

# **THE EFFECTS OF OVERSTRATIFICATION ON THE STRATIFIED LOG RANK TEST FOR SURVIVAL ANALYSIS**

by

**Shuting Yang**

B.S., Henan Normal University, P.R. China, 1987

M. S., Yunnan University, P. R. China, 1990

Submitted to the Graduate Faculty of  
Graduate School of Public Health in partial fulfillment  
of the requirements for the degree of  
Master of Science

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH  
GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Shuting Yang

It was defended on

December 02, 2011

and approved by

**Thesis Advisor:** Stewart Anderson, PhD, Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Hanna Bandos, PhD, Assistant Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Felicia Wu, PhD, Associate Professor,  
Department of Environmental and Occupational Health, Graduate School of Public Health,  
University of Pittsburgh

Copyright © by Shuting Yang

2012

# **THE EFFECTS OF OVERSTRATION ON THE STRATIFIED LOG RANK TEST FOR SURVIVAL ANALYSIS**

Shuting Yang M. S.

University of Pittsburgh, 2012

## Abstract

Survival analysis concerns the characterization or comparison of one or more distributions of the time to a well defined event. The log-rank test is the most common method used to compare the survival distributions of two samples. When data within the two groups are stratified according to some risk factors, then a *stratified* log-rank test is employed.

Stratified analysis is a procedure used to compare outcomes in different groups while at the same time correcting for the effects of confounders. It is one way to ensure that important prognostic factors are equally distributed among different treatments.

The ordinary log-rank test is known to be conservative when treatments have been assigned by a stratified design. The stratified log-rank test is valid even when the sizes of strata differ. Schoenfeld and Tsiatis modified the log-rank test with a variance adjustment reflecting the dependence of survival on strata size. Their method is shown to be more efficient than the ordinary stratified log rank test when the number of strata is large, and it remains valid when the censoring distributions differ across treatment groups.

In this thesis, we investigate these three log-rank tests for survival analysis. The effect of the stratum sizes on each type of analysis is evaluated using simulated data.

Our results show that the modified log rank test is beneficial for stratified survival analysis in most cases especially when there are large numbers of strata and the strata sizes get small. The statistical power of the modified log-rank test is relatively stable even with very small strata sizes and high strata effects.

The public health relevance of this thesis is that the modified log-rank test we investigated and implemented using the R programming language provides an alternative and more efficient way to accommodate higher amounts of stratification in analyzing survival data. More efficient statistical methods indirectly have public health impact as such methods lead to analyses which better identify treatments, interventions or factors that influence health outcomes. Such analyses are commonly used in clinical trials and other studies which influence public health.

## TABLE OF CONTENTS

PREFACE .....	XII
1.0 INTRODUCTION .....	1
1.1 LOG-RANK TESTS FOR NON-STRATIFIED AND STATIFIED DATA .....	3
1.2 SIMULATION IN SURVIVAL ANALYSIS .....	4
2.0 DATA GENERATION AND ANALYSIS .....	5
2.1 DATA GENERATION ... ..	5
2.2 DATA ANALYSIS .....	7
3.0 RESULTS.....	8
3.1 BASELINE OF STATISTICAL POWER FOR THE THREEMETHODS ACCORDING TO STRATA EFFECTS.....	8
3.2 STATISTICAL POWER OF THE THREE METHODS FOR VARIOUS STRATA EFFECTS.....	10
3.3 STATISTICAL POWER COMPARISONS FOR DIFFERENT STRATUM SIZE.....	14
3.3.1 STATISTICAL POWER COMPARISONS WHEN THE METHODS WERE APPLIED TO THE LARGE SAMPLE (N=100) .....	14
3.3.2 STATISTICAL POWER COMPARISONS WHEN THE METHODS WERE APPLIED TO THE SMALL SAMPLE (N=60) .....	20

4.0 DISCUSSION.....	24
5.0 CONCLUSION.....	26
APPENDIX A MATH FORMULAS.....	27
APPENDIX B R-CODE.....	28
BIBLIOGRAPHY .....	41

## LIST OF TABLES

Table 1. Statistical power at baseline for different stratum size (n=100) .....	8
Table 2. Statistical power at baseline for different stratum size (n=60) .....	10
Table 3. Statistical power various according to strata effects (n=100).....	13
Table 4. Statistical power various according to strata effects (n=60, HR=1.857).....	14
Table 5. Statistical power various according to stratum size ( n=100, D=1).....	16
Table 6. Statistical power various according to stratum size ( n=100, D=2).....	16
Table 7. Statistical power various according to stratum size ( n=100, D=5).....	17
Table 8. Statistical power various according to stratum size ( n=100, D=10).....	17
Table 9. Statistical power various according to stratum size ( n=60, D=1).....	21
Table 10. Statistical power various according to stratum size ( n=60, D=2).....	21
Table 11. Statistical power various according to stratum size ( n=60, D=5).....	21
Table 12. Statistical power various according to stratum size ( n=60, D=10).....	22



## LIST OF FIGURES

Figure 1. The strategy of data generation .....	5
Figure 2. Baseline of statistical power (D=1,HR=1,n=100).....	9
Figure 3. Baseline of statistical power (D=1,HR=1,n=60).....	11
Figure 4. Statistical power varies according to strata effects.....	15
Figure 5. Statistical power varies according to number of strata (n=100): noncensored .....	18
Figure 6. Statistical power varies according to number of strata (n=100): censored .....	19
Figure 7. Statistical power varies according to number of strata (n=60): balanced design .....	22
Figure 8. Statistical power varies according to number of strata (n=60): unbalanced design .....	23

## **PREFACE**

I would like to acknowledge Dr. Stewart Anderson, who has been my academic mentor for guiding me throughout the study period. I would also like to acknowledge Dr. Hana Bandos, and Dr. Felicia Wu, who have provided me with technical guidance and support that enabled me to conduct this study. I also would like to thank all my teachers and classmates who had been shared moments of encouragement and inspiration.

And most of all, I would like to thank my husband, Guoying Yu, for his love, encouragement and support at all times during the study program.

## **Nomenclature**

ULR: unstratified log-rank test

SLR: stratified log-rank test

MLR: modified log-rank test

HR: Hazard Ratio

## 1.0 INTRODUCTION

Survival analysis is used to measure follow-up time to the occurrence of a well defined event from a defined starting point. It involves estimation of the survival function and sometimes involves comparison of survival curves in different groups (treatments). Since survival data may involve incomplete observation of the failure times, and the underlying distribution is rarely normal, special techniques are required for comparing the risk for death (or event) associated with different treatments. The most commonly used statistical test for comparing survival distributions of two treatments is called the log-rank test. It is used to test the null hypothesis that the probability an event occurring at any time point is the same for each treatment. Log-rank tests are appropriate to use when data are right skewed and censored (Mantel, 1966).

In prospective, randomized studies where treatments are being compared, stratification is often employed to ensure that treatments groups are reasonably balanced with respect to variables that confound results. The basic idea of stratification is to divide the sample space into subgroups called strata, and attempt to force treatment balanced within each subgroup among variables other than treatment which could affect outcome. Fewer observations are required for a stratified analysis to achieve a particular precision of errors compared to an un-stratified analysis (Podgor & Gastwirth, 1994). Stratified analyses are used to compare outcome in different groups while at the same time correcting for the effects of the “confounders” (or variables other than treatment that are related to outcome). Many clinical trials are often designed to compare the survival distributions of different treatment groups, and the treatment effects are assessed through

significance tests (Peduzzi, Henderson, Hartigan, & Lavori, 2008). The basic comparison is made within such groups, and the discrepancy found is accumulated across all strata, the test statistic is derived from these accumulated statistics. Stratification in analysis allows a factor to be adjusted for without estimating its effect.

Stratification is very important and necessary in many randomization clinical trial designs, but the amount of stratification is crucial. The number of strata and the size of each stratum are two essential issues to concern for stratification analysis (Akazawa, et al., 1997). A stratified analysis will be very efficient as long as the number of strata is relatively small; however, as the number of strata gets large, this method becomes more inefficient partially as the size of each stratum gets smaller. Under-stratification risks bias due to imbalance of confounder across treatments, while over-stratification may reduce the precision of estimated effects and lowers the power of the analyses (De Stavola & Cox, 2008).

Previous work was developed to study the power loss of the log-rank test under certain conditions (Lagakos, 1988; Akazawa, Nakamura, & Palesch, 1997; De Stavola & Cox, 2008). Their results and others (Feng, Wang, & Tu, 2010 ) indicate that the power loss of the stratified log-rank test due to the fact that the stratum sizes may be too small or alternatively due to the existence of heterogeneity within strata.

## 1.1 LOG-RANK TESTS FOR NON-STRATIFIED AND STATIFIED DATA

The ordinary or unstratified log-rank test (Mantel, 1966) is used to compare the survival curves for the two treatment groups. Typically, the null hypothesis for log-rank test is that the probability of an event occurring at any time point is the same for each treatment group. Under the null hypothesis, the risk of death can be calculated from the combined data for both groups. Statistics are calculated based on the formula in Appendix A ( A(1)). This test is known to be conservative when treatments have been assigned by a stratified design. Additionally, when the number of patients in each subgroup is uneven, the test will be biased.

The stratified log-rank test (Green & Byar, 1978) is applied when we wish to compare two groups but taking into account the other variables such as age, sex or center. The null hypothesis is that there is no difference between the groups in the survival distribution for each stratum. This test is valid even when the patients in subgroups are different in number (but balanced by treatments), and can be very efficient as long as the size of each stratum is relatively large. However, as the number of strata gets large and/or the size of each stratum gets relatively small, this method of analysis can become substantially less efficient (Day, Byar, & Green, 1980; Neuhauser & Becher, 1997).

The modified log-rank test (Schoenfeld & Tsiatis, 1987) studied in this thesis is similar to the ordinary log-rank test, but employs a variance adjustment to reflect the dependence of survival on strata. This method is shown to be more efficient than the stratified log rank test when the number of strata is large, and it remains valid when the censored distributions depend on the treatment group.

The unstratified log rank test is always asymptotically more powerful than the stratified log rank test in a homogeneous sample (within stratum). The loss of power of a stratified log rank test is due to the loss of information from unnecessary stratification. (Akazawa, et al., 1997; Feng & Tu, 2010; Liao, 1998)

## **1.2 SIMULATION IN SURVIVAL ANALYSIS**

Simulation is increasingly being applied to solve a variety of scientific problems. Simulation studies are increasingly employed in evaluating the validity of statistical models. Statistical simulation is a numerical method of solving mathematical problems. It is a way to model random events. To test the desired phenomenon of a model, a mathematical model of the phenomenon is constructed, stochastic errors are added and the needed quantities are determined by a statistical analysis of the model. The process is performed repeatedly, so that estimates of various parameters or other features can be accurately determined. By observing simulated outcomes, one can gain insight on real world problems.

In this thesis, the effect of over stratification on the survival analysis was evaluated using three types of log rank tests with simulated survival data. For the three log-rank tests of interest, we evaluated how different strata sizes affected the efficiency for the analyses. To do this, we first generated both censored and uncensored data with hazard ratio of 1 and 1.86. We also varied sample and stratum sizes. A detailed description of our simulation strategy is given in section of 2.1 of this thesis. Our strategy followed that of Schoenfeld and Tsiatis (1987).

