Baltic J. Modern Computing, Vol. 10 (2022), No. 2, 170-184
https://doi.org/10.22364/bjmc.2022.10.2.06

Mortality Rate Estimation Models for Patients with Prostate Cancer Diagnosis

Vytautas KRAUJALIS[1], Tomas RUZGAS[1], Daimantas MILONAS[2]

[1]Department of Applied Mathematics, Kaunas University of Technology, K. Donelaicio str. 73, LT-44249 Kaunas, Lithuania
[2]Medical Academy, Department of Urology, Lithuanian University of Health Sciences, A. Mickeviciaus str. 9, LT-44307 Kaunas, Lithuania

vytautas.kraujalis@ktu.edu, tomas.ruzgas@ktu.lt, daimantas.milonas@lsmuni.lt

Abstract. Malignant tumors remain third main reason of Lithuanian men deaths. Malignancy of the prostate gland in Lithuania is the most common oncological illness among the men. The mortality from this disease in other European countries tends to be decreasing but Lithuania still experiences annual increase in prostate cancer mortality. Article presents models to assess the long-term mortality risk among patients with diagnosed malignant tumor of the prostate gland. The latter models complement popular decision support methods in clinical practice used so far, gives significant information about prostate malignant tumors and their aggressiveness potential. Cox proportional hazards regression and Fine-Gray competing risk regression comparison showed that Cox method excels not only in higher accuracy and AUC value, but also in higher sensitivity which is particularly important in clinical trials.

Keywords: prostate cancer; competing risk analysis; proportional hazards model

1. Introduction

In 1853, J. Adams, a surgeon at The London Hospital, described the first case of prostate cancer (Adams, 1853). Adams noted in his report that this condition was "a very rare disease". Remarkably, 150 years later, prostate cancer has become a significant health problem – second by frequency and sixth by men mortality oncological disease in the whole world (Center et al., 2019). In the recent 3 decades, a number of men with prostate cancer diagnosis has increased 5 times in Lithuania. A significant increase in prostate cancer diagnosis can be explained by introduction of the Early Prostate Cancer Detection Programme (EPCDP) in 2006. Overall during the 2001 – 2017 period prostate cancer has always been the third most common Lithuanian men mortality cause and made up to 20.8% of all men deaths. Every year 2.7% of men aged 45 and older experienced death from prostate cancer. Though the overall men mortality rate (standardized) since 2007 decreased almost every year, cancer specific mortality did not change in the 2001 – 2017 period but the mortality from prostate cancer increased by 1% each year (Meksriunaite et al. 2018).

Comparing Lithuanian and other European Union (EU) countries men prostate cancer specific mortality standardized rates during 2011 – 2015 period, each year

Lithuania exceeds the EU average by 42 – 72%. Men mortality rate from prostate cancer during this period was only higher in Estonia and Latvia, in some years, the mortality rate was higher in Slovenia (2011, 2015), Norway (2011, 2013 – 2015), Sweden and Denmark (2011, 2013, 2014), Iceland (2013) and Liechtenstein (2011). Other EU countries during the 2011 – 2015 period had a lower men prostate cancer mortality rates than Lithuania[1]. In 2020 first European Urology issue, the newest data on prostate cancer morbidity and mortality rates across the globe were presented. In (Culp et al., 2020) it is published that in 2008 – 2012 men morbidity of prostate cancer rates in Lithuania were highest in the world. Factors such as age, genetics, family health history and other surrounding factors highly affects the causes of prostate cancer emergence (Parkin et al., 2011). An important note on prostate cancer is that early diagnosis is crucial as most of the time, an early treatment proceeds to a very good results.

Correct data affects the reliability of analysis on cancer diagnosis and decision making. Storage of medical information and statistical analysis has been used ever since the middle ages. The first medical statistical journal was published in London, 1662 (Connor, 2022). In 1863, F. Nightingale, the pioneer of nowadays nursing raised a problem, related to the medical statistical data collection and unstructured storage in hospitals. According to F. Nightingale, the later problem was the consequence of limited financing and ineffective treatment. In 1977, USA congress published a scientific paper on the benefits of medical information system[2]. This paper states that unified medical information system can be a beneficial tool not only in trainings but also can help medical and health specialists to acquire more knowledge on healthcare, institutional planning, optimization and management. Besides, a unified medical information system is a valuable tool for researchers and government health institutions. Since the year of 2000, together with information technology breakthrough, a regional and national health record systems were started to be established. The main goal of these systems were to save valuable patients data. Information systems of medical data stores structured information about the patient, such as diagnosis, demographics, vital functions, examination results and so on. A clever analysis of patient's records helps with answering questions about quicker disease diagnosis, choosing optimal treatment, application of treatment and prognosis, also estimation of complication risk and optimization of resources in healthcare institutions. Various analytical methods are used for that. One of such groups are risk estimation algorithms (Spruance et al., 2004; Macek et al., 2020). Understanding hazard of mortality and learning from deaths is an important component of good clinical practice but current approaches and measures are complex, controversial and difficult to understand (Stewart et al., 2016).

