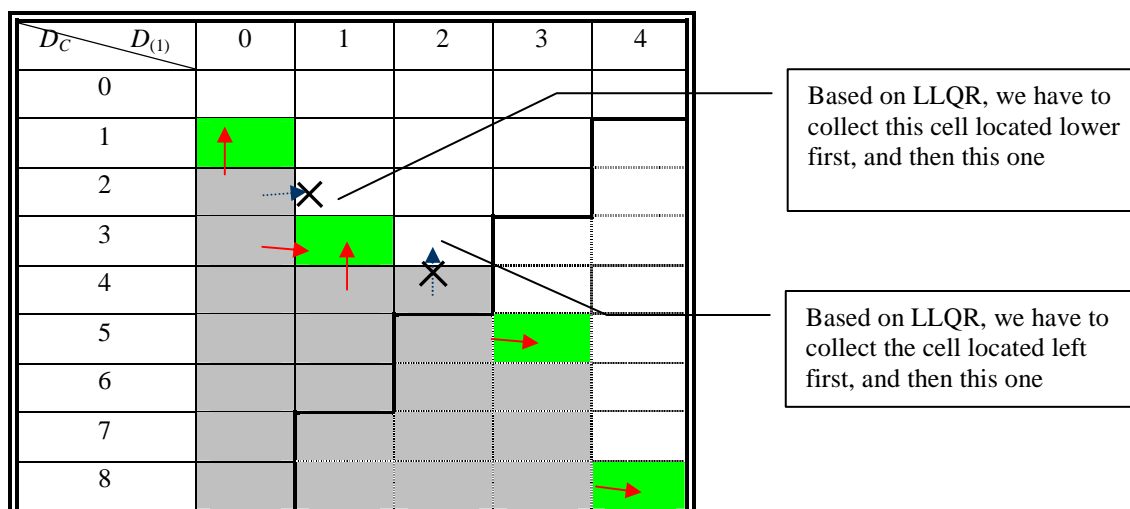


Illustration 3.5.1 Possible Candidate Cells Given Current Rejection Region; gray colored cells are current rejection region, and the green colored cells are the candidates to be accumulated to the new rejection region

Here, we construct a simple algorithm to find the candidate cells based on LLQR which could be programmed through computer directly. The cell directly above a cell in the current rejection region with $D_{(1)} = 0$ will always be considered to be a candidate but for other cells to be the candidate cells should be satisfied two conditions below:

- (a) Consider only cells that are the lowest possible cell in a column that is not in the rejection region .
- (b) But do not consider cells in (a) if there is a cell in the same row to the left of that cell which is not in the current rejection region

Once the cells have been included in (a) and (b) compare all the candidate cells according to the criteria being used and find the one which has the best value. For instance, we would pick the cell satisfying (a) and (b) that has the smallest p-value for “cumulative binomial” metric. If the criteria scores are tied in for two candidates, we could pick the one with the smallest $D_{(1)}$ to break the tie as this more directly suggests that one specific treatment works better than the control. Or alternatively, we could pick the one from the largest D_c as this more directly suggests that all treatments work better than the control, or use some other tie breaking criteria.

Form the previously used example of $K=2$, $D=10$, $\rho=1$ under H_0 and H_a : ($1: r_1 = 0.2, 1$), the cumulative binomial metric being used to form a rejection region, suppose the shaded area is the current rejection region where this current rejection region has been obtained through several iteration of the proposed 2 steps (a) and (b) above. The cells that are the lowest possible cells in each column that are not

in the current rejection region and do not have a cell in the row to the left which is also not in the current region are $\{(D_C = 4, D_{(1)} = 0), (D_C = 6, D_{(1)} = 1)\}$. Pick the cell to be in the new rejection region as the value of criteria metric corresponding to the cells $\{(D_C = 4, D_{(1)} = 0), (D_C=6, D_{(1)} = 1)\}$ with the smallest values.

We found that the values are tied and pick the one with smaller $D_{(1)}$ as the alternative of interests is (1: $r_1 = 0.2, 1$).

$D_C \backslash D_{(1)}$	0	1	2	3	4	5
0	CB=1.00000 size <0.0001	CB=1.00000 Size=0.0003	CB=1.00000 size =0.0015	CB=1.00000 size =0.0041	CB=1.00000 size =0.0071	CB=1.00000 size =0.0043
1	CB=0.50000 size =0.0003	CB=0.75000 size =0.0031	CB=0.87500 size =0.0122	CB=0.93750 size =0.0285	CB=0.96875 size =0.0427	CB=0.98438
2	CB=0.25000 size =0.0015	CB=0.50000 size =0.0122	CB=0.68750 size =0.0427	CB=0.81250 size =0.0854	CB=0.89063 size =0.0534	CB=0.93750
3	CB=0.12500 size =0.0041	CB=0.31250 size =0.0285	CB=0.50000 size =0.0854	CB=0.65625 size =0.1423	CB=0.77344	CB=0.85547
4	CB=0.06250 size =0.0071	CB=0.18750 size =0.0427	CB=0.34375 size =0.1067	CB=0.50000 size =0.0711	CB=0.63672	CB=0.74609
5	CB=0.03125 size =0.0085	CB=0.10938 size =0.0427	CB=0.22656 size =0.0854	CB=0.36328	CB=0.50000	CB=0.62305
6	CB=0.01563 size =0.0071	CB=0.06250 size =0.0285	CB=0.14453 size =0.0213	CB=0.25391	CB=0.37695	CB=0.50000
7	CB=0.00781 size =0.0041	CB=0.03516 size =0.0122	CB=0.08984	CB=0.17188	CB=0.27441	CB=0.38721
8	CB=0.00391 size =0.0015	CB=0.01953 size =0.0015	CB=0.05469	CB=0.11328	CB=0.19385	CB=0.29053
9	CB=0.00195 Size=0.0003	CB=0.01074	CB=0.03271	CB=0.07300	CB=0.13342	CB=0.21198
10	CB=0.00098 size <0.0001	CB=0.00586	CB=0.01929	CB=0.04614	CB=0.08978	CB=0.15088

Illustration 3.5.2 Candidate cells (bolded) for the next rejection region given current rejection region (gray colored)

In the step 2, we could calculate the size of the new rejection region which is the size of the current rejection region plus the probability of the combination $D_C, D_{(1)}$ under H_0 . Since the size of the new rejection region is 0.0424 which is less than 0.05, the new rejection region are considered to the current rejection region plus the chosen candidate cells and continue to the step 1 till the size of the new rejection region is larger than 0.05.

3.6 Robustness of the Different Rejection Regions to Pathological Null and Pathological Alternatives

While one of the cardinal assumptions that is made is none of the new treatments can be worse than the control, it can never be ruled out that a new treatment could be equivalent or worse and this has happened before (Hirschel, 2003).

We refer to settings where a new treatment is worse than the control as “pathological”. For the pathological setting of $(1:1, R)$, where $R > 1$ we call a *pathological null* (as nothing is superior to the control) and the setting of $(1: r, R)$, we call a *pathological alternative* since one treatment is better than the control. In these pathological cases, as R become larger, both the D_C and $D_{(1)}$ are pushed to zero with large probability as all of the cases tend to accumulate to the R treatment arm. Tests based on nulls of $(1:1,1)$ and alternatives of $(1:r_1, r_2)$ fail in pathological settings. While one would expect that the problem of one treatment performing worse than the control would become apparent to investigators either post-hoc or preferably in interim analyses and the extremely bad treatment should be indicated obviously and excluded in the trial. Here, we will discuss the impact of pathological cases focus on $R=2$ (and $R=5$) using the rejection regions based on the previously described methods. Also, we will focus on comparison of two new treatment groups to one control for equal allocation ($\rho = 1$).

3.6.1 Robustness of Rejection Regions to False Positive Findings in the Pathological Null Setting

For $K=2$, pathological nulls are single $(1:1, R)$ for $R>1$, i.e. there is one treatment which is worse than the control or double $(1: R_1, R_2)$ for $1 < R_1 \leq R_2$, i.e. both treatments are worse than the control. For the double pathological null $(1: R_1, R_2)$ where $R_1 \approx R_2$, D_C will tend to be smaller and $D_{(1)}$ will tend to be larger than for the null $(1:1,1)$ which pushes the distribution away from the rejection region for any method considered here. For $R_1 \ll R_2$ (R_1 is much smaller than R_2) the setting becomes similar to $(1:1, R)$. Hence, we focus on comparing two treatments ($K=2$) to the control under the single pathological null $(1:1, R)$ for $R > 1$ in the examples and check if there are rejection regions that exceed the requested size level (i.e. make it falsely appear a new treatment is successful).

For comparing the mass of D_C and $D_{(1)}$ under $(1:1, R)$ to this mass under $(1:1,1)$, the mass under $(1:1, R)$ tends to have higher probability to be located towards smaller minimum treatment (as well as fewer controls). Hence rejection regions that focus on small minimum treatments might have larger size than the required level α . Rejection regions based on “minimum treatment only” or “minimum treatment boundary

augmented” reject small minimum treatment numbers (with or without small control numbers) might exceed the nominal α level. On the other hand, the rejection region based on “controls only” could always controls the type I error under pathological null since the conditional distribution of $D_C | D \sim \text{Binomial}(D, 1/(2+R))$ and the probability to reject the pathological null based on the size- α test under regular null of (1:1,1) is $P_{\text{pathological}}(D_C > c | D) \leq P_{\text{regular}}(D_C > c | D) \leq \alpha$ as the conditional distribution under regular null hypothesis is $D_C | D \sim \text{Binomial}(D, 1/3)$ and $1/(2+R) < 1/3$.

$D_{(1)} \backslash D_C$	0	1	2	3	4	5	6	7
0	$P_1 < 0.0001$ $P_2 < 0.0001$	$P_1 < 0.0001$ $P_2 = 0.0002$	$P_1 < 0.0001$ $P_2 = 0.0008$	$P_1 = 0.0001$ $P_2 = 0.0017$	$P_1 = 0.0002$ $P_2 = 0.0026$	$P_1 = 0.0004$ $P_2 = 0.0030$	$P_1 = 0.0007$ $P_2 = 0.0027$	$P_1 = 0.0009$ $P_2 = 0.0023$
1	$P_1 < 0.0001$ $P_2 = 0.0002$	$P_1 < 0.0001$ $P_2 = 0.0016$	$P_1 = 0.0002$ $P_2 = 0.0052$	$P_1 = 0.0008$ $P_2 = 0.0105$	$P_1 = 0.0021$ $P_2 = 0.0145$	$P_1 = 0.0042$ $P_2 = 0.0152$	$P_1 = 0.0063$ $P_2 = 0.0134$	$P_1 = 0.0036$ $P_2 = 0.0061$
2	$P_1 < 0.0001$ $P_2 = 0.0008$	$P_1 = 0.0002$ $P_2 = 0.0052$	$P_1 = 0.0011$ $P_2 = 0.0157$	$P_1 = 0.0042$ $P_2 = 0.0289$	$P_1 = 0.0105$ $P_2 = 0.0369$	$P_1 = 0.0188$ $P_2 = 0.0362$	$P_1 = 0.0251$ $P_2 = 0.0322$	
3	$P_1 = 0.0001$ $P_2 = 0.0017$	$P_1 = 0.0008$ $P_2 = 0.0104$	$P_1 = 0.0042$ $P_2 = 0.0288$	$P_1 = 0.0140$ $P_2 = 0.0485$	$P_1 = 0.0314$ $P_2 = 0.0571$	$P_1 = 0.0502$ $P_2 = 0.0537$	$P_1 = 0.0293$ $P_2 = 0.0251$	
4	$P_1 = 0.0002$ $P_2 = 0.0026$	$P_1 = 0.0021$ $P_2 = 0.0143$	$P_1 = 0.0105$ $P_2 = 0.0361$	$P_1 = 0.0314$ $P_2 = 0.0554$	$P_1 = 0.0628$ $P_2 = 0.0604$	$P_1 = 0.0879$ $P_2 = 0.0564$		
5	$P_1 = 0.0004$ $P_2 = 0.0029$	$P_1 = 0.0042$ $P_2 = 0.0144$	$P_1 = 0.0188$ $P_2 = 0.0327$	$P_1 = 0.0502$ $P_2 = 0.0456$	$P_1 = 0.0879$ $P_2 = 0.0470$	$P_1 = 0.0527$ $P_2 = 0.0226$		
6	$P_1 = 0.0007$ $P_2 = 0.0024$	$P_1 = 0.0063$ $P_2 = 0.0108$	$P_1 = 0.0251$ $P_2 = 0.0222$	$P_1 = 0.0586$ $P_2 = 0.0282$	$P_1 = 0.0879$ $P_2 = 0.0282$			
7	$P_1 = 0.0009$ $P_2 = 0.0015$	$P_1 = 0.0072$ $P_2 = 0.0062$	$P_1 = 0.0251$ $P_2 = 0.0114$	$P_1 = 0.0502$ $P_2 = 0.0134$	$P_1 = 0.0314$ $P_2 = 0.0067$			
8	$P_1 = 0.0009$ $P_2 = 0.0008$	$P_1 = 0.0063$ $P_2 = 0.0028$	$P_1 = 0.0188$ $P_2 = 0.0045$	$P_1 = 0.0314$ $P_2 = 0.0050$				
9	$P_1 = 0.0007$ $P_2 = 0.0003$	$P_1 = 0.0042$ $P_2 = 0.0010$	$P_1 = 0.0105$ $P_2 = 0.0014$	$P_1 = 0.0070$ $P_2 = 0.0007$				
10	$P_1 = 0.0004$ $P_2 = 0.0001$	$P_1 = 0.0021$ $P_2 = 0.0003$	$P_1 = 0.0042$ $P_2 = 0.0003$					
11	$P_1 = 0.0002$ $P_2 < 0.0001$	$P_1 = 0.0008$ $P_2 = 0.0001$	$P_1 = 0.0006$ $P_2 < 0.0001$					
12	$P_1 = 0.0001$ $P_2 < 0.0001$	$P_1 = 0.0002$ $P_2 < 0.0001$						
13	$P_1 < 0.0001$ $P_2 < 0.0001$	$P_1 < 0.0001$ $P_2 < 0.0001$						
14	$P_1 < 0.0001$ $P_2 < 0.0001$							
15	$P_1 < 0.0001$ $P_2 < 0.0001$							

About 60% of mass under pathological null of (1:1,2) in the bolded area

Area of 60% mass under the null of (1:1,1) in red area

Illustration 3.6.1 Probability mass for $D = 15$. The top probability P_1 in the cell is the probability mass under regular null hypothesis (1:1,1) while the second probability P_2 in the cell is the probability mass under pathological hypothesis (1:1,2). 60% of the probability mass for each hypothesis is marked as bolded (red colored is under regular null and the black colored is under pathological null)

For example consider $K=2$, size level 0.05 and follow up until $D = 15$. We first present the distribution mass of $D_C, D_{(1)}$ under null hypothesis (1:1,1) and then the distribution mass of $D_C, D_{(1)}$ under

the pathological null hypothesis (1:1,2). The Illustration 3.6.1 shows the probability mass for $D_C, D_{(1)}$ under both the null hypothesis of (1:1,1) and pathological null hypothesis of (1:1,2). The top number, P_1 , in each cell is the probability mass that is under null hypothesis (1:1,1) and the second number, P_2 , in each cell is the probability mass under pathological null (1:1,2). The area of highest density containing ~60% of mass is bolded in black under pathological null while the region of highest density with ~60% mass from (1:1,1) is bolded in red.

Under pathological null hypothesis of (1:1,2), mass concentration is more spread than the one under the null hypothesis (1:1,1) and is shifted towards less controls and less minimal treatment. The probability mass is higher in the small control and small minimum treatment than the ones from the null hypothesis.

But what is most important in terms of size is the probability under pathological null hypothesis of (1:1,2) for $D_C, D_{(1)}$ to fall into a given α level (say $\alpha=0.05$) rejection region for (1:1,1). If this probability to reject H_0 under the pathological (1:1,2) is (substantially) larger than 0.05 for a given method, then rejection region based on that method is failure since the type I error is not well controlled under a pathological null.

Method of Rejection Region applied to a rejection region based on (1:1,1)	Probabilities to reject (1:1,2)
Control Only	0.0042
Control Boundary Accumulated	0.0077
Minimum Treatment Only	0.0807 > 0.05
Minimum Treatment Boundary Accumulated	0.0810 > 0.05
Difference	0.0085
Difference Boundary Accumulated	0.0172
Cumulative Binomial	0.0200

Table 3.6.1 Probability to reject the pathological null (1:1,2) using the $\alpha=0.05$ level on a rejection region based on regular null hypothesis (1:1,1)

The Table 3.6.1 presents probabilities to reject the pathological null (1:1,2) using an $\alpha=0.05$ rejection region for (1:1,1) for $D=15$. We find for this example, except the methods based on “minimum treatment only” and “minimum treatment boundary accumulated”, other methods have smaller probability to reject under the pathological null hypothesis of (1:1,2) than $\alpha=0.05$ with the (1:1,1) based region and in fact smaller than the probability to reject under the null hypothesis of (1:1,1 (Data not shown)). The methods based on the “minimum treatment only” and “minimum treatment boundary accumulated” are

larger than the required level 0.05. Hence, the (1:1,1) rejection region based on “minimum treatment only” or “minimum treatment boundary” is not robust to the pathological null hypothesis.

The Table below (Table 3.6.2) gives actual probabilities to reject the null hypothesis under pathological null hypotheses (1:1,2) and (1:1,5) for $\alpha=0.01$, $\alpha=0.025$ and $\alpha=0.05$ using the minimum treatment only which deterministically is at least as large as the minimum treatment boundary accumulated. In all of these settings the other methods maintain the nominal Type 1 error (data not shown). For the rejection region using “minimum treatment only”, most sizes exceeded nominal type I error and this type-1 error was increasing with D . Also, we found that as R as large as 5 and the follow up size until 30, the probability to reject the null hypothesis is larger than 35% no matter which nominal size level is. This results in high probability to find the false positive and is not good for our design as using “minimal treatment only” or “minimal treatment boundary augmented” to construct our rejection region.