## 2.0 DATA SET GENERATION AND ANALYSIS

### 2.1 DATA GENERATION

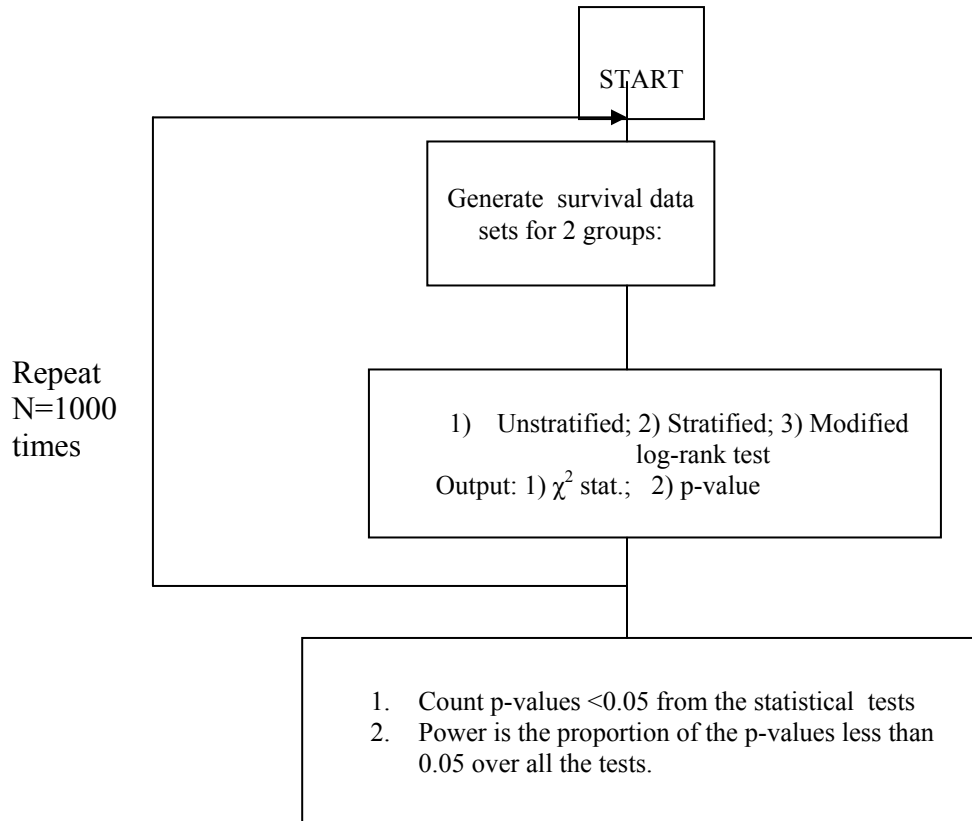


Figure 1. The strategy of data generation

The strategy of the data generation had been shown in Figure1. In the simulations, we first used a total sample size of 100, 50 in each group (treatment A and treatment B). In these data sets, we developed the uncensored and censored data. The censored data had 15% censored information. We also generated data with a smaller sample size of 60 observations, one with 30 observations in each group, and one with 40 observations in one group and 20 observations in the other group. Following the procedures in 1987, the failure times were generated according to a hazard rate model with the relationship:  $\lambda_{A_j} = \lambda_j(1 + \frac{1}{2}\theta)$  for treatment A, and  $\lambda_{B_j} = \lambda_j(1 - \frac{1}{2}\theta)$  for treatment B. Here  $j$  is the stratum indicator, and  $\lambda_j$  is the baseline hazard rate in stratum  $j$  within the treatments. The hazard rate is sometimes called as mortality rate, failure rate or death intensity, it is denoted as the instantaneous rate of failure for the survivors to time  $t$  during the next instant of time. The hazard rate is calculated from the formula:

$$\lambda = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X + \Delta x | X \geq x)}{\Delta x}.$$

The function of hazard rate can have different shapes; the most common type is the exponential distribution. In our distributions, the hazard ratio for treatment A over B in each stratum is given by

$$\lambda = \frac{\lambda_{A_j}}{\lambda_{B_j}} = (1 + \frac{1}{2}\theta) / (1 - \frac{1}{2}\theta)$$

with the null hypothesis being represented by  $\theta = 0$ , which corresponding to  $\lambda = 1$ . The simulations were also computed with  $\theta = 0.6$ , corresponding to a hazard ratio of

$$\lambda = (1 + \frac{0.6}{2}) / (1 - \frac{0.6}{2}) = 1.857.$$



For any given simulation, a D-fold strata effects means that the ratio of the larger strata hazard rate to the smaller strata hazard rate is D. Hazard rates were distributed uniformly among the strata. The formula  $\lambda_j = \lambda \{1 + (D-1)(j-1)/(K-1)\}$  was used for all j stratum.

## 2.2 DATA ANALYSIS

For each set of data, three tests, ordinary (unstratified) log-rank test (ULR), stratified log-rank test (SLR) and modified log-rank test (MLR) were performed and the null hypotheses were rejected if these test statistics exceeded the critical value for  $p=0.05$  (1.645). The statistical power is the percentage of the p-values less than 0.05 over all the tests (number of simulations). For all the tables, “stratum size” means the number of patients in each stratum, “no censored” means there was no censored in the data set, “censored” means there were 15% censored in the data set. Finally, each experiment was simulated 1000 times. All the data generation, simulation and statistical analysis were done by R. We used the built in R-programs to perform the unstratified and stratified log-rank test and used our own R programs to perform the modified log-rank test. A sample simulation program is displayed in Appendix B.

### 3.0 RESULTS

#### 3.1 BASELINE OF STATISTICAL POWER OF THE THREE METHODS ACCORDING TO STRATA SIZE

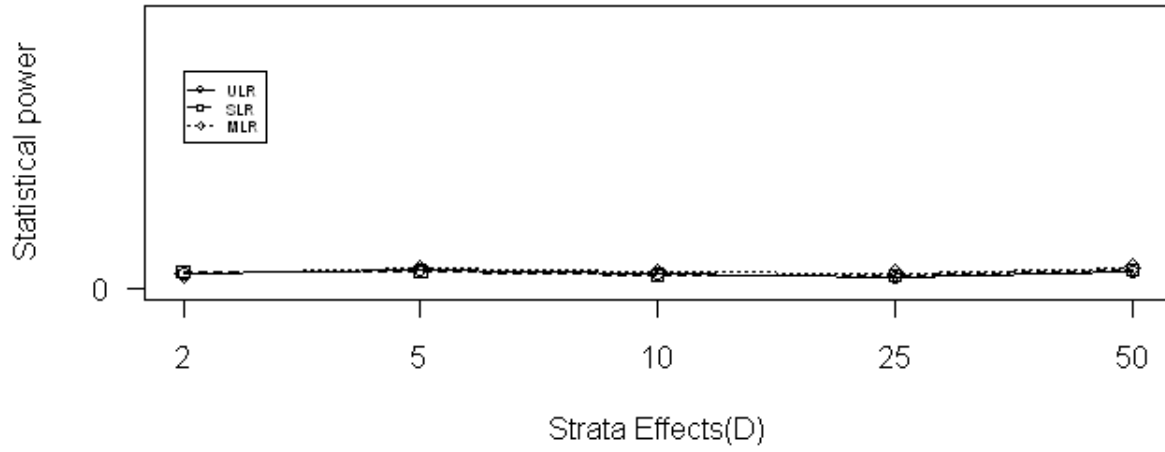
We initially checked the baseline statistical power which with no stratification effects or treatment effects. Table 1 and Figure 2 summarize the statistical power for the larger sample size (n=100). All of the three log rank tests including un- stratified log-rank test (ULR), stratified log-rank (SLR) test and modified log rank test have statistical power which is near to the correct value of 0.05 across all the different stratum sizes for uncensored data and the censored data. Similar results were generalized with the relatively smaller sample size (n=60), and again, the baseline statistical power is near to 0.05 across all of different stratum size with balanced and unbalanced designs (Tables 2 and Figure 3). For the unbalanced data with small strata, the average power was a bit higher than the normal level of 0.05 (Table 2).

Table 1. Statistical power at baseline for different stratum size (n=100)

Number of stratum	Stratum size	Uncensored			Censored*		
		ULR	SLR	MLR	ULR	SLR	MLR
K=2	50	0.053	0.061	0.052	0.052	0.057	0.062
K=5	20	0.067	0.062	0.074	0.048	0.052	0.052
K=10	10	0.054	0.049	0.059	0.044	0.04	0.052
K=25	4	0.042	0.046	0.054	0.062	0.062	0.066
K=50	2	0.063	0.069	0.075	0.062	0.05	0.068

\*There are 15% censored observations

Baseline:  $D=1, HR=1$  (uncensored,  $n=100$ )



Baseline:  $D=1, HR=1$  (censored,  $n=100$ )

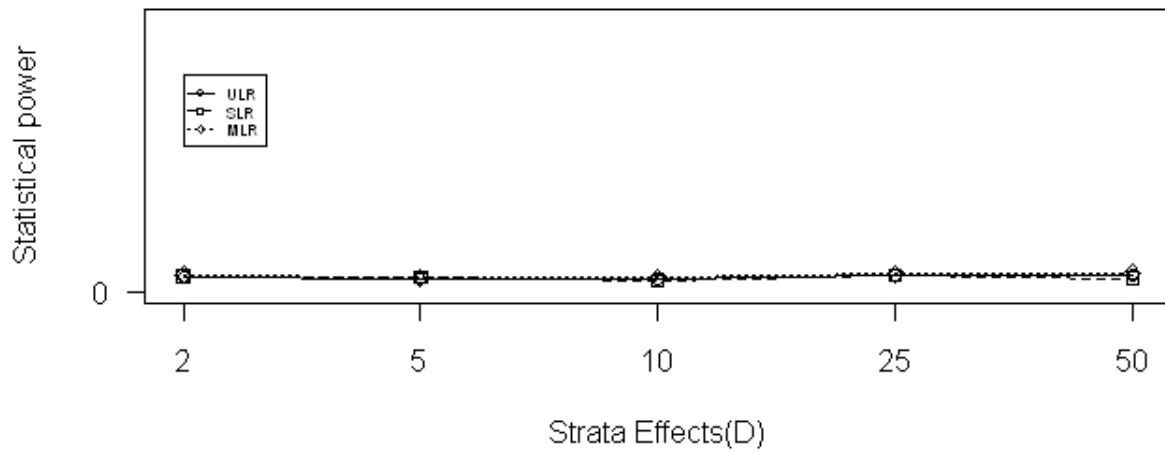


Figure 2. Baseline of Statistical power ( $D=1, HR=1, n=100$ )

Table 2. Statistical power at baseline for different stratum size (n=60)

with balanced and unbalanced design

Number of stratum	Stratum size	Balanced Design*			Unbalanced Design**		
		MLR	ULR	SLR	MLR	ULR	SLR
K=2	30	0.08	0.064	0.06	0.082	0.062	0.06
K=4	15	0.08	0.068	0.072	0.054	0.054	0.06
K=5	12	0.086	0.066	0.066	0.078	0.07	0.066
K=10	6	0.08	0.058	0.048	0.062	0.06	0.042
K=15	4	0.07	0.076	0.06	0.058	0.048	0.052
K=30	2	0.054	0.066	0.044	0.082	0.066	0.052

\*Balanced design: 30 observations in each group;\*\*Unbalanced design: 40 observations in one group and 20 observations in the other group.

