A COMPARISON OF TESTS FOR HOMOGENEITY OF THE RISK DIFFERENCE WHEN DATA ARE SPARSE

PATRANUCH SAPCHOOKUL

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE (BIOSTATISTICS)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2005

ISBN 974-04-6318-5

Thesis
Entitled

A COMPARISON OF TESTS FOR HOMOGENEITY OF THE RISK DIFFERENCE WHEN DATA ARE SPARSE

……………………………………………..….
Miss Patranuch Sapchookul,
Candidate

…………………………...……...……….
Assoc. Prof. Chukiat Viwatwongkasem,
M.Sc. (Biostatistics)
Major-Advisor

…………………………………….....…....
Assoc. Prof. Thavatchai Vorapongsathorn,
Ph.D. (Research & Stat in Ed.)
Co-Advisor

…………………………………….
Assit. Prof. Pratana Satitvipawee,
M.P.H. (Biostatistics)
Co-advisor

…………………………………………….
Prof. Dankmar Böhning,
Ph.D. (Mathematics)
Co-Advisor

…………………………………….
Assoc. Prof. Rassamidara Hoonsawat,
Ph.D. (Physics)
Dean
Faculty of Graduate Studies

……………….……………………….
Assoc. Prof. Piangchan Rojanavipart,
M.H.S. (Biostatistics)
Chair
Master of Science Programme
In Biostatistics
Faculty of Public Health

Thesis
Entitled

A COMPARISON OF TESTS FOR HOMOGENEITY OF THE RISK DIFFERENCE WHEN DATA ARE SPARSE

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science (Biostatistics)
on
June 6, 2005

……………………………………………….
Miss Patranuch Sapchookul,
Candidate

……………………………………………….
Assoc. Prof. Chukiat Viwatwongkasem,
M.Sc. (Biostatistics)
Chair

……………………………………………….
Assoc. Prof. Thavatchai Vorapongsathorn,
Ph.D. (Research & Stat in Ed.)
Member

……………………………………………
Assoc. Prof. Dusit Sujirarat,
M.Sc. (Biostatistics)
Member

……………………………………………….
Dr. Rujirek Boosarawongse,
Ph.D. (Statistics)
Member

……………………………………………
Assoc. Prof. Rassamidara Hoonsawat,
Ph.D. (Physics)
Dean
Faculty of Graduate Studies
Mahidol University

……………………………………………….
Assoc. Prof. Chalermchai Chaikittiporn,
Dr. P.H. (Epidemiology)
Dean
Faculty of Public Health
Mahidol University

ACKNOWLEDGEMENT

First of all, I would like to acknowledge my heartfelt gratitude and appreciation to my major advisor, Assoc. Prof. Chukiat Viwatwongkasem for his kind instruction, invaluable suggestion and encouragement throughout this study. I am equally grateful to my co-advisor, Prof. Dankmar Böhning, Assoc. Prof. Thavatchai Vorapongsathorn, and Assist. Prof. Pratana Satitvipawee for their constructive comments, supervision and generous assistance. They were never lacking in kindness and support.

I wish to thank Dr. Rujrek Boosarawongse and Assoc. Dusit Sujirarat for their comments, suggestions for improvement, and time to read the thesis and who were the external examiner of the thesis defense.

I am also sincerely grateful to all the lecturers of the Biostatistics Department, Faculty of Public Health, who imparted their knowledge and experiences that really guided me through this study. Thanks also to the staff of the Biostatistics Department for their helps.

Finally and above all I wish to thank my friends who have given me assistance and supports throughout this research period. Last but not least, I would like to acknowledge my appreciation and gratitude to my family for their financial support, entirely care, understanding, endless encouragement, and love through my life.

Patranuch Sapchookul

A COMPARISON OF TESTS FOR HOMOGENEITY OF THE RISK DIFFERENCE WHEN DATA ARE SPARSE

PATRANUCH SAPCHOOKUL 4437166 PHBS/M

M.Sc. (BIOSTATISTICS)

THESIS ADVISORS: CHUKIAT VIWATWONGKASEM, M.Sc.(Biostatistics), DANKMAR BÖHNING, Ph.D.(Mathematics), THAVATCHAI VORAPONGSATHORN, Ph.D. (Research & Stat in Ed.), PRATANA SATITVIPAWEE, M.P.H.(Biostatistics).

ABSTRACT

The objective of this research was to compare the Type I error and the power of tests for homogeneity of risk difference by using the statistical tests following the work of Lipsitz et al. (1998). Hypothesis testing considered 0.01, 0.05 and 0.10 levels of significance for one-sided and two-sided tests and when the number of centers K = 8, 16, 32, and 48 and mean treatment group per center equaled 4, 8, 16, 32, and 64. This study considered when the mean sample size in each treatment group for equal and unequal cases. The data was simulated by Monte Carlo simulation technique compatible with FORTRAN 90 language. A FORTRAN program was designed to calculate the probability of the Type I error and the power of all five tests in 10,000 replications for each condition. When the estimated variance in each center was zero, the study adjusted the statistics by adding 0.5 to each cell (in 2x2 tables) and dropping centers with zero variance.

Results indicated that the Q_{WLS} test performs well when mean treatment group is large ($n_{ij} \geq 32$) with respect to the Type I error and the $Z^2_{WLS,R}, Z^2_V$ tests perform well when mean treatment group is less than moderate ($n_{ij} \leq 16$) by the method of dropping case, regardless of the number of centers. For the adding constant method, only the Z^2_{WLS} test performs well when mean treatment group is less than moderate ($n_{ij} \leq 16$). With respect to the power, by the dropping case method the Q_{WLS} test has the highest power in every number of centers and mean treatment group size. For the adding constant if mean treatment group is less than moderate then the Z^2_K test has the highest power, but if the treatment group is large then the Q_{WLS} test has the highest power. In cases where the number of centers is small and/or the mean treatment group in each center is sparse, the weight least squares statistic (Q_{WLS}) test appears to perform the best.

KEY WORDS: RISK DIFFERENCE / TYPE I ERROR / POWER OF THE TEST / HOMOGENEITY / SPARSE DATA

180 P. ISBN 974-04-6318-5

การเปรียบเทียบสถิติทดสอบสำหรับการเท่ากัน (Homogeneity) ของ Risk Difference เมื่อข้อมูลมีขนาดเล็ก (A COMPARISON OF TESTS FOR HOMOGENEITY OF THE RISK DIFFERENCE WHEN DATA ARE SPARSE)

ภัทรานุช ทรัพย์ชูกุล 4437166 PHBS/M

วท.ม. (ชีวสถิติ)

คณะกรรมการควบคุมวิทยานิพนธ์ : ชูเกียรติ วิวัฒน์วงศ์เกษม, M.Sc.(Biostatistics), Dankmar Böhning, Ph.D. (Mathematics), ธวัชชัย วรพงศธร, Ph.D. (Research & Stat in Ed.), ปรารถนา สถิตย์วิภาวี, M.P.H.(Biostatistics).

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของสถิติภายใต้การเท่ากันของความแตกต่างของอัตราเสี่ยง ด้วยสถิติทดสอบห้าตัวที่ใช้ในงานวิจัยของ Lipsitz et al. ที่ระดับนัยสำคัญ 0.01, 0.05 และ 0.10 โดยพิจารณาทั้งแบบทางเดียวและแบบสองทาง โดยจำนวนศูนย์รักษา คือ 8, 16, 32, 48 ค่าเฉลี่ยของขนาดตัวอย่างในกลุ่มทดลอง และกลุ่มควบคุมมีแบบเท่ากันและไม่เท่ากัน คือ 4, 8, 16, 32 และ 64 ข้อมูลถูกสร้างขึ้นด้วยเทคนิคมอนติคาร์โล โดยใช้ภาษาฟอร์แทรน 90 ในการคำนวณหาค่าความคลาดเคลื่อนชนิดที่ 1 และอำนาจการทดสอบของสถิติทดสอบทั้งห้า ซ้ำๆ กัน 10,000 ครั้ง ในแต่ละสถานการณ์

ผลการศึกษาพบว่า สถิติทดสอบ Q_{WLS} สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ดีเมื่อค่าเฉลี่ยของขนาดตัวอย่างมีขนาดใหญ่ และสถิติทดสอบ $Z^2_{WLS,R}$ และ Z^2_V สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ดีเมื่อค่าเฉลี่ยของขนาดตัวอย่างมีขนาดเล็ก ด้วยการจำลองข้อมูลด้วยวิธี dropping center โดยที่จำนวนศูนย์รักษาจะมีขนาดเท่าไรก็ตาม และสำหรับการจำลองข้อมูลด้วยวิธี adding constant จะมีเพียงสถิติทดสอบ Z^2_{WLS} ที่สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ดี เมื่อค่าเฉลี่ยของขนาดตัวอย่างมีขนาดเล็ก

สำหรับอำนาจการทดสอบพบว่า การจำลองข้อมูลด้วยวิธี dropping center, สถิติทดสอบ Q_{WLS} จะมีอำนาจทดสอบสูงที่สุดในทุกขนาดของศูนย์รักษา และทุกค่าเฉลี่ยของขนาดตัวอย่าง ทั้งแบบเท่ากันและไม่เท่ากัน การจำลองข้อมูลด้วย adding constant, เมื่อค่าเฉลี่ยของตัวอย่างมีขนาดเล็ก สถิติทดสอบ Z^2_K จะมีอำนาจการทดสอบสูงที่สุด แต่ถ้าค่าเฉลี่ยของตัวอย่างมีขนาดใหญ่ สถิติทดสอบ Q_{WLS} จะมีอำนาจทดสอบสูงที่สุด ในกรณีที่ศูนย์รักษามีจำนวนน้อย และ/หรือ ค่าเฉลี่ยของตัวอย่างในแต่ละกลุ่มทดลองมีจำนวนน้อย สถิติทดสอบ Q_{WLS} จะทำได้ดีในเรื่องของการควบคุมความคลาดเคลื่อนชนิดที่ 1 และมีอำนาจการทดสอบที่สูง

180 หน้า. ISBN 974-04-6318-5

CONTENTS

CONTENTS (CONTINUE)

LIST OF TABLES

LIST OF TABLES (CONTINUE)

LIST OF TABLES (CONTINUE)

LIST OF FIGURES

CHAPTER I
INTRODUCTION

1.1 Research Background and Problems

Clinical trials typically involve a comparison of a standard treatment versus a new treatment. Often, the outcome for a subject is a binary outcome such as success or failure, death or life. Patients enrolled in the trial are randomly assigned to one of these two treatment groups, and the success rates on the treatments are compared to determine which treatment is more efficacious.

One of interest in the clinical trial is in estimating the difference in the success rates of the two treatments, sometimes called **the risk difference** or **treatment difference**. The risk difference is certainly one of the most important epidemiologic indices to measure actual gains that can be expected in terms of percentage of patients treated. The risk difference has the advantage of ease of interpretation because it is the natural measure of the gain of one treatment over another. In other words, the risk difference is often used as the measure of effect in practice, it has a nice interpretation in a clinical trial, and it provides the actual gain that expected in terms of percentage of patients treated.

It is common in epidemiology to encounter a problem that is “sparse” in that a large number of events in sample sizes are small or zero. To illustrate, the following example of a multicenter study with high sparsity in the data; the sparse data table arises in practice since each center may not recruit many patients. For example, consider the data in Table1, which is taken from an 8 centers randomized trial investigating the efficacy of two topical cream preparations (active drug, control) in curing an infection addressed by Bilter and Landis (1985). A total of 273 eligible patients were recruited in 8 centers; an average of 34.1 patients per centers. Thus, this example has eight 2x2 tables and, on average, each table has 34.1 patient.

Suppose K independent centers in which two treatments are compared and the outcome measures are binary, the *risk difference* is defined by $\tau_i = \pi_{i1} - \pi_{i2}$ for center i = 1, 2, …, K where π_{i1} be the probability for positive response in treatment 1, π_{i2} be the probability for positive response in treatment 2. The maximum likelihood estimate of τ_i is provided by

$$\hat{\tau}_i = \hat{\pi}_{i1} - \hat{\pi}_{i2} = \frac{X_{i1}}{n_{i1}} - \frac{X_{i2}}{n_{i2}}$$

the X_{ij} following the binomial distribution, X_{i1} are the number of positive responses out of the sample size n_{i1} for treatment 1(treatment group) and X_{i2} are the number of positive responses out of the sample size n_{i2} for treatment 2 (control group).

Table 1 Available Data of Response to Active Drug and Control Treatment.

Center	Treatment Group			Control Group			
	X_{i1}	n_{i1}	$\hat{\pi}_{i1}$	X_{i2}	n_{i2}	$\hat{\pi}_{i2}$	$\hat{\tau}_i = \hat{\pi}_{i1} - \hat{\pi}_{i2}$
1	11	36	0.306	10	37	0.207	0.036
2	16	20	0.800	22	32	0.688	0.112
3	14	19	0.737	7	19	.0368	0.369
4	2	16	0.125	1	17	0.059	0.066
5	6	17	0.353	0	12	0.000	0.353
6	1	11	0.091	0	10	0.000	0.091
7	1	5	0.200	1	9	0.111	0.089
8	4	6	0.667	6	7	0.857	-0.190

Table 1 exhibits a potential difficulty that commonly occurs with multicenter clinical trials: the sample sizes for the treatments in many of centers are medium, and the corresponding cell counts are relatively small or zero. The success probabilities are slight and consistent but the risk differences vary considerably across the centers.

The second example is a response to therapy in a trial for 22 centers given by Mehta et al. The data given in Table 2 is adapted from Table 4 of Mehta et al. Since no details of the trial were revealed. The analysis in Mehta et al. was based on computation of Zelen's exact test for constancy of the odds ratio (Mehta, 1988).

Table 2 Response to Therapy in a Trial at 22 Centers.

Center	Control Drug			New Drug			$\hat{\tau}_i = \hat{\pi}_{i1} - \hat{\pi}_{i2}$
	X_{i1}	n_{i1}	$\hat{\pi}_{i1}$	X_{i2}	n_{i2}	$\hat{\pi}_{i2}$	
1	0	15	0.00	0	15	0.00	0.00
2	6	38	0.16	0	39	0.00	0.16
3	3	21	0.14	1	21	0.05	0.09
4	2	17	0.12	1	15	0.07	0.05
5	2	21	0.09	1	21	0.05	0.04
6	2	12	0.17	0	12	0.00	0.17
7	10	48	0.21	3	49	0.06	0.15
8	2	19	0.11	0	19	0.00	0.11
9	0	15	0.00	1	15	0.07	-0.07
10	2	28	0.07	2	28	0.07	0.00
11	2	20	0.10	0	19	0.00	0.10
12	1	12	0.08	0	12	0.00	0.08
13	5	24	0.21	0	24	0.00	0.21
14	2	13	0.15	2	12	0.17	-0.02
15	11	14	0.79	0	14	0.00	0.79
16	4	50	0.08	0	48	0.00	0.08
17	0	20	0.00	0	20	0.00	0.00
18	0	21	0.00	0	21	0.00	0.00
19	1	49	0.02	1	49	0.02	0.00
20	1	14	0.07	0	13	0.00	0.07
21	1	14	0.07	0	13	0.00	0.07
22	0	21	0.00	0	21	0.00	0.00

This situation has received considerable interest. Therefore, this research will demonstrate simulation study of comparison of common risk difference tests where the sparse data occurrence is taken into account. The null hypothesis is denoted by H_o: $\tau_1 = \tau_2 = \cdots = \tau_K = \tau$; it means that there is homogeneity of the risk difference across studies. However, this study is desirable to examine whether the assumption of the tests for homogeneity of the risk difference over centers holds. Only following five statistical tests are considered in this study.

There are some statistical tests, which proposed from time to time, and can provide markedly different tests of the true common risk difference and lead to potentially different conclusion. Lipsitz et al. and team (1998) provided a systematic and stimulating discussion on testing the homogeneity of the risk difference for a series of 2x2 tables. Besides considering the commonly used weighted least squares statistic suggested elsewhere, they considered the test procedure using transformation of the weighted least square statistic *Q*. Their several other weighted tests are Z^2_{WLS} , $Z^2_{WLS,R}$, Z^2_V , and Z^2_K . They further used Monte Carlo simulation to demonstrate that their newly proposed statistics could outperform the weighted least squares test statistic.

A weighted least square (*WLS*) statistic is commonly used to test for equality or homogeneity of the risk difference across centers. The *WLS* statistic can be considered the classical test statistic that is used for testing for homogeneity. The *WLS* statistic compares the observed risk differences with an estimate of the common risk difference. This study will simulate a comparison of the different test statistics for homogeneity followed work of Lipsitz et al.

Lipsitz et al. (1998) suggested that the data from center *i* be dropped whenever $\hat{\omega}_i \left[= \hat{\pi}_{i1}\left(1-\hat{\pi}_{i1}\right)/(n_{i1}-1) + \hat{\pi}_{i2}\left(1-\hat{\pi}_{i2}\right)/(n_{i2}-1)\right]$ equals zero. Because, it has some formula that $\hat{\omega}_i$ is the divisor, such as, the Q_{WLS} statistics test $\left(= \sum_{i=1}^{K} (Y_i - \hat{\tau})^2 / \hat{\omega}_i\right)$. Practically, the divisor could not be zero. Actually, dropping centers is not a good solution because if the number of centers are small then the simulation must delete the center where $\hat{\omega}_i = 0$ so the number of centers has a very small size, example, from Table 2 has 22 center in the trial and it has 4 centers (center 1, 17, 18 and 22) were dropped case as $\hat{\omega}_i = 0$, so to remains 18 centers in the trial. To avoid this problem, this study may apply the commonly used method of adding a constant to each cell of 2x2 tables whenever $X_{ij} = 0$ or $X_{ij} = n_{ij}$ take effect $\hat{\omega}_i = 0$ by adding the X_{ij} with 0.5 and n_{ij} with 1.

From the literature review, there are a few studies that focused on comparison of this common risk difference tests. Previous studies did not attempt to determine which testing method gives higher power for different cases of dropping centers and adding a constant to each cell.

Liang and Self (1985) have investigated tests for homogeneity of the odds ratio across centers when data are sparse. Although Liang and Self's test are very useful, homogeneity on the odds ratio scale does not imply homogeneity on the risk difference scale. Thus, if homogeneity of the risk difference is of interest, one of the test statistic discussed in this study should be used.

This study will show that can improve their test procedures by using both a one-sided test and a two-sided test, because Lipsitz et al. did not consider the effect of treatments on one-sided test. Finally, this study will include general guidelines about which test statistic is preferable for a variety of situations of the zero estimated variance. This study compare the performance of tests both of dropping center situation and adding a constant to each cell of 2x2 tables.

This study is designed to compare the empirical performance of the five common risk difference tests under various situations**.** This study also reviews and describes the tests that have been publicized in the statistical and epidemiological literature. Then, this study will evaluate other test statistics for homogeneity that may be more appropriate than the *WLS* statistic when the intra center sample sizes are small. Finally, this study will compare the Type I error and statistical power of several tests of homogeneity that can be used in meta-analysis.

Understanding the performance of these tests and the result of this study will guide investigators in selecting the most appropriate test for overall treatment effect estimation in different situations.

1.2 Objective

To compare the performance in terms of the size of the Type I error and the power of tests of the classical weighted least square test (Q_{WLS}) and their several other weighted tests namely $Z^2_{WLS}, Z^2_{WLS,R}, Z^2_V$ and Z^2_K for homogeneity of the risk difference.

1.3 Hypothesis

It is hypothesized that:

1. Q_{WLS} is appropriate when the mean in each treatment group size (n_{ij}) is large.

2. Z^2_{WLS} is appropriate when both the number of centers (K) and mean treatment group size (n_{ij}) are large.

3. $Z^2_{WLS,R}$ is appropriate when both the number of center (K) and mean treatment group size (n_{ij}) are large.

4. Z^2_V is appropriate when both the number of centers (K) and mean treatment group size (n_{ij}) are large.

5. Z^2_K is appropriate when the number of centers (K) is large, regardless of the mean treatment group size (n_{ij}).

6. When comparing all test statistics with respect to Type I error rate and the power of the test statistics, $Z^2_{WLS,R}$ and Z^2_V may be the most appropriate test statistics to use for a test for clinical trials.

1.4 Scope of study

1. This study will compares results between the simulation by deleting case where $\hat{\omega}_i = 0$ (following the work of Lipsitz et al.) and adding a constant.

2. The study mean sample sizes are 4, 8, 16, 32, and 64 where the mean sample sizes in each treatment group are equal and unequal sizes. The number of centers varies between 8, 16, 32, and 48.

3. Different levels of significance (α) are 0.01, 0.05, and 0.10 when considering one- sided and two-sided tests.

4. Each condition is tested for 10,000 iterations.

1.5 Definition of terms

The following terms have been defined to clarify their meaning within the context of this research

Clinical Trial: It means a research study in human volunteers to answer specific health questions about vaccines or new therapies or new ways of using known treatment. Clinical trials (also called medical research and research studies) are used to determine whether new drugs or treatments are both safe and effective. Carefully conducted clinical trials are the fastest and safest way to find treatments that work in people and ways to improve health. Interventional trials determine whether experimental treatments or new ways of using known therapies are safe and effective under controlled environment. Observational trials address health issues in large groups of people or populations in a natural setting.

Trials are in four phase: Phase I tests a new drug or treatment in a small group; Phase II expands the study to a larger group of people; Phase III expands the study to an even larger group of people; and Phase IV takes place after the drug or treatment has been licensed and marketed.

Ideas for clinical trials usually come from researchers. Once researchers test new therapies or procedure in the laboratory and get promising results, they begin planning Phase I clinical trials. New therapies are tested on people only after laboratory and animal studies show promising results (University of Miami School of Medicine, 2002).

Multi-center Studies: A multi-center study is defined to be a study involving two or more fields sites may consist of hospitals, clinics, or other locations that the clinical trial takes place.

Risk Difference (RD) or Treatment Difference: (synonym: absolute risk reduction) It means the absolute difference in the event rate between two comparison groups. A risk difference of zero indicates no difference between comparison groups. For undesirable outcomes a RD that is less than zero indicates that the intervention was effective in reducing the risk of that outcome (University of Miami School of Medicine, 2003)**.**

Fixed effect model: It is a statistical model that stipulates that the units under analysis (e.g. people in a trial or study in a meta-analysis) are the ones of interest, and thus constitute the entire population of units. Only within-study variation is taken to influence the uncertainty of results (as reflected in the confidence interval) of a meta-analysis using a fixed effect model. Variation between the estimates of effect from each study (heterogeneity) does not effect the confidence interval in a fixed effect model (University of Miami School of Medicine, 2003).

Random effects model: It is a statistical model sometimes used in meta-analysis in which both within-study sampling error (variance) and between-studies variation are included in the assessment of the uncertainty (confidence interval) of the results of a meta-analysis. If there is significant heterogeneity among the results of the included studies, random effects models will give wider confidence intervals than fixed effect models (University of Miami School of Medicine, 2003).

Homogeneity: In systematic reviews homogeneity refers to the degree to which the results of studies included in a review are similar. "Clinical homogeneity" means that, in trials included in a review, the participants, interventions and outcome measures are similar or comparable. Studies are considered "statistically homogeneous" if their results vary no more than might be expected by the play of chance. See heterogeneity (University of Miami School of Medicine, 2003).

Heterogeneity: In systematic reviews heterogeneity refers to variability or differences between centers in the estimates of effects. A distinction is sometimes made between "statistical heterogeneity" (differences in the reported effects), "methodological heterogeneity" (differences in study design) and "clinical heterogeneity" (differences between centers in key characteristics of the participants, interventions or outcome measures). Statistical tests of heterogeneity are used to assess whether the observed variability in center results (effect sizes) is greater than that expected to occur by chance. However, these tests have low statistical power (University of Miami School of Medicine, 2003).

Null hypothesis: It is a statistical hypothesis that one variable (e.g. whether or not a study participant was allocated to receive an intervention) has no association with another variable or set of variables (e.g. whether or not a study participant died), or that two or more population distributions do not differ from one another. In simplest terms, the null hypothesis states that the results observed in a study are no different from what might have occurred as a result of the play of chance (University of Miami School of Medicine, 2003).

Type I error: It is a statistical error (said to be "of the first kind" or alpha error) made in testing an hypothesis when it is concluded that a treatment or intervention is effective when it really is not. Sometimes referred to as a false positive (The Animated Software Company, 2002). That is, in a given statistical tests, the probability of a type

I error is equal to the value you have set for alpha. Alpha is the probability of rejecting the hypothesis tested when that hypothesis is true (Decision).

On the other hand, in a given statistical tests, the probability of a type II error is equal to the value calculated for beta. Beta is the probability of accepting the hypothesis tested when the alternative hypothesis is true.

Power: It is the probability that the null hypothesis will be rejected if it is indeed false. In studies of the effectiveness of healthcare interventions, power is a measure of the certainty of avoiding a false negative conclusion that an intervention is not effective when in truth it is effective. The power of a study is determined by how large it is (the number of participants), the number of events (e.g. strokes) or the degree of variation in a continuous outcome (such as weight), how small an effect one believes is important (i.e. the smallest difference in outcomes between the intervention and the control groups that is considered to be important), and how certain one wants to be of avoiding a false positive conclusion (i.e. the cut-off that is used for statistical significance) (University of Miami School of Medicine,2002).

Simulation: It is a quantitative technique that utilizes a computerized mathematical model in order to represent actual making under conditions of uncertainty for evaluating alternative courses of action based upon facts and assumptions (Thieraut & Klekamp, 1975).

In general, simulation is modeling of a process or phenomenon. In statistics, Monte Carlo simulation is often used to model outcomes of a random experiment. This kind of simulation rests on generation of pseudo-random numbers. That is numbers which behave like truly random numbers, though generated by a deterministic (non-random) algorithm.

FORTRAN: It is a programming language for computer. Fortran, which stands for FORmula TRANslation, was the first high level programming language. It made it possible to use symbolic names to represent mathematical quantities, and to write mathematical formulate in a reasonably comprehensible form. The idea of Fortran was proposed in late 1953 by John Backus, in New York, and the first Fortran program was run in April 1957 (Brain, 1994).

1.6 Benefit of the study

For understanding performance of these tests and the results of this study will guide the investigators in selecting the appropriate tests which test statistic should be used in a variety of situation.

CHAPTER II
LITERATURE REVIEW

This chapter will describe the classical weighted least square test and their several other weighted tests based on theories and relevant research literature.

2.1 Research and Test Statistics

Consider K independent institutions participate in the clinical trial are compared and the outcome measures are binary. In institution i (i = 1,…,K), there are X_{i1} successes out of n_{i1} patients on treatment 1; similarly, there are X_{i2} successes out of n_{i2} patients on treatment 2. This study assume, for the jth (j = 1,2) treatment in the ith (i = 1,…,K) center, the data X_{ij} follow a binomial distribution with sample size n_{ij} and sample estimate of this probability π_{ij}, with X_{i1} and X_{i2} independent. Then the risk difference for institution i is

$$\tau_i = \pi_{i1} - \pi_{i2}$$

The null hypothesis of equal risk difference over institutions is

$$H_o : \tau_1 = \tau_2 = \ldots = \tau_k = \tau.$$

And against the alternative that there is a difference between at least two strata

$$H_1 : \tau_i \neq \tau_j \text{ some } 1 \leq i \leq j \leq k$$

The study is interested here in estimating τ efficiently. Estimating τ makes the most sense if there is homogeneity of the risk difference across study centers.

The maximum likelihood estimate of τ_i is the difference in the proportion of patients with a successful response to the treatments,

$$Y_i = \hat{\pi}_{i1} - \hat{\pi}_{i2}$$

where $$\hat{\pi}_{i1} = \frac{X_{i1}}{n_{i1}}, \hat{\pi}_{i2} = \frac{X_{i2}}{n_{i2}}$$

Then Y_i is an unbiased estimate of $\tau_{i,}$ i.e., $E(Y_i) = \tau_i$ and, since X_{i1} and X_{i2} are independent,

$$\omega_i = \text{var}\left(Y_i\right) = \text{var}\left(\frac{X_{i1}}{n_{i1}} - \frac{X_{i2}}{n_{i2}}\right) = \frac{\pi_{i1}\left(1-\pi_{i1}\right)}{n_{i1}} + \frac{\pi_{i2}\left(1-\pi_{i2}\right)}{n_{i2}}$$

An unbiased estimate of ω_i is

$$\hat{\omega}_i = \frac{\hat{\pi}_{i1}\left(1-\hat{\pi}_{i1}\right)}{n_{i1}-1} + \frac{\hat{\pi}_{i2}\left(1-\hat{\pi}_{i2}\right)}{n_{i2}-1}$$

Note that $\hat{\omega}_i$ is also an unbiased estimate of ω_i under the null.

In this setting, each center contributes a 2x2 table of treatment by binary response. However, the data in each table may be sparse since each center may not accrue many patients. For example, in this study return to the data considered previously by Lipsitz et al.(1998). The data in Table 3 are from the Cancer and Leukemia Group B (CALGB) randomized trial comparing two chemotherapy treatments with respect to survival (lived / died by the end of the study) in patients with multiple myeloma (Cooper et al.1993). A total of 156 eligible patients were accrued in the 21 centers. The sample size is very small in the two treatment arms, which means; with 156 patients, they have an average of 7.4 patients per center. Thus, they have twenty-one 2x2 tables and, on average, each table has only 7.4 patients.

Table 3 Center Success Rates from **CALBG** Study.

Centers	n_{i1}	$\hat{\pi}_{i1}$	n_{i2}	$\hat{\pi}_{i2}$	$\hat{\pi}_{i1}-\hat{\pi}_{i2}$	$\hat{\omega}_i=\frac{\hat{\pi}_{i1}(1-\hat{\pi}_{i1})}{n_{i1}-1}+\frac{\hat{\pi}_{i2}(1-\hat{\pi}_{i2})}{n_{i2}-1}$
1	4	0.75	3	0.33	0.42	0.17
2	4	0.75	11	0.73	0.02	0.26
3	2	1.00	3	0.67	0.33	0.22
4	2	1.00	2	1.00	0.00	0.00
5	2	1.00	3	0.00	1.00	0.00
6	3	0.33	3	0.67	-0.34	0.33
7	2	1.00	3	0.67	0.33	0.22
8	5	0.20	4	1.00	-0.80	0.04
9	2	1.00	3	0.67	0.33	0.22
10	2	0.00	3	0.67	-0.67	0.22
11	3	1.00	3	1.00	0.00	0.00
12	2	1.00	2	0.00	1.00	0.00
13	4	0.25	5	0.20	0.05	0.22
14	3	0.67	4	0.50	0.17	0.36
15	4	0.50	6	0.67	-0.17	0.30
16	12	0.33	9	0.33	0.00	0.24
17	2	0.50	3	0.67	-0.17	0.47
18	3	1.00	4	0.25	0.75	0.19
19	4	0.25	3	0.67	-0.42	0.28
20	3	0.00	2	0.00	0.00	0.00
21	4	0.50	5	0.20	0.30	0.24

2.2 Related Studies

Lipsitz et al. (1998) provided a systematic and stimulating discussion on testing the homogeneity of the risk difference for a series of 2x2 tables. Besides considering the commonly used weighted least square statistic suggested elsewhere, they proposed and evaluates several other weighted test statistics. They further used Monte Carlo simulation to demonstrate that their newly proposed statistics could out perform the weighted least squares test statistic. Lipsitz et al. (1998) discussed testing the homogeneity of the risk difference for a series of 2x2 tables, and they proposed and evaluated several weighted least squares test statistics. On the basis of this study about simulation, as long as the mean treatment group size per center is moderate or large ($n \geq 16$), this simple test statistic along with the use of the commonly used adjustment procedure for sparse data performs well when the number of center is small or moderate ($K \leq 32$). In fact, this simulation also shows that this test statistic outperforms all the statistics considered by Lipsitz et al. (1998).

Lui and Kelly (2000) considered testing the homogeneity of risk ratio over a series of 2x2 tables. In addition to the classical weighted least squares (*CWLS*) test procedure, they considered two test procedure using simple transformation of the CWLS statistic and developed three other asymptotically weighted test procedures.

Takkouche (1999) evaluated the old and new tests of heterogeneity in epidemiologic meta-analysis by used a large simulation study patterned from the key features of five (Dersimonian and Laird's test, Z^2_{WLS}, $Z^2_{WLS,R}$, Z^2_K and the likelihood ratio test) published epidemiologic meta-analysis to investigate the type I error and statistical power of five previously proposed asymptotic homogeneity tests, a parametric bootstrap version of each of the tests and τ^2-bootstrap, a tests proposed by the authors.

Liang and Self (1985) considered three statistics for testing the homogeneity of odds ratio for a series of 2x2 tables when the data are sparse. They suggested that a score test, based on the assumption that the log odds ratio are generated from some unknown distribution, is shown to be more powerful than the other tests.

Böhning and Sarol (2000) considered the case of efficient estimation of the risk difference in a multicenter study allowing for baseline heterogeneity. He considered the optimally weighted estimator for the common risk difference and show that this estimator has considerable bias when the true weights. In addition, he proposed a new estimator for this situation of the Mantel-Haenzel type that is unbiased and, in addition, has a smaller variance for small sample sizes within the study centers.

Viwatwongkasem and Böhning (2002) was interested in combining risk difference of several centers under homogeneity of equal risk difference across centers and compared the efficiency of six estimators for the common risk difference. The six estimators consist of the pooling method ignoring the stratification of centers, several popular sets of different weights, and a new estimator based on the idea of two-stage random effect models. The new inverse – variance weight consists of the variation within a center and the variation between centers (heterogeneity variance across centers). A simulation study was done to compare bias, variance, and mean square error. The major result can summarize that the new estimator is an attractive compromise between the estimators of the set of the center – specific sample size weights and the estimators of the set of the inverse – variance weights. It is not an optimal strategy, but it widely extends to cover heterogeneity cases.

2.3 Test Statistic for the Equality of the Risk Difference

Under homogeneity of $\tau_1 = \tau_2 = ... = \tau_k = \tau$, the parameter of interest is the common risk difference that is assumed to be a constant under homogeneity of risk difference across centers. In case there is effect heterogeneity, τ is defined as the true average value of overall treatment effect.

- **Weighted Least Squares Test Statistic (*WLS*)**

Under the null hypothesis $H_o : \tau_1 = \tau_2 = \ldots = \tau_k = \tau$, $E(Y_i) = \tau$. When all n_{ij} are large. $\hat{\omega}_i$ is approximately nonrandom and the variance of Y_i is known , i.e., $var(Y_i) \approx \hat{\omega}_i$, Fleiss (1981) suggested using the weighted least squares test statistic,

$$Q_{WLS} = \sum_{i=1}^{K} \frac{\left(Y_i - \hat{\tau}\right)^2}{\hat{\omega}_i} \qquad ①$$

Note that this test statistic compares the estimated risk difference Y_i is institution i to the pooled estimate $\hat{\tau}$ under the null. When $\hat{\omega}_i = 0$ in Q_{WLS}, Lipsitz et al. (1998) suggested deleting the data in center *i*. On the other hand, when each institution has a large sample size (so that $\hat{\omega}_i$ is approximately nonrandom), regardless of the size of *K*, under the null hypothesis of no institution effects, ① is approximately Chi-square with (*K*-1) degrees of freedom. However, when each institution has a relatively small sample size so that $\hat{\omega}_i$ is highly variable. Then this Chi-square approximation for the distribution of Q_{WLS} may be poor.

- **Lipsitz et al.'s Tests**

Lipsitz et al. (1998) proposed and evaluated several other weighted tests for homogeneity of the risk difference when data are sparse. He considered the test procedure using transformation of the WLS statistics *Q*. Their several other weighted tests are Z^2_{WLS} , $Z^2_{WLS,R}$, Z^2_V and Z^2_K .

Z^2_{WLS}

Lipsitz et al. (1998) considered the test procedure using transformation of the WLS statistic *Q* (Dersimonian and Laird's Q statistic) (1986). As an alternative to this Chi-square approximation, when K is large and all n_{ij} is large. The test statistic is

$$Z_{WLS} = \frac{\left[Q_{WLS} - (K-1)\right]}{\sqrt{2(K-1)}} \quad \text{or} \quad Z^2_{WLS} = \frac{\left[Q_{WLS} - (K-1)\right]^2}{2(K-1)} \qquad ②$$

The test is approximately Chi-square with one degree of freedom under homogeneity, when both $n_{ij} \to \infty$ and $K \to \infty$. If the null hypothesis is not true, we will expect a large value of *Q*. This suggest that we reject H_o at α – level when Z_{WLS} > Z_∞ where Z_∞ is the upper 100(α^{th}) percentile of the standard normal distribution. Or we can say, when the assumption of homogeneous risk difference does not hold true, test statistic Q_{WLS} is expected to be larger than that under H_o. In the other hand, we should reject H_o only when Q_{WLS} is large. On the other hand, in application of test statistic Z^2_{WLS} a small value of Q_{WLS} can lead to a value of Z^2_{WLS} that is large enough to reject H_o. McCullagh and Nelder (1989) discussed that the F-distribution with 1 and (*K*-1) degrees of freedom is a better approximation in finite samples than the Chi-square distribution with 1 degree of freedom. Here, we reject the null hypothesis H_o at α-level when Z^2_{WLS} > $F_{\infty(1,K-1)}$ where $F_{(1,K-1)}$ is the upper 100(α^{th}) percentile of the F-distribution with 1 and *K* degrees of freedom.

$Z^2_{WLS,R}$

The $Z^2_{WLS,R}$ test was also derived from Dersimonian and Laird's *Q* statistic by Lipsitz et al. (1998). The test statistic is

$$Z_{WLS,R}^{2}=\frac{\left[Q_{WLS}-K\right]^{2}}{\sum_{i=1}^{K}\left[\frac{\left(Y_{i}-\hat{\tau}\right)^{2}}{\hat{\omega}_{i}}-1\right]^{2}} \quad ③$$

where the summation is over those centers in which the resulting $\hat{\omega}_i \neq 0$. Here, the subscript *WLS, R* is a mnemonic for *WLS* with a robust variance estimate. The $Z_{WLS,R}^{2}$ is approximately Chi-square with one degree of freedom under the null of homogeneity when both K and n_{ij} are large. Thus ③ is based on the distribution of Q_{WLS} when K and n_{ij} are large. Similarly, since the F-distribution with 1 and (*K*-1) degrees of freedom are a better approximation in finite samples than Chi-square with 1 degree of freedom. Thus, Lipsitz et al. suggested approximating the finite sample distribution of $Z_{WLS,R}^{2}$ with an F- distribution with 1 and (*K*-1) degrees of freedom. Here, we reject the null hypothesis H_o at α - level when $Z_{WLS,R}^{2} > F_{\infty(1,K-1)}$.

Z_V^2

Test statistic which the expectation of increasing the power of ③, Lipsitz et al. (1998) proposed the test statistic

$$Z_{V}^{2}=\frac{\left\{\sum_{i=1}^{K}\left[\left(Y_{i}-\hat{\tau}\right)^{2}-\hat{\omega}_{i}\right]/a_{i}\right\}^{2}}{\sum_{i=1}^{K}\left[\left(Y_{i}-\hat{\tau}\right)^{2}-\hat{\omega}_{i}\right]^{2}/a_{i}^{2}} \quad ④$$

where $\boldsymbol{a_i}$ is equal to $\operatorname{var}\left[\left(Y_i-\tau\right)^2\right]$. This study can show that

$$\operatorname{var}\left[\left(Y_{i}-\tau\right)^{2}\right]=\operatorname{var}\left\{\left[\left(\frac{X_{i1}}{n_{i1}}-\pi_{i1}\right)-\left(\frac{X_{i2}}{n_{i2}}-\pi_{i2}\right)\right]^{2}\right\}$$

$$=\frac{\pi_{i1}\left(1-\pi_{i1}\right)\left[1+3\pi_{i1}\left(1-\pi_{i1}\right)\left(n_{i1}-2\right)\right]}{n_{i1}^{3}}+\frac{\pi_{i2}\left(1-\pi_{i2}\right)\left[1+3\pi_{i2}\left(1-\pi_{i2}\right)\left(n_{i2}-2\right)\right]}{n_{i2}^{3}}$$

$$+\frac{6\pi_{i1}\left(1-\pi_{i1}\right)\pi_{i2}\left(1-\pi_{i2}\right)}{n_{i1}n_{i2}}-\left[\frac{\pi_{i1}\left(1-\pi_{i1}\right)}{n_{i1}}+\frac{\pi_{i2}\left(1-\pi_{i2}\right)}{n_{i2}}\right]^{2}$$

Here, the subscript *V* is a mnemonic for variance – weighted. When applying ④, Lui and Kelly (2000) substituted $\hat{\pi}_{ij}$ for π_{ij}, also, as with Z_{WLS}^{2}, will approximate the sampling distribution of Z_V^2 by the F-distribution with 1 and (K-1) degree of freedom. Note that the above formula for $\mathrm{var}\left[\left(Y_i-\tau\right)^2\right]$ is different from formula (14) in Lipsitz et al. (1998, page 152). Because the expectation $E\left(X_{ij}-n_{ij}\pi_{ij}\right)^4=$ $3\left[n_{ij}\pi_{ij}\left(1-\pi_{ij}\right)\right]^2+n_{ij}\pi_{ij}\left(1-\pi_{ij}\right)\left[1-6\pi_{ij}\left(1-\pi_{ij}\right)\right]$, where X_{ij} follows the binomial distribution wit n_{ij} and π_{ij} (Johnson and Kotz, 1969), formula (14) by Lipsitz et al. is incorrect. Although the test statistic Z_V^2 considered by Lipsitz et al. using a weight not equal to $\mathrm{var}\left[\left(Y_i-\tau\right)^2\right]$ is still asymptotically valid, they used the correct formula for $\mathrm{var}\left[\left(Y_i-\tau\right)^2\right]$ to explore the performance of Z_V^2 in the following simulation.

Z_K^2

The Z_K^2 test avoid the sensitivity of the above test statistic to small n_{ij}, when K is large, and derived from Dersimonian and Laird's Q statistic by Lipsitz et al. (1). The test statistic is

$$Z_K^2=\frac{\left\{\sum_{i=1}^{K}\left[\left(Y_i-\hat{\tau}\right)^2-\hat{\omega}_i\right]/\left[n_{i1}^{-1}+n_{i2}^{2}\right]^2\right\}^2}{\sum_{i=1}^{K}\left[\left(Y_i-\hat{\tau}\right)^2-\hat{\omega}_i\right]^2/\left[n_{i1}^{-1}+n_{i2}^{2}\right]^4} \qquad ⑤$$

where
$$\hat{\tau}=\frac{\sum_{i=1}^{K}Y_i/\left[n_{i1}^{-1}+n_{i2}^{-1}\right]}{\sum_{i=1}^{K}1/\left[n_{i1}^{-1}+n_{i2}^{-1}\right]}$$

here, the subscript K means that the asymptotic distribution only depends on K. As noted by Lipsitz et al. (1998), the weight used in ⑤ is no longer a function of random variables. The Z_K^2 test has an approximate F-distribution with 1 and (K-1) degrees of freedom under the null, only assuming that $K\to\infty$. And it is not necessary that within study sample size be large. Here, we reject the null hypothesis H_o at α-level when $Z_K^2>F_{\infty(1,K-1)}$.

CHAPTER III
METHODOLOGY

This study is mainly concerned with the Monte Carlo simulation technique on microcomputer. The simulations are performed under both the null (no center effect) and the alternative. The objective of the simulation study is to compare the Type I error and the power of the test of the statistic Q_{WLS}, Z^2_{WLS}, $Z^2_{WLS,R}$, Z^2_V, and Z^2_K. The important issues in this chapter would cover the Monte Carlo method, the criteria for simulation and the simulation procedures.

3.1 The Monte Carlo method

The term "Monte Carlo" was introduced by Von Neumann and Ulam during World War II, as a code word for the secret work at Los Alamos; it was suggested by the gambling casinos at the city of Monte Carlo in Monaco. The Monte Carlo method was then applied to problems related to the atomic bomb. The work involved direct simulation of behavior concerned with random neutron diffusion in fissionable material.

Historically, the Monte Carlo method was considered to be a technique using random or pseudorandom numbers, for solutions of a model. Random numbers are essentially independent random variables uniformly distributed over the unit interval [0,1]. Actually, what are available at computer centers are arithmetic codes for generating sequences of pseudorandom digits, where each digit (0 through 9) occurs with approximately equal probability (likelihood). Consequently, the sequences can model successive flips of a fair ten-sided die. Such codes are called random number generators. Grouped together, these generated digits yield pseudorandom numbers with any required number of elements (Rubinstein, 1981).

3.2 The criteria for simulation

The conditions are determined for comparing the tests between five statistics tests. The criteria are set as follows:

1. This study performed the simulations using two methods, that is
   - Dropping center *i* when $\hat{\omega}_i$ equal zero
   - Adding constant when $\hat{\omega}_i = 0$
2. Numbers of centers (*K*) are 8, 16, 32, and 48.
3. Mean sample sizes in each treatment groups have equal size and unequal size where the mean sample sizes are 4, 8, 16, 32 and 64.
4. Level of significance (*α)* are 0.01, 0.05 and 0.10 where testing considers at one- sided and two-sided tests.
5. Each condition is tested for 10,000 iterations on a microcomputer compatible with FORTRAN 90 language.

3.3 The simulation procedures

To simulate the data according to the determined criteria, the Monte Carlo method will be used on a microcomputer with FORTRAN language. Its important steps will be conducted as follows:

3.3.1 Random number generation

The value of the number of sample sizes in each treatment group with positive response to treatment *j*, X_{ij}, and number of sample sizes in each treatment *j*, n_{ij}, will be generated in this step.

- **Generate the random numbers from Uniform Distribution.**

Actually, the first step for random number generation is to generate uniform random numbers. After that, the uniformly distributed random numbers will be transformed to be whatever distributed data are needed. Especially for this study, FORTRAN Powerstation includes subroutines that are easily used for various functions in the International Mathematical and Statistical Library (IMSL).

A subroutine for the generation of uniform distribution random numbers from the IMSL stat library is required. The subroutine is used by issuing the statement:

CALL RNUN (NR, R)

Arguments

NR – Number of random numbers to generate. (Input)

R **–** Vector of length NR containing the random uniform (0, 1) deviates. (Output)

Algorithm

Routine RNUN generates pseudorandom numbers from a uniform (0, 1) distribution. The values returned in R by RNUN are positive and less than 1.0. However, values in R may be smaller than the smallest relative spacing, however. Hence, it may be the case that some value *R (i)* is such that 1.0 **–** *R(i)* = 1.0.

This study generates the random number from a uniform [0.1, 0.8]. Therefore, deviates from the distribution with uniform density over the interval (A, B) can be obtained by scaling the output from RNUN. The following statements (in single precision) would yield random deviates from a uniform (A, B) distribution:

CALL RNUN (NR, R)

CALL SSCAL (NR, B-A, R, 1)

CALL SADD (NR, A, R, 1)

- **Generate the random numbers from Binomial Distribution.**

For generating binomial distributed random numbers of sample sizes with positive response to treatment j (X_{ij}), the following FORTRAN subroutine produces a number X_{ij} which is

CALL RNBIN (NR, N, P, IR)

Arguments

NR **–** Number of random numbers to generate. (*Input*)

N **–** Number of Bernoulli trials. (*Input*)

P **–** Probability of success on each trial. (*Input*)

P must be greater than 0.0 and less than 1.0.

IR – Vector of length NR containing the random binomial deviates. (*Output*)

Algorithm

Routine RNBIN generates pseudorandom numbers from a binomial distribution with parameters N and P. N and P must be positive, and P must be less than 1.

3.3.2 Calculation the type I error and power of test

There are two types of error occurring by statistical hypothesis testing, which are summarized in the following table:

Event	Decision	
	Reject H_o	Accept H_o
H_o is True	Type I Error (α) Significance Level	Correct Decision ($1-\alpha$) Confidence Coefficient
H_o is False	Correct Decision ($1-\beta$) Power of Test	Type II Error (β)

Neyman proposed guidelines for selecting the statistical test. The first step was to consider the capability to control the type I error of the test. Namely, the actual type I error would not exceed the nominal level of significance (the level of significance defined in the study). After that, the type II error was considered; the statistical test with the least probability of accepting a false null hypothesis (the least probability of type II error) was then selected. That led to the conclusion that the selected test statistic was the most powerful one, because consideration of the capability to control the type I error is necessary for comparing the power of the statistical test.

This study will considered the type I error and power of test in each method. That is, the first method is simulation following the Lipsitz et al (dropping cases where $\hat{\omega}_i = 0$) and the second method is simulation by adding a constant in each X_{ij} and n_{ij} when $\hat{\omega}_i = 0$. The calculation of the statistic tests will be compared with the critical value for decision making and acceptance or rejection of a null hypothesis. A number of null hypothesis rejections will be counted for calculating the actual type I error and power of test.

$$\text{The actual type I error} = \frac{\#\ H_o \text{ rejection of the null hypothesis which is true}}{10{,}000}$$

$$\text{The Power of test} = \frac{\#\ H_o \text{ rejection of the alternative hypothesis which is true}}{10{,}000}$$

The proportion between the number of null hypothesis rejections is true and the number of replications (=10,000) is the actual type I error, whereas the power of test is the proportion between the number of null hypothesis rejections and the number of replications (=10,000).

- **Simulation for studying the type I error**

The simulation consists of the following steps:

1. The baseline-heterogeneity risks π_{i2} are generated from uniform distribution on $[0.1, 0.8]$.
2. Set $\pi_{i1} = \pi_{i2} + \tau = \pi_{i2} + 0.1$ (this is equivalent to assuming that the underlying common risk difference τ is 0.1 of centers overall).
3. Let n_{ij} be a random variable with a limited range. For a given simulation and treatment group in a center, assume that the sample size n_{ij} varies according to the probability function $f(n_{ij})$ = .20 for $n_{ij} = n\text{-}s$, where n is a given fixed constant and s = 2, 1, 0, -1, and –2;

$$f(n_{ij}) = \begin{cases} n-2 & \text{with probability } 0.2 \\ n-1 & \text{with probability } 0.2 \\ n & \text{with probability } 0.2 \\ n+1 & \text{with probability } 0.2 \\ n+2 & \text{with probability } 0.2 \end{cases}$$

   for a given value of n. Thus, for a given simulation, when mean sample sizes are equal then $E(n_{i1}) = E(n_{i2}) = n$, and when mean sample sizes are unequal then $E(n_{i1}) = n_1$ and $E(n_{i2}) = n_2$. This study performs different simulations for n = 4, 8, 16, 32, or 64. Note that, the test statistics developed in this study are fixed n_{ij} and π_{ij}; however, in practice, this study might expect n_{ij} to be random (over a limited range of values) and π_{ij} to vary randomly across centers.

4. Binomial variates X_{ij} with parameters n_{i1}, π_{i2} and binomial variates X_{i2} with parameters n_{i2}, π_{i1} are generated for each center i.
5. Compute all statistics tests in which K = 8, 16, 32, or 48, perform 10,000 repeat samples and calculate the percentage of these samples which rejects H_o for each statistic under consideration.

