

Supplementary material for: Survival Following Pancreatoduodenectomy in England: Perspectives from the HES Database

Zhangdaihong Liu

September 23, 2015

S1 General Analysis

In this section, Figure S1 shows the number of PDs operated each year from 2001 to 2014; Figure S2 is the number of surgeons that operating PDs each year from 2001 to 2014; Figure S3 is the number of centres that providing PDs each year from 2001 to 2014 and Figure S4 is the boxplot of centre volume between 2001 and 2014. Figure S5 shows the changes in total number of PDs in each volume group over the 14-year study period. Figure S1 - S5 are plotted from the data of whole PD cohort.

Figure S6 displays the 90-day mortality rate on pancreatic cancer patients during the study period. The low starting point (year 2001) might be caused by the incomplete data used in that year (study period starts in March, 2001).

Figure S7 is the annual 90-day mortality rate for each centre volume group. The sharp drops in dark blue and red lines are due to the small or zero sample size in the respective years. This is because in the last few years of study period, there were very few centres falling into low and low-medium volume groups.

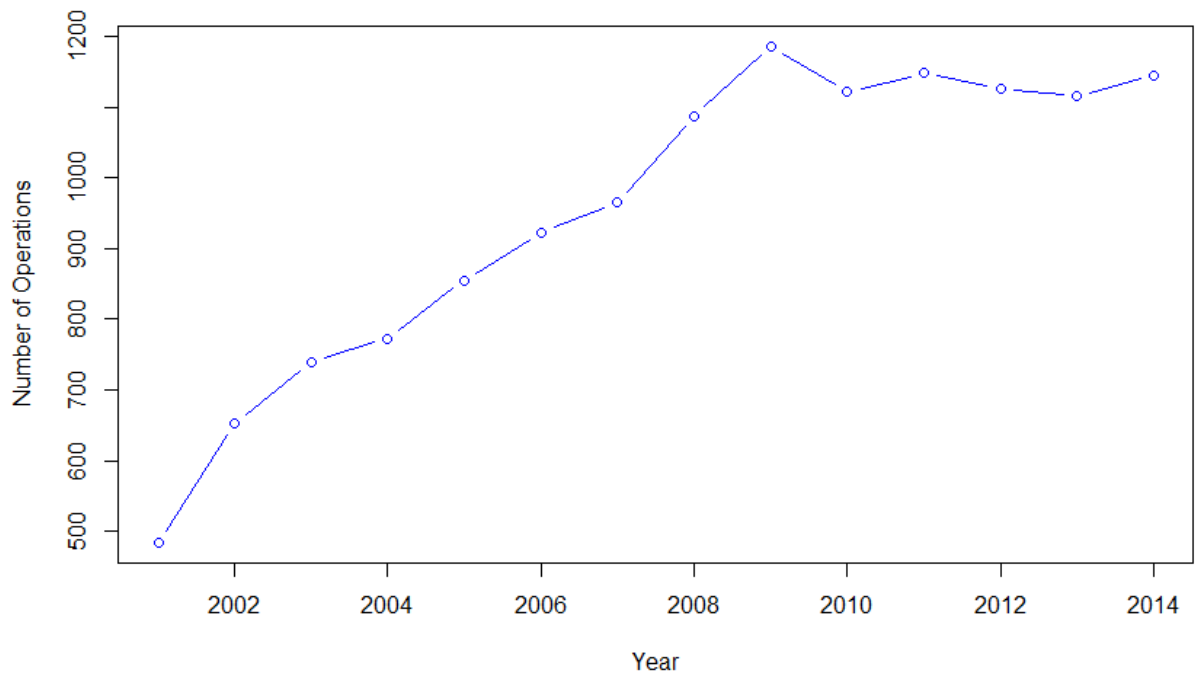


