Analysis of a Bladder Cancer Clinical Trial Using a Shared Frailty Model

Jiang Yin

February 18, 2017

Acknowledgement

I am very grateful to my graduate study mentors, Dr. Douglas Wiens and Dr. Narasimha Prasad for their support, guidance and patience. Dr. Prasad taught my first graduate level statistics course and it was such a positive experience that inspired me into thinking of taking on a graduate study in this discipline. Dr. Wiens was very kind and accommodative to have accepted me as his graduate student. The mentorship by both of them remained instrumental in encouraging me to complete my graduate study amongst my increasingly busier work responsibilities at my non-statistical day job.

Abstract

Globally, bladder cancer is ranked No. 4 and No.7 in terms of cancer incidence rate among men and all genders, respectively. With an 80% recurrence rate, it is one of the most expensive cancers to treat on a per patient basis. Bladder cancer is a very heterogeneous disease with varying clinical manifestation. More than half of the bladder cancer cases upon first diagnosis belong to the non-muscle-invasive bladder cancer type (NMIBC), in which the tumorous growth is mostly limited to the inner walls of the bladder; about 25% of new cases are muscle-invasive bladder cancer (MIBC), i.e., clinically more malignant and with cancerous penetration into the deeper layers of bladder or nearby tissues such as lymph nodes; rarely (~4%) a new bladder cancer case is found to have metastasized to distant parts of the body. Smoking status, age, ethnicity, and occupational exposure to specific chemicals have been known as risk factors for bladder cancer. NMIBC patients often receive transurethral resection (surgical removal of inner tissue of bladder) or instillation therapy (insertional drug delivery), however, recurrence after treatment and progression to MIBC are not uncommon. MIBC patients are treated with radical cystectomy (removal of bladder), which often is accompanied with the removal of surrounding tissue/organ such as prostate or uterus and ovaries. The mortality rate for MIBC patients is 40% within the first five years and there is very limited choice of efficacious treatment regime for those with cancer relapse. Thus there is a clinical emphasis on finding the most appropriate therapy for NMIBC patients that reduces the likelihood of recurrence and progression to MIBC.

In a previously published study, Lammers et al. described a randomized clinical trial with the aim to compare two treatments for MNIBC patients, i.e., Keyhole limpet hemocyanin (KLH) vs mitomycin (MM)(Lammers, Witjes, Janzing-Pastors, Caris and Witjes 2012). This

prospective, multi-centered, randomized clinical trial was conducted in eighteen Dutch institutions. The main findings were: 1. KLH was not as effective as MM in preventing NMIBC recurrences and 2. KLH was more effective in blocking progression to MIBC than MM. However, the analysis was carried out without considering the potential heterogeneity among patients, in particular, some unobserved covariates at play differentially in each institution that could have a significant impact on the outcomes of the clinical trial. In this study, I explore the possibility of improving the analysis of the trial data using a frailty model that accommodates these unobserved covariates as random effect terms. A brief introduction on the frailty model is given in the next section followed by the analysis results and discussion on the limitations of the study.

Contents

List of Tables 6
List of Figures 7
Background 8
Data and Methods 14
Results 15
Discussion 24
References 26
Appendix A. Data summary 28
Appendix B. Categorized variable age is not a significant predictor for the RFS outcome. 29
Appendix C. Analysis codes 30

List of Tables

Table 1. Effect of variable Hospital on the hazard function in a Cox proportional model..... 15

Table 2. Shared frailty model for recurrence-free survival with all predictor variables..... 17

Table 3. Shared frailty model for recurrence-free survival with only significant predictors. 18

Table 4. Shared frailty model for progression-free survival..... 22

Table 5. Survival analysis using both recurrence and progression as event types. 23

List of Figures

Figure 1. Kaplan-Meier Survival Curves by Hospitals (event is recurrence).. 16

Figure 2.Test of proportionality. Test of proportionality.. 19

Figure 3. Proportionality plot by treatment type.. 20

Figure 4. Comparison of goodness of fit. . 21

Background

Frailty Model

With recent advancements in personalized medicine, modern clinical trials have been carried out with the aim to evaluate simultaneously the benefits and potential harms of drugs, treatments, or medical devices at the levels of population and individuals. It is now widely accepted that genetic heterogeneity exists in many major diseases, e.g., cancer, diabetes, autism, cystic fibrosis, Alzheimer's disease, Parkinson's disease, etc. This heterogeneity not only refers to the notion that multiple genes are involved in the pathogenesis but also points to the varying clinical presentation of symptoms as result of the different genetic defects in affected individuals. To quantitatively validate and compare the efficacy and/or effectiveness of treatments, such genetic heterogeneity is often of concern. In initial efficacy clinical trials, it is often more preferable to narrowly target a small subpopulation of patients with specific genetic, biological, or social characteristics. This reduces the risk of an overall trial failure for drug makers while at the same time echoing tightening regulatory policies on clinical trials around the globe. More stringent selection criteria means fewer patients are now eligible to enter clinical trial at one location. Consequently, many clinical trials take place in multiple centres. These centres could be geographically far apart; being located in different regions/countries/countries, this could lead to non-designed heterogeneity in the genetic makeup of the patients receiving the trial. Other regionally and politically sensitive factors, such as the diversity in healthcare systems and their regulatory policies, language and culture, climate, etc., could also play a role in influencing the outcomes of multi-centre clinical trials. Quite often, these factors, along with some presumed unobserved covariates, are not the primary consideration for important predictors in clinical trials. The traditional Cox proportional model thus becomes inadequate for the analysis of survival data

such as those in a clinical trial because it assumes a homogeneous population, a condition that is rarely satisfied in modern healthcare setting.

To accommodate the effects of unmeasured and unobserved covariates, the traditional Cox proportional hazard model is augmented by the addition of random effects, termed frailty parameters, that may have differentiated among various trial centres. In essence, frailty offers a convenient means of absorbing random effects, associations and unobserved but presumed heterogeneity into statistical models for survival data such as those in a clinical trial. Mathematically, it is treated as an unobserved proportionality factor for the hazard function of an individual with correlated measurements or of a cluster of individuals with some sort of relatedness, either specified or otherwise.

There are three main types of frailty models, each differentiated by the way the frailty term is defined: univariate frailty model; shared frailty model; correlated frailty model. The latter two are also grouped under multivariate frailty models.