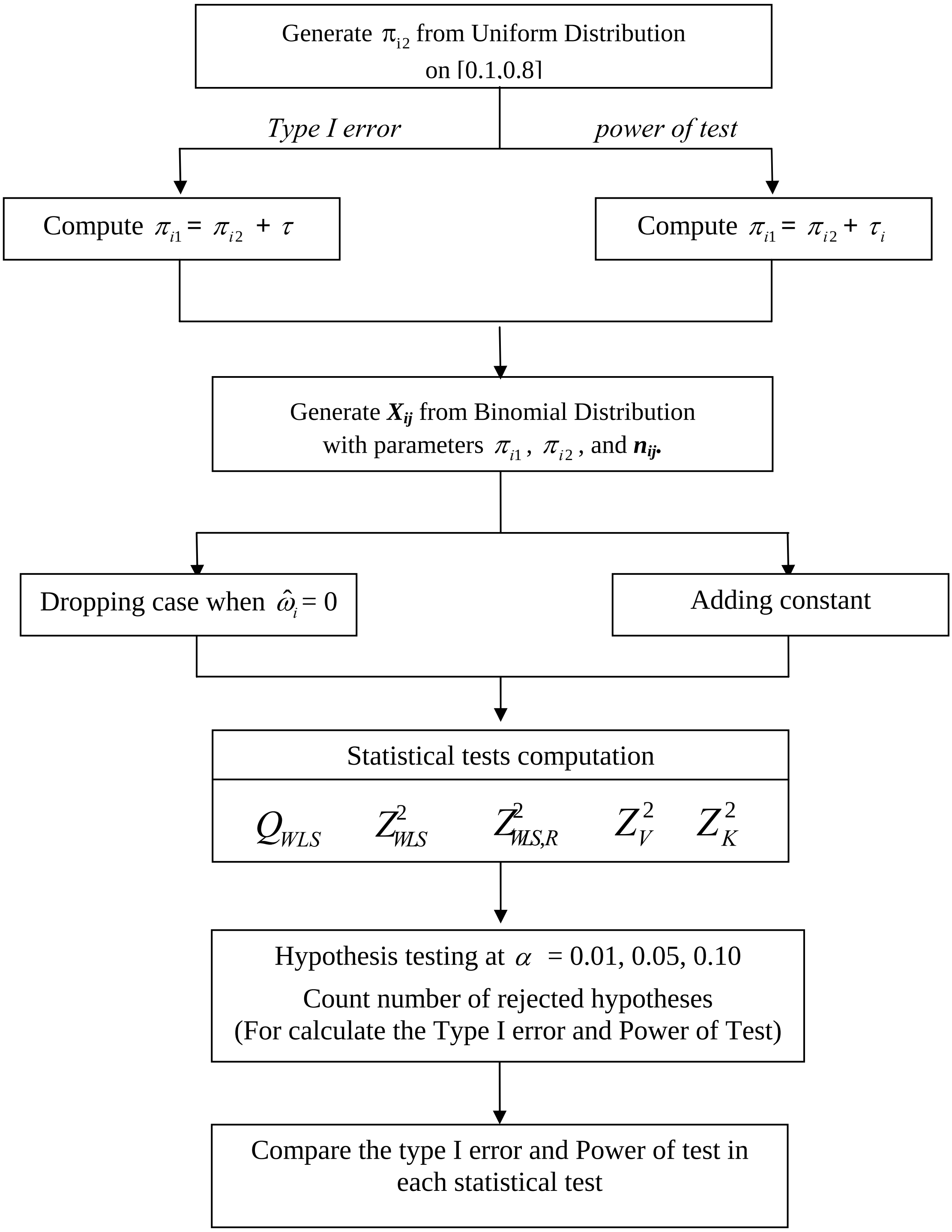
- **Simulation for studying the power of tests**

To compare the power of the test statistics, all statistics should be calibrated to have the same Type I error rate under the null hypothesis. The research will compared the power of the test statistics for a random effects alternative model. This study can classify the simulation procedures in the following cases:

Uniform case:

1. The baseline-heterogeneity risks π_{i2} are generated from uniform distribution on [0.1, 0.8] .
2. Set $\pi_{i1} = \pi_{i2} + \tau_i$. Let $\tau_i = 0.1 + c\left(2U - 1\right)$ where U is a uniform [0,1] random variable. Using results from the uniform distribution, $E(\tau_i) = \tau = .1$ and $\sigma_\tau^2 = \text{var}(\tau_i) = (2c)^2/12$. For c = 0, τ_i = .1 for all i, the null of homogeneity holds and where c = .02, .04, .06, .08.
3. Similarly, other steps compare the Type I error. That is, let n_{ij} be a random variable with a limited range. For a given simulation and treatment group in a center, assume that the sample size n_{ij} varies according to the probability function $f(n_{ij})$ = .20 for $n_{ij} = n-s$, where n is a given fixed constant and s = 2, 1, 0, -1, and –2.
4. Generate X_{i1} with parameters n_{i1}, π_{i1} and X_{i2} with parameters n_{i2}, π_{i2} for each center i.
5. Finally, calculate the tests statistics from the simulation data are the performance of times in the 10,000 simulations, to estimate the power of the test statistics at the given value of c, which rejects the null hypothesis to determine the power of tests.

The simulation procedures are summarized as follows:



CHAPTER IV
RESULTS

The main issues in this chapter are the results of comparing the type I error and power of tests between five statistics tests of this study; $Q_{WLS}, Z^2_{WLS}, Z^2_{WLS,R}, Z^2_V$ and Z^2_K tests for different situations. The results of this study will be presented in two parts: the first part is comparing the type I error and the second part is comparing power of the test. Both of two parts will concentrate simulation data from dropping case when $\hat{\omega}_i = 0$ and adding constant when $X_{ij} = 0$ or $X_{ij} = n_{ij}$ take effect $\hat{\omega}_i = 0$ by adding the X_{ij} with 0.5 and n_{ij} with 1. In each method, this study will consider for one-sided, two-sided of significance level which mean sample size in each treatment group are equal and unequal case.

4.1 Comparing the Type I Error

Type I error comparison between the five statistical tests is considered by comparing the actual type I error (α*) with the nominal level of significance (α). That is the departure of the actual type I error from the nominal level of significance must not exceed the precise limit, which is based on Chi-square test in the Q_{WLS} test, and based on F test in $Z^2_{WLS}, Z^2_{WLS,R}, Z^2_V$ and Z^2_K. In this study, the robustness evaluation is based on Cochran limits as follow:

two-sided: at 0.01 significance level, α* value is between [0.005, 0.015]
at 0.05 significance level, α* value is between [0.040, 0.060]
at 0.10 significance level, α* value is between [0.081, 0.119]
one-sided: at 0.01 significance level, α*/2 value is between [0.0025, 0.0075]
at 0.05 significance level, α*/2 value is between [0.0175, 0.0325]
at 0.10 significance level, α*/2 value is between [0.0425, 0.0575]

α* is the probability of a Type I error or the empirical alpha.

α is the nominal level of significance or the theoretical alpha.

The statistical test can control Type I error when its actual Type I error lies within those precise limits. If the actual Type I error lies over the limits, it will indicate that the statistic tests cannot control Type I error. Consequently, it can be divided into two cases as follow:

1. The probability of Type I error of the test statistic is greater than the nominal level of significance when the actual Type I error is greater than the upper bound of those limits.

2. The probability of Type I error of the test is less than the nominal level of significance when the actual Type I error is less than the lower bound of those limits.

The results of actual Type I error of the five statistical tests ($Q_{WLS}, Z^2_{WLS}, Z^2_{WLS,R}, Z^2_V, Z^2_K$) distributed data at 0.01, 0.05 and 0.10 significance levels considering one-sided and two-sided tests, number of centers, K = 8, 16, 32, and 48 and mean sample size in each center, n_{ij} = 4, 8, 16, 32, and 64 and when the mean sample size of each treatment group are unequal ($n_{i1} \neq n_{i2}$) and equal ($n_{i1} = n_{i2}$) sizes. First, this study will show the results of Type I error simulated from the dropping method where hypothesis testing considers the significance in one–sided and two–sided tests.

4.1.1 Dropping case when $\hat{\omega}_i = 0$

- Comparing the Type I error when significance level considers the two-sided test

The results in this step simulate by the dropping case $\hat{\omega}_i$ = 0 method where the mean sample size in each treatment group are equal ($n_{i1} = n_{i2}$) and unequal ($n_{i1} \neq n_{i2}$). For this step, this study will present the nominal level in the two-sided test. Table 4.1 considers 0.01, 0.05, and 0.10 significance levels and when the mean treatment group sizes are equal but Table 4.2 – 4.5 consider when the mean sample size of each treatment group is unequal in any number of centers (K).

This study assign some symbols for ease of understanding; that is if the number of centers equals 8, 16, 32 and 48, this study will label the number of centers as small, moderate, large and very large, respectively. In other words, if mean treatment group sizes equal 4, 8, 16, 32, and 64, this study will label mean treatment group size as very small, small, moderate, large, and very large, respectively.

From **Table 4.1** indicates that

- **significance level at 0.01**

Each statistic test can control the Type I error in different cases. That is, numbers of centers have many sizes, but mean treatment group is large to very large so the Q_{WLS} and Z^2_{WLS} tests can control the Type I error. If the numbers of centers is moderate to large ($K \geq 16$), then the $Z^2_{WLS,R}$, Z^2_V and Z^2_K tests can control the Type I error when the mean sample size of the treatment group are small ($n_{ij} \leq 8$).

- **significance level at 0.05**

Almost every test (except the Z^2_K test) can control the Type I error. In every case of the number of centers, if mean treatment group size is large, ($n_{ij} \geq 32$), then the Q_{WLS} and Z^2_{WLS} tests can control the Type I error. Meanwhile, if mean treatment group sizes are small to moderate, ($n_{ij} \leq 16$), then the $Z^2_{WLS,R}$ and Z^2_V tests can control the Type I error.

- **significance level at 0.10**

For this table, the Q_{WLS} test can control the Type I error for every number of centers when the mean treatment group size is very large, ($n_{ij} = 64$). Next, the Z^2_{WLS} test can control the Type I error for one-sided test of significance level when the number of centers has the same size as mean treatment group size; that is, if the number of centers is small and mean treatment group is also small, or when both the number of centers and mean treatment group are moderate or large. The $Z^2_{WLS,R}$ test can control the Type I error when the mean treatment group is small sizes in every number of centers. Lastly, the Z^2_V test can control the Type I error when the number of centers are large to very large but mean treatment group is moderate.

Table 4.1 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when mean treatment group size are equal ($n_{i1} = n_{i2}$).

α	K	$n_{i1} = n_{i2}$	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.01	8	4	10.25	7.41	0.00	0.00	0.00
		8	4.23	3.02	0.00	0.00	0.00
		16	2.25	1.33*	0.00	0.00	0.00
		32	1.42*	0.83*	0.00	0.00	0.00
		64	1.27*	0.69*	0.00	0.00	0.00
	16	4	17.94	16.74	1.36*	0.81*	1.39*
		8	6.72	6.15	3.11	2.11	3.19
		16	2.82	2.49	3.87	2.93	3.62
		32	1.85	1.57	4.05	3.49	3.80
		64	1.36*	1.13*	4.09	3.66	3.90
	32	4	29.74	28.87	0.58*	0.36	3.89
		8	10.54	10.10	2.13	1.46*	4.06
		16	3.55	3.18	3.37	2.67	3.94
		32	1.71	1.53	3.89	3.41	4.13
		64	1.33*	1.28*	4.15	3.91	4.33
	48	4	42.69	41.59	0.80*	2.21	4.43
		8	13.53	12.69	1.10*	0.74*	3.46
		16	4.23	3.91	2.42	1.72	3.52
		32	1.88	1.83	2.81	2.37	3.39
		64	1.42*	1.31*	3.13	3.07	3.47

Note: * denotes the statistic test can control the type I error.

.

Table 4.1 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when mean treatment group size are equal ($n_{i1} = n_{i2}$). (Continued)

α	K	$n_{i1} = n_{i2}$	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.05	8	4	20.31	14.83	5.92*	3.45	3.03
		8	11.27	7.44	10.52	7.72	8.31
		16	7.61	4.61*	12.03	9.90	10.42
		32	6.36	3.68	12.21	10.74	11.26
		64	5.65*	2.91	11.84	10.72	10.96
	16	4	32.07	26.98	5.37*	4.34*	10.83
		8	15.36	12.13	9.73	7.82	12.14
		16	9.31	6.53	10.97	9.86	12.05
		32	6.83	4.67*	11.94	11.45	12.40
		64	5.42*	3.57	12.34	12.26	12.57
	32	4	48.03	42.61	5.36*	9.81	11.98
		8	21.26	17.43	5.71*	4.87*	9.45
		16	10.82	8.25	7.40	6.27*	9.35
		32	7.69	5.65*	8.55	7.86	9.80
		64	5.91*	4.49*	9.70	9.25	10.07
	48	4	60.37	55.07	11.02	26.65	11.80
		8	26.27	22.01	4.74*	6.42	8.44
		16	12.04	9.61	6.34	5.14*	8.96
		32	8.01	6.19	7.62	6.69	9.10
		64	6.04	4.66*	7.53	7.14	8.11

Note: * denotes the statistic test can control the type I error.

Table 4.1 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when mean treatment group size are equal ($n_{i1} = n_{i2}$). (Continued)

α	K	$n_{i1} = n_{i2}$	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.10	8	4	28.58	19.84	16.22	13.86	15.52
		8	17.88	11.20*	19.99	18.03	20.28
		16	13.07	7.57	22.35	20.88	22.09
		32	11.73	6.32	22.55	21.91	22.31
		64	11.13*	5.65	22.31	22.03	22.14
	16	4	41.88	33.63	10.91*	11.28*	18.41
		8	22.84	17.10	14.54	12.83	18.54
		16	15.67	11.18*	16.21	14.97	18.18
		32	12.72	8.80*	17.39	16.87	18.25
		64	11.00*	7.41	18.22	17.80	18.71
	32	4	57.66	50.15	15.60	28.79	18.75
		8	30.00	23.96	10.50*	11.85	14.93
		16	18.20	13.66	12.10	10.74*	14.50
		32	13.65	10.86	13.66	12.90	14.97
		64	11.74	9.01*	14.41	13.89	15.27
	48	4	69.81	62.08	26.04	51.60	18.50
		8	36.47	29.09	10.49*	16.85	13.60
		16	20.01	15.62	11.68*	10.70*	14.28
		32	14.68	11.78*	12.68	11.88*	14.07
		64	11.45*	9.35*	12.34	12.22	13.35

Note: * denotes the statistic test can control the type I error.

Next, this study will consider the percentage of simulations in which the null hypothesis of homogeneity is rejected when the null hypothesis is true (Type I error) where tests are done at α = 0.01, 0.05, and 0.10 in any K and mean sample sizes are unequal ($n_{i1} \neq n_{i2}$).

From **Table 4.2** indicates that

- **significance level at 0.01**

  Only the Z^2_{WLS} statistic test can control the Type I error; that is, when the mean treatment group size has large sample sizes ($n_{ij} \geq 32$). Meanwhile, the other tests cannot control the Type I error in every case of mean treatment group size. Furthermore, the percentage of simulation of the $Z^2_{WLS,R}$, Z^2_V, and Z^2_K tests are equals zero in every case of mean treatment group size.

- **significance level at 0.05**

  When the number of center is small, (K = 8), almost all tests (except the Q_{WLS} test) can control the Type I error. That is, the Z^2_{WLS} test can control the Type I error when mean sample size is moderate to very large. Meanwhile, when the mean sample size of treatment group is very small and mean sample size of the control group is moderate to large then the $Z^2_{WLS,R}$ and the Z^2_K tests can control the Type I error, but if both the mean group size is small, then the Z^2_V test can control the Type I error. On the contrary, the Q_{WLS} test cannot control the Type I error for all mean sample sizes.

- **significance level at 0.10**

  When the mean sample sizes of treatment groups are very small and the mean sample size of the control group is large, then the $Z^2_{WLS,R}$ test can control the Type I error. However, the Z^2_V test can control the Type I error in every case of mean sample size of control groups when the mean sample size of treatment group is very small.

Table 4.2 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when number of center (*K*) = 8 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.01	4	8	20.16	16.84	0.00	0.00	0.00
		16	43.12	39.32	0.00	0.00	0.00
		32	59.67	57.50	0.00	0.00	0.00
		64	66.03	64.61	0.00	0.00	0.00
	8	16	6.46	4.75	0.00	0.00	0.00
		32	13.14	11.05	0.00	0.00	0.00
		64	19.39	17.28	0.00	0.00	0.00
	16	32	3.01	2.02	0.00	0.00	0.00
		64	5.00	3.81	0.00	0.00	0.00
	32	64	1.70	1.03*	0.00	0.00	0.00
0.05	4	8	32.63	26.86	6.62	3.96	4.67*
		16	54.70	49.66	5.88*	2.90	4.85*
		32	66.28	63.34	4.40*	1.96	3.88
		64	71.78	69.36	4.41*	2.44	4.32*
	8	16	14.81	10.48	10.38	7.91	8.59
		32	21.40	17.53	10.64	7.97	8.89
		64	28.47	24.04	9.66	7.78	8.32
	16	32	8.46	5.73*	12.14	10.17	10.85
		64	11.99	8.53	11.53	10.02	10.43
	32	64	6.7	3.95	11.68	10.55	10.72
0.10	4	8	40.72	32.43	13.34	11.32*	16.01
		16	60.60	54.62	11.53*	9.20*	15.31
		32	70.07	66.25	10.94*	8.89*	14.06
		64	74.78	71.76	12.35	8.81*	13.99
	8	16	21.22	14.75	20.17	17.76	20.54
		32	27.84	21.35	19.04	16.16	20.24
		64	34.74	28.41	18.61	15.13	19.45
	16	32	14.35	8.41*	21.93	20.62	21.80
		64	18.10	11.94	20.78	19.71	21.29
	32	64	12.03	6.69	22.23	21.61	22.03

Note: * denotes the statistic test can control the type I error.

From **Table 4.3** indicates that

- **significance level at 0.01**

When the numbers of centers have moderate sizes, then the $Z^2_{WLS,R}$, Z^2_V, and Z^2_K tests can control the Type I error, but the Q_{WLS} and Z^2_{WLS} tests cannot control the Type I error in every case of mean treatment groups size. That is, the $Z^2_{WLS,R}$ test can control the Type I error when the mean sample size of treatment groups is small to moderate ($n_{ij} \leq 16$). Also, when the mean group size is small, then the Z^2_V test can control the Type I error. Furthermore, when mean treatment group size is small and another mean treatment group is large then the Z^2_K test can control the Type I error.

- **significance level at 0.05**

For this number of centers, the Z^2_{WLS}, $Z^2_{WLS,R}$, and Z^2_V tests can control the Type I error. However, the Q_{WLS} and Z^2_K tests cannot control the Type I error for all of the mean sample sizes. That is, when mean sample sizes of treatment groups are large, ($n_{ij} \geq 32$), then the Z^2_{WLS} test can control the type I error. When both mean sample size for treatment groups are small to moderate, ($n_{ij} \leq 16$), then the $Z^2_{WLS,R}$ test can control the Type I error. The Z^2_V test can control the Type I error when mean treatment group size is small and another mean treatment group is large.

- **significance level at 0.10**

The Z^2_{WLS} and $Z^2_{WLS,R}$ tests can control the Type I error when the number of centers is moderate, (*K*=16). That is, the Z^2_{WLS} test can control the Type I error when either mean treatment group size is large size. Meanwhile, the $Z^2_{WLS,R}$ test can control the Type I error when both mean treatment groups are small.

Table 4.3 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when number of center (*K*) = 16 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.01	4	8	36.53	35.15	1.28*	0.65*	1.87
		16	70.07	68.99	0.65*	0.28	1.57
		32	87.43	87.11	0.38	0.20	1.31*
		64	92.29	92.07	0.39	0.21	1.40*
	8	16	10.36	9.61	3.17	2.16	3.53
		32	24.84	23.85	2.58	1.67	3.06
		64	37.97	36.93	2.30	1.60	2.84
	16	32	3.63	3.26	3.67	3.04	3.70
		64	7.99	7.33	3.79	3.17	3.89
	32	64	1.72	1.58	4.09	3.71	3.85
0.05	4	8	51.84	47.06	4.02*	3.82*	9.85
		16	79.45	76.65	5.29*	8.10	8.90
		32	90.94	90.09	9.12	12.18	8.12
		64	94.13	93.61	12.26	12.68	8.18
	8	16	21.41	17.54	9.06	7.14	12.21
		32	37.14	33.15	7.45	5.56*	11.36
		64	48.05	44.86	6.51	4.60*	10.92
	16	32	10.79	7.96	11.24	10.01	12.58
		64	16.47	13.27	10.23	9.31	11.88
	32	64	7.01	4.94*	11.89	11.40	12.81
0.10	4	8	60.15	53.30	11.48*	14.05	16.33
		16	83.83	80.22	20.78	29.60	15.74
		32	92.57	91.43	33.69	41.13	14.53
		64	95.07	94.40	40.73	42.66	14.42
	8	16	29.54	23.40	14.09	12.05	18.45
		32	44.56	38.63	13.14	10.83*	17.62
		64	54.38	49.61	13.79	8.83*	17.06
	16	32	17.32	12.50	16.59	15.03	18.44
		64	23.46	18.54	15.63	13.99	18.23
	32	64	12.95	8.55*	17.50	16.89	18.35

Note: * denotes the statistic test can control the type I error.

From **Table 4.4** indicates that

- **significance level at 0.01**

The $Z^2_{WLS,R}$ and Z^2_V tests can control the Type I error when the number of centers are large, ($K = 32$). Meanwhile, the Q_{WLS}, Z^2_{WLS}, and Z^2_K tests cannot control the Type I error in every case of mean sample size. That is, when the mean treatment group size is small and another group is large then $Z^2_{WLS,R}$ and Z^2_V tests can control the Type I error. Furthermore, the Z^2_V test also can control the Type I error when both mean treatment group sizes are small, ($n_{ij} \leq 8$).

- **significance level at 0.05**

In this case, the $Z^2_{WLS,R}$ and Z^2_V tests can control the Type I error. That is, when the mean treatment group is small to moderate and mean group size of the control group is small to moderate, too, ($n_{ij} \geq 8$) then the $Z^2_{WLS,R}$ test can control the Type I error. Meanwhile, when both mean treatment groups are small to large, ($n_{ij} \leq 64$), then the Z^2_V test can control the type I error.

- **significance level at 0.10**

The $Z^2_{WLS,R}$ and Z^2_V tests can control the Type I error. When the mean treatment group size is small and another group is moderate then the $Z^2_{WLS,R}$ test can control the Type I error**.** Also, when the mean treatment group is moderate and another is large to very large then the Z^2_V test can control the Type I error.

Table 4.4 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when number of center (*K*) = 32 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.01	4	8	60.58	59.77	0.48	0.90*	2.70
		16	92.43	92.10	2.59	7.73	2.60
		32	98.75	98.69	8.95	17.05	2.51
		64	99.47	99.47	13.94	21.82	2.36
	8	16	16.64	16.00	1.75	1.09*	3.50
		32	42.67	42.02	1.12*	0.71*	3.66
		64	63.28	62.68	0.86*	0.53*	3.37
	16	32	5.53	5.21	2.97	2.38	4.13
		64	12.68	12.25	2.44	1.76	3.89
	32	64	2.60	2.42	3.50	3.19	4.27
0.05	4	8	74.59	70.65	10.69	20.44	8.62
		16	95.93	94.93	35.39	50.18	8.01
		32	99.22	99.05	58.21	63.69	8.06
		64	99.69	99.62	66.42	65.43	7.88
	8	16	30.97	26.47	5.35*	4.57*	9.27
		32	56.92	53.49	6.33	4.92*	9.04
		64	71.93	69.68	7.38	4.20*	9.16
	16	32	13.89	11.06	7.31	5.87*	9.77
		64	24.14	21.04	6.17	4.57*	9.72
	32	64	8.46	6.74	8.89	8.10	10.60
0.10	4	8	81.16	75.93	29.34	45.16	14.26
		16	97.21	96.24	62.96	74.63	13.53
		32	99.41	99.27	82.12	84.40	13.03
		64	99.75	99.69	87.60	86.08	13.14
	8	16	40.72	33.78	11.29*	13.45	14.74
		32	64.27	58.96	16.59	17.10	14.37
		64	76.69	73.43	22.96	17.86	14.64
	16	32	21.13	17.03	12.51	10.68*	14.86
		64	32.16	27.06	11.77*	9.51*	14.55
	32	64	14.42	11.73*	14.18	12.96	15.82

Note: * denotes the statistic test can control the type I error.

From **Table 4.5** indicates that

- **significance level at 0.01**

When the number of centers is very large, (K = 48), the $Z^2_{WLS,R}$ and Z^2_V test have done well, that is, the $Z^2_{WLS,R}$ and Z^2_V tests can control the Type I error in many cases of mean sample size of treatment group, especially, when the mean sample group is small and the another mean sample group is moderate to very large then the $Z^2_{WLS,R}$ and Z^2_V tests can control the Type I error.

- **significance level at 0.05**

Two statistic tests ($Z^2_{WLS,R}$ and Z^2_V) can control the Type I error, that is, the other tests cannot control the Type I error in every case of mean treatment group size. Also, when the mean sample size of treatment group is moderate to large, ($n_{i1} \geq$ 16), and mean sample size in another group is large to very large then the $Z^2_{WLS,R}$ test can control the Type I error, but the Z^2_V test can control the Type I error when the mean sample size of control is large, ($n_{ij} \geq 32$).

- **significance level at 0.10**

There are two tests that can control the Type I error. That is, when the mean sample size in each group is greater than moderate, ($n_{ij} \geq 16$), then the $Z^2_{WLS,R}$ and Z^2_V test can control the Type I error. On the other hand, the Q_{WLS}, Z^2_{WLS}, and Z^2_K tests cannot control the Type I error in every case of mean sample size.

Table 4.5 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when number of center (*K*) = 48 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.01	4	8	75.23	74.54	2.42	6.75	3.02
		16	98.07	97.95	17.71	33.61	2.34
		32	99.87	99.86	40.67	51.57	2.26
		64	99.97	99.97	54.17	59.06	2.10
	8	16	22.09	21.12	1.00*	0.64*	2.95
		32	55.49	54.70	0.77*	0.67*	3.12
		64	78.49	78.08	0.90*	0.91*	3.39
	16	32	6.44	5.94	2.06	1.53	3.55
		64	16.19	15.57	1.63	0.95*	3.46
	32	64	2.57	2.29	2.49	2.12	3.14
0.05	4	8	86.02	83.16	26.95	44.72	7.99
		16	99.13	98.91	67.81	77.13	7.13
		32	99.92	99.92	86.85	85.89	7.34
		64	99.99	99.98	91.63	88.17	7.17
	8	16	38.75	34.00	5.54*	8.25	8.49
		32	68.85	64.75	10.44	12.25	8.35
		64	84.47	82.78	16.69	12.90	8.52
	16	32	16.54	13.43	5.98*	4.76*	8.96
		64	28.85	25.24	5.39*	3.95	8.77
	32	64	9.28	7.31	6.34	5.73*	8.07
0.10	4	8	90.06	99.17	50.86	70.07	13.07
		16	99.43	99.17	86.68	89.65	11.79*
		32	99.95	99.92	96.41	94.29	12.60
		64	100.00	99.99	98.24	96.01	12.48
	8	16	49.06	41.27	13.84	22.05	13.87
		32	75.20	70.46	26.06	32.62	13.72
		64	87.49	85.32	38.05	35.06	13.35
	16	32	24.92	20.06	11.10*	10.13*	14.04
		64	37.50	31.79	11.50*	10.44*	14.18
	32	64	16.12	12.57	11.06*	10.05*	12.80

Note: * denotes the statistic test can control the type I error.

- **Comparing the type I error when significance level considers the one-sided test.**

The results in this step of the simulation still use the dropping case $\hat{\omega}_i = 0$ method. Now, hypothesis testing will consider significance level of the one-sided test. Table 5.1 considers equal mean treatment group size ($n_{i1} = n_{i2}$) and Table 5.2 – 5.5 consider when the mean sample size of each treatment group is unequal ($n_{i1} \neq n_{i2}$).

From **Table 5.1** indicates that

- **significance level at 0.01(one-sided test)**

Almost all statistic tests can control the Type I error. That is, when the numbers of center are any size and mean treatment group is very large then Q_{WLS} and Z^2_{WLS} test can control the Type I error. Also, when the number of centers is moderate to large size and mean treatment group is very small then the $Z^2_{WLS,R}$ test can control the Type I error. The Z^2_V test can control the Type I error when the number of centers is very large size and mean treatment group size is small size.

- **significance level at 0.05 (one-sided test)**

The Q_{WLS} and Z^2_{WLS} tests can control the Type I error when the mean treatment group is large to very large in every case of the number of centers. The $Z^2_{WLS,R}$ test can control the Type I error when the number of centers is greater than large, ($K \geq 32$), and mean treatment group size is small. The Z^2_V test can control the Type I error when the number of centers is moderate to large, ($K \geq 16$), and mean treatment group size is small.

- **significance level at 0.10 (one-sided test)**

The Q_{WLS} test can control the Type I error when the number of center is small to moderate but mean treatment group is very large. The Z^2_{WLS} test can control the Type I error when the mean treatment group is moderate to very large for every number of center. The $Z^2_{WLS,R}$ and Z^2_V tests can control the Type I error when the number of centers is greater than moderate but mean treatment group is small.

Table 5.1 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when mean treatment group size are equal ($n_{i1}=n_{i2}$).

α	K	$n_{i1}=n_{i2}$	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.01	8	4	7.70	5.26	0.00	0.00	0.00
		8	2.93	1.93	0.00	0.00	0.00
		16	1.22	0.65*	0.00	0.00	0.00
		32	0.93	0.45*	0.00	0.00	0.00
		64	0.64*	0.35*	0.00	0.00	0.00
	16	4	14.32	13.76	0.54*	0.16	0.24
		8	4.51	4.38	1.45	0.76	0.96
		16	1.67	1.63	1.87	1.27	1.58
		32	1.04	0.98	1.97	1.59	1.86
		64	0.65*	0.58*	2.17	1.71	1.90
	32	4	25.26	25.67	0.45*	0.13	2.18
		8	7.21	7.44	1.25	0.76	2.12
		16	1.87	1.93	1.74	1.33	2.39
		32	1.04	1.11	2.25	1.90	2.55
		64	0.68*	0.71*	2.67	2.42	2.74
	48	4	35.85	36.14	0.18	0.46*	2.63
		8	9.65	9.84	0.79	0.51*	2.09
		16	2.47	2.53	1.60	1.11	2.43
		32	1.04	1.06	2.22	1.81	2.47
		64	0.66*	0.68*	1.83	1.80	2.19

Note: * denotes the statistic test can control the type I error.

Table 5.1 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when mean treatment group size are equal ($n_{i1}=n_{i2}$). (Continued)

α	K	$n_{i1}=n_{i2}$	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.05	8	4	14.94	11.09	0.00	0.00	0.00
		8	7.22	5.07	0.00	0.00	0.00
		16	4.38	2.62*	0.00	0.00	0.00
		32	3.50	2.14*	0.00	0.00	0.00
		64	2.75*	1.63	0.00	0.00	0.00
	16	4	24.88	21.83	3.24*	2.29*	5.39
		8	10.73	8.89	6.31	5.01	7.36
		16	5.40	4.24	7.30	6.25	7.82
		32	3.77	2.95*	7.88	7.52	7.88
		64	2.69*	2.10*	8.21	8.13	8.27
	32	4	39.54	36.66	1.99*	2.86*	7.59
		8	15.23	13.13	3.41	2.36*	6.24
		16	6.52	5.30	4.79	3.72	6.33
		32	4.17	3.37	5.77	5.04	6.58
		64	3.02*	2.38*	6.54	6.21	6.96
	48	4	52.14	48.42	3.77	10.58	7.60
		8	19.15	16.95	2.53*	2.39*	5.39
		16	7.59	6.18	3.89	3.02*	6.06
		32	4.43	3.63	4.84	4.45	5.84
		64	2.98*	2.40*	4.74	4.52	5.20

Note: * denotes the statistic test can control the type I error.

Table 5.1 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when mean treatment group size are equal ($n_{i1}=n_{i2}$). (Continued)

α	K	$n_{i1}=n_{i2}$	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.10	8	4	20.31	14.83	5.92	3.45	3.03
		8	11.27	7.44	10.52	7.72	8.31
		16	7.61	4.61*	12.03	9.90	10.42
		32	6.36	3.68	12.21	10.74	11.26
		64	5.65*	2.91	11.84	10.72	10.96
	16	4	32.07	26.98	5.37*	4.34*	10.83
		8	15.36	12.13	9.73	7.82	12.14
		16	9.31	6.53	10.97	9.86	12.05
		32	6.83	4.67*	11.94	11.45	12.40
		64	5.42*	3.57	12.34	12.26	12.57
	32	4	48.03	42.61	5.36*	9.81	11.98
		8	21.26	17.43	5.71*	4.87*	9.45
		16	10.82	8.25	7.40	6.27	9.35
		32	7.69	5.65*	8.55	7.86	9.80
		64	5.91	4.49*	9.70	9.25	10.07
	48	4	60.37	55.07	11.02	26.65	11.80
		8	26.27	22.01	4.74*	6.42	8.44
		16	12.04	9.61	6.34	5.14*	8.96
		32	8.01	6.19	7.62	6.69	9.10
		64	6.04	4.66*	7.53	7.14	8.11

Note: * denotes the statistic test can control the type I error.

From **Table 5.2** indicates that

- **significance level at 0.01 (one-sided test)**

Only the Z^2_{WLS} statistic test can control the Type I error, that is, when the mean treatment group is large, ($n_{ij} \geq 32$), the Z^2_{WLS} test can control the Type I error. The other tests cannot control the type I error in every case of mean treatment group size.

- **significance level at 0.05 (one-sided test)**

The result from significance level at 0.05 is similar the result from significance level at 0.01. That is, only the Z^2_{WLS} statistic test can control the type I error. When the mean treatment group is large to very large, ($n_{ij} \geq 32$), the Z^2_{WLS} test can control the Type I error. The other tests cannot control the type I error in every case of mean treatment group size.

- **significance level at 0.10 (one-sided test)**

When the mean sample size of the treatment group is very small and mean sample size of control group is any size then the Z^2_K test can control the Type I error. The $Z^2_{WLS,R}$ test can control the Type I error when the mean sample size of treatment group is very small and mean of another group is large to very large. Lastly, when the mean treatment group is moderate to large, then the Z^2_{WLS} test can control the Type I error.

Table 5.2 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when number of center (*K*) = 8 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α/2	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.005	4	8	16.73	13.90	0.00	0.00	0.00
		16	38.97	35.15	0.00	0.00	0.00
		32	57.31	54.83	0.00	0.00	0.00
		64	64.48	62.86	0.00	0.00	0.00
	8	16	4.55	3.32	0.00	0.00	0.00
		32	10.83	8.95	0.00	0.00	0.00
		64	17.05	14.99	0.00	0.00	0.00
	16	32	1.96	1.18	0.00	0.00	0.00
		64	3.66	2.70	0.00	0.00	0.00
	32	64	0.99	0.57*	0.00	0.00	0.00
0.025	4	8	26.51	21.83	0.00	0.00	0.00
		16	49.06	45.11	0.00	0.00	0.00
		32	63.13	60.76	0.00	0.00	0.00
		64	69.11	67.04	0.00	0.00	0.00
	8	16	10.10	7.50	0.00	0.00	0.00
		32	17.23	14.33	0.00	0.00	0.00
		64	23.72	20.76	0.00	0.00	0.00
	16	32	5.46	3.67	0.00	0.00	0.00
		64	8.28	6.07	0.00	0.00	0.00
	32	64	3.74	2.23*	0.00	0.00	0.00
0.05	4	8	32.63	26.86	6.62	3.96	4.67*
		16	54.70	49.66	5.88	2.90	4.85*
		32	66.28	63.34	4.40*	1.96	3.88
		64	71.78	69.36	4.41*	2.44	4.32*
	8	16	14.81	10.48	10.38	7.91	8.59
		32	21.40	17.53	10.64	7.97	8.89
		64	28.47	24.04	9.66	7.78	8.32
	16	32	8.46	5.73*	12.14	10.17	10.85
		64	11.99	8.53	11.53	10.02	10.43
	32	64	6.7	3.95	11.68	10.55	10.72

Note: * denotes the statistic test can control the type I error.

From **Table 5.3** indicates that

- **significance level at 0.01 (one-sided test)**

The $Z^2_{WLS,R}$ and Z^2_K tests can control the Type I error. That is, when the mean treatment group is very small and for every mean size of another group then the $Z^2_{WLS,R}$ and Z^2_K tests can control the Type I error. The Q_{WLS}, Z^2_{WLS} and Z^2_V tests cannot control the Type I error in every case of mean treatment group size.

- **significance level at 0.05 (one-sided test)**

The Z^2_{WLS} test can control the Type I error when both mean treatment group sizes are large to very large sizes, ($n_{ij} \geq 32$). The $Z^2_{WLS,R}$ test can control the type I error when the mean treatment group is very small size and another group is small or very large. Also, when the mean treatment group is very small but another group is large or very large size then the Z^2_V test can control the Type I error.

- **significance level at 0.10 (one-sided test)**

When mean treatment group sizes are large to very large size then the Z^2_{WLS} test can control the Type I error. If the mean treatment group is very small and another group is moderate then the $Z^2_{WLS,R}$ test can control the Type I error. Also, the Z^2_V test can control the Type I error when the mean treatment group is small and another group is large to very large.

Table 5.3 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when number of center (K) = 16 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α/2	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.005	4	8	31.60	30.94	0.55*	0.19	0.47*
		16	66.27	65.84	0.29*	0.10	0.44*
		32	86.04	85.88	0.28*	0.11	0.40*
		64	91.55	91.46	0.26*	0.08	0.43*
	8	16	7.77	7.54	1.56	0.86	1.16
		32	20.97	20.60	1.24	0.80	1.10
		64	34.85	34.73	1.11	0.76	1.09
	16	32	2.37	2.29	1.80	1.20	1.56
		64	6.08	5.98	1.89	1.39	1.61
	32	64	0.95	0.93	1.88	1.45	1.56
0.025	4	8	44.37	41.20	2.22*	1.60	5.40
		16	75.25	73.54	1.51	1.47	4.61
		32	89.53	88.77	1.54	2.24*	4.47
		64	93.24	92.90	2.28*	2.55*	4.25
	8	16	15.80	13.49	5.85	4.53	7.62
		32	31.04	28.56	5.07	3.54	6.78
		64	43.29	41.14	4.12	3.08	6.55
	16	32	6.84	5.36	7.47	6.56	8.16
		64	11.63	9.98	6.83	6.18	7.75
	32	64	3.95	2.96*	8.00	7.40	8.18
0.05	4	8	51.84	47.06	4.02	3.82	9.85
		16	79.45	76.65	5.29*	8.10	8.90
		32	90.94	90.09	9.12	12.18	8.12
		64	94.13	93.61	12.26	12.68	8.18
	8	16	21.41	17.54	9.06	7.14	12.21
		32	37.14	33.15	7.45	5.56*	11.36
		64	48.05	44.86	6.51	4.60*	10.92
	16	32	10.79	7.96	11.24	10.01	12.58
		64	16.47	13.27	10.23	9.31	11.88
	32	64	7.01	4.94*	11.89	11.40	12.81

Note: * denotes the statistic test can control the type I error.

From **Table 5.4** indicates that

- **significance level at 0.01 (one-sided test)**

Two statistic tests can control the Type I error. That is, when the mean treatment group is very small and another group is moderate then the $Z^2_{WLS,R}$ test can control the Type I error, and when the mean treatment group is small and another group is very large then the $Z^2_{WLS,R}$ test can control, too. The Z^2_V test can control the Type I error when the mean treatment group is a small and for every mean size of another group.

- **significance level at 0.05 (one-sided test)**

The $Z^2_{WLS,R}$ test can control the Type I error when the mean treatment group is small and for every mean size of another group. The Z^2_V test can control the Type I error when the mean treatment group is small and another group is moderate, and in another case, when the mean treatment group is moderate and mean control group is very large.

- **significance level at 0.10 (one-sided test)**

The $Z^2_{WLS,R}$ test can control the Type I error when both mean treatment group sizes are small to moderate. Also, the Z^2_V test can control the Type I error when the mean treatment group are small to large and when mean treatment group is moderate and another group is very large. Meanwhile, the Q_{WLS}, Z^2_{WLS}, and Z^2_K tests cannot control the Type I error in every case of mean treatment group size.

Table 5.4 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when number of center (K) = 32 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α/2	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.005	4	8	55.16	55.59	0.15	0.13	1.54
		16	90.59	90.74	0.45*	2.09	1.50
		32	98.49	98.49	2.17	6.64	1.41
		64	99.45	99.45	3.90	9.32	1.36
	8	16	12.88	13.17	1.16	0.67*	2.23
		32	37.99	38.26	0.79	0.43*	2.16
		64	60.11	60.34	0.60*	0.38*	2.20
	16	32	3.77	3.83	2.04	1.55	2.75
		64	9.96	10.13	1.67	1.22	2.53
	32	64	1.44	1.47	2.42	2.10	2.80
0.025	4	8	68.50	65.69	2.92*	6.34	5.32
		16	94.39	93.73	14.54	26.68	4.92
		32	98.98	98.91	32.08	41.62	4.84
		64	99.61	99.56	40.84	47.22	4.58
	8	16	23.65	21.15	3.00*	2.14*	6.16
		32	50.87	48.13	2.28*	1.57	6.11
		64	68.11	66.48	2.36*	1.55	6.23
	16	32	9.04	7.84	4.87	3.80	6.96
		64	18.55	16.56	3.96	3.05*	6.75
	32	64	5.16	4.34	6.05	5.34	7.18
0.05	4	8	74.59	70.65	10.69	20.44	8.62
		16	95.93	94.93	35.39	50.18	8.01
		32	99.22	99.05	58.21	63.69	8.06
		64	99.69	99.62	66.42	65.43	7.88
	8	16	30.97	26.47	5.35*	4.57*	9.27
		32	56.92	53.49	6.33	4.92*	9.04
		64	71.93	69.68	7.38	4.20	9.16
	16	32	13.89	11.06	7.31	5.87	9.77
		64	24.14	21.04	6.17	4.57*	9.72
	32	64	8.46	6.74	8.89	8.10	10.60

Note: * denotes the statistic test can control the type I error.

From **Table 5.5** indicates that

- **significance level at 0.01 (one-sided test)**

The $Z^2_{WLS,R}$ and Z^2_V tests can control the Type I error. That is, the $Z^2_{WLS,R}$ test can control the Type I error when both mean treatment group are very small to small and when mean treatment group is small in the treatment arm and in moderate to very large size control arm have. The Z^2_V test can control the Type I error when both mean treatment groups are small to moderate, and when mean sample size of the treatment arm is moderate and another arm is very large size.

- **significance level at 0.05 (one-sided test)**

Two statistic tests can control the Type I error. That is, the $Z^2_{WLS,R}$ and Z^2_V tests can control the Type I error when both mean treatment groups are small to moderate, and when mean treatment group is moderate in the treatment arm and in another arm is very large.

- **significance level at 0.10 (one-sided test)**

The $Z^2_{WLS,R}$ test can control the Type I error when the mean treatment groups are small to moderate size and when mean treatment group is moderate in the treatment arm and in another arm is very large. The Z^2_V test can control the Type I error when the mean treatment group is moderate to large and when mean treatment groups are large to very large.

Table 5.5 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when number of center (K) = 48 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α/2	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.005	4	8	70.51	70.80	0.58*	1.96	1.91
		16	97.47	97.49	6.80	18.15	1.50
		32	99.84	99.84	21.53	34.70	1.21
		64	99.96	99.96	33.27	42.42	1.23
	8	16	17.63	17.81	0.59*	0.34*	2.01
		32	50.52	50.71	0.37*	0.22	2.04
		64	76.11	76.27	0.29*	0.15	2.13
	16	32	4.27	4.36	1.41	1.09	2.36
		64	12.88	13.04	1.09	0.66*	2.32
	32	64	1.43	1.48	1.66	1.49	1.92
0.025	4	8	81.35	79.37	10.37	23.02	5.19
		16	98.77	98.49	43.97	59.82	4.29
		32	99.90	99.89	69.47	73.31	4.37
		64	99.98	99.98	79.10	77.81	4.26
	8	16	30.90	27.79	2.16*	2.61*	5.33
		32	62.46	60.00	3.48	3.85	5.55
		64	81.76	80.64	5.61	5.09	5.74
	16	32	11.00	9.16	3.61	2.65*	5.89
		64	22.51	20.11	3.10*	1.92*	5.81
	32	64	5.59	4.66	4.14	3.60	5.09
0.05	4	8	86.02	83.16	26.95	44.72	7.99
		16	99.13	98.91	67.81	77.13	7.13
		32	99.92	99.92	86.85	85.89	7.34
		64	99.99	99.98	91.63	88.17	7.17
	8	16	38.75	34.00	5.54*	8.25	8.49
		32	68.85	64.75	10.44	12.25	8.35
		64	84.47	82.78	16.69	12.90	8.52
	16	32	16.54	13.43	5.98	4.76*	8.96
		64	28.85	25.24	5.39*	3.95	8.77
	32	64	9.28	7.31	6.34	5.73*	8.07

Note: * denotes the statistic test can control the type I error.

4.1.2 Adding a constant when $\hat{\omega}_i = 0$

- Comparing the Type I error when significance level consider for the two-sided test.

This step shows the result from adding constant when X_{ij} = 0 or $X_{ij} = n_{ij}$ take effect $\hat{\omega}_i = 0$ by adding the X_{ij} with 0.5 and n_{ij} with 1. The results consider at the 0.01, 0.05, and 0.10 significance levels where the mean sample size in each treatment group are equal ($n_{i1} = n_{i2}$) and unequal ($n_{i1} \neq n_{i2}$).

For this step, this study will present at the two-sided test of significance level. Table 6.1 presents when the mean treatment group sizes are equal and Tables 6.2 – 6.5 present when mean sample sizes of each treatment group are unequal.

From **Table 6.1** indicates that

- **significance level at 0.01**

The Q_{WLS} and Z^2_{WLS} tests can control the Type I error. The Q_{WLS} test can control the Type I error when mean treatment groups are very large for every number of centers. Also, the Z^2_{WLS} test can control the Type I error when the mean treatment groups size is small to very large where the number of centers is small. Also, when the number of centers is moderate to large and mean treatment group is very large, then the Z^2_{WLS} test also can control the Type I error.

- **significance level at 0.05**

The Q_{WLS} test can control the Type I error when the mean treatment group is small and very large where the number of centers is small to moderate. However, if the number of centers is greater than 32 then the Q_{WLS} test can control the Type I error when mean treatment group is very large. The Z^2_{WLS} test can control the Type I error when the mean treatment group is greater than small for every number of centers. In Addition, the Z^2_{V} test can control the Type I error when the number of centers is large and mean treatment groups equal moderate size.

- **significance level at 0.10**

The Q_{WLS} test can control the Type I error when the mean treatment group is small and very large for every number of centers. The Z^2_{WLS} test can control the Type I error when the mean treatment group size is moderate to large when the number of centers is moderate to large. Also, the Z^2_V test can control the Type I error when the number of center is large and mean treatment group are moderate.

Table 6.1 Comparison of the actual type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when mean treatment group size are equal ($n_{i1} = n_{i2}$).

α	K	$n_{i1} = n_{i2}$	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.01	8	4	0.52*	0.29	0.00	0.00	0.00
		8	1.57	1.01*	0.00	0.00	0.00
		16	1.66	0.97*	0.00	0.00	0.00
		32	1.69	0.97*	0.00	0.00	0.00
		64	1.28*	0.65	0.00	0.00	0.00
	16	4	0.29	0.24	21.73	19.47	20.30
		8	1.95	1.68	6.21	4.78	5.38
		16	2.14	1.85	4.17	3.34	4.02
		32	1.77	1.48*	4.00	3.47	3.81
		64	1.29*	1.05*	4.37	3.87	4.09
	32	4	0.08	1.30*	29.59	25.17	33.67
		8	2.12	1.99	6.67	5.47	6.98
		16	2.29	2.02	3.21	2.61	4.03
		32	2.00	1.71	3.38	3.02	3.91
		64	1.25*	1.15*	3.91	3.74	4.13
	48	4	0.06	5.71	33.91	26.37	42.71
		8	2.00	1.96	5.67	4.51	7.18
		16	2.62	2.43	2.99	2.29	3.97
		32	1.91	1.74	3.14	2.63	3.64
		64	1.27*	1.12*	2.87	2.65	3.07

Note: * denotes the statistic test can control the type I error.

Table 6.1 Comparison of the actual type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when mean treatment group size are equal ($n_{i1} = n_{i2}$). (continued)

α	K	$n_{i1} = n_{i2}$	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.05	8	4	1.69	1.03	32.83	30.33	29.96
		8	5.32*	3.15	15.05	12.30	12.61
		16	6.60	3.78	12.52	10.44	10.91
		32	6.30*	3.61	12.23	10.77	11.28
		64	5.65*	2.91	11.84	10.72	10.96
	16	4	1.35	2.62	40.92	38.32	42.62
		8	6.30*	4.55*	16.84	15.50	17.48
		16	7.62	5.31*	11.76	10.71	12.69
		32	6.72	4.55*	11.98	11.49	12.41
		64	5.42*	3.57	12.34	12.26	12.57
	32	4	0.63	16.79	45.41	39.00	52.17
		8	6.76	6.06*	14.11	12.02	15.91
		16	8.52	6.41*	8.39	7.09	9.95
		32	7.54	5.45*	8.63	7.92	9.82
		64	5.91*	4.48*	9.70	9.25	10.07
	48	4	0.32	27.17	49.16	39.87	61.21
		8	6.23*	6.31*	12.27	9.93	15.44
		16	9.07	7.57	7.42	5.91*	9.81
		32	7.80	6.04*	7.73	6.75	9.12
		64	6.04*	4.66*	7.53	7.15	8.11

Note: * denotes the statistic test can control the type I error.