### 3.2 STATISTICAL POWER OF THE THREE METHODS FOR VARIOUS STRATA EFFECTS

Table 3 summarizes the stratification effects on the statistical power for the three methods with larger sample (n=100). Generally, for each method, the statistical power tends to decrease with the increasing of the strata effects. With the lower strata effects (D =1 or 2), the power of SLR is slightly lower or close to ULR, the power of MLR is also close to ULR when stratum size was the same. While with the higher strata effects (D=5 or10), both the statistical power of both the MLR and SLR are higher than that of ULR when the stratum sizes are the same. The ratio of SLR over ULR tends to increase with the increase of strata effect when the methods are applied

to the relatively larger size of stratum. When the stratum size getting bigger, for the stratum size ( $n_j$ ) equal to  $n_j = 10(k=10)$ ,  $n_j=20$  ( $k=5$ ) or  $n_j=50$  ( $k=2$ ), the ratio of the statistical power of SLR over ULR increased from 0.99 to 1.63 ( $k=2$ ), 1.03 to 1.53 ( $k=5$ ) and 0.9 to 1.22 ( $k=10$ ) for  $D=1$  to  $D=10$ . The ratio of SLR over ULR tends to decrease with the increase of strata effect when the methods are applied to the small size of stratum. In the case of uncensored data, for the sample size equal to 4 ( $k=25$ ), the ratio of statistical power of SLR over ULR is decreased from

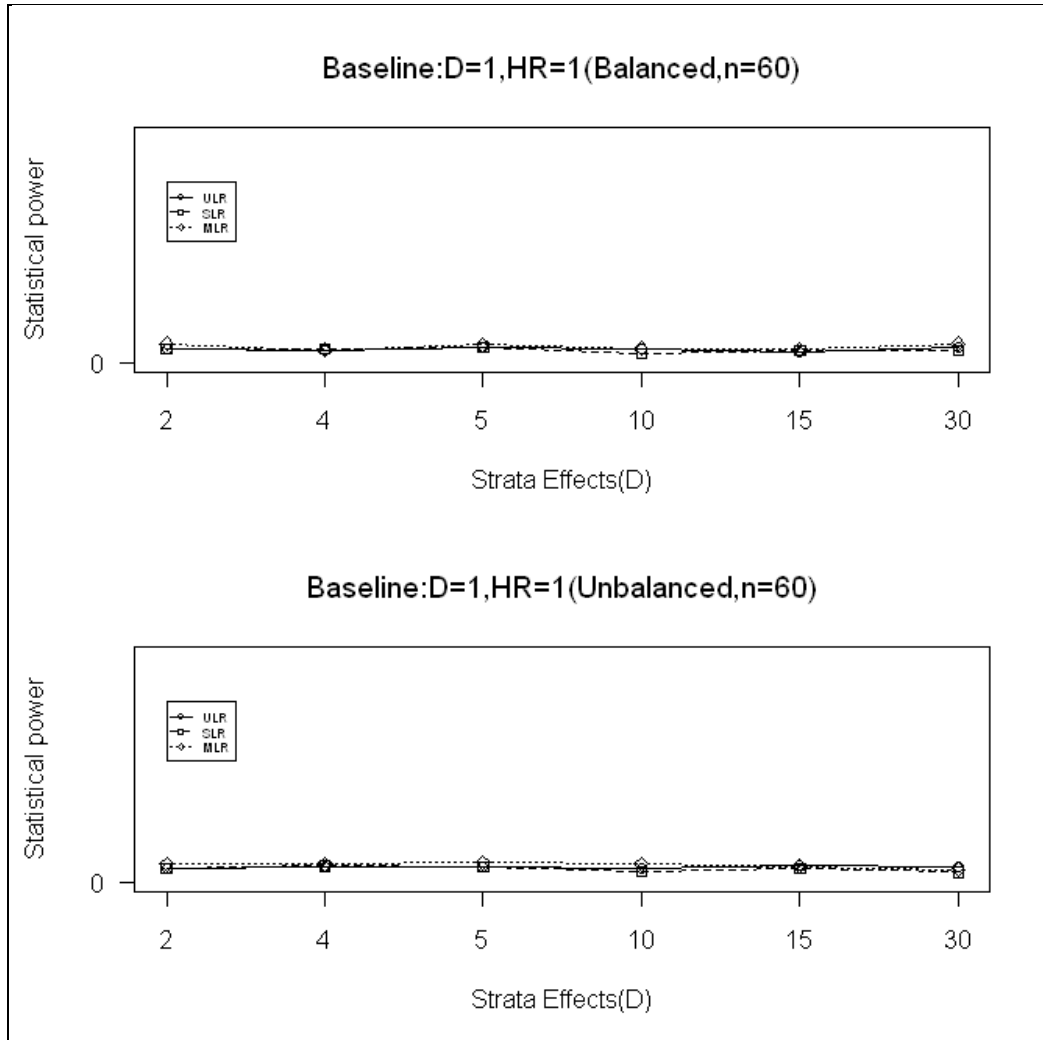


Figure 3. Baseline of Statistical power ( $D=1$ ,  $HR=1$ ,  $n=60$ )

1.23 for  $D=1$  and to 0.99 for  $D=10$ , and for the sample size equal to 2 ( $k=50$ ), the ratio of the statistical power of SLR over ULR is decreased from 1.32 for  $D=1$  to 1.18 for  $D=10$ . But the ratio of MLR over ULR kept increasing with the increase of strata effects for relatively larger stratum sizes while it kept relatively stable with smaller stratum size  $n_j=4$  ( $k=25$ ) or  $n_j=2$  ( $k=50$ ). The censored and uncensored data showed nearly the same trend for the change, while the statistical power for censored data is generally lower than that of uncensored data set. And the ratio of MLR to ULR for censored data is even higher compared to that of uncensored data at the same condition.

Table 4 shows the results when the statistical methods were applied to the relatively smaller sample size ( $n=60$ ). With the balanced design, even though the statistical power was dramatically lower compared to the larger size sample ( $n=100$ ), they both have the same trends.

The statistical power decreases with the increasing of strata effects, and both the ratio of MLR over ULR and the ratio of SLR over ULR were increase with the increasing of strata effects, while the ratio of MLR over ULR was always greater than “1” across all the different stratum number with different strata effects, while the ratio of SLR over ULR was less than “1” when the strata number getting larger.

With the unbalanced design, the statistical power was even lower than that of balanced design. At the lower  $k$  ( $k \leq 10$ ), the ratio of MLR to ULR and the ratio of SLR to ULR closed or greater than “1”, and both of them were increased with the increasing of strata effects. At a higher number of strata ( $k > 10$ ), the results appeared no stable trends and in most cases, the ratios of MLR to ULR and the ratio of SLR to ULR were less than “1”, indicating there was no significant

advantage for either MLR or SLR over ULR. Fig.4 showed the general trends for statistical power according to the strata effects. Even with some exceptions, we still can state that in generally, the statistical power dropped with the increasing of strata effects.

Table 3. Statistical power with various strata effects ( n=100, HR=1.857)

N=100;			Uncensored					Censored				
Stratum Number	Stratum Size	Strata Effects	MLR	ULR	SLR	M/U *	S/U **	MLR	ULR	SLR	M/U *	S/U **
K=2	50	D=1	0.89	0.865	0.86	1.03	0.99	0.815	0.78	0.775	1.04	0.99
		D=2	0.845	0.83	0.86	1.02	1.04	0.69	0.66	0.71	1.05	1.08
		D=5	0.79	0.675	0.85	1.17	1.26	0.675	0.535	0.795	1.26	1.49
		D=10	0.79	0.54	0.88	1.46	1.63	0.61	0.4	0.775	1.52	1.94
K=5	20	D=1	0.86	0.835	0.775	1.03	0.93	0.805	0.8	0.76	1.01	0.95
		D=2	0.87	0.845	0.79	1.03	0.93	0.835	0.795	0.8	1.05	1.01
		D=5	0.795	0.68	0.785	1.17	1.15	0.675	0.565	0.73	1.19	1.29
		D=10	0.76	0.535	0.82	1.42	1.53	0.57	0.445	0.75	1.28	1.69
K=10	10	D=1	0.87	0.875	0.79	0.99	0.90	0.755	0.76	0.735	0.99	0.97
		D=2	0.85	0.815	0.81	1.04	0.99	0.815	0.775	0.725	1.05	0.94
		D=5	0.82	0.765	0.77	1.07	1.01	0.735	0.65	0.73	1.13	1.12
		D=10	0.81	0.66	0.805	1.23	1.22	0.69	0.53	0.78	1.3	1.47
K=25	4	D=1	0.865	0.705	0.865	1.23	1.23	0.785	0.64	0.76	1.23	1.19
		D=2	0.815	0.67	0.8	1.22	1.19	0.81	0.585	0.75	1.38	1.28
		D=5	0.83	0.655	0.755	1.27	1.15	0.69	0.565	0.655	1.22	1.16
		D=10	0.78	0.67	0.66	1.16	0.99	0.665	0.59	0.57	1.13	0.97
K=50	2	D=1	0.875	0.655	0.865	1.34	1.32	0.82	0.52	0.81	1.58	1.56
		D=2	0.875	0.645	0.815	1.36	1.26	0.81	0.495	0.775	1.64	1.57
		D=5	0.82	0.64	0.73	1.28	1.14	0.74	0.505	0.645	1.47	1.28
		D=10	0.815	0.595	0.705	1.37	1.18	0.715	0.49	0.635	1.46	1.30

\*M/U: the ratio of MLR over ULR; \*\*S/U: the ratio of SLR over ULR

### 3.3 STATISTICAL POWER COMPARISON FOR DIFFERENT STRATA SIZE

#### 3.3.1 Statistical power comparisons when the methods applied to the larger sample (n=100)

The effects of the stratum size on the statistical power are shown in Tables 5 to Table 8 and Figure 5. We generated the simulation data with both uncensored and censored data with 15% of censoring. Given stratification effect level, the statistical power always increased with the increasing of the stratum size, in the other way to say, the statistical power dropped with the increasing of number of strata. With no strata effects (D=1), the power of MLR and SLR were relatively stable with the increase of number of strata. For censored data, the powers of MLR for

Table 4. Statistical power with various strata effects ( n=60, HR=1.857)

N=60; N.sim=1000			Balance Design					Unbalanced Design				
Stratum Number	Stratum size	Strata Effect	MLR	ULR	SLR	M/U*	S/U*	MLR	ULR	SLR	M/U*	S/U*
K=2	30	D=1	0.674	0.64	0.62	1.05	0.97	0.576	0.588	0.536	0.98	0.91
		D=2	0.646	0.568	0.622	1.14	1.1	0.522	0.508	0.552	1.03	1.09
		D=5	0.56	0.42	0.642	1.33	1.53	0.464	0.334	0.562	1.39	1.68
		D=10	0.566	0.3	0.662	1.89	2.21	0.414	0.23	0.58	1.8	2.52
K=5	12	D=1	0.66	0.642	0.548	1.03	0.85	0.562	0.594	0.514	0.95	0.87
		D=2	0.68	0.612	0.572	1.11	0.93	0.57	0.566	0.544	1.01	0.96
		D=5	0.554	0.452	0.542	1.23	1.2	0.466	0.392	0.498	1.19	1.27
		D=10	0.554	0.36	0.594	1.54	1.65	0.404	0.242	0.516	1.67	2.13
K=10	6	D=1	0.678	0.638	0.524	1.06	0.82	0.578	0.572	0.432	1.01	0.76
		D=2	0.638	0.616	0.5	1.04	0.81	0.56	0.568	0.448	0.99	0.79
		D=5	0.586	0.488	0.502	1.2	1.03	0.482	0.41	0.452	1.18	1.1
		D=10	0.544	0.368	0.504	1.48	1.37	0.444	0.286	0.446	1.55	1.56
K=15	4	D=1	0.652	0.62	0.438	1.05	0.71	0.27	0.576	0.408	0.47	0.71
		D=2	0.622	0.576	0.474	1.08	0.82	0.226	0.564	0.352	0.4	0.62
		D=5	0.606	0.516	0.478	1.17	0.93	0.182	0.534	0.352	0.34	0.66
		D=10	0.534	0.378	0.446	1.41	1.18	0.152	0.474	0.32	0.32	0.68
K=30	2	D=1	0.698	0.656	0.368	1.06	0.56	0.078	0.556	0.228	0.14	0.41
		D=2	0.666	0.616	0.332	1.08	0.54	0.074	0.622	0.42	0.12	0.68
		D=5	0.586	0.506	0.344	1.16	0.68	0.048	0.512	0.612	0.09	1.2
		D=10	0.542	0.39	0.332	1.39	0.85	0.028	0.458	0.738	0.06	1.61

