NONPARAMETRIC CONDITIONAL MEAN CUMULATIVE FUNCTIONS FOR COMPARISON OF RANDOM-INTERVAL COUNTING PROCESSES WITH APPLICATIONS TO PANEL COUNT DATA ANALYSIS IN MEDICAL STUDY

TAN PEI LING

IPM 2019 8

NONPARAMETRIC CONDITIONAL MEAN CUMULATIVE FUNCTIONS FOR COMPARISON OF RANDOM-INTERVAL COUNTING PROCESSES WITH APPLICATIONS TO PANEL COUNT DATA ANALYSIS IN MEDICAL STUDY

By

TAN PEI LING

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirements for the Doctor of Philosophy

December 2018

DEDICATIONS

This thesis is gratefully dedicated to
my family, my supervisors, friends
and those who contributed to the manuscript along the way.

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

NONPARAMETRIC CONDITIONAL MEAN CUMULATIVE FUNCTIONS FOR COMPARISON OF RANDOM-INTERVAL COUNTING PROCESSES WITH APPLICATIONS TO PANEL COUNT DATA ANALYSIS IN MEDICAL STUDY

By

TAN PEI LING

December 2018

Chairman :**Prof. Noor Akma Ibrahim, PhD**
Faculty :**Institute for Mathematical Research**

In medical study, patients are treated with different treatments may have different follow-up schedule. Patients will visit clinic periodically, the actual time-to-event occurrence of disease is unknown, and only the number of occurrence between two consecutive visits is recorded. This is also referred as panel count data. Whenever the event have multiple occurrence, the event process is referred as recurrence process and is treated as a random point process over the follow-up time. A broad range of test procedures have been proposed for continuous observation processes in time-to-event data analysis, but only a few test procedures are applicable for discrete time observations when only panel count data are available. In practice, the number of clinical visits and clinical visit times are different for each patient. i.e, the observation processes are not identical. The number of patients assigned in each treatment group could also be imbalanced. However, most of the existing nonparametric test procedures proposed for treatment effectiveness comparison assume that each treatment has identical observation processes and are conducted for balanced sample size. When the observation processes between treatments are different, the existing test procedures are less significance in detecting the departure from the null hypothesis and provide misleading results. To address this, the study is focused on the development of a nonparametric test procedure which is constructed based on the integrated weighted differences between the mean cumulative function of the recurrences event with condition on treatment group. The test procedure is also extended to take into account multivariate recurrence processes, when the recurrent process has multi-type events. The empirical power of the proposed test statistics in detecting the departure from the null hypothesis are evaluated via Monte Carlo simulation study. The findings show that the proposed method works well under the tested situations. For efficiency comparison, the proposed test is evaluated through real data analysis and the results are in line with earlier research.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

FUNGSI KUMULATIF MIN BERSYARAT TAK BERPARAMETER BAGI PERBANDINGAN PROSES PENGIRAAN SELANG RAWAK DENGAN APLIKASI KE ATAS ANALISIS DATA KIRAAN PANEL DALAM KAJIAN PERUBATAN

Oleh

TAN PEI LING

Disember 2018

Pengerusi :**Prof. Noor Akma Ibrahim, PhD**
Fakulti :**Institut Penyelidikan Matematik**

Dalam kajian perubatan, pesakit yang dirawat dengan rawatan yang berbeza mungkin mempunyai jadual susulan yang berlainan. Pesakit akan melawat klinik secara berkala, masa kejadian penyakit tidak diketahui, dan bilangan kejadian di antara dua lawatan berturut-turut direkod. Ini juga dirujuk sebagai data pengiraan panel. Apabila kejadian adalah pelbagai, proses kejadian dirujuk sebagai proses berulang dan dianggap sebagai proses titik rawak sepanjang masa rawatan susulan. Pelbagai prosedur ujian telah dicadangkan untuk proses pemerhatian yang berterusan dalam analisis data masa ke masa, tetapi hanya beberapa prosedur ujian dicadangkan untuk pemerhatian masa diskret apabila yang ada hanya data pengiraan panel. Bilangan lawatan klinikal dan masa lawatan klinikal adalah berbeza bagi setiap pesakit. Iaitu, proses pemerhatian tidak sama. Bilangan pesakit dalam setiap kumpulan rawatan mungkin tidak seimbang. Walau bagaimanapun, kebanyakan prosedur ujian tak berparameter yang sedia ada yang dicadangkan untuk perbandingan keberkesanan rawatan mengandaikan bahawa setiap rawatan mempunyai proses pemerhatian yang sama dan dijalankan untuk saiz sampel yang seimbang. Apabila proses pemerhatian antara rawatan berbeza, prosedur ujian sedia ada kurang cekap dalam mengesan sisihan dari hipotesis nol dan memberikan keputusan yang silap. Untuk menangani masalah ini, kajian ini memberi tumpuan kepada lanjutan prosedur ujian tak berparameter yang dibina berdasarkan perbezaan berwajaran terintegrasi antara fungsi kumulatif min bagi kejadian berulang bersyarat pada kumpulan rawatan. Prosedur ujian juga diperluas untuk mengambil kira proses berulang pelbagai pembolehubah, apabila proses berulang mempunyai pelbagai jenis peristiwa. Kuasa empirikal statistik ujian yang dicadangkan dalam mengesan sisihan dari hipotesis nol dinilai melalui kajian simulasi Monte Carlo. Hasil kajian simulasi menunjukkan bahawa kaedah yang dicadangkan adalah baik di bawah situasi yang diuji. Bagi kecekapan perbandingan, ujian yang dicadangkan dinilai melalui analisis data sebenar dan keputusannya selaras dengan kajian terdahulu.