Table 6.1 Comparison of the actual type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when mean treatment group size are equal ($n_{i1} = n_{i2}$). (continued)

α	K	$n_{i1} = n_{i2}$	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.10	8	4	3.33	1.68	48.69	47.20	47.66
		8	9.92*	5.30	27.00	25.73	26.47
		16	11.90*	6.54	23.18	21.84	23.01
		32	11.62*	6.26	22.59	21.95	22.34
		64	11.11*	5.65	22.31	22.03	22.14
	16	4	2.43	13.16	49.31	45.99	52.65
		8	10.44*	8.66*	23.70	22.03	25.03
		16	13.45	9.63*	17.29	16.33	18.86
		32	12.63	8.69*	17.49	16.95	18.34
		64	11.00*	7.41	18.22	17.80	18.71
	32	4	1.37	31.30	53.16	46.08	61.22
		8	11.09*	11.99	19.42	17.73	22.30
		16	14.93	11.38*	12.98	11.88	15.62
		32	13.41	10.69*	13.74	12.94	15.01
		64	11.72*	9.01*	14.41	13.89	15.27
	48	4	0.78	42.85	58.13	47.35	70.63
		8	11.14*	12.91	17.12	14.56	21.26
		16	15.90	13.00	12.50	11.11*	15.07
		32	14.42	11.67*	12.79	11.96	14.16
		64	11.44*	9.35*	12.34	12.22	13.35

Note: * denotes the statistic test can control the type I error.

From **Table 6.2** indicates that

- **significance level at 0.01**

The Q_{WLS} test can control the Type I error when the mean treatment size in the treatment arm is very small and another arm is moderate to large. Also, the Z_{WLS}^2 test can control the Type I error when the mean sample size in the treatment arm is very small and for every mean sample size in the control arm. Also, when mean treatment group sizes are small to moderate or moderate to large or large to very large then the Z_{WLS}^2 test can control the Type I error, too.

- **significance level at 0.05**

The Q_{WLS} test can control the Type I error when the mean sample size in the treatment arm is very small where the mean sample size in control arm is moderate to very large and when mean sample size in each group is large to very large. The Z_{WLS}^2 test can control the Type I error when the mean sample size in each group is small to very large.

- **significance level at 0.10**

Only the Z_{WLS}^2 test can control the Type I error. That is, when the mean sample size in the treatment arm is small or moderate but mean sample size in the control arm is very large, then the Z_{WLS}^2 test can control the Type I error.

Table 6.2 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when number of center (*K*) = 8 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.01	4	8	0.78	0.56	0.00	0.00	0.00
		16	1.11*	0.71*	0.00	0.00	0.00
		32	1.23*	0.67	0.00	0.00	0.00
		64	1.58	0.85*	0.00	0.00	0.00
	8	16	1.84	1.22*	0.00	0.00	0.00
		32	2.76	1.91	0.00	0.00	0.00
		64	3.91	2.57	0.00	0.00	0.00
	16	32	2.18	1.35*	0.00	0.00	0.00
		64	3.30	2.19	0.00	0.00	0.00
	32	64	1.65	0.98*	0.00	0.00	0.00
0.05	4	8	2.97	1.82	24.35	21.68	21.06
		16	3.71	2.33	23.40	21.38	20.11
		32	3.83	2.43	24.27	22.56	20.16
		64	4.80*	2.92	24.57	22.71	20.28
	8	16	6.61	3.85	13.20	11.16	11.12
		32	7.85	5.20*	13.80	11.70	11.39
		64	10.61	6.98	12.78	10.86	10.50
	16	32	7.33	4.67*	12.56	10.50	11.20
		64	9.94	6.41*	11.75	10.19	10.58
	32	64	6.30	3.61	12.23	10.77	11.28
0.10	4	8	5.60	2.96	38.85	37.63	38.09
		16	6.54	3.70	37.48	36.53	36.63
		32	6.84	3.80	37.93	36.78	37.27
		64	8.21*	4.78	37.76	36.79	37.30
	8	16	11.59*	6.59	24.82	23.54	24.43
		32	13.42	7.81	24.22	23.09	24.12
		64	16.52	10.59*	23.54	21.97	23.71
	16	32	12.88	7.28	22.50	21.43	22.25
		64	15.80	9.89*	21.32	20.46	21.82
	32	64	11.98	6.62	22.26	21.67	22.04

Note: * denotes the statistic test can control the type I error.

From **Table 6.3** indicates that

- **significance level at 0.01**

When the mean sample size in the treatment arm is very small and for every mean sample size in the control arm then the Q_{WLS} and Z^2_{WLS} tests can control the Type I error. Furthermore, the Z^2_{WLS} test can control the Type I error when mean treatment group sizes are large to very large.

- **significance level at 0.05**

The Q_{WLS} test can control the Type I error when the mean sample size in the treatment group is very small but mean sample size in the control arm is large to very large. The Z^2_{WLS} test can control the Type I error when the mean sample size in the treatment group is small and mean size of the control arm is moderate or when mean sample size in the treatment group is moderate and mean size of the control arm is large or mean sample size in each group is large to very large.

- **significance level at 0.10**

Only the Z^2_{WLS} test can control the Type I error. That is, when the mean sample size in the treatment group is a very small in every mean size of the control arm, or when the mean sample size in the treatment group is small to moderate but the control group is moderate to large.

Table 6.3 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when number of center (*K*) = 16 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.01	4	8	0.66	0.62	13.98	12.45	12.84
		16	0.84*	0.71*	12.54	11.07	11.55
		32	1.09*	0.97*	12.15	10.74	11.60
		64	1.59	1.47*	12.25	10.78	11.84
	8	16	2.43	2.10	5.31	4.23	5.01
		32	3.70	3.30	4.59	3.71	4.59
		64	5.51	4.86	4.34	3.54	4.27
	16	32	2.29	2.02	3.89	3.29	3.89
		64	4.71	4.30	4.06	3.36	3.98
	32	64	1.67	1.50*	4.09	3.71	3.86
0.05	4	8	2.59	2.56	29.71	27.77	30.16
		16	3.25	2.65	27.98	25.72	29.90
		32	3.81	3.02	26.98	24.52	30.15
		64	4.93*	4.00*	25.83	23.28	29.82
	8	16	7.83	5.53*	14.32	12.93	15.83
		32	10.66	8.00	12.97	11.44	15.56
		64	13.37	10.44	11.85	10.36	14.74
	16	32	8.44	5.98*	11.98	10.91	12.97
		64	12.54	9.51	11.02	10.16	12.37
	32	64	6.80	4.82*	11.95	11.43	12.83
0.10	4	8	5.14	8.89*	37.56	35.39	39.44
		16	5.80	8.54*	35.37	33.08	38.80
		32	6.69	9.46*	34.47	31.40	39.17
		64	8.33*	10.90*	32.95	29.72	39.25
	8	16	13.20	9.85*	20.28	18.68	22.81
		32	16.58	12.75	18.37	16.66	22.04
		64	20.07	15.61	17.08	15.14	21.87
	16	32	14.41	10.23*	17.46	16.20	18.91
		64	18.82	14.59	16.39	15.16	18.73
	32	64	12.78	8.35*	17.58	16.94	18.37

Note: * denotes the statistic test can control the type I error.

From **Table 6.4** indicates that

- **significance level at 0.01**

The Q_{WLS} and Z^2_{WLS} tests can control the Type I error when the mean sample size of the treatment group is very small and for every case of mean sample size of the control group.

- **significance level at 0.05**

The Q_{WLS} test can control the Type I error when the mean sample size in the treatment group is very small but another group is very large. Also, the Z^2_V test can control the Type I error when the mean sample in the treatment group is small to moderate but another group is large to very large.

- **significance level at 0.10**

Only the Z^2_V test can control the Type I error. That is, if the mean sample size in the treatment group is small to moderate but another group is large to very large then the Z^2_V test can control the Type I error. On the other hand, the Q_{WLS}, Z^2_{WLS}, $Z^2_{WLS,R}$, and Z^2_K tests can control the Type I error.

Table 6.4 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when number of center (*K*) = 32 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.01	4	8	0.71*	0.86*	17.06	14.61	18.97
		16	0.70*	0.86*	14.61	12.17	18.14
		32	0.93*	1.12*	13.80	10.77	18.66
		64	1.25*	1.42*	13.35	10.28	19.52
	8	16	3.00	2.83	4.28	3.33	5.63
		32	4.87	4.60	3.73	2.74	5.90
		64	7.94	7.59	3.29	2.22	5.90
	16	32	3.40	3.14	3.36	2.77	4.38
		64	6.69	6.21	2.86	2.11	4.22
	32	64	2.48	2.32	3.53	3.20	4.30
0.05	4	8	2.35	8.23	29.59	25.78	34.36
		16	2.64	7.49	26.49	21.79	33.12
		32	3.20	7.84	24.36	19.33	33.84
		64	3.93	8.06	23.57	17.89	34.95
	8	16	8.60	7.15	10.23	8.33	13.15
		32	12.90	10.70	8.64	6.49	13.37
		64	17.63	14.90	7.33	5.86*	13.78
	16	32	10.12	8.12	8.11	6.93	10.31
		64	16.22	13.20	6.92	5.58*	10.24
	32	64	8.32	6.64	8.93	8.16	10.65
0.10	4	8	4.05	18.18	37.15	32.38	42.83
		16	4.66	16.42	33.24	27.90	42.11
		32	5.81	16.57	31.18	24.90	42.66
		64	6.90	17.09	30.03	23.34	43.53
	8	16	14.24	12.14	15.33	13.31	19.41
		32	19.86	16.72	13.39	11.42*	19.18
		64	25.47	21.20	12.71	11.27*	20.29
	16	32	17.03	13.48	13.05	11.99	15.59
		64	24.15	19.40	11.96	10.87*	15.43
	32	64	14.13	11.58*	14.21	13.02	15.87

Note: * denotes the statistic test can control the type I error.

From **Table 6.5** indicates that

- **significance level at 0.01**

When the mean sample size of the treatment group is very small, if the mean sample size of the control group is large to very large, then Q_{WLS} test can control the Type I error but if the mean sample size of the control group is moderate to large, then the Z^2_{WLS} test can control the Type I error. The Z^2_V test can control the Type I error when the mean sample size of the treatment group is small to moderate size but the control group is very large.

- **significance level at 0.05**

When the mean sample size of the treatment group is small to large, then the $Z^2_{WLS,R}$ test can control the Type I error when the mean sample size of the control group is only very large but the Z^2_V test can control the Type I error when the mean sample size of the control group is large to very large.

- **significance level at 0.10**

The result of significance level at 0.10 is similar to the result of significance level at 0.05. That is, when the mean sample size of treatment group is small to large, then the $Z^2_{WLS,R}$ test can control the Type I error when the mean sample size of the control group is only very large size but the Z^2_V test can control the type I error when the mean sample size of the control group is large to very large.

Table 6.5 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 significance levels when number of center (*K*) = 48 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.01	4	8	0.40	1.80	18.48	14.98	22.80
		16	0.42	1.42*	15.31	11.61	21.58
		32	0.63	1.33*	13.78	9.94	22.83
		64	1.05*	1.98	12.17	8.12	22.90
	8	16	3.18	2.89	3.59	2.37	5.45
		32	5.37	5.06	2.73	1.59	5.75
		64	9.44	8.97	2.00	1.17*	5.97
	16	32	3.71	3.41	2.50	1.93	3.83
		64	7.61	7.24	2.10	1.39*	3.80
	32	64	2.39	2.15	2.52	2.17	3.14
0.05	4	8	1.38	12.59	30.55	25.16	38.61
		16	1.70	10.58	26.95	20.69	37.38
		32	2.47	10.49	24.70	16.66	38.96
		64	3.37	10.56	22.18	14.17	38.65
	8	16	9.20	8.10	8.88	6.72	12.68
		32	14.37	11.82	6.94	5.22*	12.82
		64	20.65	17.17	6.22	5.28*	13.19
	16	32	11.45	9.14	6.79	5.47*	9.37
		64	17.92	14.72	5.75*	5.09*	9.40
	32	64	8.96	7.08	6.38	5.85*	8.07
0.10	4	8	2.79	24.05	38.24	31.47	47.84
		16	3.49	20.57	34.63	26.78	46.43
		32	4.64	20.45	31.78	22.12	48.25
		64	6.27	19.65	28.76	19.01	47.91
	8	16	15.24	13.54	13.74	11.56*	19.01
		32	22.41	18.48	11.61*	11.13*	19.14
		64	29.61	24.26	11.36*	11.37*	19.44
	16	32	19.41	15.19	11.69*	10.72*	14.47
		64	26.76	21.34	10.96*	11.30*	14.77
	32	64	15.69	12.31	11.17*	10.11*	12.84

Note: * denotes the statistic test can control the type I error.

- Comparing the Type I error when significance level consider for the one-sided test.

The results in this step of the simulation still add the constant (0.5) in each cell of X_{ij}, n_{ij} by consider at one-sided test of significance level. Table 7.1 shows when the mean treatment group size is equal ($n_{i1} = n_{i2}$) but Tables 7.2 – 7.5 show when the mean sample size of each treatment group is unequal ($n_{i1} \neq n_{i2}$).

From **Table 7.1** indicates that

- **significance level at 0.01(one-sided test)**

The Q_{WLS} and Z^2_{WLS} tests can control the Type I error. That is, when the number of centers is small and the mean treatment group sizes are very small and very large then the Q_{WLS} test can control the Type I error. Also, when the number of centers is moderate to large and the mean treatment group is very large then the Q_{WLS} test also can control the Type I error. The Z^2_{WLS} test can control the Type I error when the number of centers is small and mean treatment group is small to very large and when the number of centers is moderate or large and the mean treatment group is very large.

- **significance level at 0.05(one-sided test)**

In every number of centers the Q_{WLS} test can control the Type I error where the mean treatment groups is very large. Also, the Z^2_{WLS} test can control the Type I error when the number of centers is small and moderate and the mean treatment group is small to large. Also, when the number of centers and the mean treatment group are large to very large, then the Z^2_{WLS} test can control the Type I error, too.

- **significance level at 0.10(one-sided test)**

The Q_{WLS} test can control the Type I error when the number of centers is small to moderate and the mean treatment groups is very large. When the number of centers is moderate and the mean treatment group is small to large, then the Z^2_{WLS} test can control the Type I error. Also, when the number of centers is greater than large and the mean treatment group is large to very large, then the Z^2_{WLS} test also can control the Type I error.

Table 7.1 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when mean treatment group are equal ($n_{i1}=n_{i2}$).

α	K	$n_{i1}=n_{i2}$	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.01	8	4	0.27*	0.19	0.00	0.00	0.00
		8	0.95	0.59*	0.00	0.00	0.00
		16	0.95	0.52*	0.00	0.00	0.00
		32	0.90	0.43*	0.00	0.00	0.00
		64	0.64*	0.35*	0.00	0.00	0.00
	16	4	0.14	0.12	13.29	11.32	11.60
		8	1.21	1.20	3.01	1.97	2.11
		16	1.23	1.19	2.02	1.42	1.65
		32	1.02	0.96	1.97	1.60	1.86
		64	0.65*	0.58*	2.17	1.71	1.90
	32	4	0.06	0.21	24.22	20.68	27.23
		8	1.35	1.36	4.55	3.60	4.84
		16	1.21	1.27	2.06	1.71	2.66
		32	0.99	1.05	2.25	1.92	2.56
		64	0.68*	0.71	2.67	2.42	2.74
	48	4	0.02	2.20	28.20	22.02	35.65
		8	1.28	1.32	4.12	3.25	4.89
		16	1.55	1.58	2.10	1.43	2.75
		32	0.99	1.01	2.22	1.87	2.52
		64	0.66*	0.68*	1.83	1.80	2.19

Note: * denotes the statistic test can control the type I error.

Table 7.1 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when mean treatment group are equal ($n_{i1}=n_{i2}$). (Continued)

α	K	$n_{i1}=n_{i2}$	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.05	8	4	0.93	0.62	0.00	0.00	0.00
		8	2.99*	1.91*	0.00	0.00	0.00
		16	3.59	2.19*	0.00	0.00	0.00
		32	3.43	2.11*	0.00	0.00	0.00
		64	2.75*	1.63	0.00	0.00	0.00
	16	4	0.69	0.50	32.51	29.97	33.13
		8	3.66	2.98*	11.88	10.41	11.82
		16	4.39	3.29	7.77	6.99	8.25
		32	3.72	2.93*	7.91	7.56	7.91
		64	2.69*	2.10*	8.21	8.14	8.27
	32	4	0.25	6.85	37.99	32.47	44.25
		8	3.94	3.63	10.20	8.59	11.18
		16	4.87	3.87	5.43	4.55	6.85
		32	4.08	3.28*	5.82	5.15	6.61
		64	3.02*	2.38*	6.54	6.21	6.96
	48	4	0.13	15.39	42.23	33.61	52.99
		8	3.79	3.75	8.91	6.94	11.10
		16	5.46	4.52	4.66	3.84	6.71
		32	4.27	3.54	4.92	4.52	5.87
		64	2.96*	2.40*	4.74	4.52	5.20

Note: * denotes the statistic test can control the type I error.

Table 7.1 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when mean treatment group are equal ($n_{i1}=n_{i2}$). (Continued)

α	K	$n_{i1}=n_{i2}$	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.10	8	4	1.69	1.03	32.83	30.33	29.96
		8	5.32*	3.15	15.05	12.30	12.61
		16	6.60	3.78	12.52	10.44	10.91
		32	6.30	3.61	12.23	10.77	11.28
		64	5.65*	2.91	11.84	10.72	10.96
	16	4	1.35	2.62	40.92	38.32	42.62
		8	6.30	4.55*	16.84	15.50	17.48
		16	7.62	5.31*	11.76	10.71	12.69
		32	6.72	4.55*	11.98	11.49	12.41
		64	5.42*	3.57	12.34	12.26	12.57
	32	4	0.63	16.79	45.41	39.00	52.17
		8	6.76	6.06	14.11	12.02	15.91
		16	8.52	6.41	8.39	7.09	9.95
		32	7.54	5.45*	8.63	7.92	9.82
		64	5.91	4.84*	9.70	9.25	10.07
	48	4	0.32	27.17	49.16	39.87	61.21
		8	6.23	6.31	12.27	9.93	15.44
		16	9.07	7.57	7.42	5.91	9.81
		32	7.80	6.04	7.73	6.75	9.12
		64	6.04	4.66*	7.53	7.15	8.11

Note: * denotes the statistic test can control the type I error.

From **Table 7.2** indicates that

- **significance level at 0.01 (one-sided test)**

When the mean sample size of the treatment group is very small and the mean sample size of the control group is small to very large then the Q_{WLS} and Z^2_{WLS} tests can control the Type I error. Furthermore, when the mean sample sizes of the treatment group is moderate to large and mean sample size of the control group is large to very large, then the Z^2_{WLS} test can control the Type I error, too.

- **significance level at 0.05 (one-sided test)**

The Q_{WLS} test can control the Type I error when the mean sample size of the treatment group is very small and mean sample size of the control group is moderate to very large. Also, the Z^2_{WLS} test can control the Type I error when the mean sample size of the treatment group is very small size and mean sample size of the control group is very large, and when the mean sample size in each treatment group is small to large.

- **significance level at 0.10 (one-sided test)**

The Q_{WLS} test can control the Type I error when the mean sample size in each treatment group is very small and very large. Also, the Z^2_{WLS} test can control the Type I error when the mean sample size in each treatment group is small and large or moderate and large.

Table 7.2 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when number of center (*K*) = 8 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α/2	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.005	4	8	0.55*	0.37*	0.00	0.00	0.00
		16	0.64*	0.34*	0.00	0.00	0.00
		32	0.65*	0.37*	0.00	0.00	0.00
		64	0.73*	0.55*	0.00	0.00	0.00
	8	16	1.17	0.77	0.00	0.00	0.00
		32	1.80	1.33	0.00	0.00	0.00
		64	2.50	1.71	0.00	0.00	0.00
	16	32	1.31	0.74*	0.00	0.00	0.00
		64	2.12	1.48	0.00	0.00	0.00
	32	64	0.94	0.53*	0.00	0.00	0.00
0.025	4	8	1.73	0.95	0.00	0.00	0.00
		16	2.25*	1.39	0.00	0.00	0.00
		32	2.35*	1.53	0.00	0.00	0.00
		64	2.83*	1.96*	0.00	0.00	0.00
	8	16	3.66	2.42*	0.00	0.00	0.00
		32	4.94	3.21*	0.00	0.00	0.00
		64	6.68	4.55	0.00	0.00	0.00
	16	32	4.40	2.80*	0.00	0.00	0.00
		64	6.16	4.18	0.00	0.00	0.00
	32	64	3.66	2.18*	0.00	0.00	0.00
0.05	4	8	2.97	1.82	24.35	21.68	21.06
		16	3.71	2.33	23.40	21.38	20.11
		32	3.83	2.43	24.27	22.56	20.16
		64	4.80*	2.92	24.57	22.71	20.28
	8	16	6.61	3.85	13.20	11.16	11.12
		32	7.85	5.20*	13.80	11.70	11.39
		64	10.61	6.98	12.78	10.86	10.50
	16	32	7.33	4.67*	12.56	10.50	11.20
		64	9.94	6.41	11.75	10.19	10.58
	32	64	6.63	3.87	11.70	10.54	10.74

Note: * denotes the statistic test can control the type I error.

From **Table 7.3** indicates that

- **significance level at 0.01 (one-sided test)**

The Q_{WLS} and Z^2_{WLS} tests can control the Type I error when the mean sample size in the treatment group is very small and the mean sample size of the control group is small to large.

- **significance level at 0.05 (one-sided test)**

The Q_{WLS} test can control the Type I error when the mean sample size in the treatment group is very small and the mean sample size of the control group is moderate to very large. Also, the Z^2_{WLS} test can control the Type I error when the mean sample size in each group is very small and very large or large and very large.

- **significance level at 0.10 (one-sided test)**

The Q_{WLS} test can control the Type I error when the mean sample size in the treatment group is very small and the mean sample size of the control group is very large. Also, the Z^2_{WLS} test can control the Type I error when the mean sample size in each group is small and moderate or large and very large.

Table 7.3 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when number of center (*K*) = 16 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α/2	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.005	4	8	0.44*	0.42*	8.15	6.68	6.54
		16	0.46*	0.43*	7.14	5.97	5.67
		32	0.64*	0.62*	7.17	5.92	5.46
		64	1.03	1.00	7.19	6.08	5.60
	8	16	1.30	1.24	2.61	1.88	1.99
		32	2.47	2.36	2.25	1.55	1.83
		64	3.82	3.74	2.28	1.54	1.77
	16	32	1.41	1.36	1.95	1.26	1.63
		64	3.29	3.15	2.02	1.46	1.71
	32	64	0.88	0.85	1.89	1.46	1.56
0.025	4	8	1.52	1.17	22.56	21.07	22.12
		16	1.76*	1.37	20.91	19.50	21.55
		32	2.07*	1.61	20.28	18.39	21.26
		64	2.92*	2.38*	19.50	17.62	21.22
	8	16	4.60	3.62	9.80	8.60	10.52
		32	6.75	5.56	9.05	7.69	9.73
		64	9.21	7.68	7.98	6.92	9.54
	16	32	5.12	3.76	7.91	7.02	8.50
		64	7.95	6.54	7.34	6.71	8.05
	32	64	3.82	2.87*	8.04	7.43	8.23
0.05	4	8	2.59	2.56	29.71	27.77	30.16
		16	3.25	2.65	27.98	25.72	29.90
		32	3.81	3.02	26.98	24.52	30.15
		64	4.93*	4.00	25.83	23.28	29.82
	8	16	7.83	5.53*	14.32	12.93	15.83
		32	10.66	8.00	12.97	11.44	15.56
		64	13.37	10.44	11.85	10.36	14.74
	16	32	8.44	5.98	11.98	10.91	12.97
		64	12.54	9.51	11.02	10.16	12.37
	32	64	6.80	4.82*	11.95	11.43	12.83

Note: * denotes the statistic test can control the type I error.

From **Table 7.4** indicates that

- **significance level at 0.01 (one-sided test)**

The Q_{WLS} and Z^2_{WLS} tests can control the Type I error when the mean sample size in the treatment group is very small and the mean sample size in the control group is small to large.

- **significance level at 0.05 (one-sided test)**

The Q_{WLS} test can control the Type I error when the mean sample size in each treatment group is very small and large to very large. Also, the Z^2_{WLS} test can control the Type I error when the mean sample size in each group is very small and moderate to large.

- **significance level at 0.10 (one-sided test)**

There is only one case of the mean treatment group size where the statistic test can control the Type I error. That is, the Z^2_{V} test can control the Type I error when the mean sample size in each treatment group is moderate and very large.

Table 7.4 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when number of center (K) = 32 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α/2	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.005	4	8	0.47*	0.49*	13.31	11.37	14.36
		16	0.42*	0.51*	11.50	9.31	13.72
		32	0.56*	0.63*	10.89	8.24	14.11
		64	0.81	0.86	10.23	7.93	14.44
	8	16	1.92	2.00	2.83	2.17	3.71
		32	3.34	3.41	2.56	1.86	3.91
		64	5.85	5.95	2.25	1.49	3.92
	16	32	2.00	2.04	2.31	1.84	2.96
		64	4.44	4.57	1.97	1.51	2.71
	32	64	1.39	1.42	2.43	2.10	2.81
0.025	4	8	1.36	2.91	23.69	20.43	26.59
		16	1.43	3.07*	20.45	16.98	25.87
		32	1.84*	3.12*	19.10	15.17	26.11
		64	2.23*	3.48	18.40	14.12	27.26
	8	16	5.36	4.66	7.15	5.60	9.17
		32	8.56	7.34	6.06	4.31	9.24
		64	12.53	10.96	5.12	3.76	9.69
	16	32	6.34	5.29	5.54	4.55	7.27
		64	10.92	9.50	4.65	3.62	7.14
	32	64	5.02	4.24	6.10	5.33	7.20
0.05	4	8	2.35	8.23	29.59	25.78	34.36
		16	2.64	7.49	26.49	21.79	33.12
		32	3.20	7.84	24.36	19.33	33.84
		64	3.93	8.06	23.57	17.89	34.95
	8	16	8.60	7.15	10.23	8.33	13.15
		32	12.90	10.70	8.64	6.49	13.37
		64	17.63	14.90	7.33	5.86	13.78
	16	32	10.12	8.12	8.11	6.93	10.31
		64	16.22	13.20	6.92	5.58*	10.24
	32	64	8.32	6.64	8.93	8.16	10.65

Note: * denotes the statistic test can control the type I error.

From **Table 7.5** indicates that

- **significance level at 0.01 (one-sided test)**

Two statistic tests can control the Type I error. That is, the Q_{WLS} test can control the Type I error when the mean sample size in each group is very small and moderate to very large. Also, the Z_{WLS}^2 test can control the Type I error when the mean sample size in each group is very small and small to large.

- **significance level at 0.05 (one-sided test)**

The Q_{WLS} and Z_V^2 tests can control the Type I error. That is, when the mean sample size in each group is very small and very large, then the Q_{WLS} test can control the Type I error. Also, the Z_V^2 test can control the Type I error when the mean sample size in the treatment group is small to moderate and mean sample size in another group is large to very large.

- **significance level at 0.10 (one-sided test)**

The $Z_{WLS,R}^2$ and Z_V^2 tests can control the Type I error. That is, when the mean sample size in each group is moderate and very large, then the $Z_{WLS,R}^2$ test can control the Type I error. Also, the Z_V^2 test can control the Type I error when the mean sample size in the treatment group is small to moderate and mean sample size in another group is large to very large.

Table 7.5 Comparison of the actual Type I error of the five statistical tests at 0.01, 0.05, 0.10 (one-sided) significance levels when number of center (K) = 48 and mean treatment group size are unequal ($n_{i1} \neq n_{i2}$).

α/2	n_{i1}	n_{i2}	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
0.005	4	8	0.20	0.65*	14.88	12.07	18.11
		16	0.25*	0.60*	11.98	8.88	16.70
		32	0.38*	0.59*	10.88	7.82	17.87
		64	0.59*	0.86	9.52	6.30	17.97
	8	16	1.98	2.04	2.47	1.63	3.65
		32	3.67	3.75	1.86	1.05	3.96
		64	7.02	7.08	1.36	0.80	4.35
	16	32	2.22	2.33	1.68	1.35	2.57
		64	5.16	5.26	1.49	0.96	2.61
	32	64	1.38	1.41	1.70	1.52	1.93
0.025	4	8	0.80	6.09	25.00	20.12	30.56
		16	0.85	4.75	21.46	16.13	29.48
		32	1.29	5.07	19.28	13.14	31.17
		64	1.90*	5.16	17.17	11.13	31.25
	8	16	5.84	5.21	5.96	4.27	9.12
		32	9.50	8.24	4.29	2.78*	9.14
		64	14.72	12.74	3.68	2.17*	9.34
	16	32	6.91	5.78	4.24	3.20*	6.39
		64	12.34	10.68	3.53	2.56*	6.29
	32	64	5.39	4.43	4.17	3.64	5.11
0.05	4	8	1.38	12.59	30.55	25.16	38.61
		16	1.70	10.58	26.95	20.69	37.38
		32	2.47	10.49	24.70	16.66	38.96
		64	3.37	10.56	22.18	14.17	38.65
	8	16	9.20	8.10	8.88	6.72	12.68
		32	14.37	11.82	6.94	5.22*	12.82
		64	20.65	17.17	6.22	5.28*	13.19
	16	32	11.45	9.14	6.79	5.47*	9.37
		64	17.92	14.72	5.75*	5.09*	9.40
	32	64	8.96	7.08	6.38	5.85	8.07

Note: * denotes the statistic test can control the type I error.

4.2 Comparing the Power of test

The study compare the power of the test statistics for a random effects alternative model; that is the risk difference in each center (τ_i) generated from the Uniform distribution with mean = 0.1 and $\text{var}(\tau_i) = (2c)^2/12$, and where c = .02, .04, .06, and .08. The cases for c = 0 correspond to those for the homogeneity of risk difference under H_o. Hypothesis testing considered at 0.01, 0.05, and 0.10 significance levels and when the number of centers (K) equal 8, 16, 32, and 48 and the mean treatment group size equal 4, 8, 16, 32, and 64. Furthermore, this study considers when mean sample sizes in each group are equal ($n_{i1} = n_{i2}$) and unequal cases ($n_{i1} \neq n_{i2}$).

Data are manipulated for better comprehension by assigning each value of percent of simulations about calculated the power of tests by ranking. One will be assigned to the maximum value of the power of tests and 5 will be assigned to the minimum value of the power. According to the ranking, the method that gives the most power has the highest power.

This study considers the power of tests of five statistical tests by plotting graph for better comprehension and easier understanding. The y-axis is the power of tests and x-axis is the mean treatment group size in any size represented by numbers 1 to 5 when mean treatment group size is equal and number 1 to 10 when mean treatment group size is unequal. That is, when mean sample sizes have the same size in each treatment group, therefore

Number **1** means the mean treatment groups sizes are very small (n_{ij}=4)

Number **2** means the mean treatment groups sizes are small (n_{ij}=8)

Number **3** means the mean treatment groups sizes are moderate (n_{ij}=16)

Number **4** means the mean treatment groups sizes are large (n_{ij}=32)

Number **5** means the mean treatment group sizes are (n_{ij}=64).

When mean sample sizes have different size in each treatment group therefore

Number **1** means n_{i1} = 4, n_{i2} = 8

Number **2** means n_{i1} = 4, n_{i2} = 16

Number **3** means n_{i1} = 4, n_{i2} = 32

Number **4** means n_{i1} = 4, n_{i2} = 64

Number **5** means n_{i1} = 8, n_{i2} = 16

Number **6** means n_{i1} = 8, n_{i2} = 32

Number **7** means n_{i1} = 8, n_{i2} = 64

Number **8** means n_{i1} = 16, n_{i2} = 32

Number **9** means n_{i1} = 16, n_{i2} = 64

Number **10** means n_{i1} = 32, n_{i2} = 64.

Note that, the meaning of the number on the x-axis (both mean treatment group sizes are equal and unequal) will be used in every graph of this study.

The simulation of this study was at 0.01, 0.05, 0.10 significance level, but because the results of 0.01, 0.05, and 0.10 significance level are similar, so that this chapter presents the result of the 0.05 significance level. The results of 0.01 and 0.10 significance level are presented in Appendix B.

4.2.1 Dropping case where $\hat{\omega}_i = 0$

- Comparing the power of test when significance level consider at two-sided

The results in this step simulate from deleting case $\hat{\omega}_i$ = 0 method. For this step, this study will present the significance level at two-sided test.

Table 8.1 indicates that mean ranking of power of test, with its respective standard error (S.E.) of mean in next column, from the Q_{WLS} , Z^2_{WLS} , $Z^2_{WLS,R}$, Z^2_V , and Z^2_K statistic tests at the 0.05 significance level for each number of centers where c = .02, .04, .06, .08 in any number of centers and the mean sample size in each treatment group is equal. Table 8.2 indicates that the mean sample size in each treatment group is unequal**.**

According to Table 8.1 and Figure 1, when mean treatment group sizes are equal and c is less than 0.04 then the Q_{WLS} test has the highest power in every K, but n_{ij} are very small to small sizes ($n_{ij} \leq 8$). However, if n_{ij} are moderate to very large ($n_{ij} \geq 16$) and K is small (K = 8) then the $Z^2_{WLS,R}$ test is the highest power. In addition, if n_{ij} are moderate to very large, ($n_{ij} \geq 16$), and K is greater than moderate, ($K \geq 16$), then the Z_K^2 test is the highest power. Moreover, if c is greater than 0.04 in every K and every n_{ij} then the Q_{WLS} test has the highest power.

Table 8.1 Mean ranking of power of tests in each number of centers, at the 0.05 significance level and $n_{i1} = n_{i2}$.

c	*K*	Q_{WLS}	Z_{WLS}^2	$Z_{WLS,R}^2$	Z_V^2	Z_K^2
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	2.8 (.74)	4.2 (.58)	1.6 (.40)	3.6 (.40)	2.8 (.58)
	16	2.8 (.74)	3.8 (.74)	2.8 (.49)	3.8 (.49)	1.8 (.49)
	32	2.0 (.63)	3.4 (.68)	3.4 (.60)	4.2 (.37)	2.0 (.45)
	48	1.2 (.20)	2.8 (.58)	4.2 (.37)	4.2 (.37)	2.6 (.51)
0.04	8	3.4 (.60)	4.4 (.60)	1.8 (.80)	3.2 (.20)	2.2 (.20)
	16	2.0 (.63)	3.8 (.74)	3.2 (.49)	4.0 (.45)	2.0 (.45)
	32	1.0 (.00)	2.0 (.00)	4.2 (.20)	4.8 (.20)	3.0 (.00)
	48	1.0 (.00)	2.0 (.00)	4.2 (.20)	4.4 (.40)	3.4 (.40)
0.06	8	1.4 (.25)	3.4 (.68)	2.0 (.45)	4.4 (.25)	3.8 (.37)
	16	1.0 (.00)	3.0 (.63)	3.6 (.25)	4.8 (.20)	2.6 (.25)
	32	1.2 (.20)	1.8 (.20)	4.0 (.32)	4.6 (.25)	3.4 (.40)
	48	1.0 (.00)	2.0 (.00)	4.2 (.20)	4.0 (.45)	3.8 (.49)
0.08	8	1.0 (.00)	3.0 (.55)	2.4 (.25)	4.6 (.25)	4.0 (.32)
	16	1.0 (.00)	2.0 (.00)	4.0 (.00)	5.0 (.00)	3.0 (.00)
	32	1.0 (.00)	2.0 (.00)	4.0 (.32)	4.6 (.25)	3.4 (.40)
	48	1.4 (.25)	3.2 (.20)	4.8 (.20)	4.0 (.32)	1.6 (.25)

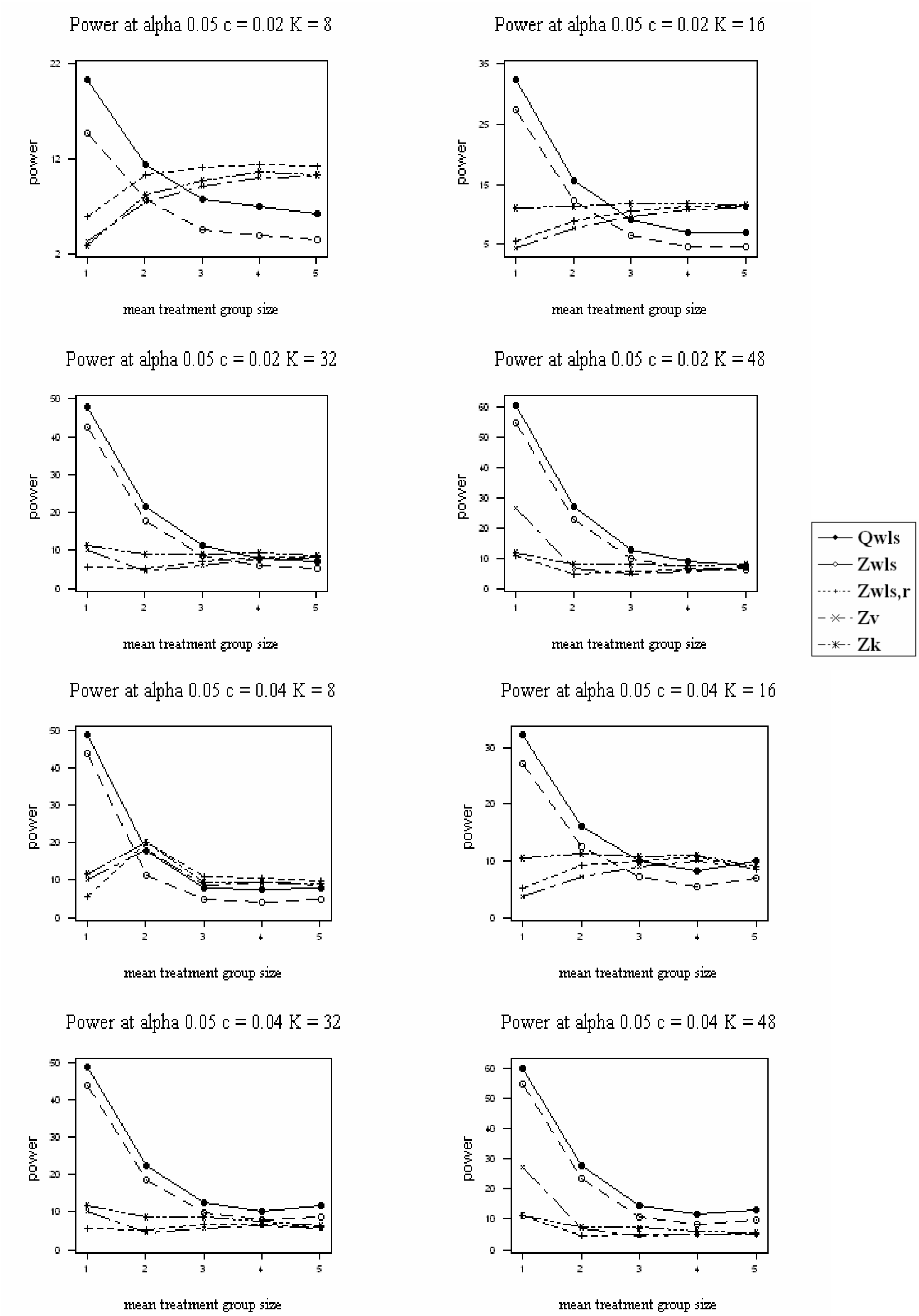


Figure 1 Comparison of the power of five statistical tests by *c*, level of significance and number of center (*K*) where mean treatment group sizes are equal.

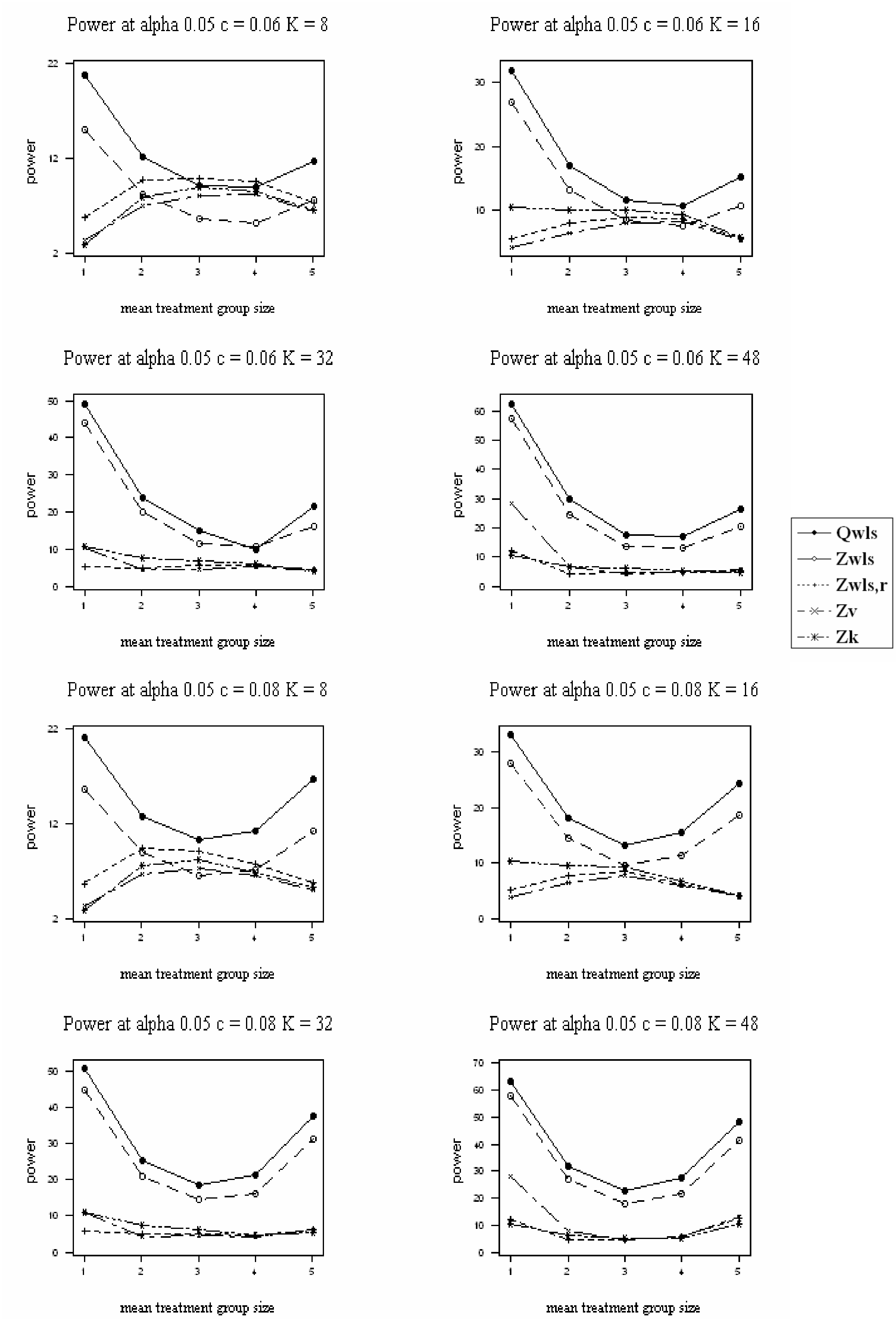


Figure 1 Comparison of the power of five statistical tests by *c*, level of significance and number of center (*K*) where mean treatment group sizes are equal. (Continued)

Table 8.2 shows that at c = 0.02, 0.04 when the number of center is small and moderate, (K = 8, 16), and both cases when n_{ij} are large to very large, ($n_{ij} \geq$ 32), then the Z_K^2 test has the highest power. Moreover, the Q_{WLS} test has the highest power, that is, at c = 0.02, 0.04 when the number of center is small and moderate, (K = 8, 16), but mean sample sizes in each treatment group is very small to large, ($n_{ij} \leq 32$), and at c = 0.06, 0.08 and all of K and $n_{ij.}$ This can be clearly demonstrated by Figure 2.

Table 8.2 Mean ranking of power of tests in each number of centers, at the 0.05 significance level and $n_{i1} \neq n_{i2}$.

c	K	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.7 (.40)	3.0 (.45)	2.4 (.28)	4.5 (.27)	3.5 (.27)
	16	1.4 (.31)	2.6 (.40)	3.8 (.25)	4.2 (.29)	3.0 (.42)
	32	1.0 (.00)	2.2 (.20)	3.8 (.13)	4.3 (.30)	3.7 (.37)
	48	1.0 (.00)	2.0 (.00)	3.8 (.20)	4.0 (.26)	4.2 (.33)
0.04	8	1.4 (.31)	3.2 (.49)	3.1 (.41)	4.5 (.22)	2.8 (.13)
	16	1.0 (.00)	2.5 (.34)	3.9 (.19)	4.4 (.26)	3.2 (.33)
	32	1.0 (.00)	2.0 (.00)	3.8 (.13)	4.3 (.33)	3.9 (.32)
	48	1.3 (.25)	2.1 (.16)	3.6 (.34)	3.7 (.34)	4.3 (.30)
0.06	8	1.0 (.00)	2.6 (.40)	2.9 (.18)	4.8 (.13)	3.7 (.15)
	16	1.0 (.00)	2.0 (.00)	4.1 (.10)	4.4 (.31)	3.5 (.27)
	32	1.0 (.00)	2.0 (.00)	3.8 (.13)	4.3 (.33)	3.9 (.32)
	48	1.1 (.07)	1.9 (.07)	3.7 (.21)	3.9 (.28)	4.4 (.27)
0.08	8	1.0 (.00)	2.0 (.00)	3.1 (.10)	5.0 (.00)	3.9 (.10)
	16	1.0 (.00)	2.0 (.00)	4.1 (.10)	4.4 (.31)	3.5 (.27)
	32	1.0 (.00)	2.0 (.00)	3.6 (.16)	4.3 (.30)	4.1 (.28)
	48	1.1 (.07)	1.9 (.07)	3.4 (.16)	3.6 (.16)	5.0 (.00)

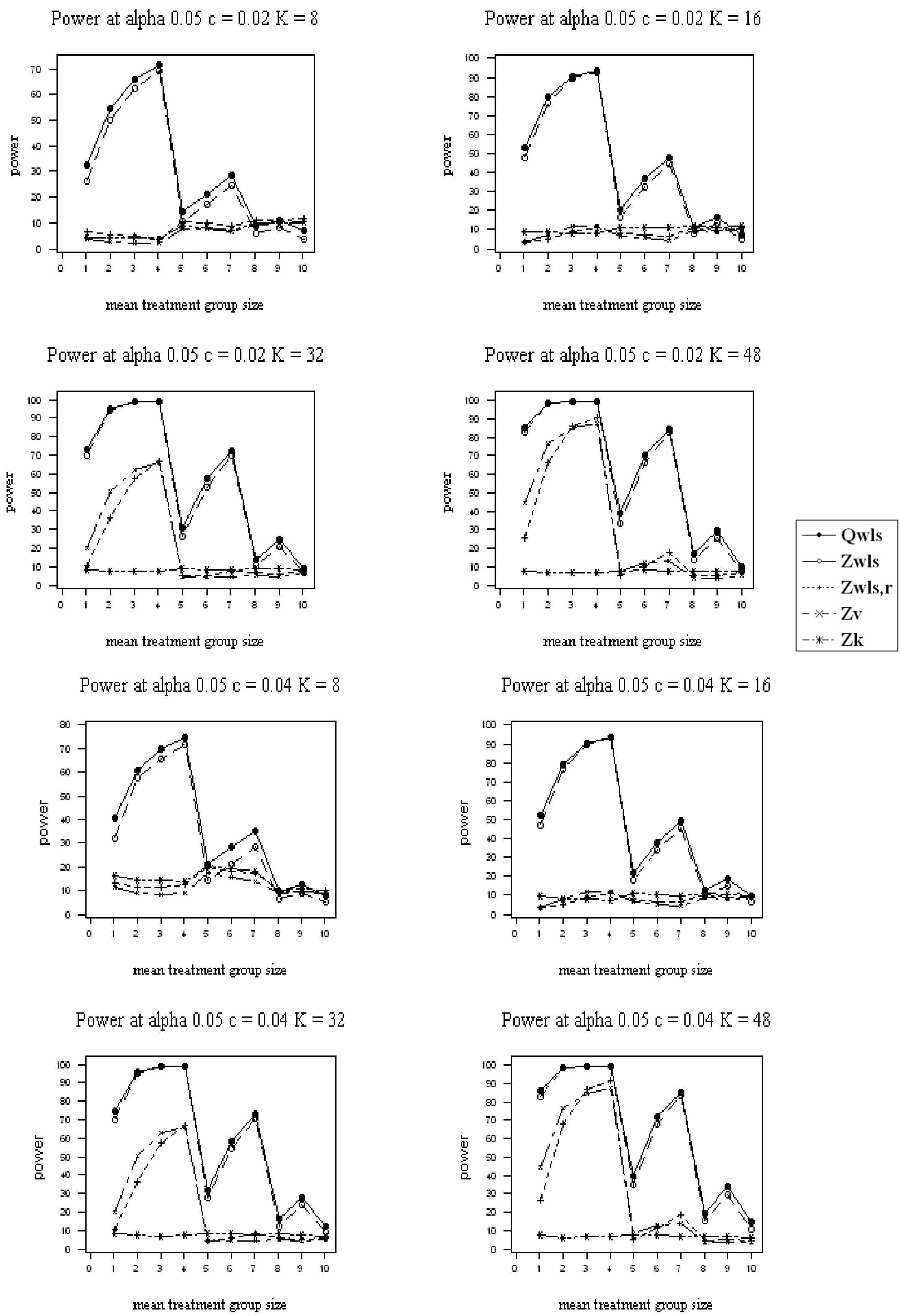


Figure 2 Comparison of the power of five statistical tests by *c*, level of significance and number of center (*K*) where mean treatment group sizes are unequal.