\*M/U: the ratio of MLR over ULR; \*\*S/U: the ratio of SLR over ULR



all set stratum were closing to 0.88, and the power of SLR is closing to 0.86, while the power of ULR decreased from 0.86 at  $k=2$  to 0.655 at  $k=50$ . The power for the censored data was slightly lower than that for uncensored data. For a given stratum size, SLR was generally higher than

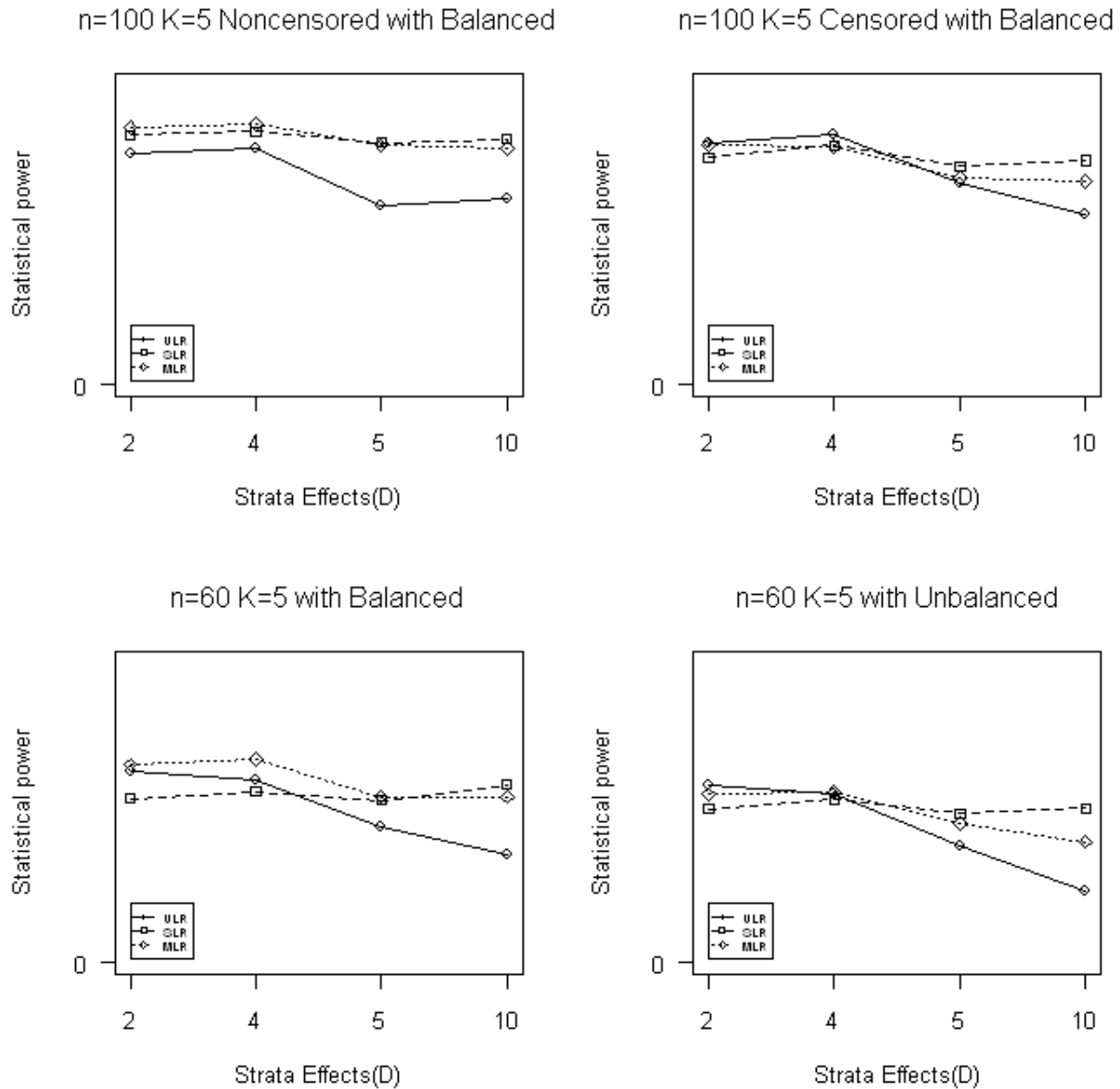


Figure 4. Statistical power varies according to strata effects

ULR, and MLR was generally higher than SLR. The ratio of SLR/ULR for uncensored data was increased from 1.01 (K=2) to 1.32 (k=50), and the ratio of MLR/ULR for uncensored data was increased from 1.03 (K=2) to 1.34 (k=50). The ratio of MLR/ULR for censored data increased from 1.05 (k=2) to 1.58(k=50), SLR/ULR for censored data increased from 1.06 (k=2) to 1.56 (k=50). These results indicate that both the MLR and SLR might be more advantages with censored data as compared to the uncensored data fort this condition (Table 5). With the strata effects for D=2, the statistical power of the three test methods have the nearly same pattern as that with the strata effects for D=1, even when the ratios of MLR over ULR and SLR over ULR are very close at each stratum (Table 6).

Table 5. Statistical power various according to stratum size ( n=100, D=1)

D=1		Uncensored					Censored***				
Number of stratum	Stratum size	MLR	ULR	SLR	M/U *	S/U **	MLR	ULR	SLR	M/U *	S/U **
K=2	50	0.89	0.86	0.865	1.03	1.01	0.815	0.775	0.78	1.05	1.01
K=5	20	0.86	0.775	0.835	1.11	1.08	0.805	0.76	0.8	1.06	1.05
K=10	10	0.87	0.79	0.875	1.1	1.11	0.755	0.735	0.76	1.03	1.03
K=25	4	0.865	0.705	0.865	1.23	1.23	0.785	0.64	0.76	1.23	1.19
K=50	2	0.875	0.655	0.865	1.34	1.32	0.82	0.52	0.81	1.58	1.56

\*M/U: the ratio of MLR over ULR; \*\*S/U: the ratio of SLR over ULR; \*\*\*15% censored observations

Table 6. Statistical power various according to stratum size ( n=100, D=2)

D=2		Uncensored					Censored				
Number of stratum	Stratum size	MLR	ULR	SLR	M/U *	S/U **	MLR	ULR	SLR	M/U *	S/U **
K=2	50	0.845	0.86	0.83	0.98	0.97	0.69	0.66	0.71	1.05	1
K=5	20	0.87	0.79	0.845	1.1	1.07	0.835	0.8	0.795	1.04	0.99
K=10	10	0.85	0.81	0.815	1.05	1.01	0.815	0.725	0.775	1.12	1.07
K=25	4	0.815	0.67	0.8	1.22	1.19	0.81	0.585	0.75	1.38	1.28
K=50	2	0.875	0.645	0.815	1.36	1.26	0.81	0.495	0.775	1.64	1.57

\*M/U: the ratio of MLR over ULR; \*\*S/U: the ratio of SLR over ULR

Table 7. Statistical power various according to stratum size ( n=100, D=5)

D=5		Uncensored					Censored				
Number of stratum	Stratum size	MLR	ULR	SLR	M/U *	S/U **	MLR	ULR	SLR	M/U *	S/U **
K=2	50	0.79	0.675	0.85	1.17	1.26	0.675	0.535	0.795	1.26	1.49
K=5	20	0.795	0.68	0.785	1.17	1.15	0.675	0.565	0.73	1.19	1.29
K=10	10	0.82	0.765	0.77	1.07	1.01	0.735	0.65	0.73	1.13	1.12
K=25	4	0.83	0.655	0.755	1.27	1.15	0.69	0.565	0.655	1.22	1.16
K=50	2	0.82	0.64	0.73	1.28	1.14	0.74	0.505	0.645	1.47	1.28

\*M/U: the ratio of MLR over ULR; \*\*S/U: the ratio of SLR over ULR

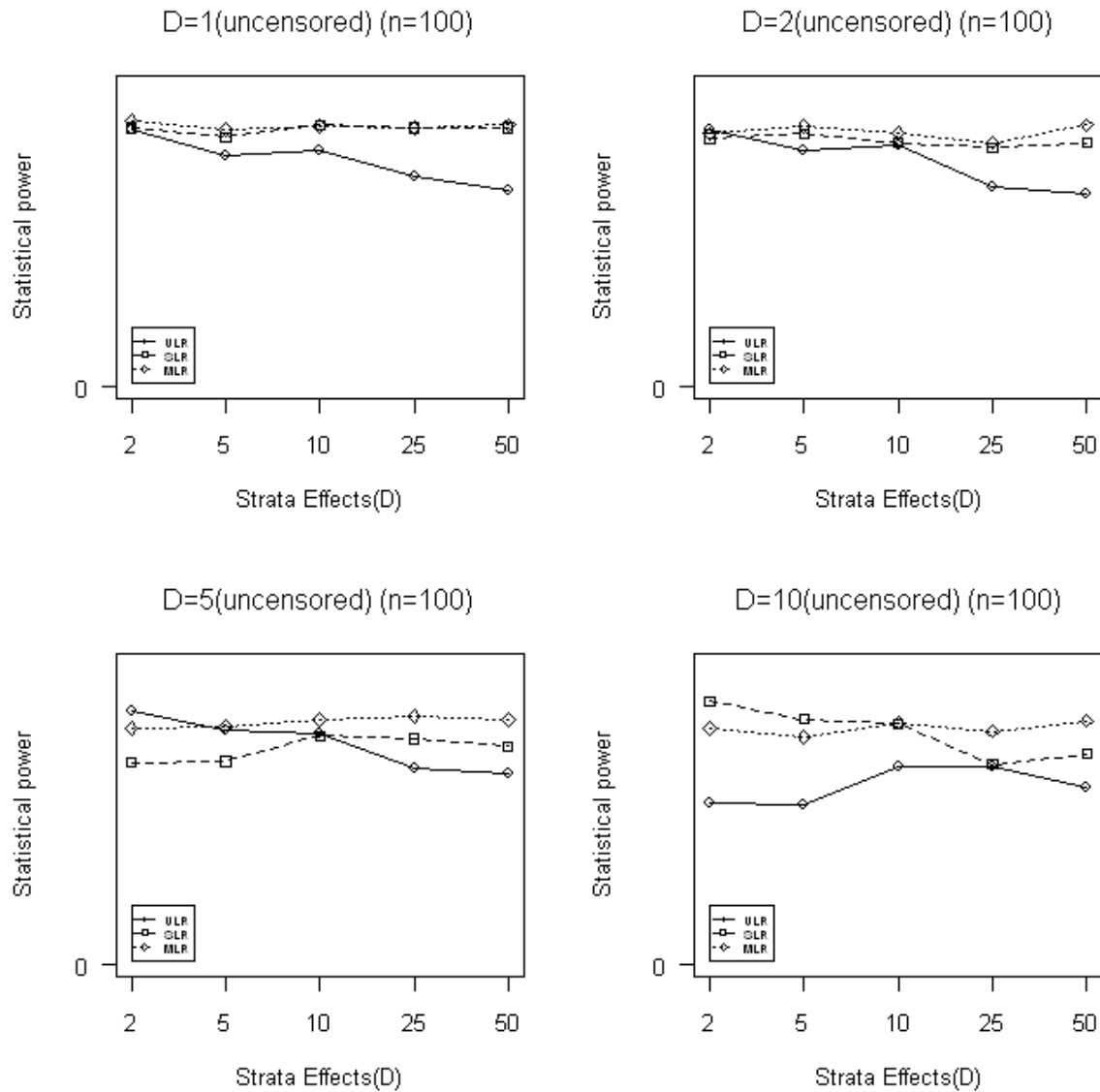
Table 8. Statistical power various according to stratum size ( n=100, D=10)