ACKNOWLEDGEMENTS

This thesis is the research work I have done under the supervision of my advisor Prof. Dr. Noor Akma Ibrahim and my co-supervisors Assoc. Prof. Dr. Jayanthi Arasan as well as Assoc. Prof. Dr. Mohd. Bakri Adam. I am deeply grateful to them for their patient guidance, insightful recommendations and suggestions. I sincerely appreciate all of their support for the past three years journey of getting my degree of Doctor of Philosophy.

This thesis is supported by MyBrain15 fund from Ministry of Higher Education of Malaysia. I would like to thank the Ministry for the financial support during my Ph.D. study. I also would like to express my sincere gratitude to Institute for Mathematical Research (INSPEM) for providing a convenience environment and facilities for conducting research.

Last but not least, a deep gratitude to my family and friends for their unconditional love and support, which have really encouraged me a lot when I was facing hardships and barriers.

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature:_____ Date:_____

Name and Matric No: Tan Pei Ling, GS39073

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: _____
Name of Chairman of Supervisory Committee:
Noor Akma Ibrahim

Signature: _____
Name of Member of Supervisory Committee:
Mohd Bakri Adam

Signature: _____
Name of Member of Supervisory Committee:
Jayanthi Arasan

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MCF	Mean Cumulative Function
NPMLE	Nonparametric Maximum Likelihood Estimator
NPMPLE	Nonparametric Maximum Pseudo Likelihood Estimator
IRE	Isotonic Regression Estimator
BCE	Basal Cell Epithelioma
BCC	Basal Cell Carcinoma
SCC	Squamous Cell Carcinoma
DFMO	Difluoromethylornithine
cheno	Chenodeoxycholic acid
HPP	Homogeneous Poisson Process
NHPP	Nonhomogeneous Poisson Process
ACS	The American Cancer Society
VACURG	Veterans Administration Cooperative Urological Research Group

CHAPTER 1

INTRODUCTION

This chapter provides some preliminary of the study and summary of research work. Section 1.1 briefly describes the background and Section 1.2 is on the motivation of the study. The purpose of study and research objectives are listed in Section 1.3, followed by conclusion in Section 1.4. The topics discuss in the remaining chapters are provided in Section 1.5.

1.1 Background

This thesis deals with a particular kind of random process. The central idea is to study random collections of point occurrences in medical follow-up study. For the most part, the points are considered as occurring along the time axis. Some examples arise from medical follow-up study are outline here to illustrate the breadth of potential applications. One may consider the sequence of time points at which systems failed, such as the development of new cancer types, growth of tumour cells and so forth. Whenever the event may have multiple occurrence, the event process is referred as recurrence process. In several of these examples, each point may be classified into one of several types or classes. For example, in non-melanoma skin cancer study, skin cancer can be distinguish into two types, Squamous cell carcinoma and Basal cell carcinoma. Alternatively, one may wish to consider several types of sickness symptoms, vomiting, high fewer or severe/chronic pain. A point process with several types or classes of point is called multivariate.

This research deals with the study of several processes of potential importance in applications, including the development of nonparametric techniques for statistical analysis of data arise from medical follow-up study and investigating its performance by assuming reasonable conditions of all properties under studied. We consider the case where the processes are independent of each other and the censoring process are non-informative. Nevertheless, the nonparametric technique is generalized to compare multiple samples from different treatments (processes, environments, system designs, operating conditions, and so forth).

1.2 Motivation

In medical follow-up study, disease process or the recurrent events evolve in continuous time, and patients are often monitored at irregular time points depend on treatment assignments. In such case, each subject is observed only at several distinct time points, where the exact time of occurrence of disease or recurrent event is unknown and only the number of events that have occurred prior to each observation

time is recorded. These produce panel count data which consist only the number of occurrences of events over a range of time between visits, it is also referred as interval count data of recurrent events. Due to the lack of information, it is common to work with the mean of recurrences or recurrence rate.

Nelson (2003) describes the nonparametric mean cumulative function (MCF) estimator for recurrent event data on continuous observation process and also on periodical observation process with some example arising from product repairs, disease recurrences and warranty claims. The MCF is not commonly used in panel count data as nonparametric maximum likelihood estimator (NPMLE) and isotonic regression estimator (IRE). Based on Wellner and Zhang (2000), NPMLE is more efficient than IRE. However, NPMLE is much more computationally complex and time consuming. IRE is restricted to non-decreasing step function of the recurrent events. However, in practice, some events may have an increasing or decreasing step functions. The MCF estimator can be applied to both increase or decrease functions and it is useful on observing an event changes over times.