The standard hazard function of the Cox proportional model has the form (Cox 1972):

$$\mu(t, X) = \mu_0(t, X)\exp(\beta^T X)$$

where $\mu_0(t)$ denotes the baseline hazard function, assumed to be unique for all individuals in a homogeneous population. X is the vector of observed covariates and b is the corresponding coefficients to be estimated. The mathematical clarity and attractiveness of this model lies in the separation of age effects ($\mu_0(t)$) from those of the predictive covariates ($\exp(\beta^T X)$).

However, as mentioned above, it is often impractical to assume a homogeneous population under study. Furthermore, not all important covariates could be measured, observed or even known.

Thus, heterogeneity caused by unknown or unobserved covariates contributes to the unpredictable part of the variability in the survival data. Ignoring potentially influential covariates in a proportional hazard model can bias the regression coefficients of the remaining covariates and the hazard rate, in particular, this will underestimate the hazard function over time/age (Bretagnolle and Huber-Carol 1988). It then becomes clear that specifying the effect of heterogeneity in the model will lead to better estimated regression parameters corresponding to those observed covariates.

Univariate frailty model

The univariate frailty model adds an multiplicative frailty term, Z, to the baseline hazard function in the traditional Cox model:

$$\mu(t, Z, X) = \mathrm{Z}\mu_0(t, X)\exp(\beta^T X)$$

Here, Z is a random variable changing among individuals. When Z is less than 1, it leads to reduced individual risk. Likewise, when Z is greater than 1, the individual assumes a higher risk. Correspondingly, the survival function is given by

$$S(t|Z, X) = \exp(-\mathrm{Z}\int_0^t \mu_0(s)ds \exp(\beta^T X))$$

this computes the portion of surviving subjects at time t given the vector of covariates X and frailty X. Since the survival at individual level is not observable, this function can otherwise be used to calculate the mean survival function of the population averaged over individuals with different Z. It is important to note that the population and individual survival functions could have very different shapes.

Shared frailty model

Shared frailty model deals with survival data that involve related individuals or multiple measurements of the same individual. For multi-centre clinical trials, it is often convenient to consider the patients at all trial centres as independent clusters each sharing the same frailty factor within that cluster. This model assumes the same hazard function for an individual in a cluster as the one in the univariate frailty model. Due to the common value of Z assigned to all individuals in the cluster, there exists dependence in the survival times of these individuals. The multivariate survival function then adopts the form

$$S(t_1, t_2, \dots, t_k|Z) = \prod_{1}^{k} S_i(t_i)^Z$$

The dependence among clustered individuals is determined by some degenerate distribution g(Z) ($\hat{Z}$=some constant typically 1, σ_z^2). When σ_z^2 is zero, the survival times of all individuals are independent. Otherwise, σ_z^2 is invariably positive due to the way the models are established. Typically, Z is assumed to follow a prescribed distribution such as the gamma distribution with mean of 1 and variance of σ_z^2. This then produces the joint survival function of the k_i individuals in the ith cluster as follows

$$S(t_{i1}, t_{i2}, \dots, t_{ik_i}) = \Pr\left(T_{i1} > t_{i1}, T_{i2} > t_{i2}, \dots, T_{ik_i} > t_{ik_i}\right)$$

$$= \int_0^{\infty} \prod_{j=1}^{k_i} Pr(T_{ij} > t_{ij}|Z_i) g(z_i) dz_i = \left[\sum_{j=1}^{k_i} S_{ij}\left(t_{ij}\right)^{-\sigma^2} - 1\right]^{1/\sigma^2}$$

$$= \left[\frac{1}{\sigma^2} \sum_{j=1}^{k_i} \widehat{\mu_0}(t_{ij}) exp(\beta^T X_{ij}) + 1 \right]^{1/\sigma^2}$$

β, σ^2, and $\widehat{\mu_0}(t)$ are often estimated by the EM (Expectation-Maximization) algorithm (Dempster, Laird and Rubin 1977).

Correlated frailty model

Sometimes a more versatile model is needed to describe the correlation structure in the data. Most of the correlated frailty models were developed for application to bivariate data such as twin data or parent-child data. In this model, the two individuals in a pair are each assigned a frailty variable and the constraint for them to be the same is removed. These frailties are dependent and are often included in the model as independent additive variables carrying some common components. The advantage of this model over the shared frailty model lies in that there is now no restriction on the types of correlation between the two frailties. The forms of survival functions are necessarily more complex than that of the shared frailty model and similarly depend on the choice of the distribution of the frailty variables. Previous studies have given the formulae for bivariate survival functions conditional on different frailty distributions such as correlated gamma, log-normal, etc. (Pickles, et al. 1994, Yashin and Iachine 1995, Yashin, Vaupel and Iachine 1995, Zahl and Tretli 1997, Yashin and Iachine 1999, Vaida and Xu 2000)

Bladder cancer clinical trial

For non-muscle invasive bladder cancer (NMIBC) patients, recurrence and progression to more malignant muscle invasive bladder cancer (MIBC) after initial treatment are of grave concern. Two instillation treatments, keyhole limpet hemocyanin (KLH) and mitomycin (MM), were

compared in the randomized clinical trial that recruited 553 patients in total at 18 centres (Lammers, et al. 2012). Primary outcome was recurrence-free survival (RFS). Secondary outcomes were progression-free survival (PFS), adverse effects (AE), and the effect of delayed-type hypersensitivity (DTH) response on clinical outcome. In the original study, the Kaplan-Meier method and one-sided log-rank test were used to statistically test the differences in RFS and PFS between KLH and MM treatments. To analyze the influence of patient and tumor characteristics on RFS and PFS, the authors used crude hazard ratio (HRs) obtained from univariate Cox regression. Adjusted HRs from multivariate Cox regression were obtained for the correction of confounding factors. Two-sided $\chi 2$ test was used to compare patient and tumor characteristics. Patient recurrence was compared using one-sided Fisher's exact test, whereas number of recurrent tumors was compared using one-sided Mann-Whitney U test. Using a log-rank test, it was found that KLH was less effective than MM regarding RFS ($P < 0.001$). Progression events were scarce ($n = 20$). While univariate Cox regression analysis indicated that KLH was more effective than MM to prevent progression, a multivariate Cox regression analysis could not support this notion. In this report, I will focus on RFS and PFS outcomes as frailty model is not really applicable to count data such as AE and DTH outcomes. I will first analyze the survival times at each centre separately and then use the shared frailty model to analyze the data assembled from all test centres. The results will be compared against those in the original study and the limitation of the method used in this study will be discussed.