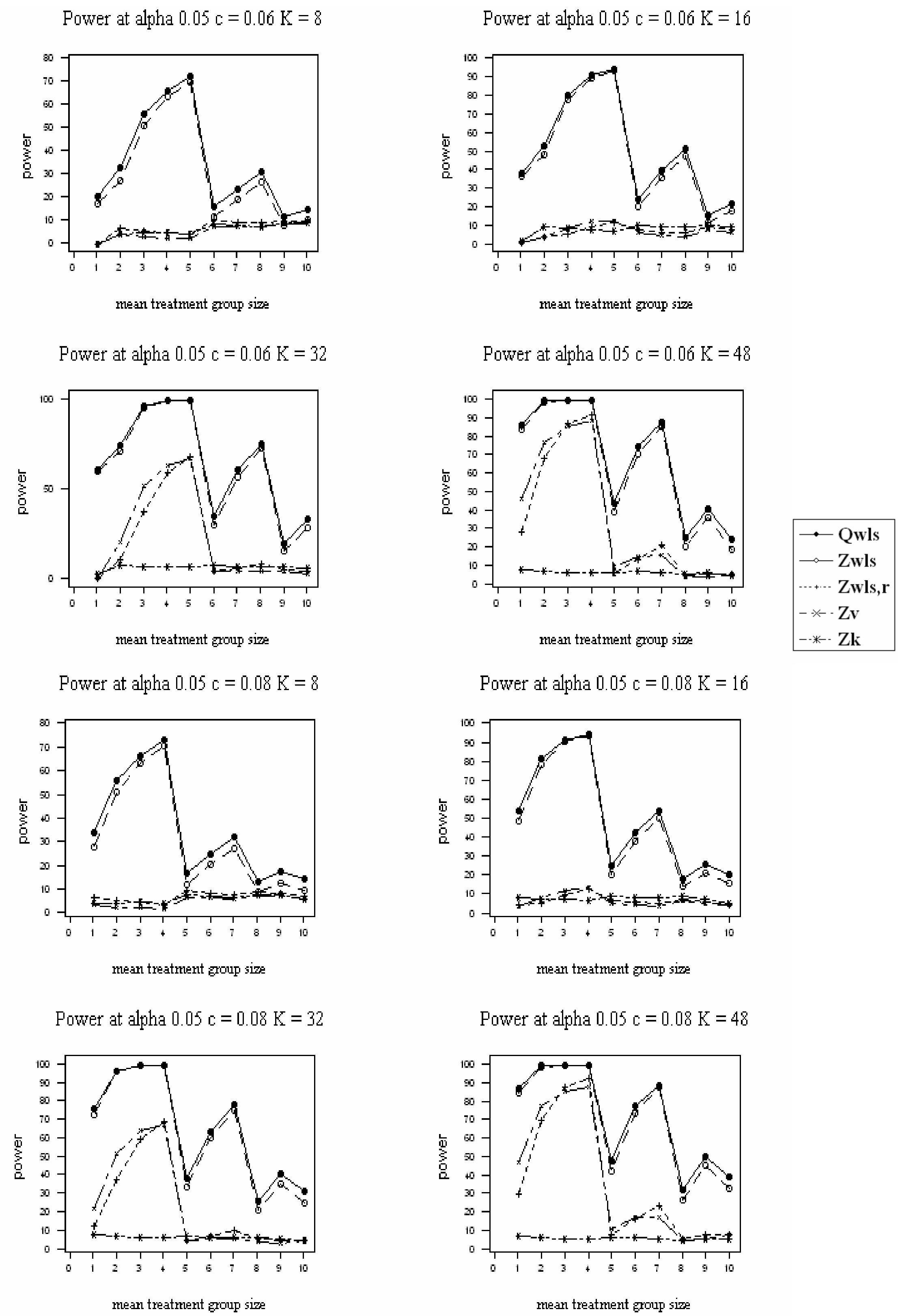


Figure 2 Comparison of the power of five statistical tests by *c*, level of significance and number of center (*K*) where mean treatment group sizes are unequal. (Continued)

- Comparing the power of test when significance level considered for one-sided test.

The results in this step still simulate using the dropping case $\hat{\omega}_i = 0$ method. However, hypothesis testing will consider the significance level for the one-sided test where any conditions are similar in significance to the two-sided test.

Table 9.1 indicates the mean ranking power of test, with its respective standard error (S.E.) of mean in next column, from the Q_{WLS}, Z^2_{WLS}, $Z^2_{WLS,R}$, Z^2_V, and Z^2_K statistic tests where c = .02, .04, .06, .08 in any number of centers where the mean treatment group size is equal, and Table 9.2 indicates when the mean treatment group size is unequal at the 0.05 significance level. Furthermore, Figure 3 and 4 present a comparison of the power of five statistical tests on the y-axis showing the power in each test and x-axis shows the mean treatment group sizes from Number 1 to Number 5 and Number 1 to Number 10 when mean treatment group sizes are equal and unequal, respectively. The meaning of the number on the x-axis has the same meaning of the numbers in Figure 1 and 2.

The results from Table 9.1 and Figure 3 that simulate by dropping case and considering the one-sided test of significance level are similar to the results from considering the two-sided test of significance level. That is, the Q_{WLS} test has highest power for almost every case of c, number of centers and mean treatment group size. However, when c = 0.02 and the number of centers is moderate to large, (K = 16, 32), where both mean treatment group sizes are moderate to very large, ($n_{ij} \geq 16$), then the Z^2_K test has the highest power. When c and the number of centers (K) increase then the trend of the power of Q_{WLS} and Z^2_{WLS} tests increase, too; on the other hand, the trend of the power of other tests ($Z^2_{WLS,R}, Z^2_V, Z^2_K$) decrease to nearly zero.

Table 9.1 Mean ranking of power of tests in each number of centers, at the 0.05 significance level (one-sided) and $n_{i1} = n_{i2}$.

c	***K***	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.8 (.74)	3.8 (.74)	2.8 (.49)	3.8 (.49)	1.8 (.49)
	32	2.2 (.74)	3.4 (.68)	3.4 (.60)	4.0 (.45)	2.0 (.45)
	48	1.4 (.40)	2.8 (.58)	4.0 (.55)	4.2 (.37)	2.6 (.51)
0.04	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.2 (.58)	3.8 (.74)	2.6 (.60)	4.2 (.37)	2.2 (.49)
	32	1.0 (.00)	2.2 (.20)	4.2 (.20)	4.8 (.20)	2.8 (.20)
	48	1.0 (.00)	2.0 (.00)	4.4 (.25)	4.4 (.40)	3.2 (.20)
0.06	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.0 (.00)	3.0 (.63)	3.4 (.40)	4.8 (.20)	2.8 (.20)
	32	1.0 (.00)	2.0 (.00)	4.2 (.20)	4.8 (.20)	3.0 (.00)
	48	1.0 (.00)	2.6 (.60)	4.2 (.37)	3.8 (.58)	3.4 (.25)
0.08	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.0 (.00)	2.0 (.00)	3.8 (.20)	5.0 (.00)	3.2 (.20)
	32	1.0 (.00)	2.0 (.00)	4.0 (.32)	3.6 (.25)	3.4 (.40)
	48	1.0 (.00)	2.0 (.00)	4.0 (.45)	4.0 (.32)	4.0 (.45)

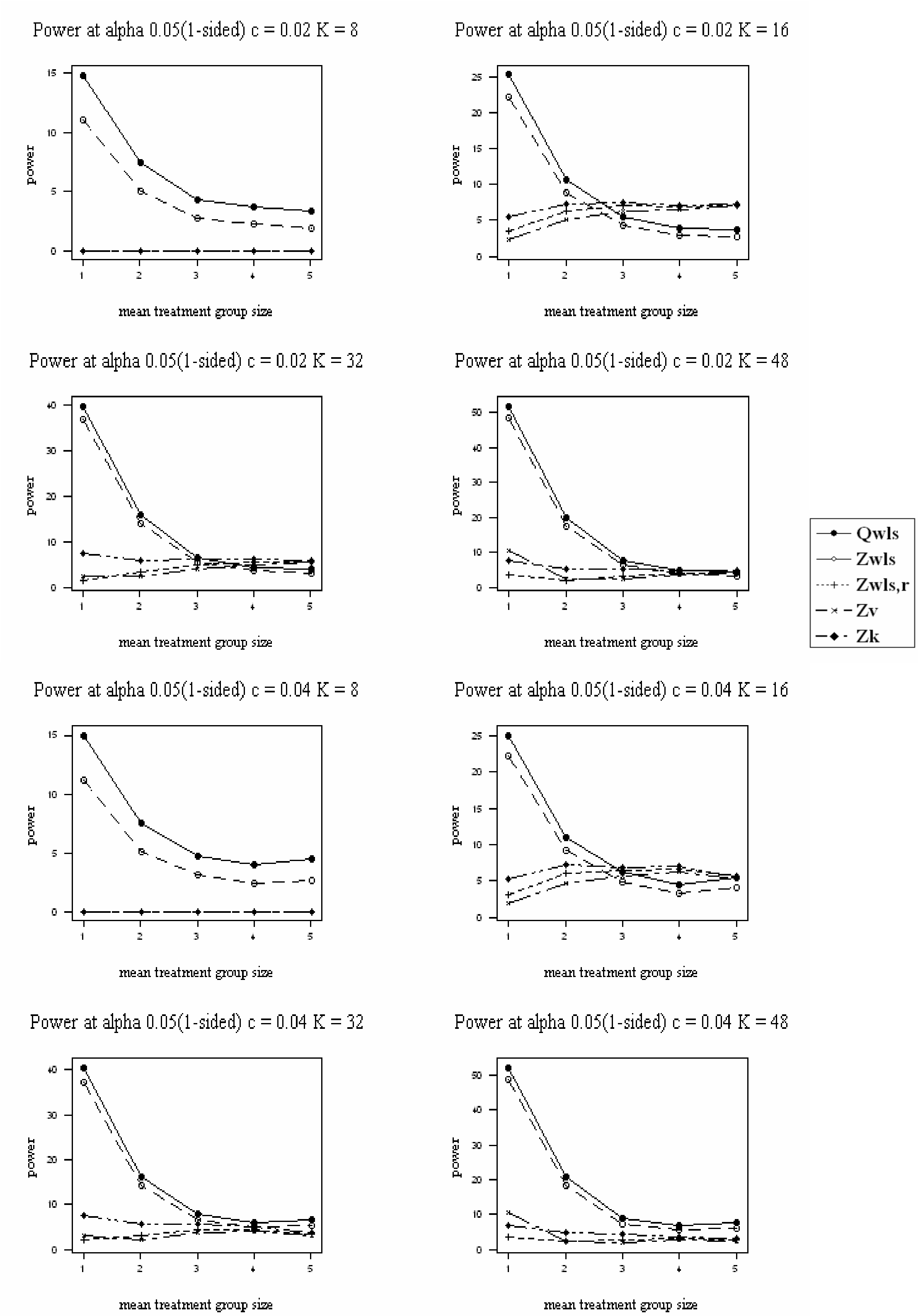


Figure 3 Comparison of the power of five statistical tests by *c*, level of significance (one-sided) and number of center where mean treatment group sizes are equal.

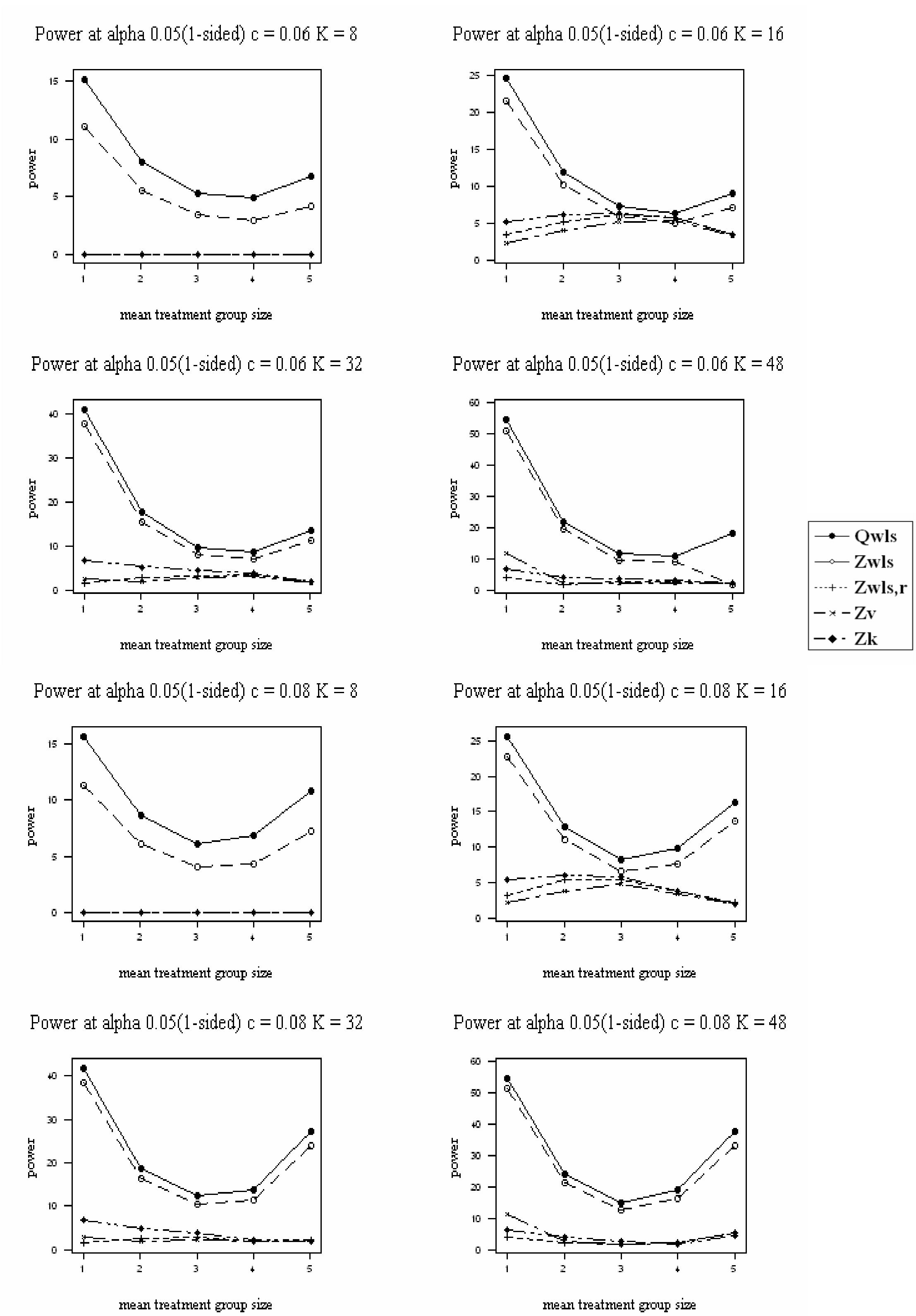


Figure 3 Comparison of the power of five statistical tests by *c*, level of significance (one-sided) and number of center where mean treatment group sizes are equal. (Continued)

Table 9.2 indicates at the one-sided of the 0.05 significance level when the mean sample size in each treatment group is unequal. The results from Table 13.2 and Figure 4 are similar to testing of the two-sided test. That is, when c = 0.06, and 0.08, for every size of K and every size of mean treatment group, the Q_{WLS} test has the first ranking (the highest power). For the case of c = 0.02 and 0.04, when K = 16 and 32, where both n_{ij} are large to very large, ($n_{ij} \geq 32$), then the Z_K^2 test has the first ranking (the highest power); furthermore, the Q_{WLS} test has the highest power.

Table 9.2 Mean ranking of power of tests in each number of centers, at the 0.05 significance level (one-sided) and $n_{i1} \neq n_{i2}$.

c	***K***	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.5 (.34)	2.6 (.40)	3.8 (.33)	4.5 (.22)	2.6 (.27)
	32	1.1 (.10)	2.3 (.30)	4.0 (.15)	4.1 (.32)	3.5 (.40)
	48	1.1 (.10)	2.1 (.12)	4.0 (.21)	4.0 (.26)	3.9 (.38)
0.04	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.2 (.20)	2.5 (.32)	3.9 (.26)	4.7 (.15)	2.7 (.21)
	32	1.0 (.00)	2.0 (.00)	4.1 (.10)	4.2 (.33)	3.7 (.30)
	48	1.1 (.07)	1.9 (.07)	3.9 (.18)	4.1 (.28)	4.0 (.33)
0.06	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.0 (.00)	2.0 (.00)	4.2 (.13)	4.8 (.13)	3.0 (.00)
	32	1.0 (.00)	2.0 (.00)	4.1 (.10)	4.2 (.33)	3.7 (.30)
	48	1.1 (.07)	1.9 (.07)	3.9 (.18)	3.9 (.28)	4.2 (.33)
0.08	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.0 (.00)	2.0 (.00)	4.3 (.15)	4.7 (.15)	3.0 (.00)
	32	1.0 (.00)	2.0 (.00)	4.1 (.10)	4.2 (.33)	3.7 (.30)
	48	1.1 (.07)	1.9 (.07)	3.6 (.22)	3.9 (.23)	4.5 (.27)

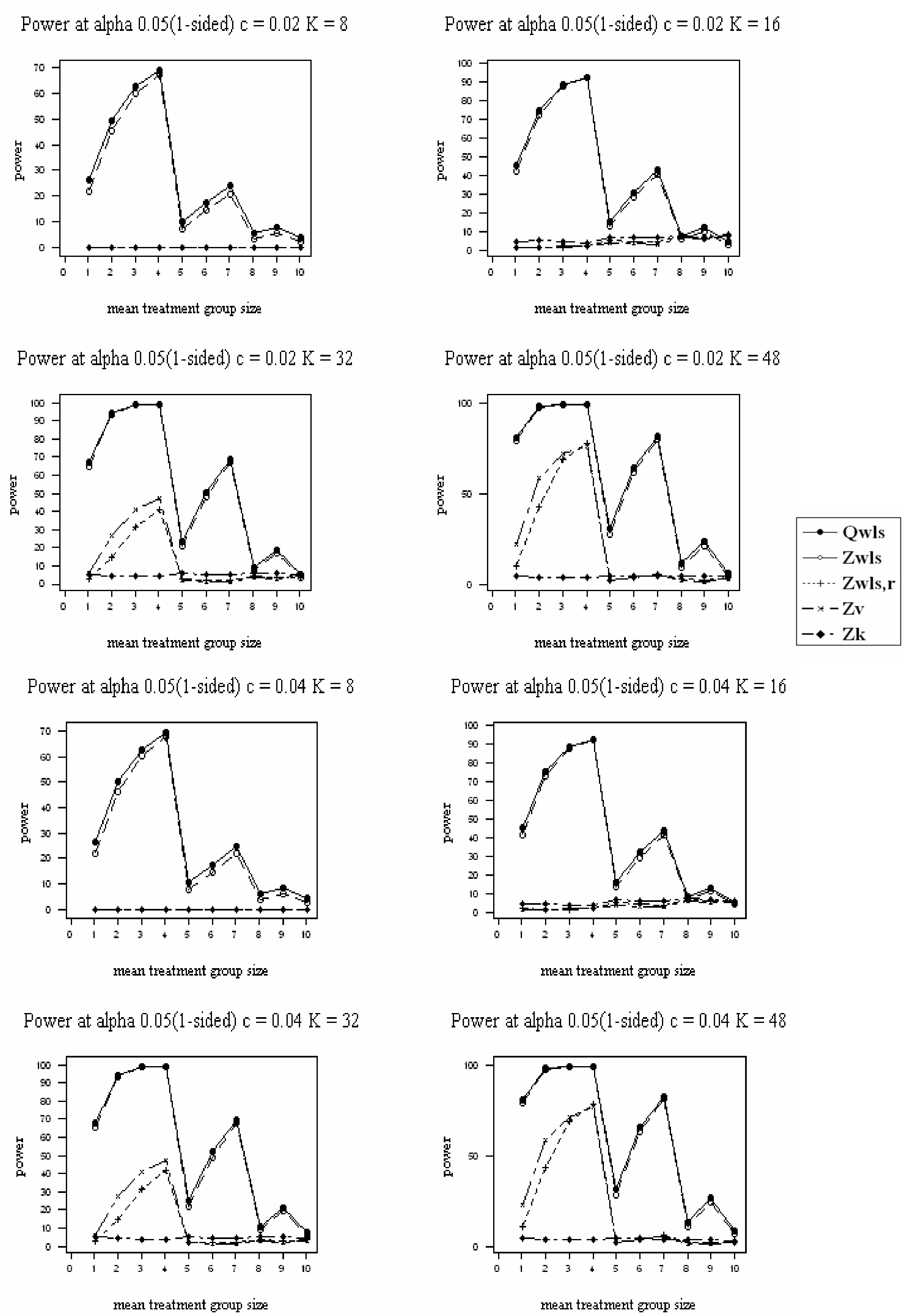


Figure 4 Comparison of the power of five statistical tests by *c*, level of significance (one-sided) and number of center (*K*) where mean treatment group sizes are unequal.

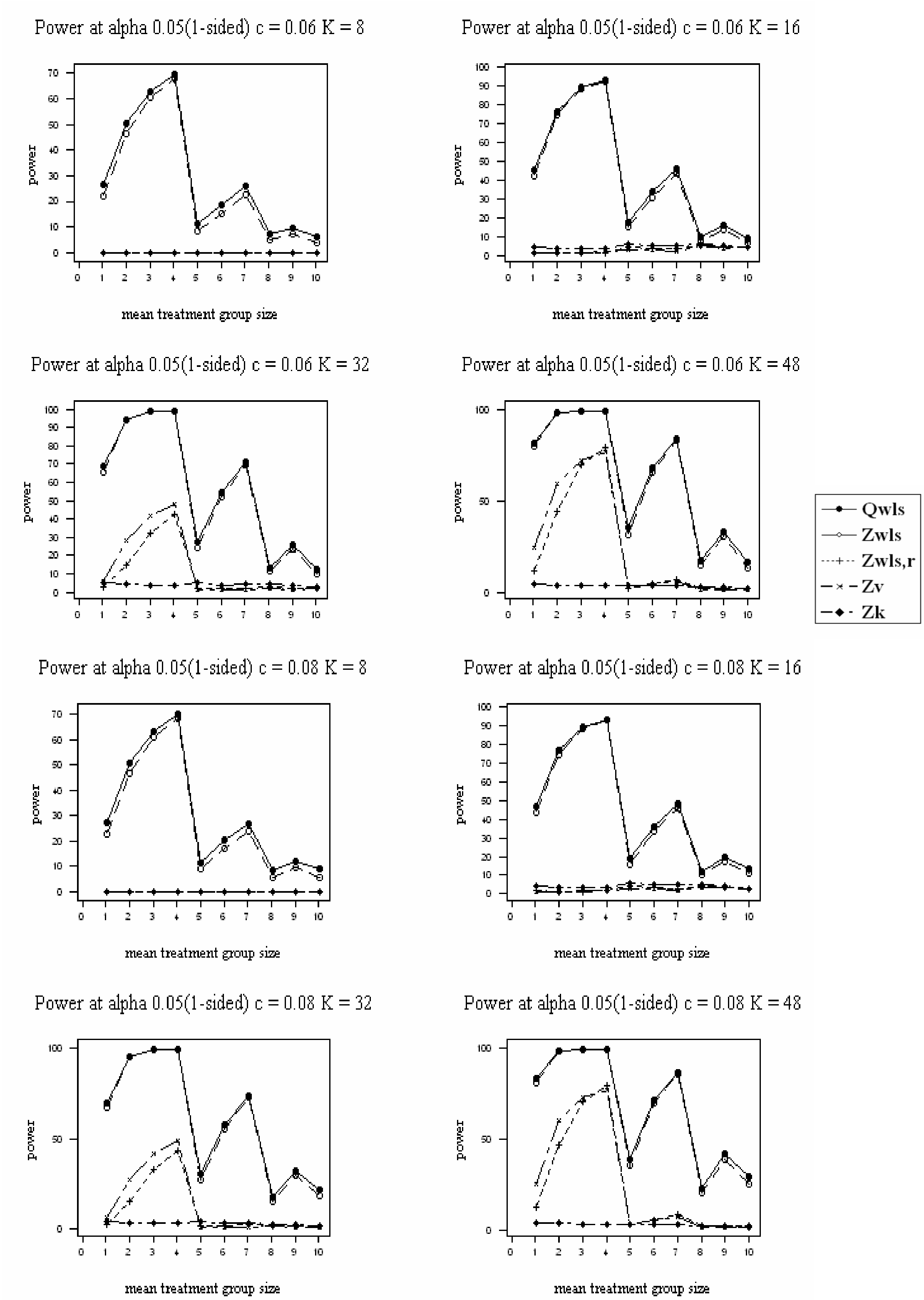


Figure 4 Comparison of the power of five statistical tests by *c*, level of significance (one-sided) and number of center (*K*) where mean treatment group sizes are unequal. (Continued)

4.2.2 Adding constant to each cell of X_{ij}, n_{ij} when $\hat{\omega}_i = 0$

- Comparing the power of test when significance level considered at two-sided test

This step shows the result from adding a constant (0.5) in each cell of X_{ij} and n_{ij} when $\hat{\omega}_i = 0$ (X_{ij} = 0 or X_{ij} = n_{ij}). The results consider when mean sample size in each treatment group are equal ($n_{i1} = n_{i2}$) and unequal ($n_{i1} \neq n_{i2}$).

Table 10.1 indicates the mean ranking of power of test, with its respective standard error (S.E.) of mean in the next column, from the Q_{WLS}, Z^2_{WLS}, $Z^2_{WLS,R}$, Z^2_V, and Z^2_K statistic tests where c = .02, .04, .06, .08 in any number of centers where the mean treatment group size is equal, and Table 10.2 shows when mean treatment group is unequal at the 0.05 significance level (0.01 and 0.10 significance level in Appendix).

The figure in this step still mean the value on the x-axis (number 1 to 5 when the mean treatment group is equal and number 1 to 10 when mean treatment group is unequal) like the figure from method of dropping cases.

From Table 14.1 and Figure 5, the results from adding a constant are different from the results of dropping case. That is,

when c = 0.02, when K = 8 and for every size of mean treatment group, then the $Z^2_{WLS,R}$ test has the highest power (first ranking), followed by Z^2_V and Z^2_K tests. When $K \geq 16$, then the Z^2_K test has the highest power, followed by $Z^2_{WLS,R}$ and Z^2_V test. However, when K = 48 and n_{ij} = 64, then the Q_{WLS} test has the highest power. When c, K and n_{ij} increase then the trend of the power of the Q_{WLS} and Z^2_{WLS} tests increase but trend of power of the $Z^2_{WLS,R}$, Z^2_V, and Z^2_K tests decrease.

When c = 0.04, 0.06, and 0.08, when K = 8 and $n_{ij} \leq 16$ then the $Z^2_{WLS,R}$ test is the first ranking (the highest power). When $K \geq 16$ and $n_{ij} \leq 16$ then the Z^2_K test is the first ranking (the highest power). Moreover, in any K but when $n_{ij} \geq 32$ then the Q_{WLS} test has the highest power.

Table 10.1 Mean ranking of power of tests in each number of centers, at the 0.05 significance level and $n_{i1} = n_{i2}$.

c	*K*	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	4.0 (.00)	5.0 (.00)	1.0 (.00)	2.8 (.20)	2.2 (.20)
	16	4.2 (.20)	4.8 (.20)	2.0 (.00)	3.0 (.00)	1.0 (.00)
	32	3.8 (.49)	4.8 (.20)	2.2 (.20)	3.2 (.20)	1.0 (.00)
	48	2.8 (.92)	3.8 (.37)	3.0 (.45)	4.0 (.45)	1.4 (.25)
0.04	8	4.0 (.00)	5.0 (.00)	1.0 (.00)	2.8 (.20)	2.2 (.20)
	16	3.4 (.68)	4.6 (.25)	3.0 (.63)	2.8 (.20)	1.2 (.20)
	32	2.4 (.87)	3.2 (.58)	3.2 (.49)	4.2 (.49)	2.0 (.45)
	48	2.4 (.87)	3.2 (.58)	3.2 (.49)	4.2 (.49)	2.0 (.45)
0.06	8	3.0 (.63)	4.6 (.40)	1.2 (.20)	3.4 (.51)	2.8 (.37)
	16	2.8 (.80)	4.2 (.58)	2.6 (.40)	3.8 (.37)	1.6 (.40)
	32	2.4 (.87)	3.0 (.63)	3.0 (.45)	4.0 (.45)	2.6 (.75)
	48	2.4 (.87)	3.0 (.63)	3.2 (.49)	3.8 (.49)	2.6 (.75)
0.08	8	2.4 (.68)	4.0 (.63)	1.6 (.40)	3.8 (.58)	3.2 (.37)
	16	2.4 (.87)	3.6 (.68)	3.0 (.45)	4.0 (.45)	2.0 (.45)
	32	2.4 (.87)	3.0 (.63)	3.0 (.45)	4.0 (.45)	2.6 (.75)
	48	2.2 (.80)	2.8 (.49)	3.2 (.49)	3.8 (.49)	3.0 (.89)

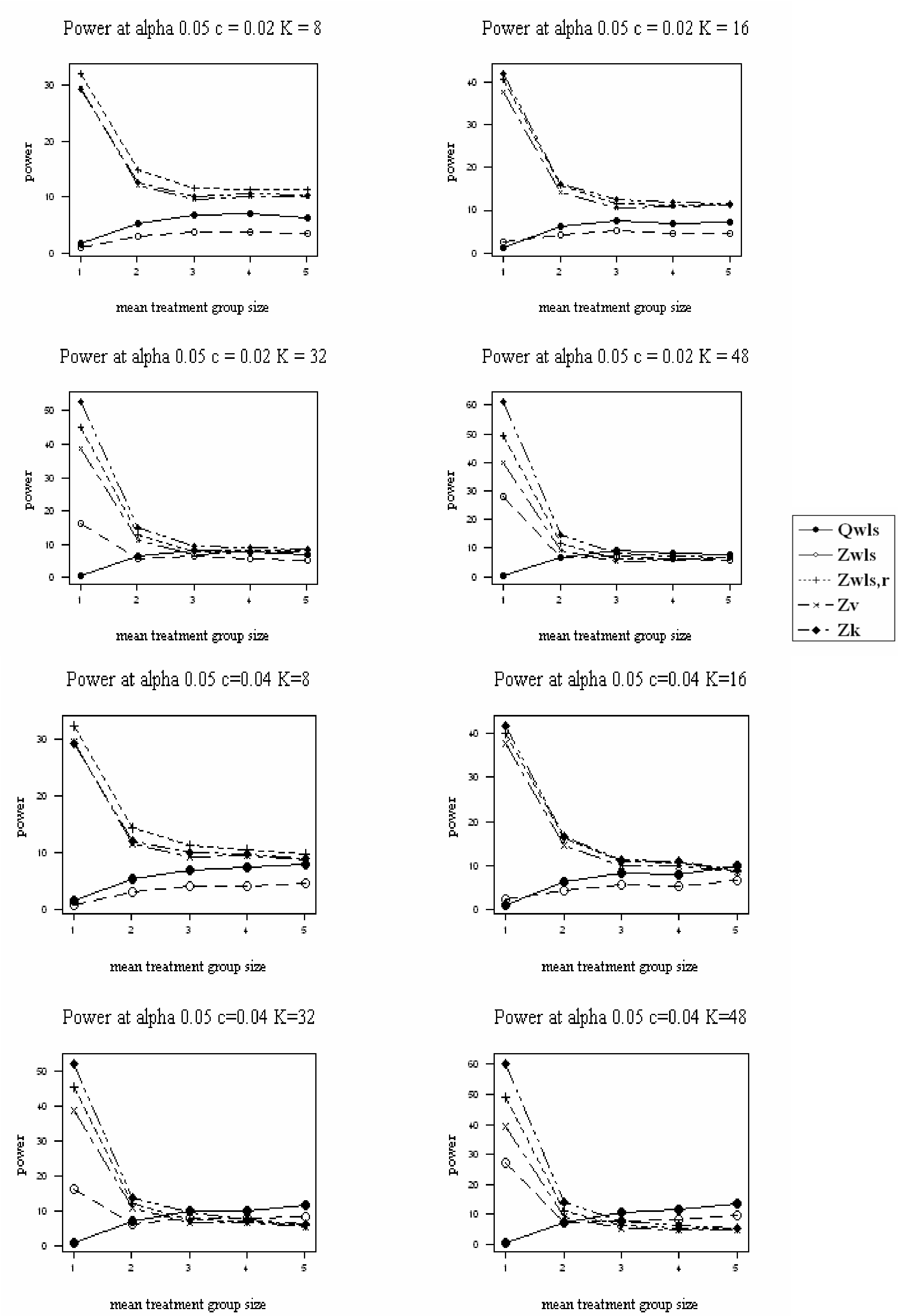


Figure 5 Comparison of the power of five statistical tests by *c*, level of significance and number of center (*K*) where mean treatment group sizes are equal.

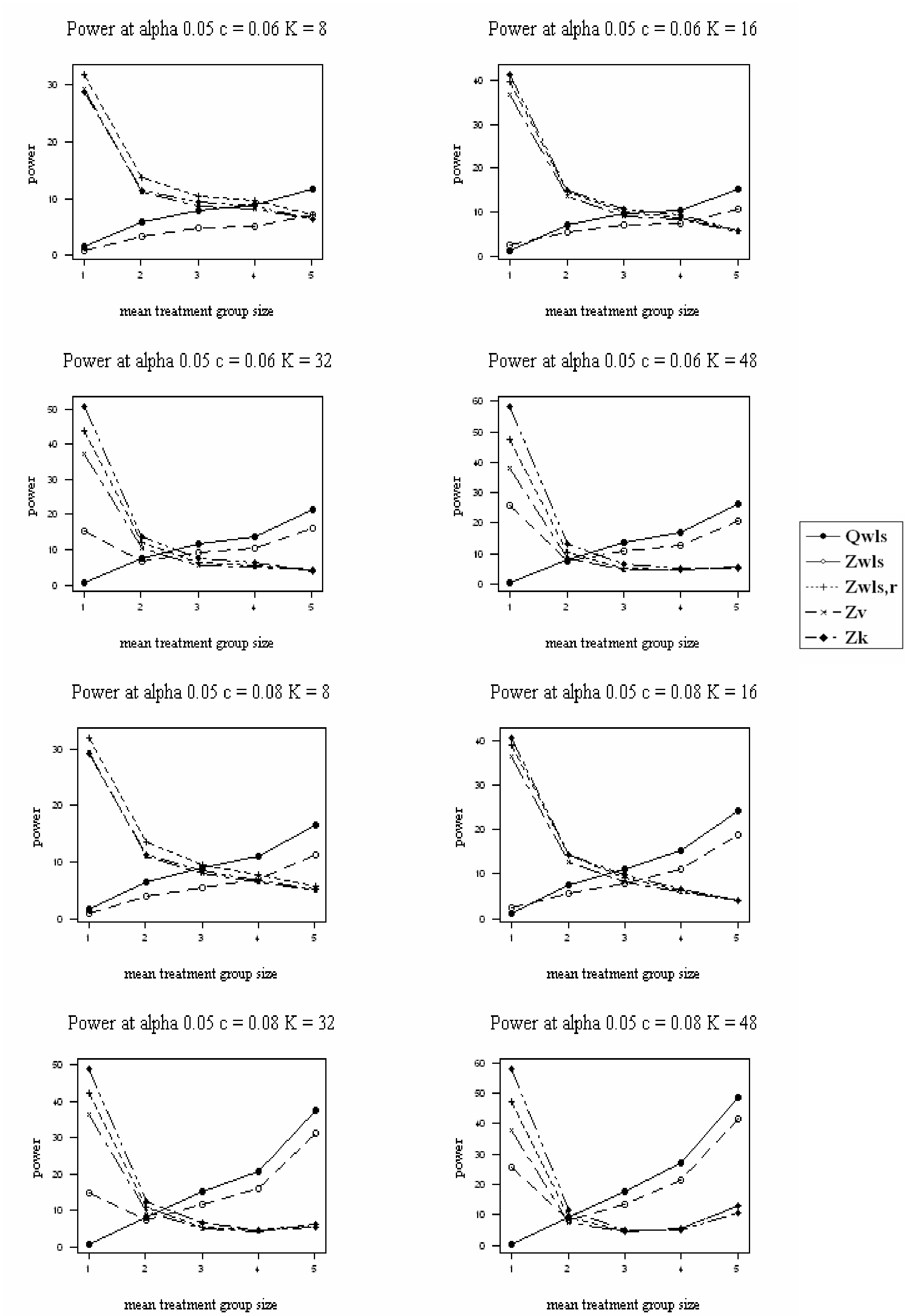


Figure 5 Comparison of the power of five statistical tests by *c*, level of significance and number of center (*K*) where mean treatment group sizes are equal. (Continued)

Next, this study will consider when the mean sample size in each treatment group is unequal shown in Table 10.2 and Figure 6. First, at K = 8 in any c and for every size of mean treatment group, the $Z^2_{WLS,R}$ test has the highest power. When c = 0.02, and 0.04 and K is moderate to large, (K = 16, 32), then the Z^2_K test has the highest power, but if both K and n_{ij} are large to very large then the Q_{WLS} test has the highest power. Especially, when c = 0.06, and 0.08 and K is greater than 16, ($K \geq$ 16), where the mean sample size in the treatment group is greater 8 and the mean sample size in another group is greater than 32, then the Q_{WLS} test is the first ranking; that is, the Q_{WLS} test has the highest power.

Table 10.2 Mean ranking of power of tests in each number of centers, at the 0.05 significance level and $n_{i1} \neq n_{i2}$.

c	***K***	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	3.8 (.20)	5.0 (.00)	1.0 (.00)	2.5 (.17)	2.7 (.21)
	16	3.5 (.34)	5.0 (.00)	2.2 (.13)	3.2 (.13)	1.1 (.10)
	32	2.8 (.63)	3.6 (.34)	2.9 (.32)	4.0 (.30)	1.7 (.26)
	48	2.7 (.63)	3.1 (.32)	3.1 (.32)	4.2 (.33)	1.9 (.32)
0.04	8	3.6 (.27)	5.0 (.00)	1.0 (.00)	2.6 (.22)	2.8 (.20)
	16	3.1 (.41)	4.7 (.24)	2.5 (.24)	3.6 (.27)	1.2 (.13)
	32	2.8 (.63)	3.2 (.36)	3.0 (.33)	4.1 (.32)	1.9 (.32)
	48	2.7 (.63)	2.9 (.31)	3.3 (.37)	4.1 (.32)	2.0 (.33)
0.06	8	3.0 (.42)	5.0 (.00)	1.2 (.13)	2.7 (.30)	3.1 (.10)
	16	2.6 (.48)	4.2 (.42)	2.8 (.29)	3.8 (.29)	1.6 (.27)
	32	2.7 (.63)	3.0 (.33)	3.1 (.32)	4.2 (.33)	2.0 (.33)
	48	2.6 (.65)	2.9 (.32)	3.2 (.36)	4.1 (.32)	2.2 (.36)
0.08	8	2.6 (.48)	4.3 (.40)	1.6 (.27)	3.0 (.39)	3.5 (.22)
	16	2.5 (.50)	3.6 (.48)	3.0 (.33)	4.0 (.33)	1.9 (.32)
	32	2.6 (.65)	3.2 (.36)	3.0 (.30)	4.0 (.30)	2.2 (.42)
	48	2.6 (.65)	2.8 (.33)	3.3 (.37)	3.5 (.27)	2.8 (.55)

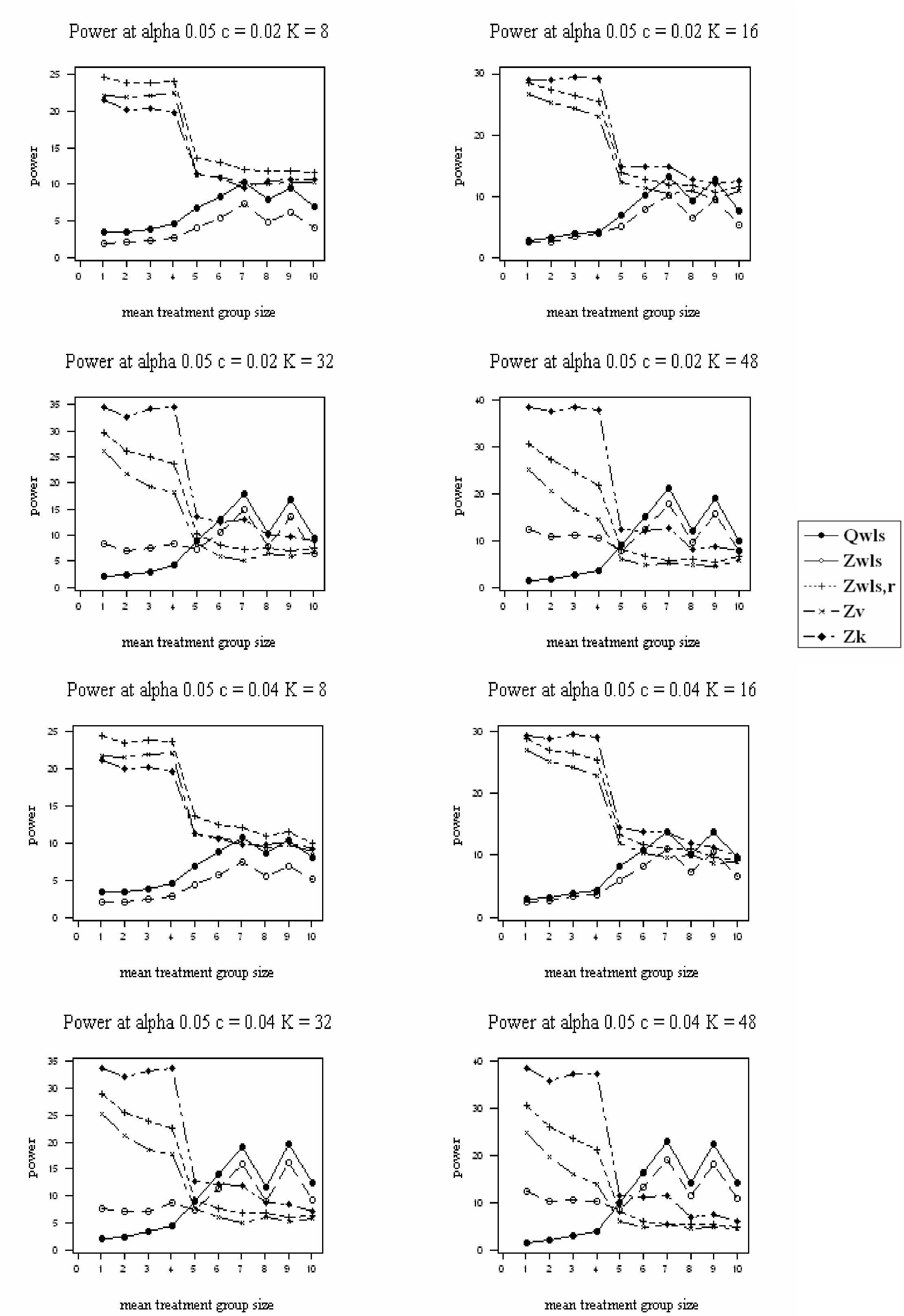


Figure 6 Comparison of the power of five statistical tests by *c*, level of significance and number of center (*K*) where mean treatment group sizes are unequal.

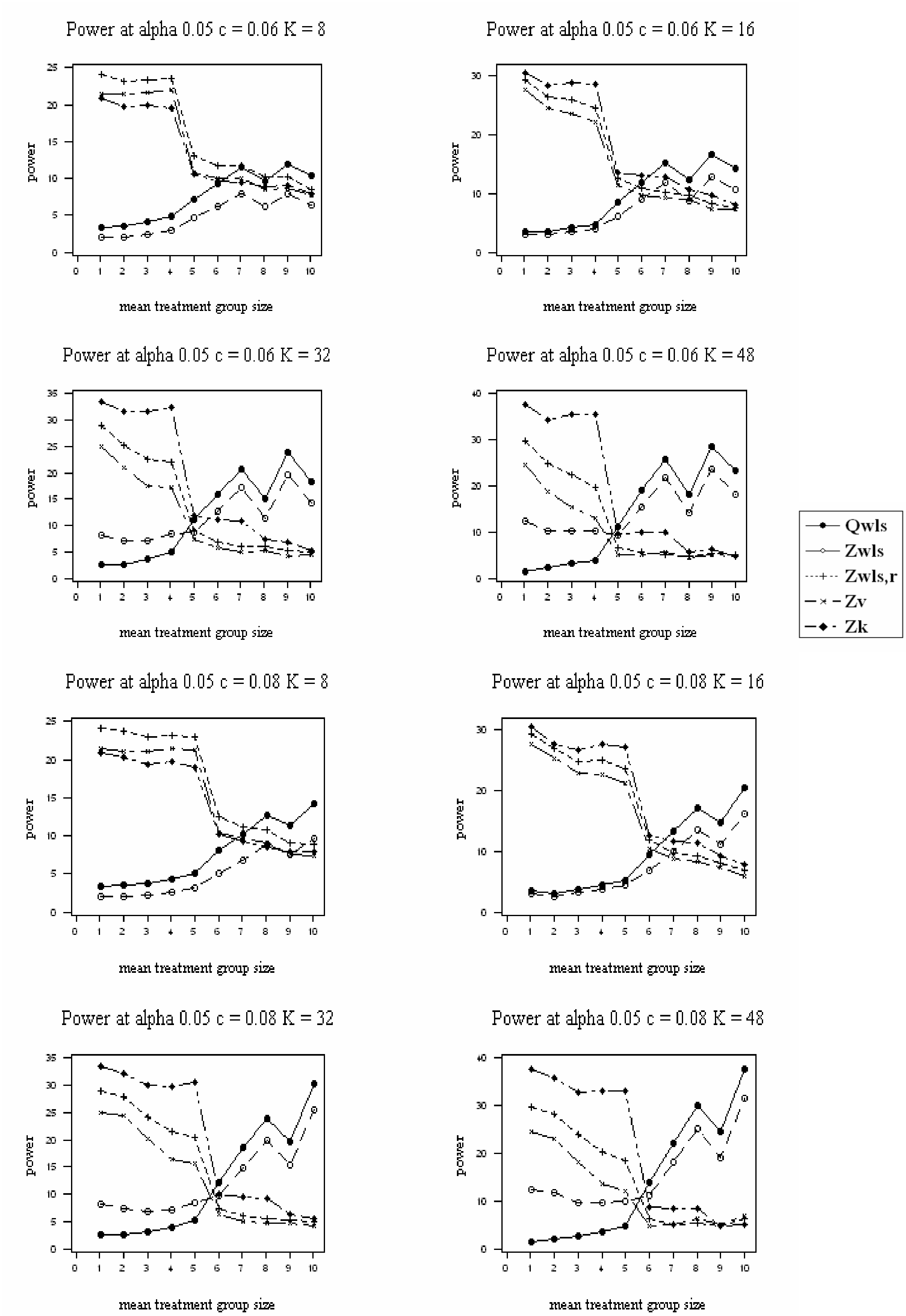


Figure 6 Comparison of the power of five statistical tests by *c*, level of significance and number of center (*K*) where mean treatment group sizes are unequal.(Continued)

- Comparing the power of test when significance level considered for the one-sided test

The results in this step still simulate by adding a constant (0.5) in each cell of X_{ij} and n_{ij} when $\hat{\omega}_i = 0$. The results consider a mean sample size in which each treatment group are equal ($n_{i1} = n_{i2}$) and unequal ($n_{i1} \neq n_{i2}$). However, hypothesis testing will consider the significance level for one-sided test where any conditions are similar in significance to the two-sided test.

Table 11.1 indicates the mean ranking of power of test, with its respective standard error (S.E.) of mean in next column, from five statistical tests where c = .02, .04, .06, .08 in any number of centers where the mean treatment group is equal and Table 11.2 shows the mean treatment group is unequal at the 0.05 significance level.

The results are shown in Table 11.1 and Figure 7 simulated by the adding a constant method and testing for one-sided of significance level. When K is small (K = 8) in any c and for every mean size of n_{ij} the Q_{WLS} test has the highest power. Moreover, when c = 0.02, and 0.04 and K is greater than 16, then the Z_K^2 test has the highest power, followed by the $Z_{WLS,R}^2$ and Z_V^2 test. However, if c = 0.06, and 0.08 and K is greater than moderate ($K \geq 16$) and both mean treatment group sizes are less than 8 ($n_{ij} \leq 8$), then the Z_K^2 test has the highest power, followed by the $Z_{WLS,R}^2$ and Z_V^2 test. However, when mean treatment group sizes are greater than 16 ($n_{ij} \geq 16$) then the Q_{WLS} test has the highest power, followed by the Z_{WLS}^2 test.

Table 11.1 Mean ranking of power of tests in each number of centers, at the 0.05 significance level (one-sided) and $n_{i1} = n_{i2}$.

c	*K*	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	4.0 (.00)	5.0 (.00)	1.8 (.20)	3.0 (.00)	1.2 (.20)
	32	4.0 (.32)	4.8 (.20)	2.0 (.00)	3.2 (.20)	1.0 (.20)
	48	3.2 (.58)	4.2 (.37)	2.6 (.40)	4.0 (.45)	1.0 (.00)
0.04	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	3.6 (.40)	5.0 (.00)	1.6 (.25)	3.2 (.20)	1.6 (.40)
	32	2.6 (.81)	3.4 (.51)	3.2 (.49)	4.2 (.49)	1.6 (.40)
	48	2.4 (.87)	3.2 (.58)	3.2 (.49)	4.2 (.49)	2.0 (.45)
0.06	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.6 (.68)	4.4 (.60)	2.2 (.49)	3.8 (.37)	2.0 (.45)
	32	2.4 (.87)	3.0 (.63)	3.2 (.49)	4.2 (.49)	2.2 (.49)
	48	2.4 (.87)	3.0 (.63)	3.2 (.49)	3.8 (.49)	2.6 (.75)
0.08	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.2 (.74)	3.6 (.68)	2.6 (.51)	4.2 (.49)	2.4 (.51)
	32	2.4 (.87)	3.0 (.63)	3.0 (.45)	4.0 (.45)	2.6 (.75)
	48	2.2 (.80)	3.0 (.63)	2.8 (.37)	4.0 (.32)	3.0 (.89)

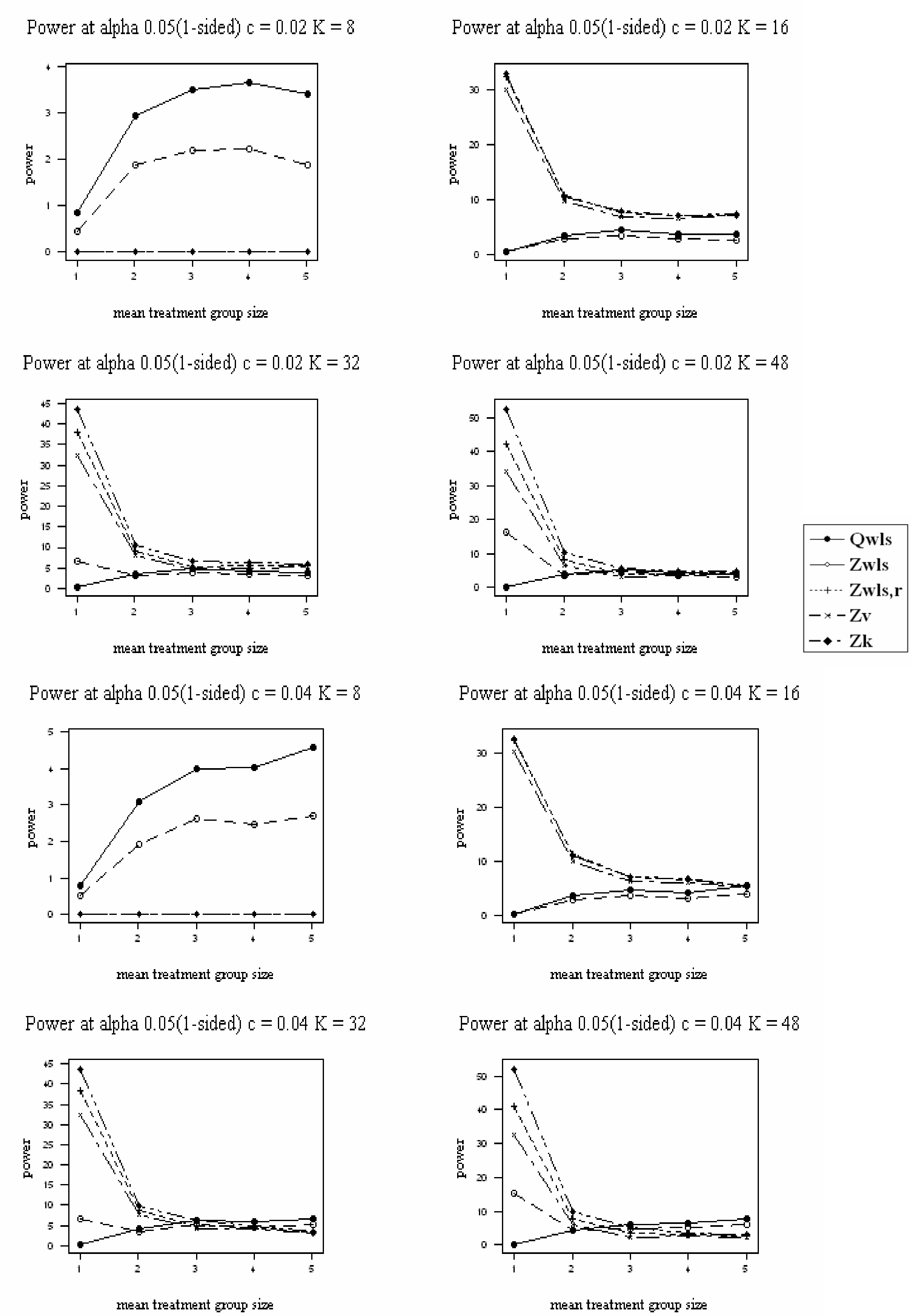


Figure 7 Comparison of the power of five statistical tests by *c*, level of significance (one-sided) and number of center where mean treatment group sizes are equal.