A broad range of nonparametric test procedures for comparing recurrent event processes based on panel count data can be divided into two types with respect to the estimator of the mean function of the processes used in test statistics. Most common method used based on the nonparametric maximum pseudolikelihood estimator (NPMPLE) or IRE which is generally the least-square method includes the test used in Balakrishnan and Zhao (2010a), Li et al. (2010), Park et al. (2007), Sun and Fang (2003), Zhang (2006), and Zhao and Sun (2011); the other is constructed based on the nonparametric maximum likelihood estimator (NPMLE) which is generally more computationally complex, it involves iteration process and does not always reach convergence (Balakrishnan and Zhao, 2009). These methods share the same property where they require identical observation processes for all subjects except that given in Zhao and Sun (2011) and Li et al. (2014). Zhao and Sun (2011) incorporate Welch procedure for invariance test into IRE based on two sample hypothesis testing to allow unequal observation processes across treatment groups. Whilst, Li et al. (2014) proposed a class of nonparametric test for univariate and multivariate case by comparing its sample mean responses of the underlying recurrent process given covariates.

Note that the observation times are rarely equally spaced over a discrete time interval, it can be different on each subject, and the interval between each observation can also be random. Due to this, the discrete time Markov process assumption can not be used, and it is usual that, the disease process is characterized by counting process with Poisson structure. Furthermore, most of the existing methods assumed that the group indicators are identical and independent (i.e., the sample sizes are equal or balance across treatment groups). Therefore, there arises the need to extend the existing methods to analyse panel count data with unequal observation process for both univariate and multivariate problems.

1.3 Problem Statement

In medical research, clinical follow-up study plays an important role to study the safety and efficacy of the clinical interventions; mainly used to evaluate new approaches, such as surgery or radiotherapy and new ways to improve the diagnosis of diseases, and test the new drugs or combinations of drugs on patient to improve patients' quality of life and potentially expand lifespan. This kind of study is usually produces panel count data, which involves recurrent events and multi-type events and each patient is observed or examined only at periodic follow-up assessments.

The main interest is to compare the effectiveness of different treatments in medical follow-up studies. Panel count data involves more than one observation times point for each subject and which may differ from subject to subject. The clinical follow-up schedule for different treatments might have different follow-up patterns. In this setting, the observation scheme could be different or unequal observation process. The times interval between two subsequent treatments are unequal, the mean recurrence might be nonmonotonic. The existing methods which assumed identical observation processes and/or required recurrent processes to be non-decreasing monotonic function may not be appropriate and inefficient. A lot of effort has been spent on deriving the estimators and modeling the panel data, see Li et al. (2011), Li et al. (2013), Hua and Zhang (2012) and Hua et al. (2014) among others. The works are considering the effect of covariate measurements on the intensity of a recurrent phenomenon for single type of recurrent event. It is useful to have statistical methods which allowing both the systematic and random observation scheme in univariate or multivariate cases, and motivate us to develop a statistic test for panel count data analysis.

It is usual that medical research study done with restricted inclusion criteria in order to obtain a homogeneous sample of patients, representative only of specific sub-groups of population of a limited size. In control clinical studies or experimental studies, the samples which are compared are similar in those characteristics which may have effect on the response. Due to the constraints on resources to observe the process and collect the data in medical follow-up study, it is difficult to have data with both large number of observational units and a long time of observation period. In this study, we explore and evaluate the proposed test for samples with relative small size and imbalanced, as well as the effect of length of observation period on the performance of proposed method.

1.3.1 Objectives

The research intends to achieve the following objectives:

1. Derive and develop a test statistic and its variance computation which allowed unequal observation processes across treatment groups.
2. Evaluate the performance of the test statistic by its statistical power for different weight processes via simulation studies.
3. Compare the performance of the test statistic when the sample sizes are imbalanced and or relatively small between treatment groups, and explore the impact of the length of follow-up period and the degree of unequal observation on the test statistic.
4. Extend the test statistics to multiple samples and multivariate panel count data in a more general set-up of multivariate counting processes allowing for multiple jumps and furthermore allowing unequal observation processes across treatment groups.
5. Compare the performance of the test statistic with existing test through the real data analysis in medical follow-up study.

1.3.2 Scope of study

In ordinary continuous time recurrent event or survival data analysis, the actual occurrence times of the event are known, the censoring process is usually modeled to allow the random observation process. The censoring process can be non-informative or informative. When the censoring process is dependent of the observation process and or the recurrent process, the censoring is informative, one might model the censoring process. Study on the degree of censoring on the performance of test can be done. In this study, we consider non-informative censoring process, when the censoring is independent of observation process and recurrent process, and the censoring times are the longest follow-up time for each patient. In this case, we did not model the censoring process and study on the censoring process is limited. The proportion of censoring can be computed as number of censored over total number sample size. We consider the correlation between subjects as the random effect modeled by mixed Poisson process which is independent of the observation process and censoring process. However, the recurrent process might be correlated with censoring process or the observation process. For informative censoring processes, it is necessary to model the censoring processes and study the effect of the heavy censoring observations on the performance of the test statistic. On the other hand, the observation process are continuous in ordinal time-to-event data analysis, sufficient test is conducted to decide whether or not the period of observation has been long enough to detect the presence of cured (immune) individuals in the study, as in cured model analysis. As the length of follow-up study will affect the sufficiency of making inferences about the cure rates. In medical follow-up study, the length of follow-up