Data and Methods

There are a total of 164 variables collected in the original dataset. For the present study, 21 variables were selected and their general description is listed in Appendix A. The distributions of many of these variables in the two treatment arms had been discussed in the original publication therefore are omitted here. Variable age was included as a continuous covariate in the Cox proportional hazard (CPH) model and it was found that the effect of age was not significant. In this study, it was initially converted to a categorical variable with 5 levels (20-50, 50-60, 60-70, 70-80, 80-). However, it was discovered that this did not change the previous conclusion that age does not play a significant role in the present CPH model (Appendix B). Variables for the size of tumors and the number of tumors at the start of the study were converted to bi-level categorical variables with corresponding levels (<15mm vs > 15mm) and (single tumor vs multiple tumors), respectively.

Two software packages were used in this study, Stata12 (Stata Corp, College Station, TX, USA) and R version 3.3.2 (www.r-project.org). Stata has a built-in module in its CPH survival analysis tool that allows the shared frailty model using the keyword "shared()". For R, an extension library "Frailtypack" was installed and the function "frailtyPenal" was used to build a shared frailty model. Both Stata and R gave closely similar results, however, each has its limitations. Frailtypack did not allow multivariate (multiple outcome variable) models whereas Stata only permitted Gamma distribution for the frailty component. For the sake of brevity and clarity, results will be presented without mentioning its software origin and most of the results shown in this report were produced using Stata.

Results

We started by comparing the Kaplan-Meier survival functions among different hospitals. As suspected, there existed significant differences among some of the hospitals (Table 1 and Figure 1).

Table 1. Effect of variable Hospital on the hazard function in a Cox proportional model.

Variable	Haz. Ratio	p-value	95% CI lower	upper
Treatment (reference, KLH)				
MM	0.420	<0.0001	0.323	0.546
Hospital (reference, CWZ, Nijmegen)				
St. Jansdal Ziekenhuis, Harderwijk	0.589	0.084	0.323	1.074
Catharina Ziekenhuis, Eindhoven	0.427	0.028	0.199	0.914
St. Elisabeth Ziekenhuis, Tilburg	0.909	0.755	0.499	1.656
Laurentius Ziekenhuis, Roermond	0.471	0.017	0.254	0.874
Rijnstate Ziekenhuis, Arnhem	0.492	0.026	0.264	0.917
OLVG, Amsterdam	0.463	0.047	0.217	0.991
RKZ, Den Haag	0.615	0.223	0.281	1.345
Diaconessenhuis, Leiden	0.600	0.118	0.316	1.139
VieCuri, Venlo	0.211	<0.0001	0.099	0.451
Ziekenhuis Koningin Beatrix, Winterswijk	0.311	0.005	0.138	0.701
TweeSteden Ziekenhuis, Tilburg	0.373	0.014	0.170	0.816
Medisch Centrum Haaglanden, Den Haag	0.425	0.038	0.189	0.954
UMC St. Radboud, Nijmegen	0.372	0.013	0.170	0.814
Martini Ziekenhuis, Groningen	0.481	0.076	0.214	1.081
Bethesda Ziekenhuis, Hoogeveen	0.356	0.043	0.131	0.967
Wilhelmina Ziekenhuis, Assen	0.245	0.003	0.096	0.624
Isala Klinieken, Zwolle	0.200	0.004	0.067	0.595

With CWZ, Nijmegen as reference, 12 other hospitals showed significantly lower hazard with a ratio ranging from ~0.2 to 0.42 (Table 2). Log-rank test gave a p-value of 0.0031 for the variable Hospital as a factor. This motivated the use of a shared frailty model to account for *Hospital* effects. The results are shown in Table 2.