D	$\alpha=0.01$		$\alpha=0.025$		$\alpha=0.05$	
	(1:1,2)	(1:1,5)	(1:1,2)	(1:1,5)	(1:1,2)	(1:1,5)
15	0.0134	0.0990	0.0134	0.0990	0.0807	0.3466
30	0.0375	0.3617	0.0979	0.5691	0.0979	0.5691
60	0.0859	0.7684	0.1476	0.8593	0.2316	0.9211
90	0.1653	0.9501	0.2356	0.9723	0.3189	0.9855
120	0.2328	0.9898	0.3821	0.9973	0.4650	0.9987

Table 3.6.2 Size based on the rejection region using “minimal treatment only” for its corresponding follow up cases and the pathological null under nominal level α equals to 0.01 and 0.025

Again, the reason that the rejection probabilities using minimum treatment rejection regions derived from (1:1,1) exceed the nominal α when the pathological null hypothesis (1:1, R) hold is that $D_{(1)}$ tends to be smaller for pathological null hypothesis: (1:1, R) compared to the null hypothesis: (1:1,1). The method based on “control only” or “control boundary augmented” does not suffer this problem since it tends to have less observed controls for pathological null hypothesis compared to the null which pushes further away from the rejection region. The rejection regions based on methods of “Difference only”, “Difference boundary augmented” and “Cumulative Binomial” tend to reject the small $D_{(1)}$ with large D_C and while we could not find a universal proof for this observation due to discreteness of the binomial, we have seen empirically that those probabilities under (1:1,2), (1:1,5) (and for other (1:1, R)) tend to be smaller than are the ones under (1:1,1). Hence, the probabilities to fall into (1:1,1) rejection regions based on “control only”, “control boundary augmented”, “difference only”, “difference boundary augmented”,

and “cumulative binomial” under the pathological nulls $(1:1, R)$ are typically smaller than the nominal alpha α .

3.6.2 Robustness of the Rejection Region to Pathological Alternatives

A reasonable rejection region is to be an unbiased test, i.e. the power of the test should be much greater than the type I error. We have shown that under the regular hypothesis testing problems, the rejection region satisfying LLQR will be an unbiased test. For regular alternative under null hypothesis $(1:1,1)$ and $K=2$, the rejection regions based on the methods we proposed in section 3.4 are unbiased and is shown in Appendix A. However, the unbiased structure might be failed for pathological alternative and should be aware of before the analysis.

Here, we are looking for a rejection region with high probability to reject the null under the alternatives. Here, we look at cases for $K=2$ under alternative $(1: r_1 = r, 1)$, $r < 1$ and pathological alternatives $(1: r_1 = r, R)$ for $r < 1$ and $R > 1$. The power of different rejection method should be high for small r . In the case of $(1: r_1 = r, 1)$, most probabilities tend to be moved towards a smaller minimum treatment ($D_{(1)}$) while D_C tends to be larger making any of the rejection regions perform well in terms of power. In the case of pathological alternatives $(1: r_1 = r, R)$ for $R > 1$, D_C tends to be smaller (when $r+R > 2$) as $D_{(1)}$ tends to be even smaller. The rejection region which tends to reject on large controls might suffer from small power while rejection regions that consider the minimum treatment perform well.

For the rejection region based on “control only”, $R_{\alpha-c}$, the probability to reject any treatment is smaller than the type I error in the pathological alternative $(1: r_1 = r, R)$ if $r+R > 2$ (which means $1+r+R > 3$) since the distribution of $D_C \sim \text{Bin}(D, 1/3)$ under the null hypothesis and $D_C \sim \text{Bin}(D, 1/(1+r+R))$ under the pathological alternative $(1: r_1 = r, R)$ and the probability of $D_C > c$ is less under pathological alternatives as $r + R > 2$ than the probability under the null hypothesis no matter how large the follow up cases D is. As the result, the rejection method based on “control only” is biased against rejection when there is a treatment that performs better than the control. For the rejection region using “control boundary augmented”, the probability to reject any treatment is larger than the “control only” rejection region but does not substantially increase power.

$D_C \backslash D_{(1)}$	0	1	2	3	4	5	6	7
0	$P_1 < 0.0001$ $P_2 = 0.0014$	$P_1 < 0.0001$ $P_2 = 0.0010$	$P_1 < 0.0001$ $P_2 = 0.0004$	$P_1 < 0.0001$ $P_2 = 0.0001$	$P_1 < 0.0001$ $P_2 < 0.0001$	$P_1 < 0.0001$ $P_2 < 0.0001$	$P_1 < 0.0001$ $P_2 < 0.0001$	$P_1 < 0.0001$ $P_2 < 0.0001$
1	$P_1 = 0.0002$ $P_2 = 0.0105$	$P_1 = 0.0003$ $P_2 = 0.0073$	$P_1 = 0.0002$ $P_2 = 0.0024$	$P_1 = 0.0001$ $P_2 = 0.0005$	$P_1 < 0.0001$ $P_2 = 0.0001$	$P_1 < 0.0001$ $P_2 < 0.0001$	$P_1 < 0.0001$ $P_2 < 0.0001$	$P_1 < 0.0001$ $P_2 < 0.0001$
2	$P_1 = 0.0015$ $P_2 = 0.0367$	$P_1 = 0.0020$ $P_2 = 0.0238$	$P_1 = 0.0012$ $P_2 = 0.0071$	$P_1 = 0.0004$ $P_2 = 0.0013$	$P_1 = 0.0001$ $P_2 = 0.0002$	$P_1 < 0.0001$ $P_2 < 0.0001$	$P_1 < 0.0001$ $P_2 < 0.0001$	
3	$P_1 = 0.0067$ $P_2 = 0.0794$	$P_1 = 0.0080$ $P_2 = 0.0477$	$P_1 = 0.0044$ $P_2 = 0.0131$	$P_1 = 0.0015$ $P_2 = 0.0022$	$P_1 = 0.0003$ $P_2 = 0.0002$	$P_1 = 0.0001$ $P_2 < 0.0001$	$P_1 < 0.0001$ $P_2 < 0.0001$	
4	$P_1 = 0.0200$ $P_2 = 0.1191$	$P_1 = 0.0220$ $P_2 = 0.0655$	$P_1 = 0.0110$ $P_2 = 0.0164$	$P_1 = 0.0033$ $P_2 = 0.0025$	$P_1 = 0.0007$ $P_2 = 0.0002$	$P_1 = 0.0001$ $P_2 < 0.0001$		
5	$P_1 = 0.0441$ $P_2 = 0.1310$	$P_1 = 0.0441$ $P_2 = 0.0655$	$P_1 = 0.0198$ $P_2 = 0.0147$	$P_1 = 0.0053$ $P_2 = 0.0020$	$P_1 = 0.0009$ $P_2 = 0.0002$	$P_1 = 0.0001$ $P_2 < 0.0001$		
6	$P_1 = 0.0735$ $P_2 = 0.1092$	$P_1 = 0.0661$ $P_2 = 0.0491$	$P_1 = 0.0264$ $P_2 = 0.0098$	$P_1 = 0.0062$ $P_2 = 0.0011$	$P_1 = 0.0010$ $P_2 = 0.0001$			
7	$P_1 = 0.0945$ $P_2 = 0.0702$	$P_1 = 0.0756$ $P_2 = 0.0281$	$P_1 = 0.0265$ $P_2 = 0.0049$	$P_1 = 0.0053$ $P_2 = 0.0005$	$P_1 = 0.0007$ $P_2 < 0.0001$			
8	$P_1 = 0.0945$ $P_2 = 0.0351$	$P_1 = 0.0661$ $P_2 = 0.0123$	$P_1 = 0.0199$ $P_2 = 0.0018$	$P_1 = 0.0036$ $P_2 = 0.0002$				
9	$P_1 = 0.0735$ $P_2 = 0.0137$	$P_1 = 0.0441$ $P_2 = 0.0041$	$P_1 = 0.0111$ $P_2 = 0.0005$	$P_1 = 0.0015$ $P_2 < 0.0001$				
10	$P_1 = 0.0441$ $P_2 = 0.0041$	$P_1 = 0.0221$ $P_2 = 0.0010$	$P_1 = 0.0048$ $P_2 = 0.0001$					
11	$P_1 = 0.0200$ $P_2 = 0.0009$	$P_1 = 0.0081$ $P_2 = 0.0002$	$P_1 = 0.0012$ $P_2 < 0.0001$					
12	$P_1 = 0.0067$ $P_2 = 0.0002$	$P_1 = 0.0022$ $P_2 < 0.0001$						
13	$P_1 = 0.0016$ $P_2 < 0.0001$	$P_1 = 0.0003$ $P_2 < 0.0001$						
14	$P_1 = 0.0002$ $P_2 < 0.0001$							
15	$P_1 < 0.0001$ $P_2 < 0.0001$							

~60% mass in this area under (1: $r_1 = 0.1, R = 2$)

About 54% mass in this area under (1: $r_1 = 0.1, 1$)

Illustration 3.6.2 Probability mass for $D = 15$. The top probability P_1 in the cell is the probability mass under regular alternative hypothesis (1: $r_1 = 0.1, 1$) while the second probability P_2 in the cell is the probability mass under pathological alternative (1: $r_1 = 0.1, R = 2$). 60% of the probability mass for each hypothesis is marked as bolded (red colored is under regular alternative and the black colored is under pathological alternative)

Continuing the example with $K=2$, size level 0.05 and follow up until $D=15$ with a regular alternative (1: $r_1 = 0.1, 1$) or pathological alternative (1: $r_1 = 0.1, R = 2$), the ideal rejection region methods should have high power to reject both alternatives. In Illustration 3.6.2, P_1 is the probability mass of D_C , $D_{(1)}$ under alternative (1: $r_1 = 0.1, 1$) while P_2 is the probability mass under pathological alternative (1: $r_1 = 0.1, R = 2$). The area of highest 60% probability mass is put in bold for the probability under pathological alternative (1: $r_1 = 0.1, R = 2$) and the corresponding area of highest 54% of probability mass from the non-pathological alternative (1: $r_1 = 0.1, 1$) is colored in red which is in the small minimum treatment and in the middle of the control. Comparing the alternative of (1: $r_1 = 0.1, R = 2$) to (1: $r_1 = 0.1, 1$), the area with approximated 60% major probability mass shifts to the up and right, i.e. smaller D_C and smaller $D_{(1)}$ and the

probability mass become much smaller for large D_C . In this case, the rejection region using the “control only” or “control boundary augmented” becomes less favorable under the pathological alternative. Therefore, the power is much smaller under the alternative of $(1: r_1 = 0.1, R = 2)$ than for $(1: r_1 = 0.1, 1)$ when using “control only” or “control boundary augmented” as the rejection method. Based on the rejection region using “control boundary augmented”, $R_{\alpha\text{-CBA}} = \{(D_C, D_{(1)}) \mid D_C > 8 \text{ or } (D_{(1)} \leq 1 \text{ as } D_C = 8)\}$, it only covers very few major probability mass while the rejection region using “minimal treatment boundary augmented”, $R_{\alpha\text{-TBA}} = \{(D_C, D_{(1)}) \mid D_{(1)} \leq 1 \text{ or } (D_C > 9 \text{ as } D_{(1)} = 2)\}$, covers most major probability mass under pathological alternative. As the result, the rejection region based on “minimal treatment” tends to have higher power as the rejection region using “control” tends to have smaller power than other methods.

We computed the probability to reject any treatment under the non-pathological alternative $(1: r_1 = 0.1, 1)$ and the pathological alternative for the same $r_1 = 0.1$ and $R = 2$, i.e., $(1: r_1 = 0.1, R = 2)$ for size level $\alpha = 0.05$ and null hypothesis $(1: 1, 1)$ (Table 3.6.3). For $D = 15$ and alternative $(1: r_1 = 0.1, 1)$, we found that the rejection region based on “control boundary augmented” results in much lower probability comparing to other rejection methods. As pathological alternative $(1: r_1 = 0.1, R = 2)$, the probability to reject any treatment under the “control boundary augmented” is 0.0722 which is not far from level 0.05. For other rejection methods, we found that the small probability to reject under $(1: r_1 = 0.1, R = 2)$ as using the rejection region based on “difference only”, “difference boundary augmented” and “cumulative binomial” since it is caused by not enough follow up cases. Hence, for $D = 30$, we found that probability to reject null under $(1: r_1 = 0.1, R = 2)$ increase for the rejection region based on “difference only”, “difference boundary augmented”, and “cumulative binomial”. However, the probabilities to reject null under $(1: r_1 = 0.1, R = 2)$ do not increase as using rejection region based on “control boundary augmented”. Hence, we will do more analysis of the rejection region using “control only” (as shown that it is biased test, and the probability is always less than the size level) and “control boundary augmented” (the probability to reject remain low in this two follow up cases).

Rejection methods	Probability to reject null under (1:0.1,1)	Probability to reject null under (1:0.1,2)
<i>D</i> = 15		
Control Only	0.2415	0.0249
Control boundary augmented	0.4021	0.0722
Minimum treatment Only	0.8421	0.9172
Minimum treatment boundary augmented	0.8474	0.9174
Difference Only	0.4951	0.1424
Difference boundary augmented	0.6441	0.2797
Cumulative binomial	0.6882	0.4107
<i>D</i> = 30		
Control Only	0.4677	0.0327
Control boundary augmented	0.6100	0.0705
Minimum treatment Only	0.9871	0.9975
Minimum treatment boundary augmented	0.9923	0.9976
Difference Only	0.9125	0.5266
Difference boundary augmented	0.9440	0.6619
Cumulative binomial	0.9679	0.8419

Table 3.6.3 Probabilities to be reject under the hypothesis of (1: $r_1 = 0.1$, $R = 2$) and under (1: $r_1 = 0.1, 1$) based on different types of rejection region for $D=15$ and $D=30$ for size level $\alpha = 0.05$

Although the probability to reject using “control boundary augmented” increases to 61% under (1: $r_1 = 0.1, 1$), the probability to reject null under pathological alternative (1: $r_1 = 0.1$, $R = 2$) is still as low as 0.0705 when using “control boundary augmented” for the rejection region method comparing to the one, 0.0722, as $D=15$. As the result, the rejection method using “control boundary augmented” is hard to achieve the required power under the pathological alternative (1: $r_1 = 0.1$, $R = 2$). From the table below (Table 3.6.4), we could find that under the pathological alternatives of (1: $r_1 = 0.1$, $R = 2$) or (1: $r_1 = 0.1$, $R = 5$), the power under the rejection region using “control only” and “control boundary augmented” does not increase and remains very low no matter how large the follow up cases are. When the follow up cases are larger than 60 and $R=5$, the probability to reject pathological alternative (1: $r_1 = 0.1$, $R = 5$) is very small and although there is one treatment works much better than the control it is hard to determine the treatment which is better than the control.

<i>D</i> ($\alpha=0.05$)	Control only			Control boundary augmented		
	(1:0.1,1)	(1:0.1,2)	(1:0.1,5)	(1:0.1,1)	(1:0.1,2)	(1:0.1,5)
15	0.2415	0.0249	0.0002	0.4021	0.0722	0.0011
30	0.4677	0.0327	≈ 0	0.6100	0.0705	0.0001
60	0.7031	0.0264	≈ 0	0.7860	0.0472	≈ 0
90	0.8711	0.0300	≈ 0	0.9105	0.0481	≈ 0
120	0.9191	0.0190	≈ 0	0.9434	0.0296	≈ 0

Table 3.6.4 Probability to reject under the alternative (1: $r = 0.1, 1$) and two pathological alternatives (1: $r_1 = 0.1$, $R = 2$) and (1: $r_1 = 0.1$, $R = 5$) based on the rejection region using “control only” and “control boundary augmented” for size level 0.05 and follow up cases 15, 30, 60, 90, or 120

In this section we have shown that rejection regions based on “control only” and “control boundary augmented”, are problematic under pathological alternative hypotheses while rejection regions based on “minimum treatment only”, and “minimum treatment boundary augmented” are problematic for pathological null hypotheses. Other rejection region methods based on “difference boundary augmented” and “cumulative binomial” appear to be robust methods for different type of null hypotheses and alternative hypotheses including pathological. From now on, we focus on the rejection methods using “difference boundary augmented”, $R_{\alpha\text{-DBA}}$, and “cumulative binomial”, $R_{\alpha\text{-CB}}$, for the remaining study since the power of rejection method based on “difference only” is always smaller comparing to the method based on “difference boundary augmented”, $R_{\alpha\text{-DBA}}$, we only focus on the later.

3.7 Summary

In this chapter, we have proposed a method to construct the rejection region for testing the global hypothesis to find if any treatment is better than the control. The probability mass of $(D_C, D_{(1)})$ is computed using stochastic matrices and the statistic $(D_C, D_{(1)})$ is used to construct two dimensional rejection regions based on several criteria metrics combining with LLQR properties. For $K = 2$, the rejection regions based on LLQR is unbiased as is proved in Appendix A. We believe this unbiased property remains true for $K > 2$ although we could not show it analytically. Since there is no UMP test for all possible alternatives, we addressed several criteria metrics based on 1) large numbers of control cases; 2) small numbers of cases in the minimum new treatment; 3) large differences between the number of control cases and the number of cases in the minimal new treatment group; 4) a small cumulative binomial metric based on the number of control cases and minimal treatment group cases in addition to LLQR to construct rejection regions. And we studied whether any of the criteria metric we proposed failed to control overall type I error or resulted in smaller power under pathological settings where some of the new treatments were worse than the control..