is planned according to the patient' health conditions and type of treatments. In this study, we assume that the observation processes (the follow-up) is independent of the recurrent processes. In other word, the length of follow-up does not affect the recurrence rate between two consecutive visits. In this case, we only model the follow-up process as exponential function which is proportional between treatment group, and the length of follow-up period is the longest follow-up time of the patients in the pooled sample. The test for sufficient follow-up is not directly relevant to the main problem in our study. However, we do explore the effect of the length of follow-up on the empirical power of the proposed test.

1.4 Novelty and Contributions

It is a need to develop a test statistic that allows for unequal observation process and relax the assumption on the identical treatment indicator. Thus, a new class of statistical test procedure is proposed based on MCF with conditional on treatment groups which allow the different observation scheme and imbalanced sample size across treatment groups. Based on simulation study, the proposed test works well for both equal and unequal observation process and even when the sample size are imbalanced. The test statistics based on conditional MCF estimator is efficiently comparable with existing NPMLE and IRE based tests. The results are also compatible with existing tests. The proposed test statistic is also extended to address multiple sample and multivariate comparison problems.

1.5 Thesis Outline

Chapter 2 gives some fundamental about panel count data and the background knowledge related to nonparameric mean cumulative function (MCF) estimator. This chapter reviews some of the existing nonparametric test statistics which are commonly used in the analysis of panel count data and limitations of those methods are highlighted. Chapter 2 further presents some practical constrains in medical follow up study including multi-sample and multivariate problems.

Chapter 3 extends the nonparametric Wilcoxon type of test statistics which allowed unequal observation processes when the recurrent events followed nonhomogeneous Poisson processes and mixed Poisson processes. The development of the test statistics based on MCF given treatment groups as well as it asymptotic variance calculation are presented in Chapter 3.

Chapter 4 demonstrates the efficiency of the test statistics for two samples comparisons for Poisson and mixed Poisson recurrent processes with identical observation processes. The proposed test is applied to compare treatment efficiency for the data arise from Gallstone study which is a pre-schedule follow-up study. Whilst, Chapter 5 generalizes the test to multi-sample comparisons with unequal observation

processes, where the number of observations between treatments are different. The test statistic is applied to panel count data arise from bladder tumor study, where the observation processes are not identical and comparisons are made between two or more treatment groups. The asymptotic normality of the proposed test statistic are presented through quantile plots.

Chapter 6 extends the approach to multivariate panel count data problem. Findings of the study are outlined in Chapter 7. Conclusion summarizes the scope of study with further discussion and outlines some future works to extend this study are included in Chapter 7. The overall research work is given in Figure 1.1.