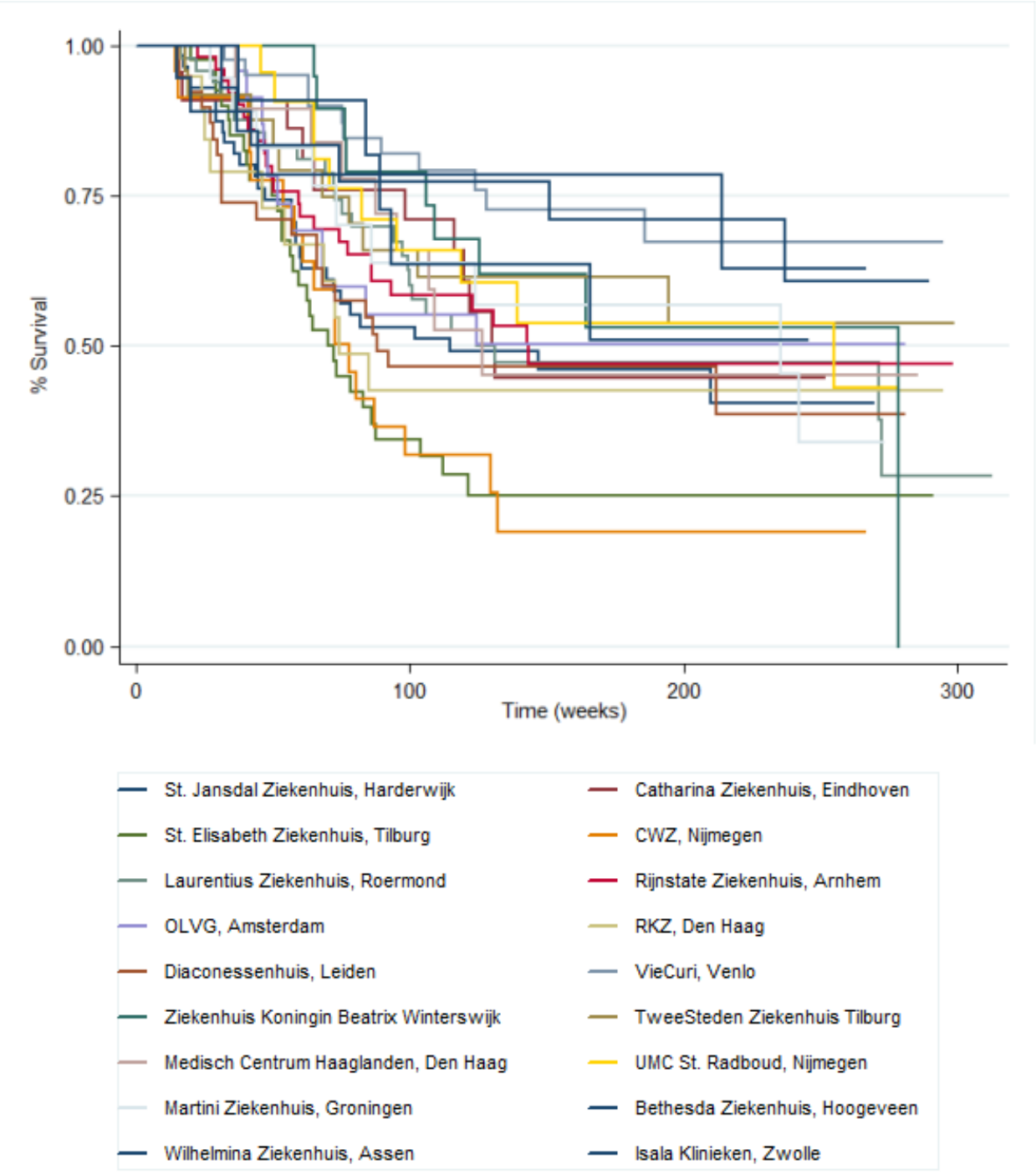

Figure 1. Kaplan-Meier Survival Curves by Hospitals (event is recurrence). The number of patients at the eighteen trial centres ranged from 11 to 58, averaging 29 patients per centre with a standardized spread of 13.

Table 2. Shared frailty model for recurrence-free survival with all predictor variables.

	Regular CPH				Shared Frailty			
Variable	Haz. Ratio	p-value	95% CI		Haz. Ratio	p-value	95% CI	
Treatment (ref., KLH)								
MM	0.37	<0.0001	0.28	0.48	0.39	<0.0001	0.29	0.50
Age	1.01	0.42	0.99	1.02	1.01	0.37	0.99	1.02
Sex (ref., male)								
Female	1.08	0.69	0.76	1.53	1.05	0.78	0.74	1.48
Primrec (ref., primary)								
Recurrent	1.79	<0.0001	1.33	2.42	1.80	<0.0001	1.33	2.42
tumor stage (ref., pTa)								
pT1	1.37	0.10	0.94	2.00	1.31	0.15	0.91	1.87
tumor grade (ref., grade 1)								
Grade 2	2.00	<0.0001	1.43	2.79	1.88	<0.0001	1.37	2.59
Grade 3	1.18	0.53	0.70	1.98	1.14	0.60	0.69	1.88
size of tumor (ref., < 15mm)								
larger than 15mm	1.03	0.87	0.76	1.38	1.05	0.75	0.79	1.39
Number of tumor (ref., single)								
Multiple	2.11	<0.0001	1.54	2.88	2.12	<0.0001	1.56	2.87

Overall, the frailty model did not significantly change the coefficients or hazard ratios of the other variables compared to a regular CPH model and the frailty component was significant (p-value = 0.003). According to this model, age, gender, initial tumor stage and tumor size were not significant predictors for the outcome, so a subsequent survival analysis was performed after dropping these predictors (Table 3). Comparing the results in Tables 2 and 3, the corresponding coefficients and hazard ratios were close in both value and significance, especially those for the treatment effect. Tumor grade measures the dissimilarity between cancer cells and normal cells; the higher the grade, the greater the dissimilarity and the greater rate at which cancer is likely to grow and recur. In bladder cancer, grade 1 (or papilloma) is also called benign papillary urothelial neoplasm of low malignant potential (PUNLMP) and denotes a cancer type that may

recur but has a low risk of progression; grade 2 (low grade) is more likely to recur and progress than PUNLMP; grade 3 (high grade) is most likely to recur and progress. Interestingly, while grade 2 is significantly more likely to negatively impact recurrence-free survival time compared to grade 1, grade 3 is not significantly different from grade 1. The assumption of proportionality for most variables in the regression models seems to be reasonable for both the CPH and shared frailty models (Figure 2). Although the variable treatment is not entirely horizontal, the survival curves of the two treatments remain largely parallel to each other (Figure 3). When examined for their goodness of fit, the shared frailty model appears to be more appropriate for the data; there exist apparently larger deviations between model and data towards the longer times in the regular CPH model (Figure 4), although such phenomenon is quite common in survival data due to censoring at the end of the study.

Table 3. Shared frailty model for recurrence-free survival with only significant predictors.

	Regular CPH				Shared Frailty			
Variable	Haz. Ratio	p-value	95% CI		Haz. Ratio	p-value	95% CI	
Treatment (ref., KLH)								
MM	0.38	<0.0001	0.29	0.50	0.39	<0.0001	0.30	0.51
Primrec (ref., primary)								
Recurrent	1.68	<0.0001	1.28	2.22	1.68	<0.0001	1.28	2.21
tumor grade (ref., grade 1)								
Grade 2	2.08	<0.0001	1.51	2.88	1.96	<0.0001	1.44	2.67
Grade 3	1.34	0.24	0.83	2.18	1.29	0.29	0.80	2.06
Number of tumor (ref., single)								
Multiple	2.03	<0.0001	1.49	2.76	2.05	<0.0001	1.51	2.77