To evaluate proportional hazards in medical research, it is not uncommon to find classical methods such as either Fine-Gray competing risk model (Fine and Gray, 1999) or semi-parametric Cox regression model (Cox, 1972) being used. In an article which developed multivariate competing-risk model to analyse postoperative prostate cancer-related death (Tosco et al., 2018), authors studied 2823 patients with high risk prostate cancer which is defined as clinical stage T3-4, prostate specific antigen >20 ng/ml or biopsy Gleason score 8-10. Multivariate analysis was used to evaluate pathologic stage, pathologic Gleason score, lymph nodes status and surgical margin status as independent predictors of prostate cancer-related death. Later authors applied a multivariate model to evaluate pT ≥ 3b (pathologic stage), pN1 (lymph node status) and pGS ≥ 8 (pathologic

[1] WEB, a: Eurostat database, product code „hlth_cd_asdr2"
[2] WEB, b: Report by the US Congress Office of Technology Assessment, 1977

Gleason score) as independent predictors of prostate cancer-related death. Variables were combined into 8 possible combinations and then 3 subgroups which gave the following cumulative risk: good prognosis subgroup 7 year cumulative risk is just 0.01 (95% CI 0.00 – 0.05), intermediate – 0.09 (95% CI 0.05 – 0.15) and poor – 0.18 (95% CI 0.11 – 0.29). Another article analyzed grouping system of Gleason score (Berney et al., 2016) which is based on the International Society of Urological Pathology (ISUP). Analysis was performed on men aged <76 during the diagnosis of prostate cancer. Authors used semi-parametric Cox proportional hazards regression model with the response – prostate cancer-related death. Multivariate analysis included overall and worst Gleason score (GS) clinical T stage, prostate specific antigen, volume of disease and method of treatment. Grouping into 5 prognostic groups was made using the new proposed grading system (Epstein et al., 2016) with both overall and worst GS. Both overall and worst GS in five grade groups showed significant results. Article presents that when GS is grouped into five grade groups on either overall or worst GS, all the hazard ratios (keeping 1st grade group as a reference) are significant. Authors conclude that these results, for the first time, shows interpretation of GS using modern criteria as an effective prognostic grade group when prostate cancer death is our outcome.

Lithuania cancer registry database each year adds up more than 1300 prostate cancer (adenocarcinoma) records. Those records could be a useful structured data source for future scientific research. An important factor in such registry databases is long-term patient observation. Usually, separate clinic's databases which were being gathered much earlier are more significantly informative although they include less amount of patient data than national registry (Carlsson et al., 2020). Although Lithuania's morbidity of prostate cancer in the recent year is declining, the mortality rate is increasing. One of the main reasons is late diagnosis. Regular examination of prostate specific antigen level in blood can decrease the amount of untreated and already spread prostate cancer. The main survivability factor is early diagnosis and timely radical treatment of the primary tumour. Surgical removal of the tumour in its early stage increases survival probability for additional 10 years up to 90 – 97% (Braun et al., 2004). However, some prostate cancers are potentially of aggressive form and despite the treatment, patients die because of disease progression. Such cancer form detection and timely adjunctive therapy could prolong survival or even save the life.

The purpose of this paper is to explore the prostate cancer related survival of Lithuanian men and train two mortality rate estimation models – semi-parametric Cox proportional hazards regression and Fine-Gray competing risk regression. To estimate the hazard ratios and identify significant predictors on prostate cancer related mortality and other cause mortality. The two mortality rate estimation models will be trained, tested and compared to each other using statistical methods.

2. Material and Methods

A dataset from „Kauno Klinikos" clinic (Kaunas, Lithuania) consisting of 2410 patients treated by radical prostatectomy (RP) for clinically localized prostate cancer (PCa) between 2001 and 2017 is used. Six characteristics were analysed: age, preoperative prostate specific antigen (PSA), metastatic lymph nodes, pathologic stage (pT), lymph node status (pN) and surgical margin status (SM). For each case of the patient death, either a cancer specific mortality (CSM) or other cause mortality (OCM) is registered.

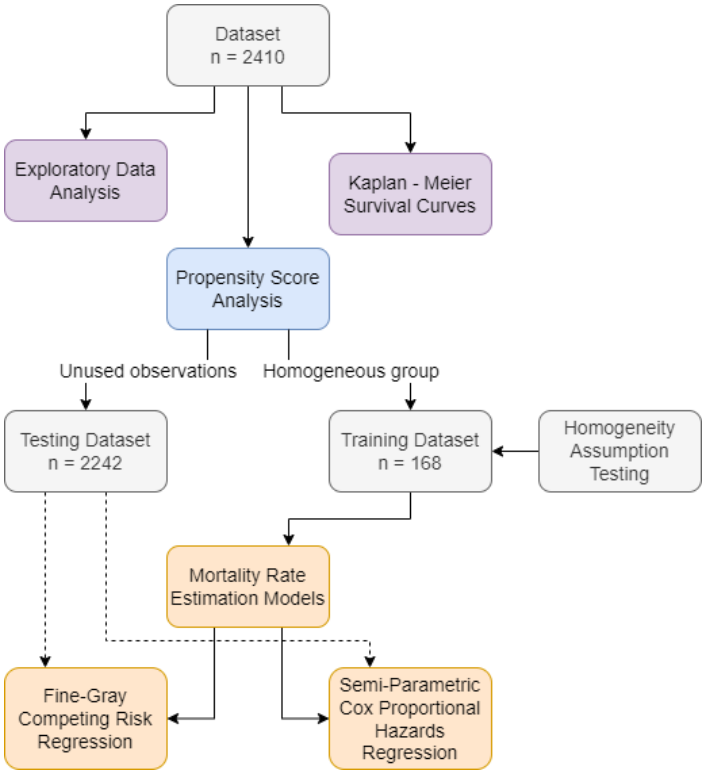

Figure 1. Research workflow diagram.

