Analysis of Clustered Competing Risks with Application to a Multicentre Clinical Trial

By

Mary Ajibola Familusi

Supervisor: Dr Freedom N. Gumedze

Submitted to

The Department of Statistical Sciences

In fulfilment

of the requirements for the degree of

Masters in Mathematical Statistics

28 July 2016

To my all in all, the great and mighty God.

Abstract

The usefulness of time-to-event (survival) analysis has made it gain a wide applicability in statistically modelling research. The methodological developments of time-to-event analysis that have been widely adopted are: (i) The Kaplan-Meier method, for estimating the survival function; (ii) The log-rank test, for comparing the equality of two or more survival distributions; (iii) The Cox proportional hazards model, for examining the covariate effects on the hazard function; and (iv) The accelerated failure time model, for examining the covariate effects on the survival function. Nonetheless, in time-to-event endpoints assessment, if subjects can fail from multiple mutually-exclusive causes, data are said to have competing risks. For competing risks data, the Fine and Gray proportional hazards model for sub-distributions has gained popularity due to its convenience in directly assessing the effect of covariates on the cumulative incidence function. Furthermore, sometimes competing risks data cannot be considered as independent because of a clustered design; for instance, in registry cohorts or multi-centre clinical trials. The Fine and Gray model has been extended to the analysis of clustered time-to-event data, by including random-centre effects or frailties in the sub-distribution hazard.

This research focuses on the analysis of clustered competing risks with an application to the investigation of the management of pericarditis clinical trial (IMPI) dataset. IMPI is a multi-centre clinical trial that was carried out from 19 centres in 8 African countries with the principal objective of assessing the effectiveness and safety of adjunctive prednisolone and *Mycobacterium indicus pranii* immunotherapy, in reducing the composite outcome of death, constriction or cardiac tamponade, requiring pericardial drainage in patients with probable or definite tuberculous pericarditis. The clinical objective in this thesis is therefore to analyse time to these outcomes. In addition, the risk factors associated with these outcomes were determined, and the effect of the prednisolone and M.*indcus pranii* was examined, while adjusting for these risk factors and considering centres as a random effect.

Using Cox proportional hazards model, it was found that age, weight, New York Heart Association (NYHA) class, hypotension, creatinine, and peripheral oedema show a statistically significant association with the composite outcome. Furthermore, weight, NYHA class, hypotension, creatinine and peripherial oedema show a statistically significant association with

death. In addition, NYHA class and hypotension show a statistically significant association with cardiac tamponade. Lastly, prednisolone, gender, NYHA class, tachycardia, haemoglobin level, peripheral oedema, pulmonary infiltrate and HIV status show a statistically significant association with constriction. A value of 0.1 significance level was used to identify variables as significant in the univariate model using forward stepwise regression method.

The random effect was found to be significant in the incidence of composite outcomes of death, cardiac tamponade and constriction, and in the individual outcome of constriction, but this only slightly changed the estimated effect of the covariates as compared to when the random effect was not considered. Accounting for death as a competing event to the outcomes of cardiac tamponade or constriction, does not affect the effect of the covariates on these outcomes. In addition, in the multivariate models that adjust for other risk factors, there was no significant difference in the primary outcome between patients who received prednisolone, and those who received placebo, or between those who received *M. indicus pranii* immunotherapy, and those who received placebo.

Acknowledgments

Much appreciation to my supervisor, Dr. Freedom Gumedze, for his support and constructive criticism throughout this research work. Special thanks to Professor Jane Hutton who introduced me to the important field of Survival Analysis. Heartfelt gratitude to Professor Neil Geoffery Turok, the founder of the African Institute for Mathematical Sciences (AIMS), for his generosity and excellent vision to empower talented scholars.

This research was supported by the African Institute for Mathematical Sciences (AIMS), University of Cape Town (UCT) and South African Centre for Epidemiological Modelling and Analysis (SACEMA). I am immensely grateful for their financial support.

I also like to acknowledge my parents, siblings and husband, for their unending love and support throughout my academic pursuit. To my little angel, Ifeoluwa Owolabi-Yusuf, who arrived miraculously during this research phase, you have been a source of inspiration and joy. Sincere thanks to my family at all levels, brethren in Christ, friends and colleagues, for their love, prayers, encouragement and advices during each phase of this research and throughout.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter gives a general introduction to this study and the motivation for this work. The objectives and goals of this research are also stated. An overview of the research organisation is documented in order to orient the intended audience about current research.

1.1 Introduction

Competing risks has been an active research area in Survival Analysis. Survival Analysis which is also referred to as time-to-event data analysis, is a term used for describing data that measure the time to a given event of interest. In medical science, an event can be the onset of an illness, recovery from an illness, death due to a specific cause etc. The two main objectives of Survival Analysis are to estimate the time-to-event of interest and to assess the relationship between explanatory variables and time-to-event. The first objective is often analysed using Kaplan Meier estimate, which is a non-parametric method of estimating survival function. In addition, the log-rank test can be used to examine the differences between the survival distribution of two groups. The second objective is often analysed by the use of Cox's proportional hazards model (Rao and Schoenfeld, 2007). Cox's proportional hazards model is a semi-parametric method of analysis which does not require the specification of any distribution. Alternatively, it is the parametric proportional hazard model, which assumes that the event time follows a pre-specified distribution. The parametric forms of the Cox proportional hazards model are available only for a few regression models such as exponential and Gompertz and Weibull. In a situation where the proportionality assumption is not satisfied, accelerated failure time model is the most appropriate model to consider. It provides estimates in terms of survival instead of hazards of the outcome. The most commonly used distributions are log-normal,

log-logistics, exponential and Weibull distributions where the choice of these distributions depends on the highest value of likelihood or lowest value of Akaike's Information Criterion (AIC).

The standards Survival Analysis focuses on the analysis of the time to an event of interest in the presence of censoring. A subject is said to be censored if it does not experience the event of interest at the time of study. There are three specific types of censoring: *Right censoring*: It refers to when patients do not experience the event of interest until the end of the study, drop-out of the study or loss to follow-up. *Left censoring*: Patients are said to be left censored if they have already experienced the event of interest before the study begins. *Interval censoring*: It refers to when an event occurs at an interval between two points.

In addition, in time-to-event data analysis, there is a possibility of the occurrence of an event other than the event of interest and this event is referred to as competing risks. Competing risk is an event which either preludes the event of interest or prevents the event of interest from occurring, or changes the probability of its occurrence (Pintilie, 2006). A sub-hazard model, also known as a Fine Gray model (Fine and Gray, 1999), and a cause-specific hazard model (Prentice et al., 1978) are the two broad models for analysing competing risks data. These two models were developed from the Cox proportional hazard model (Prentice et al., 1978). A sub-hazard model is the model that directly associates a covariate effect with the cumulative incidence function, whilst a cause-specific hazard model associates the covariates effect with the cause-specific hazard function (Do Ha et al., 2014)

In many applications involving competing risks, individuals may be correlated within clusters because of the unobserved shared factors across individuals. This is referred to as 'clustered competing risks,' with 'clustering' referring to the potential dependence across individuals within clusters, (Fine and Gray, 1999). For example, in family studies of hereditary cancer, parents and children may share genetic and environmental factors. These can lead to familial correlations in the disease onset. Also, in multi-centre studies, the patient population and referral pattern in each centre may result in correlated outcomes within centres. The Fine-Gray model, however, does not take such correlation into account, which can be modelled by the frailty (or random effect). Thus, Katsahian et al. (2006) have extended the Fine Gray model to a sub-distribution hazard frailty model with a random centre effect.

1.2 Motivation

The work in this thesis is motivated by a recent multicentre clinical trial of investigation of the management of pericarditis (IMPI). The main purpose of this trial was to assess the efficacy and safety of adjunctive prednisolone and *M. indicus pranii* immunotherapy in reducing the composite outcome of death, constriction or cardiac tamponade requiring pericardial drainage in patients with probable or definite tuberculous pericarditis. Therefore, in the case where a composite outcome is chosen as the primary outcome, it is strongly recommended that the individual components be analysed separately by means of the corresponding cause-specific hazards and sub-distribution hazards, in order to check whether the observed effect for the composite is clinically meaningful (Chi, 2005). Furthermore, in a multicentre clinical trial, a clustering problem is inevitable. Indeed, usually not all the important risk factors can be measured and it is reasonable to assume that at least some of them are shared by patients treated in the same centre. Therefore, when data arise from multi-centre clinical trials, a correlation induced by clustering should be taken into account in the analysis, (Katsahian et al., 2006).

1.3 Aim and objectives

Due to the importance of recognising the effect of frailty in clustered survival data and in clustered competing risks settings, the main aim of this thesis is to account for centre or cluster effects in a competing risk analysis. The theoretical objectives of this thesis are as follows:

1. Reviewing the usefulness of various standard survival models and competing risks models on clustered data.

2. Using appropriate models to explore the effect of different explanatory variables on the time to an event.

3. Assessing the impact of clustering on the estimate of the explanatory variables.

4. Comparing the use of models with the cluster effect to the use of models without the cluster effect.

The clinical objectives of this thesis are to:

1. Obtain risk factors that are associated with the individual and composite outcome of death, constriction or cardiac tamponade requiring pericardial drainage in patients with probable or definite Tuberculosis Pericarditis.

2. Examine the effect of the prednisolone and *M. indicus pranii* treatments while adjusting for the risk factors and considering centres as cluster effect.

1.4 Structure of the thesis

This thesis report is organized as follows: Chapter 1 presents the general introduction, motivation and objectives for this thesis. In Chapter 2, the research problem is highlighted. Furthermore, the data source and features are discussed. Chapter 3 discusses relevant functions and models in the standard event time analysis with one possible outcome and their application to the IMPI data. In Chapter 4, parametric models in the standard event time analysis and application to IMPI data are presented. In Chapter 5, the concept of competing risks settings are described, relevant functions and models used for competing risks data are documented. The application of the described functions and models on IMPI data followed. Clustered competing risks concepts are discussed in Chapter 6, and the important models used in a clustered competing risks setting explained. Also presented in this chapter is the application of the IMPI data. Chapter 7 re-establishes the focus and objective of the thesis, presents the concluding facts discovered in this thesis, and future work is also presented.

1.5 Chapter summary

The discussion in this chapter offers a general overview of the thesis. Combining several endpoints of interest into a composite increases the effect size and thereby reduces the required sample size, and the dataset from multiple centres enhance the generality of the result. There is a need to analyse composite outcomes individually and take into account any correlation induced by clustering in a multi-centre clinical trial. Hence this research focuses on the analysis of clustered competing risks with an application to the investigation of the management of a pericarditis clinical trial (IMPI) dataset. The general outline of the dissertation is also explicitly stated.

Chapter 2

Background to research problem and data introduction

In spite of anti-tuberculosis chemotherapy, tuberculous (TB) pericarditis is one of the main causes of death or disability in nearly half of those affected. Attenuation of the inflammatory response in TB pericarditis may improve outcome by reducing cardiac tamponade and pericardial constriction, but there is uncertainty as to whether adjunctive immunomodulation with corticosteroids and *Mycobacterium indicus pranii* can safely reduce mortality and morbidity. This chapter provides background to the research problem, presents important features of the IMPI clinical trial that was used to assess the effectiveness and safety of prednisolone and *M. indicus pranii* in reducing the composite outcome of death, constriction, or cardiac tamponade requiring pericardial drainage in patients with probable or definite tuberculous pericarditis. The preliminary analysis on IMPI data is also performed in order to examine the difference in the mean and proportion of the baseline characteristics observed in the trial.

2.1 Background to research problem

Tuberculosis (TB) is a major global health challenge. It causes illness among millions of people each year and has almost the same rank as the human immunodeficiency virus (HIV), which is a dominant cause of death in the world. In 2014, the World Health Organization disclosed that the approximate number of new TB cases were 9.6 million of which 3.2 million were found among women, 5.4 million among men and 1.0 million among children. They also reported that there were 1.5 million TB-related deaths (1.1 million in HIV-negative people and 0.4 million in HIV-positive people), of which about 890 000 were men, 480 000 were women and 140 000

were children (WHO, 2015).

A rare manifestation of tuberculosis disease is tuberculous pericarditis, which infects the pericardial membrane i.e. the pericardium that covers the heart. Tuberculous pericarditis which is caused by mycobacterium tuberculosis, is found in about 1% of deaths that are TB-related and in 1% to 2% of cases of pulmonary TB (Mayosi et al., 2005). It is a common cause of pericardial effusion, cardiac tamponade, and constrictive pericardial in developing countries. Pericardial effusion is characterised by an abnormal accumulation of fluid in the pericardial cavity (Mayosi et al., 2005). Cardiac tamponade is characterized by the pressure on the heart that occurs when blood or fluid builds up in the space between the heart muscle and the outer covering sac of the heart, and constrictive pericarditis is chronic inflammation of the pericardium. Tuberculous pericarditis is found to be associated with HIV infection (Mayosi et al., 2006). For instance, in sub-Saharan Africa, at least half the patients with large pericardial effusions are infected with HIV (Ntsekhe et al., 2008). In the Western Cape Province of South Africa, TB pericarditis accounted for 70% of cases with large pericarditis effusion (Reuter et al., 2005).

Therapy for TB pericaditis consists of medical treatment with rifampicin, isoniazid, pyrazinamide, and ethambutol for 6months, pericardial drainage for cardiac tamponade, and pericardiectomy for pericardial constriction (Mayosi et al., 2002). In spite the use of anti-tuberculosis chemotherapy, death caused by TB pericarditis remains high (Mayosi, 2007) but, it is even higher among people with acquired immunodeficiency syndrome (Mayosi et al., 2008). Additionally, it was found that the use of glucocorticoid therapy in patients with tuberculous pericarditis to reduce the inflammatory response, may improve outcomes and decrease the risk of death by reducing cardiac tamponade and pericardial constriction, but the clinical potency of adjunctive glucocorticoids is unclear (Mayosi et al., 2014). Meanwhile, Ntsekhe et al. (2003) indicated that there was no assurance as to whether adjunctive glucocorticoid could safely reduce mortality and morbidity. A meta-analysis of randomized, controlled trials of glucocorticoid therapy for tuberculous pericarditis showed a nonsignificant reduction in mortality, (Mayosi et al. (2002), Ntsekhe et al. (2003)), but the numbers of events and patients included were very small. Mayosi et al. (2014) hypothesized that there would be a large benefit of adjunctive prednisolone for TB pericarditis patients and that intradermal *Mycobacterium indicus pranii* could be effective in subduing inflammation and its sequelae in patients with tuberculous pericarditis. *M. indicus prani*i is a nonpathogenic, saprophytic, rapidly growing atypical mycobacterium species that has shown clinical benefit when administered as a heat-killed intradermal formula-

tion in patients with leprosy, and it may have benefits in patients with pulmonary tuberculosis and HIV infection. There is preliminary evidence indicating that repeated doses of intradermal heart-killed *M. indicus pranii* immunotherapy may lessen the inflammation associated with tuberculosis and increase the CD4+ T-cell count in people infected with HIV (Mathur, 2006).

2.2 Data source

The investigation of the management of pericarditis (IMPI -pronounced "ee-mp-ee" for warrior in Zulu) is a multicentre international double-blind placebo-controlled randomized trial with 2 x 2 factorial design. It was carried out with the principal objective of assessing the effectiveness and safety of adjunctive prednisolone and *M. indicus pranii* immunotherapy, in reducing the composite outcome of death, constriction or cardiac tamponade requiring pericardial drainage in patients with probable or definite tuberculous pericarditis . The trial was carried out between January 2009 and February 2014 in 1400 patients from 19 centres in 8 African countries. The number of patients per centre varied from 1 to 350 patients. Patients were grouped into experimental and control groups using central concealed randomization. The experimental group received oral prednisolone for 6 weeks and an *M. indicus pranii* injection for 3 months while the control group received a placebo for the same duration. The other two groups were those that received the combination of the two therapies and those that received the combination of the two placebos. The primary outcome is the first occurrence of death, pericardial constriction, or cardiac tamponade requiring pericardiocentesis and, the secondary outcome is the safety of immunomodulatory treatment measured by the effect on opportunistic infections (Mayosi et al., 2014). The IMPI trial was the first largest international multi-centre clinical trial of patients with TB pericarditis that was adequately powered to determine the effects of adjunctive immunotherapy on major effectiveness and safety outcomes. The trial design is described in detail in (Mayosi et al., 2013). The results from the analysis of the data using the standard Cox proportional hazard model on the effect of the two treatment, shows that the use of prednisolone for 6 weeks and the use of *M. indicus pranii* for three months had no significant effect on the combined outcome of death from all causes, cardiac tamponade requiring pericardiocentesis or constrictive pericarditis. Prednisolone tablets and *M. indicus pranii* injections were associated with an increased risk of HIV-associated malignancy. The use of prednisolone reduced the incidence of constrictive pericarditis and hospitalization. The beneficial effects of prednisolone on constriction and hospitalization were similar in HIV-positive and HIV-negative

patients.

2.2.1 Exploratory data analysis

A total of 1400 patients were enrolled for the comparison of prednisolone with placebo; 706 were assigned to receive prednisolone and 694 to receive placebo. The median follow-up period was 636.5 days (interquartile range, 317.5 to 1085.5). Also, a total of 1250 patients were enrolled for the comparison of *M. indicus pranii* with placebo; 625 were assigned to receive *M. indicus pranii* and 625 to receive placebo. The median follow-up period was 720.5 days (interquartile range, 368.0 to 1095.0). The baseline characteristics of the treatment groups were similar across the treatment groups (see Table 2.1 - Table 2.2, p-value >0.05). Out of 1400 patients, 336 patients experienced the composite event, (i.e. death, cardiac tamponade or constriction), 246 patients experienced death, 50 patients experienced cardiac tamponade and 85 patients experienced constriction. Furthermore, out of 50 patients who had cardiac tamponade, 36% experienced death and out of 85 patients that have constriction, 27% experience death. Out of 246 patients that died, 92.7% did not have cardiac tamponade and 90.7% did not have constriction. Approximately two thirds of the patients had a large pericardial effusion, about 32.9% had pulmonary infiltrate, and two thirds were HIV-positive. 72.4% of the total patients were from South African centres.

Table 2.1: Summary statistics for the baseline characteristics of patients randomised to either prednisolone group or placebo group.

Characteristic	Overall (N = 1400)	Prednisolone (N = 706)	Placebo (N = 694)	p-value
Age (years), Median (IQR)	35.56 (17.66)	35.90 (17.63)	35.38 (17.70)	0.763
Duration of symptoms (days), Median (IQR)	30.00 (28.00)	30.00 (46.00)	30.00 (22.00)	0.396
Weight (kg), Median (IQR)	58.00 (16.00)	57.35 (16.00)	58.00 (15.70)	0.705
	N (%)	N (%)	N (%)	
Gender				0.493
Male	784 (56.00)	389 (55.10)	395 (56.92)	
Female	616 (44.00)	317 (44.90)	299 (43.08)	
Country				0.779
South Africa	1014 (72.43)	509 (72.10)	505 (72.77)	
other	386 (27.57)	197 (27.90)	189 (27.23)	
NYHA Class				0.566
I	256 (18.32)	137 (19.43)	119 (17.20)	
II	694 (49.68)	342 (48.51)	352 (50.87)	
III	330 (23.62)	163 (23.12)	167 (24.13)	
IV	117 (8.38)	63 (8.94)	54 (7.80)	
Hypotension (SBP)				0.259
$\leq$ 90 mmHg	106 (7.58)	59 (8.37)	47 (6.77)	
$>$ 90 mmHg	1293 (92.42)	646 (91.63)	647 (93.23)	
White blood count				0.098
$\leq$ 10 /mm3	1297 (92.91)	662 (94.03)	635 (91.76)	
$>$ 10 /mm3	99 (7.09)	42 (5.97)	57 (8.24)	
Tachycardia (HR)				0.398
$\leq$ 100	608 (43.49)	314 (44.60)	294 (42.36)	
$>$ 100	790 (56.51)	390 (55.40)	400 (57.64)	
Heamoglobin				0.454
$\geq$ 10 g/dl	645 (46.30)	319 (45.31)	326 (47.31)	
$<$ 10 g/dl	748 (53.70)	385 (54.69)	363 (52.69)	
Creatinine				0.865
$\leq$ 105 umol/l	1112 (87.42)	564 (87.58)	548 (87.26)	
$>$ 105 umol/l	160 (12.58)	80 (12.58)	80 (12.74)	
Palpable paradoxus				0.512
No	1130 (80.71)	565 (80.03)	565 (81.41)	
Yes	270 (19.29)	141 (19.97)	129 (18.59)	
Peripheral oedema				0.917
No	825 (58.93)	417 (59.07)	408 (58.79)	
Yes	575 (41.07)	289 (40.93)	286 (41.21)	
Pulmonary infiltrates on CXR				0.427
No	856 (67.08)	422 (66.04)	434 (68.13)	
Yes	420 (32.92)	217 (33.96)	203 (31.87)	
Atrial fibrillation on ECG				0.249
No	1001 (94.61)	501 (93.82)	500 (95.42)	
Yes	57 (5.39)	33 (6.18)	24 (4.58)	
Effusion size				0.672
small	106 (7.80)	50 (7.31)	56 (8.30)	
medium	331 (24.36)	172 (25.15)	159 (23.56)	
large	922 (67.84)	462 (67.54)	460 (68.15)	

Table 2.1 continued ...

Characteristic	Overall (N = 1400) N (%)	Prednisolone (N = 706) N (%)	Placebo (N = 694) N (%)	p-value
HIV status				0.916
Positive	939 (67.07)	474 (67.14)	465 (67.00)	
Negative	431 (30.79)	218 (30.88)	213 (30.69)	
Unknown	30 (2.14)	14 (1.98)	16 (2.31)	
Tamponade at presentation				0.912
No	423 (43.21)	213 (43.38)	210 (43.03)	
Yes	556 (56.79)	278 (56.62)	278 (56.97)	
Constriction at presentation				0.481
No	559 (55.90)	279 (54.81)	280 (57.03)	
Yes	441 (44.1)	230 (45.10)	211 (42.79)	

* SD: Standard Deviation; NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure; CXR: Chest Xray.

* Mann-Whitney test was used to test for a significant difference in the means of the continuous variables and chi-square was used to test for a significant difference in the proportion of discrete variables.

Table 2.2: Summary statistics for the baseline characteristics of patients randomized to either *M. indicus pranii* group or placebo group.

Characteristic	Overall (N = 1250)	*M. indicus pranii* (N = 625)	Placebo (N = 625)	p-value
Age (years), Median (IQR)	35.56 (17.60)	34.94 (16.94)	35.72 (19.14)	0.171
Duration of symptoms (days), Median (IQR)	30.00 (28.00)	30.00 (41.00)	30.00 (27.00)	0.243
Weight (kg), Median (IQR)	58.00 (16.00)	57.00 (15.40)	58.00 (15.00)	0.160
	N (%)	N (%)	N (%)	
Gender				0.099
Male	695 (55.60)	333 (53.28)	362 (57.92)	
Female	555 (44.40)	292 (46.72)	263 (42.08)	
Country				0.802
South Africa	896 (71.68)	446 (71.36)	450 (72.00)	
other	354 (28.32)	179 (28.64)	175 (28.00)	
NYHA Class				0.890
I	234 (18.77)	121 (19.39)	113 (18.14)	
II	617 (49.48)	309 (49.52)	308 (49.44)	
III	290 (23.26)	144 (23.08)	146 (23.43)	
IV	106 (8.50)	50 (8.01)	56 (8.99)	
Hypotension (SBP)				0.088
$\leq$ 90 mmHg	94 (7.53)	55 (8.80)	39 (6.25)	
$>$ 90 mmHg	1155 (92.47)	570 (91.20)	585 (93.75)	
White blood count				0.975
$\leq$ 10 /mm3	1156 (92.78)	576 (92.75)	580 (92.80)	
$>$ 10 /mm3	90 (7.22)	45 (7.25)	45 (7.25)	
Tachycardia (HR)				0.397
$\leq$ 100	562 (45.03)	274 (43.84)	288 (46.23)	
$>$ 100	686 (54.97)	351 (56.16)	335 (53.77)	
Heamoglobin				0.209
$\geq$ 10 g/dl	575 (46.22)	276 (44.44)	299 (47.99)	
$<$ 10 g/dl	669 (53.78)	345 (55.56)	324 (52.01)	

Table 2.2 continued ...

Characteristic	Overall (N = 1250) N (%)	*M. indicus pranii* (N = 625) N (%)	Placebo (N = 625) N (%)	p-value
Creatinine				0.526
≤ 105 umol/l	981 (87.20)	491 (87.84)	490 (86.57)	
> 105 umol/l	144 (12.80)	68 (12.16)	76 (13.43)	
Palpable paradoxus				0.478
No	1002 (80.16)	506 (80.96)	496 (79.36)	
Yes	248 (19.84)	119 (19.04)	129 (20.64)	
Peripheral oedema				0.909
No	730 (58.40)	366 (58.56)	364 (58.24)	
Yes	520 (41.60)	259 (41.44)	261 (41.76)	
Pulmonary infiltrates on CXR				0.218
No	753 (66.05)	366 (64.32)	387 (67.78)	
Yes	387 (33.95)	203 (35.68)	184 (32.22)	
Atrial fibrillation on ECG				0.777
No	891 (94.69)	446 (94.89)	445 (94.48)	
Yes	50 (5.31)	24 (5.11)	26 (5.52)	
Effusion size				0.071
small	97 (8.02)	57 (9.47)	40 (6.58)	
medium	294 (24.30)	154 (25.58)	140 (23.03)	
large	819 (67.69)	391 (64.95)	428 (70.39)	
HIV status				0.112
Positive	840 (67.20)	437 (69.92)	403 (64.48)	
Negative	384 (30.72)	175 (28.00)	209 (33.44)	
Unknown	26 (2.08)	13 (2.08)	13 (2.08)	
Tamponade at presentation				0.102
No	386 (43.67)	201 (46.44)	184 (40.98)	
Yes	498 (56.33)	233 (53.56)	265 (59.02)	
Constriction at presentation				0.349
No	503 (55.70)	242 (54.14)	261 (57.24)	
Yes	400 (44.30)	205 (45.86)	195 (42.76)	

* SD: Standard Deviation; NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure; CXR: Chest Xray.

* Mann-Whitney test was used to test for a significant difference in the means of the continuous variables and chi-square was used to test for a significant difference in the proportion of discrete variables.

2.3 Chapter summary

This chapter discussed tuberculous pericarditis being a rare case of tuberculosis, its association with HIV infection, and some of therapy available for treating tuberculous pericarditis and its shortcomings. We also talk about the proposal made by Mayosi et al. (2014) that there would be a large benefit of adjunctive prednisolone for TB pericarditis patients, and that intradermal Mycobacterium indicus pranii could be effective in subduing inflammation and its sequelae in patients with tuberculous pericarditis. Presented also in this chapter are the baseline characteristics of the IMPI clinical trial used to examine the effect of these treatments on the probable and definite tuberculous pericarditis, and a preliminary analysis on the dataset was performed. Using the Mann-Whitney test and the Chi-square test of difference in means and proportion of the patients' baseline characteristics with respect to the treatment groups, we found that there were no statistical significant differences between the treatments in the

baseline characteristics of the patients.