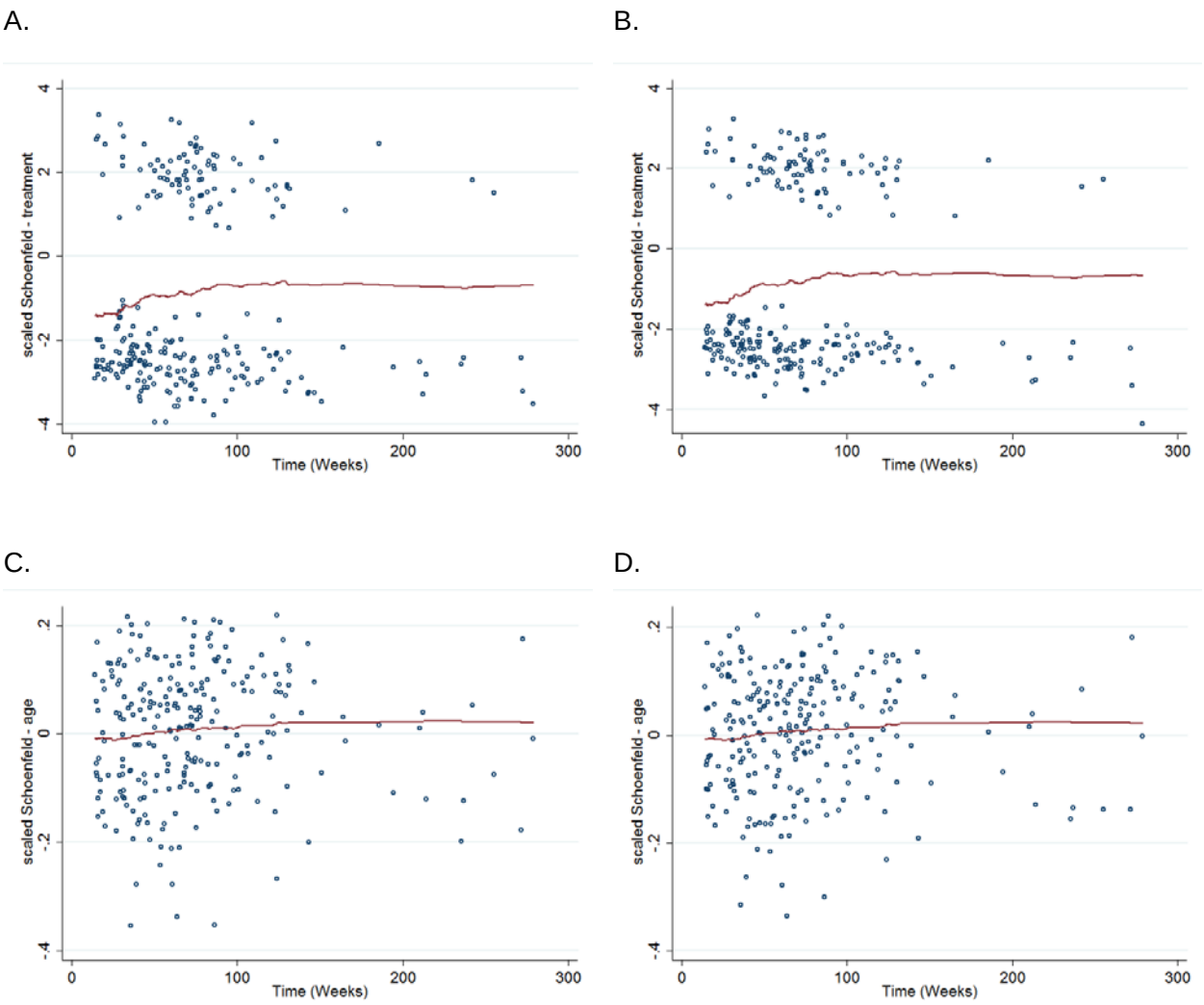

Figure 2.Test of proportionality. Test of proportionality. A) and C) CPH model. B) and D) Shared frailty model. Scaled Schoenfeld residuals were plotted against time. For brevity, the plots for the other variables were omitted here.

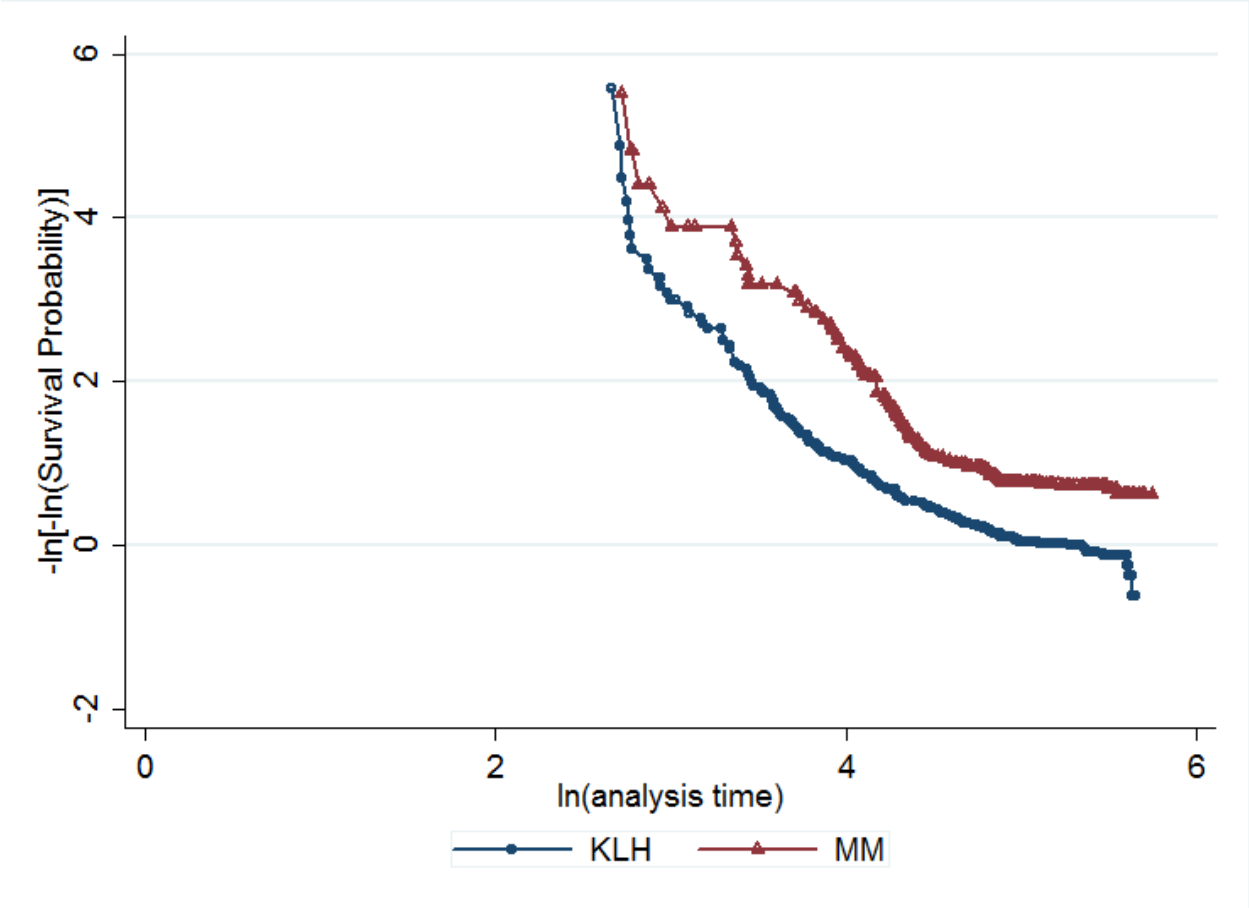

Figure 3. Proportionality plot by treatment type. The few events proximal to the beginning of the study (curves) seemed a little too close, but the remainder of the curves are relatively parallel to each other.