D=10		Uncensored					Censored				
Number of stratum	Stratum size	MLR	ULR	SLR	M/U *	S/U **	MLR	ULR	SLR	M/U *	S/U **
K=2	50	0.79	0.54	0.88	1.46	1.63	0.61	0.4	0.775	1.52	1.94
K=5	20	0.76	0.535	0.82	1.42	1.53	0.57	0.445	0.75	1.28	1.69
K=10	10	0.81	0.66	0.805	1.23	1.22	0.69	0.53	0.78	1.3	1.47
K=25	4	0.78	0.67	0.66	1.18	1.02	0.665	0.59	0.57	1.13	0.97
K=50	2	0.815	0.595	0.705	1.37	1.18	0.715	0.49	0.635	1.46	1.3

\*M/U: the ratio of MLR over ULR; \*\*S/U: the ratio of SLR over ULR

As the strata effects get higher, the patterns of statistical power for these log rank tests were becoming more and more complicated. With lower strata effects (D=1, 2), the power of all the three test methods were very close for k=2, but the power of ULR dropped dramatically with the increasing number of strata, while the power of the MLR and SLR tests were relatively stable. Also the ratio of the power of MLR over that of ULR and the power of SLR over that of ULR

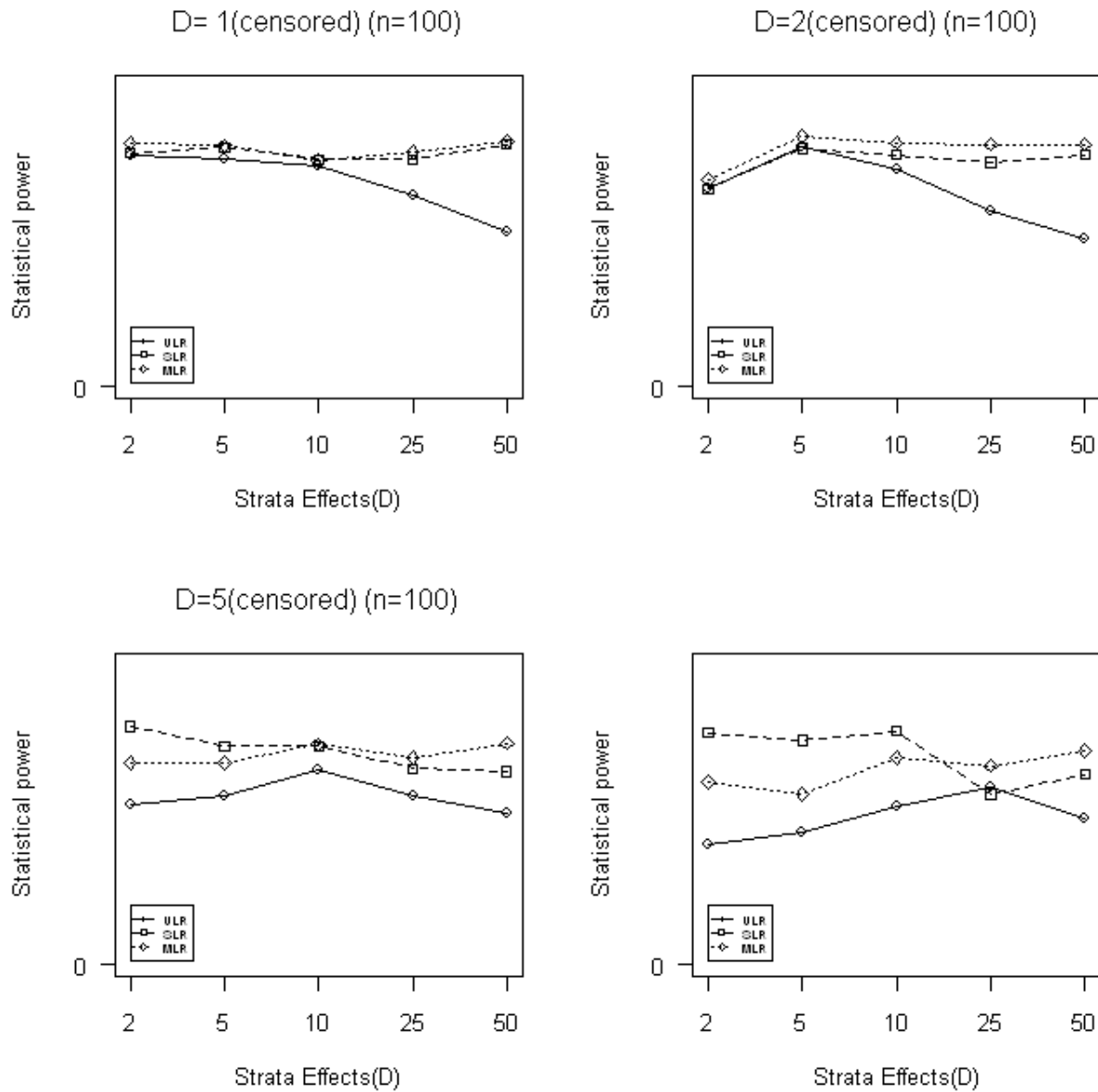
increased with the increasing of the number of strata. When the strata effects were high ( $D=5, 10$ ), the power of ULR was low for nearly all number of strata; the power of SLR was higher at the lower number of strata but dropped quickly when the number of strata getting bigger; but the



**Figure 5. Statistical power varies according to the number of stratum (HR=1, n=100)**

power of MLR kept stable higher for all the number of strata (Tables 7, 8 and Figure 5). For the censored data, the power of ULR dropped even sharply, the power of SLR was also significantly

decrease, only the power of MLR kept relatively higher and stable cross all the number of strata (tables 9,10 and Figure 6).



**Figure 6. Statistical power varies according to number of stratum ( n=100): Censored data**

### **3.3.2 Statistical power comparisons when the methods applied to the smaller sample (n=60)**

In this section, we compare the statistical power of the three different log rank tests when they were applied to the relatively smaller sample size (n=60). The simulation data were generated without censoring but with the different designs which were balanced design and unbalanced design. For the balanced design, we separated the total 60 patients evenly into two groups; for the unbalanced design, we signed 40 patients in one group, and the other 20 patients in the different groups. It is very interesting that for the balanced design, at D=1 (no strata effects) and D=2 (low strata effect), the power of SLR had dramatic change, it decreased from the same level as the powers of ULR and MLR at k=2 to the very lower level at higher number of strata (k=30). Generally, the ratio of SLR over to ULR is less than “1”, and the ration of MLR over ULR is higher than “1” (Tables 9, 10). When the strata effects were high (D=5,10), the power of SLR changed dramatically; it started at very high level at the lower number of strata, but dropped significantly with the increase of the number of strata, while the power of MLR was higher and stable. The power of ULR was relatively stable, but it was low across all the strata. For the unbalanced data, the results were hard to explanation, especially for the higher stratification. We might need further extensive studies on this issue (Tables 9,10 and Figure 7 ).

Table 9. Statistical power various according to stratum size ( n=60, D=1)

Uncensored ( HR=1.857 D=1): Balanced and unbalanced design

D=1		Balanced Design					Unbalanced Design				
Number of stratum	Stratum size	MLR	ULR	SLR	M/U *	S/U **	MLR	ULR	SLR	M/U *	S/U **
K=2	30	0.674	0.64	0.62	1.05	0.97	0.576	0.588	0.536	0.98	0.91
K=5	12	0.66	0.642	0.548	1.03	0.85	0.562	0.594	0.514	0.95	0.87
K=10	6	0.678	0.638	0.524	1.06	0.82	0.578	0.572	0.432	1.01	0.76
K=15	4	0.652	0.62	0.438	1.05	0.71	0.27	0.576	0.408	0.47	0.71
K=30	2	0.698	0.656	0.368	1.06	0.56	0.078	0.556	0.228	0.14	0.41

\*M/U: the ratio of MLR over ULR; \*\*S/U: the ratio of SLR over ULR

Table 10. Statistical power various according to stratum size ( n=60, D=2)

Uncensored ( HR=1.857 D=2): Balanced and unbalanced design

D=2		Balanced Design					Unbalanced Design				
Number of stratum	Stratum size	MLR	ULR	SLR	M/U *	S/U **	MLR	ULR	SLR	M/U *	S/U **
K=2	30	0.646	0.568	0.622	1.14	1.1	0.522	0.508	0.552	1.06	1.09
K=5	12	0.68	0.612	0.572	1.11	0.93	0.57	0.566	0.544	0.91	0.96
K=10	6	0.638	0.616	0.5	1.04	0.81	0.56	0.568	0.448	0.76	0.79
K=15	4	0.622	0.576	0.474	1.08	0.82	0.226	0.564	0.352	0.72	0.62
K=30	2	0.666	0.616	0.332	1.08	0.54	0.074	0.622	0.42	0.37	0.68

\*M/U: the ratio of MLR over ULR; \*\*S/U: the ratio of SLR over ULR

Table 11. Statistical power various according to stratum size ( n=60, D=5)

Uncensored ( HR=1.857 D=5): Balanced and unbalanced design

D=5		Balanced Design					Unbalanced Design				
Number of stratum	Stratum size	MLR	ULR	SLR	M/U *	S/U**	MLR	ULR	SLR	M/U *	S/U*
K=2	30	0.56	0.42	0.642	1.14	1.10	0.464	0.334	0.562	1.39	1.68
K=5	12	0.554	0.452	0.542	1.11	0.93	0.466	0.392	0.498	1.19	1.27
K=10	6	0.586	0.488	0.502	1.04	0.81	0.482	0.41	0.452	1.18	1.10
K=15	4	0.606	0.516	0.478	1.08	0.82	0.182	0.534	0.352	0.34	0.66
K=30	2	0.586	0.506	0.344	1.08	0.54	0.048	0.512	0.612	0.09	1.20

\*M/U: the ratio of MLR over ULR; \*\*S/U: the ratio of SLR over ULR

Table 12. Statistical power various according to stratum size ( n=60, D=10)

Uncensored ( HR=1.857 D=10): Balanced and unbalanced design

D=10		Balanced Design					Unbalanced Design***				
Number of stratum	Stratum size	MLR	ULR	SLR	M/U*	S/U*	MLR	ULR	SLR	M/U*	S/U*
K=2	30	0.566	0.3	0.662	1.89	2.21	0.414	0.23	0.58	1.80	2.52
K=5	12	0.554	0.36	0.594	1.54	1.65	0.404	0.242	0.516	1.84	2.13
K=10	6	0.544	0.368	0.504	1.48	1.37	0.444	0.286	0.446	1.67	1.56
K=15	4	0.534	0.378	0.446	1.41	1.18	0.152	0.474	0.32	1.55	0.68
K=30	2	0.542	0.39	0.332	1.39	0.85	0.028	0.458	0.738	0.32	1.61

