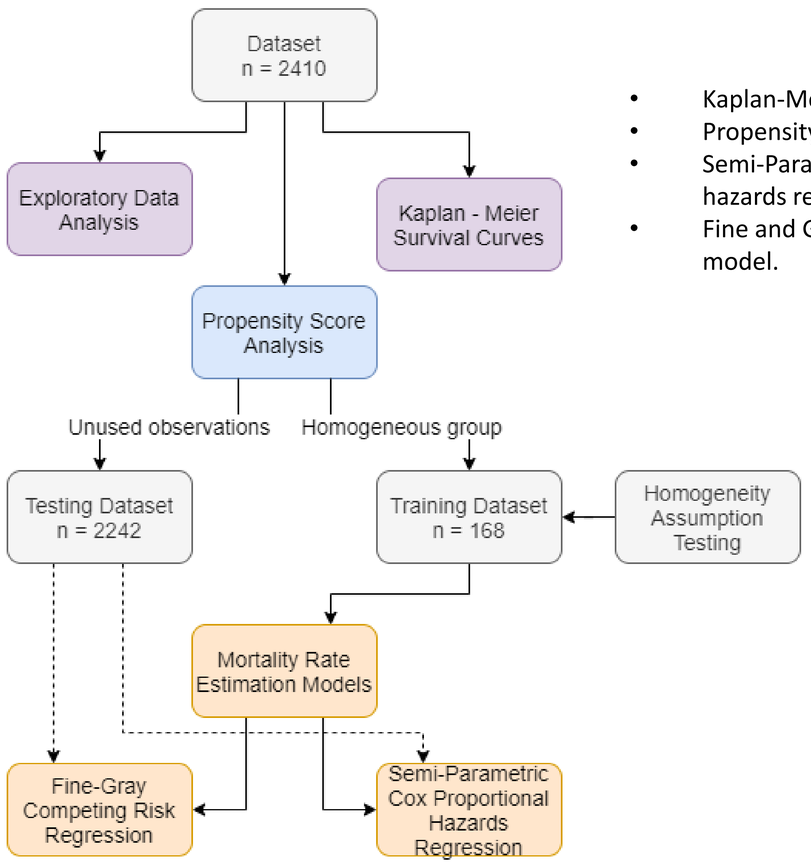
Mortality rate estimation models of patients with prostate cancer diagnosis

kaunas university of technology
ktu 1922

ABSTRACT

Prostate cancer is one of the most frequent type of male cancer all around the world, including Lithuania. Prostate cancer diagnosis takes up to 25% of all lithuanian men cancer diagnosis in the country. Every year, ~3000 new prostate cancer cases are diagnosed and ~500 lithuanian men die from this illness. It is necessary for a medical professional to be able to distinguish a fatal and non-fatal cancer in time, statistical methods and models could be implemented to help with this case and in our work we will try to implement those methods and models using data from “Kauno Klinikos” clinic (Kaunas, Lithuania). During the research we used well known Kaplan-Meier survival curves as well as compared 2 best known hazard estimation models: Cox and Fine-Gray models. In this dataset, there were 56 deaths reported from cancer specific causes and 294 from other causes, the median age of a patient was 64 years (n - 2410). During the analysis of Kaplan-Meier survival curves, we found the worst survival prognosis associated with patients, who’s lymph nodes were damaged by cancer or patients with 5 metastatic lymph nodes. It was also discovered that patients with 1 or 2 metastatic lymph nodes are much more likely to experience death from one of the causes – cancer specific cause or other causes than men with 4 metastatic lymph nodes. Significant hazard ratio was also found between men, who’s cancer is developing in the prostate area and men, who’s cancer has already spread outside the prostate cancer, with the later one having the worse survival prognosis. Patients with cancer damaged lymph nodes have a higher mortality rate than men with untreated lymph nodes only from cancer specific causes while the hazard ratio linked with man’s age was found significant only in deaths from other causes. Fine and Gray hazard estimation model distinguished less significant risk factors and usually the hazard ratios were reported smaller than the ones in the Cox model. Training and testing datasets were used to test the performance of both models and Cox model was found to be optimal on both datasets in response to the ROC curves analysis.