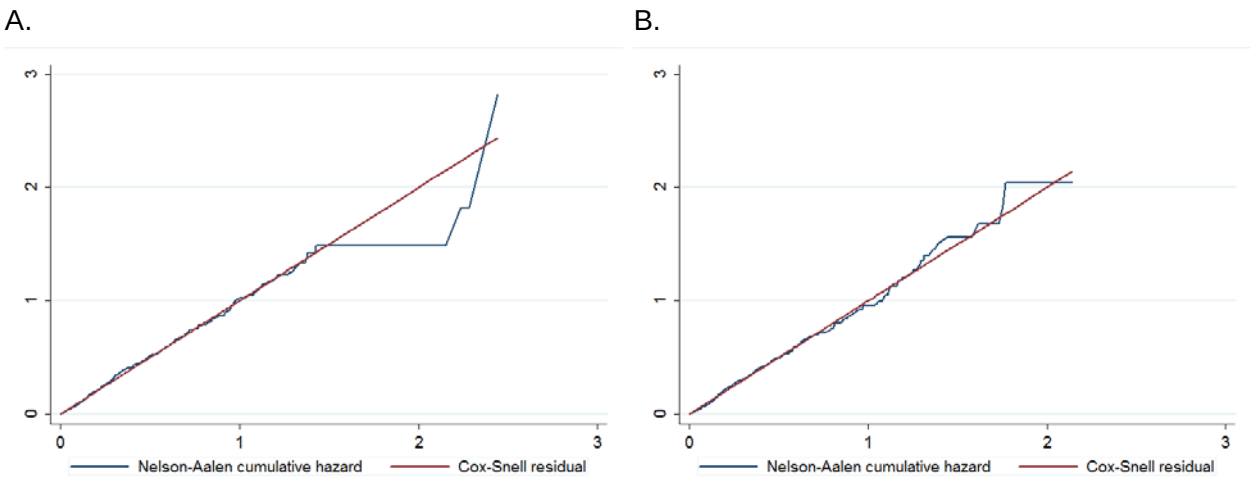

Figure 4. Comparison of goodness of fit. A) Regular CPH model. B) The shared frailty (hospital) model. The data used were recurrence-free survival time (in weeks). After fitting the data with respective models, Cox-Snell residuals were generated and these were used to create the Nelson-Aalen cumulative hazard function, which was then plotted against the Cox-Snell residuals. A plot closely matching the 45° diagonal indicates satisfactory model fit of the data.

When using progression-free survival as the outcome, however, the shared frailty model often did not converge to a solution -- the optimization routine failed likely due to the low number of failure events (20 in total out of 523 patients) (Table 4). To accommodate both recurrence and progression in the same model, we used Stata to perform survival analysis denoting recurrence and progression as two different types of failure events. The results are shown in Table 5. There is no built-in mechanism to include a shared frailty component in this

model, therefore it was not attempted. This model does not deviate much from the RFS model represented in Table 3, with but one difference in that the initial tumor stage now became significant with a p-value of 0.04. Given that the initial tumor grade was found to be a significant factor in all the models presented in this study, the new finding in Table 5 seems to suggest that the effects on survival of bladder cancer stage and grade are somewhat associated, a notion also supported by a weak association between the two variables ($R^2 = 0.17$, p-value <0.0001) and some previous report(Singh, et al. 2016).

Table 4. Shared frailty model for progression-free survival.

	Regular CPH				Shared Frailty			
Variable	Haz. Ratio	p-value	95% CI		Haz. Ratio	p-value	95% CI	
Treatment (ref., KLH)								
MM	2.66	0.059	0.96	7.37	2.61	0.06	0.95	7.19
Age	1.03	0.24	0.98	1.08	-*	-	-	-
Sex (ref., male)								
Female	1.48	0.49	0.49	4.50	1.19	0.76	0.40	3.56
Primrec (ref., primary)								
Recurrent	0.69	0.48	0.25	1.91	-	-	-	-
tumor stage (ref., pTa)								
pT1	4.37	0.01	1.53	12.43	-	-	-	-
tumor grade (ref., grade 1)								
Grade 2	7.60	0.06	0.93	61.81	-	-	-	-
Grade 3	23.12	0.01	2.64	202.46	-	-	-	-
size of tumor (ref., < 15mm)								
larger than 15mm	0.48	0.13	0.19	1.24	0.73	0.48	0.30	1.75
Number of tumor (ref., single)								
Multiple	1.63	0.34	0.60	4.44	-	-	-	-

*, optimization routine failed to converge.

Table 5. Survival analysis using both recurrence and progression as event types.

	Regular CPH			
Variable	Haz. Ratio	p-value	95% CI	
Treatment (ref., KLH)				
MM	0.42	<0.001	0.33	0.55
Age	1.01	0.38	0.99	1.02
Sex (ref., male)				
Female	1.13	0.48	0.80	1.60
Primrec (ref., primary)				
Recurrent	1.73	<0.001	1.28	2.32
tumor stage (ref., pTa)				
pT1	1.47	0.04	1.01	2.12
tumor grade (ref., grade 1)				
Grade 2	2.06	<0.001	1.51	2.81
Grade 3	1.47	0.15	0.87	2.46
size of tumor (ref., < 15mm)				
larger than 15mm	0.95	0.72	0.70	1.28
Number of tumor (ref., single)				
Multiple	2.10	<0.001	1.55	2.86

Discussion

The potential of Keyhole limpet hemocyanin (KLH) in treating urothelial cancer was first discovered in 1974 when it was noticed that patients immunized with KLH benefited a marked drop in recurrence rate of non-invasive urothelial cancer. The molecular mechanism underlying KLH immunotherapy for cancer treatment is still poorly understood, although the prevailing hypothesis hinges on the position that strong antigenic molecules such as KLH enhances immunity and body's natural anti-tumor defence in general. In bladder cancer, several agents of non-human origin have been investigated for its immunotherapeutic effectiveness: Bacillus Calmette-Guérin (BCG) (Pirzada, et al. 2017), KLH (Lammers, et al. 2012), Rubratin (de Reijke, de Boer, Schamhart and Kurth 1997), and mistletoe lectin (Elsasser-Beile, Ruhnau, Freudenberg, Wetterauer and Mengs 2001). Despite the prevalence of BCG in contemporary treatment regimens, KLH has the advantage of much lower toxicity thus remains a popular choice in clinical studies for its effectiveness against cancer cells. Several preclinical studies largely confirmed that KLH can reduce tumor growth (Lamm, DeHaven, Riggs, Delgra and Burrell 1993, Lamm, DeHaven, Riggs and Ebert 1993), although one study did not find any difference between control and KLH treated groups in either tumor growth or survival (Walsh, Tomashefsky, Olsson and deVere White 1983). More recently, a fairly large phase III study by Lammers, et al showed that KLH is inferior to adjuvant mitomycin C instillation in the prevention of tumor recurrence and progression (Lammers, et al. 2012). It is this study that provided the raw data for the present study reported here.