The pathological null and pathological alternative hypothesis cases are presented in section 3.6. The overall type I error seems to be controlled and the power remain high for the rejection regions based on the criteria “difference boundary augmented” or “cumulative binomial” metrics under pathological cases. Other rejection methods such as using controls will result in lower power for pathological alternative or such as using minimal new treatment that failure to control type I error for pathological null are not recommended to apply to this type of design study. From now on, we will only recommend using these

criteria “difference boundary augmented” and “cumulative binomial” metric for global hypothesis testing for comparative Poisson trial and will only consider these criteria in subsequent section . In the next chapter, we will propose a post-hoc testing procedure to find which treatment(s) is(are) better than the control as comparing two new treatments to one control.

Chapter 4 METHODS TO TEST WHICH TREATMENT(S) IS (ARE) BETTER THAN THE CONTROL

As discussed in the section 3.1, let r_k denote the ratio of the incident rate in the new treatment group T_k to the incident rate in the control group. Our interest would be to find if there is any new treatment that is superior to the control. The global hypothesis is expressed as

$$H_0: r_1 = r_2 = \dots = r_K = 1 \text{ vs } H_a: \text{some of } r_k \text{ are less than 1 for } k=1, \dots, K$$

When the null hypothesis that no treatment is better than the control is rejected, the next immediate question in the minds of investigators would be “which of the new treatments are better than the control” and “how to determine this”. This is equivalent to testing each individual hypothesis:

$$H_{01}: r_1 = 1 \text{ vs } H_{a1}: r_1 < 1,$$

$$H_{02}: r_2 = 1 \text{ vs } H_{a2}: r_2 < 1,$$

$$\vdots \qquad \qquad \qquad \vdots$$

$$H_{0K}: r_K = 1 \text{ vs } H_{aK}: r_K < 1.$$

In this chapter, we want to identify a procedure to decide which specific treatment(s) is (are) better than the control for the multiple comparative Poisson when the global hypothesis has been rejected. Methods to do this developed for comparing multiple new treatments to a control based on independent normal outcomes can't be applied directly to this setting as unlike the normal setting because the constrained number of disease cases sum to D , the outcomes here are not independent.

As before, our focus here will be on the comparison of two treatments ($K=2$) to the control. We will discuss one approach which would control overall type I error at level α for $K=2$ no matter what combination of true/false H_{01} and H_{02} is. Then, we will compare this approach to Bonferroni adjustment of individual test in the absence of a global comparison.

4.1 Fisher's L.S.D Adapted Approach (FLSDA)

Let d_C be the number of disease cases in the control group, d_1 and d_2 be the numbers of disease cases in the new treatment 1 and new treatment 2 groups, respectively. Let $d_{(1)} = \min(d_1, d_2)$. We adapted the concept from Fisher's Least Significant Difference for the normal distribution problem (Hsu 1996) to the multiple comparative Poisson model. We call this approach FLSDA (the “A” being for “Adapted”) and do this test as the following two steps for level α .

Step one: Test the global hypothesis, i.e. test the hypothesis:

$$H_0: r_1 = r_2 = 1 \text{ vs } H_a: \text{some of } r_k \text{ are less than 1 for } k=1,2.$$

The rejection region of this global test was discussed in Section 3.4, 3.5, and 3.6. If the observed $(D_C = d_C, D_{(1)} = d_{(1)})$ is not rejected in this step, we stop the test and conclude that there is no new treatment that is superior to the control. If the observed $(D_C = d_C, D_{(1)} = d_{(1)})$ is rejected in step one, we continue to step two to decide which (if any) of the new treatments can be deemed superior to the control.

Step Two: Test each individual hypothesis at the same α as the overall test

$$H_{01}: r_1 = 1 \text{ vs } H_{a1}: r_1 < 1$$

$$H_{02}: r_2 = 1 \text{ vs } H_{a2}: r_2 < 1$$

Reject the null hypothesis H_{0k} if p-value of the k th hypothesis, p_k , is less than or equal to α where $p_k = P(D_k \leq d_k \mid D_C + D_k = d_C + d_k)$, the conditional distribution of D_k given $D_C + D_k = d_C + d_k$ under null hypothesis H_{0k} is Binomial($d_C + d_k, 1/2$) for equal allocation and Binomial($d_C + d_k, 1/\rho + 1$) for unequal allocation with allocation ratio ρ . If we reject H_{0k} in this step, we conclude that the k th new treatment is better than the control. In most case, we consider balanced design, i.e. equal allocation to each new treatment group and the control group. We will focus on equal allocation in this section.

For $K=2$, FLSDA approach could control type I error with any combination of true/false H_{01} and H_{02} . This is because: (1) when H_{01} and H_{02} are both true, the type I error = Prob(Reject new treatment 1 or new treatment 2) under the setting (1:1,1) is less or equal to α by step one. (2) when H_{01} is true but H_{02} is false, the type I error = P(Reject new treatment 1) is less than or equal to α by step two. (3) similarly when H_{01} is false but H_{02} is true, the type I error = P(Reject new treatment 2) which is less than or equal to α by step two. (4) when H_{01} and H_{02} are false, the type I errors for each new treatment are undefined since there will be no false rejection of H_{01} and H_{02} .

For example with $K=2, D=15$ with size level 0.05 and equally allocate if we observed $D_1 = 2, D_2=5$, and $D_C = 8$ as we end the trial. Here, the rejection region accumulates through conditional binomial metric discussed in section 3.4.4 for the global test and the rejection region (shaded in gray) is shown as following (Illustration 4.1.1). In step one, the observed combination $(D_C = 8, D_{(1)}=2)$ does not reject the global null hypothesis. Hence, we stop and do not conclude that any of the new treatments are better than as the control.

$D_C \backslash D_{(1)}$	0	1	2	3	4	5	6	7
0								
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								

Illustration 4.1.1 Rejection Region (shaded in gray) in the step one of FLSDA with $\alpha=0.05, K=2, D=15$

Assume now we observed $D_C = 9, D_1 = 2$, and $D_2 = 4$ in the end of trial in the previous example. In step one, the observed $(D_C = 9, D_{(1)}=2)$ rejects the global hypothesis. Then, we continue to step two and compare individual treatment to the control. The p-value of $(D_C = 9, D_1=2)$ for individual test H_{01} could be

calculate as $p_1 = \sum_{D_1=0}^2 \binom{11}{D_1} \left(\frac{1}{2}\right)^{11} = 0.0327$ which is less than $\alpha = 0.05$. We conclude that the new treatment 1

is better than the control. For the new treatment 2, the p-value of $(D_C = 9, D_1 = 4)$ for the individual test H_{02}

is calculated as $p_2 = \sum_{D_1=0}^4 \binom{13}{D_1} \left(\frac{1}{2}\right)^{13} = 0.1334$ which is greater than 0.05 and could not reject the individual

null hypothesis H_{02} .

4.2 Comparison of FLSDA to Bonferroni Adjustment

A Bonferroni approach that would guarantee an overall type 1 error $< \alpha$ and individual type 1 errors of $< \alpha/K$ could also be used. This would test each individual hypothesis,

$$H_{01} \text{ vs } H_{a1}, \dots, H_{0K} \text{ vs } H_{aK},$$

comparing to the size level α/K and reject the null H_{0k} if the corresponding p-value p_k is less than or equal to α/K where $p_k = P(D_k \leq d_k | D_C + D_k = d_C + d_k)$, the conditional distribution of D_k given $D_C + D_k = d_C + d_k$ under null hypothesis H_{0k} is Binomial($d_C + d_k, 1/2$) as we focus on equal allocation here.

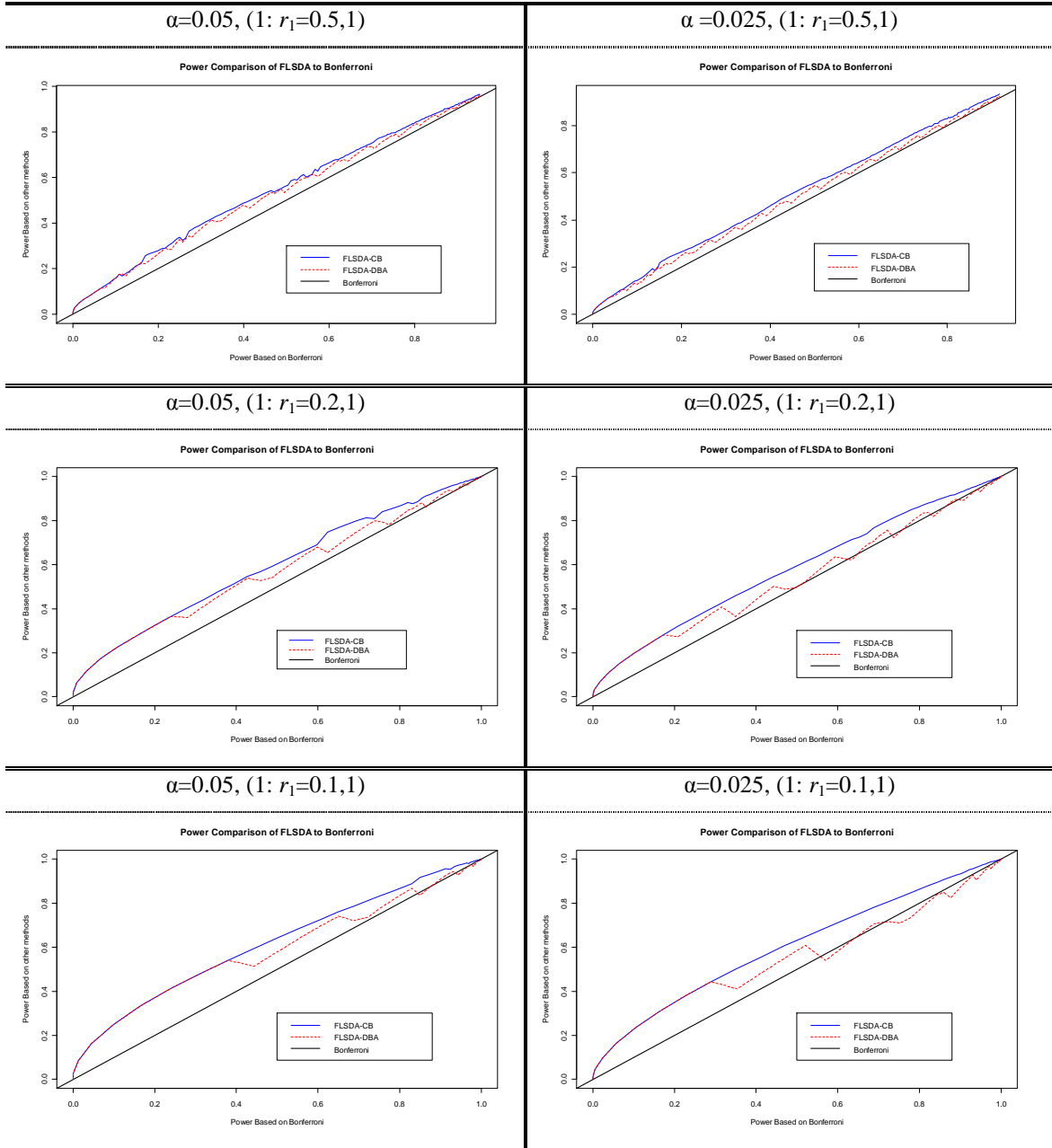


Figure 4.2.1 Power Comparison based on different $\alpha = 0.05, 0.025$ levels under the setting $(1: r_1=0.1,1)$, $(1: r_1=0.2,1)$ and $(1: r_1=0.5,1)$ under equal allocation

In this section, we will compare the power of the FLSDA approach to Bonferroni adjustment under the alternative $(1: r_1 = r, 1)$ for different value of $r < 1$ under equal allocation. The power computation method used here is discussed in Appendix C. The above figures show the power based on FLSDA vs the power based on Bonferroni for $r = 0.5, 0.2$ or 0.1 and $\alpha=0.05$ or 0.025 (x-axis is the power based on Bonferroni and the y-axis the power based on FLSDA). In the figures (Figure 4.2.1), the notation “FLSDA-DBA” and “FLSDA-CB” denote the rejection region using “Difference Boundary Augmented” or

“Conditional Binomial” in the step one of FLSDA approach, respectively. The **bolded solid black line** is the reference line which shows if the power of the rejection method is equivalent to Bonferroni approach. Being above the line means the FLSDA is more powerful while below the line means the Bonferroni is more powerful. The **dashed red line** and **solid blue line** are the power comparison based on FLSDA-DBA and FLSDA-CB, respectively, to the power based on Bonferroni approach.

The figures show that FLSDA-CB approach consistently has larger power than the Bonferroni approach for each given total numbers of disease cases and is also higher than FLSDA-DBA in most cases as well. The FLSDA-DBA approach is not as good as FLSDA-CB approach and sometimes results in lower power than Bonferroni approach does especially for small r and small size level α . Therefore, it seems that FLSDA-CB will require less numbers of follow up cases than the Bonferroni approach for a given power and a size level regardless the combination of the true or false null hypotheses H_{0k} . Thus, from now on, we only consider FLSDA-CB approach in our analysis as finding the new treatment is better than the control.

For comparing the minimum disease cases required based on FLSDA-CB and Bonferroni (shown in Table 4.2.1) for at least 80% power and the same size level under different alternative settings ($1: r_1 = r, 1$), we found that FLSDA-CB will reduce about 7% - 18% the number of disease cases from Bonferroni correction. For example, under the alternative ($1: r_1 = 0.2, 1$) and size level 0.025 for at least 80% power, it requires 35 disease cases to follow up based on FLSDA-CB and it requires 40 disease cases to follow up if applying Bonferroni adjustment. The disease cases required using FLSDA-CB procedure will be reduced about $(40-35) / 40 = 12.5\%$ than applying Bonferroni adjustment.

	($1: r_1 = r, 1$)	FLSDA-CB	Bonferroni	Reduction(%)
$\alpha = 0.025$	0.1	22	25	12 %
	0.2	35	40	12.5 %
	0.5	138	148	7.2 %
$\alpha = 0.05$	0.1	19	22	13.6 %
	0.2	27	33	18.2 %
	0.5	112	124	9.7 %

Table 4.2.1 Sample size reduction from FLSDA-CB to Bonferroni adjustment on different size levels $\alpha = 0.025$ and 0.05 under the setting ($1: r_1 = r, 1$) for $r = 0.1, 0.2, \text{ or } 0.5$ for at least 80% power

4.3 Summary

In this chapter, we have proposed a testing procedure called FLSDA to find which treatment(s) is (are) better than the control for comparing $K > 2$ new treatments to one control. For $K = 2$, we have shown

that FLSDA procedure does control overall type I error and we conjecture that the FLSDA also controls type I error for $K > 2$.

The power using FLSDA approach was compared to the power using Bonferroni adjustment procedure as the alternative of interest is that the first new treatment is superior to the control while the second new treatment is the same as the control, i.e. $(1: r_1 = r, 1)$. In section 4.2, we found that while the FLSDA-CB which tests the global hypothesis using the rejection region $R_{\alpha-CB}$ first and then individual hypothesis in the second step has higher power to reject compared to the Bonferroni adjustment but FLSDA-DBA procedure did not consistently has higher power than the Bonferroni adjustment. Hence, only FLSDA-CB approach will be used for our remaining analysis/study. Follow up disease cases required to obtain 80% power when α is either 0.025 or 0.05 for different settings for $K=2$ and one of the two treatments superior to the control were compared for the procedure FLSDA-CB to Bonferroni correction. The numbers of disease cases required for using FLSDA-CB was somewhat less comparing to Bonferroni.

**Chapter 5 SAMPLE SIZE CALCULATIONS; EXPECTED WAITING TIME; WITH AN
EXAMPLE FOR $K=2$ AND EQUAL ALLOCATION DESIGN**

5.1 Disease Cases D Needed for Given Power and Size for Finding New Treatment(s) is (Are) Better Than the Control Using FLSDA-CB

In Chapter 4, we discussed FLSDA-CB to find which new treatment(s) is (are) better than the control. In the balanced/ equally allocated designed trial, we are questioned on how much the disease cases, D , required to stop the trial for the power $1-\beta$ given type I error is less than or equals to α with a specific alternative, especially for $(1:r_1 = r,1)$ as $r < 1$.

In appendix C, we discussed power computation for given numbers of follow up cases, D , size level α under null hypothesis and specific alternative hypothesis $(1: r_1 = r,1)$ for equal allocation,. The power could be seen as a function of D , denote as $pow(D)$, for fixed size level and alternative. Hence the minimum numbers of follow up cases required, D^* , for certain power, α , and alternative would be determined through $D^*(1-\beta) = \arg \min_D \{ pow(D) \geq 1-\beta \}$.

Here, we assume the level α equals to 0.05 or 0.025 with the requested power $1-\beta$ at least 80%. The alternative here are $(1: r_1 = r, 1)$ with $r = 0.1, 0.2, \dots, 0.6$ and the power, the probability to reject new treatment 1, could be at least 80%, 85% or 90%. The filled in values are the minimum disease cases required for a certain combination of setting.

Power	80%		85%		90%	
$R \backslash \alpha$	0.025	0.05	0.025	0.05	0.025	0.05
0.1	22	19	25	21	28	24
0.2	35	27	38	31	44	37
0.3	54	43	60	49	70	58
0.4	84	68	95	78	110	91
0.5	138	112	155	128	178	148
0.6	240	195	271	225	313	263

Table 5.1.1 Minimum numbers of follow up disease cases required for given power 80%, 85% or 90% and size level 0.025 or 0.05 with the alternative setting $(1: r_1 = r,1)$, where $r=0.1, 0.2, 0.3, 0.4, 0.5, \text{ or } 0.6$.

In Table 5.1.1, each filled in values are the expected numbers of follow up diseases cases required for given power, size level α , and alternative $(1: r_1 = r,1)$. The bolded value “**22**” means that the hypothesis

testing: $H_0: (1:1,1)$ v.s. $H_a: (1:r_1 = 0.1,1)$, required to follow up till 22 diseases occur to obtain at least 80% power for a size 0.025 test.