Descriptive characteristics of variables were reported. For categorical variables frequencies and proportions are given while median, interquartile range (IQR), minimum and maximum values were reported for continuous variables. Research workflow diagram can be seen in Fig. 1. Well known Kaplan-Meier curves were computed to graphically represent survival rates as well as compare cumulative survival of 5 – 10 years after prostate cancer diagnosis. The whole dataset was used to assess cancer specific survival (CSS) and other causes survival (OCS). To test the null hypothesis of no difference in survival curves, log-rank test was performed. After exploratory data analysis, propensity score analysis was chosen to acquire a subset of dataset which is homogeneous by lymph nodes status. Matching of similar observations between 2 groups was done. Finding a match for observation from another group was performed by nearest neighbour algorithm within certain caliper bounds and without replacement (one control observation can be matched only once). Propensity score calculated with logistic regression. Homogeneity assumption testing after obtaining homogenous group of patients was performed. For the continuous variables age and prostate specific antigen grouped by lymph node status Shapiro-Wilk and Kolmogorov-Smirnov tests for normality were used followed by non-parametric Kruskal-Wallis test, chi-squared test was used on metastatic lymph nodes and pathologic stage grouped by lymph node status and Fisher-Freeman-Halton exact test was used on surgical margin status grouped by lymph node status. Finally, two mortality rate estimation models were constructed: a semi-parametric Cox proportional hazards regression and Fine-Gray competing risk

regression. Each variable was assessed using hazard ratios (HR) with significance level set to 5% and receiver operating characteristic (ROC) curves were drawn with calculated area under the curve (AUC) to compare the models. All methods and models were done using R programming language (version 4.0.4 2021-02-15).

3. Results

Descriptive characteristics of patients are presented in Table 1. Median age of patients was 64 years (IQR 59 – 68) with minimum of 40 and maximum of 87 years. Median PSA level is 6.3 (IQR 4.7 – 9.47) with 0.44 as minimum and 98.4 as maximum. Out of all 2410 men, 350 (14.5%) were reported dead with 56 (2.3%) prostate cancer related death and 294 (12.2%) death from other causes. Median follow-up after RP was 7.08 years (IQR 4.17 – 10.42).

Table 1. Descriptive characteristics of 2410 prostate cancer patients.

	(N=2410)
Patient's age	
Median	64.00
25% quantile – 75% quantile	59.00, 68.00
Min - Max	40.00 - 87.00
Prostate specific antigen (PSA), ng/mL	
Median	6.30
25% quantile – 75% quantile	4.70, 9.47
Min - Max	0.44 - 98.40
Metastatic lymph nodes (1, 2, 3, 4, 5+) (GG)	
1, GS (<=6)	630 (26.1%)
2, GS (3+4)	1283 (53.2%)
3, GS (4+3)	250 (10.4%)
4, GS (=8)	109 (4.5%)
5, GS (>=9)	138 (5.7%)
Pathologic stage (pT)	
0, initial stage, cancer is not detectable by tomography	1565 (64.9%)
1, cancer is developing in the prostate area	661 (27.4%)
2, cancer has spread outside the prostate area	184 (7.6%)

Table 2. Cont.

Lymph nodes status (pN)	
0, clean	710 (29.5%)
1, damaged by cancer	83 (3.4%)
2, untreated	1617 (67.1%)
Surgical margin status (SM)	
0, clean	1606 (66.6%)
1, damaged by cancer	698 (29.0%)
2, unknown	106 (4.4%)

GS – Gleason score.

Kaplan-Meier survival analysis was performed on all 2410 patients, survival from prostate cancer and other causes after 5 – 10 years are reported in Table 3.

Table 3. Kaplan-Meier survival rates at 5 and 10 years from prostate cancer and other causes with 95% confidence intervals.

	Survival from prostate cancer		**Survival from other causes death**	
	5 years	**10 years**	**5 years**	**10 years**
Pathologic stage (pT)				
0	0.999 (0.998 – 1)	0.985 (0.977 – 0.994)	0.947 (0.935 – 0.959)	0.847 (0.825 – 0.870)
1	0.990 (0.982 – 0.999)	0.967 (0.948 – 0.987)	0.943 (0.923 – 0.962)	0.841 (0.800 – 0.884)
2	0.883 (0.831 – 0.938)	0.784 (0.707 – 0.869)	0.846 (0.789 – 0.908)	0.779 (0.706 – 0.858)
Lymph nodes status (pN)				
0	0.989 (0.981 – 0.997)	0.961 (0.943 – 0.980)	0.934 (0.914 – 0.953)	0.838 (0.805 – 0.872)
1	0.794 (0.692 – 0.911)	0.601 (0.449 – 0.804)	0.852 (0.764 – 0.949)	0.807 (0.693 – 0.939)
2	0.997 (0.994 – 1)	0.983 (0.974 – 0.992)	0.945 (0.933 – 0.957)	0.843 (0.819 – 0.868)

Table 4. Cont.