The clinical trial in question was carried out in as many as eighteen centres across the Netherlands. Upon examining the original data, we detected significant differences in cancer survival among these centres (Table 1). By including a random effect component, a shared frailty

term, we could model the variations in survival from one hospital to another using a random variable presumed to follow gamma distribution. This, in turn, may improve estimation of model residuals and precision of the regression coefficients, as seen in the slightly narrower confidence intervals of the hazard ratios (Tables 2-3).

It is of note that while shared frailty conveniently explains the correlation between subjects within a category level (e.g., trial location), it does have some limitations. To begin with, it restricts the obscure factors to be the same within the cluster (in this case, hospitals). This assumption may not always be realistic. For example, one might assume that all patients in a particular hospital share all their unobserved risk factors because of their geographical closeness, however, if a hospital is very famous, it might have attracted patients from far away or even from abroad. Furthermore, a one-dimensional frailty can only model positive association within the cluster for most cases. However, in some scenarios, the survival times for subjects within the same cluster could be negatively associated. Last but not the least, the association between survival times within the cluster is based on their marginal distributions. However, when covariates are in a CPH model with gamma distributed frailty component, the dependence parameter and the population heterogeneity could be confounded (Clayton and Cuzick 1985). This implies that the joint distribution could be identified from the marginal distributions (Hougaard 1986). Some or all of the abovementioned limitations could be at play during the present study thus remain the points of interest for future study of the clinical trial data.

References

Bretagnolle, J., and Huber-Carol, C. (1988), "Effects of Omitting Covariates in Cox's Model for Survival Data," *Scandinavian Journal of Statistics*, 15, 125-138.

Clayton, D., and Cuzick, J. (1985), "The Em Algorithm for Cox's Regression Model Using Glim," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 34, 148-156.

Cox, D. R. (1972), "Regression Models and Life-Tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, 34, 187-220.

de Reijke, T. M., de Boer, E. C., Schamhart, D. H., and Kurth, K. H. (1997), "Immunostimulation in the Urinary Bladder by Local Application of Nocardia Rubra Cell Wall Skeleton Preparation (Rubratin) for Superficial Bladder Cancer Immunotherapy--a Phase I/Ii Study," *Urol Res*, 25, 117-120.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data Via the Em Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1-38.

Elsasser-Beile, U., Ruhnau, T., Freudenberg, N., Wetterauer, U., and Mengs, U. (2001), "Antitumoral Effect of Recombinant Mistletoe Lectin on Chemically Induced Urinary Bladder Carcinogenesis in a Rat Model," *Cancer*, 91, 998-1004.

Hougaard, P. (1986), "A Class of Multivanate Failure Time Distributions," *biometrika*, 73, 671-678.

Lamm, D. L., DeHaven, J. I., Riggs, D. R., Delgra, C., and Burrell, R. (1993), "Keyhole Limpet Hemocyanin Immunotherapy of Murine Bladder Cancer," *Urol Res*, 21, 33-37.

Lamm, D. L., DeHaven, J. I., Riggs, D. R., and Ebert, R. F. (1993), "Immunotherapy of Murine Bladder Cancer with Keyhole Limpet Hemocyanin (Klh)," *J Urol*, 149, 648-652.

Lammers, R. J., Witjes, W. P., Janzing-Pastors, M. H., Caris, C. T., and Witjes, J. A. (2012), "Intracutaneous and Intravesical Immunotherapy with Keyhole Limpet Hemocyanin Compared with Intravesical Mitomycin in Patients with Non-Muscle-Invasive Bladder Cancer: Results from a Prospective Randomized Phase Iii Trial," *J Clin Oncol*, 30, 2273-2279.

Pickles, A., et al. (1994), "Survival Models for Developmental Genetic Data: Age of Onset of Puberty and Antisocial Behavior in Twins," *Genet Epidemiol*, 11, 155-170.

Pirzada, M. T., et al. (2017), "Outcomes of Bcg Induction in High-Risk Non-Muscle-Invasive Bladder Cancer Patients (Nmibc): A Retrospective Cohort Study," *Cureus*, 9, e957.

Singh, R., et al. (2016), "Epithelial-to-Mesenchymal Transition and Its Correlation with Clinicopathologic Features in Patients with Urothelial Carcinoma of the Bladder," *Clin Genitourin Cancer*.

Vaida, F., and Xu, R. (2000), "Proportional Hazards Model with Random Effects," *Stat Med*, 19, 3309-3324.

Walsh, W. G., Tomashefsky, P., Olsson, C. A., and deVere White, R. (1983), "Keyhole-Limpet Haemocyanin (Klh) Immunotherapy of Murine Transitional Cell Carcinoma," *Urol Res*, 11, 263-265.

Yashin, A. I., and Iachine, I. A. (1995), "Survival of Related Individuals: An Extension of Some Fundamental Results of Heterogeneity Analysis," *Math Popul Stud*, 5, 321-339, 377.

Yashin, A. I., and Iachine, I. A. (1999), "What Difference Does the Dependence between Durations Make? Insights for Population Studies of Aging," *Lifetime Data Anal*, 5, 5-22.

Yashin, A. I., Vaupel, J. W., and Iachine, I. A. (1995), "Correlated Individual Frailty: An Advantageous Approach to Survival Analysis of Bivariate Data," *Math Popul Stud*, 5, 145-159, 183.

Zahl, P. H., and Tretli, S. (1997), "Long-Term Survival of Breast Cancer in Norway by Age and Clinical Stage," *Stat Med*, 16, 1435-1449.