\*M/U: the ratio of MLR over ULR; \*\*S/U: the ratio of SLR over ULR

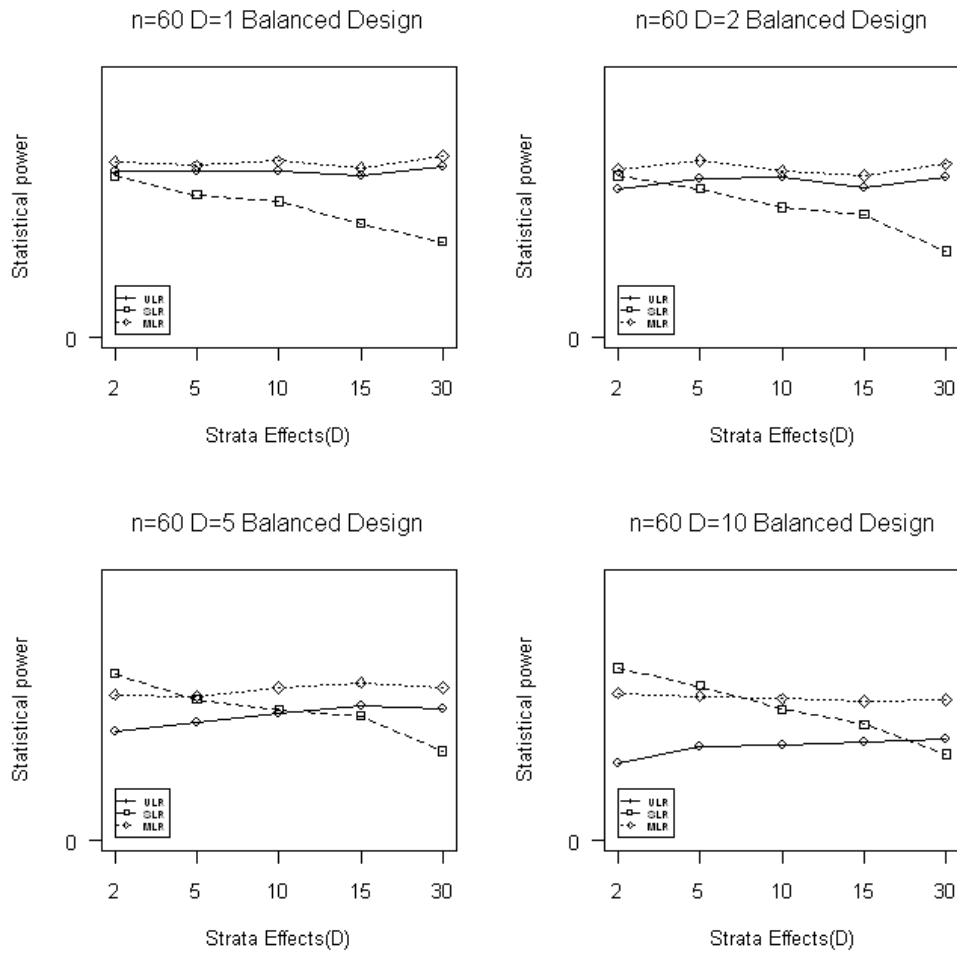
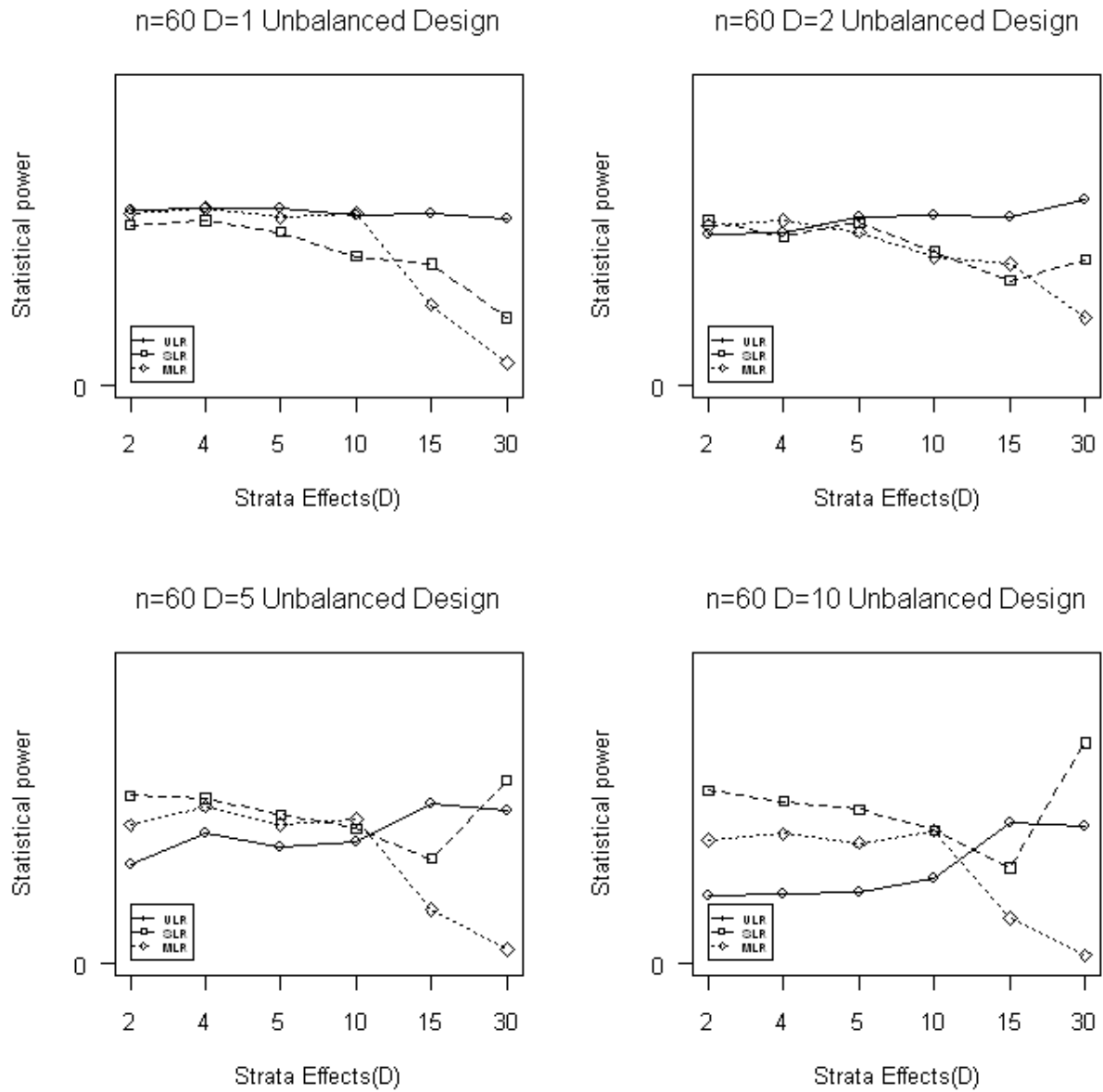


Figure 7. Statistical power varies according to number of stratum: Balanced Design





**Figure 8. Statistical power varies according to number of stratum: Unbalanced Design**

## 4.0 DISCUSSION

In this thesis, we evaluated the three log-rank tests using statistical simulations. The results showed that the stratum size is an important factor that affects the statistical power, confirming the previous results that the power loss of the stratified log-rank test can be caused by very small stratum sizes. Using our comparisons of the three log-rank tests, we confirmed that the modified log-rank test is beneficial for most cases of highly stratified data especially for the smaller sample size. Despite the fact that several previous studies indicated that the modified log-rank test tends to inflate the statistical power (Shih, 1999), the Schenfeld-Tsiatis test is, nevertheless, a very useful way to adjust for high stratification in studies, even though there might not be the advantages for using the modified log-rank test (Akazawa, et al., 1997) when the number of strata are relatively low. Concerning the suitable stratum size for applying to modified log-rank test, we might not have to worry too much about the stratum size when the balanced designs were used but we may have to be wary of the stratum size with the unbalanced design, since it appears that statistical power are very unstable when the stratum size dropped to 6 observations in our simulation data (Figure 8).

The strata effect has been also played a part in affecting statistical power. In this thesis, we addressed the strata effect from  $D=1$ , which means no strata effect to  $D=2, 5$ , and 10 four levels. The results indicate that the statistical power of all the three tests decreased with increasing of the strata effects for nearly all the set conditions, including censored and uncensored with

balanced and unbalanced design. When comparing to the unstratified log-rank test, which is invalid when there are the strata effects and the stratified log-rank test, we showed the advantages when the sizes of strata are not small but lost the advantages when the sizes of stratum got smaller, The statistical power of the modified log-rank test is relatively stable even with the very smaller sizes of stratum and the high strata effects.

Compared to the results with Schoenfeld & Tsiatis 1987, the statistical power in our simulation were close but slightly lower than that in their studies. The reasons for these difference affect the statistical power might need to be further investigated. It is also indicating that there might be still the other reasons can cause this difference.

To our knowledge, there are no statistical packages which employ the modified log-rank test suggested by Schoenfeld and Tsiatis (Schoenfeld & Tsiatis, 1987). One reason might be the complicated calculation formulas in this test. We tackled with this issue by using R and we finally got our R program to handle this relatively complicated test.

## 5.0 CONCLUSION

By comparing the three log-rank tests, we conclude that the modified log-rank test is superior to both the unstratified log-rank test and stratified log-rank test in most cases, especially when there is a large number of strata, which agree with the results of Scoenfeld.

In the current R package, there are programs for unstratified and stratified log-rank tests that have been built in. It would be useful if programs for the modified log-rank test can be built into the R package based on our current results.

## APPENDIX A

### MATH FORMULARS

#### A.1 Unstratified log-rank statistic:

$$\text{Stat}(\text{unstratification}) = \frac{\sum \Delta_i \{Z_i - \bar{Z}(X_i)\}}{[\sum \Delta_i \{X_i\} \{1 - \bar{Z}(X_i)\}]^{\frac{1}{2}}}$$

$$\bar{Z}(u) = \sum Z_j I(X_j \geq u) / \sum I(X_j \geq u)$$

#### A.2 Stratified log-rank statistic:

$$\text{Stat}(\text{stratified}) = \frac{\sum_j \sum_i \Delta_{ji} \{Z_{ji} - \bar{Z}_j(X_{ji})\}}{[\sum_j \sum_i \Delta_{ji} \{X_{ji}\} \{1 - \bar{Z}_j(X_{ji})\}]^{\frac{1}{2}}}$$

$$\bar{Z}_j(u) = \sum_{i=1}^{n_j} Z_{ji} I(X_{ji} \geq u) / \sum_{i=1}^{n_j} I(X_{ji} \geq u)$$