5.2 Expected Subject Time/Subject Years N Followed up for Given Disease Cases D

For equally allocated N subject years to each group, the distribution of total disease cases $D = D_C + D_1 + \dots + D_K$ is Poisson distributed with rate iN , where $i = i_C + i_1 + \dots + i_K$ is the total incident rate regardless of the control or new treatment groups. The waiting time for a case to happen from a Poisson distribution is exponential distributed with rate $1/i$ and each disease cases to occur are randomly and independently. Hence, time to wait for D disease cases to occur is summing over D independent exponential variable and the distribution follows a gamma with parameter D and $1/i$. Therefore, we will expect to follow up D/i total subject time for D disease cases to occur. Continue with the previous example, the setting of $H_0: (1:1,1)$ v.s. $H_a: (1:r_1 = 0.1,1)$ and 22 disease cases required to occur for at least 80% power and the 0.025 size level. If the true total incidence is 0.001 per subject year, the expected follow up subject years is $22/0.001 = 22000$ total subject times under null hypothesis and about 31429 ($22/((0.001+0.0001+0.001)/3) = 31428.57$) total subject times to each group under alternative hypothesis.

5.3 Example on Prevention of HIV Transmission

The human immunodeficiency virus (HIV) is transmitted by sharing injection needles for drug users through contaminated blood left on the needles and the incidence of HIV transmission could be considered rare. This multiple comparative Poisson design could be applied to a study of HIV transmission among a high risk group if for example two promising new interventions were being tested. Suppose that the incidence of HIV transmission for the people without intervention is 0.01 and we would apply one of the following two interventions to high risk injection drug users (IDU) that are not already infected. Here are two therapies we might be interested in: 1) Behavior intervention to reduce needle sharing along with prophylactic use of highly active antiretroviral therapy (HAART) to prevent transmission to prevent HIV infection. 2) Prophylactic use of highly active antiretroviral therapy (HAART) to prevent transmission and to prevent HIV infection from establishing even if the person has contact with contaminated blood through needles;

For each therapy, we will follow up the same numbers of high risk injection drug users that initially are not infected with HIV (HIV-) until a predetermined numbers of IDUs in all groups contract HIV (D). For example, we follow up 300 drug users that share injection needles with HIV infected people who do not have any therapy. Also, we follow up 300 drug users that share needles with HIV infected people and have behavior/HAART intervention therapy and 300 drug users that share needles with HIV infected people with HAART.

In this study, there are two interventions ($K=2$) and we want to see which intervention/therapy is better (less HIV transmission rate) than the untreated group. We would like to determine the numbers of drug users in all three arms that contract HIV from sharing injection needles with HIV infected persons required to obtain 80% under size level $\alpha=0.05$ for equal allocation. In addition to the determination of D , we also want to identify the rejection regions for our design.

If our hypothesis of interest is as follows:

$$H_0: r_1 = r_2 = 1 \text{ vs } H_a: r_1 = 0.1 \text{ and } r_2 = 1$$

While the power is 78.7% as the study only follow up till 18 people contracting HIV and the power is 81.3% for following up till 19 people present HIV under size level $\alpha=0.05$ based on our proposed procedure FLSDA-CB, we would like to end our trial after 19 people present HIV infection among the untreated and therapy groups. Then, the numbers of HIV infections in each group would be compared and tested as follows:

(1) The rejection region for global hypothesis:

$$R_{\alpha\text{-CB}} = \{(D_C, D_{(1)}) \mid (D_C \geq 5, D_{(1)}=0), (D_C \geq 7, D_{(1)}=1), (D_C \geq 9, D_{(1)}=2), (D_C \geq 11, D_{(1)}=3)\},$$

is as Illustration 5.3.1. If $(D_C, D_{(1)})$ is in $R_{\alpha\text{-CB}}$, then we continue to next step (2). Otherwise, we stop the procedure and conclude that no new treatment is better than the control.

(2) The following Illustration 5.3.2 is the rejection region for each individual hypothesis that the first row is $D_C + D_k$ which ranges from 0 to 19, and the second row is the critical value, t , that $P(D_k < t \mid D_C + D_k) \leq 0.05$ for each corresponding $D_C + D_k$. For example, if $D_C + D_k=19$, the null hypothesis H_{0k} will be rejected if we observe $D_k < 6$.

$D_C \backslash D_{(1)}$	0	1	2	3	4	5	6	7	8	9
0										
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										

Illustration 5.3.1 Rejection Region (shaded) for global hypothesis based on conditional binomial, $R_{\alpha-CB}$, for $K=2$, $D=19$, $\alpha=0.05$

D_C+D_k	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
$D_k <$	6	6	5	5	4	4	4	3	3	3	2	2	2	1	1	1	0	0	0	0

Illustration 5.3.2 Rejection region for each individual hypothesis for $K=2$, $D=19$, $\alpha=0.05$ and $k=1,2$. As $D_C + D_k = 19$, we will reject the null of H_{0k} as $D_k < 6$ and conclude that the new treatment k is better than the control

For example, assume that the following scenario we observed:

(i) 11 people in the un-treat group infected with HIV, 3 people infected in the behavior/HAART

intervention and 5 people are infected HIV in HAART group as predetermined total number of disease cases 19 infected people.

In the first step, $(D_C = 11, D_{(1)} = 3)$ is in the rejection region $R_{\alpha-CB}$, so we would conclude in the first step that at least one new treatment is better than the control and continue to the second step. For testing for the alternative that behavior/HAART is better than un-treat group, the individual null will be rejected if $D_1 < 4$ as $D_C + D_1 = 14$ for this case. Since we observe $D_C = 11, D_1 = 3$, we would reject the null hypothesis that prophylactic behavior/HAART intervention works the same as no intervention. On the other hand, for comparing HAART therapy to the un-treat group, the individual null hypothesis will be

rejected if $D_2 < 5$ as $D_C + D_2 = 16$. There is not enough evidence that the HAART is better than the un-treated group since $D_2 = 5$ and is not in the rejection region for the individual test.

Now assume that among the 19 predetermined cases overall, we observe:

(ii) 16 infected people in the un-treated group, 1 infected person in the behavior/HAART, and 2 infected people in prophylactic HAART. In the first step, $(D_C = 16, D_{(1)} = 1)$ is in the rejection region for the global hypothesis, $R_{\alpha-CB}$. We reject the null hypothesis and conclude that there is at least one new treatment that is better than the control and continue to the second step. Since $D_1 = 1 < 5$ as $D_C + D_1 = 17$ for comparing behavior/HAART to the control and $D_2 = 2 < 6$ as $D_C + D_2 = 18$ for comparing HAART only to the un-treated group, we could conclude that both therapies are better than the control.

Finally, assume that after the predetermined 19 infected cases occur, we observe;

(iii) 11 infected people in the control group, 4 infected people in behavior/HAART, and 4 infected people in HAART. Neither treatment is considered to be better than the control since $(D_C = 11, D_{(1)} = 4)$ is not rejected in the global hypothesis of Step one.

Chapter 6 OPTIMIZING ALLOCATION RATIO OF CONTROL TO TREATMENT

Until now, we have assumed in most cases that the same numbers of subjects (or subject time) will be allocated to the controls and each new treatment group. However, most of the results here derived for equal subject allocation can be extended to unequal allocation. It is difficult to think of reasons to allocate more subjects to one new treatment group versus another, say twice as many subjects into D_1 than D_2 as this would skew the comparisons towards being more precise for the treatment(s) that are allocated more subjects and less precise for those which are allocated less. Perhaps if it was already known in advance that one new treatment was likely to be worse (or more costly to study) than another, then this would inform a subject allocation. But in such a setting, the new treatment which is likely to be worse (or more costly) would not be used in the comparisons.

We, therefore, restrict our considerations here to allocate same numbers of subjects to each new treatment group and the allocation ratio that is the ratio of subjects to the control group over each new treatment group (under equal allocation into each new treatment). For example, each of the K new treatments is being compared to the same control and the new treatments are not compared to each other. Dunnet showed that with K treatments compared to one controls and each subject having a normally distributed outcome that given a constrained total number of subjects, the [Control] / [Each New Treatment] allocation ratio of $\rho = \sqrt{K}/1$ subjects minimized the variance of the test statistics $\bar{X}_k - \bar{X}_C$. While the setting differs here, i) as in Dunnet's setting, the K new treatments are compared to the same control, but the underlying distribution assumptions differ. ii) In this setting, the fact that disease cases accrue at a slower rate in the new treatment groups when the treatment(s) are good may impact the optimal control to new treatment allocation. Since in our setting for given power and size, we might tend to allocate more to the control when the new treatment(s) are really good to reduce the total follow up time. iii) Discreteness properties of the multinomial here may allow the threshold test size to be closer to (or further from) the specified α under unequal allocation schemes which would correspondingly also result in greater (or less) power from a test that meets the specified α level.

As a theoretical solution is untenable, we try to empirically identify optimal allocation ratio that maximize power per size under two settings.

A. Minimize the number of disease cases D needed for the study. Perhaps for ethical reasons and/or because disease cases are particularly expensive, it is important to minimize these. Perhaps the D can be made smaller by allocating less or more of the subject to each new treatment group (versus the controls) than the under equal allocation.

B. Minimize the expected subject follow up time. The trial could be long and costly and the costs could be directly proportional to the number of subjects or subject follow up time. In that case, minimizing the expected subjects (or subject time) needed to obtain the D cases is important. If the new treatment(s) is (are) effective, the trial length required for the specified number of cases may be shortened by allocating more to the control groups due to the more rapid accrual of cases in the enlarged control arm. Hence, we might like to allocate more to the control to minimize the cost or to minimize the total waiting time for a trial. While if it costs more for the new treatments than the control this would push the design towards allocating more subject time to the controls. However, since such cost differential will differ by each setting, we do not focus on this potential here.

In this chapter, we will thus focus on the optimum allocation ratio that will A. minimize the follow up cases required or B. minimize the expected follow up time needed to accrue the D cases (here we call it “expected waiting time”) under the alternative for a certain power and type I error as $K=2$ adapted from the method we applied to equal allocation.

6.1 Optimum Allocation Ratio that Minimizes the Follow Up Cases Required

Let ρ be the allocation ratio for the control to each of the new treatment groups. The methodologies we proposed for equal allocation ($\rho = 1$) can be modified to suit the unequal allocation. As discussed in section 3.1, the hypothesis for unequal allocation could be express as

$$H_0: r_1 = r_2 = 1 \text{ vs } H_a: r_k < 1 \text{ for some } k=1,2, \text{ where } r_k = i_k / \rho i_C$$

or the rejection region could be designed as null considered as $(\rho: 1, 1)$ and alternative considered as $(\rho:r_1,r_2)$ for some $r_k < 1$. From now on, we will focus on the setting $H_0: (\rho:1,1)$ v.s. $H_a: (\rho: r_1 = r, 1)$ for $r < 1$ for unequal allocation with allocation ratio ρ for the control to each treatment group.

Again, Appendix C, discusses the specific computations used to identify powers for given numbers of follow up cases, D , size level α , under null hypothesis and specific alternative hypothesis ($1: r_1$

$= r, 1)$ for equal allocation, $\rho = 1$ and that would be extended it to unequal allocation case easily. The power could be seen as a function of D , denote as $pow(D)$, for fixed size level and alternative, r . Denote that the minimum numbers of follow up cases required, $D^{\rho=1}$, for certain power, α , and r as

$$D^{\rho=1}(1-\beta) = \arg \min_D \{ pow(D) \geq 1-\beta \} .$$

And $D^{\rho=1}$ could be determined by i) repeating the power computation discussed in Appendix C for a given D and then ii) by increasing follow up numbers D until the power based on D exceeds $1-\beta$. For the unequal allocation setting with a specified allocation ratio ρ , we could first adjust the power computation by modifying the null hypothesis and alternative hypothesis as described above according to the allocation ratio ρ . Then we could use the same method to find the minimum disease cases required, denote as D^ρ , for requested power, size level, allocation ratio ρ , and r again in the $(1: r_1 = r, 1)$ setting.

Here, we will first show how to minimize the follow up cases needed in the case of $r = 0.1$ and $\alpha = 0.05$ with 80% power. We systematically evaluate the spectrum of possible allocation ratios ρ and find the corresponding follow up disease cases D^ρ needed to obtain power $\geq 80\%$ under size level 0.05 and alternative hypothesis $(1: r_1 = 0.1, 1)$ and the given ρ . Here, the allocation ratio ρ is selected by its logarithm, $\log(\rho)$, from -1.1 to 1.1 and increment by 0.01. Putting ρ on the log scale symmetries about 1 and the range -1.1 to 1, corresponds roughly to ρ going from 0.33 to 3 by multiples of 1.01. Note, it was clear that having $\rho < 0.33$ or $\rho > 3$ would not minimize D^ρ . The points in Figure 6.1.1 (minimum follow up cases required D^ρ as y-axis against allocation ratio ρ as the x-axis) are the minimum follow up cases required D^ρ v.s. allocation ratio ρ . Each point (ρ, D^ρ) shows that the minimum disease cases D^ρ required to have at least 80% power when the allocation ratio of control to the each treatment group is ρ given size level $\alpha=0.05$. The blue point $(\rho = 0.3753, D^\rho = 24)$ shows that for an allocation ratio 0.3753 ($\log(0.3753) = -0.98$), $D^{\rho=0.3753} = 24$ disease cases are needed to have at least 80% power. As we can see, there are 24 optimum allocation ratios that achieve the minimal follow up disease cases $D^\rho = 18$ needed for this setting due to the discreteness property of integer (D^ρ). Hence, the optimum allocation ratio is not unique.

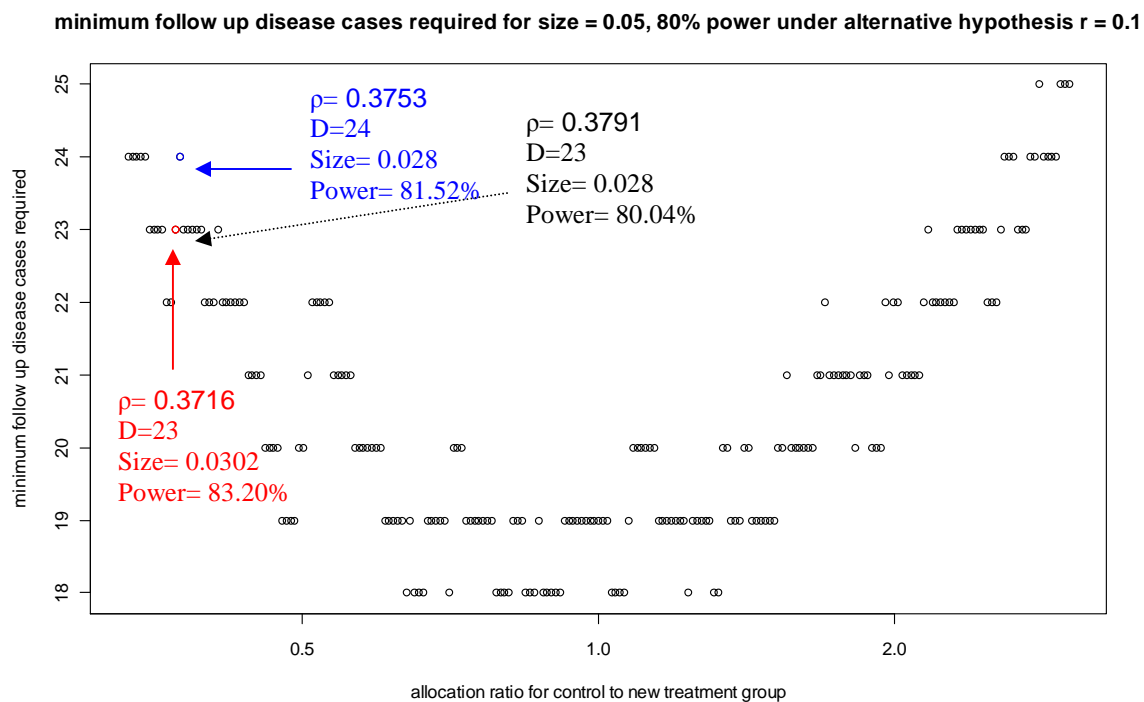


Figure 6.1.1 Minimum follow up number of disease cases required for size 0.05 with 80% power under null hypothesis when alternative hypothesis ($\rho: r = 0.1, 1$). Each point is (ρ, D^ρ) where D^ρ is the number of follow up cases required to obtain 80% power for the setting $\alpha = 0.05, r = 0.1$, and the its corresponding ρ

Also, due to the discreteness of distribution and the rejection region, the follow up disease cases needed for at least 80% power could jump up or down as allocation ratio is slightly changed at a given size level. Continuing with previous example, for allocation ratio, 0.3716 (whose log is -0.99) (shown in red in Figure 6.1.1), it needs $D^{\rho=0.3716}=23$ to obtain at least 80% power while with a slight increase of 0.01 log units in the allocation ratio to 0.3753, the disease cases required for the request power jumps to $D^{\rho=0.3753} = 24$. As allocation ratio increases another 0.01 log unit to 0.3791 (whose log is -0.07), the disease cases needed drops back to 23 again. In the practice, we would start with the optimum allocation ratio that minimized the numbers of disease cases required to have 80% power. But due to the uncontrollable participants randomly drop off during the trial, the allocation ratio might slightly vary than the predetermined setting and required more disease cases to have 80% power.

While we could try to look further to see which of the 14 allocation ratios that minimized the number of needed disease cases at 23 for this example had the maximum true power and/or minimum true size, but it is impossible to come up with an unambiguous metric on this basis to break the tie. Also, the slightly different of allocation ratio from the optimum allocation ratio might result in the increasing the

numbers of disease cases needed for given size and power. Therefore, to reduce the effect of discreteness, we tried to smooth the discreteness from the optimum allocation ratio ρ through locally weighted regression (Cleveland, 1979) (LOWESS Cleveland 1981) or smooth spline. The red / green curves in the Figure 6.1.2 are the fitted curve using function LOWESS in software R with smoothing parameters $f = 0.2$ and $f = 0.3$. Also, the blue fitted curve uses the function smooth spline in R with 5 degrees of freedom. Since the curves are V-shaped the pseudo-optimum allocation ratio by this approach could be determined as shown in the figure below.