Chapter 3

Standard statistical methods for the analysis of time-to-event data

In this chapter, we review basic functions and the Cox proportional hazards model for the standard survival analysis. Also, we review the concept of frailty in the survival analysis. We then apply the Cox model to IMPI trial data in order to explore the effect of different explanatory variables present in the data. Furthermore, we fit a frailty model to the dataset so as to examine if there is a significant random effect in the dataset and then compare the model to the standard Cox model. Finally, we perform model checking which includes the verification of the proportionality assumption.

3.1 Important functions in time-to-event analysis

There are basically four functions of interest in the time-to-event data analyses. These functions are the probability density function, survival function, hazard function, and the cumulative hazard function.

The random variable for the event time T is positive and its distribution is defined by the *probability density* function $f(t)$ or the cumulative density function $F(t)$ with

$$\begin{aligned} f(t) &= \Pr(T = t) \\ &= \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t}, \end{aligned} \tag{3.1.1}$$

where $F(t) = \Pr(T < t)$ is the probability that the event of interest has occurred before time t.

The *survival* function is defined as the probability of surviving from the beginning of a study to time t and beyond without experiencing the event of interest. It is given as

$$\begin{aligned} S(t) &= \Pr(T \geq t) \\ &= 1 - \int_0^t f(u)d(u) \\ &= 1 - F(t). \end{aligned} \tag{3.1.2}$$

The survival function can be estimated by the popular Kaplan-Meier method (Kaplan and Meier, 1958). The Kaplan-Meier estimator is the non-parametric estimator of the survival function in the presence of right censored data. It is a non-parametric estimator because it does not make assumption with respect to the structure of the distribution. An estimate for the survival function at a given time t can be obtained as:

$$\hat{S}(t) = \prod_{t_i<t} \Big(1 - \frac{d(t_i)}{n(t_i)}\Big), \tag{3.1.3}$$

where $S(0) = 1$ by definition, $d(t_i)$ is the number in the study which fail at t_i, $n(t_i)$ is the number in the study at t_i and at risk that is able to fail at t_i.

The *hazard* function is defined as the conditional probability of an event in a small time interval from t to $t + \Delta t$, given that the idividual is still at risk at the beginning of that interval,

$$\lambda(t) = \lim_{\Delta t \to 0} \Big\{ \frac{\Pr(t \leq T < t + \Delta t | T \geq t}{\Delta t} \Big\}. \tag{3.1.4}$$

The *cumulative hazard* function can be estimated using the Nelson-Aalen estimator (Nelson, 1969), by summing up the quotients of the observed events and the numbers of subjects at risk, that is

$$\hat{\Lambda}(t) = \sum_{t_i<t} \frac{d(t_i)}{n(t_i)}.$$

The probability density function $f(t)$, the survival function $S(t)$ and the hazard function $\lambda(t)$ are directly related as follows;

$$\lambda(t) = \frac{f(t)}{S(t)}$$

and

$$S(t) = 1 - F(t), \log(S(t)) = \log(1 - F(t))$$

by using rules for taking derivatives of logs,

$$\frac{d}{dt}\log(1-F(t)) = \frac{\frac{d}{dt}(1-F(t))}{1-F(t)} = -\frac{f(t)}{S(t)},$$

so we have

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}\log S(t). \tag{3.1.5}$$

3.2 The Cox proportional hazards model

The most popular regression model for event time data is the Cox proportional hazards model introduced by (David Cox, 1972). It is also known as a semi-parametric model because it makes no assumption about the baseline hazard function, $h_0(t)$, but assumes parametric form for the effect of the covariates on the hazard. The Cox proportional hazards model relies on the assumption that the hazard, given explanatory variables, is proportional to a given baseline hazard function and that each of the p covariates under consideration has a linear effect on the logarithm of the hazard rate, given the other covariates. The Cox regression model can be written as

$$h_i(t) = h_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x_i}), \tag{3.2.1}$$

where $h_0(t)$ is the baseline hazard function, $\boldsymbol{\beta}'\boldsymbol{x}_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}$, $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_p)$ are unknown regression coefficients of the p covariates $X_1, \ldots, X_p$. x_{ji} is the value of the jth covariate for the ith individual, $i = 1, \ldots, n, j = 1, \ldots, p$. The exponential coefficients e^{β_j} are known as hazard ratios.

The ratio between the hazard of two subjects x_i and x_j is constant over time. It is be computed as:

$$\frac{h(t|\boldsymbol{x}_i)}{h(t|\boldsymbol{x}_j)} = \frac{h_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}{h_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}_j)} = \exp(\boldsymbol{\beta}'(\boldsymbol{x}_i - \boldsymbol{x}_j)). \tag{3.2.2}$$

3.2.1 Parameter estimation in the Cox proportional hazards model

The regression parameters $\boldsymbol{\beta}$ in equation (3.2.1) can be estimated using the method of maximum likelihood. Let $t_1, t_2, \ldots, t_n$ be the observed survival time for n individuals, if there are r failures

at the time points $t_1 < t_2 < \cdots < t_{r-1} < t_r$ and R_j is the risk set at time t_j then the partial likelihood to be maximized is of the form

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{r} \Big\{ \frac{\exp(\beta x_j)}{\sum_{i \in R_j} \exp(\beta x_i)} \Big\}^{\delta_i}, \tag{3.2.3}$$

where $R_j = \{i : t_i \geq t_j\}$ is the risk set, i.e. the set of subject at risk, at time j and δ_i is an indicator, which is zero if the ith survival time t_i, $i = 1, 2, \ldots, n$, is right-censored, and unity otherwise.

The maximum likelihood estimates of the β-parameters in the proportional hazards model can be found by maximising the corresponding log-likelihood function of equation 3.2.3 using numerical methods. This maximisation is generally accomplished using the Newton-Raphson procedure. Most of the major statistical packages such as STATA, R, and SAS have facilities which enable the proportional hazards model to be fitted.

3.3 Proportional hazard assumption checking

The main assumption of the Cox proportional hazards model is proportional hazards. Proportional hazards implies that the hazard function of one person is proportional to the hazard function of another person, in other words, the hazard ratio is constant over time. There are a number of different ways of testing whether a model satisfies the assumption of proportionality such as adding time-dependent covariates to the Cox model, the graphical method, and tests based on the Schoenfeld residuals. In this thesis, we shall make use of the Schoenfeld residuals to examine the proportional hazards assumption which is discussed in detail in Section 3.4.4. However, if the PH assumption holds for a particular covariate then the Schoenfeld residual for that covariate will not be related to survival time. So this test is accomplished by finding the correlation between the Schoenfeld residuals for a particular covariate and the ranking of individual survival times. The null hypothesis is that the correlation between the Schoenfeld residuals and the ranked survival time is zero. Rejection of the null hypothesis concludes that the PH assumption is violated.

3.4 Model checking

In order to assess the adequacy of a fitted model, there are aspects of a model that require checking. There is a need to determine whether the model includes an appropriate set of explanatory variables, to see if there are any outlying or influential subjects, and to check if the proportionality assumption has been satisfied. The model-checking procedures below are based on residuals. In linear regression methods, residuals are determined as the difference between an observed data point and predicted or fitted values of the dependent variable. Various residuals that have been defined for the Cox regression model are the Cox-Snell residual, deviance residual, martingale residual and the Schoenfeld residual.

3.4.1 Cox-Snell residuals

Cox-Snell residuals are a useful means of checking how well the model represents the data. A survival model is adequate if it represents the survival patterns in the data to an acceptable degree. This aspect of model checking is known as goodness of fit. The Cox-Snell residual for i^{th} individual, $i = 1, 2, \ldots, n$, is given by

$$rc_i = \exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_i)\hat{H}_0(t_i) \tag{3.4.1}$$

where $\hat{H}_0(t_i)$ is the estimated integrated baseline hazard function or cumulative hazard. This is also known as Nelson-Aalen estimator. If the model fits the data well then the true cumulative hazard function condition on the covariate vector has an exponential distribution with hazard rate of one. If the hazard function follows the 45 degree line then we know that it approximately has an exponential distribution with a hazard rate of one and that the model fits the data well.

3.4.2 Martingale residuals

Martingale residuals are useful in determining the functional form of the covariates to be included in the model. If the test reveals that a covariate can not be included in a model linearly, then there is a need for such a covariate to be transformed. Assuming X follows the Cox model with covariate z and the regression coefficient β is known, the Martingale residuals is given as

$$M_j = N_j(\infty) - \int_0^\infty Y_j(s)e^{z_j(s)\beta}dH_0(s), \qquad j = 1, 2 \ldots n.$$

In practice, to compute the martingale residuals, β and H_0 will be replaced with $\hat{\beta}$ and $\hat{H}_0$. Then

$$\hat{M}_j = N_j(\infty) - \int_0^\infty Y_j(s) e^{z_j(s)\beta} d\hat{H}_0(s) = \delta_j - r_j,$$

where r_j is the Cox-Snell residual.

3.4.3 Deviance residuals

Deviance residuals are useful for examining model accuracy and identifying outliers. The skew range of Martingale residual makes it difficult to use it to detect outliers. Martingale residuals have a range between -∞ and 1. For this reason the deviance residuals, which are the Martingale residuals rescaled so that they are entered around zero, are better to use (Klein and Moeschberger, 2003). The deviance residual is defined as:

$$D_j = sign(\hat{M}_j)\{-2[M_j + \delta_j \log(\delta_j - \hat{M}_j)]\}^{\frac{1}{2}},$$

where $\hat{M}_j$ is martingale residual.

3.4.4 Schoenfeld residuals

Schoenfeld residuals are useful for checking and testing the proportional hazard assumption, examining leverage points, and identifying outliers. It can be used to descibe each explanatory variable in the fitted Cox regression model. The Schoenfeld residuals were originally called partial residuals because the Schoenfeld residuals for the i^{th} individual on the j^{th} explanatory variable X_j is an estimate of the i^{th} component of the first derivative of the logarithm of the partial likelihood function with respect to β_j. From equation (3.2.3), this logarithm of the partial likelihood function is given by

$$\frac{\partial \log L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{p} \delta_i \{x_{ij} - a_{ij}\},$$

where x_{ij} is the value of the j^{th} explanatory variable $j = 1, 2, \ldots, p$ for the i^th individual and

$$a_{ji} = \frac{\sum_{l \in R(t_i)} x_{jl} \exp(\boldsymbol{\beta}' x_l)}{\sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}' x_l)}$$

The Schoenfeld residual for i^{th} individual of X_j is given by $r_{pji} = \delta_i\{x_{ji} - a_{ji}\}$.

3.5 Application

In this section, the presented methods of the standard survival analysis were applied to the IMPI dataset with the goal of identifying various risk factors associated with the composite outcome of death, cardiac tamponade and constriction and the individual outcomes. We only present the results of the prednisolone treatment with other risk factors in this section and other application sections. The results for the *M. indicus pranii* treatment can be found in the appendix. This is done so as to reduce the volume of the main thesis that will be generated through tables and plots from the analysis. However, the interpretations are similar to the ones presented in the main thesis. The only difference is the replacement of the prednisolone treatment and its placebo group with the *M.indicus pranii* treatment and its placebo group. Also, the effect of the prednisolone and *M. indicus pranii* treatments were estimated separately because there was no significant interaction between them (Mayosi et al., 2014). R and STATA are the software used for analysis in this section and in other application sections of this thesis.

3.5.1 Non-parametric analysis

Prior to formal modelling, Kaplan-Meier plots were used to visually illustrate the survival distributions in different treatment groups in the study. In Figure 3.1, we observed that the probability estimate of experiencing the composite outcome in the predinisolone and placebo groups by 100 days is about 13% and 17% respectively, while the estimate for both groups to experience death by 100 days is about 9%. Furthermore, the probability of having cardiac tamponade at day 100 in the predinisolone group is about 3%, while 4% is the estimate in the placebo group. Also, there is about 4% and 9% probability estimate of having constriction by 100 days for the predinisolone and placebo groups respectively. The ideal way of reporting the Kaplan-Meier estimate is to present the median survival time, but from the plots in Figure 3.1, the survival curve does not drop to 0.5, hence the median survival time is not calculated. The plots show that patients in either the prednisolone treatment group or the placebo group have similar survival distribution over time. However, results from the log-rank test show a significant difference in the survival distribution of the constriction outcome ($p < 0.009$). This result is similar to those presented by Mayosi et al. (2014). In order to have a clear picture of what happens within a few days of the trial, we also examine the survival distribution over six months.(Figure 3.2).

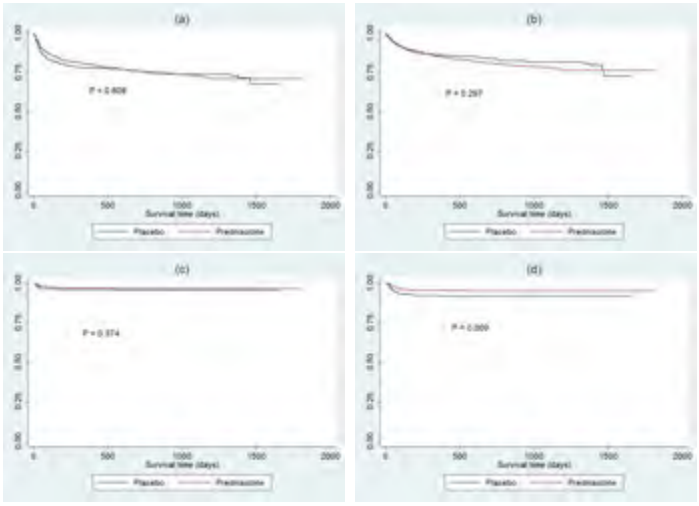

Figure 3.1: *Estimated survival function for prednisolone and placebo group until the end of study. (a) composite outcome, (b) death, (c) constriction, (d) cardiac tamponade.*

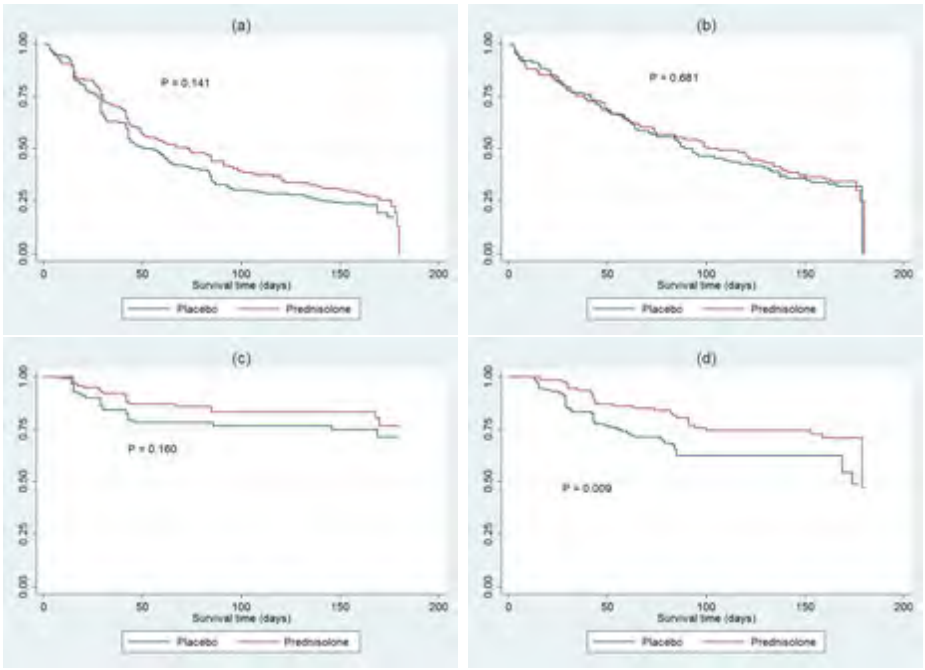

Figure 3.2: *Estimated survival function for prednisolone and placebo group at 6 months of the study. (a) composite outcome, (b) death, (c) constriction, (d) cardiac tamponade.*

3.5.2 Cox proportional hazards models

In order to determine the risk factors associated with the composite outcome, death, constriction and cardiac tamponade, univariate analysis was performed using the regression model presented in Section 3.2 before proceeding to multivariate models. A value of 0.1 significance level was used to identify variables as significant in the univariate model using forward stepwise regression method. We then fit the multivariate Cox proportional hazards model including all the risk factors and treatment groups. In the univariate and the multivariate proportional haz-

ards models, age, weight, NYHA class, hypotension (SBP), creatinine, and peripheral oedema show a statistically significant association with the composite outcome. Also, weight, NYHA class, hypotension (SBP), creatinine and peripheral oedema show a statistically significant association with death. In addition, NYHA class and hypotension (SBP) show a statistically significant association with cardiac tamponade. Lastly, prednisolone treatment, gender, NYHA class, tachycardia (HR), Haemoglobin level, peripheral oedema, pulmonary infiltrate and HIV status show a statistically significant association with constriction.

Tables 3.1 - 3.4 give the results of the four regression models examining the influence of the covariates on the hazard rate for composite outcome, death, cardiac tamponade and constriction, respectively. As observed from the Kaplan-Meier estimates and Mayosi et al. (2014) result, the prednisolone treatment effect was not associated with the risk of composite outcome, death and cardiac tamponade even after adjusting for other risk factors in the multivariate Cox PH models. But, it remained significant in the outcome of constriction as it was in the Kaplan-Meier estimates. The NYHA class had a significant effect on the hazards rate for all types of the event, though with a large confidence interval. A comparison between different levels of creatinine exposed an average hazard ratio for death of 1.82 (95%CI 1.29 to 2.58), signifying a much higher degree of death event for individuals that have creatinine levels >105 umol/l compared to patients that have ≤105 umol/l. The effect of hypotension (SBP) is similar in the death and cardiac tamponade events with a higher effect for patients that have ≤90 mmHg hypotension level compared to patients that have hypotension level of >90 mmHg. A comparison between patients that have and that do not have pulmonary infiltrate revealed an average hazard ratio for constriction of 1.55 (95%CI 0.96 to 2.50), indicating a much higher incidence of constriction for patients that have pulmonary infiltrate compared to patients that do not have pulmonary infiltrate. HIV status had a significant effect on the hazard of constriction, with a hazard ratio of 0.52 (95%CI 0.30 to 0.89). This result indicates a lesser incidence of constriction for HIV-positive patients and suggests that HIV infection is associated with a reduced incidence of the development of constrictive pericarditis. Ntsekhe et al. (2008) in his article titled "HIV Infection Is Associated with a Lower Incidence of Constriction in Presumed Tuberculous Pericarditis: A Prospective Observational Study," wrote that this observation suggested that HIV-positive patients have a lower incidence of constriction because the immune suppression associated with HIV reduces the risk of the development of pericardial fibrosis.

Table 3.1: Univariate and multivariate Cox PH model for the relative hazard of the composite outcome.

	Univariate analysis		Multivariate analysis	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	0.94 (0.76 - 1.17)	0.609	0.94 (0.74 - 1.20)	0.643
Age (years)	1.00 (1.00 - 1.02)	0.030	1.01 (1.00 - 1.02)	0.041
Weight (kg)	0.91 (0.83 - 1.00)	0.042	0.91 (0.81 - 1.00)	0.050
NYHA Class				
I	1.00		1.00	
II	1.63 (1.12 - 2.37)	0.010	1.46 (0.94 - 2.26)	0.088
III	2.73 (1.86 - 4.01)	<0.001	2.38 (1.49 - 3.81)	<0.001
IV	3.24 (2.08 - 5.07)	<0.001	2.80 (1.65 - 4.75)	<0.001
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.59 (0.42 - 0.82)	0.002	0.68 (0.47 - 1.00)	0.050
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.75 (0.60 - 0.93)	0.011	0.87 (0.67 - 1.12)	0.282
Creatinine				
$\leq$ 105 umol/l	1.00		1.00	
$>$ 105 umol/l	1.51 (1.12 - 2.05)	0.006	1.68 (1.20 - 2.35)	0.002
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	1.47 (1.15 - 1.88)	0.002	0.88 (0.65 - 1.20)	0.427
Peripheral oedema				
No	1.00		1.00	
Yes	1.75 (1.41 - 2.17)	<0.001	1.44 (1.11 - 1.85)	0.005
Pulmonary infiltrate on CXR				
No	1.00		1.00	
Yes	1.23 (0.97 - 1.55)	0.087	1.15 (0.89 - 1.49)	0.276
Effusion size				
small	1.00		1.00	
medium	1.57 (0.93 - 2.64)	0.093	1.69 (0.90 - 3.19)	0.101
large	1.64 (1.00 - 2.69)	0.048	1.55 (0.85 - 2.82)	0.151
AIC			3560.019	

NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure; CXR: Chest Xray.
CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table 3.2: Univariate and multivariate Cox PH model for the relative hazard of death.

	Univariate analysis		Multivariate analysis	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	1.14 (0.89 - 1.47)	0.298	1.14 (0.87 - 1.47)	0.341
Weight (kg)	0.85 (0.76 - 0.96)	0.006	0.88 (0.78 - 0.98)	0.026
NYHA Class				
I	1.00		1.00	
II	1.42 (0.93 - 2.16)	0.107	1.27 (0.79 - 2.05)	0.328
III	2.20 (1.42 - 3.43)	<0.001	1.87 (1.14 - 3.09)	0.014
IV	3.14 (1.91 - 5.16)	<0.001	2.69 (1.55 - 4.66)	<0.001
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.45 (0.31 - 0.64)	<0.001	0.54 (0.37 - 0.79)	0.002
Creatinine				
$\leq$ 105 umol/l	1.00		1.00	
$>$ 105 umol/l	1.79 (1.28 - 2.52)	0.001	1.82 (1.29 - 2.58)	0.001
Peripheral oedema				
No	1.00		1.00	
Yes	1.57 (1.23 - 2.02)	<0.001	1.40 (1.06 - 1.85)	0.017
AIC			2940.553	

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table 3.3: Univariate and multivariate Cox PH model for the relative hazard of cardiac tamponade.

	Univariate analysis		Multivariate analysis	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	0.78 (0.44 - 1.36)	0.376	0.57 (0.28 - 1.12)	0.104
NYHA Class				
I	1.00		1.00	
II	3.04 (0.92 - 10.10)	0.069	2.87 (0.65 - 12.58)	0.162
III	4.65 (1.36 - 15.87)	0.014	5.11 (1.15 - 22.73)	0.032
IV	4.86 (1.21 - 19.43)	0.025	6.45 (1.25 - 33.40)	0.026
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.36 (0.17 - 0.74)	0.005	0.39 (0.16 - 0.94)	0.036
Atrial fibrillation on ECG				
No	1.00		1.00	
Yes	2.41 (0.85 - 6.82)	0.098	2.27 (0.79 - 6.46)	0.126
AIC			477.3476	

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table 3.4: Univariate and multivariate Cox PH model for the relative hazard of the constriction.

	Univariate analysis		Multivariate analysis	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	0.56 (0.36 - 1.87)	0.010	0.57 (0.35 - 0.92)	0.022
Age (years)	1.01 (1.00 - 1.03)	0.046	1.00 (0.98 - 1.02)	0.898
Gender				
Male	1.00		1.00	
Female	0.52 (0.32 - 0.82)	0.005	0.57 (0.33 - 0.96)	0.036
NYHA Class				
I	1.00		1.00	
II	1.93 (0.86 - 4.36)	0.111	1.20 (0.51 - 2.81)	0.682
III	3.85 (1.70 - 8.72)	0.001	2.49 (1.02 - 6.10)	0.045
IV	3.97 (1.54 - 10.24)	0.004	3.03 (1.09 - 8.45)	0.034
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	3.26 (0.80 - 13.27)	0.098	2.61 (0.63 - 10.81)	0.186
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.46 (0.29 - 0.75)	0.002	0.49 (0.29 - 0.85)	0.011
Haemoglobin				
$\leq$ 10 g/dl	1.00		1.00	
$>$ 10 g/dl	0.40 (0.26 - 0.64)	$<$0.001	0.52 (0.30 - 0.91)	0.021
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	1.99 (1.26 - 3.14)	0.003	0.97 (0.54 - 1.74)	0.919
Peripheral oedema				
No	1.00		1.00	
Yes	2.54 (1.64 - 3.94)	$<$0.001	1.58 (0.94 - 2.67)	0.081
Pulmonary infiltrate				
No	1.00		1.00	
Yes	1.49 (0.94 - 2.37)	0.090	1.55 (0.96 - 2.50)	0.073
HIV status				
Negative	1.00		1.00	
Positive	0.41 (0.27 - 0.63)	$<$0.001	0.52 (0.30 - 0.89)	0.017
AIC			954.724	

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC:Akaike's Information Criterion

3.5.3 Model checking

After a Cox proportional hazards model is fitted, the adequacy of this model, inclusive of both the proportionality assumption and the goodness of fit, need to be examined. Using Schoenfeld residuals test (Table 3.5), there is no evidence that the proportionality assumption is violated for any of the fitted models (p-value $>$ 0.05).

Table 3.5: Schoenfeld residual global test of the multivariate models fitted to the different outcomes for checking proportionality assumption.

Outcomes	χ^2	df	Prob> χ^2
Composite	7.69	14	0.9501
Death	5.09	8	0.8002
Cardiac tamponade	3.80	6	0.7035
Constriction	14.86	14	0.3879

df: degree of freedom; χ^2 test of statistics

A plot of the Cox-Snell residuals is presented in Figures 3.3 to check the goodness of fit by residual plots. If the models fit the data perfectly, the graphs would be a straight line through the origin. In the event of death and composite outcome, the line does not deviate too much from the reference line, hence the Cox model provides a reasonably good fit for the data. However, a noticeable deviation of the curve from the 45-degree line was observed in the event of cardiac tamponade and constriction. It is very common for models with censored data to have some wiggling at large values of time and it is something which should not cause much concern. Overall we could conclude that the final models fits the data well.