	Survival from prostate cancer		Survival from other causes death	
	5 years	**10 years**	**5 years**	**10 years**
Surgical margin status (SM)				
0	0.999 (0.997 – 1)	0.982 (0.973 – 0.992)	0.942 (0.930 – 0.954)	0.846 (0.824 – 0.869)
1	0.967 (0.953 – 0.982)	0.933 (0.910 – 0.956)	0.929 (0.909 – 0.951)	0.828 (0.791 – 0.866)
2	0.978 (0.948 – 1)	0.978 (0.948 – 1)	0.943 (0.899 – 0.988)	0.825 (0.696 – 0.977)
Metastatic lymph nodes (GG)				
1	1 (1 – 1)	0.991 (0.982 – 1)	0.940 (0.922 – 0.959)	0.833 (0.802 – 0.866)
2	0.996 (0.993 – 1)	0.983 (0.972 – 0.993)	0.946 (0.932 – 0.959)	0.851 (0.824 – 0.878)
3	0.996 (0.988 – 1)	0.930 (0.875 – 0.989)	0.947 (0.916 – 0.978)	0.881 (0.801 – 0.970)
4	0.932 (0.881 – 0.987)	0.838 (0.744 – 0.944)	0.866 (0.797 – 0.941)	0.762 (0.661 – 0.878)
5	0.878 (0.812 – 0.950)	0.724 (0.598 – 0.877)	0.907 (0.851 – 0.967)	0.809 (0.669 – 0.978)

Cancer specific mortality significantly rises during longer period after diagnosis, e.g. when cancer has already spread outside prostate area (pT:2) (Fig. 2 A graph). Cumulative 5-year cancer specific survival (CSS) in pT:0 showed very high survival rate of 99.9% (95% CI 99.8 – 100) as well as in pT:1 with 99% (95% CI 98.2 – 99.9) and decreased in pT:2 with 88.3% (95% CI 83.1 – 93.8) while cumulative 10-year CSS in pT:0 decreased to only 98.5% (95% CI 97.7 – 99.4); pT:1 decreased to 96.7% (95% CI 94.8 – 98.7) and pT:2 survival decreased by almost 10 percentage points (pp.) to 78.4% (95% CI 70.7 – 86.9). Log-rank test gave chi-square value of $\chi 2 = 163$ ($p = 2e\text{-}16$) so we reject null hypothesis and conclude that there is statistically significant difference between survival curves by pathologic curves for CSS. It is also worth mentioning, that Kaplan-Meier curves by pathologic stage for other cause survival (OCS) are also statistically different, log-rank test produced $\chi 2 = 12{,}5$ ($p = 0{,}002$).

It is also very clear that mortality from prostate cancer increases drastically with years after diagnosis when lymph nodes are damaged by cancer (pN:1) (Fig. 2 B graph). Cumulative 5-year CSS in both pN:0 and pN:2 are high at respectively 98.9% (95% CI 98.1 – 99.7) and 99.7% (95% CI 99.4 – 1) compared to having lymph nodes damaged by cancer (pN:1) where survival decreased down to 79.4% (95% CI 69.2 – 91.1). Cumulative 10-year CSS decreased by 2.8 and 1.4 pp. respectively in pN:0 and pN:2 down to 96.1% (95% CI 94.3 – 98) and 98.3% (95% CI 97.4 – 99.2); in pN:1 group we

recorded 19.3 pp. drop to 60.1% (95% CI 44.9 – 80.4). Log-rank test justifies our findings, with $\chi 2 = 242$ ($p = 2e\text{-}16$) we conclude that the survival curves by lymph node status for CSS are dissimilar. However, produced Kaplan-Meier curves by the same characteristic for OCS do not show any signs of irregularities from common survival rate with log-rank statistic $\chi 2 = 2{,}9$ ($p = 0{,}2$) we can say that the curves are identical.

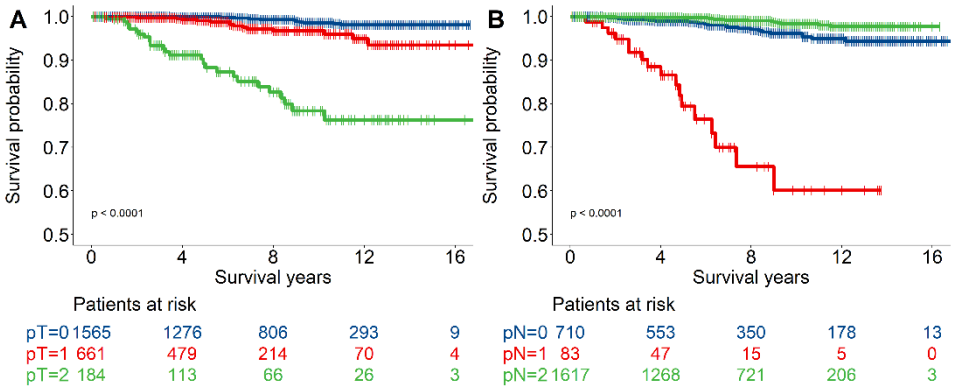

Figure 2. Kaplan-Meier survival curves for cancer specific survival.
(A) - by pathologic stage; (B) - by lymph nodes status.