Figure S1: Number of PDs operated each year from 2001 to 2014

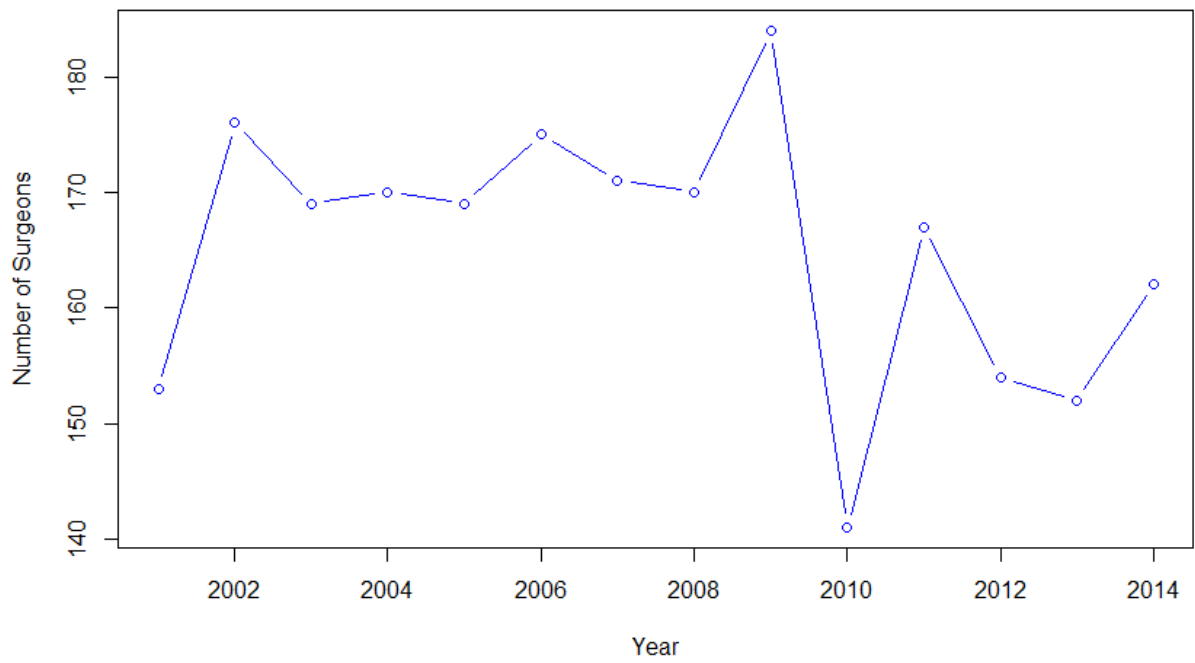


Figure S2: Number of surgeons that operating PDs each year from 2001 to 2014

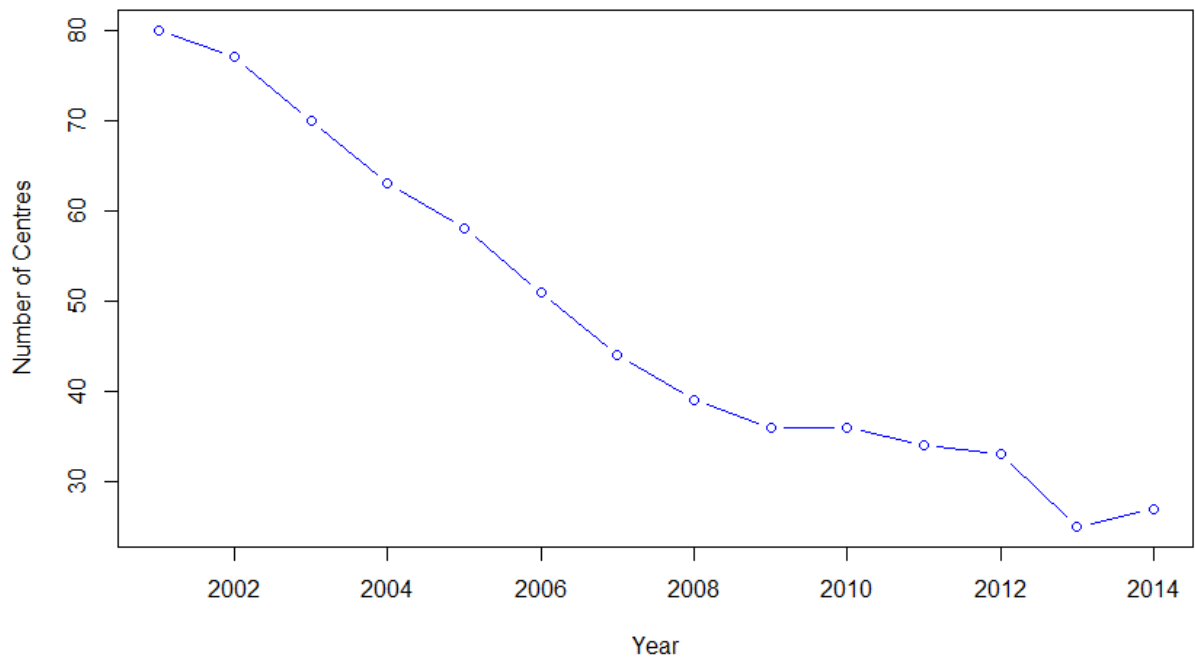


Figure S3: Number of centres that providing PDs each year from 2001 to 2014

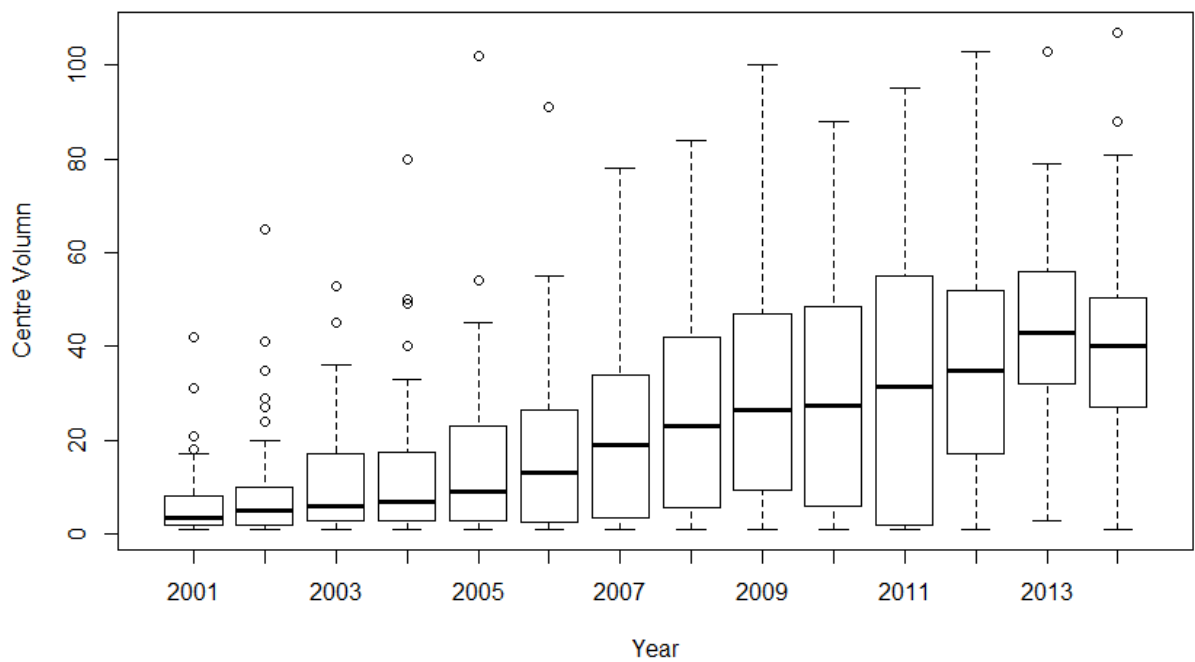


Figure S4: Boxplot of centre volume from year 2001 to 2014

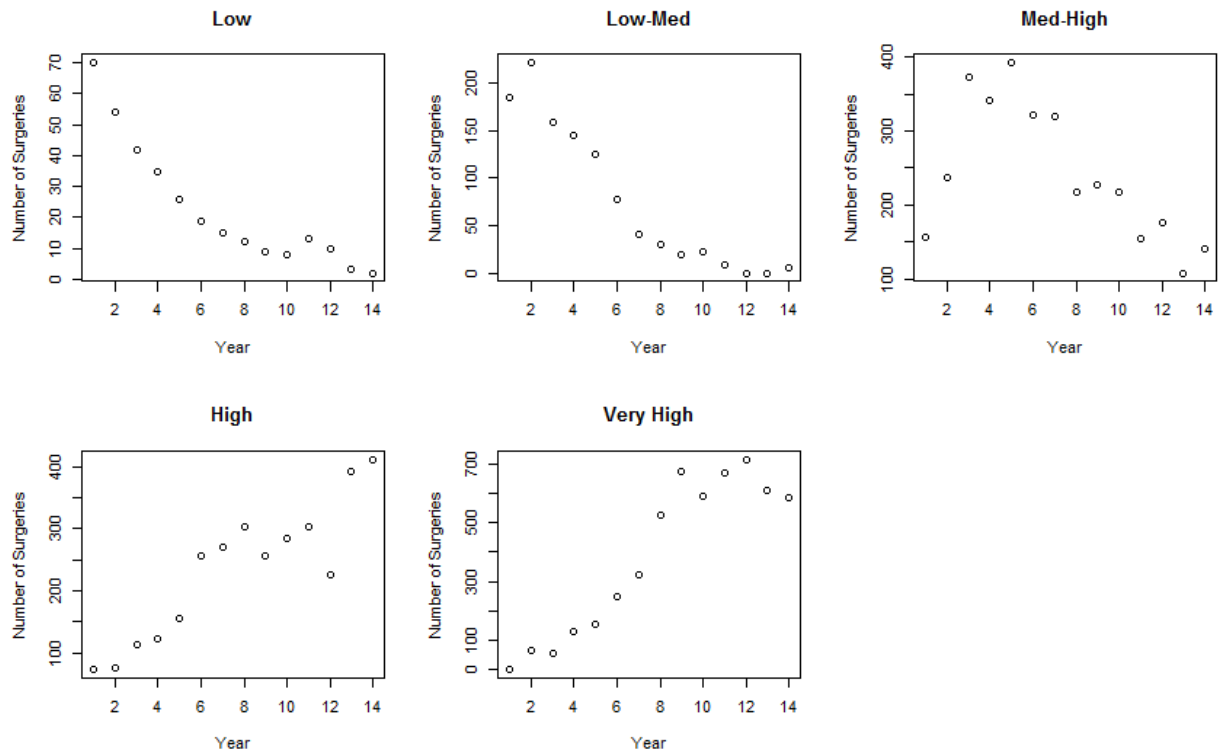


Figure S5: Number of surgeries fall into every centre group each year from 2001 to 2014. The x-axis is the n^{th} year of study period, $n = 1, \dots, 14$.