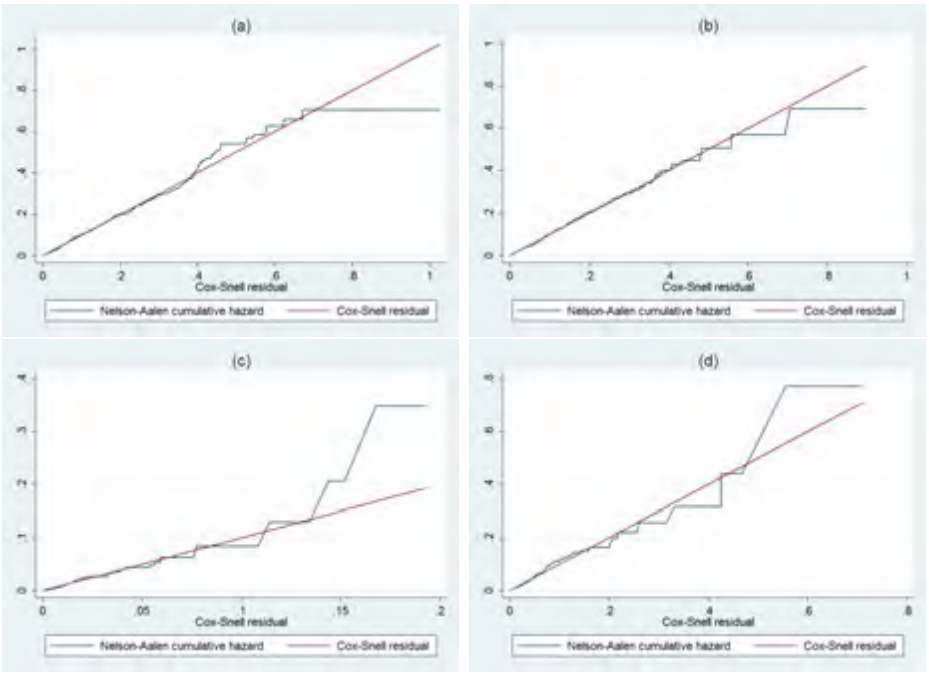

Figure 3.3: *Cox-Snell residual of Cox PH model for (a) composite outcome (b) death (c) cardiac tamponade (d) constriction.*

The plot of deviance residuals against the linear predictors for the composite outcome, death, cardiac tamponade and constriction event shows that the deviance residuals seem not to be

symmetrically distributed about zero. There are high or low deviance residuals which suggest that these subjects may be outliers; (Figure 3.4). Using likelihood displacement values to measure each subject's influence on the coefficient vector as a whole, results show some subjects are influential (Figure 3.5). Likelihood displacement values measure influence by approximating what happens to the log-likelihood when subject i is omitted. Furthermore, using delta-beta statistic the results shows that the coefficient does not change when the subjects corresponding to the largest delta-beta statistics are removed. Therefore, we do not remove the subjects from the dataset and conclude that there are no influential subjects.

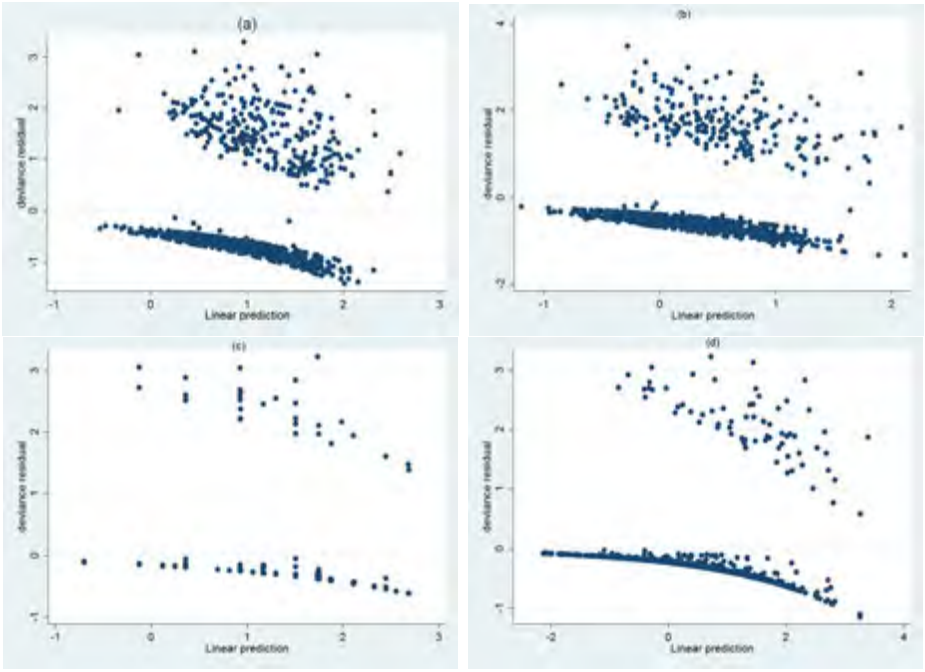

Figure 3.4: *Deviance residual of Cox PH model for (a) composite outcome (b) death (c) cardiac tamponade (d) constriction.*

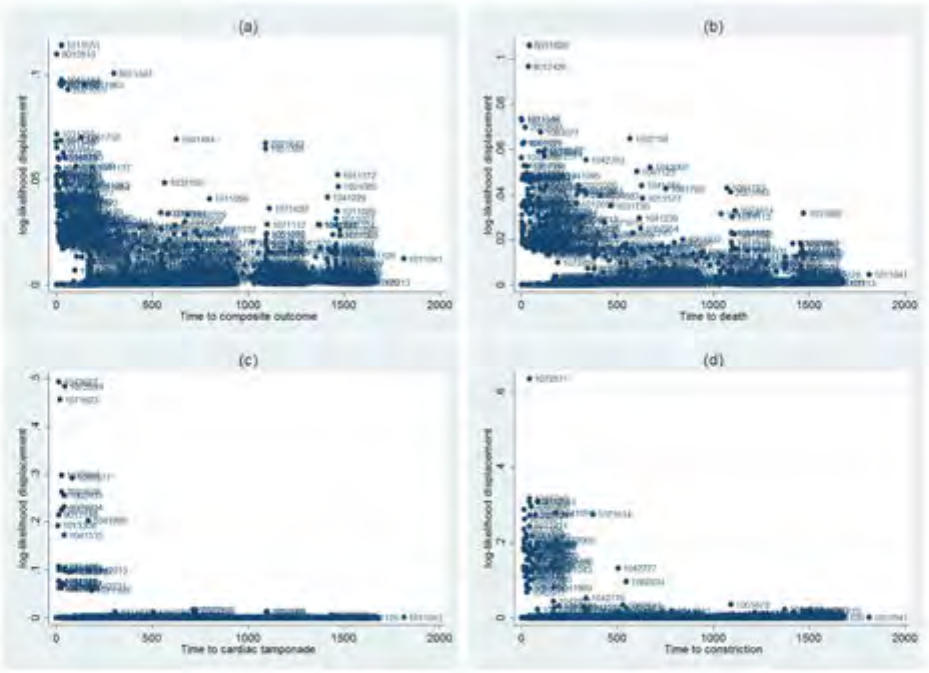

Figure 3.5: *Likelihood displacement plot of Cox PH model for (a) composite outcome (b) death (c) cardiac tamponade (d) constriction.*

3.6 Frailty model

In the analysis of survival data, situations where the survival times are not independent are frequently encountered. Such data tends to arise when different individuals have some features in common. As an example, in a multi-centre study, the survival experience of individuals from the same study centre may be more similar than for individuals from different centres. This could be because of different surgical teams in the different centres, or different nursing practices across the centres. Ignoring this heterogeneity in these centres may alter the interpretation and reporting of the treatment effect.

Frailty models are potential choices for modelling unexplained heterogeneity in a population. In its simplest form, a frailty is an unobserved random factors that modifies multiplicatively the hazard function of an individual or a group or cluster of individual. The concept of frailty offers a way to appropriately introduce random effects in the model to account for association and unseen heterogeneity. Frailty models extend Cox's proportional hazards model by introducing unobserved 'frailties' to the model. In this case, the hazard rate will not be just a function of covariates, but also a function of frailties. The term frailty was introduced by Vaupel et al. (1979) and it was used in the univariate survival models. Clayton (1978) promoted the model by its application to a multivariate situation on chronic incidence in families and it has been

further studied by Therneau and Grambsch (2000), Duchateau et al. (2002), and Duchateau and Janssen (2004), and in Hougaard (2012). The multivariate or shared frailty model is a conditional independence model in which frailty is common to all subjects in a cluster. The shared frailty model is responsible for creating dependence between event times. It assumes that, given the frailty, all event times in a cluster are independent.

The hazard function for the frailty model, either semi-parametric or parametric, can be written by adding frailty to equation (3.2.1) as

$$h_{ij}(t) = u_i h_0(t) \exp(\boldsymbol{\beta}' \boldsymbol{x}_{ij}), \tag{3.6.1}$$

where h_{ij} is the hazard function for the jth observation from the ith cluster, $h_0(t)$ is the unspecified baseline hazard, $\boldsymbol{x}_{ij}$ are the covariates associated with $\boldsymbol{\beta}'$ the regression coefficients, and u_i is the frailty of cluster i. u_i are assumed to be independently and identically distributed from a gamma distribution with mean 1 and unknown variance θ. A Large value of θ indicate a closer positive relationship between the subjects of the same group and greater heterogeneity among the groups.

Parametric or non-parametric forms of baseline hazard can be assumed in frailty models just as it is in the proportional hazards model. If a parametric form of $h_0(t)$ is assumed, then maximum likelihood estimates can be obtained by maximizing the likelihood function using the Newton Raphson method. But, if a non-parametric form is assumed for $h_0(t)$, then the semi-parametric proportional hazards model is considered and the estimates are usually obtained by using the Expectation-Maximization (EM) algorithm. These methods basically seek to solve the log-likelihood in which the frailty, will be integrated out.

3.6.1 Gamma distribution

Gamma distribution is the main frailty distribution widely used in the literature. From a computational and analytical point of view, the gamma distribution is convenient because it is easy to derive the closed form expressions of survival, density and the hazard function. This is due to the simplicity of the derivatives of the Laplace transform.

Suppose a random variable $T > 0$ is gamma distributed with scale parameter $\beta > 0$ and shape parameter $\alpha > 0$. The probability density function of a random variable T is

$$f_T(t) = \frac{\beta^\alpha t^{\alpha-1} \exp^{-\beta t}}{\Gamma(\alpha)}, t > 0,$$

where $\Gamma(\cdot)$ is the gamma function. The corresponding survival function is given by

$$S(t) = \frac{\Gamma(\alpha, \beta t)}{\Gamma(\alpha)}$$

and hazard function as

$$h(t) = \frac{\beta^\alpha}{\Gamma(\alpha, \beta t)} t^{\alpha-1} \exp^{-\beta t}, t > 0.$$

3.6.2 Application

In the models presented in Tables 3.6 - 3.9, the heterogeneity parameter θ and the p-value resulting from the likelihood ratio test comparing the models with and without frailties in the incidence of composite outcome ($\theta = 0.078$, p-value = 0.048), death ($\theta = 0.034$, p-value = 0.093), cardiac tamponade ($\theta = 0.070$, p-value = 0.326), and constriction ($\theta = 0.999$, p-value < 0.001) shows that there is a highly significant centre effect in the incidence of constriction and boardeline significant in the incidence of composite outcome. This was also illustrated by the distribution of estimated random effects, depicted in Figure 3.6

In comparing the multivariate shared frailty model with the multivariate Cox proportional hazards models, as presented in Tables (3.1, 3.2, 3.3, and 3.4), it was observed that the same set of variables are significant in the composite outcome, death, and cardiac tamponade. The variables have similar estimates. However, there is variability in the significance of variables in the constriction outcome. The Cox PH model estimates the effect of tachycardia heart rate and peripheral oedema on the hazard of constriction to be 0.49(95% CI = 0.29 - 0.85, p-value = 0.011) and 1.58(95% CI = 0.94 - 2.67, p-value 0.081), respectively. On the other hand, the shared frailty model estimates the effect of tachycardia heart rate and peripheral oedema on the hazard of constriction to be 0.66(95% CI = 0.38 - 1.17, p-value = 0.144) and 1.77(95% CI = 1.02 - 3.08, p-value = 0.044), respectively.

Table 3.6: Multivariable shared frailty model for the relative hazard of the composite outcome.

Characteristics	HR (95%CI)	p-value
Treatment		
Placebo	1.00	
Prednisolone	0.94 (0.74 - 1.19)	0.590
Age (years)	1.01 (1.00 - 1.02)	0.047
Weight (kg)	0.90 (0.81 - 1.00)	0.050
NYHA Class		
I	1.00	
II	1.44 (0.92 - 2.25)	0.107
III	2.33 (1.44 - 3.76)	0.001
IV	2.39 (1.38 - 4.14)	0.002
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	0.69 (0.47 - 1.01)	0.056
Tachycardia (HR)		
$\leq$ 100 mmHg	1.00	
$>$ 100 mmHg	0.91 (0.70 - 1.18)	0.472
Creatinine		
$\leq$ 105 umol/l	1.00	
$>$ 105 umol/l	1.69 (1.20 - 2.37)	0.002
Palpable pulsus paradoxus		
No	1.00	
Yes	0.89 (0.64 - 1.23)	0.475
Peripheral oedema		
No	1.00	
Yes	1.51 (1.15 - 1.99)	0.003
Pulmonary infiltrate on CXR		
No	1.00	
Yes	1.17 (0.91 - 1.52)	0.228
Effusion size		
small	1.00	
medium	1.62 (0.86 - 3.06)	0.135
large	1.49 (0.81 - 2.72)	0.198
θ (SE)	0.078 (0.072)	0.048

NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure; CXR: Chest Xray.
CI: Confidence Interval; HR: Hazard Ratio; SE: Standard Error; θ: Heterogeneity parameter

Table 3.7: Multivariable shared frailty model for the relative hazard of the death outcome.

Characteristics	HR (95%CI)	p-value
Treatment		
Placebo	1.00	
Prednisolone	1.14 (0.87 - 1.49)	0.328
Weight (kg)	0.87 (0.78 - 0.98)	0.024
NYHA Class		
I	1.00	
II	1.22 (0.76 - 1.98)	0.413
III	1.79 (1.08 - 2.96)	0.024
IV	2.51 (1.43 - 4.42)	0.001
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	0.54 (0.37 - 0.80)	0.002
Creatinine		
$\leq$ 105 umol/l	1.00	
$>$ 105 umol/l	1.80 (1.27 - 2.55)	0.001
Peripheral oedema		
No	1.00	
Yes	1.50 (1.12 - 2.01)	0.007
θ (SE)	0.034 (0.039)	0.093

NYHA: New York Heart Association; SBP: Systolic Blood Pressure
CI: Confidence Interval; HR: Hazard Ratio; SE: Standard Error; θ: Heterogeneity parameter

Table 3.8: Multivariable shared frailty model for the relative hazard of cardiac tamponade outcome.

Characteristics	HR (95%CI)	p-value
Treatment		
Placebo	1.00	
Prednisolone	0.57 (0.28 - 1.13)	0.105
NYHA Class		
I	1.00	
II	2.80 (0.63 - 12.35)	0.174
III	4.85 (1.08 - 21.79)	0.040
IV	5.79 (1.10 - 30.38)	0.038
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	0.38 (0.16 - 0.92)	0.034
Atrial fibrillation on ECG		
No	1.00	
Yes	2.28 (0.80 - 6.50)	0.124
θ (SE)	0.070 (0.186)	0.326

NYHA: New York Heart Association; SBP: Systolic Blood Pressure
CI: Confidence Interval; HR: Hazard Ratio; SE: Standard Error; θ: Heterogeneity parameter

Table 3.9: Multivariable shared frailty model for the relative hazard of constriction.

Characteristics	HR (95%CI)	p-value
Treatment		
Placebo	1.00	
Prednisolone	0.55 (0.34 - 0.91)	0.018
Age (years)	1.00 (0.98 - 1.02)	0.683
Gender		
Male	1.00	
Female	0.54 (0.32 - 0.93)	0.026
NYHA Class		
I	1.00	
II	1.47 (0.61 - 3.55)	0.389
III	3.36 (1.34 - 8.39)	0.010
IV	2.57 (0.87 - 7.58)	0.088
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	2.76 (0.65 - 11.67)	0.169
Tachycardia (HR)		
$\leq$ 100 mmHg	1.00	
$>$ 100 mmHg	0.66 (0.38 - 1.17)	0.144
Haemoglobin		
$\leq$ 10 g/dl	1.00	
$>$ 10 g/dl	0.54 (0.31 - 0.96)	0.035
Palpable pulsus paradoxus		
No	1.00	
Yes	0.78 (0.42 - 1.46)	0.434
Peripheral oedema		
No	1.00	
Yes	1.77 (1.02 - 3.08)	0.044
Pulmonary infiltrate		
No	1.00	
Yes	1.48 (0.90 - 2.43)	0.125
HIV status		
Negative	1.00	
Positive	0.56 (0.32 - 0.98)	0.041
θ (SE)	0.999 (0.562)	$<$0.001

NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure

CI: Confidence Interval; HR: Hazard Ratio; SE: Standard Error; θ: Heterogeneity parameter

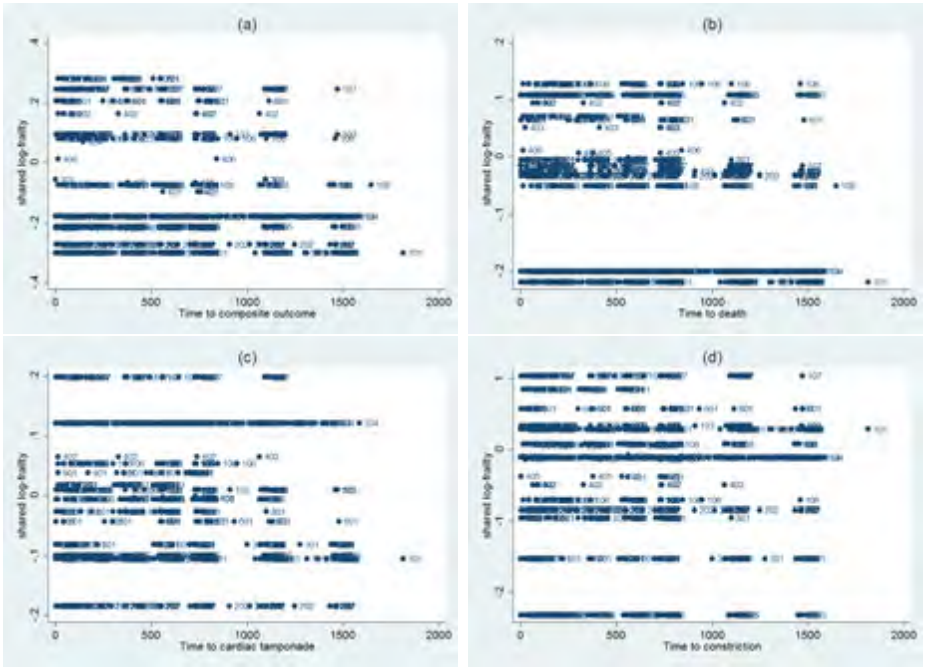

Figure 3.6: *Estimated random effect(s) (frailties) for each centre for the (a) composite outcome, (b) death, (c) cardiac tamponade, and (d) constriction. In South Africa, there are seven centres coded as, 101 = Cape Town, 103 = Port Elizabeth, 104 = Mthathia, 105 = Durban, 106 = Soweto, 107 = Pretoria, 108 = Johannesburg; Mozambique (have one centre coded as, 202 = Maputo), Malawi (have one centre coded as 301 = Lilongwe), Nigeria (have six centres coded as 401 = Ibadan, 402 = Kano, 403 = Abeokuta, 404 = Abuja, 405 = Enugun, 406 = Calabar), Sierra Leone (have one centre coded as 601 = Sierra Leone), Zimbabwe (have one centre coded as 701 = Harare), Uganda (have one centre coded as 801 = Kampala), and Kenya (have one centre coded as 901 = Nairobi).*

Comparison of Cox PH versus shared frailty model

Table 3.10 gives the log-likelihood, AIC and BIC values of the two models. From the table we can see that the shared gamma frailty model has both a minimum AIC and BIC value, and larger log-likelihood, indicating that this model fits the data better than the Cox PH model which did not take the clustering (centres) effect into account.

Table 3.10: Comparison of Cox PH and shared frailty model.

	Cox PH			Shared failty		
Outcomes	$\ell\ell$ (model)	AIC	BIC	$\ell\ell$ (model)	AIC	BIC
Composite	-1766.01	3560.019	3630.126	-1764.626	3557.252	3627.358
Death	-1462.276	2940.553	2981.587	-1461.40	2938.802	2979.837
Cardiac	-232.6738	477.3476	507.1267	-232.5717	477.1434	506.9225
Constriction	-464.3619	954.7239	1021.353	-456.3408	938.6815	1005.31

$\ell\ell$: log-likelihood; AIC: Akaike's Information Criterion; BIC: Bayesian Information Criterion

3.7 Chapter summary

In this chapter, we reviewed the probability density function, survival function, hazard function, and the cumulative hazard function as the important functions in the time-to-event data analysis. We also reviewed the concept of Cox proportional hazard (PH) model and shared frailty model for modelling heterogeneity in time-to-event data. These models were implemented to the IMPI multicentre clinical trial dataset.

Using a 0.1 level of significance, we selected variables (risk factors) that are significant to the outcomes of interest in the IMPI data i.e. composite outcome and individual outcome of death, cardiac tamponade and constrictive pericarditis. We then performed a multivariate analysis using these significant variables and treatment groups. This was considered because the previous analysis performed on the dataset by Mayosi et al. (2014) only modelled the effect of the treatments on the outcomes without considering other risk factors. The result of the multivariate analysis shows that prednisolone therapy, as compared with placebo , was associated with significant reductions in the incidence of constrictive pericarditis. This result also indicates a lesser incidence of constrictive pericarditis for HIV-positive patients and suggests that HIV infection is associated with a reduced incidence of the development of constrictive pericarditis.

The proportionality assumption check was done using Schoenfeld residuals; the result shows that the proportionality assumption was not violated for all the variables. Also, the graph of Cox-Snell residuals indicated the Cox PH model as a good fit for the data. Furthermore, results from the shared frailty models show that there is a significant random effect, because the heterogeneity factor is different from zero, but this only slightly changed the estimated effect of the covariates as compared to the Cox PH models. In the constrictive pericarditis event, Tachycardia HR was found to be significant in the Cox model but was insignificant in the shared frailty model. Also, peripheral oedema was found to be significant in the shared frailty model but was insignificant in the Cox model.

Chapter 4

Parametric models

The most common way of analysing survival data is via the Cox proportional hazards model. This might be because the model only makes the assumption of proportional hazards, but does not impose a functional form for the hazard (Collet 2003). However, when the proportional hazards' assumption is not tenable, parametric survival models which assume that the event times follow a specified distribution, can be used. Even though, in this thesis, the proportionality assumption was met for the Cox models presented in Chapter 3, we wish to explore the parametric approach for completion and comparison purposes. In this section, we review parametric models for the analysis of time-to-event data. We discuss that if the proportionality assumption is not satisfied in the Cox model, the parametric model is an alternative option. We present some event distributions used in the parametric model. We then apply the parametric proportional model and accelerated failure time model to the IMPI trial data. Furthermore, parametric proportional models are compared to parametric shared-frailty models. The frailty model estimation method has been discussed in Section 3.6 of this thesis.

4.1 Parametric proportional hazards models

A parametric proportional hazards model has the same general form as equation (3.2.1) given above for the Cox regression model. However, a parametric proportional hazards model differs from a Cox proportional hazards model, in that an event time distribution for the baseline hazard $h_0(t)$ is specified. Also, the coefficients are estimated by a partial likelihood while in the Cox model, it is estimated by a maximum likelihood. Popular event time distribution are the exponential, Weibull, and Gompertz distributions.

4.2 Accelerated failure time models

Although parametric proportional hazards models are very applicable to analyze survival data, there are relatively few probability distributions for the survival time that can be used with these models. In these situations, the accelerated failure time (AFT) model is an alternative to the parametric proportional hazard model for the analysis of survival time data. The AFT model treats the logarithm of survival time as the response variable and includes an error term that is assumed to follow a particular distribution. The equation (4.2.1) shows the log-linear survival time, $X_1 \dots X_p$ are explanatory variables with coefficients $\beta_1 \dots \beta_p$, ϵ_i represents residual or unexplained variation in the log-transformed survival times, while β_0 and σ are intercept and scale parameters, respectively.

$$\log T_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \sigma\epsilon_i, \tag{4.2.1}$$

An initial step in fitting an AFT model is determining which distribution should be specified for the survival times T_i (equation 4.2.1). Under the AFT model parametrization, the distribution chosen for T_i dictates the distribution of the error term ϵ_i. For instance, if survival times are modeled as a Weibull distribution, the error term is assumed to follow an extreme-value distribution. Likewise, if survival times are modeled using the log-logistic or log-normal distribution, the ϵ_i are assumed to be logistic or normal, respectively.

The survivor function of T_i, the random variable associated with the survival time of the i^{th} individual, is then:

$$\begin{aligned} S_i(t) &= \Pr\{T_i \geq t\} \\ &= \Pr\{\exp(\mu + \boldsymbol{\beta}' x_i + \sigma\epsilon_i) \geq t\} \\ &= \Pr\{\exp(\mu + \sigma\epsilon_i) \geq t/\exp(\boldsymbol{\beta}' x_i)\} \\ &= S_0\{t/\exp(\boldsymbol{\beta}' x_i)\}. \end{aligned} \tag{4.2.2}$$

Using the relationship between the hazard and survivor function given in equation 3.1.5 we can obtain a general hazard function for the AFT model from equation 4.2.2 as follows:

$$
\begin{aligned}
S_i(t) &= S_0(t/e^{\boldsymbol{\beta}' x_i}) \\
-\log[S_i(t)] &= -\log[S_0(t/e^{\boldsymbol{\beta}' x_i})] \text{(taking logs and multiplying by -1)} \\
-\frac{d}{dt}\log[S_i(t)] &= -\frac{d}{dt}\log[S_0(t/e^{\boldsymbol{\beta}' x_i})] \text{(taking derivatives on both sides)} \\
\lambda_i(t) &= -e^{\boldsymbol{\beta}' x_i}\lambda_0(t/e^{\boldsymbol{\beta}' x_i}) \text{from (Equation 3.1.5)}
\end{aligned}
$$

General forms of the PH and AFT Models

For proportional hazards (PH) models (based on the hazard for one group being proportional to the hazard of the other group):

$$\lambda_A(t) = \psi\lambda_B(t)$$

and equivalently

$$S_A(t) = [S_B(t)]^{\psi}$$

with $\psi = \exp(\boldsymbol{\alpha}' x)$

For the accelerated failure time (AFT) models (based on the survival time of subjects in one group being a multiple of the survival time of subjects in the other group):

$$\lambda_A(t) = \phi^{-1}\lambda_B(t)(t/\phi)$$

with $\phi = \exp(\boldsymbol{\beta}' x)$ and:
$\phi < 1$ implies that the time to the vent is accelerated (shorter) and
$\phi > 1$ implies that the time to event is decelerated (longer).

Weibull model is the commonly used model that can be parameterized in both the PH and AFT metrics.

4.3 Event time distribution

In this section some event time distributions, which are frequently considered for parametric survival models are summarized.