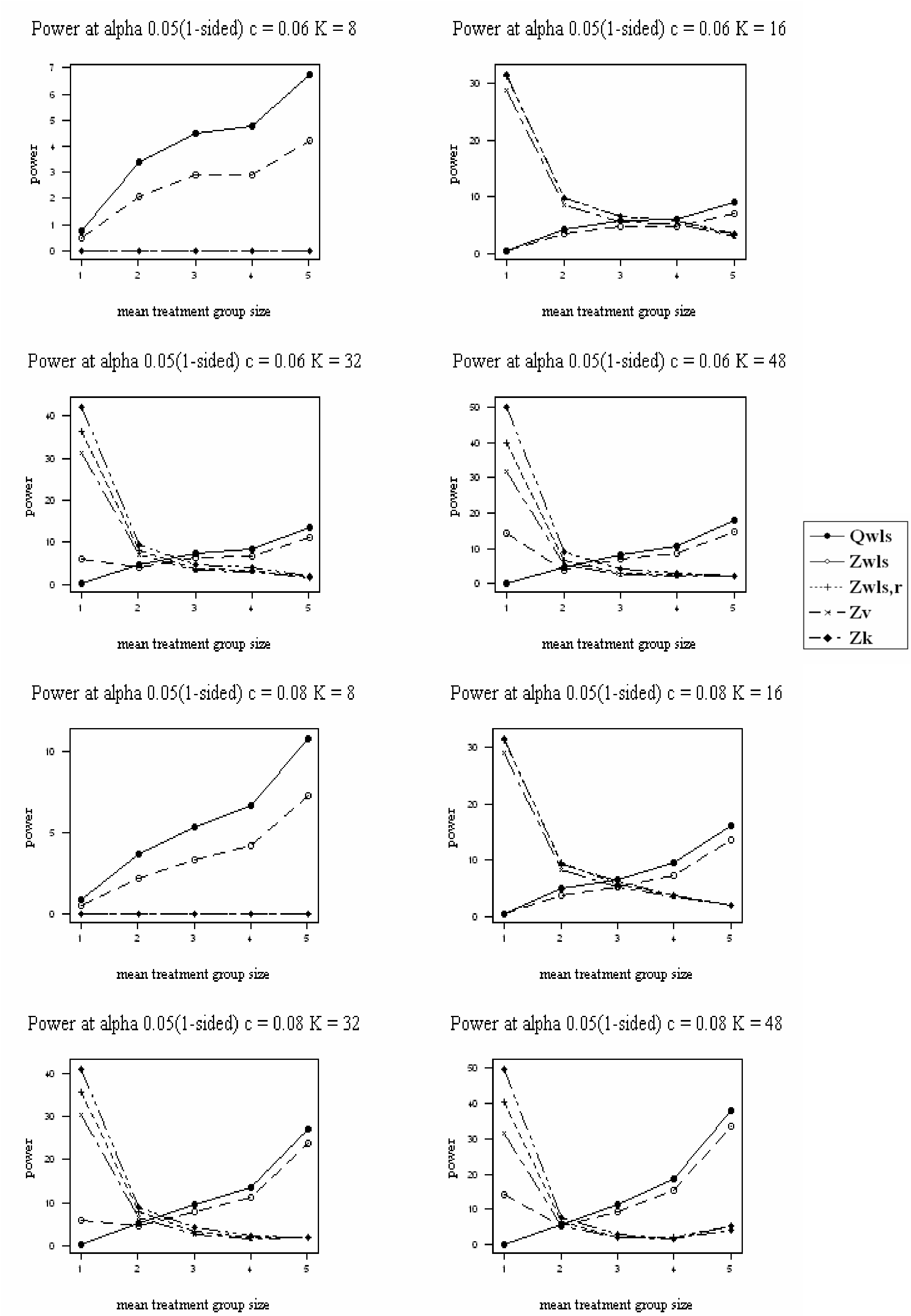


Figure 7 Comparison of the power of five statistical tests by *c*, level of significance (one-sided) and number of center where mean treatment group sizes are equal. (Continued)

This step shows the result from simulation by adding a constant method for the one-sided test of significance level when the mean sample size in each treatment group is unequal. The results when the mean treatment group sizes are equal are similar to when mean treatment group sizes are unequal. According to Table 11.2 and Figure 8, when all of c and K is small, regardless of n_{ij} then the Q_{WLS} test has the highest power**.** The Z_K^2 test has the highest power when c equals 0.02, and 0.04 and K is greater than 16, followed by the $Z_{WLS,R}^2$ and Z_V^2 test.

When c = 0.06, and 0.08, K is moderate to large ($K \geq 16$) and if mean sample size in the treatment group is very small (n_{i1} = 4) and another group has any mean size then the Z_K^2 test is the first ranking (the highest power), but if the mean sample size in the treatment group is greater than small ($n_{i1} \geq 8$) and mean sample size in another group is greater than large size ($n_{i2} \geq 32$) then the Q_{WLS} test is the first ranking (the highest power).

Table 11.2 Mean ranking of power of tests in each number of centers, at the 0.05 significance level (one-sided) and $n_{i1} \neq n_{i2}$.

c	*K*	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	3.5 (.34)	4.8 (.13)	2.1 (.18)	3.4 (.27)	1.2 (.13)
	32	3.2 (.55)	3.6 (.34)	2.9 (.32)	3.9 (.32)	1.4 (.27)
	48	2.8 (.63)	3.1 (.32)	3.0 (.33)	4.2 (.33)	1.9 (.32)
0.04	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	3.4 (.40)	4.5 (.34)	2.3 (.30)	3.4 (.27)	1.4 (.22)
	32	2.8 (.63)	3.3 (.34)	3.0 (.33)	4.1 (.32)	1.8 (.29)
	48	2.7 (.63)	3.0 (.33)	3.1 (.32)	4.2 (.33)	2.0 (.33)
0.06	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.6 (.48)	4.0 (.45)	2.7 (.34)	3.9 (.32)	1.8 (.29)
	32	2.7 (.63)	3.0 (.33)	3.1 (.32)	4.2 (.33)	2.0 (.33)
	48	2.6 (.65)	2.9 (.32)	3.2 (.33)	4.2 (.33)	2.1 (.32)
0.08	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.5 (.50)	3.6 (.48)	2.9 (.38)	4.0 (.33)	2.0 (.29)
	32	2.6 (.65)	2.8 (.33)	3.2 (.33)	4.2 (.33)	2.2 (.33)
	48	2.6 (.65)	2.8 (.33)	3.1 (.32)	4.1 (.32)	2.4 (.43)

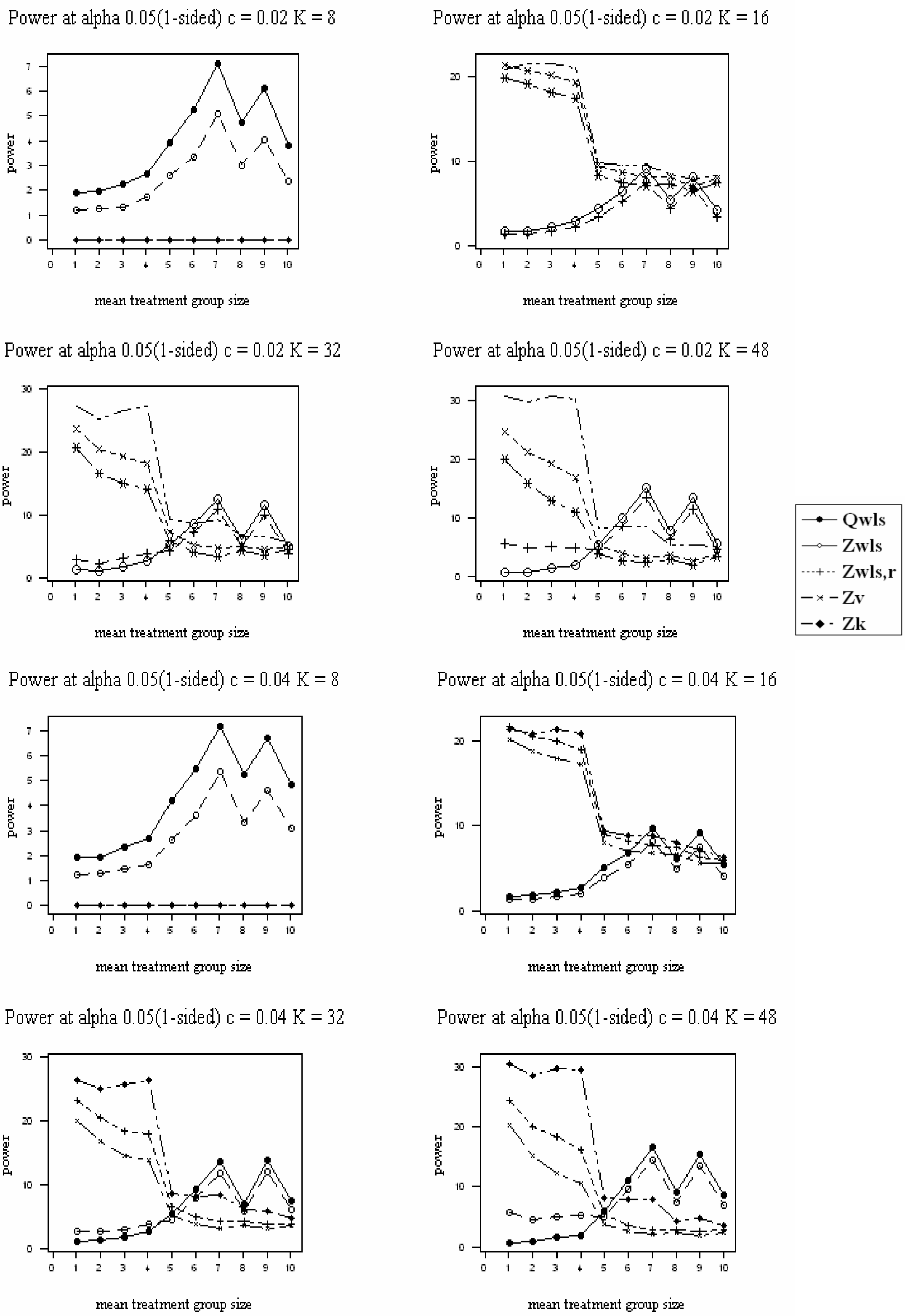


Figure 8 Comparison of the power of five statistical tests by *c*, level of significance (one-sided) and *K* where mean treatment group sizes are unequal.

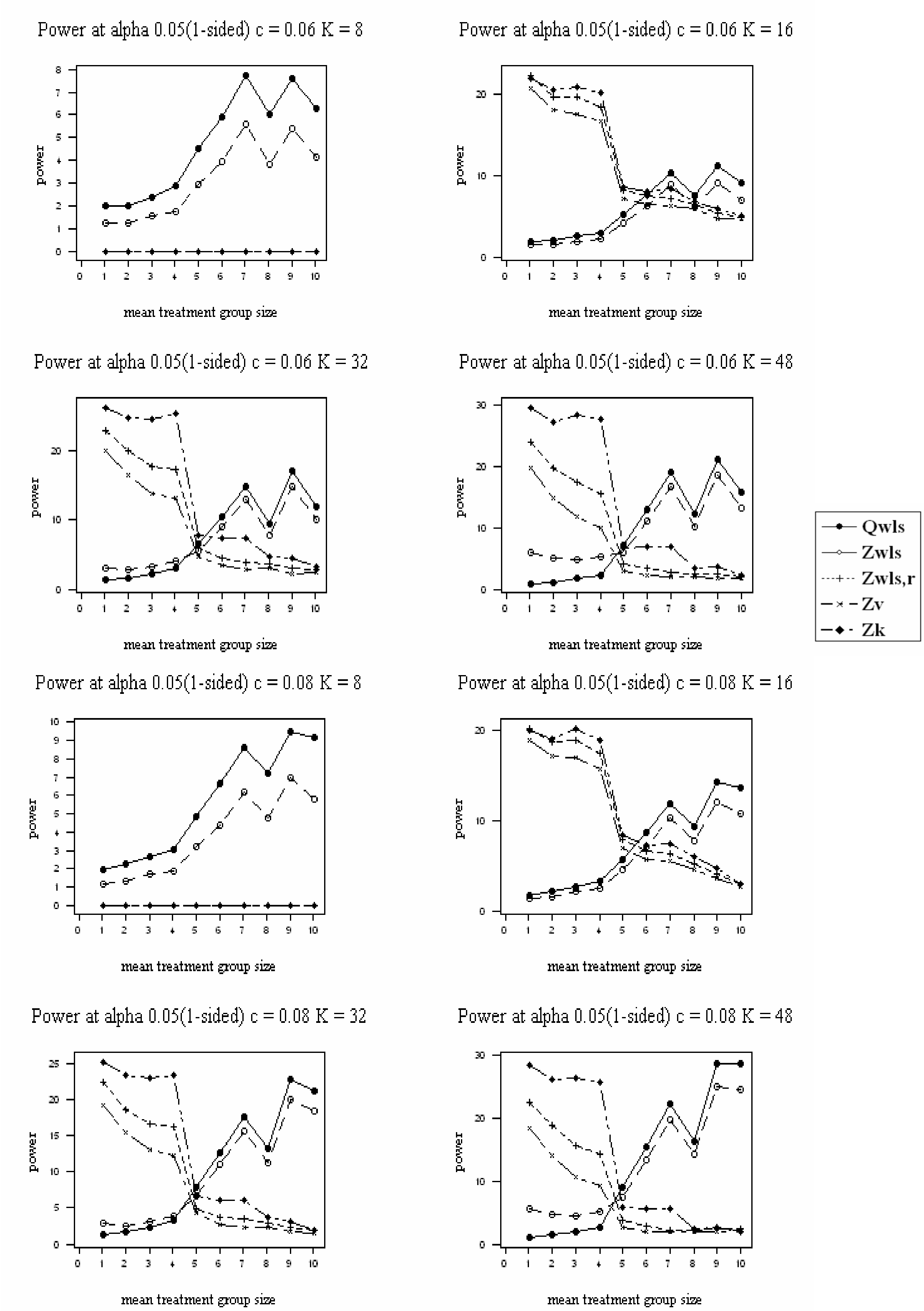


Figure 8 Comparison of the power of five statistical tests by *c*, level of significance (one-sided) and *K* where mean treatment group sizes are unequal. (Continued)

CHAPTER V
DISCUSSION

This chapter is mainly concern with the discussion for the present study. Their details will divide into two parts, one is methodology discussion, and the other is on comparison of five statistical tests under their investigation by the Type I error and the power of test.

5.1 The discussion of the methodology

This study generated data, which depended on four factors; method of simulation, the number of center, mean sample size when in each treatment group are equal and unequal, level of significance for one-sided test and two-sided test.

In criteria of simulation, this study redesigns the simulation considered by Lipsitz et al. The data in each center was generated by method of dropping case when $\hat{\omega}_i = 0$. This study improves the problems from work of Lipsitz et al. by adding a constant to X_{ij}, n_{ij} when $X_{ij} = 0$ or $X_{ij} = n_{ij}$ takes effect $\hat{\omega}_i = 0$. The data was simulated by considering for one-sided and two-sided tests of 0.01, 0.05, and 0.10 significance levels when the mean treatment groups are equal and unequal sizes.

In programming procedure, generating data in each condition is performed for 10,000 iterations with respect to the considered results in this study, whereas Lipsits et al. performed for 1825 iterations.

In statistical test, the formula of the Z_V^2 test in this study is different from formula (14) in Lipsitz et al., because, formula (14) by Lipsitz et al. is incorrect about $\mathrm{var}\left[\left(Y_i - \tau\right)^2\right]$ (Lui and Kelly, 2000). That is, this study used the formula of the Z_V^2 test followed the work of Lui and Kelly.

5.2 The discussion of the results

The results of this study are discussed as regards to the following issues:

5.2.1 The discussion of the results of the type I error comparison.

Capability to control the Type I error of the five statistical tests in this study may be slightly different from the other studies because the interval of the actual Type I error which used in each study is different. Namely, based on z-test; interval [0.004, 0.016] for 0.01 significance level, interval [0.036, 0.064] for 0.05 significance level, and interval [0.081, 0.119] for 0.10 significance level, based on Cochran criteria; interval [0.007, 0.015] for 0.01 significance level, interval [0.040, 0.060] for 0.05 significance level, and interval [0.081, 0.119] for 0.10 significance level, or based on Bradley criteria; interval [0.005, 0.015] for 0.01 significance level, interval [0.025, 0.075] for 0.05 significance level, and interval [0.050, 0.150] for 0.10 significance level.

5.2.2 The discussion of the results of the power of test comparison.

Consideration of the capability to control the Type I error is necessary for comparing the power of the statistical tests. Because if the true of Type I error value is between the nominal level of significance, therefore the power of test is considered.

Following Lpsitz et al. (1998), this study calculate the power by generating a random sequence of risk difference τ_i that equal 0.1 + $c(2U$-1) where U is a random number following the uniform distribution on [0,1] and where c = .02, .04, .06, .08. The case for c = 0 correspond to those for the homogeneity of risk difference under H_o. Note that when c is extremely small, this implies that the null hypothesis of homogeneity holds approximately. The results of the power of test were simulated by dropping case when $\hat{\omega}_i = 0$ (followed the work of Lipsitz et al.) and adding constant to X_{ij} and n_{ij} when $\hat{\omega}_i = 0$. The significance of level is 0.01, 0.05, and 0.10 for one-sided and two-sided tests.

In order to choose an appropriate statistical test, there are two factors that must be taken into account. The first is that probability of type I error value must lie within Cochran limit. Another factor is the high power of test. Comparison the power of test of statistical tests depends on the method of generating data, the number of center, and the mean treatment group sizes. The results from dropping case are different from adding constant. For example, when the number of center and the mean treatment groups are small, then the results from dropping case of the Q_{WLS} and Z_{WLS}^2 tests have the higher power than the $Z_{WLS,R}^2, Z_V^2$, and Z_K^2 tests, but results from adding constant of the Q_{WLS} and Z_{WLS}^2 tests have the lower power than the $Z_{WLS,R}^2, Z_V^2$, and Z_K^2 tests. Finally, when the mean treatment groups are large to very large then the power from both methods is nearly value.

The results and discussions presented in this study should be useful for applied statisticians and epidemiologists when testing the homogeneity of risk difference using the data in a multicenter design or in meta-analysis.

CHAPTER VI
CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

This study focused on comparing the Type I error and the power of tests of those five statistical tests. Using the five statistical tests are different situations that depend on the number of center and the mean treatment group size and method of generating data for one-sided and two-sided tests of significance level. The conclusions of result are:

6.1.1 Comparing the Type I error

6.1.1.1 By dropping case when $\hat{\omega}_i = 0$.

At the level of significance, results among two-sided and one-sided tests are similarly. Meanwhile, the significance level for one-sided has the case that can control the Type I error less than considered for two-sided test. Therefore, this study interested the number of center and the mean treatment group size.

- When mean treatment groups are equal ($n_{i1} = n_{i2}$)

- The Q_{WLS} and Z^2_{WLS} tests perform well (can control the Type I error) when the mean treatment groups are large to very large ($n_{ij} \geq 32$) in any the number of center.

- The $Z^2_{WLS,R}$ and Z^2_V tests perform well when the number of center is greater than moderate ($K \geq 16$) and the mean treatment group size are small.

For given value of the mean sample size group, the Type I error of the Q_{WLS} and Z^2_{WLS} tests tend to increase when number of center are increasing, too. For given value of number of center, the Type I error tends to decrease (closer 1, 5, 10 percent at 0.01, 0.05, 0.10 significance level, respectively) as the mean treatment group sizes increase.

For given values of mean treatment group size (n_{ij}) and the number of center (*K*), the Q_{WLS} test has the Type I error that almost always appears further away from 5 percents than the Z^2_{WLS} test, even when mean treatment group size are large. Consequently, the error in the *Chi-square* approximation for the Q_{WLS} test appears worse than the error in the *F*- approximation for the Z^2_{WLS} test.

From the review literature, to expect the $Z^2_{WLS,R}$ and Z^2_V tests to perform well when both the number of center (*K*) and the mean sample group size (n_{ij}) are large, but they also appear to perform well when the number of center are small to large, but the mean sample size group are small. However, the $Z^2_{WLS,R}$, Z^2_V and Z^2_K tests appear to perform better than the other two test statistics with respect to type I error as its the type I error are in narrow range than the other test statistics.

- When mean treatment groups are unequal ($n_{i1} \neq n_{i2}$)

- The Z^2_{WLS} test performs well when the number of center is less than moderate ($K \leq 16$) and the mean sample size in each treatment group is large.

- The $Z^2_{WLS,R}$ and Z^2_V test performs well when the number of center is small to moderate ($K \leq 16$) and the mean sample size in treatment group is very small and control group is any size. Also, when the number of center is large to very large ($K \geq 32$) and the mean sample size in treatment group is small to moderate and control group is moderate to large.

6.1.1.2 By adding constant when $\hat{\omega}_i = 0$

- When mean treatment groups are equal ($n_{i1} = n_{i2}$)

- The Q_{WLS} test performs well when the number of center is greater than moderate ($K \geq 16$) the mean treatment groups are large to very large ($n_{ij} \geq 32$).

- The Z^2_{WLS} test performs well when the number of center is less than moderate the mean treatment groups are greater than small. And when both of the number of center and mean treatment group are large.

- The Z^2_V test performs well when the number of center is large and the mean treatment groups are moderate.

- When mean treatment groups are unequal ($n_{i1} \neq n_{i2}$)

The results of the Type I error in this method have many cases that each tests have done well. The good performance depends on the number of center, the mean sample size in each treatment group are shown in Tables 7.2 – 7.5. Therefore, this study divided the conclusion in this step followed the number of center.

- When the number of center is less than moderate ($K \leq 16$): the Q_{WLS} and Z^2_{WLS} tests perform well when mean sample size in treatment group is very small and control group is very large ($n_{i1} = 4, n_{i2} = 64$). Also, The Z^2_{WLS} test performs well when both mean sample sizes are large. ($n_{ij} \geq 32$).

- When the number of center is greater than large ($K \geq 32$): the Q_{WLS} and Z^2_{WLS} tests perform well when mean sample size in treatment group is very small and control group is greater than large ($n_{i1} = 4, n_{i2} \geq 32$). The Z^2_{V} and $Z^2_{WLS,R}$ tests perform well when mean sample size in treatment group is moderate and control group is greater than large ($n_{i1} = 16, n_{i2} \geq 32$).

6.1.2 Simulation for comparing the Power of test.

About results for power of test separate to many parts, that considered at varies cases of number of center, mean treatment group sizes, and the constant *c* (*c* is a constant value for generate the statistics values of risk difference in each center and var(τ_i) depend on value of *c*, for this study *c* is equal 0.02, 0.04, 0.06, and 0.08). By the way, data are manipulated for better comprehension by assigning each value of power a ranking.

6.1.2.1 By dropping case when $\hat{\omega}_i = 0$

The trend of power at any significance level for one-sided and two sided test are similarly. When *c* and the number of center increase then the power of the Q_{WLS} and Z^2_{WLS} tests increase, but the power of other tests decrease. Because, the power depend on *c* value, so that this study focused on *c*. Almost all case of value of *c*, the number of center and mean sample size where the Q_{WLS} test has the highest power, followed by the Z_{WLS}^2 test and the Z_V^2 or Z_K^2 tests have the lowest power depend on the mean treatment groups. Except some case that the other test has the highest power, that is when c = 0.02 the number of center is small and mean treatment groups are greater than moderate then the Z_K^2 test has the highest power.

In summary, a one-sided test rather than a two-sided test should be used for statistics discussed by Lpsitz et al. (1998) in testing the homogeneity of risk difference.

6.1.2.2 By adding constant when $\hat{\omega}_i = 0$

The power in this method is different from the power from dropping case. This study concludes followed by value of c and the number of center. First, this study focused on two-sided test of significance level.

c = 0.02: when the number of center is small (K = 8) and every the mean sample size then the $Z_{WLS,R}^2$ test has the highest power.

: when the number of center is moderate and large ($K \geq 16$) and every the mean sample size then the Z_K^2 test has the highest power. Except, when the number of center is very large (K = 48) and the mean sample size is very large then the Q_{WLS} test has the highest power.

$c \geq 0.04$: when the number of center is small (K = 8) and the mean sample size is less than moderate ($n_{ij} \leq 16$) then the $Z_{WLS,R}^2$ test has the highest power.

: when the number of center is greater than moderate ($K \geq 16$) and the mean sample size is less than moderate ($n_{ij} \leq 16$) then the Z_K^2 test has the highest power.

: when the mean sample size is greater than large ($n_{ij} \geq 32$) in any the number of center then the Q_{WLS} test has the highest power.

Next, this study focused on one-sided test of significance level: when the number of center is small (K = 8) in every value of c and the mean sample size then the Q_{WLS} test has the highest power.

$c \le 0.04$: when the number of center is greater than moderate ($K \ge 16$) and every the mean sample size then the Z_K^2 test has the highest power, followed by the $Z_{WLS,R}^2$ and Z_V^2 tests.

$c \ge 0.06$: when the number of center is greater than moderate ($K \ge 16$) and the mean sample sizes are less than small ($n_{ij} \le 8$) then the Z_K^2 test has the highest power. But, if the mean sample sizes are greater than moderate ($n_{ij} \ge 16$) then the Q_{WLS} test has the highest power.

6.2 Recommendations

In the present, constraint for using the statistical tests are different in varies of situation, therefore the recommendations for further study:

- This study focused on the risk difference value, should focused on other value such as, relative risk, odds ratio.
- Try to compare these statistical tests with other tests for decrease constraint about using statistical tests.
- In part of alternative hypothesis, that is generate statistic value from other distribution such that Normal Distribution, Beta Distribution, etc.

REFERENCES

1 Ahn C. *An Evaluation of simple methods for the Estimation of a Common Odds Ratio in cluters with variable size*. **Computational Statistics & Data Analysis** 1997; 24: 47-61.

2. Beitler JP, Landis R. J. *A Mixture-Effects Model for categorical data*. **Biometrics** 1985; 41: 991-1000.

3. Böhning D, Sarol JJr. *Estimating Risk Difference in multicenter studies under baseline-risk heterogeneity*. **Biometrics** 2000; 56: 304-308.

4. Brain H. **Fortran 90 for Scientists and Engineers**, 1994. Cambridge: University Press.

5. Colditz GA., Burdick E. and Mosteller F. *Heterogeneity in Meta-Analysis of Data from Epidemiologic Studies: a Commentary*. **Am J Epidemiol** 1995; 142: 371-382.

6. Cooper MR, Dear KBG, Mcintyre OR, Ozer H, Ellerton J, Cannellos G, Duggan B, and Schiffer C. *A randomized clinical trial comparing melphalan/Prednisone with and without α-2b interferon in newly-diagnosed patients with multiple myolema: A Cancer and Leukemia Group B study*. **Journal of Clinical Oncology** 1993; 11: 155-160.

7. Dersimonial R, and Laird N. *Meta-analysis in Clinical Trials*. **Controlled Clinical Trials** 1986; 7: 177-188.

8. Fleiss JL. *Statistical methods for Rates and Proportions*, 2nd edition, 1981. New York: John Wiley & Sons.

9. Fleiss JL. *Measures of Effect Size for Categorical data* [In Cooper, H., Hedges LV,(Editors)], **The Handbook of Research Synthesis**, 1994. Sage: New York, 245-260.

10. Hasselblad V, Mosteller F, Littenberg B, Chalmers TC, Hunink MGM, Turner JA, Morton SC, Diehr P, Wong JB, Powe NR. *A Survey of Current Problems in meta-Analysis: Discussion from the agency for Health Care Policy and*

research inter-POT work group on literature review/meta-analysis. **Med. Care** 1995; 33: 202-220.

11. John M. Lachin. Biostatistical Methods: The Assessment of Relative Risks, 2000. The George Washington University. Washington, D.C: John Wiley & Sons.
12. Johnson NL. and Kotz S. *Distribution in Statistics- Discrete Distributions*. **John Wiley** 1969: New York.
13. Liang KY, and Self SG. *Tests for Homogeneity of Odds Ratio when the data are sparse*. **Biometrika** 1985; 72: 353-358.
14. Lipsitz SR, Dear KBG, Laird NM, and Molenberghs G. *Tests for homogeneity of the Risk Difference when data are sparse*. **Biometrics** 1998; 54: 148-160.
15. Lui KJ and Kelly C. *A Revisit on Tests for Homogeneity of the Risk Difference*. **Biometrics** 2000; 56: 309-315.
16. Lui KJ and Kelly C. *Tests for Homogeneity of the Risk Ratio in a Series of 2x2 Tables.* **Statistics in Medicine** 2000; 19: 2919-2932.
17. McCullagh R, and Nelder JA. **Generalized Linear Model**, 2nd edition, 1989. London: Chapman and Hill.
18. Mehta, C. R., Patel, N. R. and Senchaudhuri, P. *Importance Sampling for Estimating Exact Probabilities in Permutational Inference.* **Journal of the American Statistical Association** 1988; 83: 999-1005.
19. National Institues of Health. *Resource Information: Understanding Clini*cal *Trial.* [Online]. US: **National Library of Medicine**; 2002.
20. National Research Council. Combining Information: *Statistical Issues and Opporunities for Research.* **Washington, DC: National Academy Press**, 1992: 52.
21. O'Garman TW., Woolson RF., Jones MP., and Lemke JH. *Statistical Analysis of K 2x2 tables: a Comparative Study of Estimators / Tests Statistics for Association and Homogeneity.* **Environmental Health Perspectives** 1990; 87: 103-107.
22. O'Garman TW., Woolson RF. and Jones MP.. *A Comparison of Twos Methods of Estimating a Common Risk Difference in a Stratified Analysis of a*

Multicenter Clinical Trial. **Controlled Clinical Trials** 1994; 15(2): 135-153.

23. Ronny K. and Böhning D. *The Failure of Meta-Analytic Asymptotics for the Seemingly Efficient Estimator of the Common Risk Difference.*
24. Rubinstein RY. **Simulation and the Monte Carlo Method** 1981. New York: John Wiley & Sons.
25. Takkouche B., Cadarso-Suárez C. and Spiegelman D. *Evaluation of Old and New Tests of Heterogeneity in Epidemiologic Meta-Analysis.* **American Journal of Epidemiology** 1999; 150:206-215.
26. The Animated Software Company. **Internet Glossary of Statistical Terms**,2002.
27. Thieraut RJ, Klekamp RC. **Decision making through Operations Research**. 2nd edition, 1975. Singapore: John Wiley & Sons.
28. University of Miami School of Medicine: **Medical Dictionary**; 2003.
29. Viwatwongkasem C. *Heterogeneity Tests for Risk Difference in Meta-Analysis*, 2002.
30. Viwatwonghasem C. and Böhning W. *A Comparison of Risk Difference Estimators in Multi-Center studies under Baseline-Risk Heterogeneity*. *Computational* **Statistics & Data Analysis** 2003; 41:631-644.
31. Walker AM., Martin-Moreno JM. and Artalejo FR.. *Odd Man Out: a Graphical Approach to Meta-Analysis*. **Am J Public Health** 1988; 78:961-966.

APPENDIX

APPENDIX A
FORTRAN SUBROUTINES USED IN THIS STUDY

1. Main Program

Program main

```
! Parameter (True Values)
! RD be a parameter of interest
! i=1,2,...,k  k: be number of centers
! pcT(i): generate from Uniform Distribution
! ptT(i)=pcT(i)+RD be a Unifrom(0,1) of treatment arm in center i
! xt(i) [input] be a random vector of # of success in treatment in center i
!                    xt(i)~ Binomial(n,ptT(i))
! xc(i) [input] be a random vector of # of success in control in center i
!                    xc(i)~ Binomial(n,pcT(i))
!Note: ptT(i) and pcT(i) be parameter
!            pt(i) and pc(i) be statistics used in other subroutine
use MSIMSL
integer i, j,  m, nr, nn, iseed, k
integer xt(100),xc(100), nt(100), nc(100)
integer  ntsize(100), ncsize(100), nk(100),n1, n2
real rpc(100), pcT(100), ptT(100), p(100)
real r(100), rn1(100), rn2(100), rr(100)
real RD, Qwls, Zwls, Zwlsr, Zv, Zk
character(len=40)  name1, name2
integer rxc(100), rxt(100)
real pQw, Q5, Q1, pQw05(100), pQw01(100)
real pZw, W5, W1, pZw05(100), pZw01(100)
real pZr, R5, R1, pZr05(100), pZr01(100)
real pZv, V5, V1, pZv05(100), pZv01(100)
```

```
real pZk, K5, K1, pZk05(100), pZk01(100)

!Read Data file
name1='input1.txt'
open (1,file = name1)
read (1,*) nn                    !nn = # all of cases
do m=1,nn
   read (1,*) nk(m), ntsize(m), ncsize(m)
end do
print*,'Output Filename: Output.txt'
name2='output.txt'
open(2,file=name2)
RD = 0.1
nr = 10000
iseed = 123457
call RNSET(iseed)
! Genarate pc(i) as Uniform Distribution on [0.1,0.8]
! Set ptT(i)=pcT(i)+0.1 (assume that underly risk difference is 0.1 over all center)
! pt for using in simulation of xt
do m=1,nn
   Q1 = 0.0
   Q5 = 0.0
   W1 = 0.0
   W5 = 0.0
   R1 = 0.0
   R5 = 0.0
   V1 = 0.0
   V5 = 0.0
   K1 = 0.0
   K5 = 0.0
   do j=1,nr                            ! nr = number of replicates
      do i=1,nk(m)
```

```
            call rnun (1, rpc)
            p(i) = rpc(1)
            pcT(i) = 0.1 + (0.70*p(i))          !pcT(i) ~ uniform[0.1, 0.8]
            ptT(i)= pcT(i) + RD          !ptT(i)  = pcT(i)+ RD where RD = 0.1
      end do
!Generate nc, nt where let nc, nt be  random variable with a limited range
!nc, nt be the discrete random variable
!     n-2 with probability .2
!     n-1 with probability .2
!     n   with probability .2
!     n+1 with probability .2
!     n+1 with probability .2
      n1 = ncsize(m)
      do i=1,nk(m)
            call rnun(1, rn1)
            r(i) = rn1(1)
            if(r(i) .le. 0.2) then
                  nc(i) = n1 - 2
            else if(r(i) .gt. 0.2 .and. r(i) .lt. 0.4) then
                  nc(i) = n1 - 1
            else if(r(i) .gt. 0.4 .and. r(i) .lt. 0.6)then
                  nc(i) = n1
            else if(r(i) .gt. 0.6 .and. r(i) .lt. 0.8)then
                  nc(i) = n1 + 1
            else
                  nc(i) = n1 + 2
            end if
      end do
! Generate nt
      n2 = ntsize(m)
      do i=1,nk(m)
            call rnun(1, rn2)
```

```
            rr(i) = rn2(1)
            if(rr(i) .le. 0.2) then
                  nt(i) = n2 - 2
            else if(rr(i) .gt. 0.2 .and. rr(i) .lt. 0.4)then
                  nt(i) = n2 - 1
            else if(rr(i) .gt. 0.4 .and. rr(i) .lt. 0.6)then
                  nt(i) = n2
            else if(rr(i) .gt. 0.6 .and. rr(i) .lt. 0.8)then
                  nt(i) = n2 + 1
            else
                  nt(i) = n2 + 2
            end if
      end do
! Generate xc and xt from Binomial Distribution
      do i=1,nk(m)
            call RNBIN(1,nc(i),pcT(i),rxc)
            xc(i)= rxc(1)
            call RNBIN(1,nt(i),ptT(i),rxt)
            xt(i)= rxt(1)
      end do                                    !for i loop
! Compute the statistic tests => Qwls, Zwls, Zwlsr, Zv, Zk
      k=nk(m)
      call RDQwls(k,xt,nt,xc,nc,Qwls,pQw)
      call RDZwls(k,xt,nt,xc,nc,Zwls,pZw)
      call RDZwlsr(k,xt,nt,xc,nc,Zwlsr,pZr)
      call RDZv(k,xt,nt,xc,nc,Zv,pZv)
      call RDZk(k,xt,nt,xc,nc,Zk,pZk)
! Compare the probability of Qwls and nominal level = 0.05 and 0.01
            if (pQw .lt. 0.05) then
                  Q5 = Q5+1
            end if
            if (pQw .lt. 0.01) then
```

```
                Q1 = Q1+1
        end if
! Compare the probability of Zwls and nominal level = 0.05 and 0.01
        if (pZw .lt. 0.05) then
                W5 = W5+1
        end if
        if (pZw .lt. 0.01) then
                W1 = W1+1
        end if
! Compare the probability of Zwlsr and nominal level = 0.05 and 0.01
        if (pZr .lt. 0.05) then
                R5 = R5+1
        end if
        if (pZr .lt. 0.01) then
                R1 = R1+1
        end if
! Compare the probability of Zv and nominal level = 0.05 and 0.01
        if (pZv .lt. 0.05) then
                V5 = V5+1
        end if
        if (pZv.lt. 0.01) then
                V1 = V1+1
        end if
! Compare the probability of Zk and nominal level = 0.05 and 0.01
        if (pZk .lt. 0.05) then
                K5 = K5+1
        end if
        if (pZk .lt. 0.01) then
                K1 = K1+1
        end if
  end do                                    !for j=1,nr
```

```
    pQw05(m) = (Q5/nr)*100
    pQw01(m) = (Q1/nr)*100
    pZw05(m) = (W5/nr)*100
    pZw01(m) = (W1/nr)*100
    pZr05(m) = (R5/nr)*100
    pZr01(m) = (R1/nr)*100
    pZv05(m) = (V5/nr)*100
    pZv01(m) = (V1/nr)*100
    pZk05(m) = (K5/nr)*100
    pZk01(m) = (K1/nr)*100
  end do                                    !for m=1,nn
  write (2,*)' '
  do m=1,nn
        write(2,10) m, nk(m), ntsize(m), ncsize(m)
        write(2,11)
        write(2,12) pQw05(m),pQw01(m)
        write(2,15) pZw05(m),pZw01(m)
        write(2,20) pZr05(m),pZr01(m)
        write(2,25) pZv05(m),pZv01(m)
        write(2,30) pZk05(m),pZk01(m)
        write(2,*)' '
    end do
10  format('m=',i4,x,'nk=',i4,2x,'nt=',i4,2x,'nc=',i4)
11  format('alpha: 0.05',5x,'0.01')
12  format(2f10.4)
15  format(2f10.4)
20  format(2f10.4)
25  format(2f10.4)
30  format(2f10.4)
```

end program

1. Subroutine for Q_{WLS} statistic test

```
Subroutine RDQwls(k,xt,nt,xc,nc,Qwls,pQw)
implicit none
integer i, k , xt(1), xc(1),nt(1), nc(1), newk
real xtN(100),xcN(100),ncN(100),ntN(100) ,loss
real Qwls, t1, t2, t,rd(100), sum, df1, pQw, chidf
real pt(100), pc(100), w(100),w1(100),w2(100)
newk = 0.0
t1= 0.0
t2= 0.0
loss = 0.0
do i=1,k
        xcN(i)= xc(i)
        xtN(i)= xt(i)
        ncN(i)= nc(i)
        ntN(i)= nt(i)
                pt(i)= xtN(i)/ntN(i)
                pc(i)= xcN(i)/ncN(i)
                rd(i)= pt(i)-pc(i)
                w1(i) = (pt(i)*(1-pt(i)))/(ntN(i)-1)
                w2(i) = (pc(i)*(1-pc(i)))/(ncN(i)-1)
                w(i) = w1(i)+w2(i)
                if (w(i) > 0.0 )then
                        t1 = t1 + (rd(i)/w(i))
                        t2 = t2 + (1/w(i))
                end if
                if (w(i) .eq. 0) then
                        loss = loss+1
                end if
end do
newk = k – loss
 t = t1/t2
sum =0.0
```

```
do i=1,k
        if (w(i) >0) then
                sum = sum +((rd(i)-t)**2/w(i))
        end if
    end do
Qwls = sum
df1 = newk - 1
pQw = 1 - chidf(Qwls, df1)
return
end subroutine
```

2. Subroutine for Z^2_{WLS} statistic test

```
Subroutine RDZwls(k,xt,nt,xc,nc,Zwls,pZw)
implicit none
integer i, k ,xt(1), xc(1), nt(1), nc(1)
real Zwls, Qwls, t1, t2, t,  rd(100), sum1, loss, newk
real  pt(100), pc(100),w(100),w1(100),w2(100)
real xtN(100),xcN(100),ncN(100),ntN(100)
real df1, df2, pZw, fdf
t1= 0.0
t2= 0.0
loss = 0.0
do i=1,k
        xcN(i)= xc(i)
        xtN(i)= xt(i)
        ncN(i)= nc(i)
        ntN(i)= nt(i)
        pt(i)= xtN(i)/ntN(i)
        pc(i)= xcN(i)/ncN(i)

        rd(i)= pt(i)-pc(i)
        w1(i) = (pt(i)*(1-pt(i)))/(ntN(i)-1)
```

```
        w2(i) = (pc(i)*(1-pc(i)))/(ncN(i)-1)
        w(i) = w1(i)+w2(i)
        if (w(i)>0) then
                t1 = t1+(rd(i)/w(i))
                t2 = t2+(1/w(i))
        end if
        if (w(i) .eq. 0) then
                loss = loss+1
        end if
end do
newk = k-loss
t = t1/t2
sum1= 0.0
do i=1,k
        if (w(i) > 0) then
                sum1 = sum1+(((rd(i)-t)**2)/w(i))
        end if
end do
Qwls = sum1
Zwls = ((Qwls-(newk-1))**2)/(2*(newk-1))
df1 = 1.0
df2 = newk - 1
pZw = 1 - fdf(Zwls,df1, df2)
return
end subroutine
```

3. Subroutine for $Z^2_{WLS,R}$ statistic test

```
Subroutine RDZwlsr(k,xt,nt,xc,nc,Zwlsr,pZr)
implicit none
integer i, k,xc(1), xt(1), nt(1), nc(1)
real Zwlsr, Qwls, t1, t2, t, sum1, sum2, A, newk, loss
real pt(100), pc(100),w(100),w1(100),w2(100),rd(100)
```

```
real xtN(100),xcN(100),ncN(100),ntN(100)
real df1, df2, pZr, fdf
t1= 0.0
t2= 0.0
loss = 0.0
do i=1,k
        xcN(i)= xc(i)
        xtN(i)= xt(i)
        ncN(i)= nc(i)
        ntN(i)= nt(i)
        pt(i)= xtN(i)/ntN(i)
        pc(i)= xcN(i)/ncN(i)
        rd(i)= pt(i)-pc(i)
        w1(i) = (pt(i)*(1-pt(i)))/(ntN(i)-1)
        w2(i) = (pc(i)*(1-pc(i)))/(ncN(i)-1)
        w(i) = w1(i)+w2(i)
        if (w(i) > 0) then
                t1 = t1+(rd(i)/w(i))
                t2 = t2+(1/w(i))
        end if
        if (w(i) .eq. 0) then
                loss = loss+1
        end if
end do
newk = k-loss
t = t1/t2
sum1=0.0
sum2=0.0
do i=1,k
        if (w(i)>0) then
                sum1 = sum1+(((rd(i)-t)**2)/w(i))
                sum2 = sum2+(((((rd(i)-t)**2)/w(i))-1)**2)
```

```
        end if
end do
Qwls = sum1
A = sum2
Zwlsr = ((Qwls-newk)**2)/A
df1 = 1.0
df2 = newk-1
pZr = 1 - fdf(Zwlsr,df1, df2)
return
end subroutine
```

4. **Subroutine for Z_V^2 statistic test**

```
Subroutine RDZv(k,xt,nt,xc,nc,Zv,pZv)
implicit none
integer i, k,xc(1), xt(1), nt(1), nc(1)
real Zv, t1, t2, t, sum1,sum2, b1, b2, rd(100)
real pt(100), pc(100),w(100),w1(100),w2(100)
real C1(100), C2(100),C3(100),C4(100),a(100)
real xtN(100),xcN(100),ncN(100),ntN(100)
real df1, df2, pZv, fdf, loss, newk
t1= 0.0
t2= 0.0
loss = 0.0
do i=1,k
     xcN(i)= xc(i)
     xtN(i)= xt(i)
     ncN(i)= nc(i)
     ntN(i)= nt(i)
     pt(i)= xtN(i)/ntN(i)
     pc(i)= xcN(i)/ncN(i)
     rd(i)= pt(i)-pc(i)
     w1(i) = (pt(i)*(1-pt(i)))/(ntN(i)-1)
```

```
        w2(i) = (pc(i)*(1-pc(i)))/(ncN(i)-1)
        w(i) = w1(i)+w2(i)
        if (w(i)>0) then
                t1 = t1+(rd(i)/w(i))
                t2 = t2+(1/w(i))
        end if
        if (w(i) .eq. 0) then
                loss = loss+1
        end if
end do
newk = k-loss
t = t1/t2
sum1=0.0
sum2=0.0
do i=1,k
        if (w(i)>0) then
                C1(i) = ((pt(i)*(1-pt(i)))*(1+(3*pt(i)*(1-pt(i))*(ntN(i)-2))))/(ntN(i)**3)
                C2(i) = ((pc(i)*(1-pc(i)))*(1+(3*pc(i)*(1-pc(i))*(ncN(i)-2))))/(ncN(i)**3)
                C3(i) =     (6*pt(i)*(1-pt(i))*pc(i)*(1-pc(i)))/(ntN(i)*ncN(i))
                C4(i) = ((pt(i)*(1-pt(i)))/ntN(i))+((pc(i)*(1-pc(i)))/ncN(i))
                a(i) = C1(i) + C2(i) + C3(i) - (C4(i)**2)
                sum1 = sum1+((((rd(i)-t)**2)-w(i))/a(i))
                sum2 = sum2+(((((rd(i)-t)**2)-w(i))**2)/(a(i)**2))
        end if
end do
b1 = (sum1)**2
b2 = sum2
Zv = b1/b2
df1 = 1.0
df2 = newk-1
pZv = 1 - fdf(Zv,df1, df2)
return
```

```
end subroutine
```

5. Subroutine for Z_K^2 statistic test

```
Subroutine RDZk(k,xt,nt,xc,nc,Zk,pZk)
implicit none
integer i,k, xt(1), xc(1), nt(1), nc(1)
real Zk, t1, t2, b1, b2, tn, sum1, sum2 ,loss, newk
real xtN(100),xcN(100),ncN(100),ntN(100)
real pt(100), pc(100),w(100),w1(100),w2(100)
real M1(100), M2(100), rd(100)
real df1, df2, pZk, fdf
t1= 0.0
t2= 0.0
b1= 0.0
b2= 0.0
do i=1,k
        xcN(i)= xc(i)
        xtN(i)= xt(i)
        ncN(i)= nc(i)
        ntN(i)= nt(i)
        pt(i)= xtN(i)/ntN(i)
        pc(i)= xcN(i)/ncN(i)
        rd(i)= pt(i)-pc(i)
        t1 = t1 + (rd(i)/((1/ntN(i))+(1/ncN(i))))
        t2 = t2 + (1/((1/ntN(i))+(1/ncN(i))))
        w1(i) = (pt(i)*(1-pt(i)))/(ntN(i)-1)
        w2(i) = (pc(i)*(1-pc(i)))/(ncN(i)-1)
        w(i) = w1(i)+w2(i)
end do
tn = t1/t2
sum1=0.0
sum2=0.0
```

```
loss = 0.0
do i=1,k
        if (w(i)>0) then
                M1(i) = (rd(i)-tn)**2
                M2(i) = ((1/ntN(i))+(1/ncN(i)))**2
                sum1 = sum1+((M1(i)-w(i))/M2(i))
                sum2 = sum2+(((M1(i)-w(i))**2)/(M2(i)**2))
        end if
        if (w(i) .eq. 0) then
                loss = loss+1
        end if
end do
newk = k-loss
b1 = sum1
b2 = sum2
Zk = (b1*b1)/b2
df1 = 1.0
df2 = newk-1
pZk = 1 - fdf(Zk,df1, df2)
return
end subroutine
```

APPENDIX B

The Results of the Simulated Study

This main issue in this appendix is the results of comparing the power of tests with two methods of generating data, that is dropping case and adding constant method. With respect to the conditions, it still uses the same condition like the results of power of tests in Chapter IV. In this part, the results of the power of tests show at 0.01 and 0.10 levels of significance for one-sided and two-sided test.