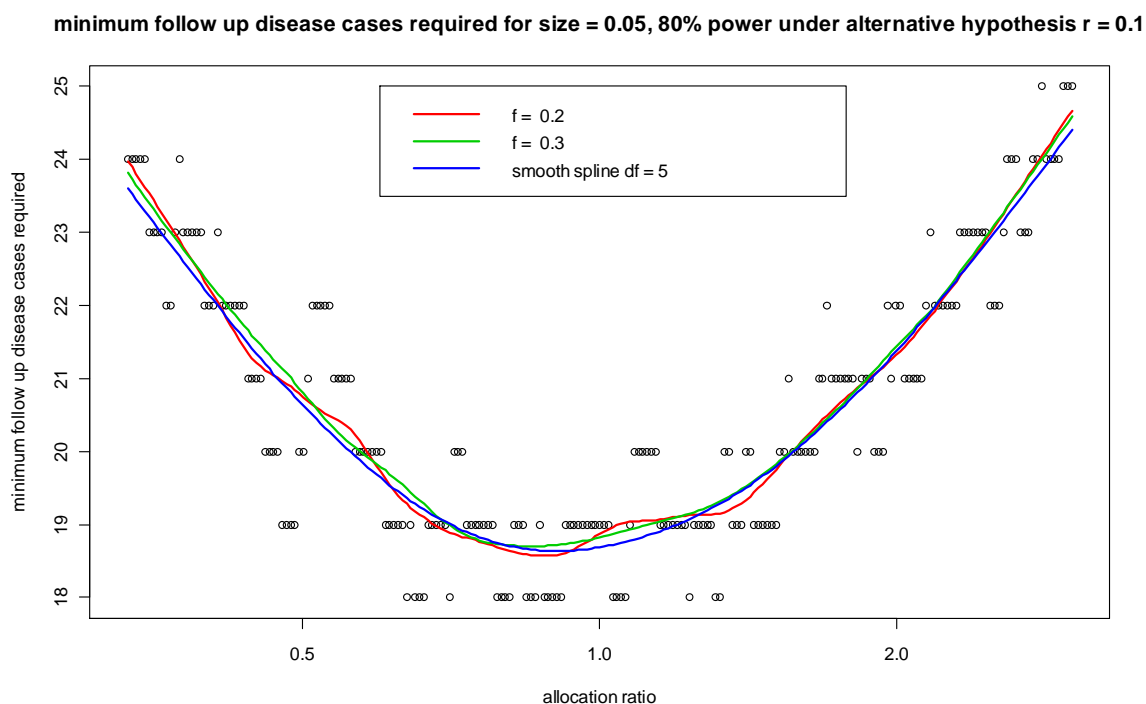


Figure 6.1.2 Minimum follow up disease cases required for size 0.05 with 80% power under null hypothesis with an alternative hypothesis ($\rho: r_1 = 0.1, 1$) and smoothing spline curves. Each point is (ρ, D^ρ) where D^ρ is the numbers of follow up cases required to obtain 80% power for the setting $\alpha = 0.05, r = 0.1$, and the its corresponding ρ . The smoothing curve is obtained through locally weighted regression by function LOWESS in R software with smoothing parameter $f=0.2$ (red) and 0.3 (green) and through smoothing spline by function `smooth.spline` in software R with degrees of freedom $df = 5$ (blue).

6.2 Optimum Allocation Ratio that Minimizing the Expected Follow Up Subjects / Time

Once the minimum follow up cases required, D^ρ , is determined for the allocation ratio ρ , the expected waiting time for D^ρ cases to occur is D^ρ/i where i is the averaged incident rate across all groups and which would be achieved by

$$i = \rho c i_C + \rho_T i_1 + \rho_T i_2 = \rho_T (\rho i_C + i_1 + i_2) = (\rho i_C + i_1 + i_2) / (\rho + 2)$$

$= (\rho+r+1) i_C / (\rho+2)$ since $i_1 = r i_C$ and $i_2 = i_C$ under any hypothesis of form $(\rho: r_1 = r, 1)$.

Hence for the $(1: r_1 = r, 1)$ setting, the expected waiting time to accrue D^ρ is $D^\rho / i = (\rho+2) D^\rho / i_C(\rho+r+1)$ which is proportional to $(\rho+2) D^\rho / (\rho+r+1)$ since i_C is fixed as the allocation ratio differs. Since the multiplicative effect of i_C is invariant to the allocation ratio, i_C can be set to 1 and the minimization of the expected waiting time is equivalent to the minimization of $(\rho+2) D^\rho / (\rho+r+1)$. Note that under the null hypothesis, $r = 1$, the optimum allocation ratio that has shortest expected waiting time is equivalent to the optimum allocation ratio that minimized follow up cases required D^ρ since $(\rho+2)/(\rho+r+1) = 1$. Except for pathological settings noted earlier where the treatment groups do worse than the controls, under the alternative hypothesis where $r < 1$, it is true that $(\rho+2)/(\rho+r+1) > 1$ for every ρ , meaning the expected waiting time will always be longer than the expected waiting time under null hypothesis.

Here, we will continue the previous example to find the optimum allocation ratio which minimizes the expected waiting time under a specific alternative hypothesis, $(\rho: r_1 = r, 1)$ for $r < 1$, for the null hypotheses tested at an overall type-1 error of 0.05 with 80% power.

To identify the optimum allocation ratio to achieve the shortest expected waiting time to obtain the given number of cased D for a given power, we could apply similar methodology as searching for the optimum allocation ratio which minimized the number of follow up disease cases to achieve that power. In Figure 6.2.1 (labeled the expected waiting time under the alternative $(\rho: r_1 = 0.1, 1): (\rho+2) D^\rho / (\rho+1.1)$ as **y-axis and ρ as x-axis**), it shows that there was one unique smoothed exact optimum allocation ratio, 1.32313, ($\log(1.32313) = 0.28$) which resulted in the shortest expected follow up time needed to accrue $D^{\rho=1.32313} = 18$ cases among the allocation ratios ρ that are selected by its logarithm, $\log(\rho)$, from -1.1 to 1.1 and increment by 0.01. But similarly as with the number of D required in 6.1, the expected waiting time to obtain the needed cases could jump up or down as allocation ratio is slightly different for request power and a given size level. Hence, in practice, the optimum allocation ratio might not guarantee the shortest expected waiting time since the optimum allocation ratio were impractical to achieve due to the random dropout and might result in slightly less or more allocation ratio than the pre-determined optimum allocation ratio. Hence, locally weighted regression and smoothing spline are applied to search for the optimum allocation ratio that is considered to find the optimum allocation ratio that shortest expected waiting time locally and averagely.

The red / green colored curves are the smoothing curves that are fitted by locally weighted regression (LOWESS; $f = 0.2$ and $f = 0.3$) and the blue curve is fitted using a smoothing spline (smooth.spline; $df = 5$). Unique smoothed optimum allocation ratios to minimize expected subject-time for this setting through the three locally weighted regression approaches (LOWESS) and smooth spline (smooth.spline) roughly coincided.

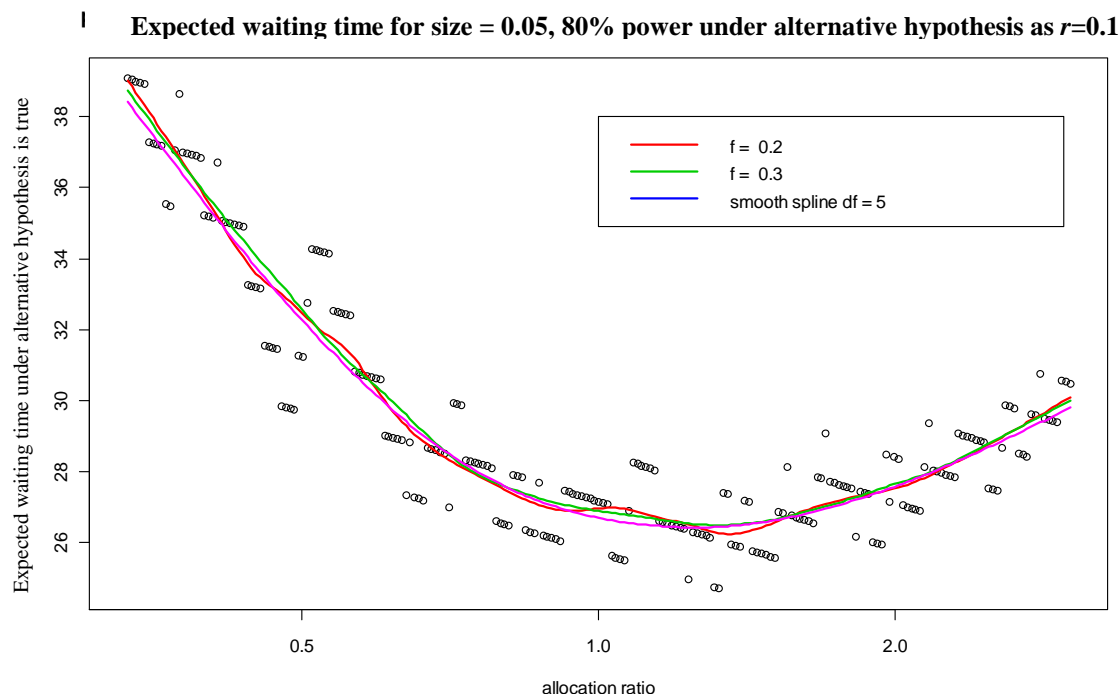


Figure 6.2.1 Expected waiting time for $\alpha = 0.05$ with 80% power under alternative hypothesis for ($\rho: r = 0.1, 1$). Each point is (ρ, D^ρ) where D^ρ is the numbers of follow up cases required to obtain 80% power for the setting $\alpha = 0.05, r = 0.1$, and its corresponding ρ . The smoothing curve is obtained through locally weighted regression by function LOWESS in R software with smoothing parameter $f=0.2$ (red) or 0.3 (green) or through smoothing spline by function smooth.spline in software R with degrees of freedom $df = 5$ (blue).

6.3 Results: Optimum Allocation Ratio For Different Alternatives

6.3.1 Optimum Allocation Ratio that Minimizes the Follow Up Disease Cases for Different

Alternatives

In this section, we are going to discuss the optimum allocation ratio that minimizing the required follow up cases for 80% power and size level $\alpha = 0.025$ or 0.05 for $K = 2$ through different smoothing methods we mentioned in Section 6.1 under different alternative settings. Noting that we are not discussing

the optimum allocation ratio based on “exact” method that allocation ratio is optimum among the allocation ratios ρ which are selected by its logarithm scale, $\log(\rho)$, from -1.1 to 1.1 and increment by 0.01 due to the tie and no unique solution.

Again, the allocation ratios ρ are selected through logarithm of ρ from -1.1 to 1.1 and incremented by 0.01 and the smoothing methodologies are applied. In Table 6.3.1, the optimum allocation ratio to minimized follow up disease cases required for 80% power under size level $\alpha = 0.05$ or 0.025 and alternative ($\rho: r_1 = r, 1$) where r is from 0.1 to 0.5 is obtained through smoothing methodologies. Most approaches show that the smoothed optimum allocation ratio to minimize the follow up cases is near 1.

$\alpha = 0.025$				$\alpha = 0.05$			
r	LOWESS		Smooth Spline	r	LOWESS		Smooth Spline
	f = 0.2	f = 0.3			f = 0.2	F = 0.3	
0.5	0.8694	0.8187	1.1201	0.5	1.1275	1.1052	1.1158
0.4	0.9608	0.9418	1.0748	0.4	1.0202	1.0202	1.0538
0.3	1.0000	1.0000	1.0140	0.3	1.0408	0.9802	1.0083
0.2	1.0408	1.0618	0.9687	0.2	0.9418	0.8869	0.9653
0.1	1.1275	1.1275	0.9164	0.1	0.8694	0.8521	0.9040

Table 6.3.1 Smoothed optimum allocation ratio to minimize number of cases required for 80% power under size level $\alpha = 0.05$ and 0.025 as alternative is $H_a: (\rho: r_1 = r, 1)$ (where $r = 0.1, 0.2, 0.3, 0.4,$ and 0.5) based on LOWESS with parameter f = 0.2 or 0.3 and smoothing spline with degrees of freedom df = 5

Using a balanced design, i.e. same numbers of people assigned to the control and to each new treatment group or $\rho = 1$, has some advantages such as being easier to apply randomization and double-blinded design trials. Since the optimum allocation ratio to minimize the follow up cases in each case is not far from 1, even when there is a penalty for having more disease cases occur in the study, we might still prefer to allocate equal number of subjects to the control and each treatment group in order to make the randomization and double-blind design easier to apply and to prevent some biased issues that might occur from unbalanced design.

6.3.2 Optimum Allocation Ratio that Minimizing the Follow Up Subjects / Time for Different

Alternatives

In this section, we are going to determine the optimum allocation ratio that would result in shortest follow up time for a certain required number of disease cases for 80% power and size level $\alpha = 0.025$ or 0.05 for $K = 2$ using the exact or smoothing methods discussed in Section 6.2 under different alternatives.

Table 6.3.2 shows the optimum allocation ratio that result in shortest the follow up time to achieve the needed number of cases for 80% power, size level $\alpha = 0.05$ or 0.025 when the alternative ($\rho: r_1 = r, 1$) is true where r is from 0.1 to 0.5 under the following methods: (1)“exact” - the optimum allocation ratio among the allocation ratios ρ that are selected by its logarithm scale, $\log(\rho)$, from -1.1 to 1.1 and increment by 0.01. (2) based on LOWESS with parameter $f = 0.2$ or 0.3, and (3) based on smoothing spline with 5 degrees of freedom. In most case, the optimum allocation ratio is located from 1.2 to 1.4. Hence, for $r = 0.1$ to 0.5 at 80% power and overall sided one type error of 0.025 or 0.05 for $K=2$, we would like to allocate 1.2 to 1.4 times more subjects to control group than each new treatment group to shorten the expected waiting time if we were convinced that the alternative hypotheses were true.

r	$\alpha = 0.025$				$\alpha = 0.05$			
	Exact	LOWESS		Smooth Spline	Exact	LOWESS		Smooth Spline
		$f = 0.2$	$f = 0.3$			$f = 0.2$	$f = 0.3$	
0.5	1.4192	1.2712	1.2712	1.2886	1.2214	1.2969	1.3231	1.2876
0.4	1.1388	1.3771	1.2461	1.2881	1.1853	1.2461	1.3231	1.2606
0.3	1.4049	1.3499	1.2969	1.2553	1.2337	1.1503	1.2712	1.2418
0.2	1.3771	1.3499	1.3231	1.2418	1.3771	1.4333	1.2712	1.2378
0.1	1.2093	1.2969	1.3231	1.2344	1.3231	1.3499	1.3231	1.2555

Table 6.3.2 Smoothed optimum allocation ratio to shorten expected waiting time for D^p cases to accrue of $\alpha = 0.05$ and 0.025 with power = 80% under $H_a: (\rho: r_1 = r, 1)$ is true (where $r = 0.1, 0.2, 0.3, 0.4,$ and 0.5) based on exact search, LOWESS with parameter $f = 0.2$ or 0.3, and smoothing spline with degrees of freedom $df = 5$

Chapter 7 CONCLUSIONS AND FUTURE WORK

Comparative Poisson trials can be long and costly, especially when there are several prophylactic intervention candidates or several different dosage options. It might be ideal to find the best candidate treatment. In this study, first, we suggested using statistic $(D_C, D_{(1)})$ to construct the rejection region for testing if there is any new treatment that is superior to the control and the probability mass of $(D_C, D_{(1)})$ could be computed through products of stochastic metrics. Since there is no UMP for this multiple comparison setting based on this statistic, we added LLQR which will reject from large numbers of disease cases in the control and from small numbers of disease cases in the minimum new treatment group. For comparing two new treatments to one control ($K = 2$), the rejection region satisfying LLQR is unbiased has been proved. For $K > 2$, we believe that this unbiased property remains true but it is hard to extend and prove mathematically. In addition to LLQR, we also discussed several criteria to construct the rejection region. The metric criteria include “controls”, “minimum new treatment”, “difference”, and “cumulative binomial”. We tried to find if the criteria could control type I error under pathological null hypothesis and remain high power under pathological alternative hypothesis. We eliminated the first two criteria since the one using “controls” results in small power under pathological alternative and the one using “minimum new treatment” could not control type I error under pathological null. The rejection region using “difference” and “cumulative binomial” metric is applied for constructing the rejection region for global hypothesis.

Second, we propose a procedure, FLSDA, to test which new treatment(s) is(are) superior to the control for comparing two new treatments or two prophylactic methods to a control. This post-hoc procedure is to test each individual hypothesis once the global hypothesis is rejected. The FLSDA procedure has been shown to control overall type I error for $K = 2$, however, this property is hard to extend to or prove for $K > 2$. The rejection region using “difference” metric for global hypothesis is eliminated since FLSDA-DBA does not have higher power in some settings compared to Bonferroni adjustment. On the other hand, FLSDA-CB has consistent higher power than the Bonferroni adjustment for the same total numbers of disease cases to end the trial. Therefore, FLSDA-CB is applied for comparative Poisson trial design to compare 2 new treatments to one control. Then, we extended the design strategy to unequal allocation case for FLSDA-CB procedure. We tried to find the optimum allocation ratio that minimized the

total disease cases required to follow up or the optimum allocation ratio that could shorten the follow up time.

In this type of design for comparative Poisson trial, some properties could be shown and proved for $K = 2$, but these properties might be hard to extend to or prove the properties for $K > 2$ cases mathematically. The difficulty results from the dependent multiple hypotheses and is caused by the setting of design A that conditioning the total numbers of disease cases in all groups.