Graphically a lower mortality is seen between surgical margin status groups (Fig. 3 C graph). There, cumulative 5-year CSS with clean removed prostate tissue (SM:0) is almost 100% - at 99.9% (95% CI 99.7 – 100); a bit lower when removed prostate tissue is damaged by cancer (SM:1) or unknown (SM:2) with survival rates respectively at 96.7% (95% CI 95.3 – 98.2) and 97.8% (95% CI 94.8 – 100). Cumulative 10-year CSS slightly decreases in SM:0 and SM:1 groups respectively down to 98.2% (95% CI 97.3 – 99.2) and 93.3% (95% CI 91 – 95.6); no changes are seen in SM:2 group. Log-rank test produced $\chi 2 = 45{,}3$ ($p = 1e\text{-}10$) which means the survival curves are not identical. Surgical margin status does not seem to make a difference in deaths from other causes as log-rank test gave us $\chi 2 = 1{,}5$ ($p = 0{,}5$) which means the survival curves are the same across the groups.

A noticeable difference in survival between groups is noticed between metastatic lymph nodes groups (Fig. 3 D graph). The cumulative 5-year CSS based on number of metastatic lymph nodes are as following:

- Metastatic lymph nodes: 1 – 100% (95% 100 – 100),
- Metastatic lymph nodes: 2 – 99.6% (95% 99.3 – 100),
- Metastatic lymph nodes: 3 – 99.6% (95% 98.8 – 100),
- Metastatic lymph nodes: 4 – 93.2% (95% 88.1 – 98.7),
- Metastatic lymph nodes: 5+ – 87.8% (95% 81.2 – 95),

and cumulative 10-year CSS:

- Metastatic lymph nodes: 1 – 99.1% (95% 97.9 – 100),
- Metastatic lymph nodes: 2 – 98.3% (95% 96.9 – 99.6),
- Metastatic lymph nodes: 3 – 93% (95% 85.8 – 100),
- Metastatic lymph nodes: 4 – 83.8% (95% 71.7 – 98),
- Metastatic lymph nodes: 5+ – 72.4% (95% 56.3 – 93.2).

It is noticeable, that for every additional metastatic lymph node, the mortality rate increases. Calculating CSS difference between cumulative 5-year and 10-year shows a 10.9 pp. decrease having 4 metastatic lymph nodes and 15.4 pp. having 5 or more. Log-rank test produced $\chi 2 = 268$ ($p = 2e\text{-}16$) so the survival curves are indeed different. Based on OCS, log-rank test gave us $\chi 2 = 5{,}2$ ($p = 0{,}3$) which concludes it does not affect other cause mortality significantly.

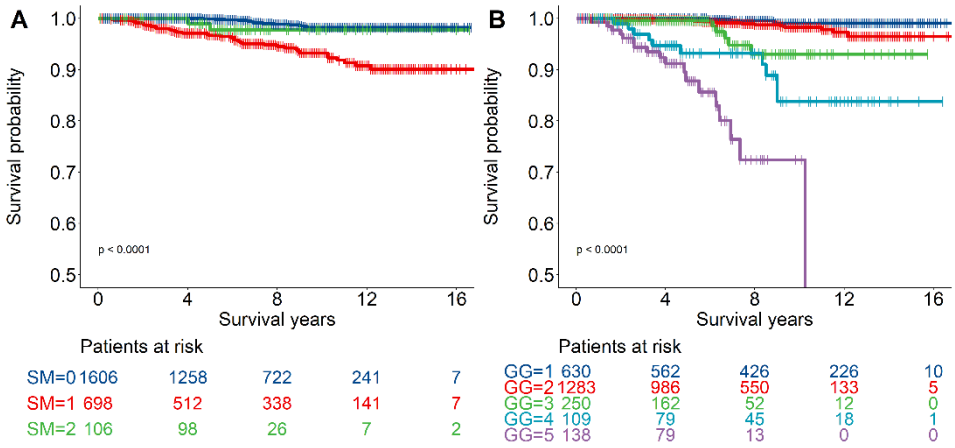

Figure 3. Kaplan-Meier survival curves for cancer specific survival.
(C) - by surgical margin status; (D) - by metastatic lymph nodes.

A propensity score analysis was performed to obtain homogenous group of patients based on lymph nodes status (pN). Our original dataset consists of 83 (3.4%) observations with cancer damaged lymph nodes and up to 1617 (67.1%) records with untreated lymph nodes (Table 1). During the analysis we will obtain a subset of patients whom other covariates will not show any statistically significant difference between pN groups. 2 iterative processes were done, because the matching algorithm works with 2 level factor and our pN variable is 3 levels factor, so firstly pN:0 with pN:1 were matched and then matched obtained subset with pN:2 group. After propensity score analysis our training set will consist of 168 patients whereas 41 of them had clean lymph nodes, 43 – cancer damaged lymph nodes and 84 – untreated lymph nodes. During homogeneity assumption testing it was noted that we have been left with too small set of 1 metastatic lymph nodes records, two groups were merged to accommodate this – 1 and 2 metastatic lymph nodes combined into single one. Non parametric "Kruskal-Wallis" test was used to check assumptions for continuous variables, for both age ($\chi 2 = 0.215$, $p = 0.898$) and PSA ($\chi 2 = 5.320$, $p = 0.070$) we accept the null hypothesis. For categorical variables, chi square test was used except for surgical margin status where Fisher-Freeman-Halton exact test was performed due to smaller set of observations in one of the factor levels. For all 3 variables, GG ($\chi 2 = 4.981$, $p = 0.546$), pT ($\chi 2 = 2.520$, $p = 0.640$) and SM ($p = 0.565$), we accept the null hypothesis about independence.