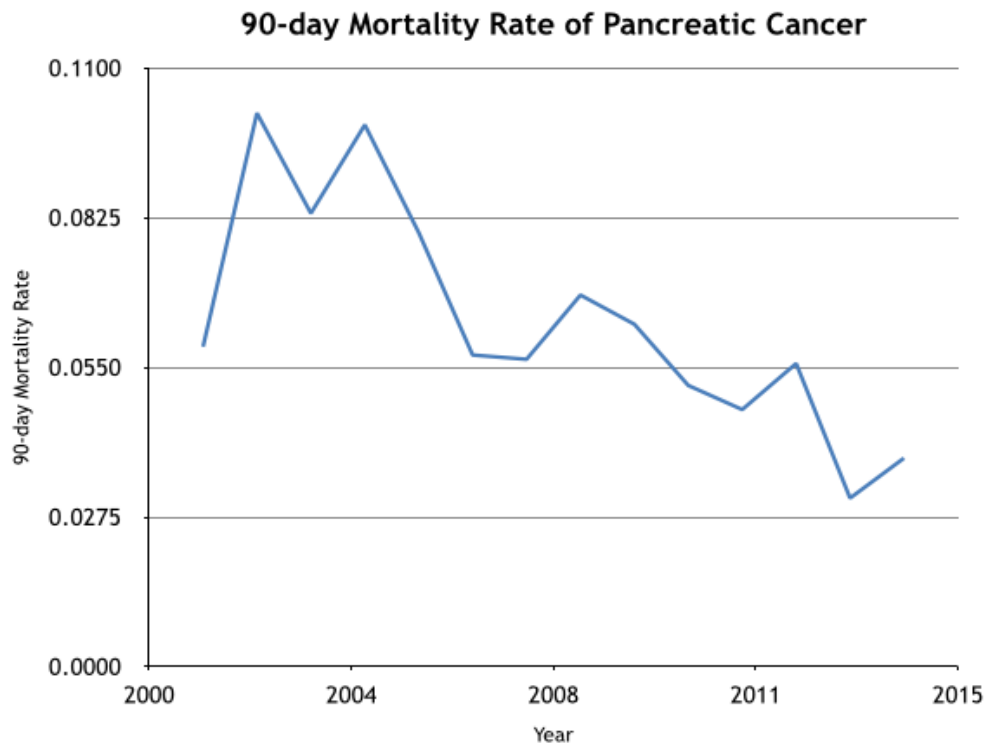


Figure S6: 90-day mortality rate of pancreatic cancer from 2001 to 2014

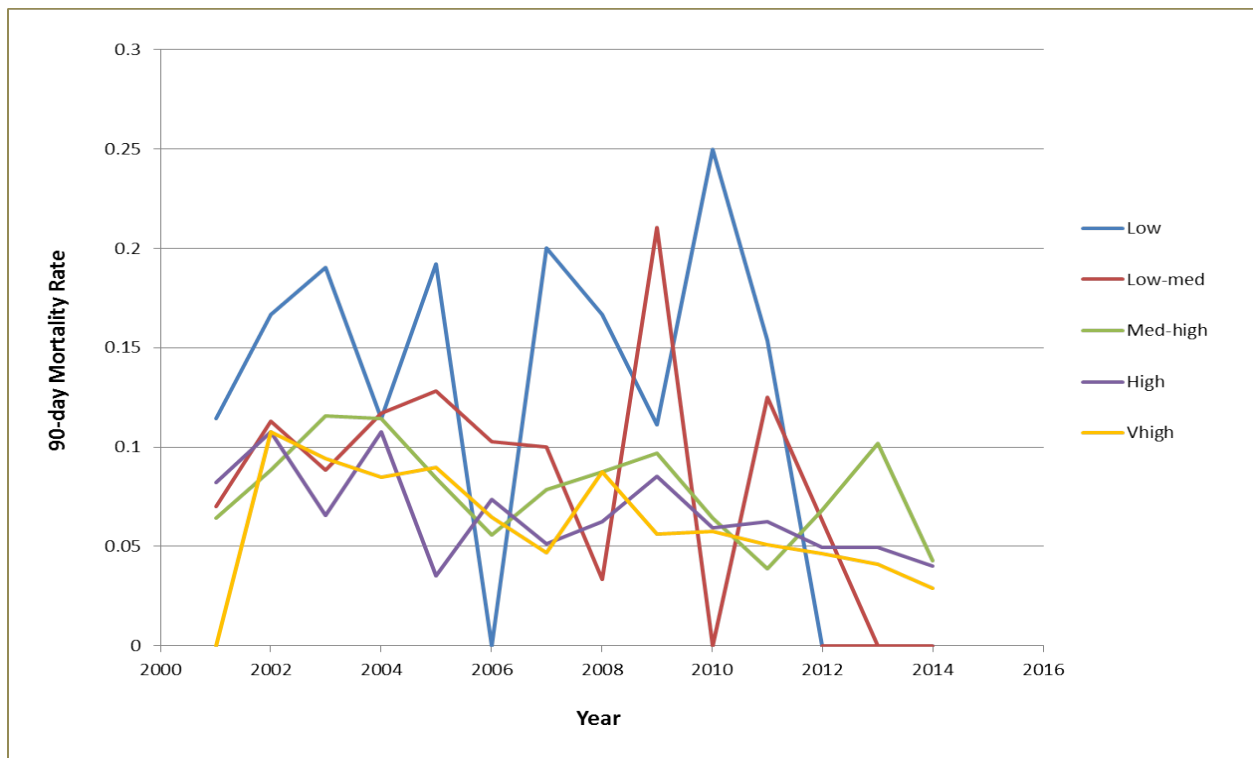
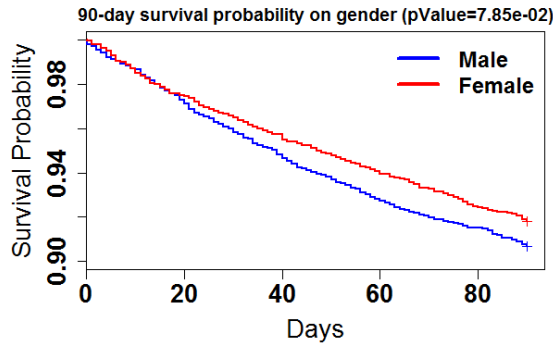


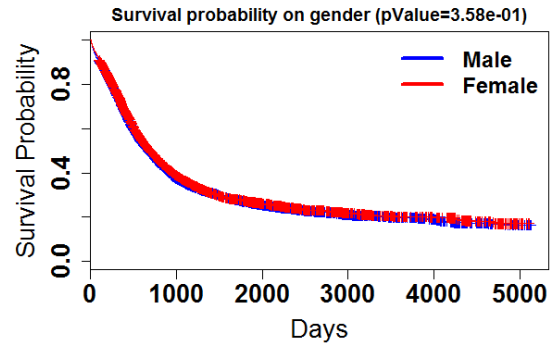
Figure S7: 90-day mortality rate on centre volume groups between 2001 and 2014. The drops on blue and red line are caused by small sample size in respective groups and years.

S2 Kaplan-Meier Curves

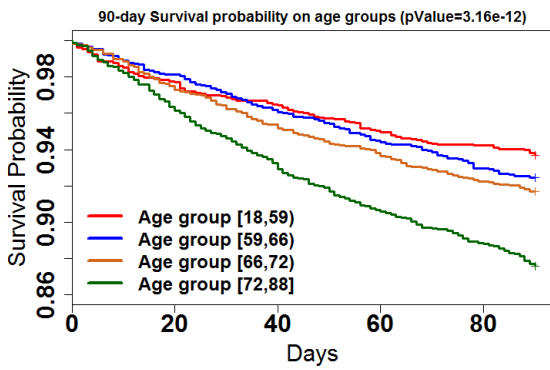
The following graphs are the Kaplan-Meier curves for all the 6 selected predictors. The 7th group is the Kaplan-Meier curves for centre volume on all PD patients. They were used for univariate survival analysis and the p-value obtained from log-rank test is attached in the plot's title. On the left are the 90-day survival curves; On the right are the long-term survival curves.