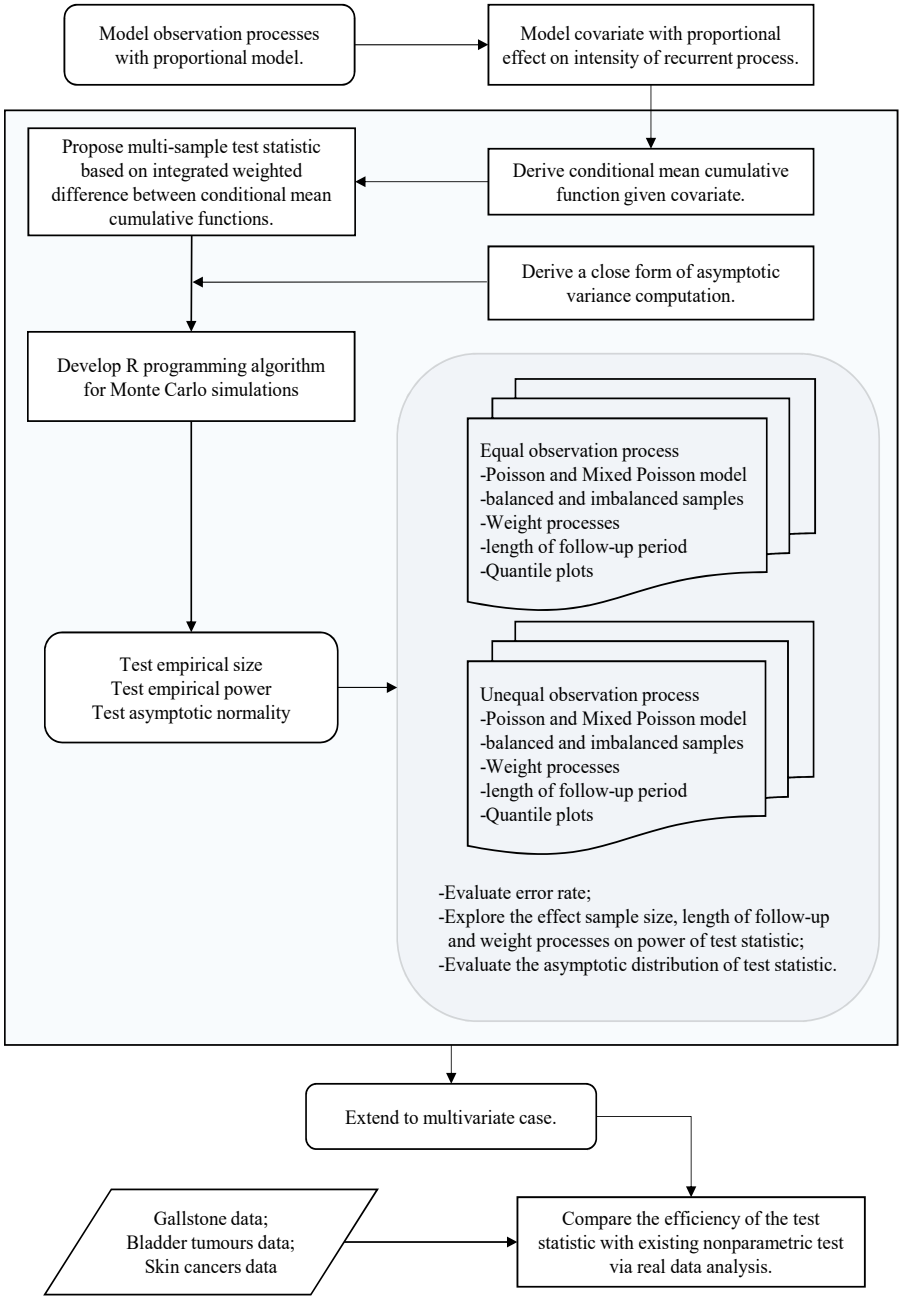

Figure 1.1: Research work flow.

BIBLIOGRAPHY

Abu-Libdeh, H. (1988). *Statistical methodology for the analysis of replicated point processes: with application to a randomized clinical trial for the prevention of skin cancer.* PhD thesis, Cornell University, Ithaca, New York.

Abu-Libdeh, H., Turnbull, B., and Clark, L. (1988). Analysis of multi-type recurrent events in longitudinal studies; application to a skin cancer preventive trial. Technical report, Cornell University, Ithaca, New York.

Abu-Libdeh, H., Turnbull, B., and Clark, L. (1990). Analysis of multi-type recurrent events in longitudinal studies: application to a skin cancer prevention trial. *Biometrics*, 46:1017–1034.

Andersen, P., Borgan, O., Gill, R., and Keiding, N. (1993). *Statistical models based on counting processes.* New York: Springer.

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, 10(4):1100–1120.

Andersen, P. and Ronn, B. (1995). A nonparametric test for comparing two samples where all observations are either left- or right- censored. *Biometrics*, 51:323–329.

Andrews, D. and Herzberg, A. (1985). *Data: A collection of problems from many fields for the student and research worker.* Number 253-259. New York: Springer.

Bagdonavi?ius, V. and Nikulin, M. (2013). Goodness-of-fit test for homogeneous markov processes. *Comptes Rendus Mathematique*, 351(3-4):149–154.

Balakrishnan, N. and Zhao, X. (2009). New multi-sample nonparametric tests for panel count data. *Annals of Statistics*, 37:1112–1149.

Balakrishnan, N. and Zhao, X. (2010a). A class of multi-sample nonparametric test for panel count data. *Annals of the Institute of Statistical Mathematics*, 63:135–156.

Balakrishnan, N. and Zhao, X. (2010b). A nonparametric test for equality of counting processes with panel count data. *Communicational Statistics and Data Analysis*, 4:135–142.

Balshaw, R. and Dean, C. (2002). A semiparametric model for the analysis of recurrent event panel data. *Biometrics*, 58:324–331.

Boucher, J., Denuit, M., and Guillen, M. (2008). Models of insurance claim count with time dependence based on generalisation of poisson and negative binomial distributions. *Variance*, 2(1):135–162.

Boucher, J. and Guillen, M. (2009). A survey on models for panel count data with applications to insurance. *Real Academia de Ciencias, Series A. Mathematics*, 103(2):277–294.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31(2):144–152.

Buzkova, P. (2010). Panel count data regression with informative observation times. *International Journal of Biostatistics*, 6(30).

Byar, D., Blackard, C., and VACURG (1977). Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage i bladder cancer. *Urology*, 10:556–561.

Cai, J. and Prentice, R. (1997). Regression estimation using multivariate failure time data and a common baseline hazard function model. *Lifetime Data Analysis*, 3:197–213. PubMed: 9384652.

Castro, M., Chen, M., Ibrahim, J., and Klein, J. (2014). Bayesian transformation models for multivariate survival data. *Scandinavian Journal of Statistics*, 41:187–199.

Chen, Y., Chen, K., and Ying, Z. (2010). Analysis of multivariate failure time data using marginal proportional hazards model. *Statistica Sinica*, 20(33):1025–1041.