4.3.1 Exponential distribution

The exponential distribution is a one-parametric event time distribution implying a time constant hazard rate $\lambda(t) = \lambda$. The density function of the exponential distribution is

$$f(t) = \lambda \exp(-\lambda t),$$

and the survival function is

$$S(t) = \exp(-\lambda t).$$

In a regression model investigating the influence of covariates on the event times, the hazard rate λ is commonly modelled using

$$\lambda = \exp(\beta' x),$$

to ensure positivity of the estimated hazard rates.

4.3.2 Weibull distribution

The Weibull distribution is defined by the two parameters λ and α, and therefore it is more flexible than the exponential distribution. Different parametrizations of the Weibull distribution exist in the literature. One possible formulation of the density function is

$$f(t) = \lambda\alpha(\lambda t)^{\alpha-1} \exp\{-(\lambda t)^{\alpha}\},$$

therefore the exponential distribution is a special case of the Weibull distribution for $\alpha = 1$. The survival function for that parametrization is

$$S(t) = \exp\{-(\lambda t)^{\alpha}\},$$

and the hazard function can be denoted as

$$\lambda(t) = \lambda\alpha(\lambda t)^{\alpha-1}.$$

4.3.3 Gompertz distribution

The survival function of the Gompertz distribution is given by

$$S(t) = \exp\{\frac{\lambda}{\theta}(1 - e^{\theta t})\}$$

and the corresponding hazard function

$$h(t) = \lambda \exp(\theta t)$$

for $0 \leq t < \infty$ and $\lambda > 0$. The shape of the hazard function is determined by the parameter θ. The survival time has an exponential distribution when $\theta = 0$, that is, the exponential distribution is also a special case of the Gompertz distibution.

4.3.4 Log-logistic distribution

The log-logistic distribution has a fairly flexible functional form. If $Y = \log(T)$ follows logistics distribution with location parameter μ and scale parameter σ having probability density function

$$f(y) = \sigma^{-1}\exp((y-\mu)/\sigma)(1+\exp(y-\mu)/\sigma)^{-2},$$

then the lifetime T follows a log-logistics distribution with scale parameter α (> 0) and shape parameter β (> 0) having a probability density function of the form,

$$f(t) = (\beta/\alpha)(t/\alpha)^{\beta-1}(1+(t/\alpha)^{\beta})^{-2}, \qquad t > 0$$

where $\alpha = \exp(\mu)$ and $\beta = 1/\sigma$. The corresponding hazard and survival functions can be written as

$$\lambda(t) = (\beta/\alpha)(t/\alpha)^{\beta-1}(1+(t/\alpha)^{\beta})^{-1}, \qquad \text{and}$$
$$S(t) = (1+(t/\alpha)^{\beta})^{-1}.$$

The general shape of the hazard function of a log-logistic distribution is very similar to that of the log-normal distribution.

4.3.5 Log-normal distribution

The log-normal distribution is defined by two parameters λ and α. The density function for a failure time t can be written as

$$f(t) = (2\pi)^{-1/2}\alpha t^{-1}\exp\left\{\frac{-\alpha^2(\log\lambda t)^2}{2}\right\},$$

and the corresponding survival function and hazard function can be written as

$$S(t) = 1 - \Phi(\alpha\log\lambda t), \qquad \text{and} \qquad \lambda(t) = \frac{f(t)}{S(t)}.$$

where $\Phi(x)$ is the cumulative density function of the standard normal distribution.

4.4 Goodness of fit test for the AFT model

The Cox-Snell residuals' diagnostic plot can be used to evaluate the overall fit of the AFT model. The Cox-Snell residuals are calculated by using the cumulative hazard $H(t_i, \beta, \sigma)$ function and standardized residual as:

$$rs_i = \frac{\log t_i - (\hat{\beta}_0 + \hat{\beta}_i X_i)}{\hat{\sigma}},$$

where $\hat{\beta}_0$, $\hat{\beta}$ and $\hat{\sigma}$ are the maximum likelihoods estimates of β_0, β and σ respectively.

The Cox-Snell residual can be applied to any parametric model. For the Weibull AFT model, the Cox-Snell residuals is

$$rc_i = \exp(rs_i).$$

Cox-Snell residuals for the Log-logistic AFT model will be

$$rc_i = \log\{1 + \exp(rs_i)\}.$$

For the Lognormal AFT model, Cox-Snell residual is

$$rc_i = \log\{1 + \Phi(rs_i)\},$$

where $\Phi(.)$ is the cumulative distribution function of the standard Normal distribution. In the Cox-Snell residuals plot, the fitting of the model is good if the plotted points lie on a line that has an intercept zero and a unit slope.

Another method of assessing the goodness of fit of an AFT model is Akaike's Information Criterion (AIC). It can be computed as follows,

$$AIC = -2\ell + 2(b + d),$$

where ℓ =log-likelihood of the model, b, number of parameters of the assumed probability distribution and d the number of coefficients excluding constant in the final model. A model with a smaller value of AIC can be considered as a better model compared to other models under consideration.

4.5 Application

As explained in Section 4.1, the exponential, Weibull, and Gompertz are the popular event-time distributions used in the parametric proportional hazards model. By using log-likelihood

and AIC to compare the three distributions, the Weibull distribution was discovered to be a suitable distribution for the fit of the model because it has the highest log-likelihood and lowest AIC, hence, Weibull is used as the underline distribution in the model presented in Tables 4.1 - 4.4.

The result of the analysis in Tables 4.1 - 4.4 shows that patients in NYHA class 1, patients that have hypotension (SBP) >90 mmHg, have creatinine ≤105 umol/l and do not have peripheral oedema, have lower hazards of the composite outcome. Also, a one-year increase in age increases the hazard of a composite outcome by 1.01 and a one-kilogram increase in weight increases the hazard of a composite outcome by 0.90. Furthermore, patients in NYHA class 1, patients that have hypotension (SBP) >90 mmHg, have creatinine ≤105 umol/l, and do not have peripheral oedema, have a lower hazards of death. Also, a one-kilogram increase in weight increases the hazard of death by 0.87. In addition, patients in NHYA class 1 that have hypotension (SBP) that is >90 mmHg, have lower hazards of cardiac tamponade. Lastly, female patients, patients in NYHA class 1, patients having tachycardia (HR) >100 mmHg, have haemoglobin >10 g/dl and patients that are HIV-negative patients have lower hazards of constriction. The Cox-Snell residuals' plot for the Weibull PH models shows a deviation from the straight line passing through the origin (see Figure 4.1) for the composite outcome and the individual outcome of death, cardiac tamponade and constriction. This indicates that the overall fit of the Weibull PH model is not a good fit for the data.

The result from the parametric shared frailty model shows that the likelihood ratio test of $H_0 : \hat{\theta} = 0$, for the hazard of composite outcome and hazard of constriction are significant (p-value 0.019 and p-value < 0.001 respectively). Meanwhile, the likelihood ratio test was not significant in the hazards of death and cardiac tamponade (p-value $= 0.072$ and p-value $= 0.212$ respectively).

By comparing the Weibull parametric proportional hazards' model with the Weibull shared frailty model, the same set of variables was found to be significant to the hazards of the composite outcome, death and cardiac tamponade. On the other hand, in the event of constriction, a tachycardia (HR) and pulmonary infiltrate was found to be significant in the Weibull parametric PH model but was not significant in the Weibull shared frailty model. Also, peripheral oedema was found to be significant in the Weibull shared frailty model, but not significant in the Weibull parametric PH model. Table 4.5 gives the log-likelihood, AIC and BIC values of

the two models. From the Table we can see that the Weibull shared gamma frailty model has both a minimum AIC and BIC value, and larger log-likelihood, indicating that this model fits the data better than the Weibull PH model which did not take into account the clustering effect.

Table 4.1: Weibull proportional hazards models with and without frailty in the composite outcome.

	Without frailty		With frailty	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	0.95 (0.74 - 1.21)	0.664	0.93 (0.73 - 1.19)	0.585
Age (years)	1.01 (1.00 - 1.02)	0.029	1.01 (1.00 - 1.02)	0.032
Weight (kg)	0.90 (0.81 - 1.00)	0.042	0.90 (0.81 - 1.00)	0.039
NYHA Class				
I	1.00		1.00	
II	1.41 (0.91 - 2.19)	0.122	1.39 (0.89 - 2.16)	0.151
III	2.33 (1.46 - 3.74)	<0.001	2.27 (1.40 - 3.67)	0.001
IV	2.72 (1.60 - 4.61)	<0.001	2.26 (1.29 - 3.97)	0.004
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.67 (0.46 - 0.98)	0.040	0.67 (0.46 - 0.99)	0.043
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.86 (0.66 - 1.10)	0.231	0.90 (0.69 - 1.17)	0.427
Creatinine				
$\leq$ 105 umol/l	1.00		1.00	
$>$ 105 umol/l	1.71 (1.22 - 2.39)	0.002	1.72 (1.23 - 2.41)	0.002
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	0.86 (0.63 - 1.17)	0.332	0.86 (0.63 - 1.19)	0.376
Peripheral oedema				
No	1.00		1.00	
Yes	1.44 (1.12 - 1.85)	0.005	1.54 (1.17 - 2.04)	0.002
Pulmonary infiltrate on CXR				
No	1.00		1.00	
Yes	1.15 (0.90 - 1.49)	0.268	1.18 (0.91 - 1.53)	0.220
Effusion size				
small	1.00		1.00	
medium	1.71 (0.91 - 3.21)	0.097	1.62 (0.86 - 3.06)	0.138
large	1.59 (0.87 - 2.88)	0.131	1.50 (0.82 - 2.76)	0.186
θ (SE)			0.103 (0.084)	0.019

NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure; CXR: Chest Xray.
CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table 4.2: Weibull proportional hazards models with and without frailty in death outcome.

	Without frailty		With frailty	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	1.13 (0.87 - 1.48)	0.366	1.14 (0.87 - 1.48)	0.350
Weight (kg)	0.87 (0.78 - 0.98)	0.023	0.87 (0.78 - 0.98)	0.020
NYHA Class				
I	1.00		1.00	
II	1.24 (0.77 - 2.00)	0.372	1.20 (0.73 - 1.94)	0.466
III	1.86 (1.13 - 3.06)	0.015	1.77 (1.07 - 2.94)	0.027
IV	2.62 (1.51 - 4.54)	0.001	2.44 (1.38 - 4.31)	0.002
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.54 (0.37 - 0.87)	0.001	0.54 (0.37 - 0.79)	0.002
Creatinine				
$\leq$ 105 umol/l	1.00		1.00	
$>$ 105 umol/l	1.83 (1.29 - 2.58)	0.001	1.80 (1.27 - 2.04)	0.007
Peripheral oedema				
No	1.00		1.00	
Yes	1.41 (1.07 - 1.86)	0.016	1.51 (1.12 - 2.04)	0.007
θ (SE)			0.039 (0.043)	0.072

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table 4.3: Weibull proportional hazards models with and without frailty in the cardiac tamponade outcome.

	Without frailty		With frailty	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	0.56 (0.28 - 1.12)	0.102	0.56 (0.28 - 1.11)	0.099
NYHA Class				
I	1.00		1.00	
II	2.85 (0.65 - 12.50)	0.164	2.74 (0.62 - 12.12)	0.184
III	5.01 (1.13 - 22.25)	0.034	4.60 (1.01 - 20.88)	0.048
IV	6.24 (1.21 - 32.28)	0.026	5.15 (0.91 - 29.22)	0.065
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.37 (0.15 - 0.89)	0.026	0.35 (0.14 - 6.82)	0.023
Atrial fibrillation on ECG				
No	1.00		1.00	
Yes	2.35 (0.82 - 6.70)	0.110	2.38 (0.83 - 6.82)	0.106
θ (SE)			0.137 (0.230)	0.212

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table 4.4: Weibull proportional hazards models with and without frailty in the constriction outcome.

	Without frailty		With frailty	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	0.56 (0.35 - 0.92)	0.021	0.55 (0.34 - 0.90)	0.018
Age (years)	1.00 (0.98 - 1.02)	0.760	1.01 (0.99 - 1.03)	0.533
Gender				
Male	1.00		1.00	
Female	0.54 (0.32 - 0.91)	0.022	0.52 (0.30 - 0.90)	0.018
NYHA Class				
I	1.00		1.00	
II	1.08 (0.46 - 2.54)	0.862	1.29 (0.53 - 3.11)	0.572
III	2.37 (0.96 - 5.80)	0.060	3.09 (1.24 - 7.72)	0.016
IV	2.70 (0.97 - 7.51)	0.057	2.30 (0.78 - 6.74)	0.129
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	2.34 (0.57 - 9.69)	0.239	2.52 (0.59 - 10.64)	0.210
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.47 (0.27 - 0.81)	0.006	0.63 (0.36 - 1.11)	0.112
Haemoglobin				
$\leq$ 10 g/dl	1.00		1.00	
$>$ 10 g/dl	0.51 (0.29 - 0.89)	0.018	0.55 (0.31 - 0.97)	0.039
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	0.93 (0.52 - 1.68)	0.821	0.73 (0.39 - 1.37)	0.331
Peripheral oedema				
No	1.00		1.00	
Yes	1.59 (0.95 - 2.66)	0.079	1.76 (1.01 - 3.07)	0.045
Pulmonary infiltrate				
No	1.00		1.00	
Yes	1.57 (0.97 - 2.54)	0.065	1.48 (0.90 - 2.44)	0.122
HIV status				
Negative	1.00		1.00	
Positive	0.53 (0.31 - 0.91)	0.022	0.55 (0.32 - 0.97)	0.040
θ (SE)			1.030 (0.577)	< 0.001

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC:Akaike's Information Criterion

Table 4.5: Comparison of Weibull PH and Shared frailty model.

	Weibull PH analysis			Weibull Shared failty analysis		
Outcomes	$\ell\ell$ (model)	AIC	BIC	$\ell\ell$ (model)	AIC	BIC
Composite	-1053.135	2138.27	2218.391	-1050.98	2135.96	2221.089
Death	-902.6241	1825.248	1876.541	-901.5556	1825.111	1881.534
Cardiac	-215.6668	447.3336	487.0391	-215.3463	448.6927	493.3614
Constriction	-370.8102	771.6205	848.4997	-363.0353	758.0706	840.0751

$\ell\ell$: log-likelihood; AIC: Akaike's Information Criterion; BIC: Bayesian Information Criterion

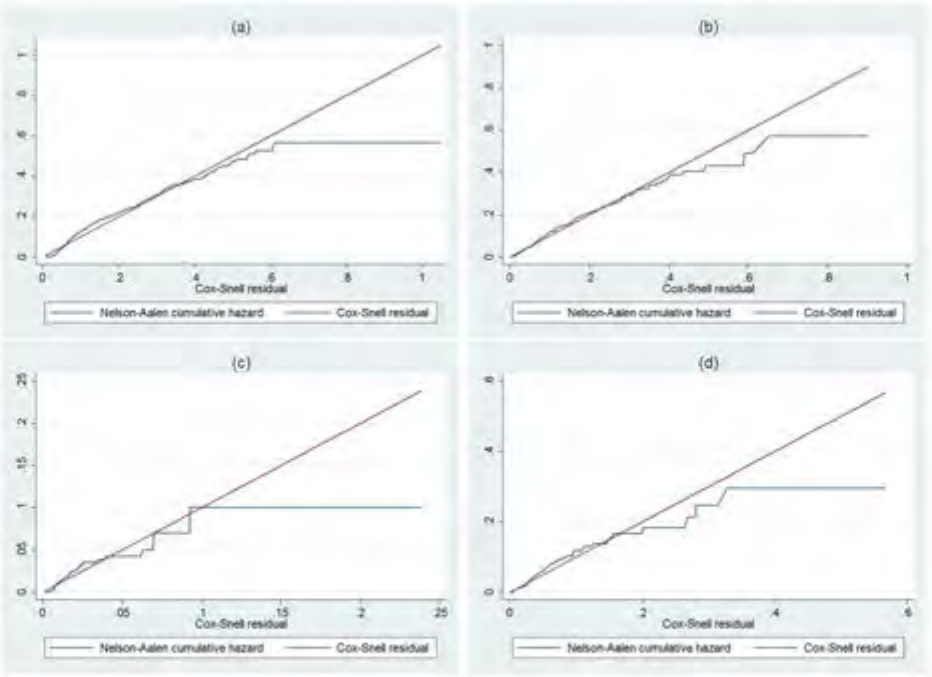

Figure 4.1: *Cox-Snell residual of Weibull proportional hazards model for (a) composite outcome (b) death (c) cardiac tamponade (d) constriction.*

Accelerated failure time model

In order to make the choice of which distribution to use, the identified significant variables in Section 3.5.2 were fitted using exponential, Weibull, log-logistic and log-normal as the underline distribution for the accelerated failure time model. The computed value of AIC for a lognormal AFT model seems to be the better choice as its AIC is less than the rest of the models (see Table 4.6).

The Q-Q plot is used to check the AFT assumption. The Q-Q plot in Figures 4.2 and 4.3 closely approximates a straight line from the origin which implies the AFT model may provide an appropriate model. Furthermore, we check the goodness of fit of the model using Cox-Snell residuals' plots. The cumulative hazard plot of the Cox-Snell residuals in a lognormal model is presented in Figure 4.4. The plotted line lies on a line that has a unit slope and zero intercept in the incidence of death, but there is a noticeable deviation from the straight line passing through the origin in the incidence of cardiac tamponade and constriction. As a result, there is a reason to doubt the suitability of this fitted lognormal model.

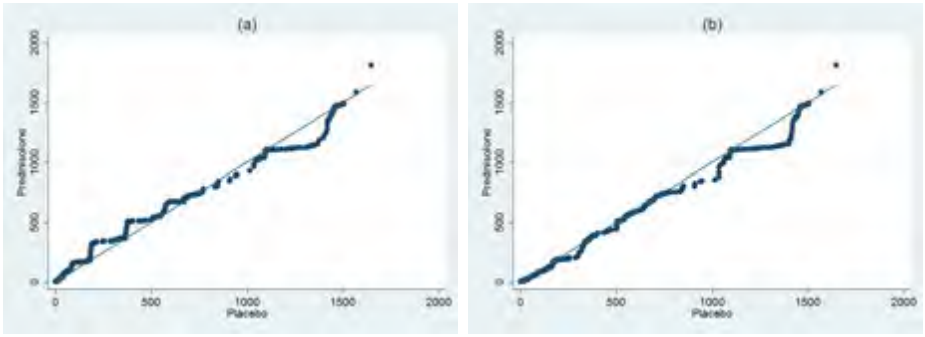

Figure 4.2: *Q-Q plot of AFT model for time to (a) composite outcome (b) death.*

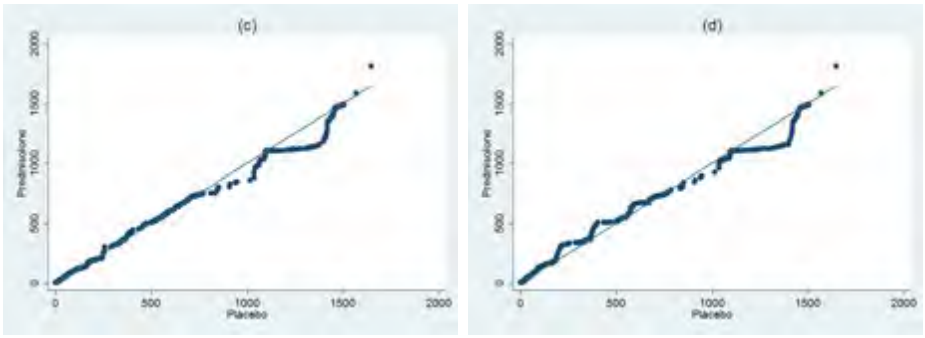

Figure 4.3: *Q-Q plot of AFT model for time to (c) cardiac tamponade (d) constriction.*

Table 4.6: Comparison of model fit statistics for AFT Models

Models	$\ell\ell$(null)	$\ell\ell$(model)	df	AIC
Composite				
Exponential	-2360.3	-2322.6	15	4675.13
Weibull	-2237.6	-2202.1	16	4436.177
Log-normal	**-2218.6**	**-2183.5**	**16**	**4399.029**
Log-logistic	-2232.2	-2195.4	16	4422.821
Death				
Exponential	-2017.1	-1986.6	9	3991.162
Weibull	-1954.9	-1925.4	10	3870.851
Log-normal	**-1945.9**	**-1916.8**	**10**	**3853.588**
Log-logistic	-1952.9	-1922.4	10	3864.849
Constriction				
Exponential	-379.1	-370.0	7	754.0975
Weibull	-346.0	-337.1	8	690.2262
Log-normal	**-343.5**	**-334.9**	**8**	**685.7354**
Log-logistic	-345.8	-336.9	8	689.7931
Cardiac tamponade				
Exponential	-732.9	-695.8	15	1421.686
Weibull	-683.5	-648.2	16	1328.412
Log-normal	**-677.8**	**-643.0**	**16**	**1318.08**
Log-logistic	-682.9	-647.1	16	1326.182

$\ell\ell$(nul): the log-likelihood for the constant-only model; $\ell\ell$(model): the log-likelihood for the model df:degree of freedom; df: degree of freedom; AIC:Akaike's Information Criterion

Table 4.7: Log-normal accelerated failure time model for the composite outcome.

Characteristics	TR (95%CI)	p-value
Treatment		
Placebo	1.00	
Prednisolone	1.15 (0.68 - 1.95)	0.596
Age (years)	0.98 (0.96 - 1.00)	0.021
Weight (kg)	1.21 (0.97 - 1.53)	0.096
NYHA Class		
I	1.00	
II	0.41 (0.17 - 0.95)	0.038
III	0.15 (0.06 - 0.39)	<0.001
IV	0.10 (0.03 - 0.33)	<0.001
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	1.99 (0.80 - 4.92)	0.139
Tachycardia (HR)		
$\leq$ 100 mmHg	1.00	
$>$ 100 mmHg	1.42 (0.81 - 2.48)	0.218
Creatinine		
$\leq$ 105 umol/l	1.00	
$>$ 105 umol/l	0.35 (0.16 - 0.77)	0.009
Palpable pulsus paradoxus		
No	1.00	
Yes	1.21 (0.61 - 2.43)	0.585
Peripheral oedema		
No	1.00	
Yes	0.46 (0.26 - 0.82)	0.008
Pulmonary infiltrate on CXR		
No	1.00	
Yes	0.80 (0.46 - 1.40)	0.440
Effusion size		
small	1.00	
medium	0.36 (0.10 - 1.28)	0.115
large	0.46 (0.14 - 1.53)	0.207

NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure; CXR: Chest Xray.

CI: Confidence Interval; TR: Time Ratio; AIC: Akaike's Information Criterion

Table 4.8: Log-normal accelerated failure time model for death outcome.

Characteristics	TR (95%CI)	p-value
Treatment		
Placebo	1.00	
Prednisolone	0.73 (0.43 - 1.23)	0.243
Weight (kg)	1.23 (0.98 - 1.54)	0.070
NYHA Class		
I	1.00	
II	0.55 (0.24 - 1.30)	0.175
III	0.27 (0.11 - 0.68)	0.006
IV	0.14 (0.05 - 0.41)	<0.001
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	3.35 (1.41 - 7.94)	0.006
Creatinine		
$\leq$ 105 umol/l	1.00	
$>$ 105 umol/l	0.30 (0.14 - 0.63)	0.002
Peripheral oedema		
No	1.00	
Yes	0.51 (0.29 - 0.87)	0.015

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; TR: Time Ratio; AIC: Akaike's Information Criterion

Table 4.9: Log-normal accelerated failure time model for cardiac tamponade outcome.

Characteristics	TR (95%CI)	p-value
Treatment		
Placebo	1.00	
Prednisolone	5.93 (0.81 - 43.47)	0.008
NYHA Class		
I	1.00	
II	0.07 (0.002 - 2.65)	0.152
III	0.01 (0.0003 - 0.63)	0.028
IV	0.01 (0.0001 - 0.65)	0.032
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	14.73 (0.77 - 284.32)	0.074
Atrial fibrillation on ECG		
No	1.00	
Yes	0.09 (0.003 - 2.77)	0.170

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; TR: Time Ratio; AIC: Akaike's Information Criterion

Table 4.10: Log-normal accelerated failure time model for the constriction outcome.

Characteristics	TR (95%CI)	p-value
Treatment		
Placebo	1.00	
Prednisolone	4.11 (1.28 - 13.23)	0.018
Age (years)	1.00 (0.95 - 1.04)	0.727
Gender		
Male	1.00	
Female	4.39 (1.26 - 15.36)	0.021
NYHA Class		
I	1.00	
II	0.61 (0.10 - 3.82)	0.598
III	0.12 (0.02 - 0.94)	0.044
IV	0.10 (0.01 - 1.12)	0.061
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	0.10 (0.004 - 2.43)	0.158
Tachycardia (HR)		
$\leq$ 100 mmHg	1.00	
$>$ 100 mmHg	5.51 (1.46 - 18.17)	0.011
Haemoglobin		
$\leq$ 10 g/dl	1.00	
$>$ 10 g/dl	3.64 (1.00 - 13.23)	0.050
Palpable pulsus paradoxus		
No	1.00	
Yes	1.05 (0.24 - 4.54)	0.947
Peripheral oedema		
No	1.00	
Yes	0.36 (0.11 - 1.19)	0.094
Pulmonary infiltrate		
No	1.00	
Yes	0.44 (0.14 - 1.44)	0.176
HIV status		
Negative	1.00	
Positive	5.52 (1.50 - 20.36)	0.010

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; TR: Time Ratio; AIC:Akaike's Information Criterion, HIV: Human Immunodeficiency Virus.

In Tables 4.7 - 4.10, the lognormal AFT model suggests that in the composite outcome, a one-year increase in age and a one kilogram increase in weight will increase the survival time by 0.98 and 1.21, respectively. Also, patients in NYHA classes I, have creatinine >105 umol/l, presence of peripheral oedema are factors showing considerable association with survival time in the composite outcome. An increase in weight, NYHA classes III & IV, hypotension (SBP) >90 mmHg, creatine >105 umol/l, presence of peripheral oedema are factors showing considerable association with survival time in the outcome of death. Furthermore, the prednisolone group, NYHA classes III & IV, hypotension (SBP) >90 mmHg, are factors showing considerable association with survival time in the outcome of cardiac tamponade. Patients in the prednisolone group, female gender, NYHA III & IV, Tachycardia (HR) >100 mmHg, haemoglobin >10 g/dl, presence of peripheral oedema, HIV-negative patients are factors showing considerable associ-

ation with survival time in the outcome of constriction.