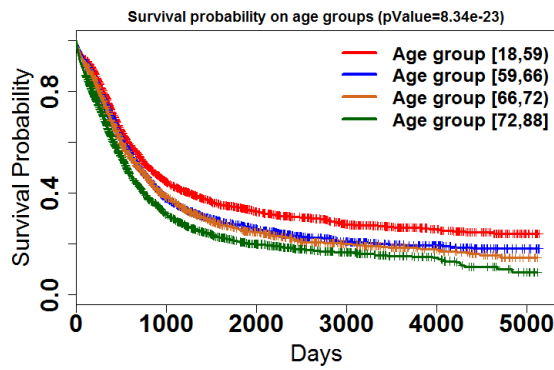
(a) Kaplan-Meier curves of gender for 90-day survival



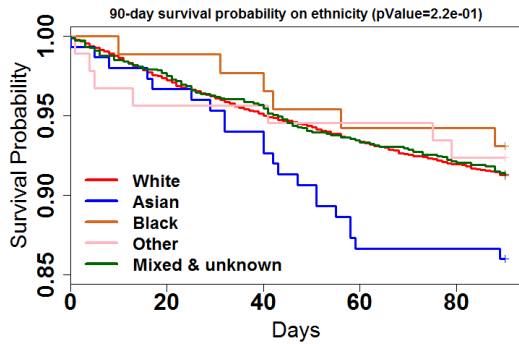
(b) Kaplan-Meier curves of gender for long-term survival



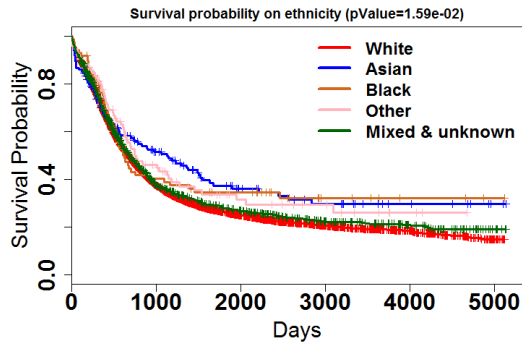
(c) Kaplan-Meier curves of age quartile groups for 90-day survival



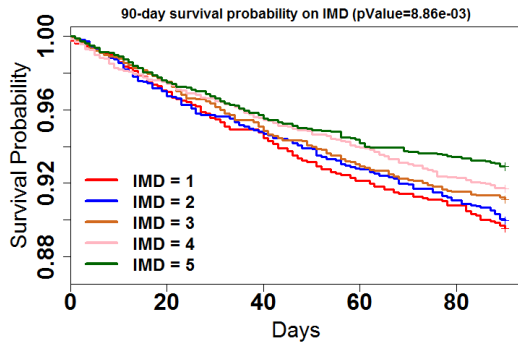
(d) Kaplan-Meier curves of age quartile groups for long-term survival



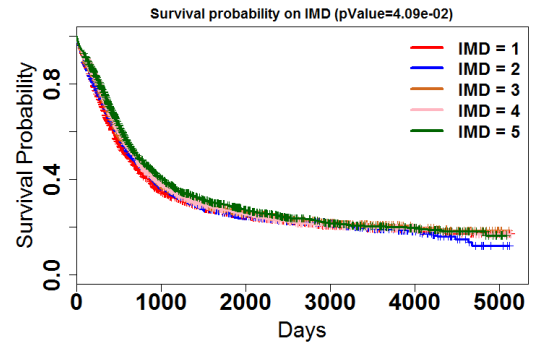
(e) Kaplan-Meier curves of ethnicity for 90-day survival



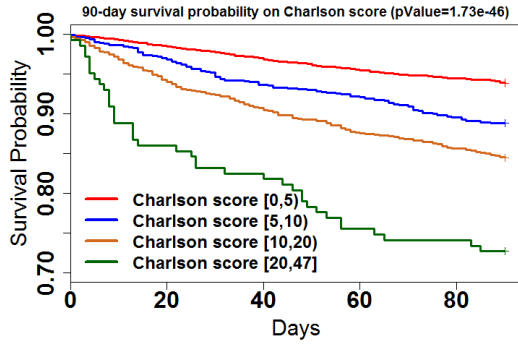
(f) Kaplan-Meier curves of ethnicity for long-term survival



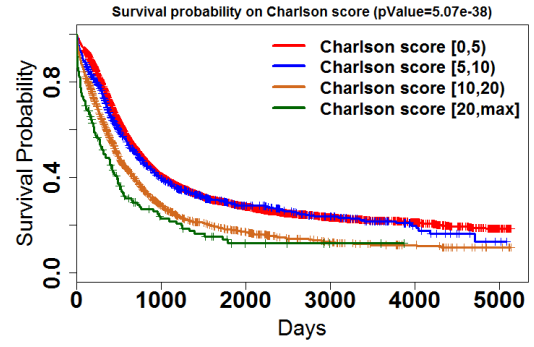
(a) Kaplan-Meier curves of IMD for 90-day survival