Clayton, D. (1978). A model for association in bivariate life tables and its application to epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141–151.

Clayton, D. (1985). Multivariate generalization of the proportional hazards model (with discussion). *Statistics Society A*, 148:82–117.

Clegg, J., Cai, J., and Sen, P. (1999). A marginal mixed baseline hazards model for multivariate failure time data. *Biometrics*, 55:805–812. PubMed:11315010.

Collett, D. (2003). *Modelling survival data in medical research*. CRC Press.

Cook, R. J. and Lawless, J. F. (1997). Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine*, 16:911–1149.

Cook, R. J. and Lawless, J. F. (2007). *The statistical analysis of recurrent events*. New York: Springer-Verlag.

Cox, D. and Isham, V. (1980). *Point Processes*. New York: Chapman and Hall.

Cox, D. and Oakes, D. (1984). *Analysis of survival data*. Chapman and Hall.

Datta, S., Pardo, M., Scheike, T., and Yuen, K. (2016). Special issue on advances in survival analysis. *Computational Statistics And Data Analysis*, 93:255–256.

Dean, C. and Balshaw, R. (1997). Efficiency lost by analysing counts rather than event times in poisson and overdispersed poisson regression models. *Journal of the American Statistical Association*, 92:1387–1398.

Garcła-Prez, M. A. (2013). Statistical criteria for parallel tests: A comparison of accuracy and power. *Behavior Research Methods*, 45(4):999–1010.

Ghosh, D. and Lin, D. (2000). Nonparametric analysis of recurrent events and death. *Biometrics*, 56:554–562.

Gray, R. J. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics.*, 16(3):1141–1154.

Hart, J., Choi, T., and Yi, S. (2016). Frequentist nonparametric goodness-of-fit tests via marginal likelihood ratios. *Computational Statistics & Data Analysis*, 96:120–132.

He, X., Tong, X., and Sun, J. (2000). Semiparametric analysis of panel count data with correlated observation and follow-up times. *Lifetime Data Analysis*, 15:177–196.

He, X., Tong, X., Sun, J., and Cook, R. J. (2008). Regression analysis of multivariate panel count data. *Biostatistics*, 9:234–248.

Holgate, P. (1964). Estimation for the bivariate poisson distribution. *Biometrika*, 51:241–245.

Hortobagyi, G., Theriault, R., Porter, L., Blayney, D., Lipton, A., Sinoff, C., Wheeler, H., Simeone, J., Seaman, J., Knight, R., Heffernan, M., and Reitsma, D. (1996). Efficacy of pamidronate in reducing skeletal complications in patients with breast cancer and lytic bone metastases. *The New England Journal of Medicine*, 335:1785–1791.

Hosmer, D. and Lemeshow, S. (1999). *Applied survival analysis: Regression modelling of time to event data*. John Wiley and Sons.

Hu, X., Lagakos, S., and Lockhart, R. (2009a). Marginal analysis of panel count through estimating functions. *Biometrika*, 96:445–456.

Hu, X., Lakagos, S., and Lockhart, R. (2009b). Generalized least squares estimation of the mean function of a counting process based on panel counts. *Statistica Sinica*, 19:561–580.

Hu, X., Sun, J., and Wei, L. (2003). Regression parameter estimation from panel counts. *Scandinavian Journal of Statistics*, 30:25–43.

Hua, L. and Zhang, Y. (2012). Spline-based semiparametric projected generalized estimating equation method for panel count data. *Biostatistics*, 13(3):440–454.

Hua, L., Zhang, Y., and Tu, W. (2014). A spline-based semiparametric sieve likelihood method for over-dispersed panel count data. *The Canadian Journal of Statistics.*, 42(2):217–245.

Huang, C. and Wang, M. (2004). Joint modelling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association*, 99(468):1153–1165.

Huang, C., Wang, M., and Zhang, Y. (2006). Analysing panel count data with informative observation times. *Biometrika*, 93(4):763–775.

Johnson, N., Kotz, S., and Balakrishnan, N. (1996). *Discrete multivairate distributions*. Wiley, New York, 2nd edition edition.

Kalbfieisch, J. and Lawless, J. F. (1985a). The analysis of panel count under a markov assumption. *Journal of American Statistics Asssociation*, 80:863–871.

Kalbfieisch, J. and Lawless, J. F. (1985b). The analysis of panel data under markov assumption. *Journal of the American Statistical Association*, 80(392):863–871.

Kalbfieisch, J. and Prentice, R. (2002). *The statistics analysis of failure time data*. John Wiley and Sons.

Klein, J. and Moeschberger, M. (2003). *Survival analysis: Techniques for censored and truncated data. Statistics for Biology and Health*. Springer-Verleg New York, 2nd edition edition.

Kleinbaum, D. and Klein, M. (2012). *Survival analysis: A self-learning text. Statistics for Biology and Health*. Springer-Verlag, New York.

Lawless, J. F. (2003). *Stistical models and methods for lifetime data*. John Wiley and Sons.

Li, N., Park, D., Sun, J., and Kim, K. (2011). Semiparametric transformation models for multivariate panel count data with dependent observation process. *Canadian Journal of Statistics*, 39:458–474.