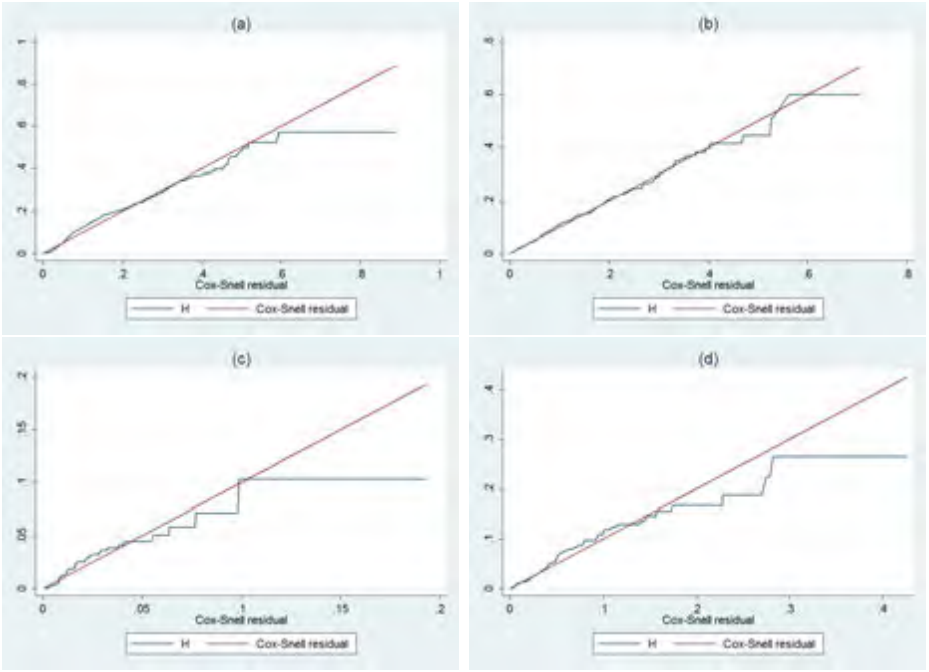

Figure 4.4: *Cox-Snell residual of AFT models for (a) composite outcome (b) death (c) cardiac tamponade (d) constriction.*

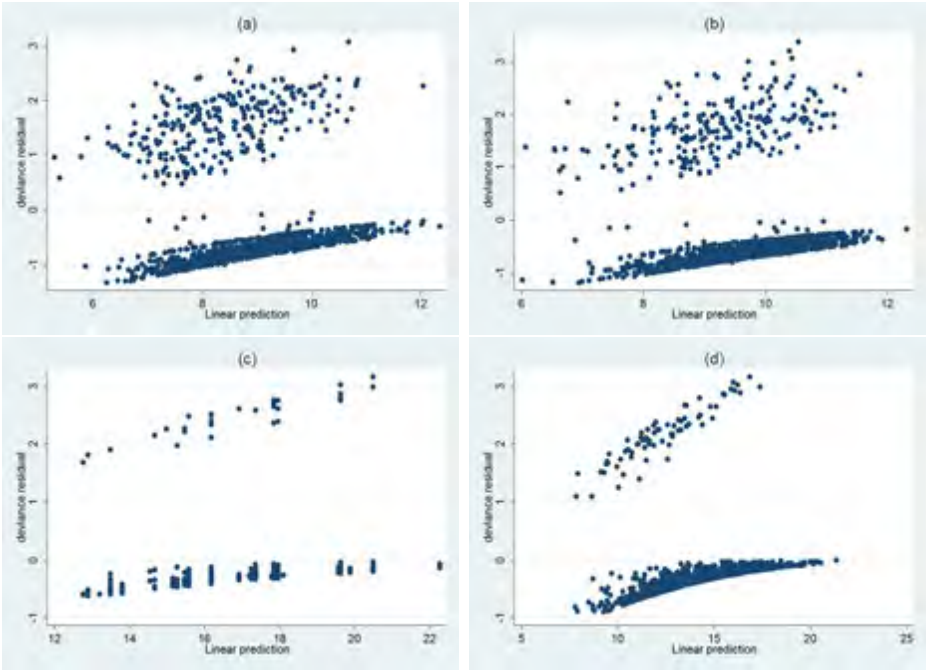

Figure 4.5: *Deviance residual of the AFT model for (a) composite outcome (b) death (c) cardiac tamponade (d) constriction.*

4.6 Chapter summary

In this chapter, we review the use of parametric models for the analysis of time-to-event data. These models are used instead of the Cox proportional hazards model when the proportionality assumption for the Cox PH model is not satisfied. In the analysis done in Chapter 3 of this thesis using a Cox proportional hazards model for IMPI clinical data, we found that the proportionality assumption was satisfied. However, we explore the use of the parametric model for completion purposes. Out of the three popular distributions available for the parametric proportional hazards modelling, the Weibull distribution was found to be a good fit because it has larger log-likelihood and minimum AIC and BIC. We found that the same sets of variables are significant in the hazards of the composite outcome, death and cardiac tamponade, but there is a variation in the significance of variables in the constriction outcome. In the event of constriction, tachycardia (HR) and pulmonary infiltrate were found to be significant in the Weibull parametric PH model but was not significant in the Weibull shared frailty model. Also, peripheral oedema was found to be significant in the Weibull shared frailty model, but not significant in the Weibull parametric PH model. Heterogeneity parameter θ was found to be significant in the composite outcome and constriction, but was not significant in the death and cardiac tamponade. Using log-likelihood and the AIC test shows that the Weibull shared frailty model fits the data better than the Weibull parametric proportional hazards model. In addition, we also analysed the dataset using the accelerated failure time model, and the lognormal distribution was found to be a better distribution for modelling the survival time of events in the dataset. Finally, we did model checking using Cox-Snell and deviance residuals.

Chapter 5

Competing risks framework

In this chapter, important functions in competing risks settings are discussed. The method of estimation for competing risks regression coefficients is presented. Regression models based on different approaches are also discussed. We then apply the cause-specific hazards model and sub-distribution hazards model to the IMPI trial data.

5.1 Background to competing risk

Competing risks data are common in medical research. They occur as a result of a subject being exposed to more than one cause of failure. The occurrence of any of the risks of failure prevents the occurrence of other competing risks or events (Gooley et al., 1999). For example, patients in a breast cancer study may fail as a result of heart disease or an accident. In this example, heart disease or an accident are regarded as competing risks because, a patients that failed as a result of heart disease cannot experience breast cancer. Furthermore, when there are composite outcomes, it is common to see that an event of interest is competing risk censored by other events. A composite outcome is as a result of combining several events of interest within a single outcome. This is done when the required sample size is too large to be realized. By combining several endpoints of interest into a composite, the intent is to increase the effect size, thereby reducing the required sample size (Freemantle et al., 2003). Often, this composite endpoint is the time to the first of several events (Wolbers et al., 2014). In the IMPI clinical trial dataset considered in this dissertation, the event of interest is the composite outcome of death, constriction or cardiac tamponade. If we consider constriction as a primary event of interest, it may be dependently censored by another competing event such as death or cardiac tamponade. Also, if we choose cardiac tamponade to be the primary event of interest, constriction or death

can compete against it. Therefore, caution is needed in estimating the probability of the event of interest occurring in the presence of so-called competing risks.

5.2 The naive Kaplan-Meier estimator

According to medical literature, often the Kaplan-Meier estimator is used in the analysis of competing risks whereby it views competing events as censored observations. This procedure, which is sometimes called the 'naive Kaplan-Meier estimator' in literature (Putter et al., 2007), violates one important assumption of the standard Kaplan-Meier estimator, that is the assumption of independence of the censoring distribution, i.e. the distribution of the time to the competing events. If the competing event time distributions were independent of the distribution of time to the event of interest, this would imply that at each point in time, the hazard of the event of interest is the same for subjects that have not yet failed and are still under follow-up, as for subjects that have experienced a competing event by that time. However, a subject that is censored because of failure from a competing risk may not experience the event of interest. Since subjects that will never fail are treated as if they could fail (they are censored), the naive Kaplan-Meier overestimates the probability of failure and hence underestimates the corresponding survival probability.

Using the naive Kaplan-Meier estimator, the probability for an event of type k up to a given time t is estimated by

$$1 - S_k(t) = \exp(\Lambda_k(t)), \tag{5.2.1}$$

where $S_k(t)$ can be estimated from the observed data, but cannot be interpreted as a marginal survival probability, $\Lambda_k(t)$ is the cumulative cause-specific hazard. The cumulative incidence function (CI) is an unbiased estimator when competing risks are present.

5.3 Important functions in the competing risks setting

There are two basic functions used in analyzing competing risk data; the cause-specific ($h_k(t)$) and the cumulative incidence function ($F_k(t)$).

The cause-specific function for event of type k is the instantaneous risk of dying from a specific cause k given that the subject is still alive at time t while the cumulative incidence function for event k, $F_k(t) = P(T \leq t, K = k)$, can be interpreted as the probability of patients

that died from cause k at time t, recognizing the fact that the patient can die from other causes.

Let $T = min(T_1, \ldots, T_K)$ be the time to the first event, and K be the total number of event types. The cause-specific hazard rate for the k^{th} event type can be defined as

$$h_k(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t, K = k | T \geq t)}{\Delta t} \tag{5.3.1}$$

$$= \frac{\frac{d}{dt} F_k(t)}{S(t)}. \tag{5.3.2}$$

Integrating both sides of equation 5.3.2, the cumulative incidence function for the k^{th} event is defined as

$$F_k(t) = \int_0^t h_k(u) S(u) du, \qquad k = 1, \ldots, K$$

where $S(t) = \Pr(T \geq t)$ is the overall survival function and $H_k(t) = \int_0^t h_k(u) du$ is referred to as the cumulative hazard function. As $t \to \infty$, $F_k(t) \to \Pr(K = k) < 1$, hence $F_k(t)$ is not a proper distribution function. It is thus referred to as a sub-distribution function.

The sub-distribution function was introduced by Gray (1988) in order to define a 'hazard-type' quantity that is directly linked to the cumulative incidence function in the presence of competing risks. The link between the cumulative incidence function and the sub-distribution hazard is known from standard survival analysis as

$$S(t) = 1 - F(t) = \exp(-\Lambda(t)), \tag{5.3.3}$$

where $S(t)$ is the survival function, $\Lambda(t)$ is the cumulative hazard rate, and $F(t)$ is the cumulative density function.

For the sub-distribution hazard rate of an event type k at time t, individuals that failed from an event other than k prior to t remain in the risk set. The sub-distribution hazard rate for the k^{th} event type can be defined as

$$h_k(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, K = k | T \geq t \cup \{T < t, K \neq k\})}{\Delta t}. \tag{5.3.4}$$

The condition in the curled brackets are subjects that experience a competing event which are not removed from the risk set, though they are not really at risk at that time. Interpretation of the sub-distribution hazard may be seen as problematic, because individuals who experienced a competing event are not necessarily at risk of experiencing an event of interest.

5.4 Regression models for the competing risks setting

In the presence of competing risks, the most commonly used approaches are the cause-specific hazards regression proposed by Prentice et al. (1978) and the sub-distribution hazards regression introduced by Fine and Gray (1999). A computational technique using pseudo-observations, which was introduced by Andersen et al. (2002) and Klein and Andersen (2005) and binomial approach suggested by Scheike and Zhang (2008) are other regression approaches in a competing risks setting, all of which are directly based on cumulative incidence function (Dianatkhah et al., 2014). Also, the multinonial logistic regression (MNL) model provides an alternative approach to estimating a competing risks model. It treats the dependent variable as polytomous qualitative choice variable (Clapp et al., 2006). In ordinary multinomial logistic regression, there is one equation for predicting each outcome but, in the competing risk set up, there is a different equation for each outcome in each time point (Jenkins, 2005). A summary of cause-specific and sub-distribution hazards approaches with an application on IMPI clinical dataset are presented in this section.

5.4.1 Cause-specific hazard regression model

The regression model for cause-specific hazard (CSH) is based on the Cox proportional hazards model (David Cox, 1972). The CSH for event type k is a multiplicative function of the baseline hazard $\lambda_{k;0}(t)$, given a single covariate x:

$$\lambda_k(t|\boldsymbol{x}) = \lambda_{k;0}(t)\exp(\boldsymbol{\beta}_k'\boldsymbol{x}), \tag{5.4.1}$$

where $\lambda_{k;0}(t)$ describes the cause-specific baseline hazard for event type k, $\boldsymbol{x}$ is the p-dimensional vector of covariates and $\boldsymbol{\beta}_k$ is the vector of regression coefficients of length p for the kth type of event.

The regression coefficients of the cause-specific hazard model can be obtained by maximizing the partial likelihood function given by

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{n} \Big\{ \frac{\exp(\beta x_j)}{\sum_{i \in R_j} \exp(\beta x_i)} \Big\}^{\delta_j}. \tag{5.4.2}$$

The risk set is: $R_j = \{i : t_i \geq t_j\}$, which includes any individual that has not failed from any event and is under observation. δ_i is a censoring indicator returning the value of zero for a censored subject and a value of one, if any event was noticed. It is an adaptation of the likelihood function used in standard survival analysis (equation 3.2.3). Hence, an estimation of regression coefficients can be carried out numerically using the Newton-Raphson method.

5.4.2 Sub-distribution hazards regression model

Fine and Gray (1999) developed another regression model for time-to-event data in the presence of competing risks. They proposed a Cox-type regression model for the sub-distribution hazard for an event of interest, in which the effect of covariates can be estimated directly with reference to the cumulative incidence function. The model is defined as,

$$\lambda_1^*(t|\boldsymbol{x}) = \lambda_{1;0}^*(t) \exp(\boldsymbol{\beta}_1' \boldsymbol{x}), \tag{5.4.3}$$

where $\lambda_1^*(t|\boldsymbol{x})$ denotes the sub-distribution hazard for the event of interest depending on the vector of covariates $\boldsymbol{x}$, $\lambda_{1;0}^*(t)$ is the baseline sub-distribution hazard for an individual with all covariates equaling zero, and $\boldsymbol{\beta}_1$ is the vector of regression coefficients.

The exponentiated regression coefficients, e^{β}, can be obtained from the ratio of two sub-distribution hazards and are thus also termed sub-distribution hazard ratios, or sub-hazard ratios. These can be interpreted in a similar way to the hazard ratios in a Cox's regression model.

The regression coefficients of the sub-distribution hazard model can be obtained by maximizing the partial likelihood function given by

$$L^*(\boldsymbol{\beta}) = \prod_{j=1}^{r} \Big\{ \frac{\exp(\beta x_j)}{\sum_{i \in R_j^*} \ w_{ij} \ \exp(\beta x_i)} \Big\}^{\delta_i}. \tag{5.4.4}$$

The risk set is: $R_j^* = \{i : (t_i \geq t_j) \cup (t_i \leq t_j)\}$, represent an individual who has not failed from the cause of interest by time t is at risk. This includes two distinct groups: those who have not failed from any cause and those who have previously failed from another cause. For

$R_j^* = \{i : (t_i \leq t_j)\}$, $w_{ij} \leq 1$: the further t_i is from t_j the smaller the weight (Pintile, M. (2006)). These weights are given by the formula: $w_{ij} = \hat{G}(t_j)/\hat{G}(min(t_j, t_i))$, where $\hat{G}(.)$ is the survivor function for the censoring distribution.

The procedure for estimation of the regression parameters in model (5.4.3) for the sub-distribution is identical to that for the cause-specific proportional hazards model, the only difference is the definitions of the risk sets (Fine and Gray, 1999).

5.4.3 Regression models based on pseudo-value approach

A pseudo-value approach is a direct method of modelling of the effects of covariates on the cumulative incidence function. It was proposed by Anderson et al (2003) as a technique for modelling state probabilities in multi-state models using pseudo-observations. This technique was adjusted for the competing risks setting as demonstrated by Klein and Andersen (2005) since, a competing risks model can be interpreted as a special case of a multi-state model. Unlike the Fine and Gray model, this approach does not need to establish a proportionality assumption. The main idea of the approach is to obtain quantities that allow the application of standard methods for data analysis without consideration of censored subjects.

Estimation method

In general, the pseudo-value approach can be considered to estimate effects of covariates on any function of events $g(T)$, if an unbiased estimator $\hat{\theta}$ exists for

$$\theta = E(g(T)). \tag{5.4.5}$$

For the pseudo-value model approach, a grid of time points $\tau_i, \ldots, \tau_M$ is selected. The pseudo-observations can be estimated for one fixed time point τ_0 or for a pre-specified number of time points $\tau_i, \ldots, \tau_M$. If multiple time points are considered, a $nXH-$matrix of pseudo observations is needed. At each grid point, the estimated cumulative incidence function is computed based on the complete data set $\hat{G}(\tau_h)$ and the estimated cumulative incidence function based on the sample size $n-1$ obtained by deleting the i^{th} observation $\hat{G}^{(i)}(\tau_h)$ then the pseudo-value for the i^{th} subject at time τ_h is defined as:

$$\hat{\theta}_{ih} = n\hat{G}(\tau_h) - (n-1)\hat{G}^{(i)}(\tau_h), \qquad i = 1, \ldots, n \quad h = 1, \ldots, M \tag{5.4.6}$$

these are the pseudo-values known from jack-knife techniques. $n\hat{G}(t)$ is the number of events of type of interest occurring prior to t. When there is no censoring, $n\hat{G}(t)$ is the number of

events of type 1 occurring prior to t. In this case, $\hat{\theta}_i = (\hat{\theta}_{ih}, h = 1, \ldots, M) = (I(T_i \leq \tau_1, \epsilon_i = 1), \ldots, I(T_i \leq T_M, \epsilon_i = 1))$ and the $\hat{\theta}_i$'s are independent. When there is censoring, the pseudo-values are close to the indicators and are approximately independent. Thus, this allows to make use of results from generalized linear models to model the effects of covariates. For regression purposes, $\hat{\theta}_{ih}$, pseudo-observations can be used as a dependent variable Klein and Andersen (2005) in a generalized linear model

$$g(\theta_{ih}|\boldsymbol{X}_1) = \alpha_h + \boldsymbol{\beta}'\boldsymbol{X}_i, \qquad i = 1, \ldots, n, \ h = 1, \ldots, M. \tag{5.4.7}$$

where $g(.)$ is a link function as the logit or the complementary log-log function and $\boldsymbol{X}_i$ is the vector of covariates of subject i. The complementary log-log function when applied to a survival function gives a proportional hazard representation. A generalized linear model was assumed with

$$f(\theta_{ih}) = \alpha_h + \gamma'\boldsymbol{Z}_i = \boldsymbol{\beta}'\boldsymbol{Z}_{ih}, \qquad i = 1, \ldots, n \ h = 1, \ldots, M. \tag{5.4.8}$$

The inverse link is define by

$$\theta_{ih} = f^{-1}(\beta' Z_{ih}) = \mu(\beta' Z_{ih}). \tag{5.4.9}$$

In the case of multiple time points the generalized estimation equation approach (GEE, Liang and Zeger, 1986) was proposed for estimation and inference to account for repeated measures on the same subjects in order to obtain robust and valid standard errors under independent censoring.

5.5 Application

So far we have analysed IMPI trial data using the standard Cox model and parametric models by considering the composite event of death, cardiac tamponade or constriction as a single event of interest and death, cardiac tamponade and constriction individually as the event of interest. However, there is a need to examine the contribution of competing risks analyses in the composite outcome setting. In this section, we considered death as a competing risk to the event of cardiac tamponade and constriction since patient can die before experiencing these events. We used two main regression approaches in competing risks' settings, because, a deeper understanding of competing risks data can be gained by performing both regression on the cumulative incidence function (CIF) and regression on the cause-specific hazard functions. This approach has been recommended by some authors, among them is (Latouche et al.,

2013). Only the result for the effect of the prednisolone treatment with alternative risk factors is presented in this section and other application sections. The results for the *M. indicus pranii* treatment can be found in the appendix. The effects of *M. indicus pranii* and prednisolone treatment were estimated separately because there was no significant interaction between them (Mayosi et al., 2014).

5.5.1 Cumulative incidence function

The primary interest in describing competing risks data is often to estimate the absolute risk of the occurrence of an event of interest up to a follow-up time point t. Observation from Figure 5.1 of which Figure 5.2 is it's reduced form shows that the probability of cardiac tamponade within 100 days is roughly 10% in the prednisolone group and near 11% in the placebo group. Both probabilities take into account the possibility that death could occur instead. Also, taking into account the possibility of death, the probability of constriction within 100 days is about 12% in the prednisolone group and roughly 15% in the placebo group. Using gray's test, the effect of prednisolone on the hazard of cardiac tamponade when death is a competing event is not statistically significant (p-value = 0.159) but was significant on the hazard of constriction when death is a competing event (p-value = 0.0089). These results are similar to when competing event were ignored (see Figures 3.2).

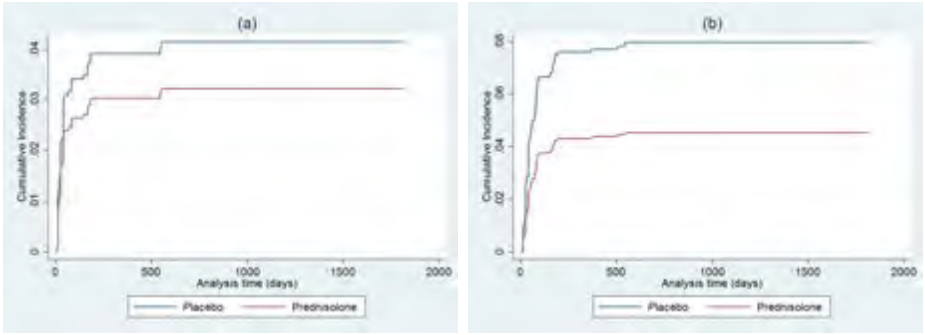

Figure 5.1: *Estimated cumulative incidence function for prednisolone and placebo group from the beginning of study until the end of the study. (a) constriction, (b) cardiac tamponade*

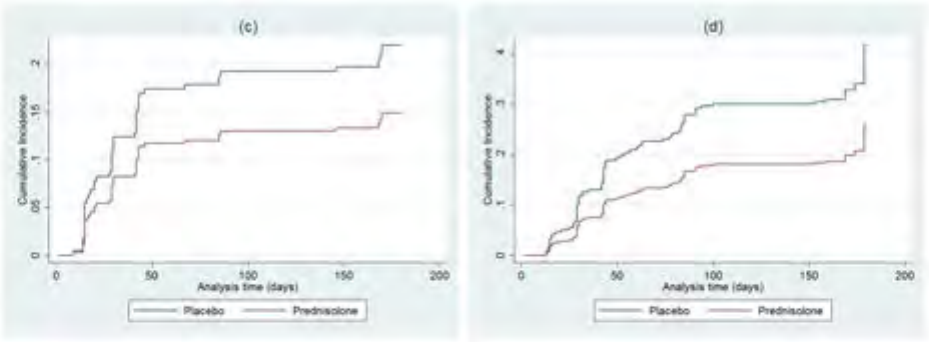

Figure 5.2: *Estimated cumulative incidence function for prednisolone and placebo group from the beginning of study until 6 months. (c) constriction, (d) cardiac tamponade*

5.5.2 Competing risks models

The results of multivariate regression models for the cause-specific hazards and the CIFs of the IMPI clinical trial are displayed in Tables 5.1 and 5.2. The significant risk factors detected in Section 3 were also used in this section. In comparing the prednisolone group and the placebo group, a sub-distribution hazard ratio for cardiac tamponade of 0.56 (95 % CI 0.28 - 1.11) was discovered, suggesting a much lower incidence of cardiac tamponade events for patients randomised to the predinisolone group compared to patients in the lacebo group. The effect of cardiac tamponade was lower in patients having hypotension (SBP) of $>$ 90mmHg, 0.40 (95 %CI 0.17- 0.94).

The prednisolone group revealed a sub-distribution hazard ratio for constriction of 0.56 (95 % CI 0.34 - 0.90), indicating a much lower incidence of constriction events for patients randomised to the predinisolone group compared to patients in the placebo group. The effect of constriction was lower in female patients compared to male patients 0.59 (95 %CI 0.34 - 1.00). The effect of constriction was lower in patients that have tachycardia (HR) >100mmHg and haemoglobin >10g/dl. The presence of pulmonary infiltrate in patients was significant to a constriction event. Also, a comparison between the HIV-positive patient and the HIV-negative patient revealed a sub-distribution hazard ratio for constriction of 0.52 (95 % CI 0.30 - 0.89), indicating a much lower incidence of constriction for patients that are HIV-positive to patients that are HIV-negative.

The effect estimates of covariates on the cause-specific hazard function and the sub-distribution hazard function are identical. Also, we observed that the estimates of the risk factors from the standard Cox PH models presented in Section three of this thesis are the same as the estimates

obtained when from the cause-specific hazard models.

Table 5.1: Multivariate regression models on the cause-specific hazards function and sub-distribution hazards function on the relative hazard of cardiac tamponade when death is a competing event.

	Cause-specific hazards analysis		Sub-distribution hazards analysis	
Characteristics	HR (95% CI)	p-value	SHR (95% CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	0.57 (0.28 - 1.12)	0.104	0.56 (0.28 - 1.11)	0.096
NYHA Class				
I	1.00		1.00	
II	2.87 (0.65 - 12.58)	0.162	2.79 (0.64 - 12.22)	0.172
III	5.11 (1.15 - 22.73)	0.032	4.88 (1.11 - 21.76)	0.038
IV	6.45 (1.25 - 30.40)	0.026	6.03 (1.17 - 31.01)	0.032
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.39 (0.16 - 0.94)	0.036	0.40 (0.17 - 0.94)	0.035
Atrial fibrillation on ECG				
No	1.00		1.00	
Yes	2.27 (0.79 - 6.46)	0.126	2.19 (0.76 - 6.33)	0.147

NYHA: New York Heart Association; SBP: Systolic Blood Pressure

CI: Confidence Interval; HR: Hazard Ratio; SHR: Subdistribution Hazard Ratio; AIC: Akaike's information criterion

Table 5.2: Multivariate regression models on the cause-specific hazards function and sub-distribution hazards function on the relative hazard of constriction when death is a competing event.

	Cause-specific hazards analysis		Sub-distribution hazards analysis	
Characteristics	HR (95% CI)	p-value	SHR (95% CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	0.57 (0.35 - 0.92)	0.022	0.56 (0.34 - 0.90)	0.017
Age (years)	1.00 (0.98 - 1.02)	0.898	1.00 (0.98 - 1.02)	0.919
Gender				
Male	1.00		1.00	
Female	0.57 (0.33 - 0.96)	0.036	0.59 (0.34 - 1.00)	0.048
NYHA Class				
I	1.00		1.00	
II	1.20 (0.51 - 2.81)	0.682	1.18 (0.50 - 2.78)	0.712
III	2.49 (1.02 - 6.10)	0.045	2.43 (0.99 - 5.95)	0.053
IV	3.03 (1.09 - 8.45)	0.034	2.70 (0.97 - 7.46)	0.056
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	2.61 (0.63 - 10.81)	0.186	2.73 (0.77 - 9.67)	0.120
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.49 (0.29 - 0.85)	0.011	0.51 (0.30 - 0.86)	0.012
Haemoglobin				
$\leq$ 10 g/dl	1.00		1.00	
$>$ 10 g/dl	0.52 (0.30 - 0.91)	0.021	0.53 (0.30 - 0.93)	0.026
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	0.97 (0.54 - 1.74)	0.919	1.00 (0.56 - 1.79)	0.996
Peripheral oedema				
No	1.00		1.00	
Yes	1.58 (0.94 - 2.67)	0.081	1.54 (0.91 - 2.61)	0.111
Pulmonary infiltrate				
No	1.00		1.00	
Yes	1.55 (0.96 - 2.50)	0.073	1.55 (0.95 - 2.52)	0.078
HIV status				
Negative	1.00		1.00	
Positive	0.52 (0.30 - 0.89)	0.017	0.52 (0.30 - 0.89)	0.019

NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure

CI: Confidence Interval; HR: Hazard Ratio; SHR: Subdistribution Hazard Ratio; AIC: Akaike's information criterion

Table 5.3: Multivariate regression on the cause-specific hazard function of the relative hazard of death.