For the future work, we would like to explore other approach design; we call Design C (Appendix D), which is to end the trial when a certain numbers of disease cases occur in the control group. This design would be an alternative and might work especially well for extended cases $K > 2$. Also, the distribution that the numbers of disease cases in a new treatment group as the trial end based on design C follows a Negative Binomial distribution and the numbers of disease cases in each new treatment group is independent as conditional on the numbers of diseases cases in the control group. Hence, there might be a better theoretical approach for this type problem for $K=2$ and easier to extend the theories to $K > 2$ new treatments. In addition, the expected waiting time to recruit/follow up a certain numbers of disease cases in control group only depends on the incidence rate in the control group (which is better known than the incidence in treatment groups that depend on treatment efficacy) based on design C. The expected waiting time for design C could thus be better predetermined.

Appendix A UNBIASED PROPERTY FOR REJECTION REGION SATISFYING LLQR FOR $K=2$

In this appendix, we prove that if a rejection region satisfies LLQR, the power will be larger than the size (unbiased property) for comparing two treatments to one control ($K=2$). We will (i) first show that under the alternative of $(1: r_1, r_2)$ as $r_1, r_2 \leq 1$ with at least one strict inequality hold, the unbiased property holds for the rejection region using number in the “minimum treatment group” only, and then (ii) we will extend the unbiased property to any rejection region as long as it satisfies LLQR. Hence, any LLQR rejection region, including rejection regions using both the number of controls and the minimum treatment we proposed in chapter 3 and 4, satisfy the unbiased property.

A.1 Unbiased Properties for the Rejection Region Using “Minimal Treatment Only”

The probability to reject using the rejection region based on “minimum treatment only” under the alternative setting $(1:r_1, r_2)$ is denoted as $P_{(1:r_1, r_2)}(D_{(1)} \leq d_{(1)} | D)$ and $P_{(1:1,1)}(D_{(1)} \leq d_{(1)} | D)$ is the size under the null hypothesis $(1:1,1)$. We will first show that three propositions: (A.1-1), (A.1-2), and (A.1-3) below hold, and then combine these properties to prove that

$$P_{(1:r_1, r_2)}(D_{(1)} \leq d_{(1)} | D) \geq P_{(1:1,1)}(D_{(1)} \leq d_{(1)} | D) \text{ for all possible } d_{(1)}.$$

Proposition A.1-1: The conditional distribution of $d_{(1)} | D$ given $D_1 + D_2 = D_T$ is stochastically smaller under the setting $(1:r_1, r_2)$ than $(1:1,1)$, i.e. $P_{(1:r_1, r_2)}(D_{(1)} \leq d_{(1)} | D_T, D) \geq P_{(1:1,1)}(D_{(1)} \leq d_{(1)} | D_T, D)$ with strictly inequality for some $d_{(1)}$

Pf:

By using the incomplete beta function, $P(X \leq k) = (n-k) \binom{n}{k} \int_0^{1-p} t^{n-k-1} (1-t)^k dt$ as $X \sim \text{Bin}(n, p)$. Since $D_1 |$

$D_1 + D_2 = D_T$ is Binomially distributed with parameters (D_T, p_1) where $p_1 = r_1/(r_1+r_2)$ for the setting $(1:r_1, r_2)$

and $p_1 = 1/2$ for setting $(1:1,1)$ and $D_2 | D_1 + D_2 = D_T$ is Binomially distributed with parameters (D_T, p_2)

where $p_2 = r_2/(r_1+r_2)$ for the setting $(1: r_1, r_2)$ and $p_2 = 1/2$ for setting $(1:1,1)$.

For $k \geq D_T/2$,

$$P(D_{(1)} \leq k | D_T) = 1.$$

For $k < D_T/2$,

$$P(D_{(1)} \leq k | D_T) = P(D_1 \leq k | D_T) + P(D_2 \leq k | D_T)$$

$$\begin{aligned}
&= (D_T - k) \binom{D_T}{k} \int_0^{1-p} t^{D_T-k-1} (1-t)^k dt + (D_T - k) \binom{D_T}{k} \int_0^p t^{D_T-k-1} (1-t)^k dt \\
&= C \cdot \left(\int_0^{1-p} t^{D_T-k-1} (1-t)^k dt + \int_0^p t^{D_T-k-1} (1-t)^k dt \right) = C \cdot h(p),
\end{aligned}$$

where $C = (D_T - k) \binom{D_T}{k}$ and $h(p) = \int_0^{1-p} t^{D_T-k-1} (1-t)^k dt + \int_0^p t^{D_T-k-1} (1-t)^k dt$.

Here, $h(0)=h(1)$, $h'(p)=0$ as $p=1/2$ and $h(0)=h(1)>h(1/2)$. (will show as follows)

The maximums (and minimums) for this function are found by setting the first derivative to zero, i.e.

$$\frac{dP(D_{(1)} \leq k | D_T)}{dp} = C \cdot \frac{dh(p)}{dp} = 0 \text{ which happens when } -(1-p)^{D_T-k-1} p^k + p^{D_T-k-1} (1-p)^k = 0 \text{ and this only}$$

happens when $p^{D_T-2k-1} = (1-p)^{D_T-2k-1}$ which for $0 \leq p \leq 1$ for all possible D_T-2k-1 . Hence $p=1-p$, i.e.

$p=1/2$ is the solution for the first derivative equals to 0 and is the minimum if the boundaries ($p=0$ and $p=1$) are both greater than this extreme value at $p=1/2$.

For $p=0$,

$$\begin{aligned}
h(0) &= \int_0^{1-0} t^{D_T-k-1} (1-t)^k dt + \int_0^0 t^{D_T-k-1} (1-t)^k dt = \int_0^1 t^{D_T-k-1} (1-t)^k dt \\
&= \int_0^{1/2} t^{D_T-k-1} (1-t)^k dt + \int_{1/2}^1 t^{D_T-k-1} (1-t)^k dt \\
&\geq 2 \int_0^{1/2} t^{D_T-k-1} (1-t)^k dt = h(1/2).
\end{aligned}$$

The last inequality of above holds because

$$\int_{1/2}^1 t^{D_T-k-1} (1-t)^k dt = \int_0^{1/2} (1-t)^{D_T-k-1} t^k dt \geq \int_0^{1/2} t^{D_T-k-1} (1-t)^k dt$$

Since $\frac{(1-t)^{D_T-k-1} t^k}{t^{D_T-k-1} (1-t)^k} = \left(\frac{1-t}{t}\right)^{D_T-k-1} \left(\frac{t}{1-t}\right)^k = \left(\frac{1-t}{t}\right)^{D_T-2k-1} \geq 1$ for $k < D_T/2$ and $t < 1/2$.

For $p=1$,

$$h(1) = \int_0^{1-1} t^{D_T-k-1} (1-t)^k dt + \int_0^1 t^{D_T-k-1} (1-t)^k dt = \int_0^1 t^{D_T-k-1} (1-t)^k dt = h(0) \geq h(1/2)$$

Therefore, $h(1)=h(0) \geq h(1/2)$.

Hence there is a global minimum at $p=1/2$ no matter what combination of D_T and k , i.e.

$$P_{(1:r_1, r_2)}(D_{(1)} \leq k | D_T) \geq P_{(1:1,1)}(D_{(1)} \leq k | D_T) \text{ under setting for any given } D_T$$

In order to show that $D_{(1)} | D_T$ is stochastically smaller under $(1:r_1, r_2)$ than the setting under $(1:1,1)$, we need to show the strict inequality $P_{(1:r_1, r_2)}(D_{(1)} < k | D_T) > P_{(1:1,1)}(D_{(1)} \leq k | D_T)$ for some k :

For $k=0$ and let $p=r_1/(1+r_1)$,

$$\begin{aligned} & P(D_{(1)}=0 | D_T) \text{ under setting } (1:r_1, r_2) \\ &= p^{D_T} + (1-p)^{D_T} > 2\left(\frac{1}{2}\right)^{D_T} = P(D_{(1)}=0 | D_T) \text{ under setting } (1:1,1) \end{aligned}$$

The inequality holds as $p^{D_T} + (1-p)^{D_T}$ has global minimum at $p=1/2$ since $\frac{d}{dp} \ln(p^{D_T} + (1-p)^{D_T}) = 0$ as

$$p=1/2 \text{ and } \frac{d^2}{dp^2} \ln(p^{D_T} + (1-p)^{D_T}) = -D_T(1-p)^{-2} - D_T p^{-2} < 0.$$

$P(D_{(1)}=0 | D_T)$ under setting $(1:r_1, r_2)$ is strictly larger than $P(D_{(1)}=0 | D_T)$ under setting $(1:1,1)$.

As the result, the distribution of $D_{(1)} | D_T$ under $(1:r_1, r_2)$ is stochastically smaller than the distribution of $D_{(1)} | D_T$ under $(1:1,1)$.

Proposition A.1-2: The conditional distribution of $D_{(1)} | D_T, D$ is stochastically smaller for smaller D_T , i.e.

$$P(D_{(1)} \leq d_{(1)} | D_T = d_T, D) \geq P(D_{(1)} \leq d_{(1)} | D_T = d_T + 1, D)$$

with strictly inequality for some $d_{(1)}$

Pf:

First, we will show that the binomial(n, p) will stochastically smaller than binomial(m, p) if $n < m$. And then extended to the cases of distribution of $D_{(1)} | D_T$.

For $T \geq 1$, let $X = D_1 | D_T = T$ and $Y = D_1 | D_T = T+1$

Let X and Y be binomial distributed, with $X \sim \text{Bin}(T, p)$ and $Y \sim \text{Bin}(T+1, p)$.

We want to show that X is stochastically smaller than Y . To achieve this, we only need to show that for

any k , $\text{Prob}(Y \leq k) \leq \text{Prob}(X \leq k)$ and that for some k $\text{Prob}(Y \leq k) < \text{Prob}(X \leq k)$. Since

$$P(Y \leq k) = pP(X \leq k-1) + (1-p)P(X \leq k) = P(X \leq k) - p\{P(X \leq k) - P(X \leq k-1)\} = P(X \leq k) - pP(X=k)$$

$< P(X \leq k)$ with “=” hold only as $p=0$.

Hence, X (i.e. $D_1 | D_T = T$) is stochastically smaller than Y (i.e. $D_1 | D_T = T+1$), that is,

$$\text{Prob}(D_1 \leq k | D_T = T+1) \leq \text{Prob}(D_1 \leq k | D_T = T) \text{ for any } k \quad (*1)$$

Similarly,

$$\text{Prob}(D_2 \leq k | D_T = T+1) \leq \text{Prob}(D_2 \leq k | D_T = T) \text{ for any } k \quad (*2)$$

Now we will extend above result to show that $D_{(1)} | D_T = T$ (minimum of the binomial distribution) is stochastically smaller than $D_{(1)} | D_T = T+1$. We only need to show that for any k ,

$$\text{Prob}(D_{(1)} \leq k | D_T = T+1) \leq \text{Prob}(D_{(1)} \leq k | D_T = T) \quad (*)$$

If $k \geq T/2$, the right side of (*) is 1 since at least one of the groups must have fewer than k , so (*) will hold.

If $k < T/2$,

$\text{Prob}(D_{(1)} \leq k | D_T = T+1) = P(D_1 \leq k | D_T = T+1) + P(D_2 \leq k | D_T = T+1)$ since only one group will be $\leq k$ with the other being $\geq D - k$

$$\text{Prob}(D_{(1)} \leq k | D_T = T) = P(D_1 \leq k | D_T = T) + P(D_2 \leq k | D_T = T)$$

By (*1) and (*2), it is clear that (*) will hold for $k < T/2$.

Hence, the distribution of $D_{(1)} | D_T$ is stochastically smaller for small D_T since for any $p > 0$ for the smaller D_T .

Proposition A.1-3: The distribution of $D_T | D$ is stochastically smaller under the setting $(1:r_1, r_2)$ than $(1:1,1)$.

Pf:

Since $D_T | D \sim \text{Bin}(D, p)$ as $p = (r_1 + r_2) / (I + r_1 + r_2)$ for setting $(1:r_1, r_2)$

and $D_T | D \sim \text{Bin}(D, 2/3)$ for setting $(1:1,1)$.

For $r_1, r_2 \leq 1$ with at least one strict inequality holds, $(r_1 + r_2) / (I + r_1 + r_2) < 2/3$.

We now show that $\text{Bin}(D, p)$ is stochastically smaller than $\text{Bin}(D, 2/3)$ for $p < 2/3$.

For any k , $\text{Prob}(D_T \leq k | D) = (D - k) \binom{D}{k} \int_0^{1-p} t^{D-k-1} (1-t)^k dt$ through incomplete Beta

And for $p < 2/3$,

$$\begin{aligned} & \text{Prob}(D_T \leq k | D, p) - \text{Prob}(D_T \leq k | D, 2/3) \\ &= (D - k) \binom{D}{k} \int_0^{1-p} t^{D-k-1} (1-t)^k dt - (D - k) \binom{D}{k} \int_0^{1-2/3} t^{D-k-1} (1-t)^k dt \\ &= (D - k) \binom{D}{k} \int_{1-2/3}^{1-p} t^{D-k-1} (1-t)^k dt > 0 \text{ since } t^{D-k-1} (1-t)^k > 0 \text{ on } (1/3, 1) \end{aligned}$$

Since $(r_1 + r_2) / (I + r_1 + r_2) < 2/3$, $\text{Prob}(D_T \leq k | D)$ under setting $(1:r_1, r_2)$ is larger than $\text{Prob}(D_T \leq k | D)$ under setting $(1:1,1)$. Hence $D_T | D$ under $(1:r_1, r_2)$ is stochastically smaller than $(1:1,1)$.

Proposition A.1-4: To prove $P_{(1:r_1, r_2)}(D_{(1)} \leq d_{(1)} | D) \geq P_{(1:1,1)}(D_{(1)} \leq d_{(1)} | D)$ for all possible $d_{(1)}$.

Pf:

Combine Proposition A.1-1, A.1-2 and A.1-3.

Let $P_{(1:r_1, r_2)}(D_{(1)} | D)$ denote $P(D_{(1)} | D)$ under the setting $(1:r_1, r_2)$ and $P_{(1:1,1)}(D_{(1)} | D)$ denote $P(D_{(1)} | D)$ under the setting $(1:1,1)$.

$$\begin{aligned} P_{(1:r_1, r_2)}(D_{(1)} \leq k | D) &= \sum_{D_T=0}^D P_{(1:r_1, r_2)}(D_{(1)} \leq k | D, D_T) P_{(1:r_1, r_2)}(D_T | D) \\ &\geq \sum_{D_T=0}^D P_{(1:1,1)}(D_{(1)} \leq k | D, D_T) P_{(1:r_1, r_2)}(D_T | D) \text{ By Proposition A.1-1} \end{aligned}$$

$$\begin{aligned} &\geq \sum_{D_T=0}^D P_{(1:1,1)}(D_{(1)} \leq k \mid D, D_T) P_{(1:1,1)}(D_T \mid D) \\ &= P_{(1:1,1)}(D_{(1)} \leq k \mid D) \end{aligned}$$

The last inequality holds since $\sum_{D_T=0}^D P_{(1:1,1)}(D_{(1)} \leq k \mid D, D_T) P_{(1:r_1, r_2)}(D_T \mid D)$ could be seen as

$E(P_{(1:1,1)}(D_{(1)} \leq k \mid D, D_T))$ given D_T is distributed as it would be under $(1:r_1, r_2)$ and the

$P_{(1:1,1)}(D_{(1)} \leq k \mid D, D_T)$ is a decreasing function on D_T by **Proposition A.1-2**. Also since D_T under setting $(1:r_1, r_2)$ is stochastically smaller than setting $(1:1,1)$ by **Proposition A.1-3**, and the proposition for stochastically smaller: If X is stochastically smaller than Y , then for every non-decreasing function g , $E(g(X)) \leq E(g(Y))$.

$$\begin{aligned} &\sum_{D_T=0}^D P_{(1:1,1)}(D_{(1)} \leq k \mid D, D_T) P_{(1:r_1, r_2)}(D_T \mid D) \\ &= E(P_{(1:1,1)}(D_{(1)} \leq k \mid D, D_T)) \text{ given } D_T \mid D \text{ is distributed as it would be under } (1:r_1, r_2) \\ &\geq E(P_{(1:1,1)}(D_{(1)} \leq k \mid D, D_T)) \text{ given } D_T \mid D \text{ is distributed as it would be under } (1:1,1) \text{ by the} \end{aligned}$$

proposition.

A.2 Unbiased Properties for Any Rejection Regions Satisfying LLQR For $K = 2$

Hence, once the rejection region satisfying LLQR is defined, for each given D_C , the rejection region satisfying LLQR is equivalent to reject if $D_{(1)} \leq t(D_C)$ where $t(D_C)$ is a nonnegative integer defined once the rejection region satisfying LLQR is decided. Also, by LLQR $t(D_C)$ is non-decreasing with D_C , i.e. $t(d_C+1) \geq t(d_C)$. Now, we use these two properties to prove that the LLQR rejection region is unbiased. One thing to be notified is that, for a given D_C and D , $D_T = D - D_C$ is also determined. Hence, we have the same information for given $\{D_C \text{ and } D\}$ or given $\{D_T \text{ and } D\}$.