#### A.3 Stratified log-rank test as modified by Schoenfeld (Schoenfeld and Tsiatis, 1987):

$$L_{jia} = \sum_{k,m} \{q_k Z_{km} (2 - \alpha) + p_k (1 - Z_{km}) (\alpha - 1)\} I(X_{km} \gg X_{ji})$$

$$S_{jia} = L_{jia} / (pL_{ji1} + qL_{ji2})$$

$$T = \sum_{j,i} \Delta_{ji} \{Z_{ji} q_j S_{ji2} - (1 - Z_{ji}) p_j S_{ji1}\}$$

$$d_{jia} = \Delta_{ji} S_{jia} - \sum_{k,m} \left[ \frac{\{p q_k Z_{km} + q p_k (1 - Z_{km})\} \Delta_{km} I(X_{km} \leq X_{ji}) S_{kma}}{\sum_{k',m'} \{p q_{k'} Z_{k'm'} + q p_{k'} (1 - Z_{k'm'})\} I(X_{k'm'} \geq X_{km})} \right]$$

$$V = \sum_{j=1}^k [p_j^2 \sum_{i=1}^{n_j} (1 - Z_{ji}) d_{ji1}^2 + q_j^2 \sum_{i=1}^{n_j} Z_{ji} d_{ji2}^2 - n_j^{-1} \{ \sum_{i=1}^{n_j} (1 - Z_{ji}) d_{ji1} \} \{ \sum_{i=1}^{n_j} Z_{ji} d_{ji2} \}]$$

$$\text{Stat}(\text{m-stratified}) = T / V^{\frac{1}{2}}$$

## APPENDIX B

### R-CODE

#### **B.1** R -code for data generation and analysis

```
library(survival)
n1=30
n2=30

n<-n1+n2

h<-0.1 ###<--lambda: baseline hazard rate
D<-2   ###<-- Stratification effect: ratio of the highest to lowest hazard
      Rate (can be changed)
e<-0.6 #####<-- Treatment effect: e=0 => HR=1;
      #<-- e=0.6 => HR=1.857

K <-2  #<----Number of strata (can be changed)
N.sim <-1000
# -----
# INITIALIZE vectors with output information
# -----
lr.stat<-rep(NA,N.sim) # ORDINARY log-rank statistic (unstratified)
lr.stat.strat<-rep(NA, N.sim) # ORDINARY log-rank statistic (stratified)
lr.stat.modi.strat<-rep(NA, N.sim) # MODIFIED log-rank statistic (stratified)
p.val<-rep(NA, N.sim)#p-value for unstratified log-rank
p.val.strat<-rep(NA, N.sim) # p-value for stratified analysis
p.val.modi.strat<-rep(NA, N.sim) #P-value for modified analysis

#####SIMULATION LOOP #####

for(l in 1:N.sim){

  Z<-c(rep(1,n1),rep(0,n2))###control and treatment group
  J.<-c(rep(1:K,n/K))
  k<-c(1:K)
  hj<-h*(1+(D-1)*(k-1)/(K-1))
  #hj<-ifelse(J.==1,h,D*h)
  haj<-hj*(1+e/2); hbj<-hj*(1-e/2)###hazard rate for treatment A and B in
stratum j
  x.1<-rexp(n1,haj)####group1 (Z1)
  x.2<-rexp(n2,hbj)####group 2 (Z2)
```

```

x<-c(x.1,x.2)#####
d.<-rbinom(n,1,1) #<-- CENSORING INDICATOR (NO CENSORED)
#d.<-rbinom(n,1,0.85) #<-- CENSORING VARIABLE (15% CENSORED)

data.<-cbind(x,d.,J.,Z)
(o <- order(x))
data.[o, ]

data.simu<-data.frame(data.[o, ]##re-order the data frame with increasing x

##### unstratified and stratified log-rank tests using R functions #####
diff.lr.sv<-survdiff(Surv(x,d.)~Z,data=data.simu)#####unstratified
check.lr.sv<-coxph(Surv(x,d.)~Z,data=data.simu)
check.lr.sv
diff.lr.stra.sv<-survdiff(Surv(x,d.)~Z+strata(J.),data=data.simu)##stratified

#####----- code for Modified Logrank test-----#####

x.<-data.simu$x; d<-data.simu$d.;J<-data.simu$J.;Z.<-data.simu$Z
data.m<-cbind(x.,d,J,Z.) ##ensure to use the same data as above

ar.ind<-matrix(0,n,K)
for (i in 1:n){

j <- data.m[i,3]
ar.ind[i,j] <- 1

}

Zj.<-Z.*ar.ind ###pts with treatment in each stratum

nn<-length(x[Z.==1]) ##total number of treatment A in all (stratum 1&2)

nnj.<- c(rep(0,K))
nj. <- c(rep(0,K))
for (j in 1:K){
nnj.[j] <-length(x[Zj.[,j]==1])
nj.[j] <-length(x[J==j])
pj.<-nnj./nj.;qj.<-1-pj.
p.<-nn/n;q.<-1-p.
}

#####-----Working on Lj1&Lj2-----#####

At.risk<-matrix(rep(1,n^2), nrow=n)

for (i in 1:n){
for (j in 1:n){
At.risk[i,j]<-ifelse(j>i,0,1)
}
}

ar.ind<-matrix(0,n,K)
for (i in 1:n){
j <- data.m[i,3]
ar.ind[i,j] <- 1

```

```

}

Lj11<-apply(At.risk*qj.*Z.,2,sum)
Lj12<-apply(At.risk*pj.*(1-Z.),2,sum)

Sj11<-Lj11/(p.*Lj11+q.*Lj12); Sj12<-Lj12/(p.*Lj11+q.*Lj12)

Tj.<-d*(Z.*qj.*Sj12-(1-Z.)*pj.*Sj11)
modi.strat.T<-sum(Tj.)

#####-----Working on the denominators of statistics (Variance)-----#####

#####denormators for dj11&dj12#####
Ddn<-apply(At.risk*p.*qj.*Z.,2,sum)+apply(At.risk*q.*pj.*(1-Z.),2,sum)
Ddn ##Denominator for dj11&dj12

rDdn<-1/Ddn

#####---Fraction term for dj11&dj12---#####
At.event<-matrix(rep(0, n^2), nrow=n)
for (i in 1:n){
for (j in 1:n){
  At.event[i,j]<-ifelse(j>=i,1,0)
}
}

fnn<-p.*qj.*Z.+q.*pj.*(1-Z.)

sum1<-apply(At.event*rDdn*d*Sj11,2,sum)      ##q1; q2 for all term
sum2<-apply(At.event*rDdn*d*Sj12,2,sum)      ##q1;q2 for all term
sum1;sum2
frac1<-fnn*sum1;frac2<-fnn*sum2
frac1;frac2

dj11<-d.*Sj11-frac1;dj12<-d.*Sj12-frac2
dj11;dj12

Ddn<-apply(At.risk*p.*qj.*Z.,2,sum)+apply(At.risk*q.*pj.*(1-Z.),2,sum)

#####---variance calculation---#####

Vj.1<-(pj.^2)*((1-Zj.)*ar.ind*dj11^2)
Vj.1s<-c(apply(Vj.1,2,sum))

Vj.2<-(qj.^2)*(Zj.*dj12^2)
Vj.2s<-c(apply(Vj.2,2,sum))

Vj.3<-((1-Zj.)*ar.ind*dj11)
Vj.3s<-c(apply(Vj.3,2,sum))

Vj.4 <-Zj.*dj12
Vj.4s<-c(apply(Vj.4,2,sum))

Vj.<-Vj.1s+Vj.2s-(Vj.3s)*(Vj.4s)/nj.

```



```

modi.strat.var<-sum(Vj.)
modi.strat.V<-sqrt(modi.strat.var)

Z.modi.strat.Stat<-modi.strat.T/modi.strat.V

Z.diff.lr.stat.R<-sqrt(diff.lr.sv$chisq)
Z.diff.lr.stra.stat.R<-sqrt(diff.lr.stra.sv$chisq)
Z.diff.lr.stat.R
Z.diff.lr.stra.stat.R
Z.modi.strat.Stat

lr.stat[1]<-diff.lr.sv$chisq
lr.stat.strat[1]<-diff.lr.stra.sv$chisq
lr.stat.modi.strat[1]<-modi.strat.T^2/modi.strat.var

p.val[1]<-1-pchisq(diff.lr.sv$chisq,1)
p.val.strat[1]<-1-pchisq(diff.lr.stra.sv$chisq,1)
p.val.modi.strat[1]<-1-pchisq(lr.stat.modi.strat[1],1)

} #<-- END of simulation loop

alpha=0.05
sig.lr<-ifelse(p.val<alpha,1,0)
power.lr<-sum(sig.lr)/N.sim
power.lr
#power.all.ULR[k]<-power.lr

sig.strat<-ifelse(p.val.strat<alpha,1,0)
power.strat<-sum(sig.strat)/N.sim
power.strat
#power.all.SLR<-power.strat

p.val.modi.strat<-na.omit(p.val.modi.strat)
sig.modi.strat<-ifelse(p.val.modi.strat<alpha,1,0)
power.modi.strat<-sum(sig.modi.strat)/N.sim
power.modi.strat
#power.all.MLR[k]<-power.modi.strat
cbind(power.modi.strat,power.lr,power.strat)

```

---

## B.2 R -code for graphing

```
par(mfrow=c(2,1))

ULR<-c(0.053,0.067,0.054,0.042,0.063)
SLR <-c(0.061,0.062,0.049,0.046,0.069)
MLR<-c(0.052,0.074,0.059,0.054,0.075)
g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c("2","5","10","25","50"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.8, c("ULR","SLR","MLR"), cex=0.5,
       pch=21:23, lty=1:3);
title(main="Baseline:D=1,HR=1(uncensored,n=100)", font.main=1)
```

```
#D=1, h=1, censored
ULR<-c(0.052,0.048,0.044,0.062,0.062)
SLR <-c(0.057,0.052,0.04,0.062,0.05)
MLR<-c(0.062,0.052,0.052,0.066,0.068)
g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c("2","5","10","25","50"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.8, c("ULR","SLR","MLR"), cex=0.5,
       pch=21:23, lty=1:3);
title(main="Baseline:D=1,HR=1(censored,n=100)", font.main=1)
```

```
#####-----#####
```

```
par(mfrow=c(2,2))
#D=1, h=1.857,n=100 uncensored

ULR<-c(0.86,0.775,0.79,0.705,0.655)
SLR <-c(0.865,0.835,0.875,0.865,0.865)
MLR<-c(0.89,0.86,0.87, 0.865,0.875)

g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c("2","5","10","25","50"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
```

```

lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
      pch=21:23, lty=1:3);
title(main="D=1(uncensored) (n=100)", font.main=1)

```

```

#D=2, h=1.857, uncensored
ULR<-c(0.86,0.79,0.81,0.67,0.645)
SLR <-c(0.83,0.845,0.815,0.8,0.815)
MLR<-c(0.845,0.87,0.85,0.815,0.875 )

```

```

g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c("2","5","10","25","50"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

```

```

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
      pch=21:23, lty=1:3);
title(main="D=2(uncensored) (n=100)", font.main=1)

```

```

#D=5, h=1.857, uncensored
ULR<-c(0.85,0.785,0.77,0.655,0.64)
SLR <-c(0.675,0.68,0.765,0.755,0.73 )
MLR<-c(0.79,0.795,0.82,0.83,0.82)

```

```

g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c("2","5","10","25","50"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

```

```

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
      pch=21:23, lty=1:3);
title(main="D=5(uncensored) (n=100)", font.main=1)

```

```

#D=10, h=1.857, uncensored
ULR<-c(0.54,0.535,0.66,0.66,0.595)
SLR <-c(0.88,0.82,0.805,0.67,0.705 )
MLR<-c(0.79,0.76,0.81,0.78,0.815)

```

```

g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)

```

```

axis(1, at=1:5, lab=c("2","5","10","25","50"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
       pch=21:23, lty=1:3);
title(main="D=10(uncensored) (n=100)", font.main=1)

##### censored #####

par(mfrow=c(2,2))
#D=1, h=1.857, n=100 censored
ULR<-c(0.775,0.76,0.735,0.64,0.52)
SLR<-c(0.78,0.8,0.76,0.76,0.81)
MLR<-c(0.815,0.805,0.755,0.785,0.82)
g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c("2","5","10","25","50"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
       pch=21:23, lty=1:3);
title(main="D= 1(censored) (n=100)", font.main=1)

#D=2, h=1.857,censored

ULR<-c(0.66,0.8,0.725,0.585,0.495)
SLR<-c(0.66,0.795,0.775,0.75,0.775)
MLR<-c(0.69,0.835,0.815,0.81,0.81)

g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c("2","5","10","25","50"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
       pch=21:23, lty=1:3);
title(main="D=2(censored) (n=100)", font.main=1)

#D=5, h=1.857, censored