Training set consists of homogenous set by lymph nodes status while testing set consists of all other patients left after propensity score analysis. Training set has 168 records where 15 have died from prostate cancer and 24 – from other causes. Testing set

consists of unused 2242 observations where 41 have died from prostate cancer and 270 – death from other causes.

Semi-parametric Cox proportional hazards regression on homogenous training dataset established a few significant predictors for CSM (Table 5). Patients whose cancer is developing in prostate area (pT:1) have 5.7 times (HR = 0.174; 95% CI 0.044 – 0.682; p = 0.012) lower mortality risk than patients whose cancer spread outside prostate area (pT:2). Men with cancer damaged lymph nodes (pN:1) are 4.8 times (95% CI 1.184 – 19.640; p = 0.028) more likely to experience death from prostate cancer than patients with untreated lymph nodes (pN:2). Patients with 1 or 2 metastatic lymph nodes have 26.3 times (HR = 0.038; 95% CI 0.004 – 0.374; p = 0.005) lower CSM rate than men with 4 metastatic lymph nodes.

The proportional hazards assumption was checked using Schoenfeld residuals. Chi square was computed for each variable as well as for global one. Predictors age (p = 0.257), PSA (p = 0.120), pT (p = 0.236), pN (p = 0.114) and GG (p = 0.669) did not violate the proportional hazards assumptions at 5% significance level except for SM (p = 0.042). Overall, the global assumption was also not violated with p = 0.109.

It is also worth mentioning that training the Cox model based on deaths from other causes also finds statistically significant hazard ratios between pT:1 vs pT:2 and 1 or 2 vs 4 metastatic lymph nodes. There, patients whose cancer is developing in the prostate area (pT:1) have 3.4 times (HR = 0.295; 95% CI 0.100 – 0.830; p = 0.021) lower OCM rate than patients whose cancer is already spread outside the prostate area (pT:2). Also, men with 1 or 2 metastatic lymph nodes are 4 times (HR = 0.247; 95% CI 0.072 – 0.846; p = 0.026) less likely to experience death from other causes than men with 4 metastatic lymph nodes. Besides, the Cox model trained on deaths from other causes finds another statistically significant predictor – age. With each additional year in patient's age, on average the death from other causes increases by 17.8% (HR = 1.178; 95% CI 5.7% – 31.2%; p = 0.003).

Table 5. Semi-parametric Cox proportional hazards regression analysis for cancer-specific mortality.

Variable	Comparable level	Coefficient	Hazard ratio	Inverse hazard ratio	95% Confidence interval	p
Age		-0.009	0.991	1.009	0.889 – 1.105	0.876
Prostate specific antigen (PSA)		-0.026	0.974	1.027	0.920 – 1.032	0.371
Pathologic stage (pT)						
	(1 vs 0)	0.219	1.245	0.803	0.106 – 14.572	0.861
	(2 vs 0)	1.967	7.151	0.140	0.638 – 80.097	0.111
	(1 vs 2)	-1.748	0.174	5.744	0.044 – 0.682	0.012
Lymph nodes status (pN)						
	(1 vs 0)	0.836	2.307	0.434	0.513 – 10.369	0.276
	(2 vs 0)	-0.738	0.478	2.091	0.104 – 2.208	0.345

Variable	Comparable level	Coefficient	Hazard ratio	Inverse hazard ratio	95% Confidence interval	p
	(1 vs 2)	1.573	4.823	0.207	1.184 – 19.640	0.028
Surgical margin status (SM)						
	(1 vs 0)	0.116	1.123	0.890	0.295 – 4.277	0.865
	(2 vs 0)	0.678	1.971	0.507	0.204 – 19.081	0.558
	(1 vs 2)	-0.562	0.570	1.755	0.058 – 5.554	0.628
Metastatic lymph nodes						
	({1,2} vs 5)	-2.174	0.114	8.790	0.012 – 1.123	0.063
	(3 vs 5)	-0.099	0.906	1.104	0.191 – 4.299	0.901
	(4 vs 5)	1.096	2.993	0.334	0.655 – 13.686	0.157
	(3 vs {1,2})	2.075	7.964	0.126	0.840 – 75.462	0.071
	(3 vs 4)	-1.195	0.303	3.304	0.072 – 1.265	0.102
	({1,2} vs 4)	-3.270	0.038	26.309	0.004 – 0.374	0.005

Here {1,2} under metastatic lymph nodes means 1 and 2 groups were combined into single group; Inverse hazard ratio – exp(-coefficient).

Fine-Gray (FG) competing risk regression was performed on the same dataset for CSM (Table 6) and OCM. According to the pathologic stage, significant hazard ratio, same as in Cox model, was also found between pT:1 and pT:2. Fine-Gray model gave a 1.3 times lower mortality rate – patients, whose cancer is developing in the prostate area have 4.3 times (HR = 0.233; 95% CI 0.066 – 0.823; p = 0.024) lower CSM rate than men whose cancer is already spread outside the prostate area. But FG model did not find statistically significant hazard ratio between pN:1 and pN:2, FG gave us hazard ratio of 2.7 (95% CI 0.812 – 8.872; p = 0.110) where Cox model produced significant hazard ratio of 4.8 (95% CI 1.184 – 19.640; p = 0.028). Based on metastatic lymph nodes, FG model, same as Cox, found a significant hazard ratio between 1 or 2 and 4 metastatic lymph nodes. This time, the FG model lowered the risk by 5.14 times. People with 1 or 2 metastatic lymph nodes have 5.1 times (HR = 0.195; 95% CI 0.040 – 0.960; p = 0.044) lower CSM rate than patients with 4 metastatic lymph nodes.