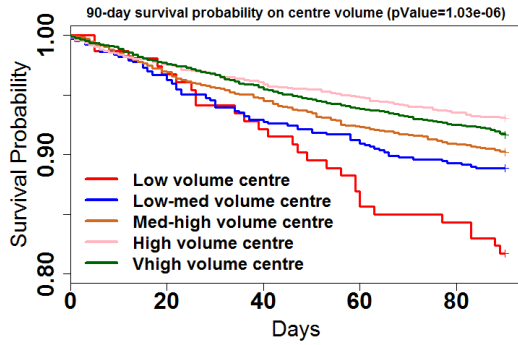
(b) Kaplan-Meier curves of IMD for long-term survival



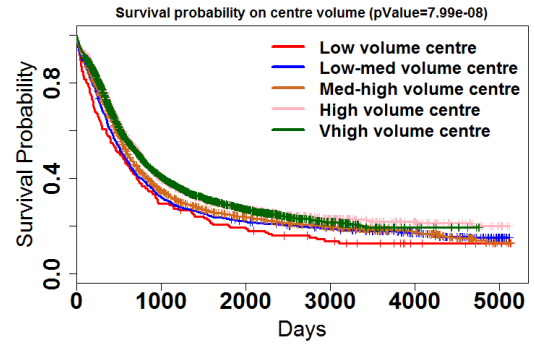
(c) Kaplan-Meier curves of Charlson score for 90-day survival



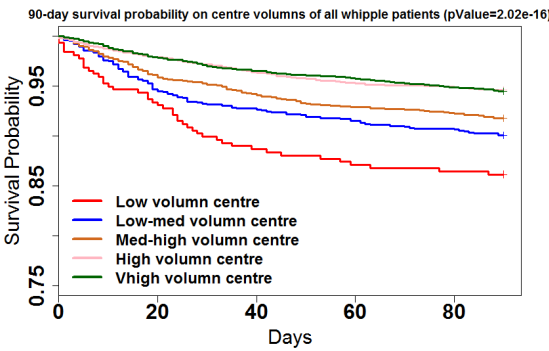
(d) Kaplan-Meier curves of Charlson score for long-term survival



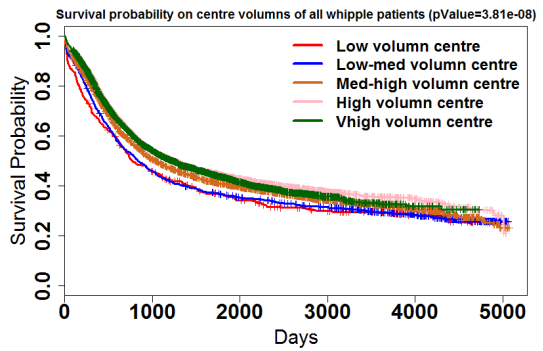
(e) Kaplan-Meier curves of centre volume for 90-day survival on pancreatic cancer patients



(f) Kaplan-Meier curves of centre volume for long-term survival on pancreatic cancer patients



(g) Kaplan-Meier curves of centre volume for 90-day survival on all PD patients



(h) Kaplan-Meier curves of centre volume for long-term survival on all PD patients

S3 Proportional Hazard Assumption Test for Cox Model

PH assumption was tested for all the variables and we found that Charlson score, IMD and centre volume violated this assumption judging from the small p-values. However, due to the large sample size, small fluctuation may easily cause this violation. Therefore, we looked at the graphical explanation (values of effect parameter β) and they are shown in Figure S10. As displayed in graphs (straight line means satisfying the assumption), IMD satisfies the assumption very well. Charlson score shows general straight lines. Centre volume violates the assumption only in the end of the time series. Therefore, they were all considered in the Cox model.