```

```

ULR<-c(0.535,0.565,0.65,0.565,0.505)
SLR<-c(0.795,0.73,0.73,0.655,0.645)
MLR<-c(0.675,0.675,0.735,0.69,0.74)

g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c("2","5","10","25","50"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
       pch=21:23, lty=1:3);
title(main="D=5(censored) (n=100)", font.main=1)

#D=10, h=1.857,censored
ULR<-c(0.4,0.445,0.53,0.59,0.49)
SLR<-c(0.775,0.75,0.78,0.57,0.635)
MLR<-c(0.61,0.57,0.69,0.665,0.715)

g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c("2","5","10","25","50"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
       pch=21:23, lty=1:3);

#####----small data set n=60(uncensored balanced data)----#####
par(mfrow=c(2,1))
#D=1, h=1, balanced(baseline) n=60
ULR<-c(0.062,0.054,0.07,0.06,0.048,0.066)
SLR <-c(0.06,0.06,0.066,0.042,0.052,0.052)
MLR<-c(0.082,0.054,0.078,0.062,0.058,0.082)
g_range<-range(0,1)
plot(ULR, type="o",ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:6, lab=c("2","4","5","10","15","30"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.8, c("ULR","SLR","MLR"), cex=0.5,

```

```

    pch=21:23, lty=1:3);
title(main="Baseline:D=1,HR=1(Balanced,n=60)", font.main=1)
#####-----unblanced design-----#####
#D=1, h=1, unblanced(baseline)
ULR<-c(0.064,0.068,0.066,0.058,0.076,0.066)
SLR <-c(0.06,0.072,0.066,0.048,0.06,0.044)
MLR<-c(0.08,0.08,0.086,0.08,0.07,0.054)
g_range<-range(0,1)
g_range<-range(0,1)
plot(ULR, type="o",ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:6, lab=c("2","4","5","10","15","30"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.8, c("ULR","SLR","MLR"), cex=0.5,
      pch=21:23, lty=1:3);
title(main="Baseline:D=1,HR=1(Unbalanced,n=60)", font.main=1)
#####-----balanced design-----#####
#D=1, h=1.875, balanced n=60
par(mfrow=c(2,2))

ULR<-c(0.64,0.642,0.638,0.62,0.656)
SLR <-c(0.62,0.548,0.524,0.438,0.368)
MLR<-c(0.674,0.66,0.678,0.652,0.698)
g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c("2","5","10","15","30"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
      pch=21:23, lty=1:3);
title(main=" n=60 D=1 Balanced Design ", font.main=1)

ULR<-c(0.568,0.612,0.616,0.576,0.616)
SLR <-c(0.622,0.572,0.5,0.474,0.332)
MLR<-c(0.646,0.68,0.638,0.622,0.666)
g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c("2","5","10","15","30"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,

```

```

    pch=21:23, lty=1:3);
title(main=" n=60 D=2 Balanced Design ", font.main=1)

ULR<-c(0.42,0.452,0.488,0.516,0.506)
SLR <-c(0.642,0.542,0.502,0.478,0.344)
MLR<-c(0.56,0.554,0.586,0.606,0.586)
g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c("2","5","10","15","30"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
      pch=21:23, lty=1:3);
title(main=" n=60 D=5 Balanced Design ", font.main=1)

ULR<-c(0.3,0.36,0.368,0.378,0.39)
SLR <-c(0.662,0.594,0.504,0.446,0.332)
MLR<-c(0.566,0.554,0.544,0.534,0.542)
g_range<-range(0,1)
plot(ULR, type="o", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c("2","5","10","15","30"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
      pch=21:23, lty=1:3);
title(main=" n=60 D=10 Balanced Design ", font.main=1)

#####----small data set n=60(uncensored unbalanced data)----#####

par(mfrow=c(2,2))

ULR<-c(0.588,0.594,0.594,0.572,0.576,0.556)
SLR <-c(0.536,0.552,0.514,0.432,0.408,0.228)
MLR<-c(0.576,0.594,0.562,0.578,0.27,0.078)
g_range<-range(0,1)
plot(ULR, type="o", col="blue", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:6, lab=c("2","4","5","10","15","30"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")

```

```

legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
      pch=21:23, lty=1:3);
title(main=" n=60 D=1 Unbalanced Design ", font.main=1)

```

```

ULR<-c(0.508,0.514,0.566,0.568,0.564,0.622)
SLR <-c(0.552,0.498,0.544,0.448,0.352,0.42)
MLR<-c(0.536,0.552,0.514,0.432,0.408,0.228)
g_range<-range(0,1)
plot(ULR, type="o", col="blue", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:6, lab=c("2","4","5","10","15","30"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

```

```

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
      pch=21:23, lty=1:3);
title(main=" n=60 D=2 Unbalanced Design ", font.main=1)

```

```

ULR<-c(0.334,0.436,0.392,0.41,0.534,0.512)
SLR <-c(0.562,0.552,0.498,0.452,0.352,0.612)
MLR<-c(0.464,0.526,0.466,0.482,0.182,0.048)
g_range<-range(0,1)
plot(ULR, type="o", col="blue", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:6, lab=c("2","4","5","10","15","30"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

```

```

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
      pch=21:23, lty=1:3);
title(main=" n=60 D=5 Unbalanced Design ", font.main=1)

```

```

ULR<-c(0.23,0.236,0.242,0.286,0.474,0.458)
SLR <-c(0.58,0.542,0.516,0.446,0.32,0.738)
MLR<-c(0.414,0.434,0.404,0.444,0.152,0.028)
g_range<-range(0,1)
plot(ULR, type="o", col="blue", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:6, lab=c("2","4","5","10","15","30"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

```

```

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
      pch=21:23, lty=1:3);
title(main=" n=60 D=10 Unbalanced Design ", font.main=1)

```



```
#####----Stiatrics power with strata effects ----#####  
par(mfrow=c(2,2))
```

```
ULR<-c(0.775,0.79,0.6,0.62)  
SLR <-c(0.835,0.845,0.81,0.82)  
MLR<-c(0.86,0.87,0.8, 0.79)
```

```
g_range<-range(0,1)  
plot(ULR, type="o", col="blue", ylim=g_range,  
      axes=FALSE, ann=FALSE)  
axis(1, at=1:4, lab=c("2","4","5","10"))  
axis(2,las=2,at=0.2*0:g_range[1])  
box()  
lines(ULR, type="o", pch=21, lty=1 )  
lines(SLR, type="o", pch=22, lty=2 )  
lines(MLR, type="o", pch=23, lty=3)  
  
title(xlab= "Strata Effects(D)")  
title(ylab= "Statistical power")  
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,  
       pch=21:23, lty=1:3);  
title(main="n=100 K=5 Noncensored with Balanced", font.main=1)
```

```
ULR<-c(0.805,0.835,0.675,0.57)  
SLR <-c(0.76,0.8,0.73,0.75)  
MLR<-c(0.8,0.795,0.69,0.68)
```

```
g_range<-range(0,1)  
plot(ULR, type="o", col="blue", ylim=g_range,  
      axes=FALSE, ann=FALSE)  
axis(1, at=1:4, lab=c("2","4","5","10"))  
axis(2,las=2,at=0.2*0:g_range[1])  
box()  
lines(ULR, type="o", pch=21, lty=1 )  
lines(SLR, type="o", pch=22, lty=2 )  
lines(MLR, type="o", pch=23, lty=3)  
  
title(xlab= "Strata Effects(D)")  
title(ylab= "Statistical power")  
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,  
       pch=21:23, lty=1:3);  
title(main="n=100 K=5 Censored with Balanced", font.main=1)
```

```
ULR<-c(0.642,0.612,0.452,0.36)  
SLR <-c(0.548,0.572,0.542,0.594)  
MLR<-c(0.66,0.68,0.554,0.554)
```

```
g_range<-range(0,1)  
plot(ULR, type="o", col="blue", ylim=g_range,  
      axes=FALSE, ann=FALSE)  
axis(1, at=1:4, lab=c("2","4","5","10"))  
axis(2,las=2,at=0.2*0:g_range[1])  
box()  
lines(ULR, type="o", pch=21, lty=1 )  
lines(SLR, type="o", pch=22, lty=2 )  
lines(MLR, type="o", pch=23, lty=3)
```

```

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
      pch=21:23, lty=1:3);
title(main="n=60 K=5 with Balanced", font.main=1)

ULR<-c(0.594,0.566,0.392,0.242)
SLR <-c(0.514,0.544,0.498,0.516)
MLR<-c(0.562,0.57,0.466,0.404)

g_range<-range(0,1)
plot(ULR, type="o", col="blue", ylim=g_range,
     axes=FALSE, ann=FALSE)
axis(1, at=1:4, lab=c("2","4","5","10"))
axis(2,las=2,at=0.2*0:g_range[1])
box()
lines(ULR, type="o", pch=21, lty=1 )
lines(SLR, type="o", pch=22, lty=2 )
lines(MLR, type="o", pch=23, lty=3)

title(xlab= "Strata Effects(D)")
title(ylab= "Statistical power")
legend(1, 0.2, c("ULR","SLR","MLR"), cex=0.5,
      pch=21:23, lty=1:3);
title(main="n=60 K=5 with Unbalanced", font.main=1)

```

#####----- End of code-----#####

## BIBLIOGRAPHY

- Akazawa, K., Nakamura, T., & Palesch, Y. (1997). Power of Logrank Test and Cox Regression Model in Clinical Trials with Heterogeneous Samples. *Statistics in Medicine*, 16, 583-597.
- Changyong Feng, Hongyue Wang, & M.Tu, X. (2010). Power Loss of Stratified Log-Rank test in Homogeneous Samples. *International Journal of Quality, Statistics, and Reliability*, 2010.
- Day, N. E., Byar, D. P., & Green, S. B. (1980). Overadjustment in case-control studies. *Am J Epidemiol*, 112(5), 696-706.
- De Stavola, B. L., & Cox, D. R. (2008). On the Consequences of Overstratification. *Biometrika* 95(4), 992-996.
- Feng, C. Y., Wang, H. Y., & Tu, X. M. (2010). Power Loss of Stratified Log-Rank test in Homogeneous Samples. *International Journal of Quality, Statistics, and Reliability*, 2010, 1-4.
- Green, S. B., & Byar, D. P. (1978). The effect of stratified randomization on size and power of statistical tests in clinical trials. *J Chronic Dis*, 31(6-7), 445-454.
- Lagakos, S. W. (1988). The loss of efficiency from misspecifying covariates in proportional hazards regression models. *biometrika*, 75, 156-160.
- Liao, H. W. (1998). A simulation study of estimation in stratified proportional hazards model. *NESUG 1998 Proceedings*, 118-125.
- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemo. Rep.*, 50, 163-170.
- Neuhauser, M., & Becher, H. (1997). Improved odds ratio estimation by post hoc stratification of case-control data. *Stat Med*, 16(9), 993-1004. Peduzzi, P., Henderson, W., Hartigan, P., & Lavori, P. (2008). Generalizability of the results of randomized trials. *Arch International Medicine*, 168(2), 133-135.
- Podgor, M. J., & Gastwirth, J. L. (1994). A Cautionary Note on Applying Scores in Stratified Data. *Biometrics*, 50(4), 1215-1218.

Schoenfeld, D. A., & Tsiatis, A. A. (1987). A modified log rank test for highly stratified data. *Biometrika*, 74(1), 167-175.

Shih, J. H. (1999). A class of permutation tests for stratified survival data. *Biometrics*, 55(4), 1156-1161.