Li, N., Zhao, H., and Sun, J. (2013). Semiparametric transformation models for panel count data with correlated observation and follow-up times. *Statistics in Medicine*, 32:3039–3054.

Li, Y., Shchy, A., and Sun, J. (2010). Nonparametric treatment comparison for current status data. *Journal of Biometrics and Biostatistics*, 1:102.

Li, Y., Zhao, H., Sun, J., and Kim, K. (2014). Nonparametric tests for panel count data with unequal observation processes. *Computational Statistics and Data Analysis*, 73:103–111.

Lin, D., Feuer, E., Etzioni, R., and Wax, Y. (1997). Estimating medical cost from incomplete follow-up data. *Biometrics*, 53:419–434.

Lin, D., Wei, L., Yang, I., and Ying, Z. (2000). Semiparametric regression for mean and rate functions of recurrent events. *Journal Royal Statistics Society Series B Stat Methodologies*, 62:711–730.

Lu, M., Zhang, Y., and Huang, J. (2009). Estimation of the mean function with panel count data using monotone polynomial b-splines. *Journal of The American Statistical Association*, 104(487):1060–1070.

Mandel, M., Una-Alvarez, J., Simon, D., and Betensky, R. (2018). Inverse probability weighted cox regression for doubly truncated data. *Journal of Biometrics*, 74:481–487.

Matloff, N. (2011). *The art of R programming: A tour of statistical software design*. No Starch Press.

Moreira, C., Una-Alvarez, J., and Meira-Machado, L. (2016). Nonparametric regression with doubly truncated data. *Computational Statistics And Data Analysis*, 93:294–307.

Morris, N. and Elston, R. (2013). A note on comparing the power of test statistics at low significance levels. *Am Statistics*, 65(3):1–5.

Namba, A. (2004). Simulation studies on bootstrap empirical likelihood tests. *Communications in Statistics, Simulation and Computation*, 33:99–108.

Nelson, W. B. (1979). How to analysis data with simple plots, in asqc basic references in quality control. *Statistics Techniques*, 1((800)952-6587).

Nelson, W. B. (1988). Graphical analysis of system repair data. *Journal of Quality Technology*, 20:24–35.

Nelson, W. B. (1995). Confidence limits for recurrences data: Applied to cost or number of product repairs. *Technometrics*, 37:147–157.

Nelson, W. B. (2003). *Recurrent events data analysis for product repairs, disease recurrences, and other applications*. Philadelphia: ASA-SIAM.

Nietert, P. and Dooley, M. (2011). The power of the sign test given uncertainty in the proportion of tied observations. *Contemporary Clinical Trials*, 32(1):147–150.

Park, D., Sun, J., and Zhao, X. (2007). A class of two-sample nonparametric tests for panel count data. *Communications in Statistics Theory and Methods*, 36:1611–1625.

Pepe, M. and Fleming, T. (1989). Weighted kaplan-meier statistics: A class of distance tests for censored survival data. *Biometrics*, 45:497–507.

Prentice, R. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, 79:495–512.

Prentice, R. and Hsu, L. (1997). Regression on hazard ratios and cross ratios in multivariate failure time analysis. *Biometrika*, 84:349–363.

Raykov, T., Patelis, T., and Marcoulides, G. A. (2011). Examining parallelism of sets of psychometric measures using latent variable modeling. *Educational and Psychological Measurement*, 71(6):1047C1064.

Rennert, L. and Xie, S. (2018). Cox regression model with doubly truncated data. *Journal of Biometrics*, 74:725–733.

Robey, R. R. and Barcikowski, R. S. (1992). Type-i error and the number of iterations in monte-carlo studies of robustness. *British Journal of Mathematical and Statistical Psychology*, 43(1):113C130.

Sainani, K. L. (2016). Introduction to survival analysis. *PM&R*, 8(6):580–585.

Shen, P. (2015). Conditional mle for the proportional hazards model with left-truncated and interval-censored data. *Statistics & Probability Letters*, 100:164–171.

Sun, J. (1999). A nonparametric test for current status data with unequal censoring. *Journal Royal Statistics Society Series B Stat Methodologies*, 61:243–250.

Sun, J. (2006). *The statistical analysis of interval-censored failure time data. Statistics for Biology and Health*. Springer Science and Business Media.

Sun, J. and Fang, H. (2003). A nonparametric test for panel count data. *Biometrika*, 90:199–208.

Sun, J. and Kalbfieisch, J. (1993). The analysis of current status data on point processes. *Journal of the American Statistical Association*, 88(424):1449–1454.

Sun, J. and Kalbfieisch, J. (1995). Estimation of the mean function of point processes based on panel count data. *Statistics Sinica*, 5:279–290.

Sun, J., Tong, X., and He, X. (2007). Regression analysis of panel count data with dependent observation times. *Biometrics*, 63:1053–1059.

Sun, J. and Wei, L. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal Royal Statistics Society Series B Stat Methodologies*, 62:293–302.

Sun, J. and Zhao, X. (2013). *The statistical analysis of panel count data.* Springer Science+Business Inc.