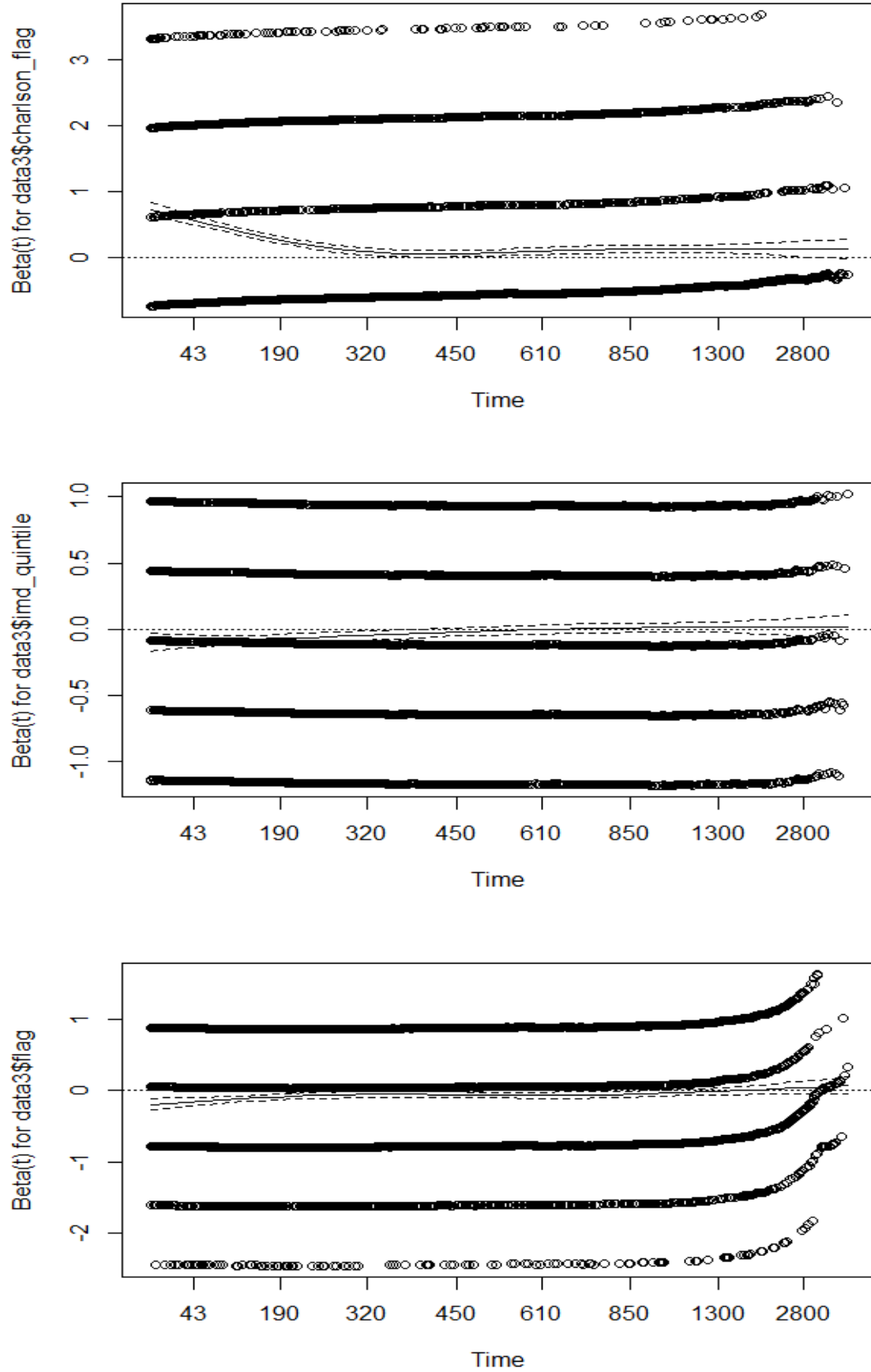


Figure S10: Graphical illustrations of PH assumption in Cox model for Charlson score (categorical), IMD and centre volume (categorical) respectively. Flat straight lines mean satisfaction of the assumption. Charlson score (top) shows a slope to the right up, but the slope angle is small and the lines are generally straight. IMD (middle) shows quite good standard of satisfaction. Centre volume (bottom) satisfies the assumption quite well until time point 1300, but violates it in the end.

S4 90-day Survival Predictions

Model fitting was also done to predict 90-day survival using subset of patients whose survival time is no longer than 90 days. Since it is not clinically very interesting, it has been put into the supplementary. This cohort has 706 patients and all died within 90 days which means the response variable is no longer censored. Cox model, GLM and random forests model were fitted to this subset of data with the same method (3-fold cross validation) and variables as used in long-term survival predictions. AIC and BIC were applied in the Cox model. However, due to the strictness of BIC and the less information provided, for one set of cross validation, BIC did not select any significant variable. Therefore, only predictions acquired from AIC were considered. The C-index was calculated for each model to compare the outcomes and is shown in Table 1. The predictive level has decreased in Cox model and GLM, and increased in random forests model.

Table 1: 90-day survival predictions

	Cox model	GLM	Random forests model
C-index	0.5230574	0.5221286	0.5136892
Standard error	0.01282064	0.01282916	0.01279546