1. Dropping cases when $\hat{\omega}_i = 0$ at 0.01and 0.10 significance levels

- **Two-sided test**

Table 1.1 Mean ranking of power of tests in each number of centers, at the 0.01 significance level and $n_{i1} = n_{i2}$.

c	K	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.6 (.68)	3.6 (.68)	1.8 (.49)	4.2 (.49)	2.8 (.49)
	32	2.4 (.68)	3.4 (.68)	3.2 (.49)	4.2 (.49)	1.8 (.49)
	48	2.0 (.63)	3.0 (.63)	3.4 (.60)	4.4 (.40)	2.2 (.49)
0.04	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.2 (.58)	3.6 (.68)	2.4 (.60)	4.4 (.40)	2.4 (.51)
	32	2.6 (.98)	3.8 (.49)	2.6 (.25)	3.2 (.92)	3.8 (.37)
	48	2.0 (.00)	2.0 (.00)	4.4 (.25)	4.6 (.25)	3.0 (.00)
0.06	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.0 (.00)	2.4 (.40)	2.8 (.20)	5.0 (.00)	3.8 (.20)
	32	1.0 (.00)	2.0 (.00)	4.2 (.20)	4.8 (.20)	3.0 (.00)
	48	1.0 (.00)	2.0 (.00)	4.2 (.20)	4.8 (.20)	3.0 (.00)
0.08	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.0 (.00)	2.0 (.00)	3.0 (.00)	5.0 (.00)	4.0 (.00)
	32	1.0 (.00)	2.0 (.00)	4.0 (.00)	5.0 (.00)	3.0 (.00)
	48	1.0 (.00)	2.0 (.00)	4.0 (.32)	4.6 (.25)	3.4 (.40)

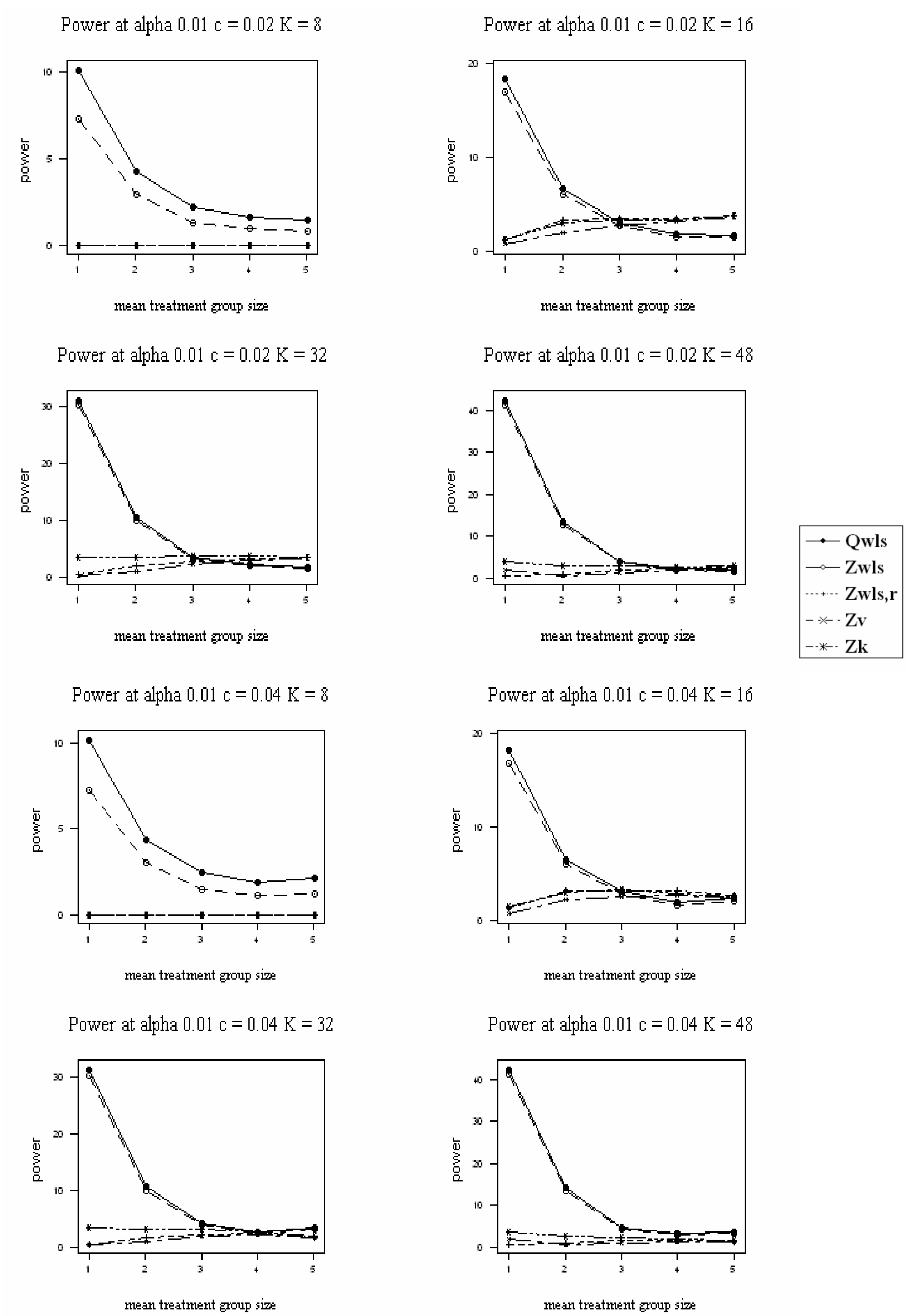


Figure 1A Comparison of the power of tests by *c*, at 0.01 level of significance and number of center (*K*) where mean treatment group sizes are equal.

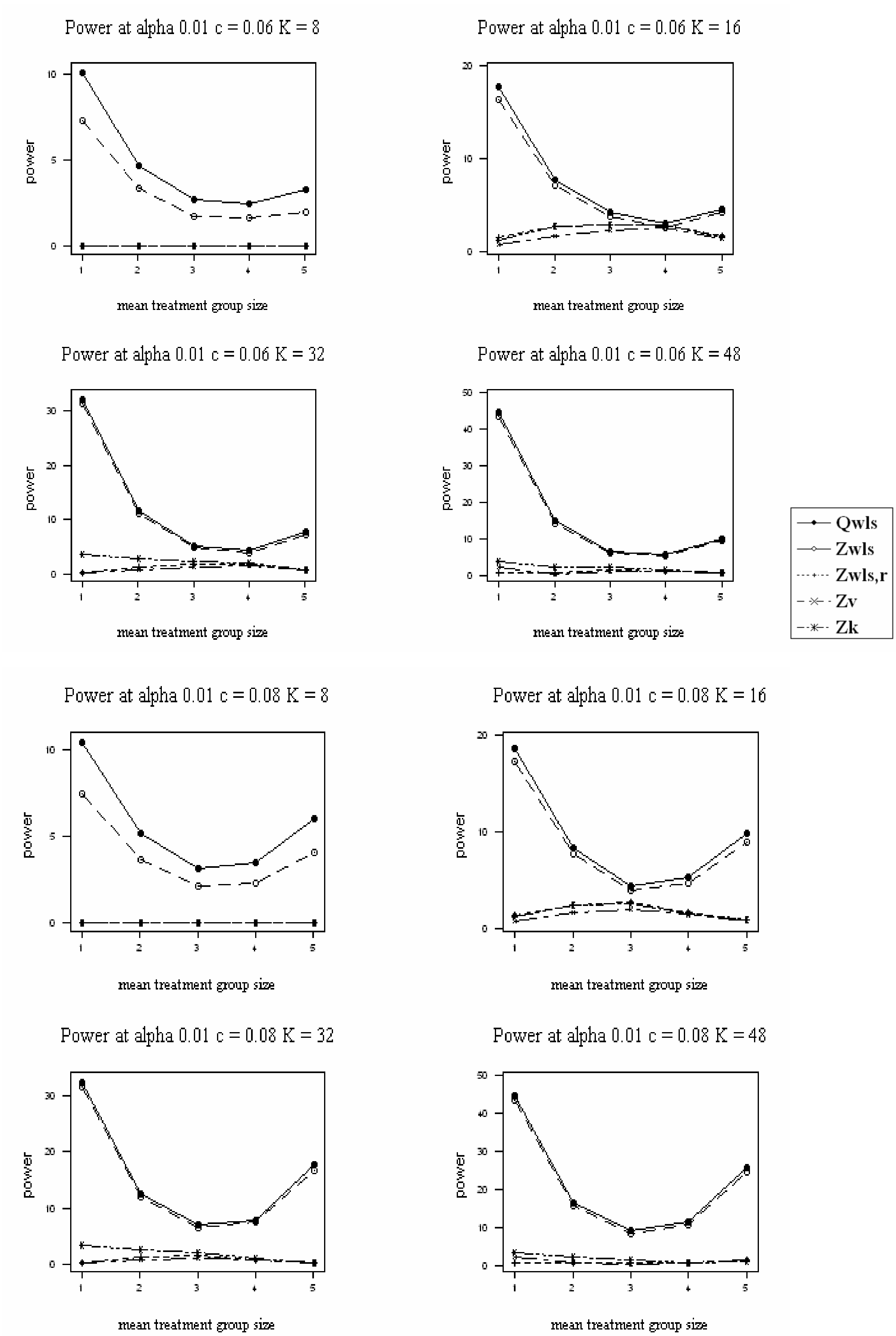


Figure 1A. Comparison of the power of tests by *c*, at 0.01 level of significance and number of center (*K*) where mean treatment group sizes are equal. (Continued)

Table 1.2 Mean ranking of power of tests in each number of centers, at the 0.10 significance level and $n_{i1} = n_{i2}$.

c	*K*	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	3.4 (.98)	3.2 (.49)	2.7 (.20)	3.1 (.40)	2.6 (.98)
	16	2.6 (.68)	3.8 (.74)	3.0 (.63)	3.8 (.37)	1.8 (.49)
	32	1.2 (.20)	3.4 (.68)	4.0 (.45)	4.0 (.32)	2.4 (.51)
	48	1.0 (.00)	2.8 (.58)	4.0 (.32)	4.0 (.45)	3.2 (.58)
0.04	8	3.2 (.58)	4.4 (.60)	1.6 (.40)	3.6 (.40)	2.2 (.49)
	16	1.8 (.58)	3.8 (.74)	3.4 (.40)	4.0 (.45)	2.0 (.45)
	32	1.0 (.00)	2.0 (.00)	4.4 (.25)	4.4 (.40)	3.2 (.20)
	48	1.0 (.00)	2.0 (.00)	4.4 (.25)	3.8 (.49)	3.8 (.37)
0.06	8	2.2 (.74)	4.4 (.60)	1.8 (.37)	3.8 (.37)	2.8 (.37)
	16	1.0 (.00)	3.0 (.63)	3.8 (.37)	4.6 (.25)	2.6 (.25)
	32	1.0 (.00)	2.0 (.00)	4.4 (.25)	4.0 (.45)	3.6 (.40)
	48	1.0 (.00)	2.0 (.00)	4.4 (.25)	3.2 (.20)	4.4 (.40)
0.08	8	1.4 (.40)	3.8 (.74)	2.4 (.25)	4.4 (.25)	3.0 (.55)
	16	1.0 (.00)	2.2 (.20)	4.0 (.32)	4.8 (.20)	3.0 (.32)
	32	1.0 (.00)	2.0 (.00)	4.0 (.45)	3.8 (.37)	4.2 (.37)
	48	1.0 (.00)	2.0 (.00)	4.0 (.00)	3.0 (.00)	5.0 (.00)

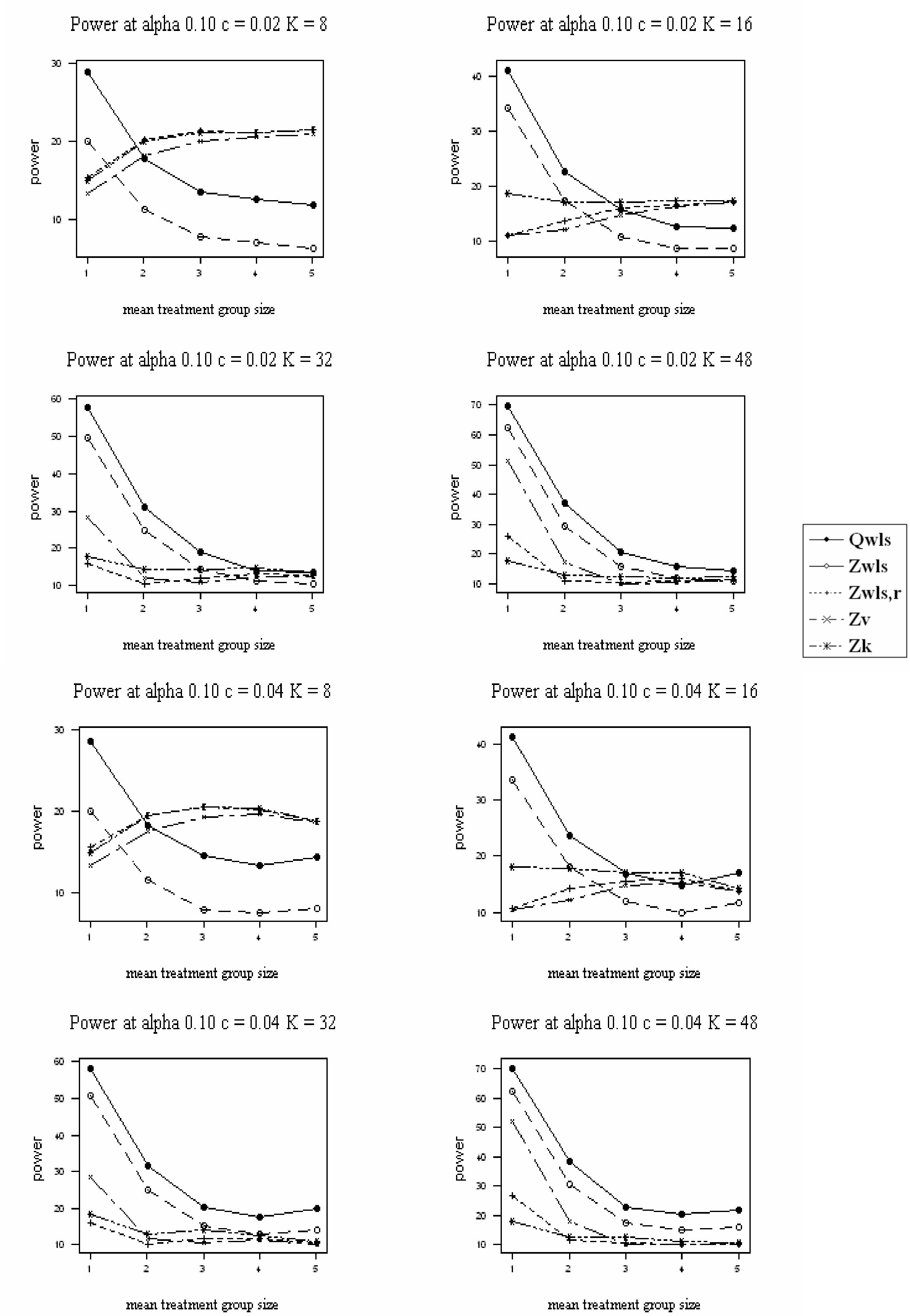


Figure 1B Comparison of the power of tests by *c*, at 0.10 level of significance and number of center (*K*) where mean treatment group sizes are equal.

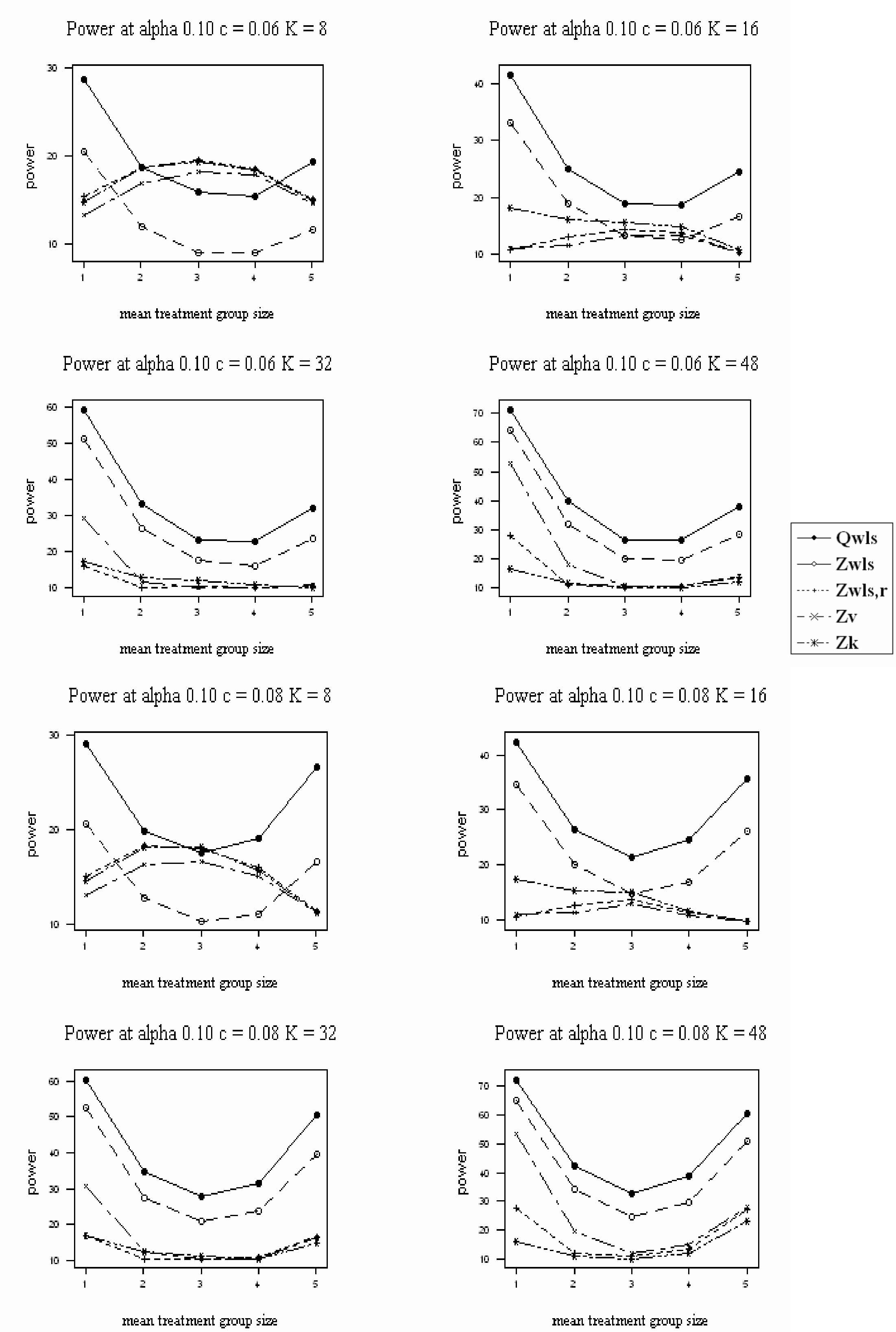


Figure 1B Comparison of the power of tests by *c*, at 0.10 level of significance and number of center (*K*) where mean treatment group sizes are equal. (Continued)

Table 1.3 Mean ranking of power of tests in each number of centers, at the 0.01 significance level and $n_{i1} \neq n_{i2}$.

c	***K***	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.3 (.30)	2.3 (.30)	3.7 (.30)	4.8 (.20)	2.9 (.10)
	32	1.2 (.20)	2.3 (.30)	3.9 (.23)	4.2 (.29)	3.4 (.40)
	48	1.1 (.07)	2.0 (.13)	4.1 (.10)	4.2 (.33)	3.6 (.34)
0.04	8	1.1 (.07)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.2 (.20)	2.3 (.30)	3.6 (.31)	4.9 (.10)	3.0 (.15)
	32	1.1 (.05)	2.0 (.05)	4.1 (.10)	4.3 (.30)	3.6 (.31)
	48	1.1 (.05)	2.0 (.05)	4.1 (.10)	4.2 (.33)	3.7 (.30)
0.06	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.0 (.00)	2.0 (.00)	3.8 (.13)	5.0 (.00)	3.2 (.13)
	32	1.0 (.00)	2.2 (.20)	4.0 (.15)	4.2 (.36)	3.6 (.31)
	48	1.1 (.05)	1.9 (.05)	4.0 (.00)	4.2 (.33)	3.8 (.33)
0.08	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.0 (.00)	2.0 (.00)	3.9 (.10)	5.0 (.00)	3.1 (.10)
	32	1.1 (.07)	1.9 (.07)	4.1 (.10)	4.3 (.30)	3.6 (.31)
	48	1.1 (.05)	1.9 (.05)	3.8 (.13)	4.2 (.32)	4.1 (.30)

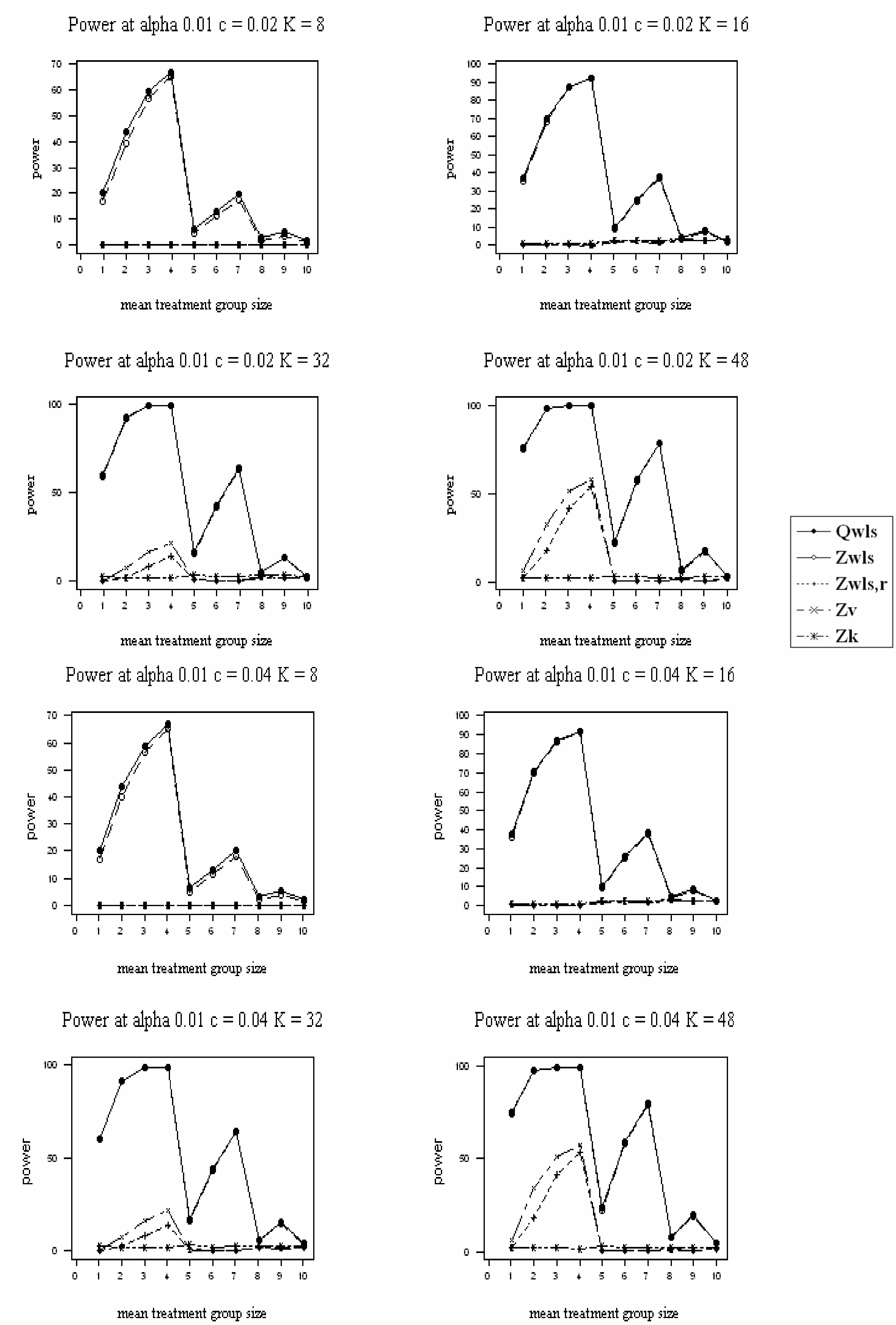


Figure 1C Comparison of the power of tests by *c*, at 0.01 level of significance and number of center (*K*) where mean treatment group sizes are unequal.

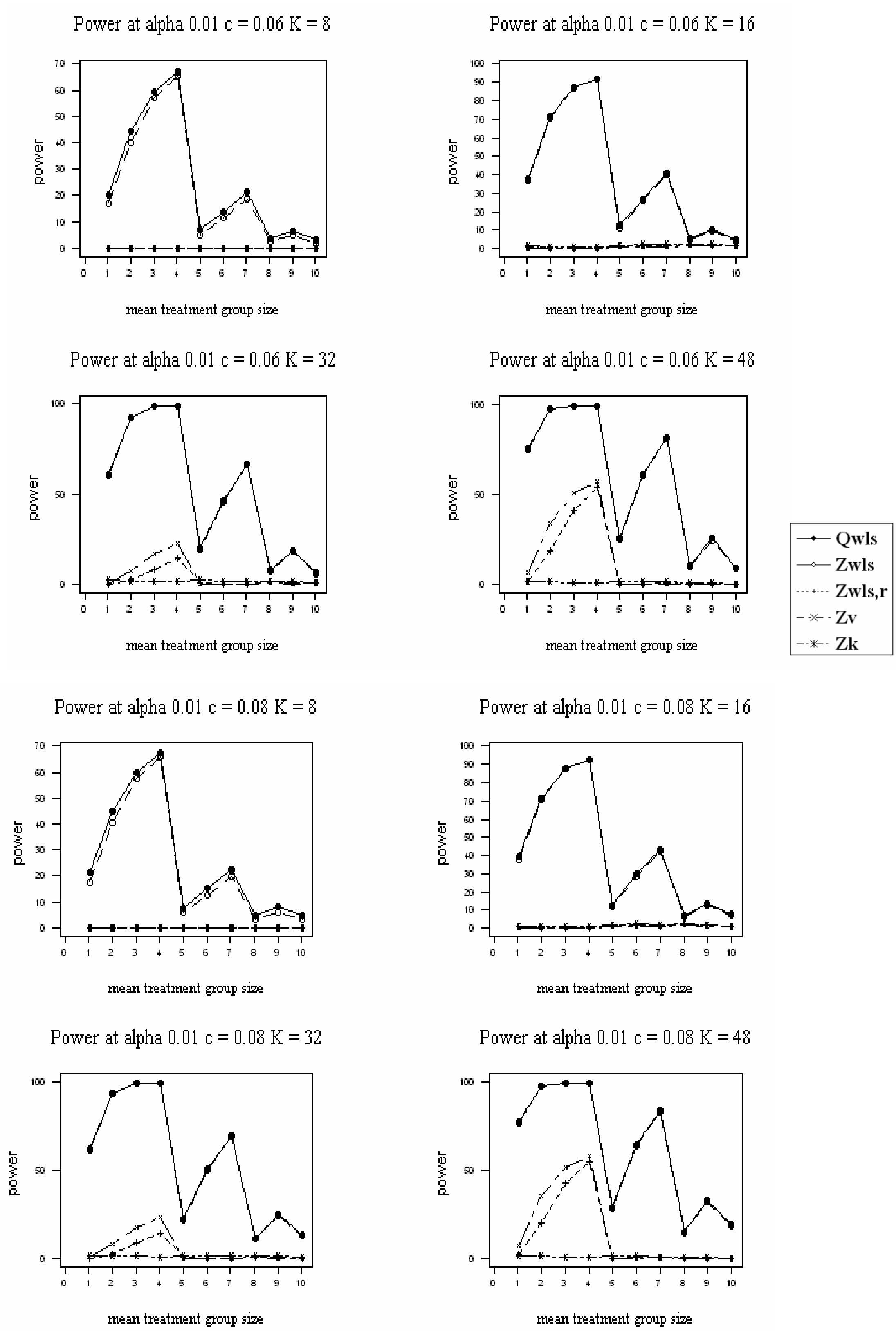


Figure 1C Comparison of the power of tests by *c*, at 0.01 level of significance and number of center (*K*) where mean treatment group sizes are unequal. (Continued)

Table 1.4 Mean ranking of power of tests in each number of centers, at the 0.10 significance level and $n_{i1} \neq n_{i2}$.

c	*K*	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.9 (.46)	3.2 (.49)	3.2 (.36)	4.3 (.30)	2.4 (.27)
	16	1.3 (.30)	2.6 (.40)	3.8 (.25)	4.0 (.29)	3.3 (.42)
	32	1.0 (.00)	2.2 (.20)	3.8 (.20)	3.9 (.28)	4.1 (.38)
	48	1.1 (.01)	1.9 (.01)	3.7 (.15)	3.9 (.28)	4.4 (.31)
0.04	8	1.8 (.42)	3.2 (.49)	3.1 (.41)	4.4 (.27)	2.5 (.22)
	16	1.0 (.00)	2.5 (.34)	3.9 (.18)	4.2 (.29)	3.4 (.37)
	32	1.0 (.00)	2.0 (.00)	3.9 (.18)	3.9 (.28)	4.2 (.33)
	48	1.1 (.01)	1.9 (.01)	3.7 (.15)	3.9 (.28)	4.4 (.31)
0.06	8	1.3 (.30)	3.2 (.49)	3.3 (.34)	4.5 (.22)	2.7 (.15)
	16	1.0 (.00)	2.0 (.00)	4.1 (.10)	4.3 (.30)	3.6 (.31)
	32	1.0 (.00)	2.0 (.00)	3.6 (.22)	3.9 (.28)	4.5 (.22)
	48	1.1 (.07)	1.9 (.07)	3.6 (.16)	3.4 (.16)	5.0 (.00)
0.08	8	1.0 (.00)	2.6 (.34)	3.7 (.21)	4.9 (.10)	2.8 (.13)
	16	1.0 (.00)	2.0 (.00)	4.0 (.15)	4.2 (.33)	3.8 (.29)
	32	1.0 (.00)	2.0 (.00)	3.3 (.15)	3.9 (.23)	4.8 (.13)
	48	1.1 (.07)	1.9 (.07)	3.6 (.16)	3.4 (.16)	5.0 (.00)

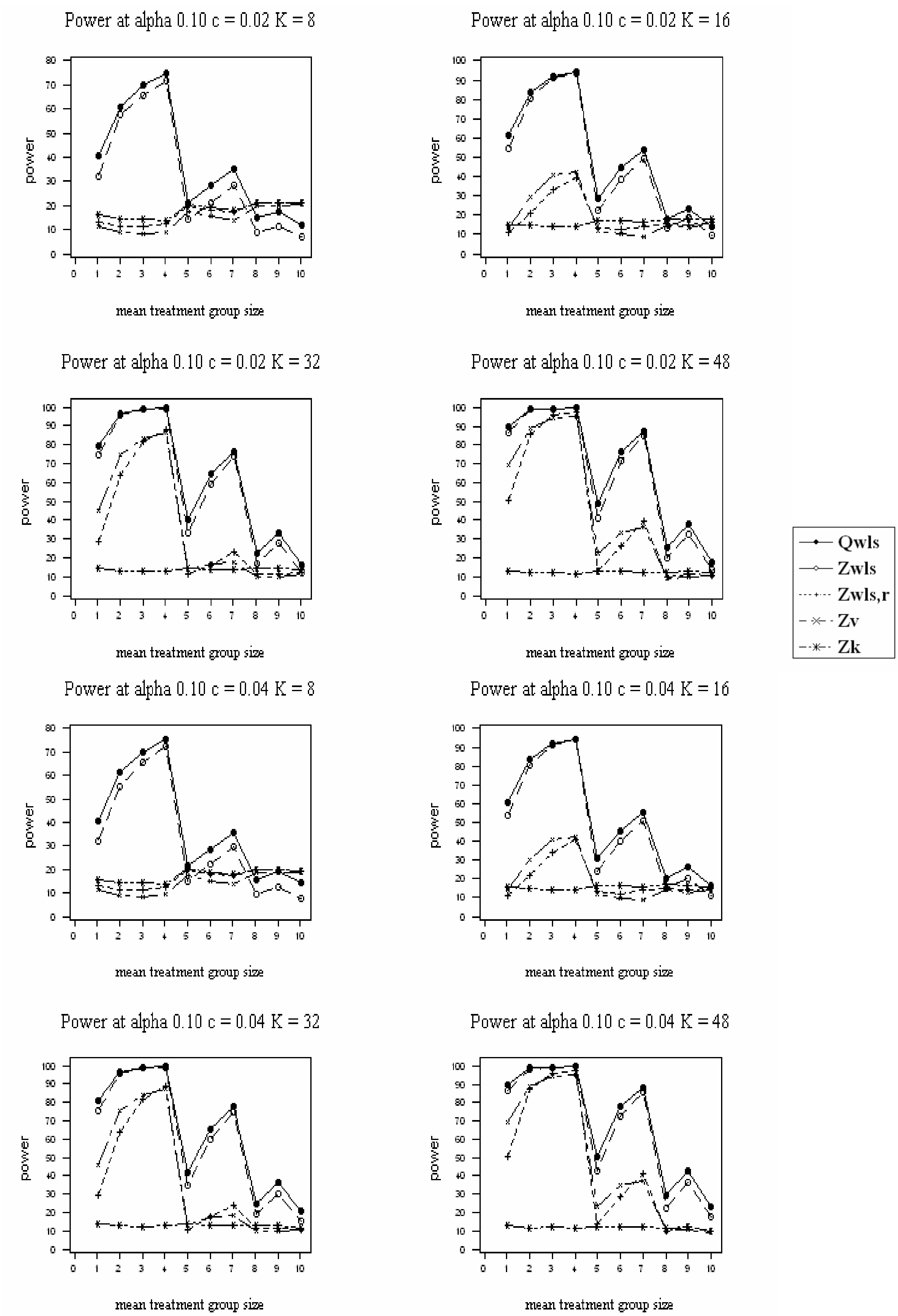


Figure 1D Comparison of the power of tests by *c*, 0.10 level of significance and number of center (*K*) where mean treatment group sizes are unequal.

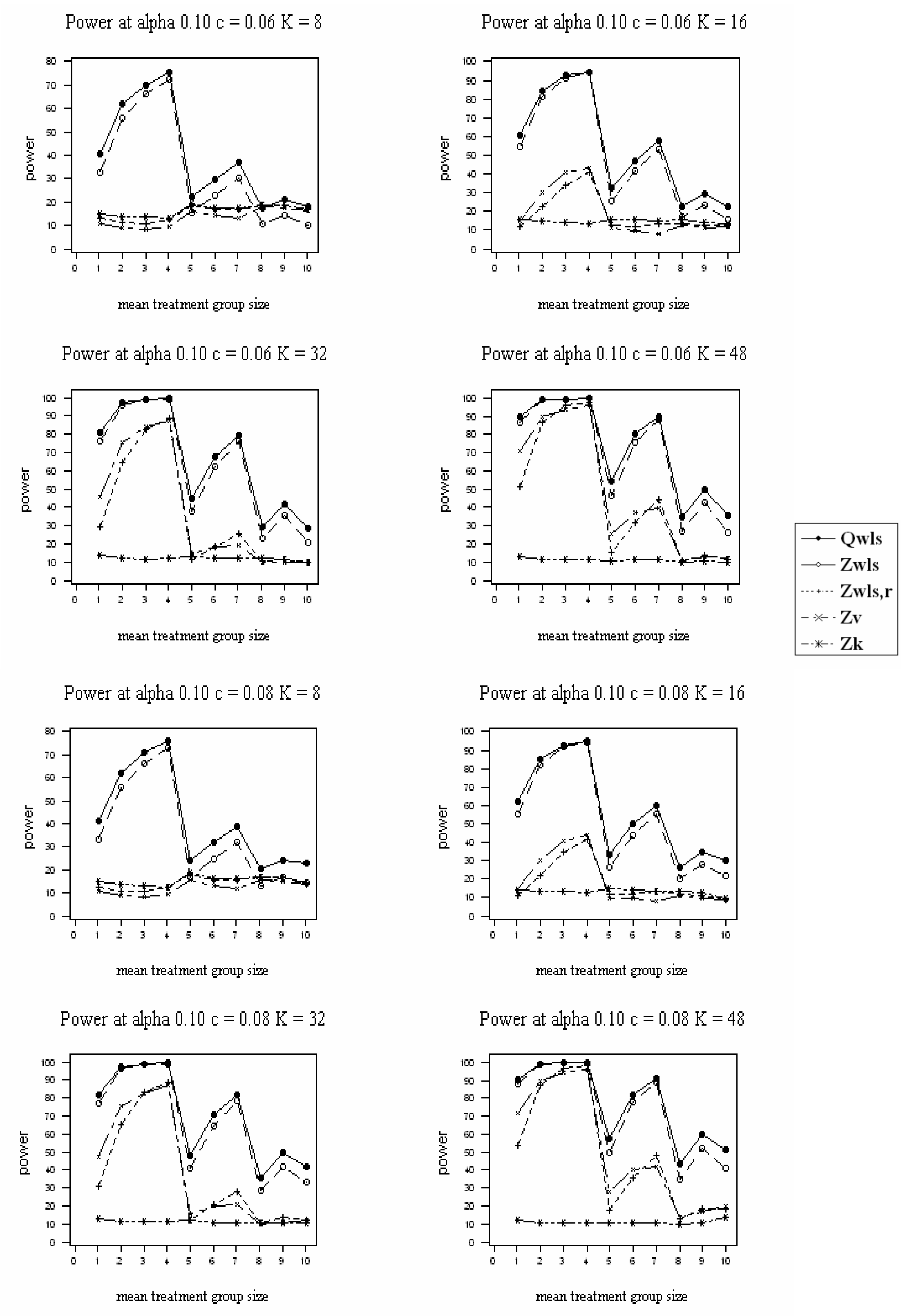


Figure 1D Comparison of the power of tests by *c*, 0.10 level of significance and number of center (*K*) where mean treatment group sizes are unequal. (Continued)

- **one-sided test**

Table 1.5 Mean ranking of power of tests in each number of centers, at the 0.01 significance level (one-sided) and $n_{i1} = n_{i2}$.

c	*K*	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.2 (.73)	3.2 (.73)	2.2 (.49)	4.0 (.45)	3.4 (.60)
	32	3.4 (.68)	2.4 (.68)	3.2 (.49)	4.2 (.49)	1.8 (.49)
	48	3.2 (.74)	2.2 (.74)	3.4 (.60)	4.0 (.45)	2.2 (.49)
0.04	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.4 (.40)	2.6 (.60)	2.6 (.40)	4.7 (.20)	3.7 (.44)
	32	2.4 (.40)	1.4 (.40)	3.6 (.40)	5.0 (.00)	2.6 (.40)
	48	2.0 (.00)	1.0 (.00)	4.4 (.25)	4.6 (.25)	3.0 (.00)
0.06	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.0 (.00)	2.0 (.00)	3.0 (.00)	5.0 (.00)	4.0 (.00)
	32	2.0 (.00)	1.0 (.00)	4.0 (.00)	5.0 (.00)	3.0 (.00)
	48	2.0 (.00)	1.0 (.00)	4.2 (.20)	4.8 (.20)	3.0 (.00)
0.08	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.0 (.00)	2.0 (.00)	3.0 (.00)	4.8 (.20)	4.2 (.20)
	32	2.0 (.00)	1.0 (.00)	4.0 (.00)	5.0 (.00)	3.0 (.00)
	48	2.0 (.00)	1.0 (.00)	4.0 (.32)	4.6 (.25)	3.4 (.40)

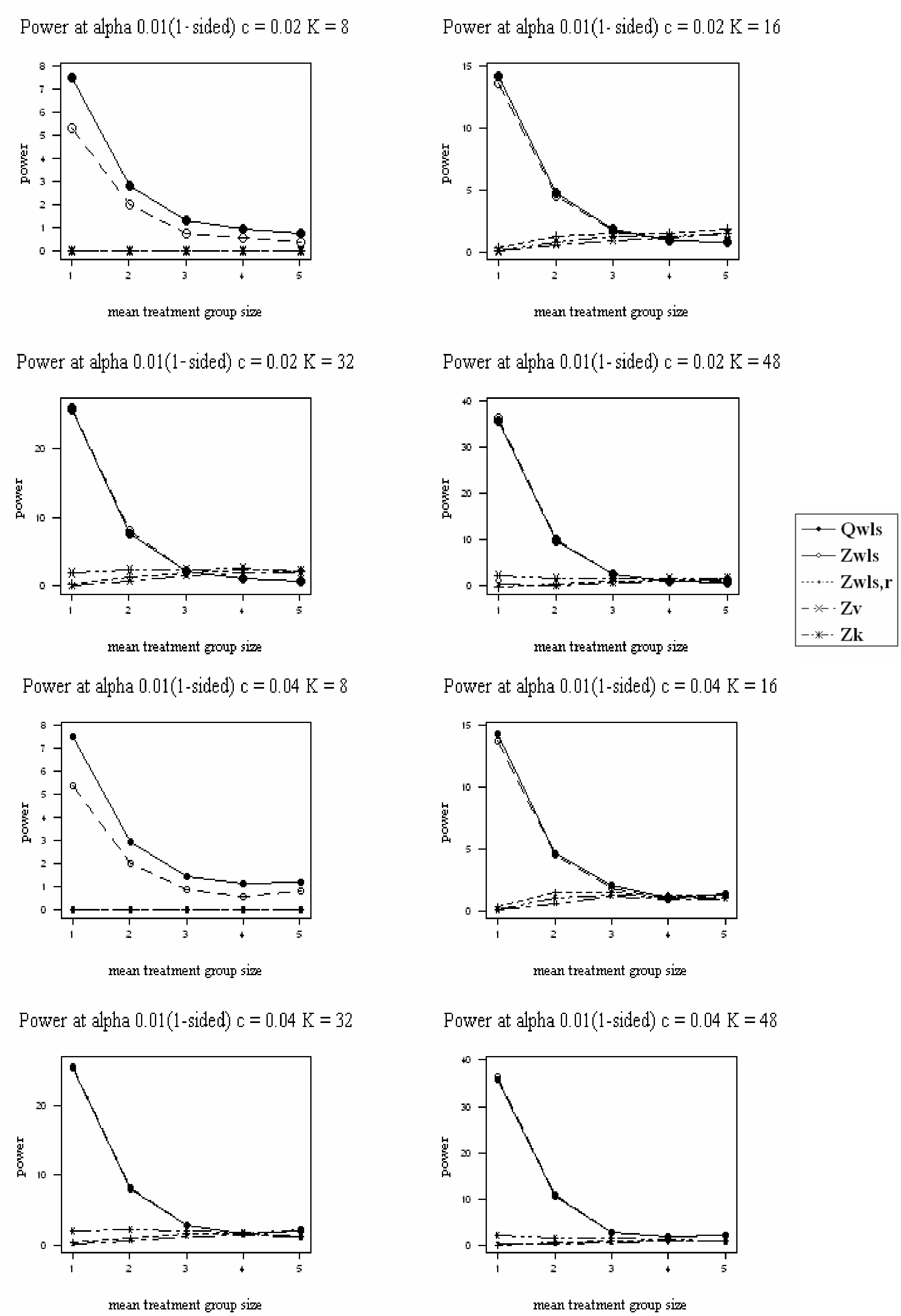


Figure 1E Comparison of the power of tests by *c*, at 0.01 level of significance for one-sided and number of center (*K*) where mean treatment group sizes are equal.

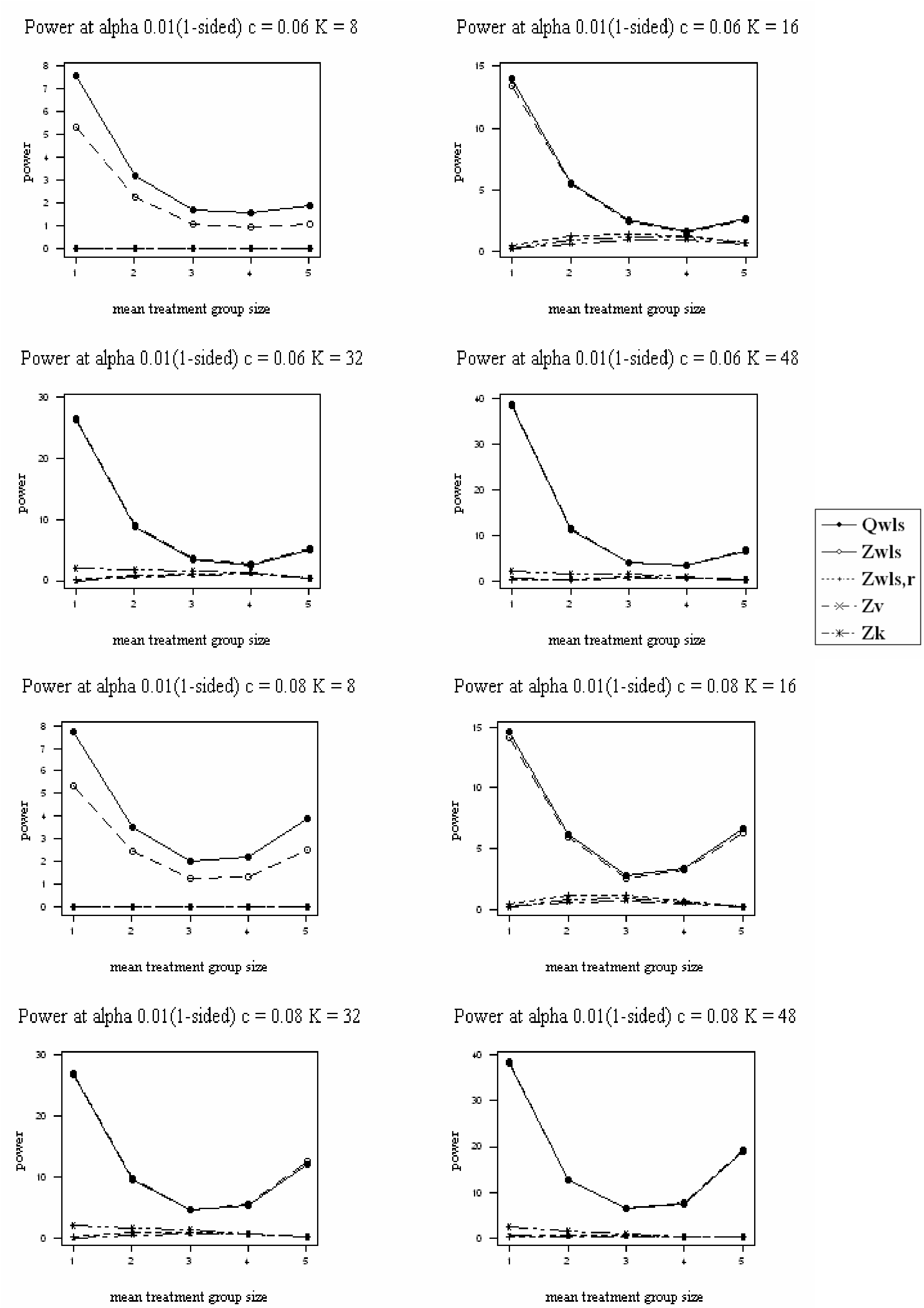


Figure 1E Comparison of the power of tests by *c*, at 0.01 level of significance for one-sided and number of center (*K*) where mean treatment group sizes are equal. (Continued)

Table 1.6 Mean ranking of power of tests in each number of centers, at the 0.10 significance level (one-sided) and $n_{i1} = n_{i2}$.

c	***K***	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	2.8 (.74)	4.2 (.58)	1.6 (.40)	3.6 (.40)	2.8 (.58)
	16	2.8 (.74)	3.8 (.74)	2.8 (.49)	3.8 (.49)	1.8 (.49)
	32	2.0 (.63)	3.4 (.68)	3.4 (.60)	4.2 (.37)	2.0 (.45)
	48	1.2 (.20)	2.8 (.58)	4.2 (.37)	4.2 (.37)	2.6 (.51)
0.04	8	3.4 (.60)	4.4 (.60)	1.8 (.80)	3.2 (.20)	2.2 (.20)
	16	2.0 (.63)	3.8 (.74)	3.2 (.49)	4.0 (.45)	2.0 (.45)
	32	1.0 (.00)	2.0 (.00)	4.2 (.20)	4.8 (.20)	3.0 (.00)
	48	1.0 (.00)	2.0 (.00)	4.2 (.20)	4.4 (.40)	3.4 (.40)
0.06	8	1.4 (.25)	3.4 (.68)	2.0 (.45)	4.4 (.25)	3.8 (.37)
	16	1.0 (.00)	3.0 (.63)	3.6 (.25)	4.8 (.20)	2.6 (.25)
	32	1.2 (.20)	1.8 (.20)	4.0 (.32)	4.6 (.25)	3.4 (.40)
	48	1.0 (.00)	2.0 (.00)	4.2 (.20)	4.0 (.45)	3.8 (.49)
0.08	8	1.0 (.00)	3.0 (.55)	2.4 (.25)	4.6 (.25)	4.0 (.32)
	16	1.0 (.00)	2.0 (.00)	4.0 (.00)	5.0 (.00)	3.0 (.00)
	32	1.0 (.00)	2.0 (.00)	4.0 (.32)	4.6 (.25)	3.4 (.40)
	48	1.4 (.25)	3.2 (.20)	4.8 (.20)	4.0 (.32)	1.6 (.25)

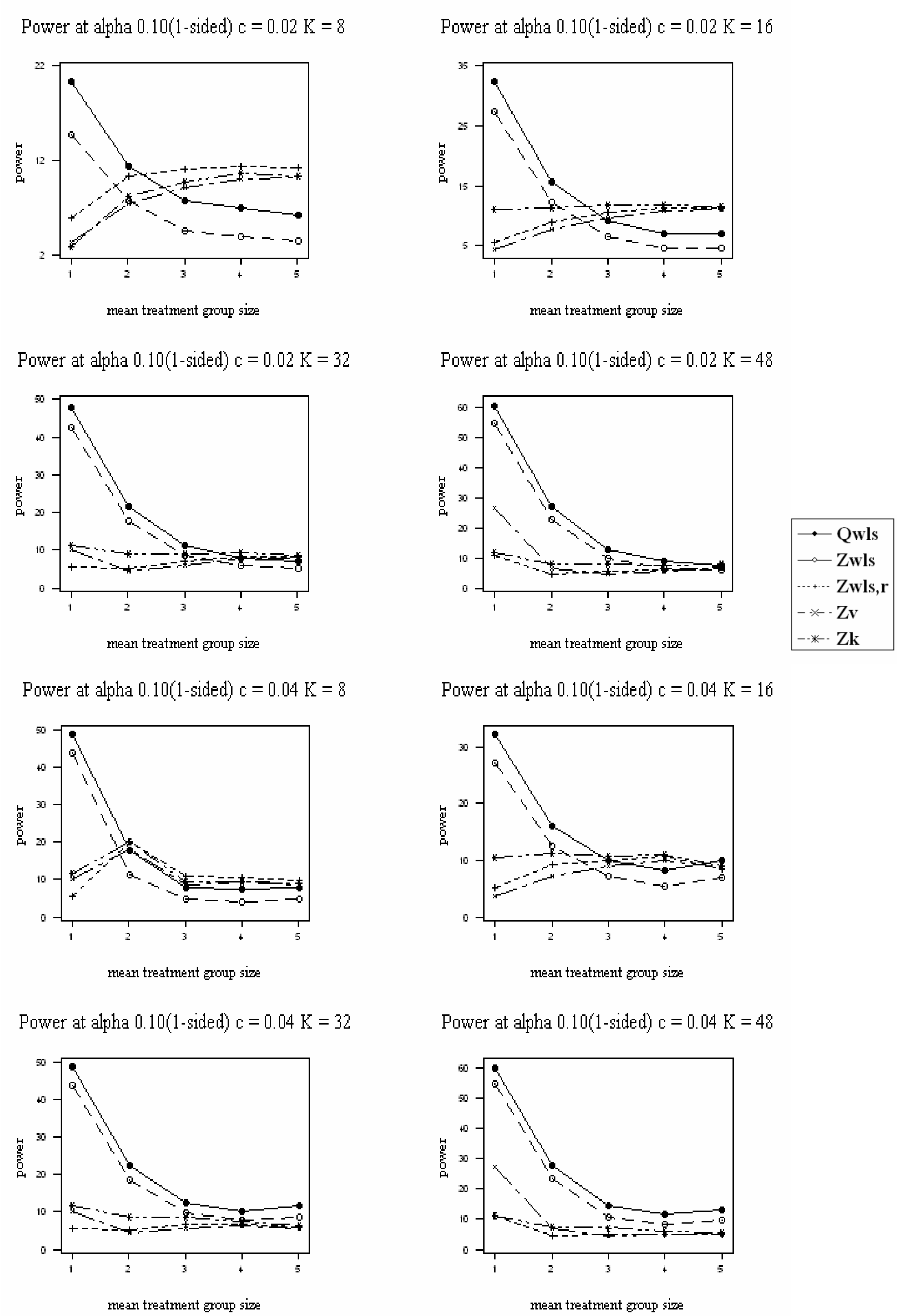


Figure 1F Comparison of the power of tests by *c*, 0.10 level of significance for one-sided and number of center (*K*) where mean treatment group are equal.