Sun, Y., Li, M., and Gilbert, P. B. (2016). Goodness-of-fit test of the stratified mark-specific proportional hazards model with continuous mark. *Computational Statistics and Data Analysis*, 93:348–358.

Synder, D. L. and Miller, M. I. (1991). *Random point processes in time and space*. Springer-Verlag, New York.

Tanizaki, H. (2004). Power comparison of empirical likelihood ratio tests: Small sample properties through monte monte-carlo studies. *Kobe University Economic Review*, 50:13–25.

Taub, M., Schwender, H., Younkin, S., Louis, T., and Ruczinski, I. (2013). On multi-marker tests for association in case-control studies. *PudMed*, 4(252):1–12.

Thall, P. and Lachin, J. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data. *Journal of American Statistics Association*, 83:339–347.

Therneau, T. and Grambsch, P. (2000). *Modeling survival data: extending the Cox model*. New York: Springer-Verlag.

Thompson, W.A., J. (1988). *Point process model with applications to safety and reliability*. Chapman and Hall.

Tian, Y. and Jiang, B. (2016). Equalities for estimators of partial parameters under linear model with restrictions. *Journal of Multivariate Analysis*, 143:299–313.

Wang, L., McMahan, C., Hudgens, M., and Qureshi, Z. (2016). A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics*, 72(1):222–231.

Wei, L., Lin, D., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling of marginal distributions. *Journal of American Statistics Asssociation*, 84:1065–1073.

Wellner, J. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Annals of Statistics*, 28:779–814.

Wellner, J. and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Annals of Statistics*, 35:2106–2142.

Wilcox, R. R. (1990). Comparing the variances of two dependent groups. *Journal of Educational Statistics*, 15(3):237–247.

Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *Annals of Mathematical Statistics*, 17(3):257C281.

Xu, R. and Harrington, D. (2001). A semiparametric estimate of treatment effects with censored data. *Biometrics*, 57:875–885.

Yang, Y. and Ying, Z. (2001). Marginal proportional hazards models for multiple event-time data. *Biometrika*, 88:581–586.

Yao, B., Wang, L., and He, X. (2016). Semiparametric regression analysis of panel count data allowing for within-subject correlation. *Computational Statistics and Data Analysis*, 97:47–59. doi:10.1016/j.csda.2015.11.017.

Yoo, Y., Sun, L., Poirier, J., Paterson, A., and Bull, S. (2017). Multiple linear combination (mlc) regression tests for common variants adapted to linkage disequilibrium structure. *PubMed*, 41(2):108–121.

Yuan, A., Zheng, Y., Huang, P., and Tan, M. (2016). A nonparametric test for the evaluation of group sequential clinical trials with covariate information. *Journal of Multivariate Analysis*, 152:82–99.

Zhang, C. (2009). Nonparametric test for general multivariate multi-sample problem. *Journal of Nonparametric Statistics.*

Zhang, H., Zhao, J., Sun. J., Wang, D., and Kim, K. (2013). Regression analysis of multivariate panel count data with an informative observation process. *Journal of Multivariate Analysis*, 119:71–80.

Zhang, Q., Dai, H., and Fu, B. (2016). A proportional hazards model for time-to-event data with epidemiological bias. *Journal of Multivariate Analysis*, 152:224–236.

Zhang, Y. (2002). Semiparametric pseudolikelihood estimation method for panel count data. *Biometrika*, 89:39–48.

Zhang, Y. (2006). Nonparametric k-sample tests with panel count data. *Biometrika*, 93:777–790.

Zhang, Y. and Jamshidian, M. (2003). The gamma frailty poisson model for nonparametric estimation of panel count data. *Biometrics*, 59:1099–1106.

Zhao, H., Li, Y., and Sun, J. (2013a). Analyzing panel count data with dependent observation process and terminal event. *The Canadian Journal of Statistics.*, 33:61–70.

Zhao, H., Li, Y., and Sun, J. (2013b). Semiparametric analysis of multivariate panel count data with dependent observation process and terminal event. *Journal of Nonparametric Statistics.*, 25:379–394.

Zhao, H. and Sun, J. (2006). Semiparametric and nonparametric analysis of recurrent events with observation gaps. *Computational Statistics and Data Analysis*, 51:1924–1933.

Zhao, H. and Sun, J. (2011). Nonparametric comparison for panel count data with unequal observation processes. *Biometrics*, 67:770–779.

Zhao, H., Virkler, K., and Sun, J. (2014). Nonparametric comparison for multivariate panel count data. *Communications in Statistics Theory and Methods*, 43:644–655.

Zhao, X., Balakrishnan, N., and Sun, J. (2011). Nonparametric inference based on panel count data. *TEST*, 20:1–42.

Zhao, X. and Tong, X. (2011). Semiparametric regression analysis for panel count data with informative observation times. *Computational Statististics and Data Analysis*, 55:291–300.

Zhao, X., Tong, X., and Sun, J. (2013c). Robust estimation for panel count data with informative observation times. *Statistica Sinica*, 5:279–290.