	Death without cardiac tamponade		Death without constriction	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	1.15 (0.87 - 1.51)	0.326	1.22 (0.92 - 1.62)	0.163
Weight (kg)	0.88 (0.78 - 0.99)	0.039	0.86 (0.76 - 0.97)	0.012
NYHA Class				
I	1.00		1.00	
II	1.29 (0.79 - 2.10)	0.311	1.27 (0.77 - 2.10)	0.349
III	1.88 (1.12 - 3.14)	0.016	1.90 (1.13 - 3.22)	0.016
IV	2.68 (1.52 - 4.74)	0.001	2.63 (1.47 - 4.71)	<0.001
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.57 (0.38 - 0.85)	0.006	0.51 (0.34 - 0.75)	0.001
Creatinine				
$\leq$ 105 umol/l	1.00		1.00	
$>$ 105 umol/l	1.75 (1.21 - 2.51)	0.003	1.79 (1.25 - 2.58)	0.002
Peripheral oedema				
No	1.00		1.00	
Yes	1.35 (1.02 - 1.81)	0.039	1.43 (1.07 - 1.92)	0.017

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

The proportional sub-hazard assumption

The Fine and Gray model relies on the key assumption of the proportionality of sub-hazard. It is assumed that the sub-distribution hazard ratio does not depend on time. The alternative way to the proportionality hazards assumption is to allow the hazard ratio to vary over time. This can be done by introducing to competing risks models 5.4.1 & 5.4.3 variables of the form $z_i(t) = z_i g(t)$, which vary continuously with time. This method is similar to the one presented in Section 3.3 of this thesis.

The proportional hazard assumption for the Fine and Gray model was investigated by testing for time by covariate interaction in the model presented in equation (5.1). The proportionality assumption for the sub-hazard of cardiac tamponade was met for all the covariates, that is, there is no significance by treatment interaction, nor by NYHA class, Hypotension and Atrial fibrillation on ECG. On the contrary, in Table 5.2, the proportionality assumption for the sub-hazard of constriction was not met for pulmonary infiltrate (borderline significant p-value = 0.046).

5.6 Chapter summary

The analysis of time to event data in the presence of competing risks, i.e. when subjects can fail from one out of two or more mutually exclusive types of event are considered and discussed. We examined the effect of the risk factors using a cause-specific hazard model and sub-distribution hazards model. We believe that a complete understanding of the effect of a risk factor on competing risk endpoints requires modelling both cause-specific hazards and sub-distribution hazards side-by-side. The estimates of these risk factors from the standard multivariate Cox proportional hazards model are the same as the estimates obtained from the cause-specific hazards model. The effect of these risk factors on both cause-specific hazards and sub-distribution hazards models are found to be very similar.

Chapter 6

Clustered competing risk models

The main focus in this research is to analysis clustered competing risks with application to a multicentre clinical trial. This Chapter discusses an overview of the clustered competing risks concept, presents the available models used in the clustered competing risks settings, and the methods of estimation. The models are then applied to the IMPI dataset.

6.1 Overview of clustered competing risks

In the analysis of competing risks data, situation arises where competing risks data cannot be considered as independent because of a clustered design, such is the case in the registry cohorts or multicentre clinical trials studies. In the analysis of survival data with clustered design, frailty models have shown to be useful, where only one risk acts on the population. Studies have proven that ignoring heterogeneity in centres may alter the interpretation of treatment effects. Katsahian et al. (2006) have extended Fine and Gray's model to the case of clustered data, by including random centre effects or frailties in the sub-distribution hazard model. The model first allows to assess the heterogeneity across clusters, then incorporates such an effect when testing the effect of a covariate of interest. Katsahian et al. (2006) used the residual maximum likelihood approach for the estimation of parameters in their proposed model for the analysis of clustered competing risks data. In 2011, Katsahian and Boudreau proposed a penalized partial log-likelihood approach as an alternative estimation method to the one used by Katsahian et al. (2006). Their proposed method is an extension of Ripatti and Palmgren (2000) and Therneau et al. (2003) to the case of competing risks. In the analysis of survival data with clustered design, a penalized partial log-likelihood has been used by Ripatti and Palmgren (2000) and Therneau et al. (2003) to fit the Gaussian frailty proportional hazards models. Katsahian and Boudreau (2011) stated that the advantages of their proposed approach

are that maximization of a penalized partial log-likelihood function can be done using existing statistical software for Gaussian frailty models and that it is less sensitive to small changes in the data than the restricted maximum likelihood (REML) method of Katsahian et al. (2006). Another approach to handling clustered competing risks data is given by Christian et al. (2016) to infer the cause-specific hazard frailty model for clustered competing risks data using the hierarchical likelihood method. The hierarchical likelihood incorporates fixed effects as well as random effects into an extended likelihood function, so that the method does not require intensive numerical methods to find the marginal distribution like the EM algorithm.

6.2 Frailty model for the cause-specific hazard

Let T_i be the time to failure and ε be the cause of failure. The cause-specific hazard function for cause k at time t is defined by

$$h_k(t) = \lim_{\Delta t \to 0} \left\{ \frac{\Pr(t \leq T < t + \Delta t, \epsilon_i = k | T \geq t)}{\Delta t} \right\} \tag{6.2.1}$$

Let $Y_{ij} = \min(T_{ij}, C_{ij})$, where T_{ij} is the failure time and C_{ij} is the corresponding censoring time for the j^{th} subject in the i^{th} cluster. Also, let $\delta_{ij} = I(T_{ij} \leq C_{ij})\varepsilon_{ij}$ and v_i be an unobserved log-frailty associated with the i^{th} cluster. Following Do Ha et al. (2001), Lee et al. (2014) assume that given v_i, C_{ij} is independent of $(T_{ij}, \varepsilon_{ij})$ conditional on the covariates $x_{ij} (j = 1, \ldots, n_i)$ and that given x_{ij} and v_i, C_{ij} is independent and non-informative on v_i. Given v_i, the cause-specific proportional hazards model with a shared frailty conditional on covariates x_{ij} is given by

$$\lambda_k(t; \boldsymbol{x}_{ij}, \boldsymbol{v}_i) = \lambda_{0k}(t) \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}_k + \boldsymbol{v}_i), \tag{6.2.2}$$

where $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kp})'$ is a p x 1 regression parameter vector for cause k, and $\lambda_{0k}(t)$ is an unspecified baseline hazard function for cause k. Lee et al. (2014) also assumed that log-frailties $v_i \sim N(0, \theta)$ are independent and follow a distribution with frailty parameter θ.

6.2.1 Estimation procedure and inference

According to Lee and Nelder (1996) and Do Ha et al. (2001), the hierarchical log-likelihood for the cause-specific proportional hazards model with a shared frailty in equation (6.2.2) is defined by

$$h = h(\beta_1, \beta_2, \lambda_{01}, \lambda_{02}, v, \theta) = \sum_{ij} l_{1ij} + \sum_{i} l_{2i}, \tag{6.2.3}$$

where $l_{1ij} = l_{1ij}(\beta_1, \beta_2, \lambda_{01}, \lambda_{02}; Y_{ij}, \delta_{ij}|v_i)$ is a logarithm of the conditional density function for Y_{ij} and δ_{ij}. l_{2i} is the logarithm of the normal density function v_i with parameter θ given by

$$l_{2i} = l_{2i}(\theta; v_i) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\theta - \frac{1}{2\theta}v_i^2.$$

In order to estimate β and v, Lee et al. (2014) used the profile h-likelihood h^*, where λ_{01} and λ_{02} in equation (6.2.3) are eliminated

$$h^* = h|_{\lambda_{01}=\hat{\lambda}_{01}, \lambda_{02}=\hat{\lambda}_{02}} = \sum_{ij} l^*_{1ij} + \sum_i l_{2i}, \tag{6.2.4}$$

where

$$\sum_{ij} l^*_{1ij} = \sum_{ij} I(\epsilon_{ij} = 1)\eta_{ij} - \sum_k d_{0(k)} \log\Big\{ \sum_{(ij)\in R_{0(k)}} \exp(\eta_{ij})\Big\}, \tag{6.2.5}$$

$l^*_{1ij} = l_{1ij}(\beta, \lambda^s_{01}; t_{ij}, \epsilon_{ij}|v_i)\Big|_{\lambda^s_{01}=\hat{\lambda}^s_{01}}$ is the logarithm of the conditional density function for (T_{ij}, ϵ_{ij}) given v_i evaluated at $\hat{\lambda}^s_{01}$ which is the nonparametric maximum HL estimator of λ^s_{01} (Do Ha et al., 2014). $d_{0(k)}$ is the number of type 1 events at $t_{(k)}$.

Following Ha and Lee (2003), the outline of the estimation procedure for fitting the model 6.2.2 is as follows; given a frailty parameter θ, the maximum h-likelihood estimators $\hat{\beta}$ and $\hat{v}$ of β and v are obtained by solving:

$$\frac{\partial h^*}{\partial \beta} = 0 \quad \text{and} \quad \frac{\partial h^*}{\partial v} = 0 \tag{6.2.6}$$

The asymptotic covariance matrix of $\hat{\beta}$ and $\hat{v} - v$ is obtained from the inverse of the observed information matrix, $H^* = -\partial^2 h^*/\partial(\beta, v)^2$, for β and v based on h^*. Thus, the joint equations (6.2.6) are solved using the Newton-Raphson method with H^*. For the estimation of θ, Lee et al. (2014) used the adjusted profile h-likelood $p_{\beta,v}(h^*)$ given by

$$p_{\beta,v}(h^*) = \Big[h^* - \frac{1}{2}\text{logdet}\{H^*/(2\pi)\}\Big]\Big|_{\beta=\hat{\beta}, v=\hat{v}} \tag{6.2.7}$$

where $\hat{\beta} = \hat{\beta}(\theta)$ and $\hat{v} = \hat{v}(\theta)$. Note that $p_{\beta,v}(h^*)$ is a function of θ only because it has already eliminated β and v from h^*. The restricted maximum likelihood (REML) estimator $\hat{\theta}$ of θ is obtained by solving the estimating equation

$$\frac{\partial p_{\beta,v}(h^*)}{\partial \theta} = 0. \tag{6.2.8}$$

Note here that we allow for the $\partial\hat{v}/\partial\theta$ term when implementing equation (6.2.8). Also, it can be shown that (6.2.8) provides a simple form of the REML estimator, given by

$$\hat{\theta} = \frac{\sum_i \hat{v}_i^2}{q - \xi} \tag{6.2.9}$$

where $\xi = -\theta \text{tr}\{\hat{H}^{-1}(\partial \hat{H}/\partial \theta)\}$ and $\hat{H} = H^*|_{\beta=\hat{\beta}, v=\hat{v}}$. In summary, the estimates of (β, v) and θ are obtained by alternating between equations (6.2.6) and (6.2.9) until convergence is achieved. After convergence, we directly compute the estimates of $\text{var}(\hat{\beta})$ using the inverse of H^*.

6.3 Frailty model for the sub-distribution hazard

Let T_i be the failure times for subject i and C_i be its corresponding right-censoring times for $i = 1, \ldots, n$. Also, let ε_i be the cause of failure for the ith subject, then $\varepsilon = 0$ if $T_i > C_i$. The cumulative incidence function of failure from cause 1 is defined as

$$F_1(t) = Pr(T_i \leq t, \varepsilon_i = 1) \tag{6.3.1}$$

and the corresponding sub-distribution hazard is given as

$$h_1(t) = \lim_{\Delta t \to 0} \left\{ \frac{P(t \leq T_i < t + \Delta t, \varepsilon_i = 1 | T_i \geq t \cup (T_i \leq t \cap \varepsilon_i \neq 1))}{\Delta t} \right\} \tag{6.3.2}$$

note from Section 4.2, that the risk set associated with $h_1(t)$ is unusual and differs from the traditional risk set at time t.

Katsahian and Boudreau (2011) extended the proposed model for the sub-distribution hazard by Fine and Gray (1999) to handle clustered data by introducing Gaussian frailties. The model is given by

$$\lambda_1^*(t; \boldsymbol{x}, \boldsymbol{u}) = \lambda_{10}^*(t) \exp(\boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{u}) \tag{6.3.3}$$

where $\lambda_{10}(t)$ is an unspecified non-negative function time, $\boldsymbol{\beta}$ is a p x 1 vector of unknown parameters, and u is a K x 1 vector of frailties. The frailties $u_1, \ldots, u_K$ are assumed to be Gaussian with mean 0 and unknown variance θ.

6.3.1 Estimation procedure and inference

This section presents the two methods of estimation proposed by (Katsahian et al., 2006) and (Katsahian and Boudreau, 2011) for the sub-distribution hazard model with frailty. The first method is the residual maximum likelihood approach and the second method is the penalized

partial log-likelihood approach.

Residual maximum likelihood approach

This approach involves two main procedures. The first is to find the best linear predictors (BLUP) of the fixed and random components. The second is to use the results to find the restricted maximum likelihood (REML) estimators. BLUP can be obtained by maximizing the sum of the two components of l_1 and l_2. l_1 is the partial log-likelihod of failure times when u_k's are fixed and l_2 is the log-likelihood of the Gaussian random effects, respectively. For the uncensored observation, l_1 can be written as

$$l_1 = l(\beta; u_k) = \sum_{i=1}^{N} I(\varepsilon_i = 1)\Big\{\beta x_i' + z_i' u - \log\Big[\sum_{j \in R_i} \exp(\beta x_j' + z_j' u)\Big]\Big\}, \tag{6.3.4}$$

where $R_i = \{j : (T_i \leq T_j) \cup (T_i > T_j \cap \varepsilon_j \neq 1)\}$ is the risk set at time T_i. It is comprised of individuals who have not failed from any cause by T_i, but also those who have previously failed from competing causes. l_2 is given as

$$l_2(\theta; u) = -\frac{1}{2}\Big[K \log(2\pi\theta) + \sum_{i=1}^{k} \frac{u_k^2}{\theta}\Big]. \tag{6.3.5}$$

Using the Newton-Raphson procedure, we can obtain BLUP estimators of β and u_k.

In order to carry out the second procedure, only l_1 was modified using an inverse probability of censoring weighting, as originally proposed by Fine and Gray.

$$l_1 = l(\beta; u_k) = \sum_{i=1}^{N} I(\varepsilon_j = 1)\Big\{\beta x_i' + z_i' u - \log\Big[\sum_{j \in R_i} w_j(t) \exp(\beta x_j' + z_j' u)\Big]\Big\}, \tag{6.3.6}$$

where $w_j(t) = I(t_j \geq t \cup \varepsilon_j > 1)\hat{G}(t)/\hat{G}(t_j \wedge t)$ and $\hat{G}$ is the Kaplan-Meier of the survival function of censoring time. $w_j(t) = 1$ if individuals have not failed, but equal to zero if they have failed from cause 1 or have been right censored, and below 1 and decreasing overtime if they failed from another cause.

Penalized partial log-likelihood approach

By extending the penalized partial likelihood (PPL) approach of Gaussian frailty models introduced by Ripatti and Palmgren (2000) and Therneau et al. (2003) to allow for competing risks, penalized partial log-likelihood (PPLL) can be obtained by summing the partial log-likelihood

in equation 6.3.4 which is the conditional partial log-likelihood given the frailties and the log-likelihood function of K i.i.d, $N(0, \theta)$ random variable which corresponds to the distribution of the frailties (equation 6.3.5).

$$l_{ppll}(\beta, \theta, u) = l(\beta; u_k) + l_2(\theta; u). \tag{6.3.7}$$

The main interest is in estimating β and θ however, u is a vector of nuisance parameter which needs to be intergrated out. Now, the integrated penalized partial likelihood of (β, θ) which is given by

$$L_{INT}(\beta, \theta) = \int_R \dots \int_R \exp(l_1 = l(\beta; u_k) + l_2(\theta; u))du, \tag{6.3.8}$$

must be computed. With normally distributed frailties, the above K-dimensional integral becomes very difficult to solve. Following the methods of (Ripatti and Palmgren, 2000) and (Therneau et al., 2003) and (Breslow and Clayton, 1993), the Laplace approximation can be use to solve the integral in the equation (6.3.8).

6.4 Application

Tables 6.1 - 6.4 reports the estimated multivariate risk value of the risk factor analysis on the patient as well as the centre level, either ignoring or incorporating the random centre effect. The observed heterogeneity across centres in the rates of constriction is remarkably large (θ = 0.95 with standard error (SE) = 0.49), in contrast to the heterogeneity in the rates of cardiac tamponade ($\theta = 0.16$ with SE = 0.18). Large heterogeneity means that event times are strongly correlated within centres. This was also illustrated by the distribution of estimated random effects, depicted in Figure 6.1. One might conclude that observed and unobserved centre-specific factors play a more substantial role in the constriction hazard rates compared to the competing event rates. Including covariates in the models led to only a small reduction in centres heterogeneity for the rate of cardiac tamponade of the competing outcomes, with θ falling to 0.08 but an increase in the rate of constriction, with θ rising to 0.99. Nevertheless, such a significant random effect only slightly changed the estimated effect of the covariates on cause-specific hazard and sub-distribution hazard of cardiac tamponade and constriction.

Table 6.1: Multivariate regression models for with and without frailty on the incidence of cardiac tamponade when death is a competing risk.

	Without frailty		With frailty	
Characteristics	HR (95% CI)	p-value	HR (95% CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	0.57 (0.28 - 1.12)	0.104	0.56 (0.28 - 1.12)	0.103
NYHA Class				
I	1.00		1.00	
II	2.87 (0.65 - 12.58)	0.162	2.86 (0.66 - 12.58)	0.158
III	5.11 (1.15 - 22.73)	0.032	4.97 (1.11 - 22.30)	0.036
IV	6.45 (1.25 - 30.40)	0.026	6.03 (1.09 - 33.33)	0.039
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.39 (0.16 - 0.94)	0.036	0.38 (0.16 - 0.93)	0.034
Atrial fibrillation on ECG				
No	1.00		1.00	
Yes	2.27 (0.79 - 6.46)	0.126	2.33 (0.83 - 6.53)	0.108
θ without covariates (SE)			0.16 (0.18)	
θ with covariates (SE)			0.08 (0.20)	

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; θ: Heterogeneity parameter; SE: standard error

Table 6.2: Multivariate regression models for with and without frailty on the incidence of cardiac tamponade when death is a competing risk.

	Without frailty		With frailty	
Characteristics	SHR (95% CI)	p-value	SHR (95% CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	0.56 (0.28 - 1.11)	0.096	0.57 (0.36 - 0.88)	0.011
NYHA Class				
I	1.00		1.00	
II	2.79 (0.64 - 12.22)	0.172	2.87 (0.85 - 9.73)	0.091
III	4.88 (1.11 - 21.76)	0.038	5.11 (1.35 - 19.39)	0.016
IV	6.03 (1.17 - 31.01)	0.032	6.45 (1.25 - 33.35)	0.026
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.40 (0.17 - 0.94)	0.035	0.39 (0.18 - 0.84)	0.016
Atrial fibrillation on ECG				
No	1.00		1.00	
Yes	2.19 (0.76 - 6.33)	0.147	2.27 (0.62 - 8.23)	0.210
θ with covariates (SE)			0.07 (0.16)	0.033

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; SHR: Subdistribution Hazard Ratio; θ: Heterogeneity parameter; SE: standard error

Table 6.3: Multivariate regression models for with and without frailty on the incidence of constriction when death is a competing risk.

	Without frailty		With frailty	
Characteristics	HR (95% CI)	p-value	HR (95% CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	0.57 (0.35 - 0.93)	0.025	0.56 (0.34 - 0.91)	0.020
Age (years)	1.00 (0.98 - 1.02)	1.000	1.00 (0.98 - 1.02)	0.685
Gender				
Male	1.00		1.00	
Female	0.57 (0.33 - 0.97)	0.039	0.54 (0.31 - 0.92)	0.023
NYHA Class				
I	1.00		1.00	
II	1.23 (0.52 - 2.90)	0.630	1.47 (0.61 - 3.55)	0.394
III	2.55 (1.04 - 6.22)	0.040	3.38 (1.35 - 8.47)	0.010
IV	3.13 (1.12 - 8.70)	0.029	2.61 (0.88 - 7.73)	0.084
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	2.45 (0.59 - 10.23)	0.218	2.76 (0.65 - 11.74)	0.169
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.52 (0.30 - 0.91)	0.021	0.66 (0.37 - 1.16)	0.145
Haemoglobin				
$\leq$ 10 g/dl	1.00		1.00	
$>$ 10 g/dl	0.52 (0.30 - 0.91)	0.022	0.54 (0.31 - 0.96)	0.035
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	0.91 (0.51 - 1.65)	0.768	0.77 (0.41 - 1.44)	0.411
Peripheral oedema				
No	1.00		1.00	
Yes	1.58 (0.94 - 2.67)	0.084	1.76 (1.01 - 3.06)	0.046
Pulmonary infiltrate				
No	1.00		1.00	
Yes	1.52 (0.94 - 2.46)	0.086	1.48 (0.90 - 2.43)	0.124
HIV status				
Negative	1.00		1.00	
Positive	0.50 (0.29 - 0.86)	0.013	0.56 (0.32 - 0.97)	0.040
θ without covariates (SE)			0.95 (0.49)	0.027
θ with covariates (SE)			0.99 (0.56)	0.038

NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; θ: Heterogeneity parameter; SE: standard error

Table 6.4: Multivariate regression models for with and without frailty on the incidence of constriction when death is a competing risk.

	Competing risk analysis		Clustered competing risk analysis	
Characteristics	SHR (95% CI)	p-value	SHR (95% CI)	p-value
Treatment				
Placebo	1.00		1.00	
Prednisolone	0.56 (0.35 - 0.91)	0.020	0.57 (0.40 - 0.80)	0.001
Age (years)	1.00 (0.98 - 1.02)	0.971	1.00 (0.99 - 1.02)	0.880
Gender				
Male	1.00		1.00	
Female	0.59 (0.34 - 1.00)	0.052	0.57 (0.37 - 0.87)	0.010
NYHA Class				
I	1.00		1.00	
II	1.21 (0.51 - 2.85)	0.662	1.20 (0.30 - 4.80)	0.800
III	2.47 (1.01 - 6.01)	0.047	2.50 (0.86 - 7.25)	0.093
IV	2.83 (1.03 - 7.76)	0.043	3.03 (0.81 - 11.43)	0.100
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	2.53 (0.72 - 8.94)	0.149	2.61 (0.66 - 10.35)	0.170
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.54 (0.32 - 0.92)	0.025	0.49 (0.30 - 0.81)	0.005
Haemoglobin				
$\leq$ 10 g/dl	1.00		1.00	
$>$ 10 g/dl	0.52 (0.30 - 0.92)	0.027	0.52 (0.35 - 0.78)	0.002
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	0.93 (0.52 - 1.67)	0.821	0.97 (0.62 - 1.53)	0.900
Peripheral oedema				
No	1.00		1.00	
Yes	1.55 (0.91 - 2.64)	0.108	1.59 (1.03 - 2.45)	0.037
Pulmonary infiltrate				
No	1.00		1.00	
Yes	1.52 (0.94 - 2.47)	0.087	1.55 (0.87 - 2.76)	0.140
HIV status				
Negative	1.00		1.00	
Positive	0.49 (0.28 - 0.86)	0.013	0.52 (0.34 - 0.79)	0.002
θ with covariates (SE)			0.99 (0.56)	0.035

NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure; CI: Confidence Interval; SHR: Subdistribution Hazard Ratio; θ: Heterogeneity parameter; SE: standard error

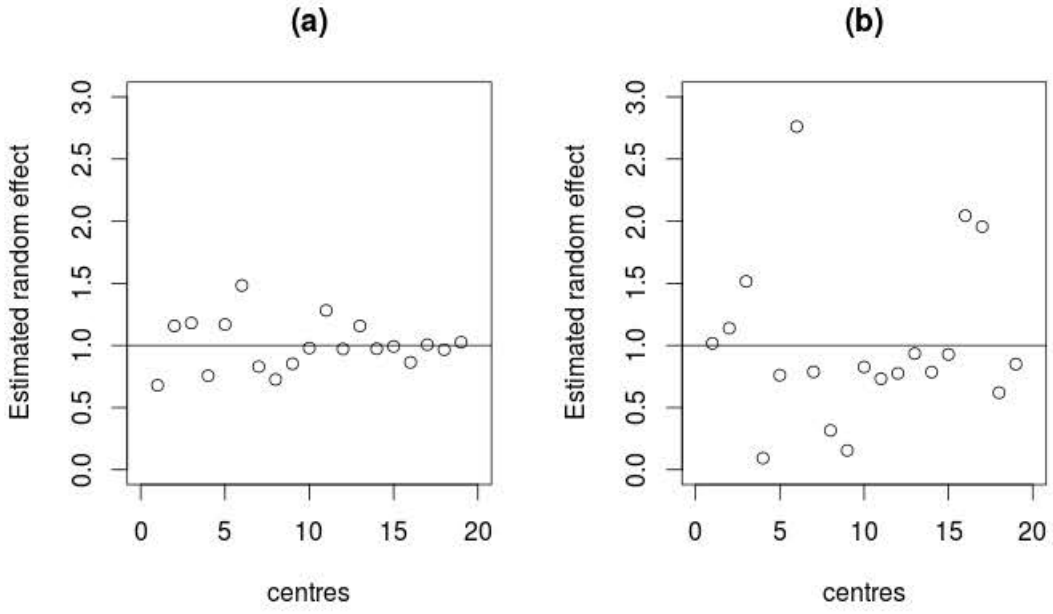

Figure 6.1: *Estimated random effect(s) (frailties) for each centre for the (a) cardiac tamponade, and (b) constriction. Data are from the null model without covariates. Each circle on the plot represent each centre from the eight African countries namely, South Africa, Mozambique, Malawi, Nigeria, Sierra Leone, Zimbabwe, Uganda, and Kenya.*

6.5 Chapter summary

Due to the complexity of the IMPI clinical trial dataset, we decided to examine the effect of the recognised risk factors on the outcomes of interest by taking into account the effect of the competing event and the centres effect. Cause-specific hazards and sub-distribution hazards models for competing risks have been extended to accommodate a random effect. This was done by Christian et al. (2016) and Katsahian et al. (2006) respectively. A remarkable heterogeneity across centres in the rates of constriction as compared to the cardiac tamponade was observed, but this only slightly changed the estimated effect of the covariates as compared to the Cox PH model and the Fine and Gray model.