Proposition A.2-1: For a given D_C ,

$$P_{(1:r_1, r_2)}(\text{reject through LLQR} \mid D_C, D) \geq P_{(1:1,1)}(\text{reject through LLQR} \mid D_C, D)$$

Pf:

By Proposition A.1-1, we have:

$$P_{(1:r_1, r_2)}(\text{reject through LLQR} \mid D_C, D) = P_{(1:r_1, r_2)}(D_{(1)} \leq t(D_C) \mid D_C, D)$$

$$\geq P_{(1,1,1)}(D_{(1)} \leq t(D_C) \mid D_C, D) = P_{(1,1,1)}(\text{reject through LLQR} \mid D_C, D)$$

Proposition A.2-2: $P(\text{reject through LLQR} \mid D_C = d_C + 1, D) \geq P(\text{reject through LLQR} \mid D_C = d_C, D)$

Pf:

The inequality holds since by Proposition A.1-2 and $t(d_C + 1) \geq t(d_C)$ (from LLQR),

$$\begin{aligned} & P(\text{reject through LLQR} \mid D_C = d_C + 1, D = d) \\ &= P(D_{(1)} \leq t(d_C + 1) \mid D_C = d_C + 1, D = d) \\ &= P(D_{(1)} \leq t(d_C + 1) \mid D_T = d - d_C - 1, D = d) \\ &\geq P(D_{(1)} \leq t(d_C + 1) \mid D_T = d - d_C, D = d) && \text{by Proposition A.1-2} \\ &= P(D_{(1)} \leq t(d_C + 1) \mid D_C = d_C, D = d) \\ &\geq P(D_{(1)} \leq t(d_C) \mid D_C = d_C, D = d) && \text{by } t(d_C + 1) \geq t(d_C) \text{ LLQR} \\ &= P(\text{reject through LLQR} \mid D_C = d_C, D = d). \end{aligned}$$

Proposition A.2-3: The distribution of $D_C \mid D$ is stochastically larger under the setting $(1:r_1, r_2)$ than $(1,1,1)$.

Pf:

Since Proposition A.1-3 and $D_C = D - D_T$, **Proposition A.2-3** is the same statement as **Proposition A.1-3**.

Proposition A.2-4: Conditional on D , $P_{(1:r_1, r_2)}(\text{Reject using LLQR}) \geq P_{(1,1,1)}(\text{Reject using LLQR})$

Pf:

Combine Proposition A.2-1, A.2-2, and A.2-3,

$$\begin{aligned} P_{(1:r_1, r_2)}(\text{Reject using LLQR} \mid D) &= \sum_{d_C=0}^D P_{(1:r_1, r_2)}(\text{Reject using LLQR} \mid D_C = d_C, D) P_{(1:r_1, r_2)}(D_C \mid D) \\ &\geq \sum_{d_C=0}^D P_{(1,1,1)}(\text{Reject using LLQR} \mid D_C = d_C, D) P_{(1:r_1, r_2)}(D_C \mid D) && \text{by Proposition A.2-1} \\ &\geq \sum_{d_C=0}^D P_{(1,1,1)}(\text{Reject using LLQR} \mid D_C = d_C, D) P_{(1,1,1)}(D_C \mid D) && (*) \\ &= P_{(1,1,1)}(\text{Reject using LLQR} \mid D) \end{aligned}$$

The last inequality (*) holds since (i) $P(\text{reject through LLQR} | D_C = d_C, D)$ is increasing on D_C by

Proposition A.2-2, (ii) by **Proposition A.2-3**, $D_C | D$ is stochastically larger under the setting $(1:r_1, r_2)$ than $(1:1,1)$, and (iii) the property for stochastically smaller: If X is stochastically smaller than Y , then for every non-decreasing function g , $E(g(X)) \leq E(g(Y))$. Therefore,

$$\begin{aligned} & \sum_{d_C=0}^D P_{(1:1,1)}(\text{Reject using LLQR} | D_C = d_C, D) P_{(1:r_1, r_2)}(D_C | D) \\ &= E\left(P_{(1:1,1)}(\text{Reject using LLQR} | D_C = d_C, D)\right) \text{ given distribution of } D_C | D \text{ under } (1:r_1, r_2) \\ &\geq E\left(P_{(1:1,1)}(\text{Reject using LLQR} | D_C = d_C, D)\right) \text{ given distribution of } D_C | D \text{ under } (1:1,1) \\ &= \sum_{d_C=0}^D P_{(1:1,1)}(\text{Reject using LLQR} | D_C = d_C, D) P_{(1:1,1)}(D_C | D) \end{aligned}$$

A.3 Unbiased Properties for Rejection Region Based On LLQR for Unequally Allocation For $K=2$

The extension of unbiased properties to unequally allocation ratio, we would like to show that

$$P_{(\rho:r_1, r_2)}(D_{(1)} \leq d_{(1)} | D) \geq P_{(\rho:1, 1)}(D_{(1)} \leq d_{(1)} | D) \text{ for all possible } d_{(1)}. \text{ As proving Proposition A.1-1, A.2-1,}$$

A.1-2, and A.2-2, the proof only use the comparison of the probability between new treatment groups and which does not related to the incident rate in the control group. And the Proposition A.1-3 and A.2-3 remain correct for unequally allocation. The proof of unbiased property for unequal allocation for $K=2$ is similar to the proposition A.1-4 or A.2-4.

Appendix B: Proof that the Cumulative Binomial Metric under Equal Allocation is LLQR

Section 3.4.4 developed the cumulative binomial metric to construct the rejection region for the global hypothesis with a smaller cumulative binomial value being more evidence to reject H_0 . We now show that the cumulative binomial rejection region for H_0 is LLQR for equal allocation. Showing the LLQR property is equivalent to show the following: (1) under H_0 for a given $D_{(1)}$, the cumulative binomial metric value will be smaller for a larger D_C ; and (2) under H_0 for a given D_C , the cumulative binomial metric value will be smaller for a smaller $D_{(1)}$.

Now, we only consider equal allocation and let

$$\text{CB}(D_C = d_C, D_{(1)} = d_{(1)}) = \text{Prob}(D_{(1)} \leq d_{(1)} \mid D_C + D_{(1)} = d_C + d_{(1)})$$

be the cumulative binomial value applied to any cell ($D_C = d_C, D_{(1)} = d_{(1)}$) given $D_{(1)} \mid D_C + D_{(1)}$ is Binomial ($d_C + d_{(1)}, 1/2$).

Proposition B.1: The cumulative binomial metric value will be smaller for larger D_C given the same $D_{(1)}$ with induction, i.e.

$$\text{CB}(D_C = d_C, D_{(1)} = d_{(1)}) \geq \text{CB}(D_C = d_C + 1, D_{(1)} = d_{(1)}).$$

Pf: Letting X corresponding to $D_{(1)} \mid D_C + D_{(1)} = n$ and Y corresponding to $D_{(1)} \mid D_C + D_{(1)} = n+1$ for $n = d_C + d_{(1)}$. The proposition B-1 is equivalent to show that for $n = d_C + d_{(1)}$,

$$\text{Prob}(X \leq t) \geq \text{Prob}(Y \leq t) \text{ where } X \sim \text{Bin}(n, 1/2) \text{ and } Y \sim \text{Bin}(n+1, 1/2)$$

and the proof is similar to the proof that has been shown in Proposition A.1-2 in Appendix A.

Proposition B.2: The cumulative binomial metric value will be smaller for smaller $D_{(1)}$ given the same D_C with induction, i.e.

$$\text{CB}(D_C = d_C, D_{(1)} = d_{(1)}) \geq \text{CB}(D_C = d_C, D_{(1)} = d_{(1)} - 1).$$

Pf: Letting X corresponding to $D_{(1)} \mid D_C + D_{(1)} = n$ and Z corresponding to $D_{(1)} \mid D_C + D_{(1)} = n-1$ for $n = d_C + d_{(1)}$. The proposition B-2 is equivalent to show that as $n = d_C + d_{(1)}$,

$$\text{Prob}(X \leq t) \geq \text{Prob}(Z \leq t-1) \text{ where } X \sim \text{Bin}(n, 1/2) \text{ and } Z \sim \text{Bin}(n-1, 1/2).$$

$$\begin{aligned} \text{Prob}(X \leq t) &= (n-t) \binom{n}{t} \int_0^{1/2} s^{n-t-1} (1-s)^t ds \\ &= (n-t) \frac{n}{t} \binom{n-1}{t-1} \int_0^{1/2} s^{n-t-1} (1-s)^{t-1} (1-s) ds \end{aligned}$$

$$\begin{aligned}
&\geq (n-t) \frac{n}{t} \binom{n-1}{t-1} \frac{1}{2} \int_0^{1/2} s^{n-t-1} (1-s)^{t-1} ds && \text{inequality holds since } (1-s) \geq 1/2 \text{ for } 0 \leq s \leq 1/2 \\
&= \frac{n}{2t} (n-t) \binom{n-1}{t-1} \int_0^{1/2} s^{n-t-1} (1-s)^{t-1} ds \\
&\geq (n-t) \binom{n-1}{t-1} \int_0^{1/2} s^{n-t-1} (1-s)^{t-1} ds && \text{inequality holds for } t \leq n/2 \\
&= \text{Prob}(Z \leq t-1)
\end{aligned}$$

By Proposition B.1 and B.2, the cumulative binomial metric has smaller value as D_C increases or $D_{(1)}$ decreases and will thus reject the null from $(D_C + i, D_{(1)})$ or $(D_C, D_{(1)} - j)$ for any possible integer i or j first before rejecting $(D_C, D_{(1)})$. The rejection region based on cumulative binomial metric thus satisfies LLQR for equal allocation.

Appendix C: Power Computations under the alternative (1:r,1)

For $K=2$ and under the alternative setting (1:r,1) for $r<1$, power is defined as the probability to reject the right new treatment, the new treatment 1, and is obtained by summing over the probabilities of all D_C , D_1 that are in the rejection region [NOTE the change here from $D_{(1)}$ to D_1 as we specifically want to be sure that treatment 1 which is better than the control is rejected as opposed to potentially rejecting treatment 2 which is not better than the control.]

$P(\text{Reject new treatment 1}|D) = P((D_C, D_1) \text{ is in the rejection region} | D)$

$$= \sum_{\substack{\text{all possible} \\ D_C=d_C, D_1=d_1}} P(D_C = d_C, D_1 = d_1 | D) \times I\{ (D_C = d_C, D_1 = d_1) \text{ is in the rejection region} | D \} \quad (\text{Eq. C.1})$$

Where $I\{ (D_C = d_C, D_1 = d_1) \text{ is in the rejection region} | D \}$ is an indicator function that equals 1 if observed $(D_C = d_C, D_1 = d_1)$ rejects the new treatment 1 and equals to 0 otherwise.

$D_C \backslash D_1$	0	1	2	3	4	5	6	7	8	9	10
0	0.0001	0.0005	0.0012	0.0016	0.0014	0.0008	0.0003	<0.0001	<0.0001	<0.0001	<0.0001
1	0.0010	0.0047	0.0094	0.0110	0.0083	0.0041	0.0014	0.0003	<0.0001	<0.0001	
2	0.0047	0.0189	0.0330	0.0330	0.0206	0.0083	0.0021	0.0003	<0.0001		
3	0.0126	0.0440	0.0661	0.0551	0.0275	0.0083	0.0014	0.0001			
4	0.0220	0.0661	0.0826	0.0551	0.0206	0.0041	0.0003				
5	0.0264	0.0661	0.0661	0.0330	0.0083	0.0009					
6	0.0220	0.0440	0.0330	0.0110	0.0014						
7	0.0126	0.0189	0.0094	0.0016							
8	0.0047	0.0047	0.0012								
9	0.0010	0.0005									
10	0.0001										

Illustration C-1 Probabilities of all possible combinations of (D_C, D_1) given $D = 10$ under the setting (1: $r_1 = 0.5, 1$)

For the first part of the (Eq. C.1), the $P(D_C = d_C, D_1 = d_1|D)$ for all possible combinations of $\{D_C, D_1\}$ is obtained through the distribution of D_C, D_1 and D_2 given total diseases cases D which is multinomial($D, 1/(2+r), r/(2+r), 1/(2+r)$). Here, we fill in the probabilities for each cell of the all possible combinations of $\{D_C, D_1\}$. Again, we illustrate this for $D=10$ cases and now under the alternative (1: $r_1 = 0.5, 1$), the Illustration C-1 contains all the possible combinations of $(D_C = d_C, D_1 = d_1)$; under the constraint $d_C + d_1 < D$. The probabilities that $(D_C = d_C, D_1 = d_1, D_2 = D - (d_C + d_1))$ from the Multinomial distribution $(D, 1/(2+0.5), 0.5/(2+0.5), 1/(2+0.5))$ are filled into the respective Table cells. For example the value 0.0094 (bolded probability) in the table below is calculated by

$$P(D_C = 1, D_{T_1} = 2 | D = 10) = \binom{10}{1 \ 2 \ 7} \left(\frac{1}{2.5}\right)^1 \left(\frac{0.5}{2.5}\right)^2 \left(\frac{1}{2.5}\right)^7.$$

For the second part of (Eq. C.1), $I\{(D_C = d_C, D_1 = d_1) \text{ is in the rejection region } | D\}$, we will now construct the rejection region for the combinations of $\{D_C, D_1\}$ using different approaches and apply them to this example. The following subsections discuss how to construct the rejection region defined by $\{D_C, D_1\}$ through the two methods (described in Chapter 4), FLSDA and Bonferroni adjustments, using combinations of (D_C, D_1) for $K=2$. After rejection region is determined, the power to reject new treatment 1 under $(1: r_1 = r, 1)$ for each approach is then the summation over all the multinomial probabilities of $\{D_C, D_1\}$ which fall in the rejection region.

C.1 Rejection Region for FLSDA for Given Total Disease Cases D

In FLSDA approach (Section 4.1), in order to conclude that new treatment 1 is superior to the control, we have to first reject the global hypothesis test that at least one of the two new treatments is better than the control and if this hypothesis is rejected then reject an individual test for new treatment 1 being superior to the control in step 2. This is equivalent to rejecting $(D_C = d_C, D_1 = d_1)$ if (1) the global hypothesis is rejected at the overall α from $(D_C = d_C, D_{(1)} = d_{(1)})$ **and** also (2) the conditional binomial p-value of $D_1 = d_1 | D_C + D_1 = d_C + d_1 \leq \alpha$, i.e. $P(D_1 \leq d_1 | D_C + D_1 = d_C + d_1) = \sum_{t \leq d_1} \binom{d_C + d_1}{t} \left(\frac{1}{2}\right)^{d_C + d_1} \leq \alpha$, where $D_1 | D_C + D_1 = d_C + d_1$ is binomial distributed with parameters $d_C + d_1, 1/2$ under null hypothesis. Here, the rejection of the global hypothesis is a prerequisite but does not guarantee that new treatment 1 will be found superior to the control. For example if the minimum value observed was from new treatment 2 then the new treatment 1 will not be rejected as being superior to the control if the conditional p-value for $D_1 \leq d_1 | (d_C + d_1)$ is greater than α even though the global hypothesis is rejected.

Now the details for implementation to construct this rejection region: STEP 1 - Among the combinations of $(D_C = d_C, D_1 = d_1)$, first identify the values of (d_C, d_1) which fall in the global rejection region (using the fact that $(d_C, d_{(1)})$ is determined by (d_C, d_1) through $d_2 = D - d_C - d_1$) and then STEP 2 - find the values of (d_C, d_1) in the global rejection region from STEP 1 for which the conditional Binomial p-value is less than α , i.e. $P(D_1 \leq d_1 | d_1 + d_C) \leq \alpha$ under the individual null hypothesis of equality.

Note that the rejection region for global hypothesis was expressed in terms of $(D_C, D_{(1)})$ in section 3.4, denoted as $R_\alpha^{(1)}$. Here, we display the same rejection region in terms of the combination of (D_C, D_1) , and denote this 2-dimensional table as R_α^1 . Now R_α^1 can be determined from $R_\alpha^{(1)}$ since the observed $(D_C = d_C, D_1 = d_1)$ will reject the global hypothesis in two situations:

- (i) D_1 is the minimum of $(D_1$ and $D_2)$ and $(D_C = d_C, D_{(1)} = d_{(1)})$ is in $R_\alpha^{(1)}$. or
- (ii) $D_2 = d - d_C - D_1$ is the minimum of $(D_1$ and $D_2)$ and $(D_C = d_C, D_{(1)} = d - d_C - d_1)$ is in $R_\alpha^{(1)}$.

$D_C \backslash D_{(1)}$	0	1	2	3	4	5
0						
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						

Illustration C-2 Rejection region using $(D_C, D_{(1)})$, $R_\alpha^{(1)}$ (shaded) for $K=2$, $D=10$ and size $\alpha=0.05$ for the global hypothesis, the cell $(D_C=6, D_{(1)}=0)$ which used in example is marked in red and the cell $(D_C=6, D_{(1)}=1)$ is marked in blue.

$D_C \backslash D_1$	0	1	2	3	4	5	6	7	8	9	10
0											
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											

Illustration C-3 Rejection region using (D_C, D_1) , R_α^1 (shaded) for $K=2$, $D=10$, and size $\alpha = 0.05$ for the global hypothesis, corresponding to $R_\alpha^{(1)}$ in Illustration C-2. For example, the cells $(D_C = 6, D_1 = 0)$ and $(D_C = 6, D_1 = 4)$ marked in red are rejected in R_α^1 since $(D_C=6, D_{(1)}=0)$ is rejected in $R_\alpha^{(1)}$ and the cells $(D_C = 6, D_1 = 1)$ or $(D_C = 6, D_1 = 3)$ marked in blue is not rejected in R_α^1 as $(D_C=6, D_{(1)}=1)$ is not rejected in $R_\alpha^{(1)}$

When $K=2$, $D=10$ and size $\alpha=0.05$, the $R_\alpha^{(1)*}$ is shown in Illustration C-2 (shaded area) and then the corresponding R_α^1 (shaded in Illustration C-3) is determined from it. For example, since the $(D_C = 6, D_{(1)}=0)$ is in the shaded rejection region of $R_\alpha^{(1)}$, the corresponding observed combinations $(D_C = 6, D_1 = 0)$ or $(D_C = 6, D_1 = 4, \text{ which means } D_2 = 0)$ are in the shaded rejection region of R_α^1 (each having $D_{(1)} = 0$). For other cells which do not reject the global hypothesis from $R_\alpha^{(1)}$, such as $(D_C = 6, D_{(1)}=1)$, the corresponding combination $(D_C = 6, D_1=1)$ or $(D_C = 6, D_1=3 \text{ and } D_2=1)$, is not in the rejection region of R_α^1 .