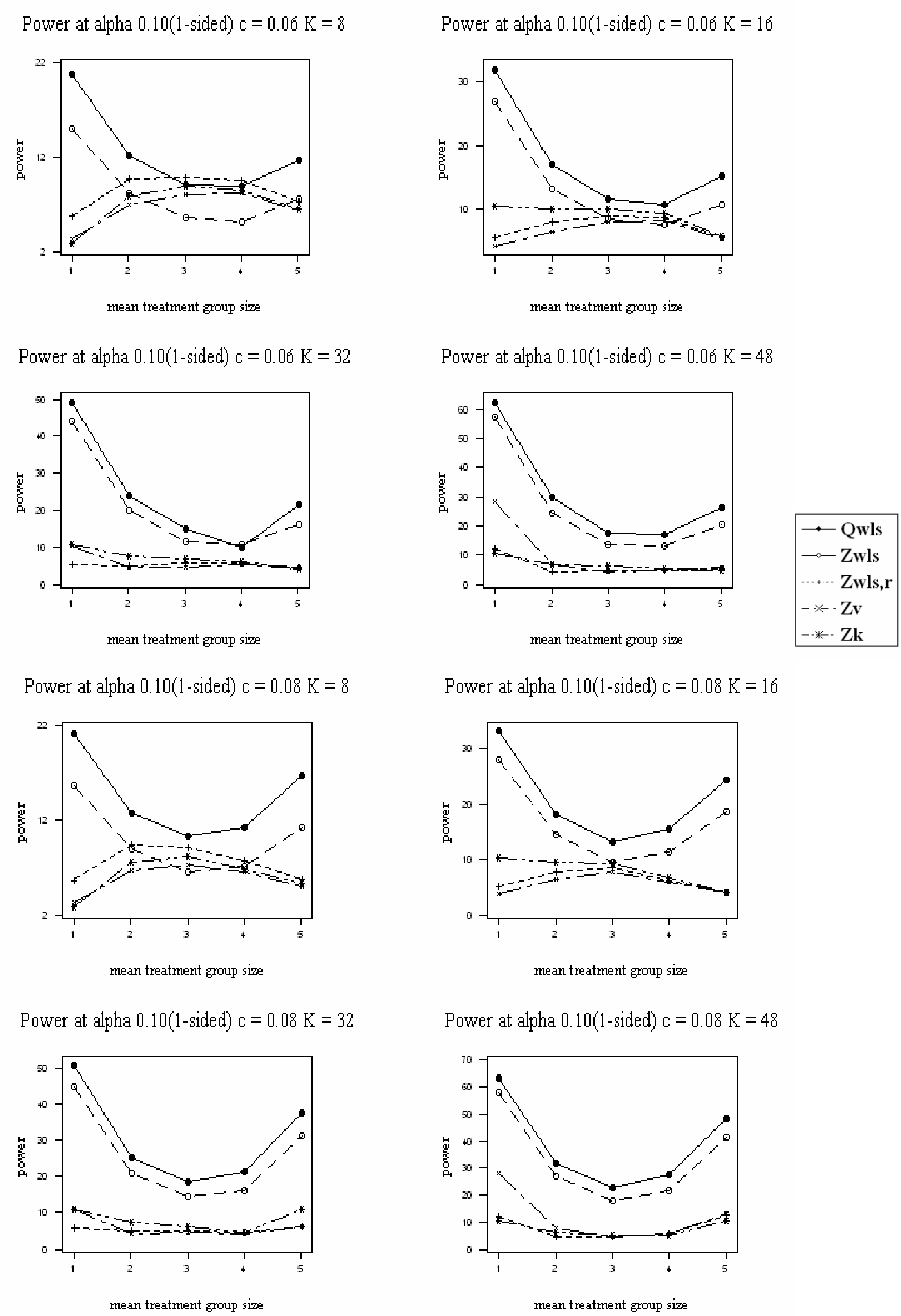


Figure 1F Comparison of the power of tests by *c*, 0.10 level of significance for one-sided and number of center (*K*) where mean treatment group are equal. (Continued)

Table 1.7 Mean ranking of power of tests in each number of centers, at the 0.01 significance level (one-sided) and $n_{i1} \neq n_{i2}$.

c	*K*	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.3 (.30)	2.3 (.30)	3.1 (.28)	4.8 (.20)	3.5 (.22)
	32	2.3 (.30)	1.3 (.30)	3.9 (.23)	4.2 (.33)	3.3 (.37)
	48	2.0 (.13)	1.2 (.11)	4.1 (.10)	4.2 (.33)	3.5 (.40)
0.04	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.0 (.00)	2.0 (.00)	3.3 (.15)	5.0 (.10)	3.7 (.15)
	32	2.0 (.00)	1.0 (.00)	4.2 (.13)	4.3 (.30)	3.5 (.27)
	48	1.9 (.07)	1.1 (.07)	4.1 (.10)	4.2 (.33)	3.7 (.30)
0.06	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.0 (.00)	2.0 (.00)	3.3 (.15)	5.0 (.00)	3.7 (.15)
	32	1.9 (.07)	1.1 (.07)	4.1 (.10)	4.4 (.31)	3.5 (.27)
	48	1.9 (.07)	1.1 (.07)	4.2 (.11)	4.2 (.32)	3.7 (.30)
0.08	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	1.0 (.00)	2.0 (.00)	3.3 (.15)	5.0 (.00)	3.7 (.15)
	32	2.0 (.00)	1.0 (.00)	4.2 (.13)	4.3 (.30)	3.5 (.27)
	48	1.9 (.07)	1.1 (.07)	4.1 (.10)	4.2 (.33)	3.7 (.30)

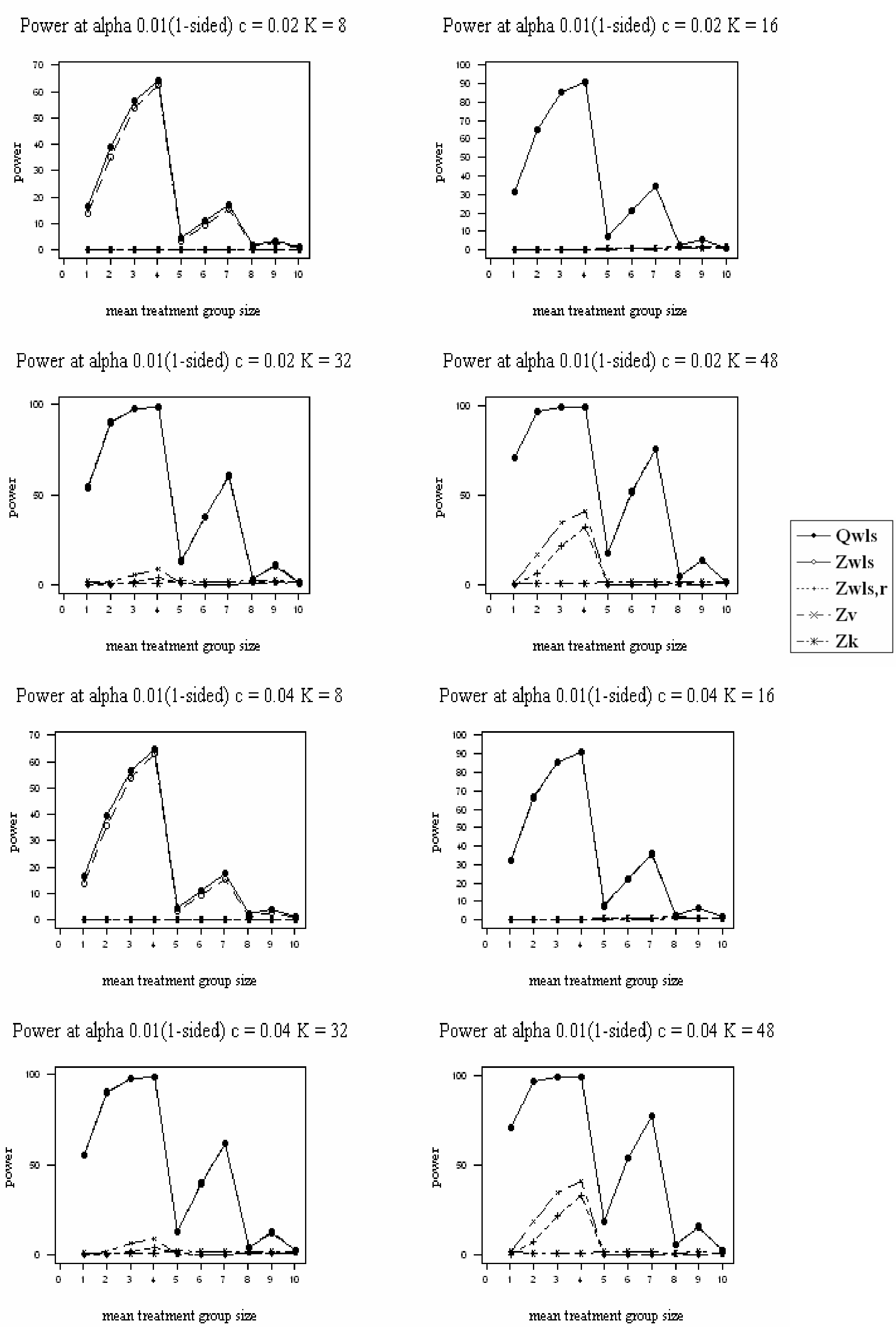


Figure 1G Comparison of the power of tests by *c*, 0.01 level of significance for one-sided and number of center (*K*) where mean treatment group are unequal.

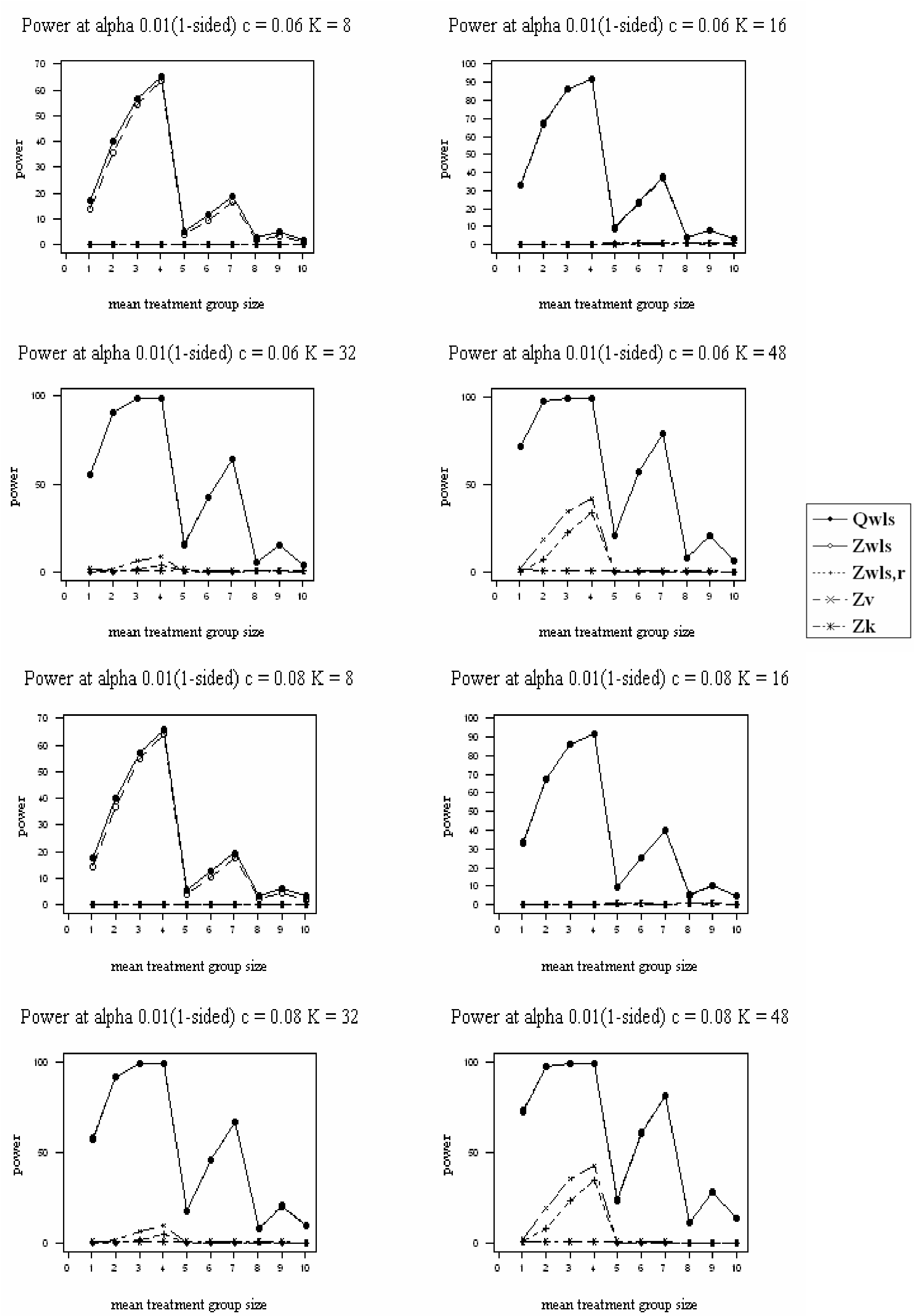


Figure 1G Comparison of the power of tests by *c*, 0.01 level of significance for one-sided and number of center (*K*) where mean treatment group are unequal. (Continued)

Table 1.8 Mean ranking of power of tests in each number of centers, at the 0.10 significance level (one-sided) and $n_{i1} \neq n_{i2}$.

c	***K***	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.7 (.40)	3.0 (.45)	2.4 (.28)	4.5 (.27)	3.5 (.27)
	16	1.4 (.31)	2.6 (.40)	3.8 (.25)	4.2 (.29)	3.0 (.42)
	32	1.0 (.00)	2.2 (.20)	3.8 (.13)	4.3 (.30)	3.7 (.37)
	48	1.0 (.00)	2.0 (.00)	3.8 (.20)	4.0 (.26)	4.2 (.33)
0.04	8	1.4 (.31)	3.2 (.49)	3.1 (.41)	4.5 (.22)	2.8 (.13)
	16	1.0 (.00)	2.5 (.34)	3.9 (.19)	4.4 (.26)	3.2 (.33)
	32	1.0 (.00)	2.0 (.00)	3.8 (.13)	4.3 (.33)	3.9 (.32)
	48	1.3 (.25)	2.1 (.16)	3.6 (.34)	3.7 (.34)	4.3 (.30)
0.06	8	1.0 (.00)	2.6 (.40)	2.9 (.18)	4.8 (.13)	3.7 (.15)
	16	1.0 (.00)	2.0 (.00)	4.1 (.10)	4.4 (.31)	3.5 (.27)
	32	1.0 (.00)	2.0 (.00)	3.8 (.13)	4.3 (.33)	3.9 (.32)
	48	1.1 (.07)	1.9 (.07)	3.7 (.21)	3.9 (.28)	4.4 (.27)
0.08	8	1.0 (.00)	2.0 (.00)	3.1 (.10)	5.0 (.00)	3.9 (.10)
	16	1.0 (.00)	2.0 (.00)	4.1 (.10)	4.4 (.31)	3.5 (.27)
	32	1.0 (.00)	2.0 (.00)	3.6 (.16)	4.3 (.30)	4.1 (.28)
	48	1.1 (.07)	1.9 (.07)	3.4 (.16)	3.6 (.16)	5.0 (.00)

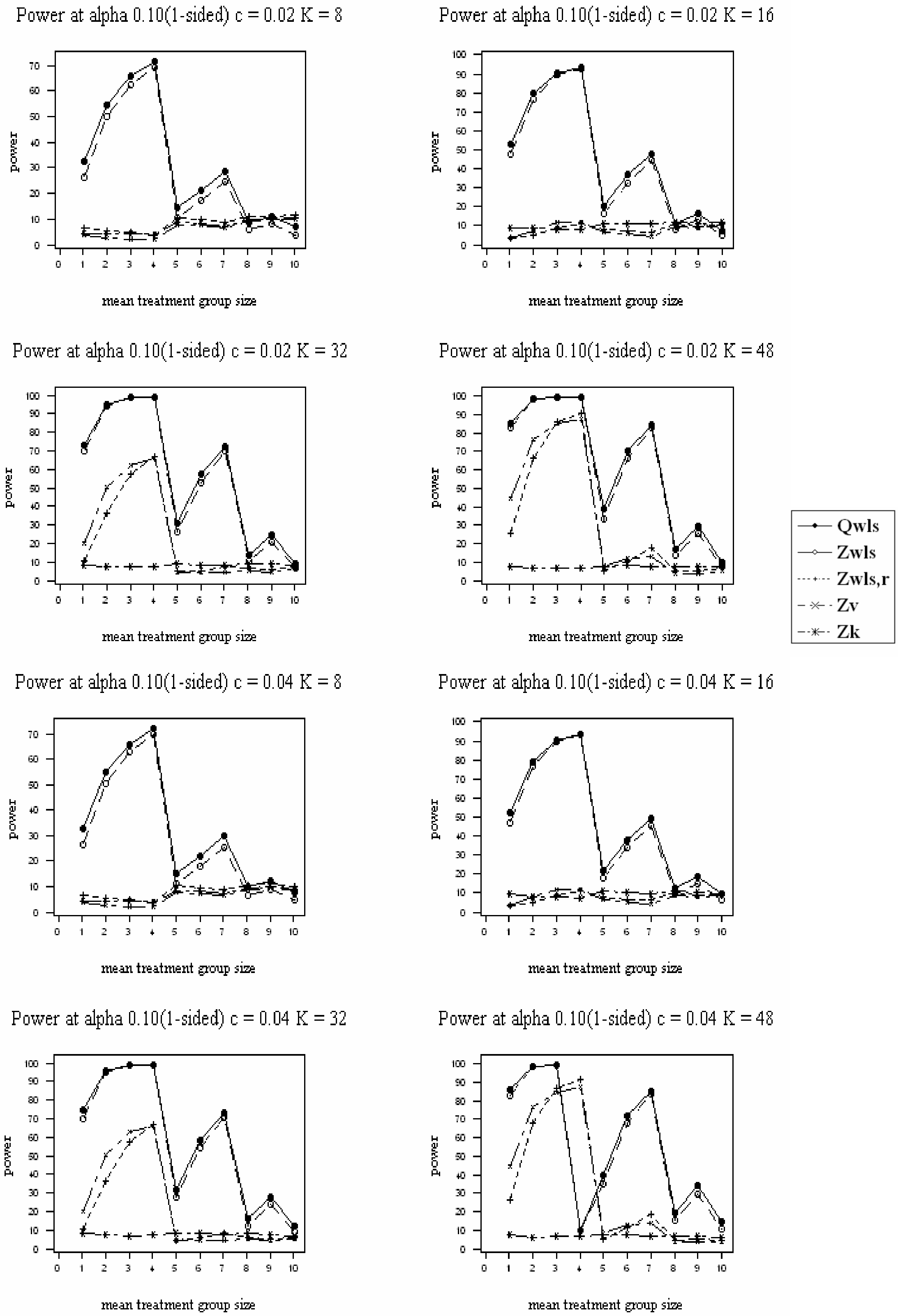


Figure 1H Comparison of the power of tests by *c*, 0.10 level of significance for one-sided and number of center (*K*) where mean treatment group are unequal.

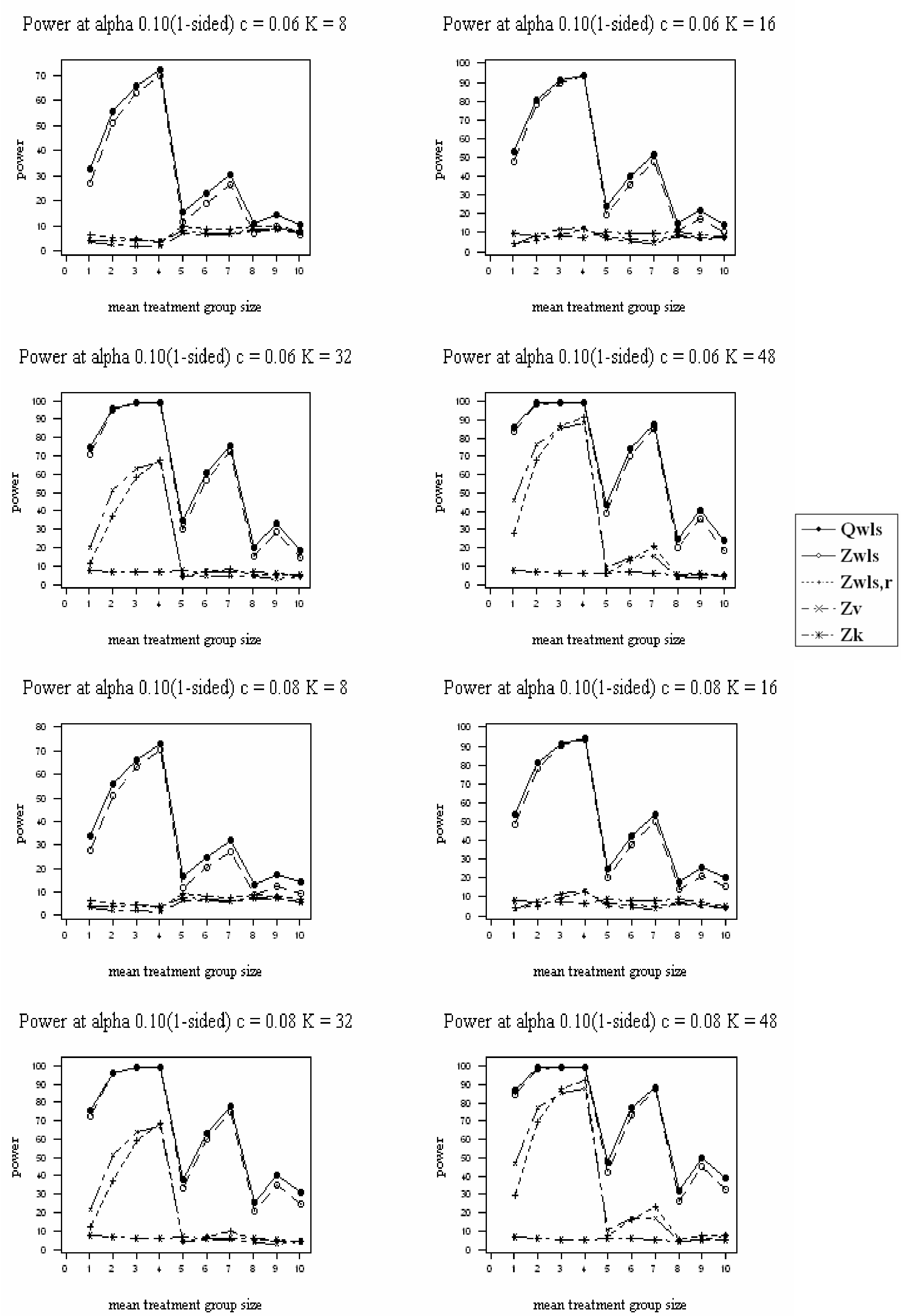


Figure 1H Comparison of the power of tests by *c*, 0.10 level of significance for one-sided and number of center (*K*) where mean treatment group are unequal.(Continued)

2. Adding constant in X_{ij} = 0 or X_{ij} = n_{ij} at 0.01 and 0.10 significance level

- **two-sided test**

Table 2.1 Mean ranking of power of tests in each number of centers, at the 0.01 significance level and $n_{i1} = n_{i2}$.

c	*K*	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	4.0 (.00)	5.0 (.00)	1.0 (.00)	3.0 (.00)	2.0 (.00)
	32	4.2 (.20)	4.8 (.20)	2.0 (.00)	3.0 (.00)	1.0 (.00)
	48	3.6 (.51)	4.2 (.37)	2.4 (.40)	3.8 (.49)	1.0 (.00)
0.04	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	3.8 (.20)	5.0 (.00)	1.0 (.00)	3.2 (.20)	2.0 (.00)
	32	2.8 (.73)	3.6 (.51)	3.2 (.58)	4.0 (.45)	1.4 (.40)
	48	2.4 (.87)	4.3 (.63)	3.4 (.60)	4.0 (.45)	2.2 (.49)
0.06	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.4 (.68)	4.2 (.58)	1.6 (.40)	4.0 (.45)	2.8 (.37)
	32	2.4 (.87)	3.0 (.63)	3.4 (.60)	4.0 (.45)	2.2 (.49)
	48	2.4 (.87)	3.0 (.63)	3.2 (.49)	4.2 (.49)	2.2 (.49)
0.08	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.2 (.74)	3.2 (.74)	2.6 (.75)	4.0 (.45)	3.0 (.45)
	32	2.4 (.87)	3.0 (.63)	3.2 (.49)	4.2 (.49)	2.2 (.49)
	48	2.4 (.87)	3.0 (.63)	3.0 (.45)	4.0 (.45)	2.6 (.75)

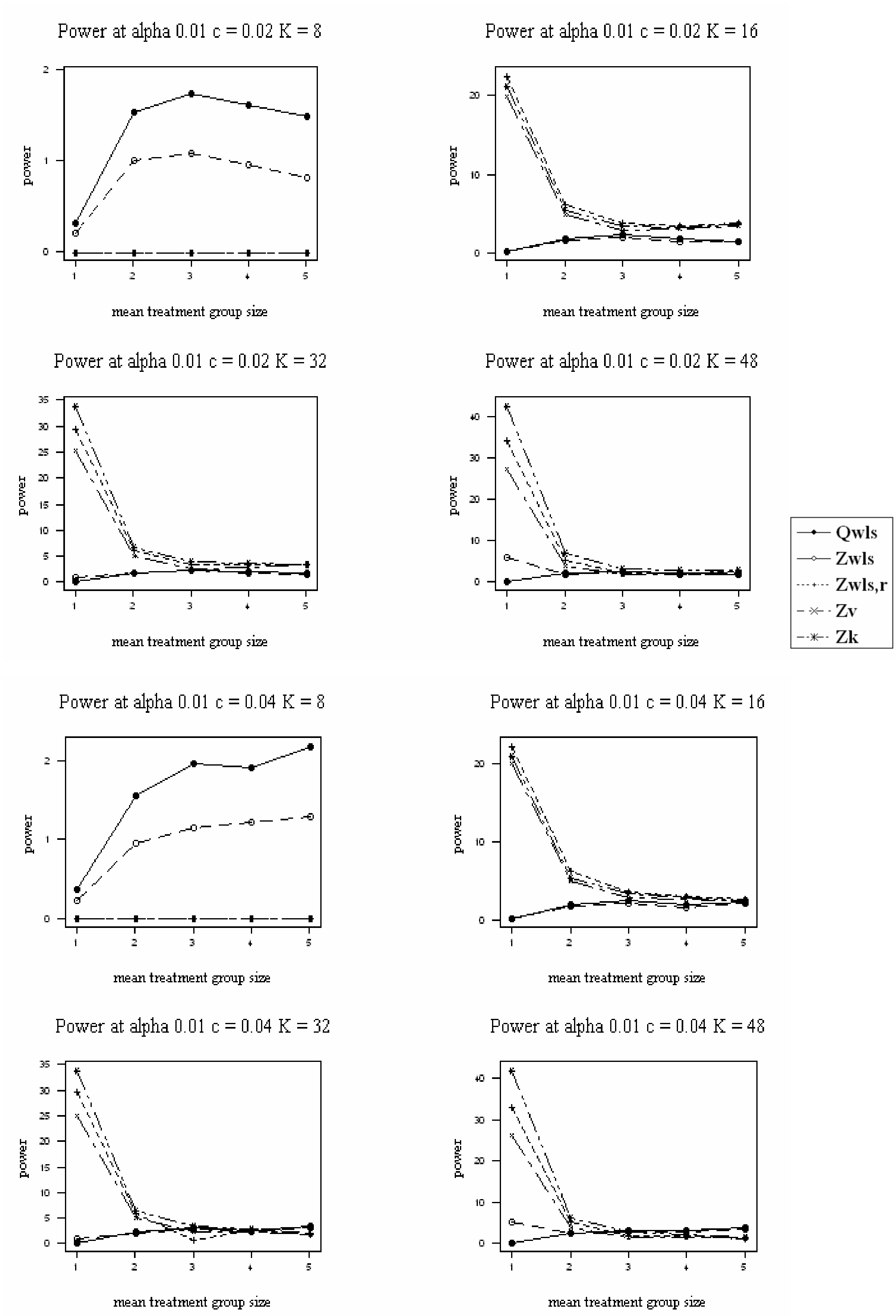


Figure 2A Comparison of the power of tests by *c*, 0.01 level of significance and number of center (*K*) where mean treatment group sizes are equal.

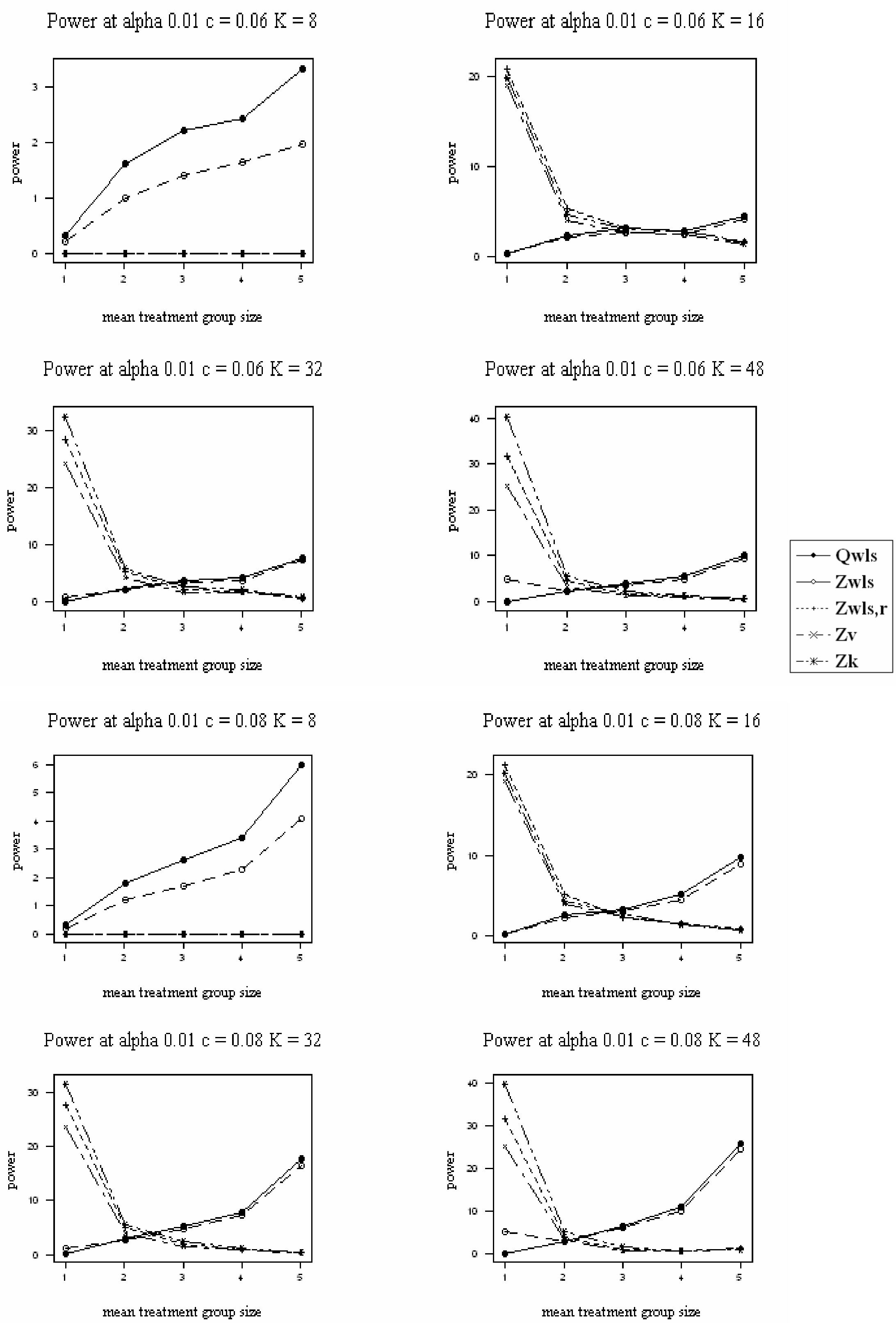


Figure 2A Comparison of the power of tests by *c*, 0.01 level of significance and number of center (*K*) where mean treatment group sizes are equal. (Continued)

Table 2.2 Mean ranking of power of tests in each number of centers, at the 0.10 significance level and $n_{i1} = n_{i2}$.

c	***K***	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	4.0 (.00)	5.0 (.00)	1.2 (.20)	3.0 (.00)	1.8 (.20)
	16	4.2 (.20)	4.8 (.20)	2.0 (.00)	3.0 (.00)	1.0 (.00)
	32	2.8 (.92)	4.6 (.25)	2.6 (.25)	3.6 (.25)	1.4 (.25)
	48	2.6 (.98)	3.8 (.37)	3.0 (.45)	4.0 (.45)	1.6 (.25)
0.04	8	4.0 (.00)	5.0 (.00)	1.2 (.20)	3.0 (.00)	1.8 (.20)
	16	3.6 (.68)	4.6 (.40)	2.8 (.49)	2.8 (.20)	1.2 (.20)
	32	2.4 (.87)	3.2 (.58)	3.2 (.49)	4.2 (.49)	2.0 (.45)
	48	2.6 (.98)	2.8 (.49)	3.4 (.60)	3.8 (.49)	2.4 (.60)
0.06	8	3.4 (.60)	5.0 (.00)	1.2 (.20)	3.2 (.20)	2.2 (.20)
	16	2.6 (.81)	4.2 (.58)	2.8 (.37)	3.8 (.37)	1.6 (.40)
	32	2.4 (.87)	3.0 (.63)	3.4 (.60)	3.6 (.40)	2.6 (.75)
	48	2.6 (.98)	2.8 (.49)	3.4 (.60)	3.2 (.20)	3.0 (.89)
0.08	8	2.6 (.68)	4.4 (.60)	1.6 (.40)	3.8 (.37)	2.6 (.40)
	16	2.4 (.87)	3.6 (.68)	2.8 (.37)	4.0 (.45)	2.2 (.58)
	32	2.6 (.98)	2.8 (.49)	3.0 (.45)	3.6 (.40)	3.0 (.89)
	48	2.0 (.78)	2.7 (.44)	3.7 (.49)	3.4 (.40)	3.2 (.92)

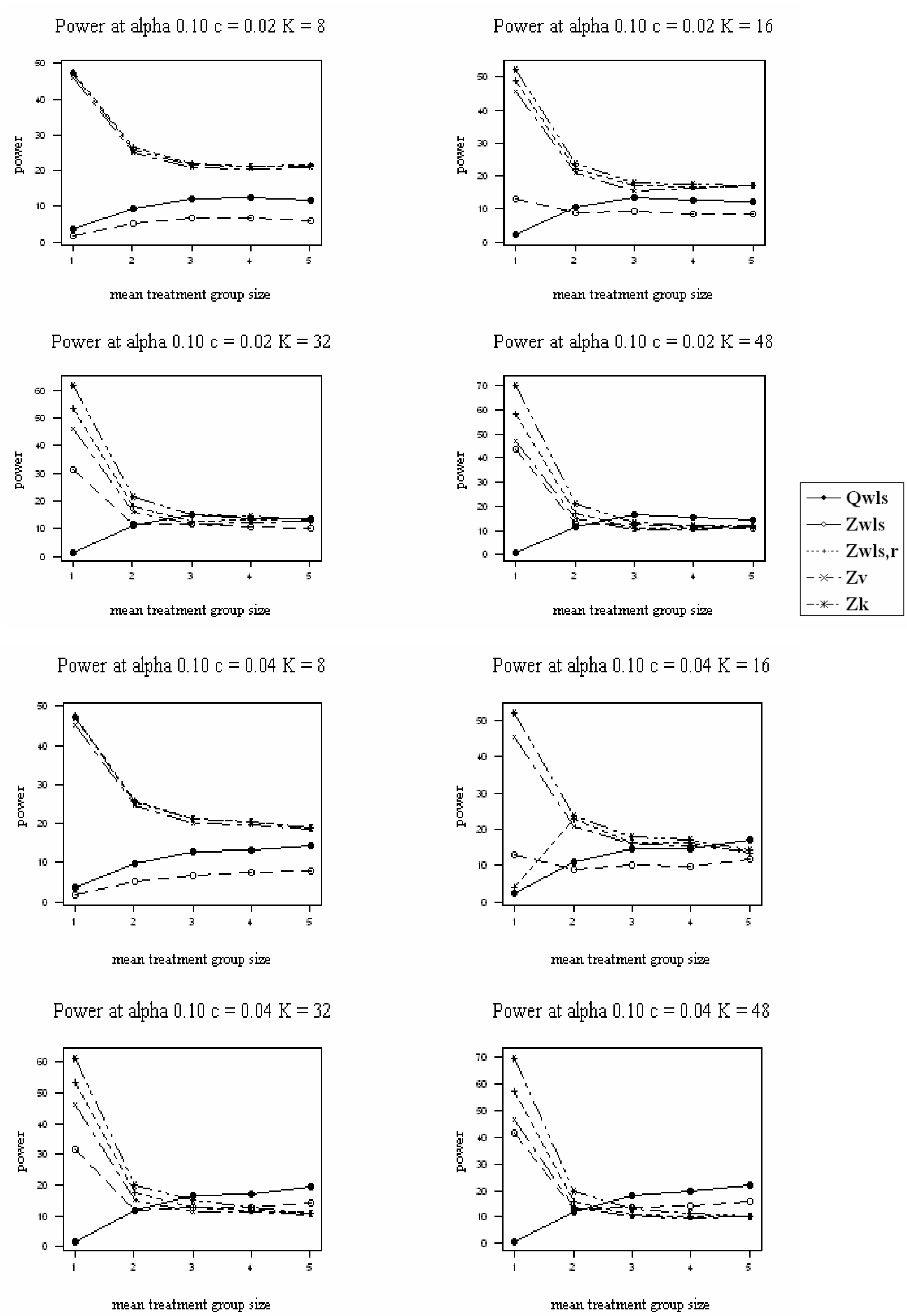


Figure 2B Comparison of the power of tests by *c*, 0.10 level of significance and number of center (*K*) where mean treatment group sizes are equal.

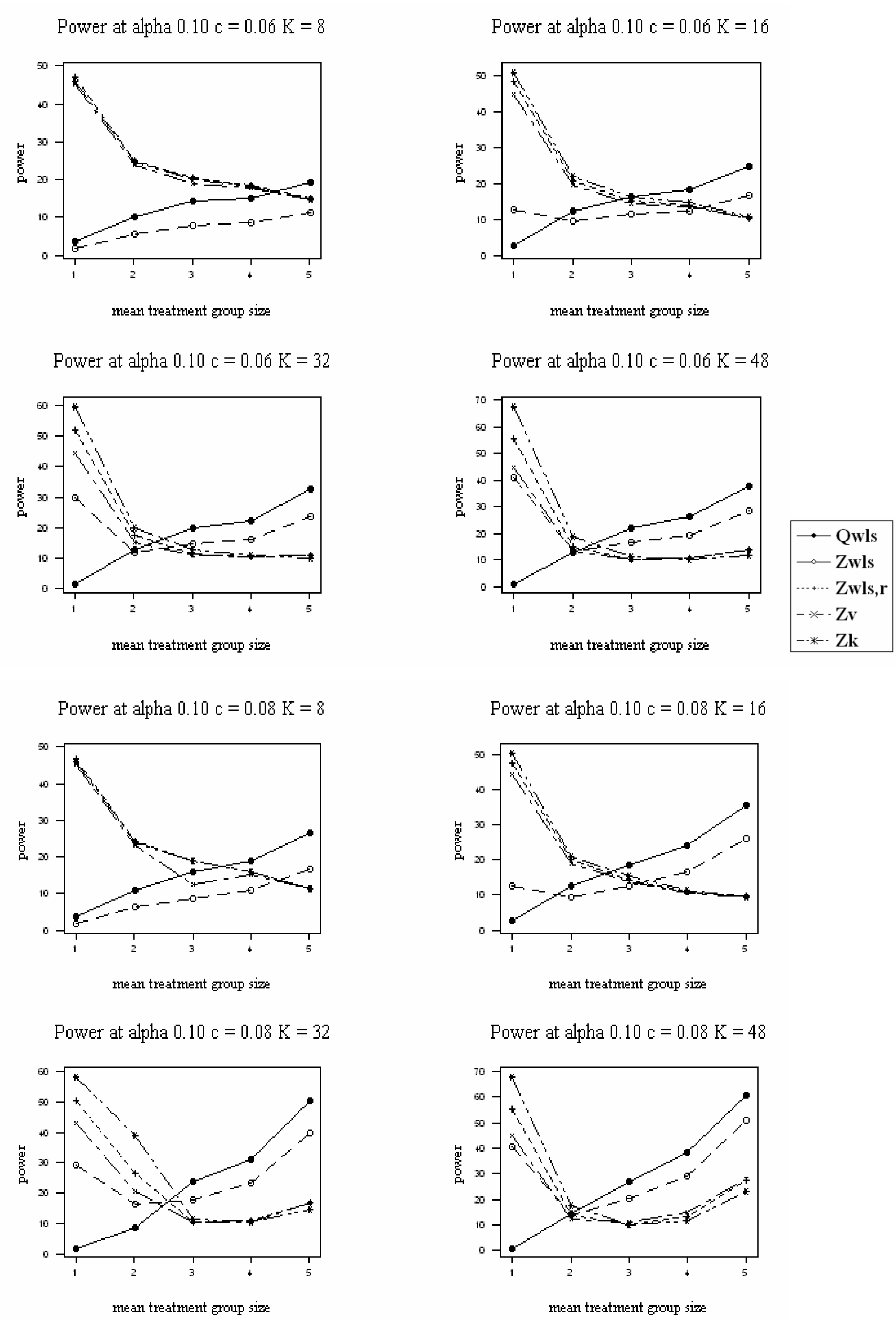


Figure 2B Comparison of the power of tests by *c*, 0.10 level of significance and number of center (*K*) where mean treatment group sizes are equal. (Continued)

Table 2.3 Mean ranking of power of tests in each number of centers, at the 0.01 significance level and $n_{i1} \neq n_{i2}$.

c	***K***	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	3.4 (.40)	4.4 (.40)	1.6 (.40)	3.4 (.27)	2.2 (.13)
	32	3.4 (.54)	3.7 (.34)	2.7 (.30)	3.8 (.33)	1.4 (.27)
	48	2.9 (.61)	3.1 (.32)	3.0 (.33)	4.2 (.33)	1.8 (.33)
0.04	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	3.1 (.38)	4.3 (.40)	1.6 (.34)	3.8 (.29)	2.2 (.25)
	32	2.9 (.64)	3.1 (.38)	3.0 (.33)	4.0 (.33)	2.0 (.33)
	48	2.8 (.63)	3.0 (.33)	3.0 (.33)	4.2 (.33)	2.0 (.33)
0.06	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.5 (.50)	3.6 (.48)	2.0 (.37)	4.0 (.33)	2.9 (.32)
	32	2.8 (.63)	3.0 (.33)	3.0 (.33)	4.2 (.33)	2.0 (.33)
	48	2.7 (.63)	2.9 (.32)	3.2 (.33)	4.2 (.33)	2.0 (.33)
0.08	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.5 (.50)	3.5 (.50)	2.4 (.40)	3.9 (.38)	2.7 (.34)
	32	2.6 (.65)	2.8 (.33)	3.2 (.33)	4.2 (.33)	2.2 (.33)
	48	2.6 (.65)	2.8 (.33)	3.1 (.32)	4.1 (.32)	2.4 (.43)

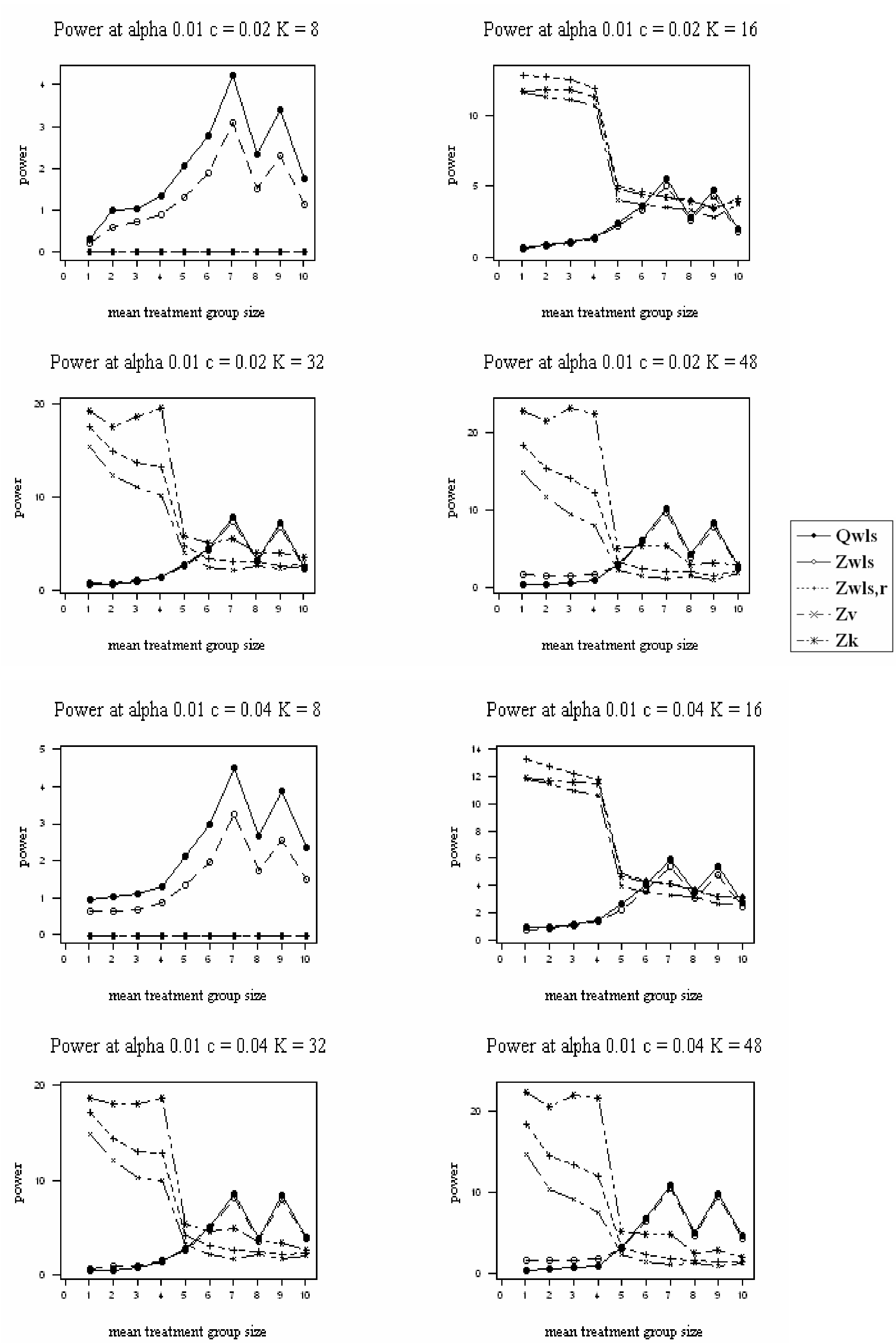


Figure 2C Comparison of the power of tests by *c*, at 0.01 level of significance and number of center (*K*) where mean treatment group sizes are unequal.