Chapter 7

Summary, conclusion and further research

This chapter summarises and concludes this research work. Possible further research is also suggested in this chapter.

7.1 Summary and conclusion

This thesis analysed the IMPI clinical trial which was carried out in 19 centres in 8 African countries with the principal objective of assessing the effectiveness and safety of adjunctive prednisolone and *M. indicus pranii* immunotherapy in reducing the composite outcome of death, constriction or cardiac tamponade, requiring pericardial drainage in patients with probable or definite tuberculous pericarditis. During the follow-up period out of 1400 patients, 50 experienced cardiac tamponade, 85 experience constriction, 246 experienced death of which 18 experienced cardiac tamponade before they died and 23 experienced constriction before they died. 336 experienced composite outcome which can either be death, cardiac tamponade or constriction. Using Mann-Whitney and Chi-square test satistics, there were no statistically significant differences between the treatments in the baseline characteristics of the patients. This was important in ensuring there was no allocation bias influencing the effect of the treatment outcome.

In univariate and multivariate Cox proportional hazards models, we found that age, weight, New York Heart Association (NYHA) class, hypotension, creatinine, peripheral oedema show a statistically significant association with the composite outcome. Furthermore, weight, NYHA

class, hypotension, creatinine and peripheral oedema show a statistically significant association with death. In addition, NYHA class and hypotension show a statistically significant association with cardiac tamponade. Lastly, prednisolone, gender, NYHA class, tachycardia, haemoglobin level, peripheral oedema, pulmonary infiltrate and HIV status show a statistically significant association with constriction. In an article written by Mayosi et al. (2014) where IMPI clinical trial data was analysed, these other significant variables were not detected because they only examined the effect of the treatments groups on the event of interest. The proportionality assumption check using Schoenfeld residuals shows that the proportionality assumption was not violated for all the variables. Also, the graph of Cox-Snell residuals indicated the Cox PH model as a good fit for the data.

In order to understand the centre effect on the incidence of composite outcome (death, cardiac tamponade or constriction) and the individual outcome, we fitted a multivariate shared gamma frailty model. There is a significant centre effect on the incidence of constriction (p-value = $<$ 0.001). Furthermore, in making a comparison of shared gamma frailty models with the Cox PH models, the variables estimate and significance of variables in the event of cardiac tamponade and death, are similar but vary in the event of constriction. The presence of pulmonary infiltrate was significant in the Cox PH model but, insignificant in the shared frailty effect model. Also, tachycardia heart rate was not significant in the Cox PH model but significant in the shared frailty model. These results are likely to be because the frailty effect was highly significant in the event of constriction compared to other events ($p < 0.001$). After fitting both the Cox PH model and the shared gamma frailty model, the goodness of fit of the models was assessed through residual plots. Both models do not fit the data too badly. A comparison of the two modelling approaches based on their log-likelihood, AIC and BIC, suggest that the shared gamma frailty model fits the data better than the Cox PH model.

Also, we explored the use of parametric models, which was done for completion and comparison purposes since the proportional hazard assumption was satisfied in the Cox PH models. The hazard ratios for the Cox proportional hazards model and the Weibull model in the proportional hazards framework were very similar. However, the Cox-Snell residuals plot shows that the overall fit of the Weibull PH model isn't sufficiently adequate for the data. The result from the Weibull frailty models indicates a high amount of variation in the incidence of constriction but not in the incidence of death nor cardiac tamponade. Again these results are similar to the Cox frailty models presented in Chapter three of this thesis. For the AFT models, the

log-normal distribution appeared to have the best model fit as compared to other distributions commonly used in the AFT model. The Cox-Snell residuals plot shows that the overall fit of the lognormal AFT model is not sufficiently adequate for the data.

We proceeded with the competing risk analysis because there is a need to examine the contribution of competing risks analyses in the composite outcome setting; reasons for this were earlier discussed. We considered death as a competing risk to the event of cardiac tamponade and constriction since a patient can die before experiencing these events. Two main modelling approaches in competing risks settings, (cause-specific and sub-distribution hazards), were used. In both analyses, the patient in NYHA class one that had hypotension of >90 mmHg, had a lower hazard or sub-hazard of cardiac tamponade. Also, the patient who received prednisolone treatment compared to those who received placebo, female patients, patients that are in NHYA class one, had tachycardia (heart rate) >100 mmHg, haemoglobin level that is <10 g/dl and is HIV-positive, had lower hazard or sub-hazard of constriction. The effect estimates of covariates on the cause-specific hazard function and the sub-distribution hazard function are fairly similar. The estimates of these risk factors from the standard Cox proportional hazards models are the same as the estimates obtained from the analysis, using a cause-specific hazards function. This result is inevitable because a cause-specific hazards function treats competing events as being censored. In this thesis, the result of competing risks analysis has limited power because the IMPI trial was designed to study a composite outcome and not powered to study competing risks.

In the analysis of clustered competing risks, there is a significant centre effect in the incidence of constriction (p-value = <0.001). The observed heterogeneity across centres in the rates of constriction is remarkably large in contrast to the heterogeneity in the rates of cardiac tamponade. In comparison with the model from the competing risks that does not account for the centre effect, we observed that the covariate effect was only slightly changed. However, in the incidence of constriction, the presence of pulmonary infiltrate was significant in the competing risk model (without frailty) but, insignificant in the clustered competing risk model (with frailty). Also, the tachycardia heart rate was not significant in the competing risk model but significant in the clustered competing risk model.

In conclusion, a comparison of the modelling approaches using different statistical techniques suggests that the models which take into account the random effect are better than those that

do not, especially in the incidence of constriction. This is based on the significant result of the clustering parameter (theta).

7.2 Further research

A further area of research would be to conduct analysis of HIV-positive patients only, since these patients are more sick. Furthermore, other competing risks models such as the binomial regression, pseudo-value, and the multi-state modelling approach could be used. In addition, bivariate competing risks in which constriction and cardiac tamponade would be considered together against death, as a competing event using copulas or the Bayesian method, can be performed. Also, of interest would be to examine the effect of time-varying covariates such as the CD4 count or ART treatment in the competing risk setting(s).

Appendix A

Analyses of data for patients who received *M. indicus pranii* only

The appendix presents the results of the analysis which include *M. indicus pranii* treatment and its placebo group. The interpretation of the models is the same as the presented models in the application sections that included prednisolone treatment and its placebo group.

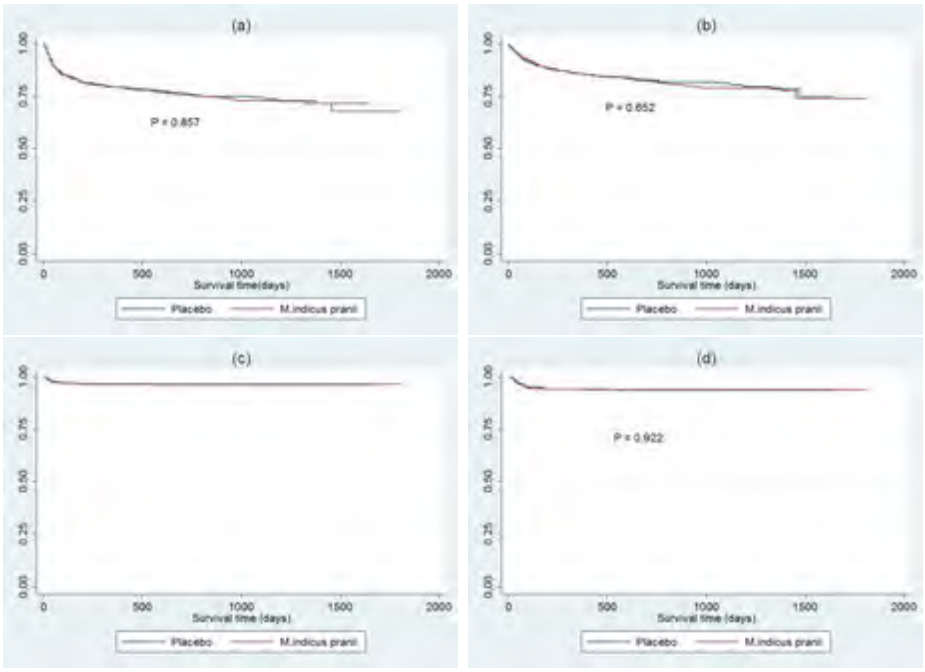

Figure A.1: *Estimated survival function for M. indicus pranii and placebo group until the end of study. (a) composite outcome, (b) death, (c) constriction, (d) cardiac tamponade.*

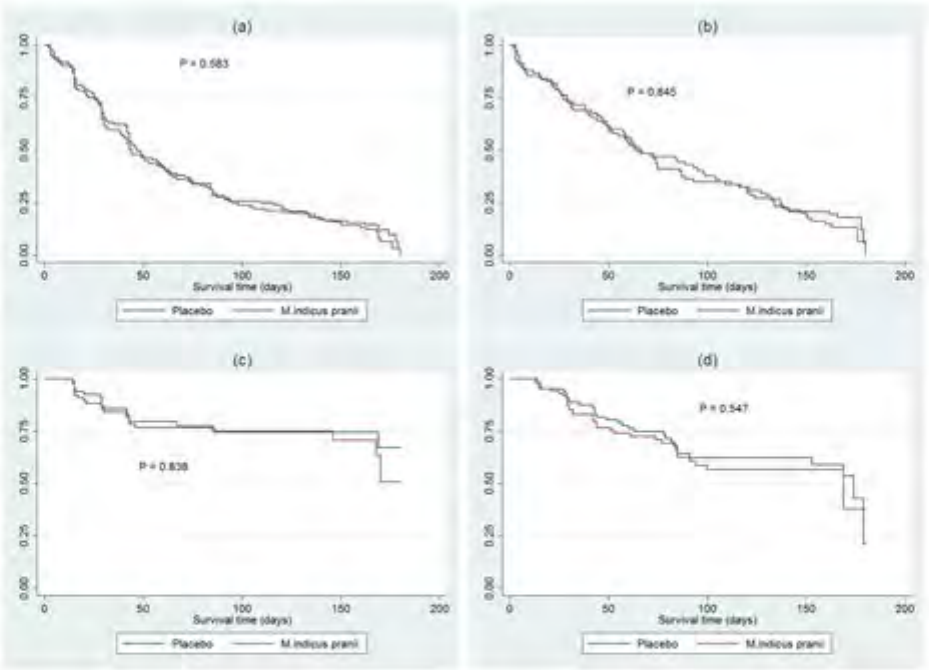

Figure A.2: *Estimated survival function for M. indicus pranii and placebo group at 6 months of the study. (a) composite outcome, (b) death, (c) constriction, (d) cardiac tamponade.*

A.1 Cox proportional hazards models

Table A.1: Univariate and multivariate Cox PH model for the relative hazard of the composite outcome.

	Univariate analysis		Multivariate analysis	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	1.02 (0.82 - 1.28)	0.857	1.12 (0.87 - 1.45)	0.379
Age (years)	1.00 (1.00 - 1.02)	0.030	1.01 (0.99 - 1.02)	0.092
Weight (kg)	0.91 (0.83 - 1.00)	0.042	0.89 (0.79 - 0.99)	0.031
NYHA Class				
I	1.00		1.00	
II	1.63 (1.12 - 2.37)	0.010	1.44 (0.91 - 2.27)	0.121
III	2.73 (1.86 - 4.01)	<0.001	2.27 (1.38 - 3.72)	0.001
IV	3.24 (2.08 - 5.07)	<0.001	2.49 (1.42 - 4.35)	0.001
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.59 (0.42 - 0.82)	0.002	0.73 (0.49 - 1.10)	0.132
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.75 (0.60 - 0.93)	0.011	0.80 (0.61 - 1.05)	0.115
Creatinine				
$\leq$ 105 umol/l	1.00		1.00	
$>$ 105 umol/l	1.51 (1.12 - 2.05)	0.006	1.76 (1.24 - 2.51)	0.002
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	1.47 (1.15 - 1.88)	0.002	0.92 (0.67 - 1.27)	0.625
Peripheral oedema				
No	1.00		1.00	
Yes	1.75 (1.41 - 2.17)	<0.001	1.52 (1.16 - 1.98)	0.002
Pulmonary infiltrate on CXR				
No	1.00		1.00	
Yes	1.23 (0.97 - 1.55)	0.087	1.12 (0.86 - 1.46)	0.399
Effusion size				
small	1.00		1.00	
medium	1.57 (0.93 - 2.64)	0.093	1.76 (0.91 - 3.39)	0.094
large	1.64 (1.00 - 2.69)	0.048	1.63 (0.87 - 3.04)	0.127
AIC			3148.256	

NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure; CXR: Chest Xray.
CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table A.2: Univariate and multivariate Cox PH model for the relative hazard of death.

	Univariate analysis		Multivariate analysis	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	1.06 (0.82 - 1.38)	0.652	1.01 (0.78 - 1.34)	0.922
Weight (kg)	0.85 (0.76 - 0.96)	0.006	0.87 (0.77 - 1.97)	0.024
NYHA Class				
I	1.00		1.00	
II	1.42 (0.93 - 2.16)	0.107	1.21 (0.74 - 1.97)	0.456
III	2.20 (1.42 - 3.43)	<0.001	1.79 (1.07 - 3.00)	0.027
IV	3.14 (1.91 - 5.16)	<0.001	2.45 (1.38 - 4.34)	0.002
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.45 (0.31 - 0.64)	<0.001	0.57 (0.38 - 0.85)	0.006
Creatinine				
$\leq$ 105 umol/l	1.00		1.00	
$>$ 105 umol/l	1.79 (1.28 - 2.52)	0.001	1.94 (1.35 - 2.77)	<0.001
Peripheral oedema				
No	1.00		1.00	
Yes	1.57 (1.23 - 2.02)	<0.001	1.43 (1.07 - 1.92)	0.016
AIC			2663.553	

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table A.3: Univariate and multivariate Cox PH model for the relative hazard of cardiac tamponade.

	Univariate analysis		Multivariate analysis	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	1.00 (0.55 - 1.81)	0.999	1.04 (0.50 - 2.16)	0.912
NYHA Class				
I	1.00		1.00	
II	3.04 (0.92 - 10.10)	0.069	2.61 (0.59 - 11.61)	0.207
III	4.65 (1.36 - 15.87)	0.014	4.65 (1.03 - 21.03)	0.046
IV	4.86 (1.21 - 19.43)	0.025	3.96 (0.66 - 23.78)	0.133
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.36 (0.17 - 0.74)	0.005	0.38 (0.14 - 0.99)	0.049
Atrial fibrillation on ECG				
No	1.00		1.00	
Yes	2.41 (0.85 - 6.82)	0.098	2.69 (0.93 - 7.80)	0.069
AIC			393.249	

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table A.4: Univariate and multivariate Cox PH model for the relative hazard of the constriction.

	Univariate analysis		Multivariate analysis	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	0.98 (0.62 - 1.55)	0.922	1.10 (0.67 - 1.81)	0.703
Age (years)	1.01 (1.00 - 1.03)	0.046	1.00 (0.98 - 1.02)	0.872
Gender				
Male	1.00		1.00	
Female	0.52 (0.32 - 0.82)	0.005	0.48 (0.27 - 0.86)	0.013
NYHA Class				
I	1.00		1.00	
II	1.93 (0.86 - 4.36)	0.111	1.33 (0.53 - 3.30)	0.544
III	3.85 (1.70 - 8.72)	0.001	2.53 (0.97 - 6.60)	0.059
IV	3.97 (1.54 - 10.24)	0.004	2.70 (0.90 - 8.13)	0.077
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	3.26 (0.80 - 13.27)	0.098	2.45 (0.59 - 10.22)	0.220
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.46 (0.29 - 0.75)	0.002	0.47 (0.26 - 0.84)	0.010
Haemoglobin				
$\leq$ 10 g/dl	1.00		1.00	
$>$ 10 g/dl	0.40 (0.26 - 0.64)	$<$0.001	0.43 (0.24 - 0.80)	0.007
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	1.99 (1.26 - 3.14)	0.003	1.24 (0.68 - 2.27)	0.485
Peripheral oedema				
No	1.00		1.00	
Yes	2.54 (1.64 - 3.94)	$<$0.001	1.63 (0.94 - 2.83)	0.081
Pulmonary infiltrate				
No	1.00		1.00	
Yes	1.49 (0.94 - 2.37)	0.090	1.50 (0.90 - 2.49)	0.120
HIV status				
Negative	1.00		1.00	
Positive	0.41 (0.27 - 0.63)	$<$0.001	0.60 (0.34 - 1.07)	0.082
AIC			834.157	

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC:Akaike's Information Criterion

A.2 Shared frailty models

Table A.5: Multivariable shared frailty model for the relative hazard of the composite outcome.

Characteristics	HR (95%CI)	p-value
Treatment		
Placebo	1.00	
M. indicus pranii	1.11 (0.86 - 1.44)	0.421
Age (years)	1.01 (0.99 - 1.02)	0.100
Weight (kg)	0.88 (0.79 - 0.99)	0.027
NYHA Class		
I	1.00	
II	1.41 (0.89 - 2.24)	0.149
III	2.19 (1.32 - 3.64)	0.002
IV	2.07 (1.16 - 3.69)	0.014
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	0.74 (0.49 - 1.12)	0.156
Tachycardia (HR)		
$\leq$ 100 mmHg	1.00	
$>$ 100 mmHg	0.85 (0.65 - 1.12)	0.253
Creatinine		
$\leq$ 105 umol/l	1.00	
$>$ 105 umol/l	1.78 (1.25 - 2.54)	0.002
Palpable pulsus paradoxus		
No	1.00	
Yes	0.94 (0.67 - 1.32)	0.736
Peripheral oedema		
No	1.00	
Yes	1.65 (1.24 - 2.20)	0.001
Pulmonary infiltrate on CXR		
No	1.00	
Yes	1.15 (0.88 - 1.51)	0.313
Effusion size		
small	1.00	
medium	1.66 (0.86 - 3.23)	0.134
large	1.55 (0.82 - 2.92)	0.177
θ (SE)	0.117 (0.090)	0.012

NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure; CXR: Chest Xray. CI: Confidence Interval; HR: Hazard Ratio; SE: Standard Error; θ: Heterogeneity parameter

Table A.6: Multivariable shared frailty model for the relative hazard of the death outcome.

Characteristics	HR (95%CI)	p-value
Treatment		
Placebo	1.00	
M. indicus pranii	1.01 (0.76 - 1.34)	0.942
Weight (kg)	0.87 (0.77 - 0.98)	0.021
NYHA Class		
I	1.00	
II	1.16 (0.71 - 1.90)	0.558
III	1.70 (1.01 - 2.90)	0.047
IV	2.27 (1.27 - 4.08)	0.006
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	0.57 (0.38 - 0.86)	0.008
Creatinine		
$\leq$ 105 umol/l	1.00	
$>$ 105 umol/l	1.91 (1.34 - 2.74)	<0.001
Peripheral oedema		
No	1.00	
Yes	1.53 (1.13 - 2.09)	0.006
θ (SE)	0.036 (0.042)	0.099

NYHA: New York Heart Association; SBP: Systolic Blood Pressure
CI: Confidence Interval; HR: Hazard Ratio; SE: Standard Error; θ: Heterogeneity parameter

Table A.7: Multivariable shared frailty model for the relative hazard of cardiac tamponade outcome.

Characteristics	HR (95%CI)	p-value
Treatment		
Placebo	1.00	
M. indicus pranii	1.05 (0.51 - 2.19)	0.887
NYHA Class		
I	1.00	
II	2.40 (0.53 - 10.85)	0.255
III	3.98 (0.86 - 18.50)	0.078
IV	2.86 (0.46 - 17.89)	0.262
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	0.36 (0.14 - 0.98)	0.045
Atrial fibrillation on ECG		
No	1.00	
Yes	2.63 (0.91 - 7.67)	0.076
θ (SE)	0.241 (0.291)	0.112

NYHA: New York Heart Association; SBP: Systolic Blood Pressure
CI: Confidence Interval; HR: Hazard Ratio; SE: Standard Error; θ: Heterogeneity parameter

Table A.8: Multivariable shared frailty model for the relative hazard of constriction.

Characteristics	HR (95%CI)	p-value
Treatment		
Placebo	1.00	
M. indicus pranii	1.10 (0.66 - 1.81)	0.722
Age (years)	1.00 (0.98 - 1.03)	0.674
Gender		
Male	1.00	
Female	0.46 (0.25 - 0.83)	0.009
NYHA Class		
I	1.00	
II	1.66 (0.65 - 4.23)	0.290
III	3.51 (1.32 - 9.35)	0.012
IV	2.26 (0.70 - 7.31)	0.174
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	2.50 (0.59 - 10.68)	0.215
Tachycardia (HR)		
$\leq$ 100 mmHg	1.00	
$>$ 100 mmHg	0.64 (0.36 - 1.16)	0.144
Haemoglobin		
$\leq$ 10 g/dl	1.00	
$>$ 10 g/dl	0.48 (0.26 - 0.88)	0.019
Palpable pulsus paradoxus		
No	1.00	
Yes	0.99 (0.52 - 1.93)	0.995
Peripheral oedema		
No	1.00	
Yes	1.84 (1.01 - 3.33)	0.045
Pulmonary infiltrate		
No	1.00	
Yes	1.41 (0.83 - 2.41)	0.206
HIV status		
Negative	1.00	
Positive	0.63 (0.35 - 1.15)	0.132
θ (SE)	0.1.316 (0.751)	$<$0.001

NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure

CI: Confidence Interval; HR: Hazard Ratio; SE: Standard Error; θ: Heterogeneity parameter

A.3 Parametric proportional hazards models

Table A.9: Weibull proportional hazards models with and without frailty in the composite outcome.

	Without frailty		With frailty	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	1.13 (0.87 - 1.46)	0.350	1.12 (0.86 - 1.44)	0.400
Age (years)	1.01 (1.00 - 1.02)	0.079	1.01 (1.00 - 1.02)	0.078
Weight (kg)	0.88 (0.79 - 0.98)	0.022	0.87 (0.78 - 0.98)	0.018
NYHA Class				
I	1.00		1.00	
II	1.37 (0.87 - 2.17)	0.173	1.35 (0.85 - 2.15)	0.207
III	2.19 (1.33 - 3.60)	0.002	2.12 (1.28 - 3.52)	0.004
IV	2.41 (1.38 - 4.22)	0.002	1.98 (1.10 - 3.56)	0.023
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.73 (0.48 - 1.09)	0.120	0.73 (0.48 - 1.10)	0.133
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.80 (0.61 - 1.05)	0.108	0.85 (0.64 - 1.12)	0.254
Creatinine				
$\leq$ 105 umol/l	1.00		1.00	
$>$ 105 umol/l	1.80 (1.26 - 2.55)	0.001	1.81 (1.27 - 2.59)	0.001
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	0.91 (0.66 - 1.25)	0.561	0.93 (0.66 - 1.30)	0.660
Peripheral oedema				
No	1.00		1.00	
Yes	1.53 (1.17 - 2.00)	0.002	1.68 (1.25 - 2.25)	0.001
Pulmonary infiltrate on CXR				
No	1.00		1.00	
Yes	1.14 (0.87 - 1.48)	0.350	1.16 (0.89 - 1.53)	0.275
Effusion size				
small	1.00		1.00	
medium	1.75 (0.91 - 3.39)	0.096	1.64 (0.84 - 3.20)	0.145
large	1.66 (0.89 - 3.10)	0.114	1.56 (0.83 - 2.94)	0.171
θ (SE)			0.135 (0.100)	0.007

NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure; CXR: Chest Xray.
CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table A.10: Weibull proportional hazards models with and without frailty in death outcome.

	Without frailty		With frailty	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	1.02 (0.77 - 1.34)	0.908	1.01 (0.77 - 1.34)	0.926
Weight (kg)	0.87 (0.77 - 0.98)	0.019	0.86 (0.76 - 0.97)	0.016
NYHA Class				
I	1.00		1.00	
II	1.18 (0.72 - 1.93)	0.513	1.14 (0.69 - 1.87)	0.615
III	1.77 (1.06 - 2.96)	0.030	1.68 (1.00 - 2.84)	0.052
IV	2.39 (1.35 - 4.23)	0.003	2.21 (1.22 - 4.00)	0.009
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.56 (0.38 - 0.85)	0.006	0.57 (0.38 - 0.86)	0.007
Creatinine				
$\leq$ 105 umol/l	1.00		1.00	
$>$ 105 umol/l	1.94 (1.36 - 2.77)	<0.001	1.91 (1.34 - 2.74)	0.<0.001
Peripheral oedema				
No	1.00		1.00	
Yes	1.44 (1.08 - 1.93)	0.014	1.55 (1.13 - 2.12)	0.007
θ (SE)			0.038 (0.044)	0.090

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table A.11: Weibull proportional hazards models with and without frailty in the cardiac tamponade outcome.

	Without frailty		With frailty	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	1.05 (0.51 - 2.19)	0.886	1.07 (0.52 - 2.22)	0.857
NYHA Class				
I	1.00		1.00	
II	2.56 (0.58 - 11.36)	0.217	2.34 (0.52 - 10.56)	0.270
III	4.46 (0.99 - 20.16)	0.052	3.77 (0.81 - 17.60)	0.091
IV	3.85 (0.64 - 23.10)	0.141	2.63 (0.39 - 17.85)	0.321
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.37 (0.14 - 0.96)	0.041	0.34 (0.13 - 0.93)	0.035
Atrial fibrillation on ECG				
No	1.00		1.00	
Yes	2.79 (0.96 - 8.10)	0.059	2.75 (0.94 - 8.01)	0.064
θ (SE)			0.301 (0.323)	0.074

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table A.12: Weibull proportional hazards models with and without frailty in the constriction outcome.