But now that we have described R_α^1 , for the STEP 2 of FLSDA approach, the rejected cell need to be only that part of R_α^1 that also satisfies the individual conditional binomial p-value,

$$P(D_1 \leq d_1 | D_C + D_1 = d_C + d_1) = \sum_{t \leq d_1} \binom{d_C + d_1}{t} \left(\frac{1}{2}\right)^{d_C + d_1} \leq \alpha.$$

Hence, the rejection region for FLSDA will be the subset of R_α^1 where $\sum_{t \leq d_1} \binom{d_C + d_1}{t} \left(\frac{1}{2}\right)^{d_C + d_1}$ are less than or equals to α .

Continued with previous example for $K = 2$, $D = 10$, and $\alpha = 0.05$, in Illustration C-4, the p-values of $D_1 \leq d_1 | D_C + D_1 = d_C + d_1$ are filled in only for the cells of $(D_C = d_C, D_1 = d_1)$ falling in R_α^1 . For example, the number 0.035 in the $D_C = 7, D_1 = 1$ cell in the table below is the conditional probability that $D_1 \leq 7$ given $d_1 + d_C = 8$ under the null hypothesis of equality for controls and new treatment 1. Since this is ≤ 0.05 , it falls into the final rejection region. However, for $D_C = 4, D_1 = 0$ cell the conditional probability that $D_1 \leq 0 | d_1 + d_C = 4$ under equality for controls and new treatment 1 is $0.062 > 0.05$. Thus, even though $d_C = 4, d_1 = 0$ falls in R_α^1 for the minimum to be less than the control, the individual test of new treatment 1 against the control is not rejected so $d_C = 4, d_1 = 0$ does not fall in the FLSDA rejection region. The bolded region is the rejection region for FLSDA which is the overlaid rejected cells for global hypothesis with the p-values ≤ 0.05 .

* $R_\alpha^{(1)}$ is the rejection region for global hypothesis based on “conditional binomial” we discussed in Section 4.4

$D_C \backslash D_1$	0	1	2	3	4	5	6	7	8	9	10
0											
1											
2											
3											
4	0.062						0.828				
5	0.031					0.623					
6	0.016				0.377						
7	0.008	0.035	0.090	0.172							
8	0.004	0.020	0.055								
9	0.002	0.011									
10	0.001										

Illustration C-4 Conditional p-values for the individual hypothesis of Treatment 1 versus the control conditioning on $d_C + d_1$; filled in on the global rejection region $R_\alpha^{(1)}$ that rejects in the global hypothesis for the minimum treatment being less than the control for $D=10$, $K=2$, and $\alpha=0.05$. The bolded enclosed region covers the conditional p-values $\leq \alpha$

C.2 Rejection Region for Bonferroni Adjustment for Given Total Disease Cases D

For Bonferroni adjustment (Section 4.2), the reject region to reject new treatment 1 for each $D_C + D_1$ is:

$$D_1 \leq \max\{d_1 : \sum_{t \leq d_1} \binom{d_C + d_1}{t} \left(\frac{1}{2}\right)^{d_C + d_1} \leq \alpha/K\}.$$

That is equivalent to the rejection region of (D_C, D_1) where the conditional binomial p-values

$$\sum_{t \leq d_1} \binom{d_C + d_1}{t} \left(\frac{1}{2}\right)^{d_C + d_1} \leq \alpha/K.$$

As in the previous example, for $K=2$, $D=10$ and $\alpha=0.05$, for $D_C + D_1=10$,

$$P(D_1 = 0 \mid D_C + D_1=10) = \binom{10}{0} \left(\frac{1}{2}\right)^{10} = 0.001 < 0.025,$$

$$P(D_1 \leq 1 \mid D_C + D_1=10) = \sum_{t \leq 1} \binom{10}{t} \left(\frac{1}{2}\right)^{10} = 0.011 < 0.025, \text{ and}$$

$$P(D_1 \leq 2 \mid D_C + D_1=10) = \sum_{t \leq 2} \binom{10}{t} \left(\frac{1}{2}\right)^{10} = 0.055 > 0.025.$$

Hence, for $D_C + D_1=10$, the rejected cells are $\{(D_C=10, D_1=0), (D_C=9, D_1=1)\}$. As we collect all rejected

cells for different $D_C + D_1 = d_C + d_1$ that $\sum_{t \leq d_1} \binom{d_C + d_1}{t} \left(\frac{1}{2}\right)^{d_C + d_1} \leq 0.025$, it is the rejection region for

Bonferroni adjustment. Those conditional p-values could be found in Illustration C-5 which is the table

filled in $\sum_{t \leq d_1} \binom{d_C + d_1}{t} \left(\frac{1}{2}\right)^{d_C + d_1}$ for each possible combination of (D_C, D_1) . In Illustration C-5, we also

identify (D_C, D_1) that the p-values that is less than or equals to $\alpha/2 = 0.025$ which is equivalent to rejection region created by Bonferroni adjustment. The bolded region the rejection region by Bonferroni adjustment with $\alpha=0.05$.

$D_C \backslash D_1$	0	1	2	3	4	5	6	7	8	9	10
0	1	1	1	1	1	1	1	1	1	1	1
1	0.500	0.750	0.875	0.938	0.969	0.984	0.992	0.996	0.998	0.999	
2	0.250	0.500	0.688	0.812	0.891	0.938	0.965	0.980	0.989		
3	0.125	0.312	0.500	0.656	0.773	0.855	0.910	0.945			
4	0.062	0.187	0.344	0.500	0.637	0.746	0.828				
5	0.031	0.109	0.227	0.363	0.500	0.623					
6	0.016	0.063	0.145	0.254	0.377						
7	0.008	0.035	0.090	0.172							
8	0.004	0.020	0.055								
9	0.002	0.011									
10	0.001										

Illustration C-5 Conditional p-values to reject New Treatment 1 being equal to the control conditioning on $d_C + d_1$ for $D=10, K=2$. The bolded region covers all conditional p-values ≤ 0.025 and is the rejection region based on Bonferroni adjustment for the case $D=10, K=2$, and $\alpha=0.05$.

C.3 Power Computation after the Rejection Region Is Determined

Once a specific D , size level α , and the rejection region are determined, the power for the alternative $(1:r_1 = r, 1)$ is the probability that the combinations (D_C, D_1) fall in the rejection region under alternative setting. And, for $K=2$, the probability of $(D_C = d_C, D_1 = d_1)$ given D for any possible combination is

$$P(D_C = d_C, D_1 = d_1 | D = d) = \binom{d}{d_C \quad d_1 \quad d - d_C - d_1} \left(\frac{1}{2+r}\right)^{d-d_1} \left(\frac{r}{2+r}\right)^{d_1}.$$

Hence, the power is obtained by summing over all probabilities that (D_C, D_1) is in the determined rejection region under $(1: r_1 = r, 1)$.

For example when $D=10, \alpha = 0.05$ and $r = 0.01$, the rejection region is given by FLSDA shown as Illustration C-4 in C.1. Under the alternative $(1:r_1=0.01,1)$, the multinomial probabilities for rejection

region cells ($D_C = d_C, D_1 = d_1$) are filled in the following table (Illustration C-6). The overall power from summing over these filled probabilities in the rejection region is about 59.7%.

$D_C \backslash D_1$	0	1	2	3	4	5	6	7	8	9	10
0											
1											
2											
3											
4											
5	0.2341										
6	0.1951										
7	0.1115	0.0033									
8	0.0418	0.0008									
9	0.0093	0.0001									
10	0.0009										

Illustration C-6 Multinomial probability of combinations of (D_C, D_1) that are in the given rejection region for given $D=10$ under the setting (1: $r_1 = 0.01, 1$).

Appendix D Introduction to Design C

Gail's Design A ended the trial at a certain numbers of follow up disease cases and Design B ended the trial after a certain number of subject time. Here, we propose a third design that ends follow up when certain numbers of disease cases accrued in the control group, D_C . Design C is similar to the following game strategy with 3 people hitting baseballs at the same rate so that the number of foul ball each person has hit at a given time is a Poisson process with the parameter based on how skillful the person is. One person (C) is already a member of our baseball team. We want to see if either of the two new persons (T_1 or T_2) hits foul balls with lower intensity than C. We continue to compete until person C hits 10 foul balls. The final results will be examining the numbers of balls hit by the lowest person T_1 or T_2 given stoppage when C hits 10 to see if there is statistical evidence that the best of these two people hits fewer foul balls than C.

For Design C, the numbers of disease cases (i.e foul balls for the previous example) in each new treatment, D_k , has a Negative Binomial distribution. Let i_C and i_k be the incidence rate per-subject time in the control group and in the k th new treatment group, respectively for $k=1,2,\dots,K$. For Design C, the numbers of disease cases in the control group, D_C , is predetermined, but the trial length for this design is unknown and depends on D_C . The waiting time for one disease case to occur in the control group could be seen as exponential with rate $1/i_C$ and each subsequential disease case in the control group occurs randomly and independent of previous cases. Therefore, the length of time for D_C disease cases to occur, N_C , is the sum of independent exponential($1/i_C$) random variables which follows a Gamma distribution with $\alpha = D_C$, and $\beta = 1/i_C$. Once the follow up time, N_C , is known, the numbers of disease cases occur in the k th new treatment group D_k during this N_C period time follows a Poisson distribution with Poisson rate $i_k N_C$ and no longer dependent on D_C . (Note that D_k only depends on the length of time, hence D_k is independent of D_C once N_C is determined; i.e. $P(D_k = d_k | D_C = d_C, N_C = t) = P(D_k = d_k | N_C = t)$). Hence, the numbers of disease cases in each new treatment group for design C follows a Negative Multinomial distribution:

$$\begin{aligned}
 N_C | D_C &\sim \text{Gamma}(D_C, 1/i_C); D_k | N_C \sim \text{Poisson}(i_k N_C) \text{ for } k = 1, 2, \dots, K \\
 P(D_1 = d_1, \dots, D_k = d_k | D_C = d_C) &= \int P(D_1 = d_1, \dots, D_k = d_k, N_C = t | D_C = d_C) dt \\
 &= \int P(D_1 = d_1, \dots, D_k = d_k | D_C = d_C, N_C = t) P(N_C = t | D_C = d_C) dt
 \end{aligned}$$

$$\begin{aligned}
&= \int P(D_1 = d_1, \dots, D_k = d_k \mid N_C = t) P(N_C = t \mid D_C = d_C) dt \\
&= \int \prod_{k=1}^K P(D_k = d_k \mid N_C = t) P(N_C = t \mid D_C = d_C) dt \\
&= \int \prod_{k=1}^K \frac{e^{-i_k t} (i_k t)^{d_k}}{d_k!} \frac{i_C^{d_C}}{\Gamma(d_C)} t^{d_C-1} e^{-i_C t} dt = \frac{1}{\Gamma(d_C)} \prod_{k=1}^K \frac{i_k^{d_k} i_C^{d_C}}{d_k!} \int t^{d_C + \sum_{k=1}^K d_k - 1} e^{-(i_C + \sum_{k=1}^K i_k) t} dt \\
&= \frac{1}{\Gamma(d_C)} \frac{\Gamma(d_C + \sum_{k=1}^K d_k)}{(i_C + \sum_{k=1}^K i_k)^{d_C + \sum_{k=1}^K d_k}} \prod_{k=1}^K \frac{i_k^{d_k} i_C^{d_C}}{d_k!} \\
&= \binom{d_C + \sum_{k=1}^K d_k - 1}{d_C - 1 \quad d_1 \quad \dots \quad d_K} \left(\frac{i_C}{i_C + \sum_{k=1}^K i_k} \right)^{d_C} \prod_{k=1}^K \left(\frac{i_k}{i_C + \sum_{k=1}^K i_k} \right)^{d_k} \sim NMN \left(d_C, \frac{i_C}{i_C + \sum_{k=1}^K i_k}, \frac{i_1}{i_C + \sum_{k=1}^K i_k}, \dots, \frac{i_K}{i_C + \sum_{k=1}^K i_k} \right)
\end{aligned}$$

Then, we could design a study testing rule based on Negative Multinomial distribution. (Note that the marginal distribution of D_k for design C follows by a Negative Binomial for each k).

Again H_0 is $i_1 = i_2 = \dots = i_C$ with the alternative being that at least one of the $i_k < i_C$. We would find the value t for which $P\{\min(D_1, D_2, \dots, D_K) < t\} : H_0 < \alpha$ and reject H_0 iff $D_{(1)} < t$. While substantial new work is needed to evaluate the properties of this design, because the $\sum D_k + D_C$ is no longer constrained to equal D , the D_k are now independent of each other which may facilitate derivation of statistical properties.

Appendix E List of R Functions Implemented In This Thesis

Following is the list of some functions that would be useful for this comparative Poisson design study. For each function, we have the function name followed by parenthesis with some arguments included. Here are some input arguments that shared in the functions:

k – total number of new treatments + 1 control in the comparative Poisson trial

$total.events$ – integer (D in our notation) - the numbers of events occur to end the Poisson trial

ρ – the allocation ratio of the portion of total population in the control group to the portion in each new treatment group (ρ in our notation)

α – size level of the type I error

power – power level

prob – vector of multinomial probabilities; first component is the incidents rate of control group and the rests are the rate from new treatment groups; prob1: multinomial probabilities under null hypothesis; prob2: multinomial probabilities under alternative hypothesis

List of functions with descriptions:

pmf.control.txmin(k, total.events, prob)

Description: Compute the probabilities of each possible combinations ($D_C, D_{(1)}$).

diff.crit(k, total.events, rho=1), Binom.crit(k, total.events, rho = 1), powratio.crit(k, total.events, prob1, prob2)

Description: Rejection region criteria for testing global hypothesis that discussed in section 3.4.

Rej.Reg.Crit(k, total.events, prob.null.matrix, crit, alpha = 0.025)

Description: Rejection region for global hypothesis that is determined through criteria discussed in Section 3.4. Here the argument *prob.null.matrix* is the probability matrix for ($D_C, D_{(1)}$) under null hypothesis.

Rej.Reg.Compar(k, total.events, prob1, prob2, rho = 1, alpha = 0.025)

Description: Construct all rejection regions discussed in Section 3.4 with corresponding size level and power.

FLSDA_Rej(k=3, alpha=0.05, total.events=10, rho=1, r=0.1)

Description: Compute the size, power for FLSDA_CB and the rejection region constructed by FLSDA_CB which is discussed in Section 4.1 for comparing 2 new treatment to 1 control under alternative ($\rho:r, 1$).

Sample_Size_FLSDA(from=10,power=.80,r=0.1,alpha=0.05,rho=1)

Description: minimum disease cases required to follow up for a certain settings with at least 80% power and size level 0.05 as alternative is $(\rho:r,1)$.

References Cited

- P.C. Allen, H.D. Danforth, and B. L. Vinyard, Development of a Protective Index to Rank Effectiveness of Multiple Treatments Within an Experiment: Application to a Cross Protection study of Several Strains of *Eimeria maxima* and a Live Vaccine., *Avian diseases*, 48:370-375, 2004
- B. Hirschel, Beware of Drug Holidays before HIV Salvage Therapy, *The New England journal of Medicine* 2003 349:827-828.
- C.C. Brown and S.B. Green, Additional Power Computations for Designing Comparative Poisson Trials, *Am. J. Epidemiology*, 115:752-758, 1982
- Corrado, Charles J., The Exact Joint Distribution for the Multinomial Maximum and Minimum and the Exact Distribution for the Multinomial Range (January 2007). Available at SSRN: <http://ssrn.com/abstract=989082>
- William S. Cleveland, A Program for Smoothing Scatterplots by Robust Locally Weighted Regression, *The American Statistician*, Vol. 35, No. 1 (Feb., 1981), p. 54
- William S. Cleveland, Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of the American Statistical Association*, Vol. 74, No. 368 (Dec., 1979), pp. 829-836
- C.W. Dunnett, A Multiple Comparison Procedure for Comparing Several Treatments with a Control, *Journal of the American Statistical Association*, Vol. 50, No. 272, (Dec., 1955), pp. 1096
- M. Gail, Power Computations for Designing Comparative Poisson Trials., *Biometrics*, Vol. 30, No. 2, (Jun., 1974) pp. 231-237
- N.I. Kumwenda, D.R. Hoover, L.M. Mofenson, M.C. Thigpen, G.Kafulafula, Q. Li, L. Mipando, K. Nkanaunena, T.Mebrahtu, M. Bulterys, M.G. Fowler, and T.E. Taha, Extended Antiretroviral Prophylaxis to Reduce Breast-Milk HIV-1 Transmission, *The New England journal of Medicine*, (July 10, 2008), Vol. 359, No. 2.
- J. C. Hsu. *Multiple comparisons: theory and methods* P. 139. Chapman & Hall, 1996
- E.L. Lehmann, Joseph P. Romano, *Testing Statistcal Hypotheses*, 3rd Edition, Springer, 2005
- E. Paulson, On the Comparison of Several Experimental Categories with a Control, *The Annals of Mathematical Statistics*, Vol. 23, No. 2 (Jun. 1952), pp.239-246.
- J. Przyborowski and H. Wilenski, Homogeneity of Results in Testing Samples from Poisson Series, *Biometrika*, Vol. 31, No. 3/4 (Mar., 1940), pp. 313-323
- J. P. Shaffer, Multiple Hypothesis Testing, *Annual Reviews in Psychology*. 1995. 46:561-84
- Hochberg Y, Tamhane AC. *Multiple comparison procedures*. Wiley & Sons. New York NY 1987
- S. M. Ross. *Introduction to probability models*. 7th edition. Harcorut-AP, 2000

Curriculum Vita**Tzu-Lin Hsu**

- 2000-2004 B.S in Mathematics, Tsing-Hua University, Taiwan
- 2005-2010 Ph. D in Statistics, Rutgers University
-
- 2007-2009 Teaching Assistant, Department of Statistics, Rutgers University
- 2009-2010 Graduate Assistant, Department of Statistics, Rutgers University