The proportional hazards assumption was checked using the same technique as previously, all FGs models variables did not violate the assumption with age (p = 0.091), PSA (p = 0.107), pT (p = 0.186), pN (p = 0.250), SM (p = 0.148) and GG (p = 0.492) as well as the global one with p = 0.165.

Fine-Gray model, trained based on deaths from other causes established statistically significant hazard ratios only on pathologic stages pT:1 vs pT:2 and patient's age. FG method models a very similar hazard ratio for patient's age as Cox model: each additional year to the patient's age increases death from other causes by 18.1% (HR = 1.181; 95% CI 6.4% – 31.1%; p = 0.002). And men whose cancer is developing in

prostate area are 2.9 times (HR = 0.345; 95% CI 0.120 – 0.970; p = 0.043) less likely to experience death from other causes than patients whose cancer is already spread outside the prostate area.

Table 6. Fine-Gray competing risk regression analysis of cancer-specific mortality.

Variable	Comparable level	Coefficient	Hazard ratio	Inverse hazard ratio	95% Confidence interval	p
Age		-0.006	0.994	1.006	0.912 – 1.084	0.900
Prostate specific antigen (PSA)		-0.023	0.977	1.024	0.897 – 1.064	0.590
Pathologic stage (pT)						
	(1 vs 0)	-0.453	0.636	1.573	0.070 – 5.733	0.690
	(2 vs 0)	1.005	2.733	0.366	0.438 – 17.054	0.280
	(1 vs 2)	-1.458	0.233	4.299	0.066 – 0.823	0.024
Lymph nodes status (pN)						
	(1 vs 0)	0.512	1.669	0.599	0.448 – 6.216	0.440
	(2 vs 0)	-0.475	0.622	1.608	0.155 – 2.499	0.500
	(1 vs 2)	0.987	2.684	0.373	0.812 – 8.872	0.110
Surgical margin status (SM)						
	(1 vs 0)	0.495	1.641	0.609	0.486 – 5.535	0.420
	(2 vs 0)	0.484	1.622	0.616	0.122 – 21.520	0.710
	(1 vs 2)	0.011	1.011	0.989	0.087 – 11.764	0.990
Metastatic lymph nodes						
	({1,2} vs 5)	-0.349	0.705	1.418	0.139 – 3.578	0.670
	(3 vs 5)	0.335	1.398	0.715	0.340 – 5.747	0.640
	(4 vs 5)	1.284	3.611	0.277	0.916 – 14.232	0.066
	(3 vs {1,2})	0.685	1.983	0.504	0.321 – 12.265	0.460
	(3 vs 4)	-0.949	0.387	2.583	0.076 – 1.973	0.250
	({1,2} vs 4)	-1.634	0.195	5.122	0.040 – 0.960	0.044

Here {1,2} under metastatic lymph nodes mean 1 and 2 groups were combined into single group; Inverse hazard ratio – exp(-coefficient).

ROC curves using training set are very similar for Cox and FG models (Fig. 4 A graph), Cox model's AUC value is 0.7708 where FG model's – 0.7673 (Table 7). Models were tested on observations left after propensity score analysis, ROC curves are presented in

Fig. 4 B graph, Cox model's AUC value is 0.6747 and FG's – 0.6578 (Table 7). It is clear that Cox model shows better results both times than FG model.

Table 7. AUC values for Cox and FG models on training and testing sets.

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

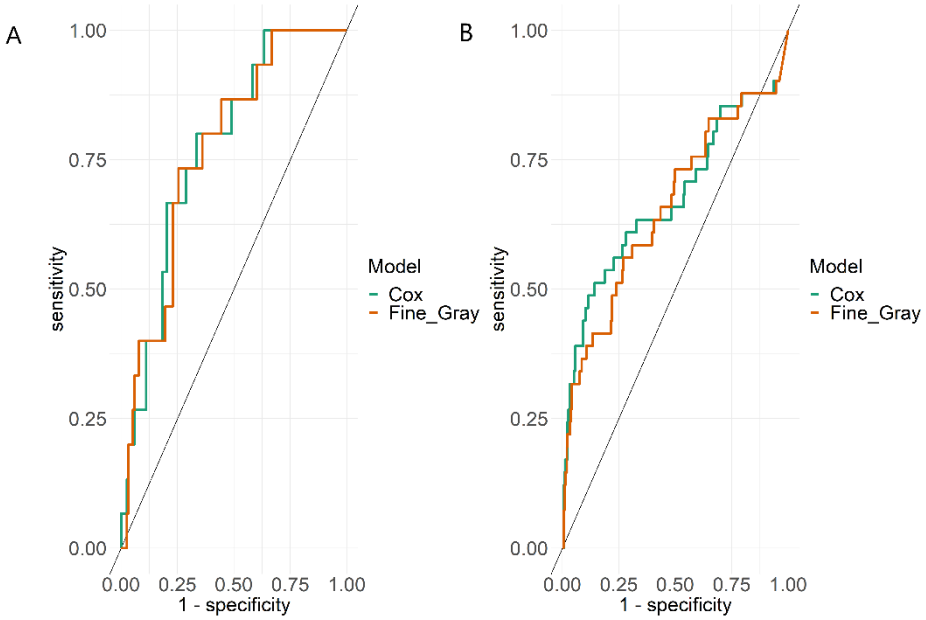

Figure 4. Cox and FG model's ROC curves; (A) - based on training set; (B) - based on testing set.