Appendix A. Data summary

Name	Type	Label
csdttur	Date	Date of TUR
Hospital	Numeric	Trial centres
osgrade	Numeric	Tumor grade
ospttur	Numeric	Tumor stage
hospital	Numeric	trial centre id
offstdt	Date	date off study
offstreas	Numeric	Reason off study
offstrec	Numeric	Reason off study due to recurrence?
cnumb	Numeric	Number of tumor (single, multiple)
osdttur	Date	date of TUR at entry
Patino	Numeric	patient id
randdt	Date	date of randomization
rectime	Date	date of recurrence (or censoring)- date TUR in study (months)
rectr	Numeric	recurrence during treatment?
primrec	Numeric	Primary or recurrent on study
EPrec	Numeric	recurrence?
recprog	Numeric	Progression?
Sex	Numeric	gender
sizecat	Numeric	Size of tumor (< 15mm, otherwise)
Treat	Numeric	treatment actually received
treatment	Numeric	treatment assigned

Appendix B. Categorized variable age is not a significant predictor for the RFS outcome.

	Regular CPH				Shared Frailty[#]			
Variable	Haz. Ratio	p-value	95% CI		Haz. Ratio	p-value	95% CI	
Treatment (reference, KLH)								
MM	0.42	<0.0001	0.32	0.54	0.43	<0.0001	0.33	0.56
Age (reference, 20-50)								
50-60	1.20	0.59	0.61	2.37	1.14	0.71	0.58	2.22
60-70	1.28	0.45	0.67	2.47	1.20	0.59	0.63	2.29
70-80	1.23	0.53	0.64	2.34	1.20	0.57	0.63	2.28
80-100	1.79	0.12	0.85	3.74	1.64	0.19	0.79	3.40
Age *	1.00	0.54	0.99	1.02	1.00	0.52	0.99	1.02

[#], LR-ratio test for theta, 8.4, p-value 0.002; *, Age is a continuous variable.

Appendix C. Analysis codes

R:

//EPrec is the censor variable 1== failure, 2 == censored

frailtyPenal(Surv(jyRecWeek, censor) ~ as.factor(SEX) + as.factor(TREAT) + as.factor(cnumb)+as.factor(sizecat)+as.factor(OSPTTUR)+as.factor(OSGRADE)+cluster(Hospital), data=dat1, n.knots=20,kappa=10000, RanDist="LogN")

Stata12:

*data exploration:

use "E:\frailty model\data\project.dta", clear

. stset jyRecWeeks RECUR

. rename treatment regimen

. label define regime 1 "KLH" 2 "MM"

. label values regime regime

stdescribe

. stsum, by(regimen)

. sts list, by(regimen) compare

. sts graph, by(regimen)

. sts test regimen

. sts test regimen, wilcox

. stcox age i.regimen

. stcox age i.Hospital i.regimen

. stcox age i.Hospital i.regimen, vce(robust)

*building models

. encode(cnumb), generate(cnumb2)

. generate recevent = .

. replace recevent = 1 if (EPrec == 1)

. replace recevent = 0 if (EPrec == 2)

. stset jyRecWeeks recevent

**regular CPH model

. stcox age SEX PRIMREC sizecat cnumb2 i.OSPTTUR i.OSGRADE i.regimen i.Hospital

**shared frailty model

. stcox age SEX PRIMREC sizecat cnumb2 i.OSPTTUR i.OSGRADE i.regimen, shared(Hospital)

**multiple outcomes, (recurrence, progression)

*import recprog data, recprogstatus lists events 1=rec 2=prog, status=event type (1=failure 0=censored)

. stset jyRecWeeks, failure(status)

. stcox i.treatment age SEX PRIMREC sizecat cnumb2 i.OSPTTUR i.OSGRADE i.Hospital, nohr efron robust strata(recprogstatus) cluster(PATNO)

. stcox i.treatment age SEX PRIMREC sizecat cnumb2 i.OSPTTUR i.OSGRADE, shared(Hospital)

*stratify on treatment

. bysort treat: stcox PRIMREC cnumb2 i.OSPTTUR, shared(Hospital)

*residual plots

. quietly stcox i.treatment age SEX PRIMREC sizecat cnumb2 i.OSPTTUR i.OSGRADE, nohr shared(Hospital) mgale(mg)

. predict cs, csnell

. stset cs, failure(recevent)

. sts generate H= na

. line H cs cs, sort xlab(0 1 to 3) ylab(0 1 to 3)

. quietly stcox i.treatment PRIMREC cnumb2 i.OSGRADE, nohr shared(Hospital) mgale(mg)

. predict cs, csnell

. stset cs, failure(recevent)

. sts generate H= na

. line H cs cs, sort xlab(0 1 to 3) ylab(0 1 to 3)

. quietly stcox i.treatment PRIMREC cnumb2 i.OSGRADE , nohr mgale(mg)

. predict cs, csnell

. stset cs, failure(recevent)

. sts generate H= na

. line H cs cs, sort xlab(0 1 to 3) ylab(0 1 to 3)

*check time-dependence of the predictors

. stcox i.treatment age SEX PRIMREC sizecat cnumb2 i.OSPTTUR i.OSGRADE, nohr tvc(treatment age SEX PRIMREC sizecat cnumb2 OSPTTUR OSGRADE) texp(ln(_t))

*proportionality tests and plots

. quietly stcox treatment age SEX PRIMREC sizecat cnumb2 OSPTTUR i.OSGRADE, shared(Hospital) schoenfeld(sch*) scaledsch(sca*)

. stphtest, detail

. stphtest, plot(age) msym(oh)

. stphtest, plot(SEX) msym(oh)

. stphtest, plot(treatment) msym(oh)

. stphtest, plot(PRIMREC) msym(oh)

. stphtest, plot(OSPTTUR) msym(oh)

. stphtest, plot(cnumb2) msym(oh)

. stphtest, plot(sizecat) msym(oh)

. stphtest, plot(OSPTTUR) msym(oh)

. stphplot, by(treatment) plot1(msym(oh)) plot2(msym(th))

. quietly stcox treatment age SEX PRIMREC sizecat cnumb2 OSPTTUR i.OSGRADE i.Hospital, schoenfeld(sch*) scaledsch(sca*)

. stphtest, detail

. stphtest, plot(age) msym(oh)

. stphtest, plot(SEX) msym(oh)

. stphtest, plot(treatment) msym(oh)

. stphtest, plot(PRIMREC) msym(oh)

. stphtest, plot(OSPTTUR) msym(oh)

. stphtest, plot(cnumb2) msym(oh)

. stphtest, plot(sizecat) msym(oh)

. stphtest, plot(OSPTTUR) msym(oh)

. stphplot, by(treatment) plot1(msym(oh)) plot2(msym(th))