	Without frailty		With frailty	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	1.14 (0.69 - 1.87)	0.616	1.11 (0.67 - 1.84)	0.673
Age (years)	1.00 (0.98 - 1.02)	0.836	1.00 (0.98 - 1.03)	0.642
Gender				
Male	1.00		1.00	
Female	0.46 (0.26 - 0.82)	0.009	0.45 (0.25 - 0.80)	0.007
NYHA Class				
I	1.00		1.00	
II	1.18 (0.47 - 2.95)	0.724	1.42 (0.56 - 3.64)	0.462
III	2.38 (0.90 - 6.24)	0.079	3.17 (1.19 - 8.45)	0.021
IV	2.46 (0.82 - 7.40)	0.108	2.06 (0.64 - 6.61)	0.224
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	2.24 (0.54 - 9.32)	0.269	2.29 (0.54 - 9.71)	0.263
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.46 (0.26 - 0.81)	0.007	0.63 (0.35 - 1.15)	0.132
Haemoglobin				
$\leq$ 10 g/dl	1.00		1.00	
$>$ 10 g/dl	0.42 (0.23 - 0.78)	0.006	0.48 (0.26 - 0.89)	0.020
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	1.18 (0.65 - 2.17)	0.584	0.93 (0.48 - 1.80)	0.840
Peripheral oedema				
No	1.00		1.00	
Yes	1.65 (0.95 - 2.86)	0.073	1.83 (1.01 - 3.31)	0.047
Pulmonary infiltrate				
No	1.00		1.00	
Yes	1.53 (0.92 - 2.55)	0.101	1.44 (0.85 - 2.45)	0.178
HIV status				
Negative	1.00		1.00	
Positive	0.62 (0.35 - 1.10)	0.100	0.63 (0.34 - 1.15)	0.131
θ (SE)			1.325 (0.759)	< 0.001

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC:Akaike's Information Criterion

A.4 Accelerated Failure Models

Table A.13: Log-normal accelerated failure time model for the composite outcome.

	AFT analysis	
Characteristics	TR (95%CI)	p-value
Treatment		
Placebo	1.00	
M. indicus pranii	0.79 (0.45 - 1.39)	0.409
Age (years)	0.98 (0.96 - 1.00)	0.077
Weight (kg)	1.27 (0.99 - 1.62)	0.057
NYHA Class		
I	1.00	
II	0.42 (0.17 - 1.04)	0.061
III	0.16 (0.06 - 0.45)	<0.001
IV	0.14 (0.04 - 0.49)	0.02
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	1.67 (0.63 - 4.46)	0.304
Tachycardia (HR)		
$\leq$ 100 mmHg	1.00	
$>$ 100 mmHg	1.71 (0.94 - 3.10)	0.077
Creatinine		
$\leq$ 105 umol/l	1.00	
$>$ 105 umol/l	0.30 (0.13 - 0.69)	0.005
Palpable pulsus paradoxus		
No	1.00	
Yes	1.08 (0.52 - 2.24)	0.843
Peripheral oedema		
No	1.00	
Yes	0.41 (0.22 - 0.74)	0.003
Pulmonary infiltrate on CXR		
No	1.00	
Yes	0.87 (0.48 - 1.58)	0.653
Effusion size		
small	1.00	
medium	0.34 (0.09 - 1.31)	0.118
large	0.44 (0.12 - 1.53)	0.195

NYHA: New York Heart Association; HR: Heart Rate; SBP: Systolic Blood Pressure; CXR: Chest Xray.

CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table A.14: Log-normal accelerated failure time model for the death outcome.

	AFT analysis	
Characteristics	TR (95%CI)	p-value
Treatment		
Placebo	1.00	
M. indicus pranii	0.92 (0.53 - 1.60)	0.762
Weight (kg)	1.25 (0.99 - 1.57)	0.066
NYHA Class		
I	1.00	
II	0.63 (0.26 - 1.51)	0.297
III	0.30 (0.11 - 0.78)	0.014
IV	0.17 (0.05 - 0.54)	0.003
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	2.92 (1.17 - 7.32)	0.022
Creatinine		
$\leq$ 105 umol/l	1.00	
$>$ 105 umol/l	0.27 (0.12 - 0.58)	0.001
Peripheral oedema		
No	1.00	
Yes	0.48 (0.27 - 0.85)	0.013

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table A.15: Log-normal accelerated failure time model for the cardiac tamponade outcome.

	AFT analysis	
Characteristics	TR (95%CI)	p-value
Treatment		
Placebo	1.00	
M. indicus pranii	1.12 (0.13 - 9.83)	0.917
NYHA Class		
I	1.00	
II	0.09 (0.002 - 4.34)	0.226
III	0.02 (0.0003 - 0.99)	0.049
IV	0.03 (0.0002 - 5.11)	0.185
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	16.77 (0.57 - 491.27)	0.096
Atrial fibrillation on ECG		
No	1.00	
Yes	0.04 (0.001 - 1.75)	0.999

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table A.16: Log-normal accelerated failure time model for the constriction outcome.

	AFT analysis	
Characteristics	TR (95%CI)	p-value
Treatment		
Placebo	1.00	
M. indicus pranii	0.69 (0.20 - 2.36)	0.552
Age (years)	0.99 (0.95 - 1.04)	0.680
Gender		
Male	1.00	
Female	6.61 (1.62 - 27.02)	0.009
NYHA Class		
I	1.00	
II	0.49 (0.06 - 3.70)	0.486
III	0.11 (0.01 - 1.06)	0.057
IV	0.11 (0.01 - 1.66)	0.112
Hypotension (SBP)		
$\leq$ 90 mmHg	1.00	
$>$ 90 mmHg	0.10 (0.004 - 2.94)	0.185
Tachycardia (HR)		
$\leq$ 100 mmHg	1.00	
$>$ 100 mmHg	6.07 (1.53 - 24.02)	0.010
Haemoglobin		
$\leq$ 10 g/dl	1.00	
$>$ 10 g/dl	6.13 (1.45 - 25.79)	0.013
Palpable pulsus paradoxus		
No	1.00	
Yes	0.61 (0.13 - 2.93)	0.538
Peripheral oedema		
No	1.00	
Yes	0.31 (0.08 - 1.18)	0.087
Pulmonary infiltrate		
No	1.00	
Yes	0.49 (0.13 - 1.80)	0.284
HIV status		
Negative	1.00	
Positive	3.84 (0.93 - 15.85)	0.063

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC:Akaike's Information Criterion

A.5 Cumulative incidence function

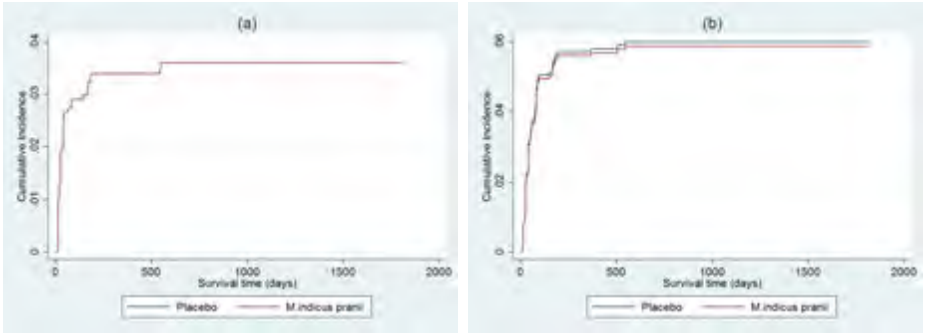

Figure A.3: *Estimated cumulative incidence function for M. indicus pranii and placebo group from the beginning of study until the end of the study. (a) constriction, (b) cardiac tamponade*

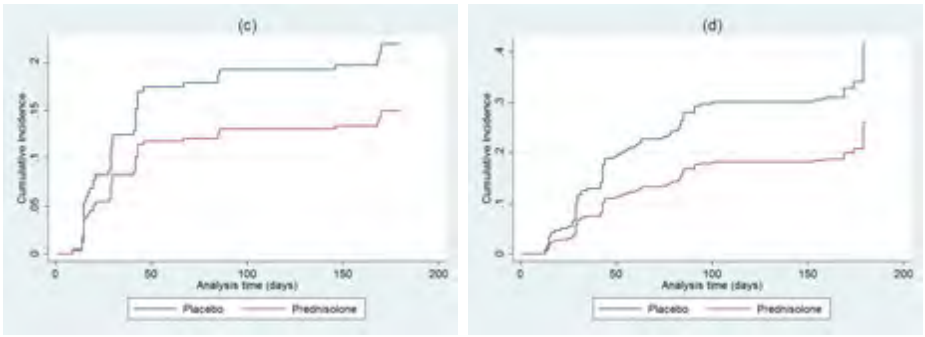

Figure A.4: *Estimated cumulative incidence function for M. indicus pranii and placebo group from the beginning of study until 6 months. (c) constriction, (d) cardiac tamponade*

A.6 Competing risks models

Table A.17: Multivariate regression models on the cause-specific hazards function and sub-distribution hazards function on the relative hazard of cardiac tamponade when death is a competing event.

	Sub-distribution hazards		Cause-specific hazards	
Characteristics	SHR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	1.04 (0.51 - 2.12)	0.919	1.04 (0.50 - 2.16)	0.912
NYHA Class				
I	1.00		1.00	
II	2.53 (0.57 - 11.19)	0.222	2.61 (0.59 - 11.61)	0.207
III	4.43 (0.98 - 20.05)	0.053	4.65 (1.03 - 21.03)	0.046
IV	3.76 (0.63 - 22.39)	0.145	3.96 (0.66 - 23.78)	0.133
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.39 (0.15 - 0.99)	0.005	0.38 (0.14 - 0.99)	0.049
Atrial fibrillation on ECG				
No	1.00		1.00	
Yes	2.50 (0.83 - 7.53)	0.104	2.69 (0.93 - 7.80)	0.069

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table A.18: Multivariate regression models on the cause-specific hazards function and sub-distribution hazards function on the relative hazard of constriction when death is a competing event.

	Sub-distribution hazards		Cause-specific hazards	
Characteristics	SHR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	1.10 (0.67 - 1.80)	0.715	1.10 (0.67 - 1.81)	0.703
Age (years)	1.00 (0.98 - 1.03)	0.894	1.00 (0.98 - 1.02)	0.872
Gender				
Male	1.00		1.00	
Female	0.50 (0.28 - 0.90)	0.021	0.48 (0.27 - 0.86)	0.013
NYHA Class				
I	1.00		1.00	
II	1.30 (0.52 - 3.26)	0.581	1.33 (0.53 - 3.30)	0.544
III	2.46 (0.93 - 6.48)	0.069	2.53 (0.97 - 6.60)	0.059
IV	2.45 (0.82 - 7.40)	0.111	2.70 (0.90 - 8.13)	0.077
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	2.46 (0.68 - 8.95)	0.171	2.45 (0.59 - 10.22)	0.220
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.49 (0.28 - 0.86)	0.013	0.47 (0.26 - 0.84)	0.010
Haemoglobin				
$\leq$ 10 g/dl	1.00		1.00	
$>$ 10 g/dl	0.43 (0.24 - 0.83)	0.009	0.43 (0.24 - 0.80)	0.007
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	1.27 (0.71 - 2.27)	0.420	1.24 (0.68 - 2.27)	0.485
Peripheral oedema				
No	1.00		1.00	
Yes	1.54 (0.89 - 2.70)	0.131	1.63 (0.94 - 2.83)	0.081
Pulmonary infiltrate				
No	1.00		1.00	
Yes	1.47 (0.88 - 2.46)	0.140	1.50 (0.90 - 2.49)	0.120
HIV status				
Negative	1.00		1.00	
Positive	0.60 (0.33 - 1.07)	0.084	0.60 (0.34 - 1.07)	0.082

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC:Akaike's Information Criterion

A.7 Clustered competing risks

Table A.19: Multivariate regression models for with and without frailty on the incidence of cardiac tamponade when death is a competing risk.

	Without frailty		With frailty	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	1.04 (0.50 - 2.16)	0.912	1.07 (0.51 - 2.21)	0.865
NYHA Class				
I	1.00		1.00	
II	2.61 (0.59 - 11.61)	0.207	2.44 (0.54 - 11.02)	0.245
III	4.65 (1.03 - 21.03)	0.046	4.01 (0.86 - 18.72)	0.077
IV	3.96 (0.66 - 23.78)	0.133	3.01 (0.45 - 20.20)	0.256
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.38 (0.14 - 0.99)	0.049	0.36 (0.13 - 0.94)	0.037
Atrial fibrillation on ECG				
No	1.00		1.00	
Yes	2.69 (0.93 - 7.80)	0.069	2.76 (0.95 - 8.01)	0.062
θ (SE)			0.259 (0.300)	0.195

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table A.20: Multivariate regression models for with and without frailty on the incidence of cardiac tamponade when death is a competing risk.

	Without frailty		With frailty	
Characteristics	SHR (95%CI)	p-value	SHR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	1.04 (0.51 - 2.12)	0.919	1.04 (0.54 - 2.01)	0.900
NYHA Class				
I	1.00		1.00	
II	2.53 (0.57 - 11.19)	0.222	2.61 (0.81 - 8.55)	0.110
III	4.43 (0.98 - 20.05)	0.053	4.65 (1.32 - 16.34)	0.017
IV	3.76 (0.63 - 22.39)	0.145	3.96 (0.82 - 19.18)	0.088
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	0.39 (0.15 - 0.99)	0.005	0.38 (0.16 - 0.87)	0.023
Atrial fibrillation on ECG				
No	1.00		1.00	
Yes	2.50 (0.83 - 7.53)	0.104	2.69 (0.65 - 11.14)	0.170

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC: Akaike's Information Criterion

Table A.21: Multivariate regression models for with and without frailty on the incidence of constriction when death is a competing risk.

	With frailty		Without frailty	
Characteristics	HR (95%CI)	p-value	HR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	1.09 (0.66 - 1.80)	0.730	1.10 (0.67 - 1.81)	0.703
Age (years)	1.00 (0.98 - 1.03)	0.894	1.00 (0.98 - 1.02)	0.685
Gender				
Male	1.00		1.00	
Female	0.45 (0.25 - 0.82)	0.008	0.48 (0.27 - 0.86)	0.013
NYHA Class				
I	1.00		1.00	
II	1.65 (0.65 - 4.22)	0.294	1.33 (0.53 - 3.30)	0.544
III	3.52 (1.32 - 9.40)	0.012	2.53 (0.97 - 6.60)	0.059
IV	2.30 (0.71 - 7.45)	0.165	2.70 (0.90 - 8.13)	0.077
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	2.51 (0.59 - 10.69)	0.212	2.45 (0.59 - 10.22)	0.220
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.64 (0.35 - 1.16)	0.144	0.47 (0.26 - 0.84)	0.010
Haemoglobin				
$\leq$ 10 g/dl	1.00		1.00	
$>$ 10 g/dl	0.48 (0.26 - 0.88)	0.018	0.43 (0.24 - 0.80)	0.007
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	0.99 (0.51 - 1.91)	0.978	1.24 (0.68 - 2.27)	0.485
Peripheral oedema				
No	1.00		1.00	
Yes	1.82 (1.01 - 3.31)	0.048	1.63 (0.94 - 2.83)	0.081
Pulmonary infiltrate				
No	1.00		1.00	
Yes	1.42 (0.83 - 2.42)	0.199	1.50 (0.90 - 2.49)	0.120
HIV status				
Negative	1.00		1.00	
Positive	0.63 (0.35 - 1.15)	0.132	0.60 (0.34 - 1.07)	0.082
θ (SE)	1.305 (0.746)	0.040		

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC:Akaike's Information Criterion

Table A.22: Multivariate regression models for with and without frailty on the incidence of constriction when death is a competing risk.

	Without frailty		With frailty	
Characteristics	SHR (95%CI)	p-value	SHR (95%CI)	p-value
Treatment				
Placebo	1.00		1.00	
M. indicus pranii	1.10 (0.67 - 1.80)	0.715	1.10 (0.85 - 1.44)	0.470
Age (years)	1.00 (0.98 - 1.03)	0.894	1.00 (0.98 - 1.02)	0.860
Gender				
Male	1.00		1.00	
Female	0.50 (0.28 - 0.90)	0.021	0.48 (0.28 - 0.82)	0.007
NYHA Class				
I	1.00		1.00	
II	1.30 (0.52 - 3.26)	0.581	1.33 (0.34 - 5.11)	0.680
III	2.46 (0.93 - 6.48)	0.069	2.53 (0.96 - 6.62)	0.059
IV	2.45 (0.82 - 7.40)	0.111	2.70 (0.76 - 9.57)	0.120
Hypotension (SBP)				
$\leq$ 90 mmHg	1.00		1.00	
$>$ 90 mmHg	2.46 (0.68 - 8.95)	0.171	2.45 (0.62 - 9.73)	0.200
Tachycardia (HR)				
$\leq$ 100 mmHg	1.00		1.00	
$>$ 100 mmHg	0.49 (0.28 - 0.86)	0.013	0.47 (0.26 - 0.84)	0.010
Haemoglobin				
$\leq$ 10 g/dl	1.00		1.00	
$>$ 10 g/dl	0.43 (0.24 - 0.83)	0.009	0.43 (0.28 - 0.67)	$<$0.001
Palpable pulsus paradoxus				
No	1.00		1.00	
Yes	1.27 (0.71 - 2.27)	0.420	1.24 (0.79 - 1.95)	0.350
Peripheral oedema				
No	1.00		1.00	
Yes	1.54 (0.89 - 2.70)	0.131	1.63 (1.07 - 2.50)	0.024
Pulmonary infiltrate				
No	1.00		1.00	
Yes	1.47 (0.88 - 2.46)	0.140	1.50 (0.73 - 3.08)	0.270
HIV status				
Negative	1.00		1.00	
Positive	0.60 (0.33 - 1.07)	0.084	0.60 (0.37 - 0.98)	0.043

NYHA: New York Heart Association; SBP: Systolic Blood Pressure; CI: Confidence Interval; HR: Hazard Ratio; AIC:Akaike's Information Criterion

Appendix B

Code

In this section the STATA and R-code used for analysis of the data from the IMPI clinical trial, which is presented in this thesis, is sketched. Some of the variables considered:

- time: Event time or censoring time in the standard survival analysis settings
- status: Indicating type of event or a censored observation in the standard survival analysis settings

  1 = composite, death, cardiac tamponade, or constriction

  0 = censored

- Time: Event time or censoring time in the competing risk settings
- Status: Indicating type of event or a censored observation in the competing risk settings

  1 = cardiac tamponade, or constriction

  2 = death

  0 = censored

- Group: Indicating treatment group

  1 = prednisolone or *M.indicus pranii*

  0 = placebo

B.1 Cox proportional hazards models

- Cox model

```
stset time, failure(status)
stcox group i.Nyhaclass Syblood ecgrthm4
```

- Shared frailty models

```
stcox group i.Nyhaclass Syblood ecgrthm4, shared(centre)
```

B.2 Parametric proportional hazards models

- Without frailty (centre effect)

```
streg group i.Nyhaclass Syblood ecgrthm4, dist(weibull)
```

- With frailty (centre effect)

```
streg group i.Nyhaclass Syblood ecgrthm4, d(weibull) frailty(gamma) shared(centre)
```

B.3 Accelerated failure time models

```
streg group i.Nyhaclass Syblood ecgrthm4, dist(lognormal) time
```

B.4 Competing risk models

- Cause-specific hazard model

```
stset Time, failure(Status=1)
stcox pred i.Nyhaclass Syblood ecgrthm4
```

- Sub-distribution hazards model

```
stcrreg pred i.Nyhaclass Syblood ecgrthm4, compete(Status == 2)
```

B.5 Clustered competing risk models

- Cause-specific hazards model with frailty (centre effect)

```
library(cmprsk); library(frailtypack)
source("C:/Users/Desktop/crr-addson.R")
cov <- cbind(group, factor2ind(Nyhaclass, 1), Syblood, ecgrthm4)

fit <- frailtyPenal(Surv(Time, Status==1) ~ cluster(centre) + cov,
data=Data, n.knots=6, kappa=10000, Frailty=TRUE, cross.validation=TRUE)
```

- Sub-distribution hazards model with frailty (centre effect)

```
library(crrSC)
fit <- crrc(Time, Status==1, cov, cluster=centre, failcode=1, na.action=na.omit)
```

Bibliography

Andersen, P. K., Abildstrom, S. Z., and Rosthoj, S. (2002). Competing risks as a multi-state model. *Statistical Methods in Medical Research*, 11(2):203–215.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25.

Chi, G. Y. (2005). Some issues with composite endpoints in clinical trials. *Fundamental & clinical pharmacology*, 19(6):609–619.

Christian, N. J., Ha, I. D., and Jeong, J.-H. (2016). Hierarchical likelihood inference on clustered competing risks data. *Statistics in medicine*, 35(2):251–267.

Clapp, J. M., Deng, Y., and An, X. (2006). Unobserved heterogeneity in models of competing mortgage termination risks. *Real Estate Economics*, 34(2):243–273.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151.

David Cox, R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 34:187–220.

Dianatkhah, M., Rahgozar, M., Talaei, M., Karimloua, M., Sadeghi, M., Oveisgharan, S., and Sarrafzadegan, N. (2014). Comparison of competing risks models based on cumulative incidence function in analyzing time to cardiovascular diseases. *ARYA atherosclerosis*, 10(1):6.

Do Ha, I., Christian, N. J., Jeong, J.-H., Park, J., and Lee, Y. (2014). Analysis of clustered competing risks data using subdistribution hazard models with multivariate frailties. *Statistical Methods in Medical Research*.

Do Ha, I., Lee, Y., and Song, J.-k. (2001). likelihood approach for frailty models. *Biometrika*, 88(1):233–233.

Duchateau, L. and Janssen, P. (2004). Penalized partial likelihood for frailties and smoothing splines in time to first insemination models for dairy cows. *Biometrics*, 60(3):608–614.

Duchateau, L., Janssen, P., Lindsey, P., Legrand, C., Nguti, R., and Sylvester, R. (2002). The shared frailty model and the power for heterogeneity tests in multicenter trials. *Computational Statistics & Data Analysis*, 40(3):603–620.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, 94(446):496–509.

Freemantle, N., Calvert, M., Wood, J., Eastaugh, J., and Griffin, C. (2003). Composite outcomes in randomized trials: greater precision but with greater uncertainty? *Jama*, 289(19):2554–2559.

Gooley, T. A., Leisenring, W., Crowley, J., Storer, B. E., et al. (1999). Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in medicine*, 18(6):695–706.

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

Ha, I. D. and Lee, Y. (2003). Estimating frailty models via poisson hierarchical generalized linear models. *Journal of Computational and Graphical Statistics*, 12(3):663–681.

Hougaard, P. (2012). *Analysis of multivariate survival data*. Springer Science & Business Media.

Jenkins, S. P. (2005). Survival analysis. *Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK*.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.

Katsahian, S. and Boudreau, C. (2011). Estimating and testing for center effects in competing risks. *Statistics in medicine*, 30(13):1608–1617.

Katsahian, S., Resche-Rigon, M., Chevret, S., and Porcher, R. (2006). Analysing multicentre competing risks data with a mixed proportional hazards model for the subdistribution. *Statistics in Medicine*, 25(24):4267–4278.

Klein, J. P. and Andersen, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, 61(1):223–229.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

Latouche, A., Allignol, A., Beyersmann, J., Labopin, M., and Fine, J. P. (2013). A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *Journal of clinical epidemiology*, 66(6):648–653.

Lee, M., Do Ha, I., and Lee, Y. (2014). Frailty modeling for clustered competing risks data with missing cause of failure. *Statistical methods in medical research*, page 0962280214545639.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 619–678.

Mathur, M. L. (2006). Potential utility of mycobacterium w vaccine in control of tuberculosis. *Current Respiratory Medicine Reviews*, 2(2):183–188.

Mayosi, B. M. (2007). Contemporary trends in the epidemiology and management of cardiomyopathy and pericarditis in sub-saharan africa. *American Heart Journal*, 93(10):1176–1183.

Mayosi, B. M., Burgess, L. J., and Doubell, A. F. (2005). Tuberculous pericarditis. *Circulation*, 112(23):3608–3616.

Mayosi, B. M., Ntsekhe, M., Bosch, J., Pandie, S., Jung, H., Gumedze, F., Pogue, J., Thabane, L., Smieja, M., Francis, V., et al. (2014). Prednisolone and mycobacterium indicus pranii in tuberculous pericarditis. *New England Journal of Medicine*, 371(12):1121–1130.

Mayosi, B. M., Ntsekhe, M., Bosch, J., Pogue, J., Gumedze, F., Badri, M., Jung, H., Pandie, S., Smieja, M., Thabane, L., et al. (2013). Rationale and design of the investigation of the management of pericarditis (impi) trial: A 2× 2 factorial randomized double-blind multicenter trial of adjunctive prednisolone and mycobacterium immunotherapy in tuberculous pericarditis. *American Heart Journal*, 165(2):109–115.

Mayosi, B. M., Ntsekhe, M., Volmink, J., and Commerford, P. (2002). Interventions for treating tuberculous pericarditis. *Cochrane Database Syst Rev*, 4.

Mayosi, B. M., Wiysonge, C. S., Ntsekhe, M., Burch, V. C., Maartens, G., Rebe, K., Commerford, P. J., Volmink, J. A., Gumedze, F., Aje, A., et al. (2008). Mortality in patients treated for tuberculous pericarditis in sub-saharan africa. *South African Medical Journal*, 98(1):36–40.

Mayosi, B. M., Wiysonge, C. S., Ntsekhe, M., Volmink, J. A., Gumedze, F., Maartens, G., Aje, A., Thomas, B. M., Thomas, K. M., Awotedu, A. A., et al. (2006). Clinical characteristics and initial management of patients with tuberculous pericarditis in the hiv era: the investigation of the management of pericarditis in africa (impi africa) registry. *BMC infectious diseases*, 6(1):2.

Ntsekhe, M., Wiysonge, C., Volmink, J., Commerford, P., and Mayosi, B. (2003). Adjuvant corticosteroids for tuberculous pericarditis: promising, but not proven. *Quality Journal of Medicine*, 96(8):593–599.

Ntsekhe, M., Wiysonge, C. S., Gumedze, F., Maartens, G., Commerford, P. J., Volmink, J. A., and Mayosi, B. M. (2008). Hiv infection is associated with a lower incidence of constriction in presumed tuberculous pericarditis: a prospective observational study. *PLoS One*, 3(6):e2253.

Pintilie, M. (2006). *Competing risks: a practical perspective*, volume 58. John Wiley & Sons.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V., and Breslow, N. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, pages 541–554.

Putter, H., Fiocco, M., Geskus, R., et al. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, 26(11):2389.

Rao, S. R. and Schoenfeld, D. A. (2007). Survival methods. *Circulation*, 115(1):109–113.

Reuter, H., Burgess, L., and Doubell, A. (2005). Epidemiology of pericardial effusions at a large academic hospital in south africa. *Epidemiology and infection*, 133(03):393–399.

Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56(4):1016–1022.

Scheike, T. H. and Zhang, M.-J. (2008). Flexible competing risks regression modeling and goodness-of-fit. *Lifetime data analysis*, 14(4):464–483.

Therneau, T. M. and Grambsch, P. M. (2000). Expected survival. In *Modeling Survival Data: Extending the Cox Model*, pages 261–287. Springer.

Therneau, T. M., Grambsch, P. M., and Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of computational and graphical statistics*, 12(1):156–175.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454.

WHO (2015). *Global tuberculosis control: World Health Organization report 2015*. World Health Organization.

Wolbers, M., Koller, M. T., Stel, V. S., Schaer, B., Jager, K. J., Leffondré, K., and Heinze, G. (2014). Competing risks analyses: objectives and approaches. *European heart journal*, page ehu131.