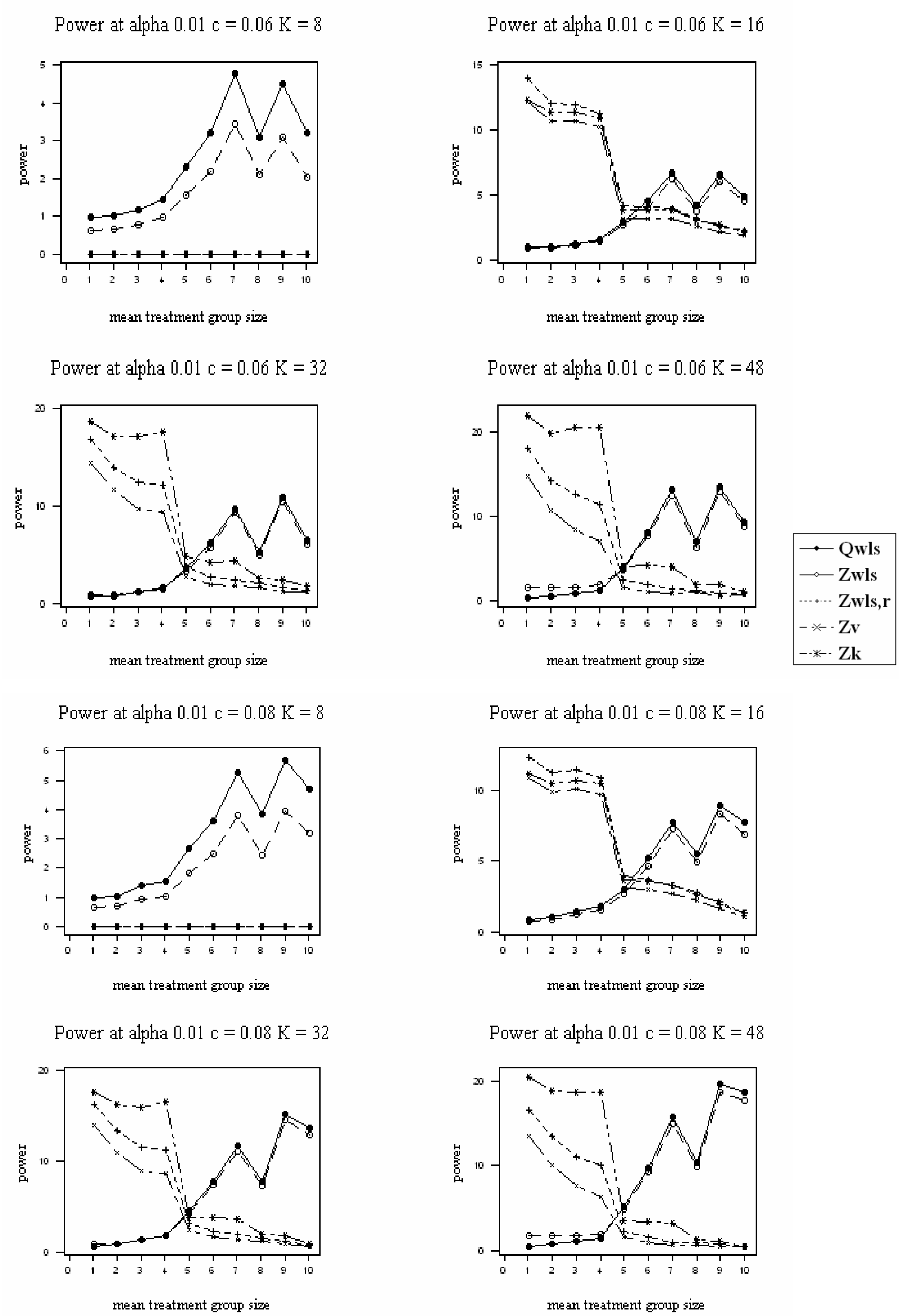


Figure 2C Comparison of the power of tests by *c*, at 0.01 level of significance and number of center (*K*) where mean treatment group sizes are unequal. (Continued)

Table 2.4 Mean ranking of power of tests in each number of centers, at the 0.10 significance level and $n_{i1} \neq n_{i2}$.

c	***K***	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	4.0 (.00)	5.0 (.00)	1.2 (.13)	3.0 (.00)	1.8 (.13)
	16	3.8 (.17)	4.5 (.17)	2.2 (.13)	3.4 (.22)	1.1 (.10)
	32	2.8 (.63)	3.5 (.31)	2.9 (.32)	4.1 (.32)	1.7 (.26)
	48	2.7 (.61)	2.9 (.28)	3.5 (.43)	4.0 (.26)	1.8 (.33)
0.04	8	4.0 (.00)	5.0 (.00)	1.1 (.10)	2.9 (.10)	2.0 (.15)
	16	3.1 (.21)	2.6 (.22)	2.6 (.22)	3.7 (.26)	1.3 (.15)
	32	2.7 (.63)	3.1 (.32)	3.1 (.32)	4.2 (.33)	1.9 (.32)
	48	2.7 (.63)	2.8 (.29)	3.6 (.45)	3.9 (.23)	2.0 (.33)
0.06	8	3.6 (.31)	5.0 (.00)	1.2 (.13)	3.1 (.18)	2.1 (.18)
	16	3.0 (.62)	3.7 (.34)	2.8 (.29)	3.9 (.32)	1.6 (.27)
	32	2.7 (.63)	2.9 (.32)	3.2 (.36)	4.1 (.32)	2.1 (.38)
	48	2.5 (.62)	2.7 (.26)	3.8 (.42)	3.3 (.26)	2.7 (.54)
0.08	8	3.1 (.46)	4.9 (.01)	1.6 (.27)	3.4 (.24)	2.1 (.16)
	16	2.8 (.63)	3.2 (.36)	3.0 (.33)	4.1 (.32)	1.9 (.32)
	32	2.6 (.65)	2.9 (.32)	3.5 (.43)	3.5 (.22)	3.5 (.50)
	48	2.6 (.65)	2.7 (.30)	3.4 (.40)	3.3 (.21)	3.0 (.58)

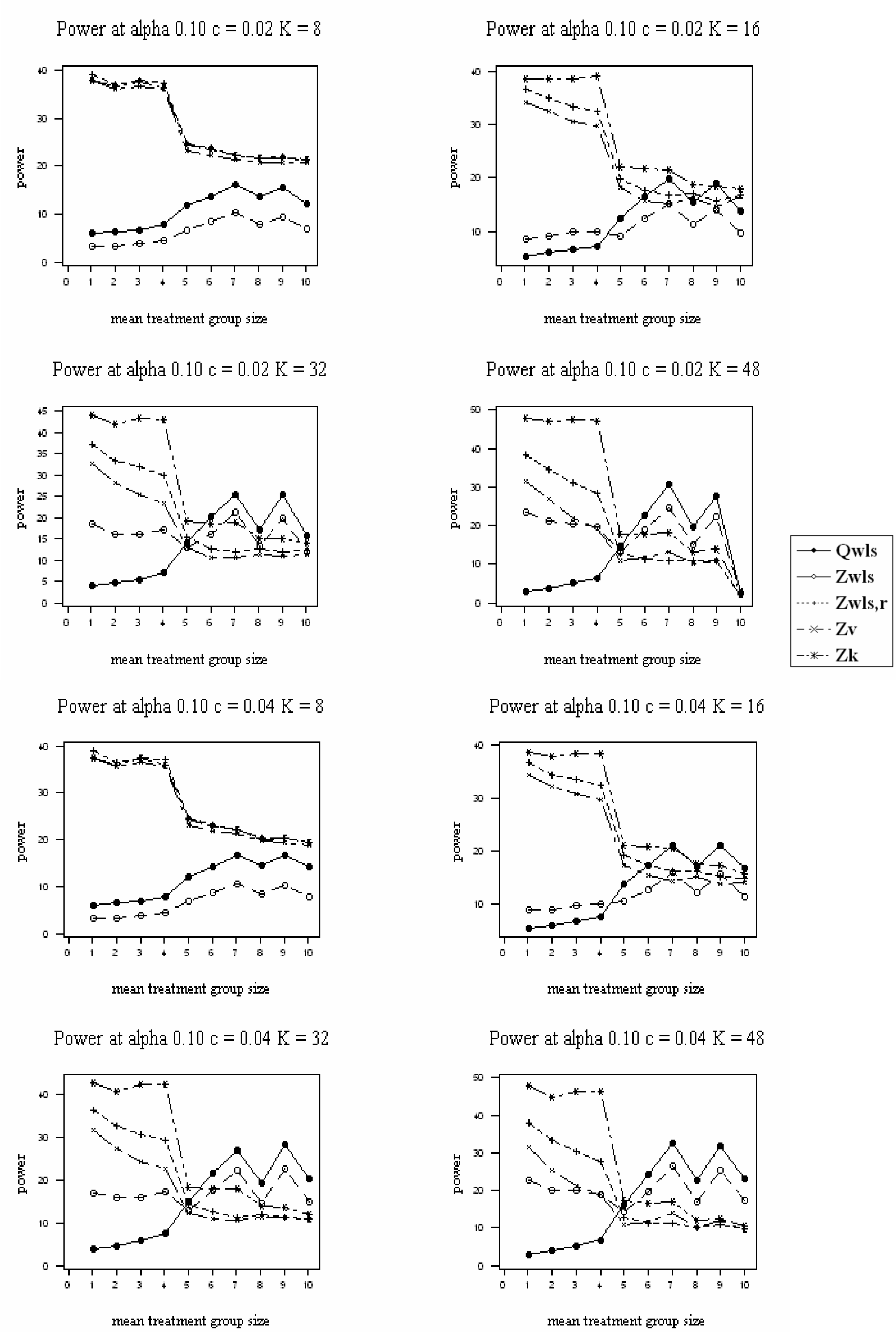


Figure 2D Comparison of the power of tests by *c*, at 0.10 level of significance and number of center (*K*) where mean treatment group sizes are unequal.

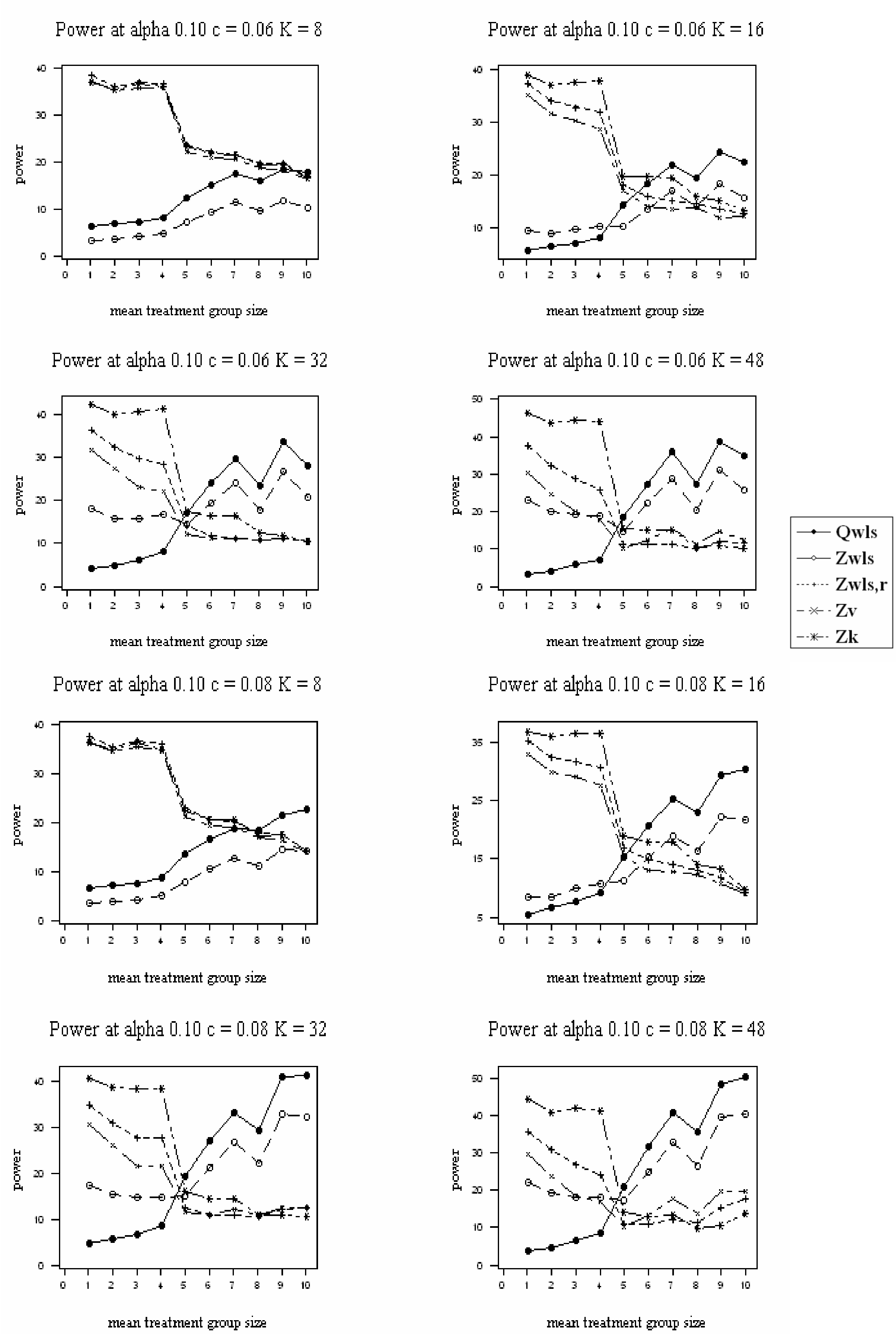


Figure 2D Comparison of the power of tests by *c*, at 0.10 level of significance and number of center (*K*) where mean treatment group sizes are unequal. (Continued)

- **One-sided test**

Table 2.5 Mean ranking of power of tests in each number of centers, at the 0.01 significance level (one-sided) and $n_{i1} = n_{i2}$.

c	***K***	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	3.6 (.40)	4.8 (.20)	1.0 (.00)	3.4 (.40)	2.2 (.20)
	32	4.9 (.10)	4.1 (.10)	2.0 (.00)	3.0 (.00)	1.0 (.00)
	48	4.6 (.40)	3.6 (.40)	2.4 (.40)	3.4 (.40)	1.0 (.00)
0.04	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	3.0 (.63)	4.2 (.58)	1.4 (.40)	3.8 (.49)	2.6 (.40)
	32	3.8 (.58)	2.8 (.58)	2.8 (.49)	4.2 (.49)	1.4 (.40)
	48	3.4 (.68)	2.2 (.74)	3.4 (.60)	4.0 (.45)	2.0 (.45)
0.06	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.3 (.80)	3.1 (.68)	2.2 (.49)	4.2 (.49)	3.2 (.49)
	32	3.2 (.74)	2.2 (.74)	3.2 (.49)	4.2 (.49)	2.2 (.49)
	48	3.2 (.74)	2.2 (.74)	3.2 (.49)	4.2 (.49)	2.2 (.49)
0.08	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.1 (.71)	3.1 (.68)	2.2 (.49)	4.2 (.37)	3.4 (.60)
	32	3.2 (.74)	2.2 (.74)	3.2 (.49)	4.2 (.49)	2.2 (.49)
	48	3.2 (.74)	2.2 (.74)	3.0 (.45)	4.0 (.45)	2.6 (.75)

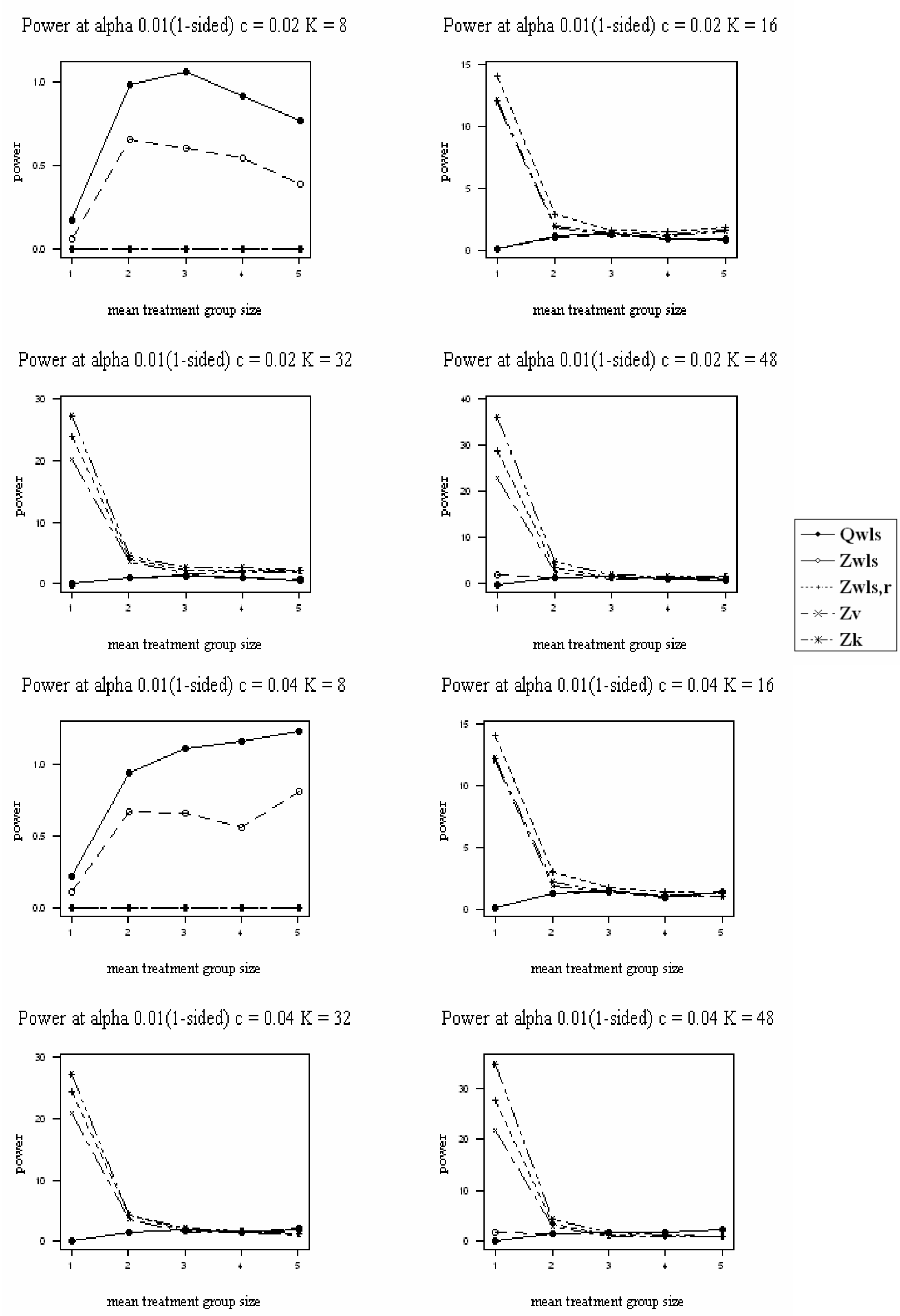


Figure 2E Comparison of the power of tests by *c*, 0.01 level of significance for one-sided and number of center (*K*) where mean treatment group are equal.

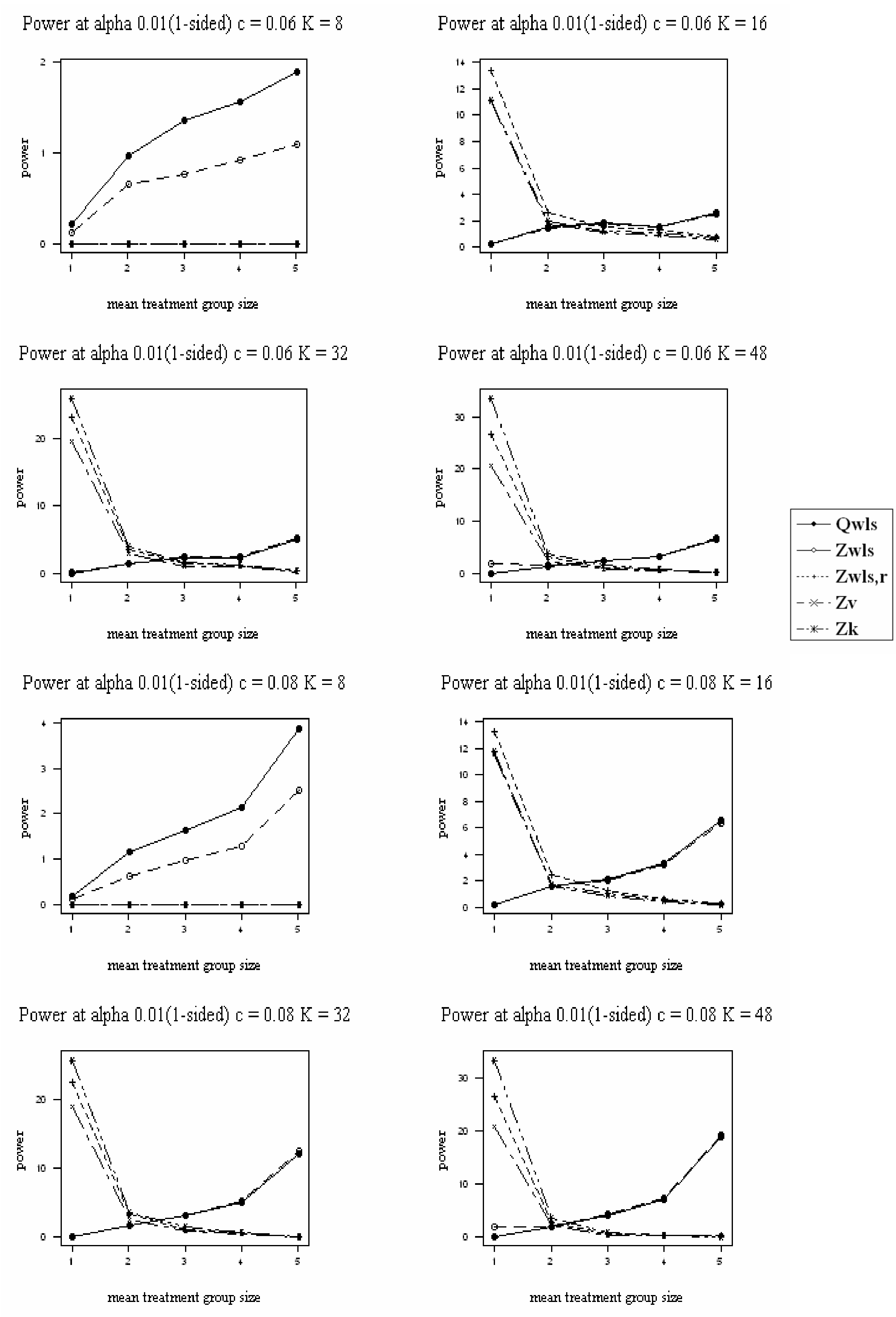


Figure 2E Comparison of the power of tests by *c*, 0.01 level of significance for one-sided and number of center (*K*) where mean treatment group are equal. (Continued)

Table 2.6 Mean ranking of power of tests in each number of centers, at the 0.10 significance level (one-sided) and $n_{i1} = n_{i2}$.

c	***K***	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	4.0 (.00)	5.0 (.00)	1.0 (.00)	2.8 (.20)	2.2 (.20)
	16	4.2 (.20)	4.8 (.20)	2.0 (.00)	3.0 (.00)	1.0 (.00)
	32	3.8 (.49)	4.8 (.20)	2.2 (.20)	3.2 (.20)	1.0 (.00)
	48	2.8 (.92)	3.8 (.37)	3.0 (.45)	4.0 (.45)	1.4 (.25)
0.04	8	4.0 (.00)	5.0 (.00)	1.0 (.00)	2.8 (.20)	2.2 (.20)
	16	3.4 (.68)	4.6 (.25)	3.0 (.63)	2.8 (.20)	1.2 (.20)
	32	2.4 (.87)	3.2 (.58)	3.2 (.49)	4.2 (.49)	2.0 (.45)
	48	2.4 (.87)	3.2 (.58)	3.2 (.49)	4.2 (.49)	2.0 (.45)
0.06	8	3.0 (.63)	4.6 (.40)	1.2 (.20)	3.4 (.51)	2.8 (.37)
	16	2.8 (.80)	4.2 (.58)	2.6 (.40)	3.8 (.37)	1.6 (.40)
	32	2.4 (.87)	3.0 (.63)	3.0 (.45)	4.0 (.45)	2.6 (.75)
	48	2.4 (.87)	3.0 (.63)	3.2 (.49)	3.8 (.49)	2.6 (.75)
0.08	8	2.4 (.68)	4.0 (.63)	1.6 (.40)	3.8 (.58)	3.2 (.37)
	16	2.4 (.87)	3.6 (.68)	3.0 (.45)	4.0 (.45)	2.0 (.45)
	32	2.4 (.87)	3.0 (.63)	3.0 (.45)	4.0 (.45)	2.6 (.75)
	48	2.2 (.80)	2.8 (.49)	3.2 (.49)	3.8 (.49)	3.0 (.89)

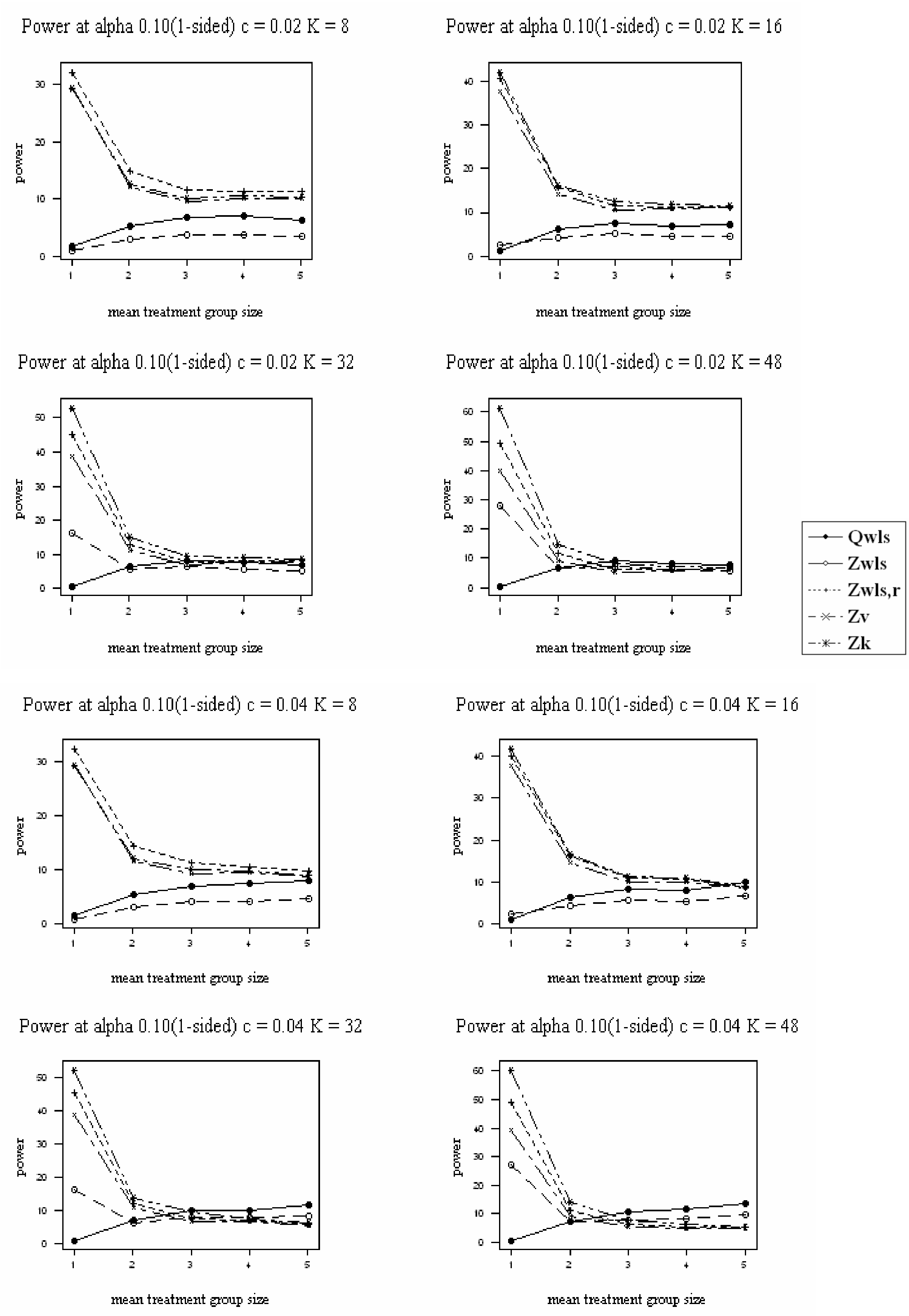


Figure 2F Comparison of the power of tests by *c*, 0.10 level of significance for one-sided and number of center (*K*) where mean treatment group are equal.

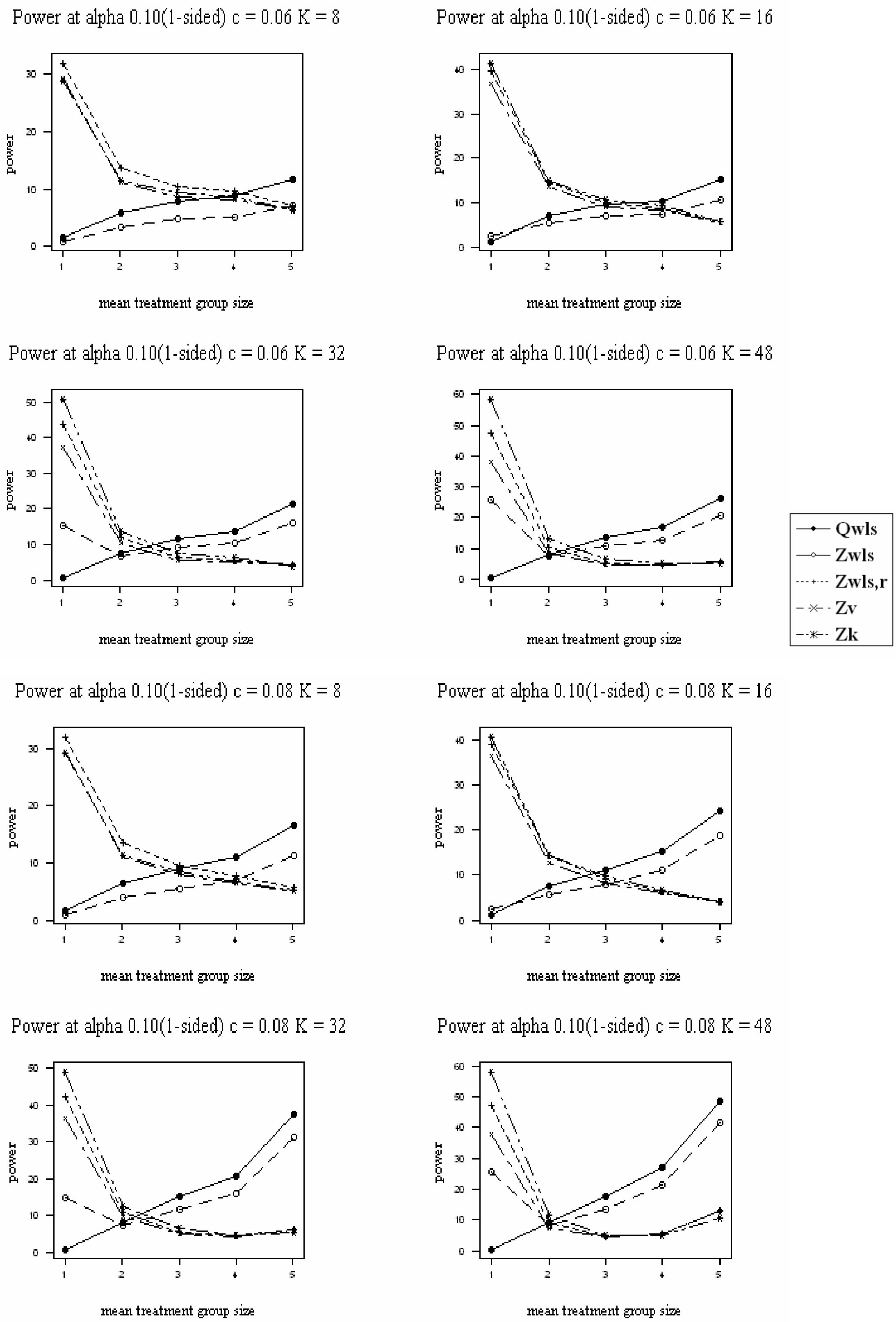


Figure 2F Comparison of the power of tests by *c*, 0.10 level of significance for one-sided and number of center (*K*) where mean treatment group are equal. (Continued)

Table 2.7 Mean ranking of power of tests in each number of centers, at the 0.01 significance level (one-sided) and $n_{i1} \neq n_{i2}$.

c	***K***	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	3.0 (.45)	4.1 (.41)	1.5 (.27)	3.4 (.44)	3.1 (.28)
	32	4.1 (.40)	3.1 (.44)	2.7 (.30)	3.8 (.33)	1.4 (.27)
	48	3.5 (.45)	2.5 (.45)	3.0 (.33)	4.2 (.33)	1.8 (.33)
0.04	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.5 (.50)	3.5 (.50)	2.0 (.33)	3.6 (.48)	3.4 (.22)
	32	3.5 (.50)	2.5 (.50)	3.1 (.32)	3.9 (.38)	2.0 (.33)
	48	3.4 (.48)	2.4 (.48)	3.0 (.33)	4.2 (.33)	2.0 (.33)
0.06	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.2 (.49)	3.2 (.49)	2.2 (.33)	3.8 (.49)	3.6 (.16)
	32	3.4 (.48)	2.4 (.47)	3.1 (.32)	4.2 (.33)	2.0 (.33)
	48	3.3 (.47)	2.3 (.47)	3.3 (.34)	4.2 (.32)	2.0 (.33)
0.08	8	1.0 (.00)	2.0 (.00)	4.0 (.00)	4.0 (.00)	4.0 (.00)
	16	2.3 (.51)	3.2 (.47)	2.2 (.33)	3.8 (.49)	3.6 (.16)
	32	3.2 (.49)	2.2 (.49)	3.2 (.33)	4.2 (.33)	2.2 (.33)
	48	3.2 (.49)	2.2 (.49)	3.2 (.33)	4.2 (.33)	2.2 (.33)

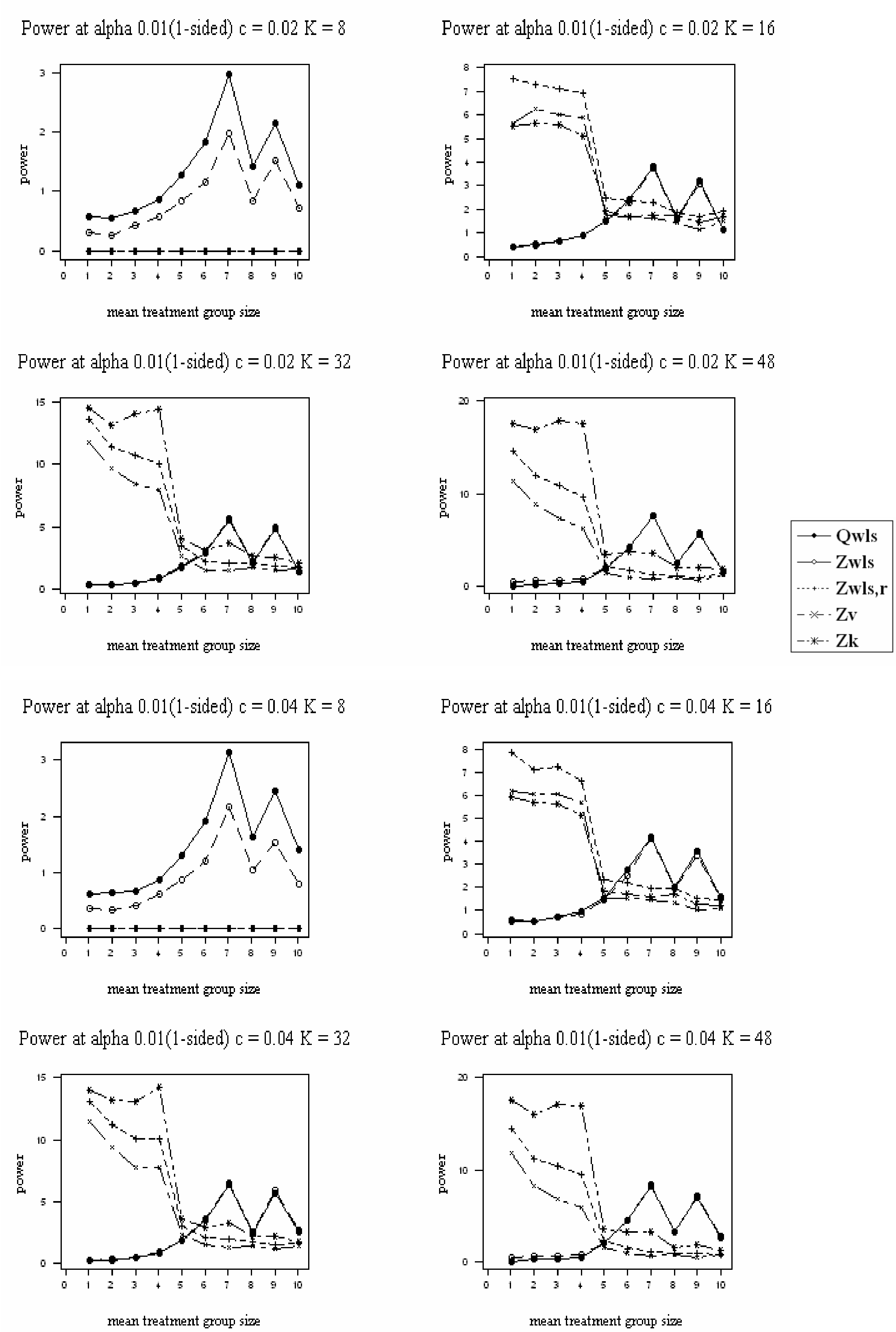


Figure 2G Comparison of the power of tests by *c*, 0.01 level of significance for one-sided and number of center (*K*) where mean treatment group are unequal.

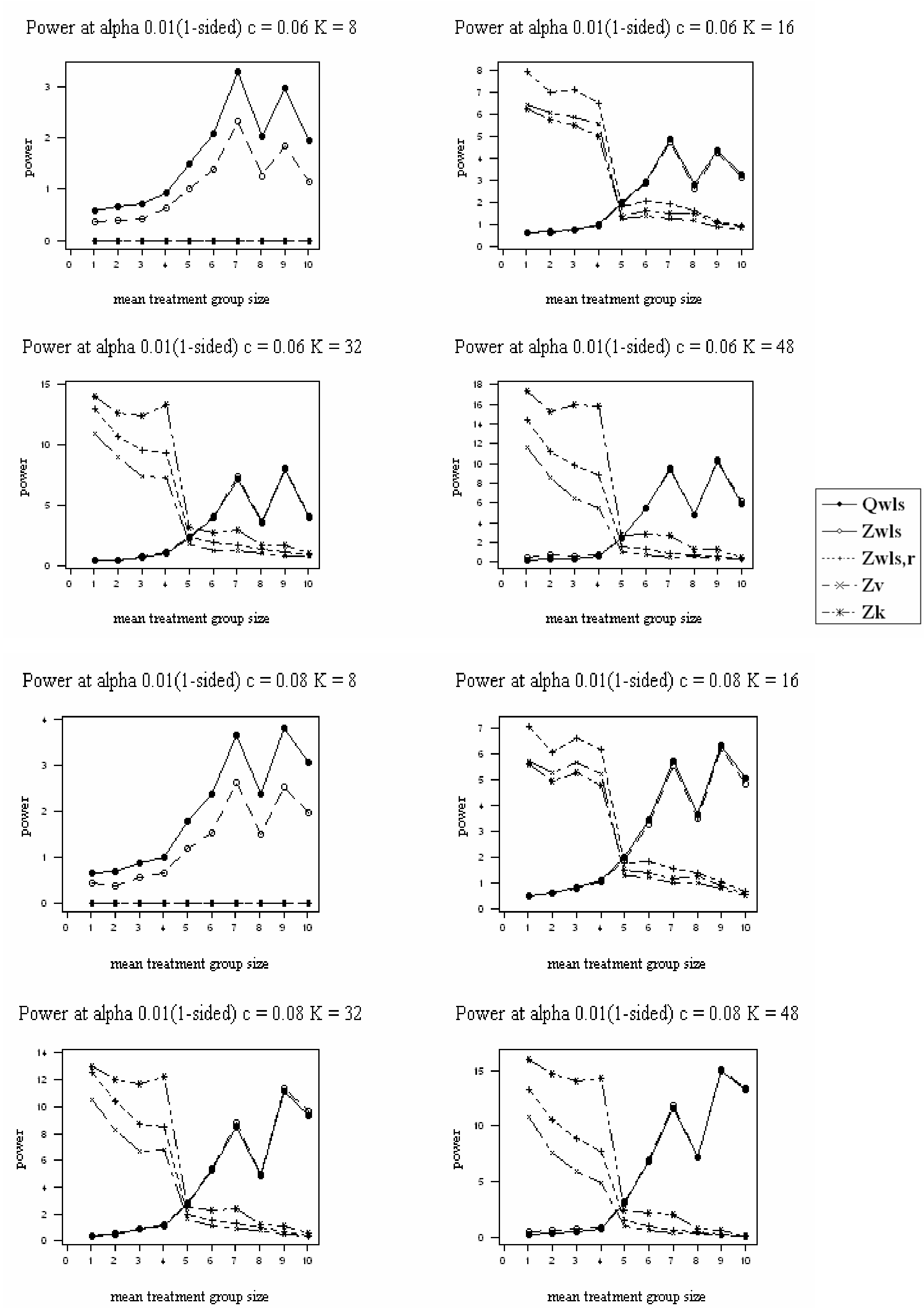


Figure 2G Comparison of the power of tests by *c*, 0.01 level of significance for one - sided and number of center (*K*) where mean treatment group are unequal. (Continued)

Table 2.8 Mean ranking of power of tests in each number of centers, at the 0.10 significance level (one-sided) and $n_{i1} \neq n_{i2}$.

c	***K***	Q_{WLS}	Z^2_{WLS}	$Z^2_{WLS,R}$	Z^2_V	Z^2_K
		Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)	Mean(S.E.)
0.02	8	3.8 (.20)	5.0 (.00)	1.0 (.00)	2.5 (.17)	2.7 (.21)
	16	3.5 (.34)	5.0 (.00)	2.2 (.13)	3.2 (.13)	1.1 (.10)
	32	2.8 (.63)	3.6 (.34)	2.9 (.32)	4.0 (.30)	1.7 (.26)
	48	2.7 (.63)	3.1 (.32)	3.1 (.32)	4.2 (.33)	1.9 (.32)
0.04	8	3.6 (.27)	5.0 (.00)	1.0 (.00)	2.6 (.22)	2.8 (.20)
	16	3.1 (.41)	4.7 (.24)	2.5 (.24)	3.6 (.27)	1.2 (.13)
	32	2.8 (.63)	3.2 (.36)	3.0 (.33)	4.1 (.32)	1.9 (.32)
	48	2.7 (.63)	2.9 (.31)	3.3 (.37)	4.1 (.32)	2.0 (.33)
0.06	8	3.0 (.42)	5.0 (.00)	1.2 (.13)	2.7 (.30)	3.1 (.10)
	16	2.6 (.48)	4.2 (.42)	2.8 (.29)	3.8 (.29)	1.6 (.27)
	32	2.7 (.63)	3.0 (.33)	3.1 (.32)	4.2 (.33)	2.0 (.33)
	48	2.6 (.65)	2.9 (.32)	3.2 (.36)	4.1 (.32)	2.2 (.36)
0.08	8	2.6 (.48)	4.3 (.40)	1.6 (.27)	3.0 (.39)	3.5 (.22)
	16	2.5 (.50)	3.6 (.48)	3.0 (.33)	4.0 (.33)	1.9 (.32)
	32	2.6 (.65)	3.2 (.36)	3.0 (.30)	4.0 (.30)	2.2 (.42)
	48	2.6 (.65)	2.8 (.33)	3.3 (.37)	3.5 (.27)	2.8 (.55)

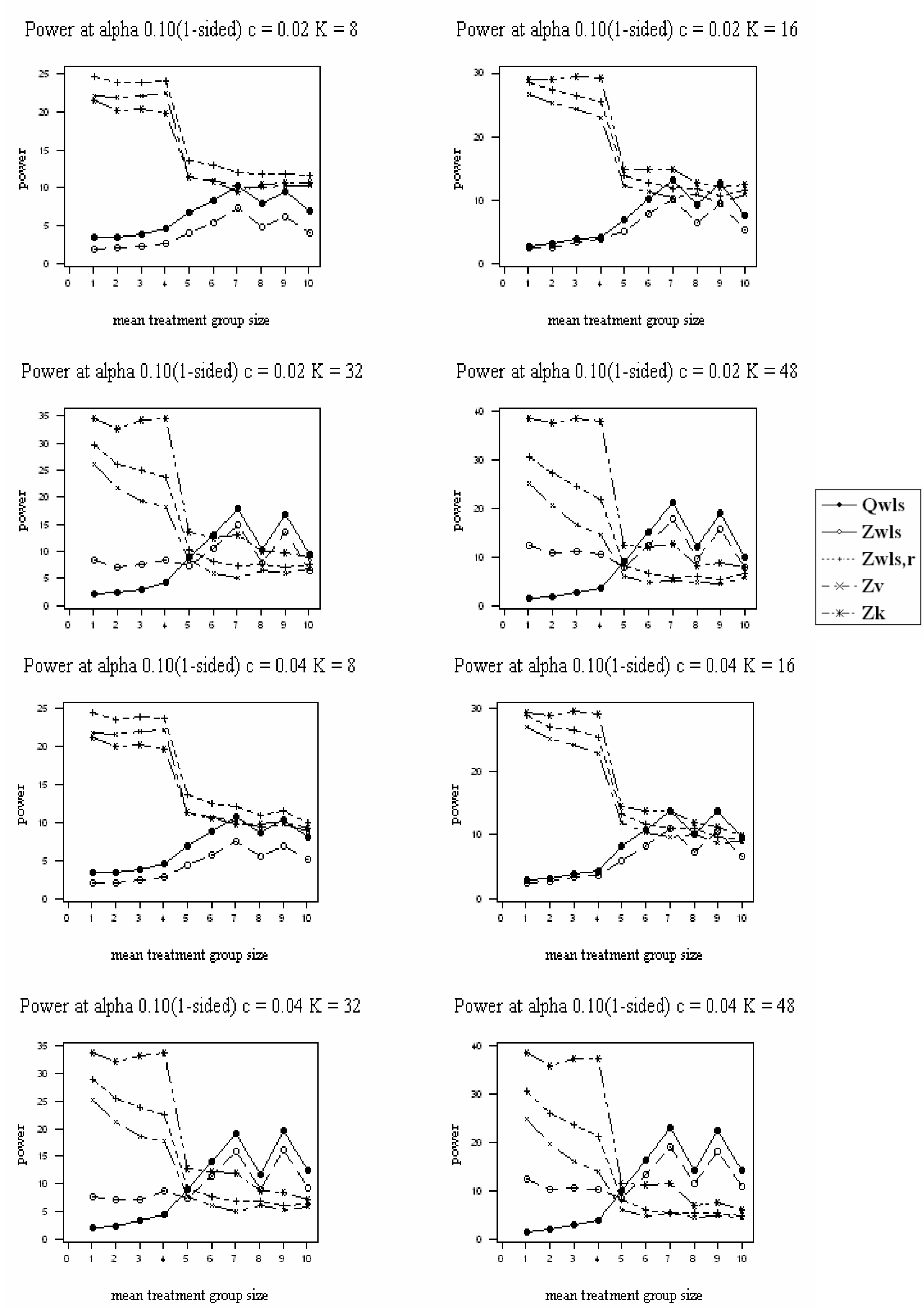


Figure 2H. Comparison of the power of tests by *c*, 0.10 level of significance for one-sided and number of center (*K*) where mean treatment group are unequal.

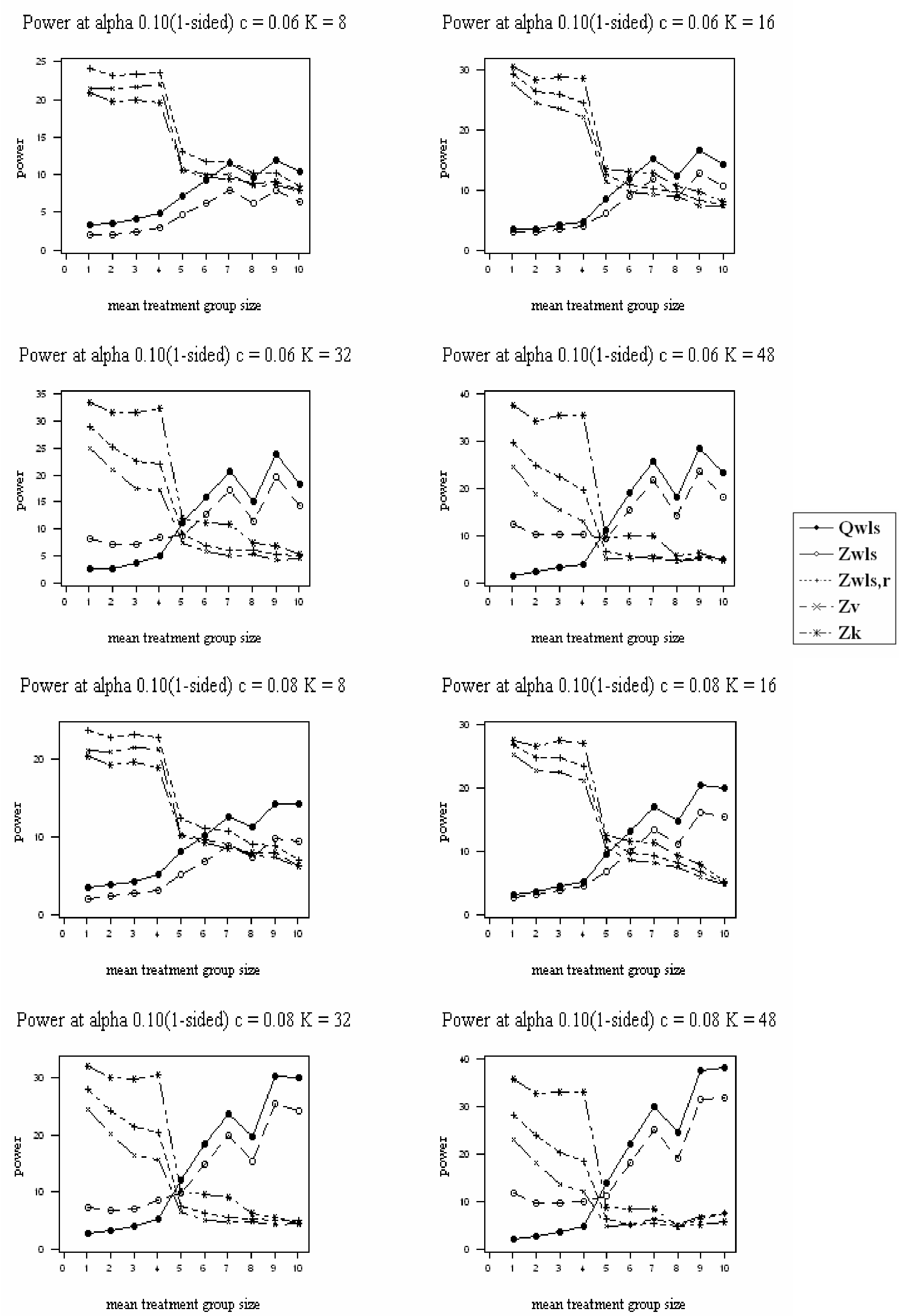


Figure 2H. Comparison of the power of tests by *c*, 0.10 level of significance for one-sided and number of center (*K*) where mean treatment group are unequal. (Continued)

BIOGRAPHY

NAME	Miss Patranuch Sapchookul
DATE OF BIRTH	June 29, 2005
PLACE OF BIRTH	Nakhonratchasima, Thailand
INSTITUTIONS ATTENDED	Srinakharinwiroj University, 1994-1997: Bachelor of Science (Mathematics) Mahidol University, 2001-2005: Master of Science (Biostatistics)
HOME ADDRESS	14 / 98 Soi. Charansanitwong 95/1 Charansamitwong Rd. Bang-Or, Bangplat, Bangkok, Thailand.