4. Conclusions

After Kaplan-Meier survival analysis we conclude:

- With 95% confidence level significant differences are detected between survival curves for prostate cancer mortality for all discrete variables.
- Statistically significant difference between factor levels are found only in pathologic stage for deaths from other causes.

- Worst case prognosis in the long-term is associated with patients whose lymph nodes are damaged with cancer or there are 5 or more metastatic lymph nodes.

After applying Cox model on sampled homogenous dataset by lymph node status we obtained:

- Men with 4 metastatic lymph nodes on average has 26 times higher prostate cancer mortality rate and 4 times higher other causes mortality than men with 1 or 2 metastatic lymph nodes.
- Patients whose cancer is developing in 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 patients whose cancer has already spread outside the prostate area.
- Men with cancer damaged lymph nodes are on average 4.8 times more likely to die from prostate cancer than men with untreated lymph nodes.
- With each additional year in patient's age, on average the death from other causes increases by 17.8%

Both Cox and Fine-Gray models are applied in medical research, yet only one of either are being used. Our comparison of Cox and Fine-Gray models on both training and testing sets showed that Cox model was found to be more accurate than Fine-Gray model based on ROC curves.

Prostate specific antigen was not found to be statistically significant predictor of prostate cancer-related death, a same trend could be seen in other studies on prostate cancer-related deaths.

Some of the metastatic lymph node groups were significant in both Cox and Fine-Gray models. Groups were also found to give the worst hazard ratios and worst long-term prognosis. This conclusion agrees with other papers which states the ISUP Gleason groups as an effective prognostic variables in modelling mortality from prostate cancer.

References

Adams, J., (1853). The case of scirrhous of the prostate gland with corresponding affliction of the lymphatic glands in the lumbar region and in the pelvis. *Lancet*;1:393.

Berney, D.M., Beltran, L., Fisher, G., North, B.V., Greenberg, D., Møller, H., Soosay, G., Scardino, P., Cuzick, J. (2016). Validation of a contemporary prostate cancer grading system using prostate cancer death as outcome. *Br J Cancer*, 114(10):1078-1083.

Braun, D.P., Gupta, D., Staren, E.D. (2012). Predicting survival in prostate cancer: the role of quality of life assessment. *Support Care Cancer* 20, 1267–1274.

Carlsson, S., Benfante, N., Alvim, R., Sjoberg, D.D., Vickers, A., Reuter, V.E., Fine, S.W., Vargas, H.A., Wiseman, M., Mamoor, M. et al. (2020). Long-Term Outcomes of Active Surveillance for Prostate Cancer: The Memorial Sloan Kettering Cancer Center Experience. *The Journal of Urology*. 203, 1122–1127.

Center, M., Jemal, A., Lortet-Tieulent, J., Ward, E., Ferlay, J., Brawley, O. et al., (2019). International Variation in Prostate Cancer Incidence and Mortality Rates. *European Urology*; 61(6):1079-1092.

Connor H., (2022). John Graunt F.R.S. (1620-74): The founding father of human demography, epidemiology and vital statistics. *Journal of Medical Biography*. p. 1-13.

Cox, D. R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society*, Series B 20:187–220.

Culp, M. B. B., Soerjomataram, I., Efstathiou, J. A., Bray, F., Jemal, A. (2020). Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates. *European Urology*, 77, 38–52.

Epstein, J.I., Egevad, L, Amin, M.B. et al. (2016). The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *The American journal of surgical pathology*. 40(2):244-252.

Fine, J. P., Gray, R. J. (1999), A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*, 94, p.496–509.

Macek, P., Biskup, M., Terek-Derszniak, M., Manczuk, M., Krol, H., Naszydlowska, E., Smok-Kalwat, J., Gozdz, S., Zak, M. (2020). Competing Risks of Cancer and Non-Cancer Mortality When Accompanied by Lifestyle-Related Factors-A Prospective Cohort Study in Middle-Aged and Older Adults. *Frontiers in Oncology*. 10: 545078.

Mekšriūnaitė, S., Gurevičius, R., Cicėnienė, V., Trakienė, A., Pošienė, A., (2018). Validity of prostate cancer as underlying cause of death in Lithuania from 2015 to 2017. *Visuomenės sveikata*, 4 (83), p.40-51.

Parkin, D., Boyd, L., Walker, L. C., (2011). The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *British Journal of Cancer*;105:2–5.

Spruance, S. L., Reid, J. E., Grace, M., Samore, M. (2004). Hazard ratio in clinical trials. *Antimicrob Agents Chemother*. 48(8):2787-2792.

Stewart, K., Choudry, M. I., Buckingham, R. (2016). Learning from hospital mortality. *Clinical medicine*, 16(6), 530–534.

Tosco, L., Laenen, A., Briganti, A. et al. (2018). The EMPaCT Classifier: A Validated Tool to Predict Postoperative Prostate Cancer-related Death Using Competing-risk Analysis. *European urology focus*, 4(3):369-375.

WEB (a). https://ec.europa.eu/eurostat/data/database

WEB (b) https://repository.library.georgetown.edu/handle/10822/707927

Received February 11, 2022, revised April 15, 2022, accepted May 26, 2022