MATHEMATICAL METHODS



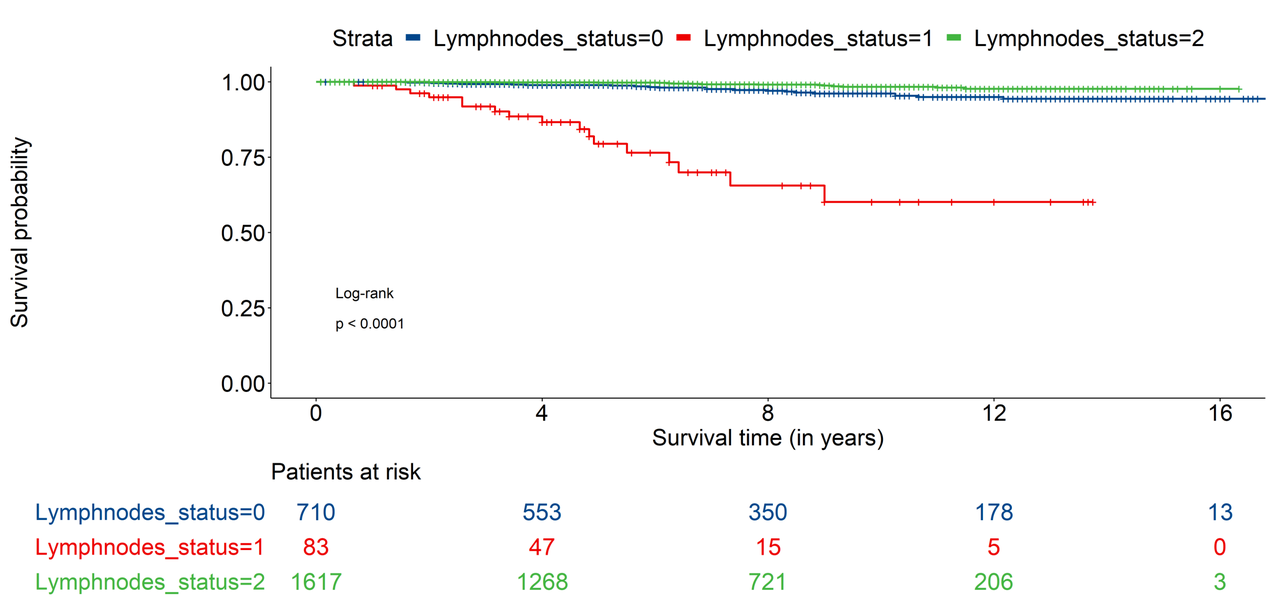
- Kaplan-Meier survival curves;
- Propensity score analysis;
- Semi-Parametric Cox proportional-hazards regression model;
- Fine and Gray competing risk regression model.

KAPLAN-MEIER SURVIVAL CURVES

Survival function:

$$S(t_i) = S(t_{i-1})\left(1 - \frac{d_i}{n_i}\right)$$

where n_i – survived (censored) individuals up to time t_i; d_i – number of events (e.g., deaths) that happened at time t_i.



Lymph nodes status: 0 – not damaged by cancer; 1 – cancer damaged; 2 – untreated lymph nodes

SEMI-PARAMETRIC COX PROPORTIONAL-HAZARDS REGRESSION MODEL

Hazard function:

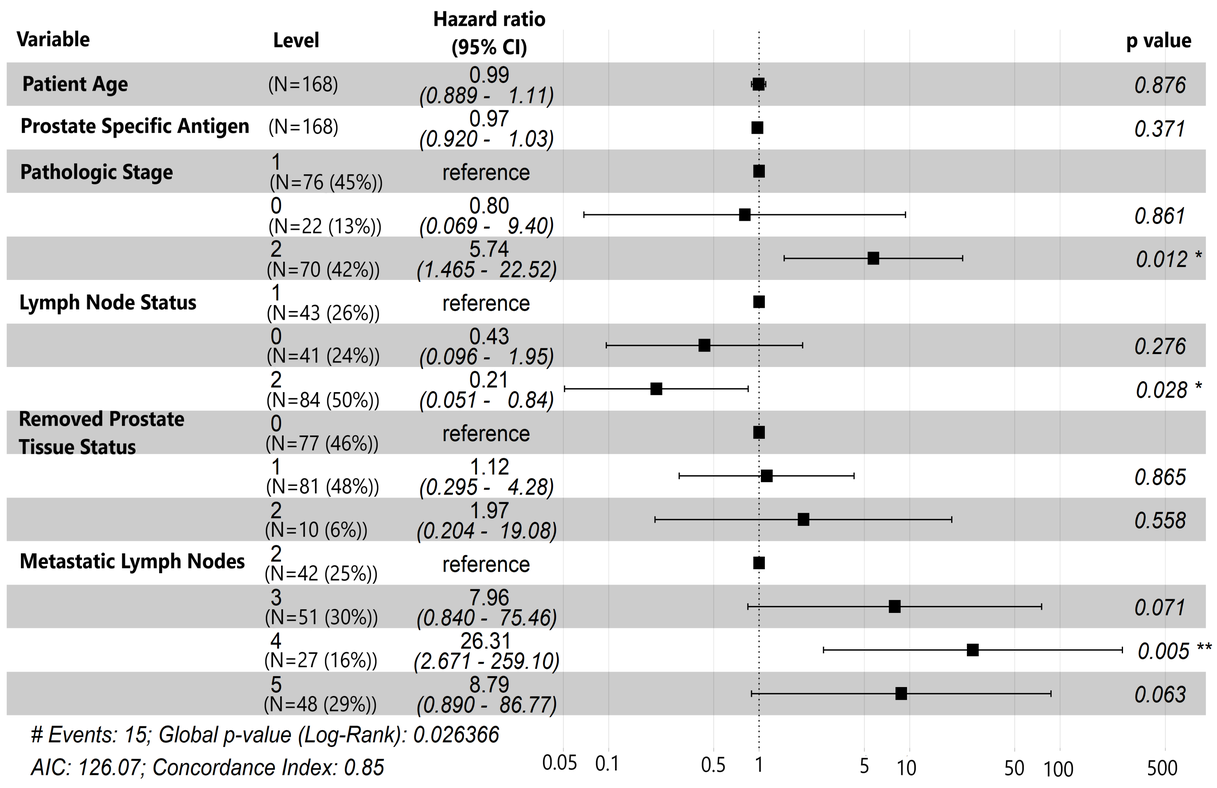
$$h_i(t, X) = h_0(t) \cdot e^{\{\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}\}}$$

where i – i th observation, t – survival time; $(\beta_1, \beta_2, \ldots, \beta_k)$ – coefficients of covariates; $(X_{i1}, X_{i2}, \ldots, X_{ik})$ – covariates; $h_0(t)$ – baseline hazard function.

Hazard ratio between two groups:

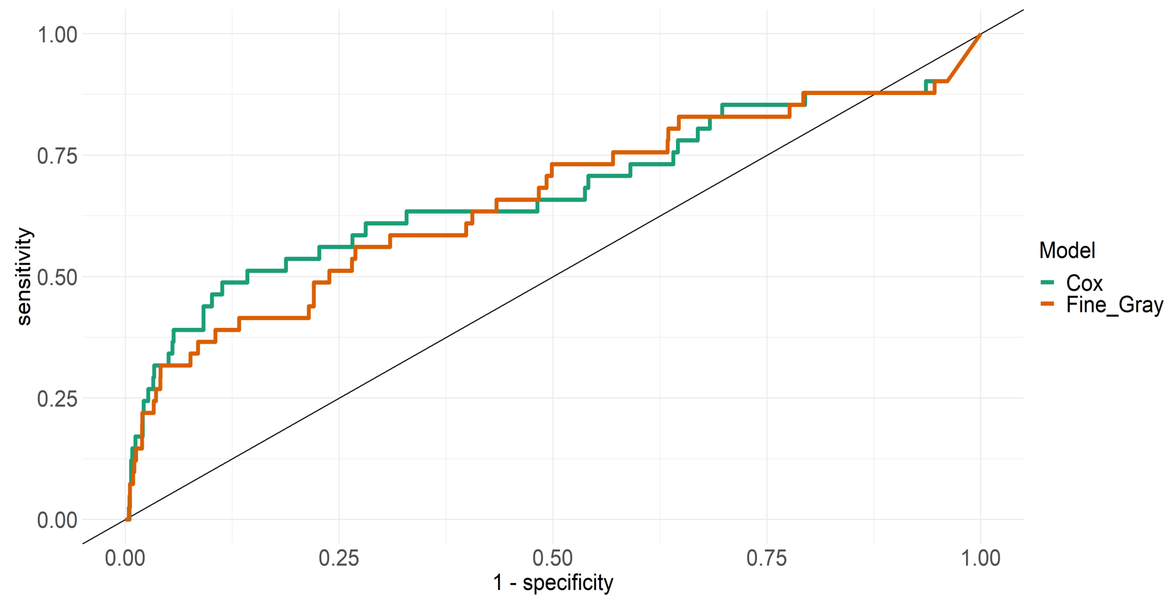
$$HR = \frac{B\ group\ hazard}{A\ group\ hazard} = \frac{h_B(t, X)}{h_A(t, X)}$$

- **HR > 1:** Hazard in group B is stronger than group’s A,
- **HR < 1:** Hazard in group B is weaker than group’s A,
- **HR = 1:** Hazards have the same effect.



COMPARISON OF MODELS

ROC curves on Test dataset



Dataset	Model	AUC
Training	Cox	0,7708
	Fine-Gray	0,7673
Testing	Cox	0,6747
	Fine-Gray	0,6578

CONCLUSIONS

- Men, who’s cancer is developing in the prostate area are on average 5.7 times less likely to die from prostate cancer and 3.4 times less likely to die from other causes than men, who’s cancer has already spread outside the prostate area.
- Cox model was found to be more accurate than Fine-Gray model.

V. Kraujalis, T. Ruzgas, D.Milonas