TITLE (Like for any scientific paper)

Group name
List of Authors

Abstract

Introduction
The problem. Literature review (papers).
Objectives
Research questions
Materials and methods
Study design. Population. Frequency Tables. Univariate Analysis. Statistical models. Stratification. Propensity scores. Validation. Graphical representation.
Results
Numbers, models, summary tables, graphs.
Conclusions
Discussion. Interpretation. Limitations. Summary conclusions (bullet points)

Appendix
Extended tables
Source code

PROJECT

Introduction
The problem. Reference (papers).

Objectives
Research questions (bullet points)

Materials and methods
Study design (datasets, years).
Population (list hospitals and the selection process. Diagnosis and procedures).
Describe all analytical process, including used R functions.

Results (TARGETED TO RESPOND TO THE RESEARCH QUESTIONS)
Frequency Tables (by hospital, by risk factors).
Univariate Analysis (ORs by hospital or risk factor).
Statistical models (depending from your question). Related to Standardization (Direct, Indirect).
Funnel plots.
Stratification. Check for correlation (GEE). Propensity scores.

Validation (n-fold, bootstrap).
Graphical representation.

Conclusions (based upon the results, response to research questions)
Discussion. Interpretation.
Limitations. Summary conclusions (bullet points)

METHODS AND TOOLS FOR HEALTH STATISTICS
MSc Stat Sciences 2022/23
Department of Statistical Sciences, University of Bologna
(Fabrizio Carinci)

GUIDELINES FOR THE INDIVIDUAL ASSESSMENT

A. PROCEDURE

Groups between 1-5 people (optimal number: 2) will be formed to conduct a project of data analysis using a dataset of choice (recommended: NY Hospital datasets). The group submits to fabrizio.carinci@unibo.it **a one-page abstract (see project sample document), at least two weeks in advance of the examination, to make sure that the plan is on the right track. The Professor (and tutor for software development) will provide hints on the conduct of the study, but will not assess the document before the examination).**
The final version of the project will be a document with a short title and an abstract of maximum half a page, followed by structured contents (see below).
The document will be sent by email to fabrizio.carinci@unibo.it at least 3 days before the exam (booked as usual on AlmaEsami) and should not be substantially changed before the presentation at the exam. At the exam, a maximum set of 5 slides can be prepared for the scope. The exam will consist of a presentation of the project results by the student(s) (max 20 mins), followed by a series of questions submitted to each student independently.
The individual assessment will be based on: a) evaluation of a data analysis project chosen by up to a maximum of two students; b) oral examination.
Marks will be based on the judgment of: a) originality and difficulty of the project; b) quality of the product; c) ability to interpret the results; c) knowledge and understanding of the methods.

B. DESCRIPTION OF THE PRODUCT

Format: Technical Report
Length: Max 5 Pages

Structure:

Background
- Research question(s): precisely stated in relation to target parameters

Materials and Methods
- Description of dataset content and reference population
- Methodological Approach
- Statistical Analysis

Results
- Presentation of main results as text, referring to:
  - max 2 Tables
  - max 2 Figures

Discussion
- Interpretation of the results

Conclusions
- Final summary response to the research question(s)

C. ANALYTICAL STEPS OF A TYPICAL ASSIGNMENT

Suggested dataset (not compulsory, different projects could be considered):
Dataset of hospital inpatient discharges (SPARCS), State of New York, Year 2015-6
https://health.data.ny.gov/browse?q=sparcs&sortBy=newest&utf8=%E2%9C%93

1 Data Management
  1.1 Calculation of Crude Outcome Indicators *for the specific problem*
2 Direct Standardization
  2.1 Calculation of Directly Standardized Rates for all *units of interest* (using the complete dataset, i.e. all units, as a reference standard)
3 Risk Adjustment Models
  3.1 Application of hierarchical model strategy (Interaction, backward elimination)
  3.2 Application of GEE regression models (if considered appropriate)
  3.3 Evaluation of the goodness of fit of alternative models (LR, AIC, ROC)
  3.4 Compare results using different regression models, cross-validation and bootstrap
4 Indirect Standardization
  4.1 Calculation of Risk Adjusted Rates for all *units of interest* (AHRQ method)
5 Performance Evaluation
  5.1 Use graphs (Forest/Funnel Plots) to describe the variability of results
  5.2 Describe the impact of different methods (e.g. regression models) on interpretation

D. EXAMPLES OF PROJECTS

Project 1. *An exploration of the variability of Mortality after Surgery among hospitals, standardizing rates through multivariate risk adjustment using different regression models.*
Contents:

- *Identification of excess mortality rates for specific hospitals and counties*
- *Risk stratification of hospitals using regression models with hospital characteristics as major exposure, while adjusting for case-mix variables*
- *Variation of results between hospitals and counties when considering patients at high and low clinical risk separately.*

Project 2. *A study of the variability of Mortality for Acute Cerebrovascular Disease among hospitals, standardizing rates through multivariate risk adjustment using different regression models.*
Contents:

- *Identification of excess mortality rates for specific hospitals and counties*
- *Assessment of the impact of within hospital correlation through the application of GEE regression models*
- *Variation of results between hospitals and counties when considering surviving patients at high and low length of stay separately*
- *Appropriateness of separate analysis for hospitals with high/low volume of cases*

Project 3. *Variability of Mortality for Acute Myocardial Infarction (AMI) among hospitals: standardizing rates through multivariate risk adjustment using different regression models.*
Contents:

- *Identification of excess mortality rates for specific hospitals and counties*
- *Variation of results between hospitals and counties when considering surviving patients at high and low length of stay separately*
- *Analysis of determinants of high cost and variability of expenditure among counties in AMI survivors, adjusting for all potential confounders.*

Project 4. *Variation of Mortality after Chronic Obstructive Pulmonary Disease (COPD) among hospitals: standardizing rates through multivariate risk adjustment using different regression models.*
Contents:

- *Identification of excess mortality rates for specific hospitals and counties*
- *Comparison of results obtained with direct and indirect standardization*
- *Evaluation of goodness of fit for alternative models (LR, AIC, ROC)*

Project 5. *Variation of Mortality after Chronic Obstructive Pulmonary Disease (COPD) among hospitals, standardizing rates through multivariate risk adjustment using different regression models.*
Contents:

- *Identification of excess mortality rates for specific hospitals and counties*
- *Assessment of the impact of within hospital correlation through the application of GEE regression models.*
- *Comparison of results obtained with direct and indirect standardization*
- *Evaluation of the goodness of fit of alternative models (LR, AIC, ROC)*
- *Appropriateness of separate analysis for hospitals with high/low volume of cases*

IN-HOSPITAL MORTALITY FOR PANCREATIC CANCER IN THE 2015 NY HOSPITAL DATABASE: A TEMPORAL-SPATIAL COMPARISON

Abstract

OBJECTIVE: Cancer of pancreas is increasingly recognized as an emerging type of cancer in the US (American Cancer Society 2017) and remains one of the deadliest cancer types, with increasing incidence (Ilic *et al.* 2016). The current study examined the association between possible predictors of this neoplasm (age, gender, race, severity of illness) and pancreatic cancer mortality; a second aim was to assess differences between hospitals in terms of performance.

METHODS: By use of NY hospital inpatient discharge dataset (2015), Generalized Linear Models and Generalized Estimating Equation Models were performed to detect possible risk factors associated with in-hospital pancreatic cancer mortality. Likelihood ratio test and accuracy were used to establish the better model between a reduced model, obtained after applying Backward elimination technique, and the full one. Direct and indirect standardization methods were applied to evaluate performances in four major-volume hospitals. Comparisons among the hospitals, between hospital and areas and in time were developed.

RESULTS: A total of 2389 subjects with pancreatic cancer were identified from the 2015 New York Dataset of hospital inpatient discharges (SPARCS). Calvary Hospital was recognized as the worst performing among the six, also after standardized by age. It was decided to remove from the following analyses because it is a end-of-life specialized stricter, which hosts only terminal high-risky patients. The other three structures seems to perform quite good when compared to the NY overall performance. Population mortality rates is slightly decreasing in the last six years. Being 70 or older (OR=2.50 (1.31, 4.74)), presenting an extreme severity of cancer (OR=3.08 (2.36, 4.02)) appeared positively independently correlated with pancreatic cancer risk of mortality. Mortality does not seem to follow statistically different patterns among males and females.

CONCLUSIONS: Our results indicated a significantly positive association between age, severity of illness and length of stay and pancreatic cancer. Among hospitals, some variability is observed, with a general decreasing mortality trend. A better understanding of the aetiology and identifying the risk factors is essential for the primary prevention of this disease.

Introduction

Pancreatic cancer: aetiology,classification and epidemiology

The **pancreas** is a 6-inch long organ located behind the stomach in the back of the abdomen, near the gall bladder. It contains glands that create pancreatic juices, hormones, and insulin. Cancer can affect either the endocrine or the exocrine glands in the pancreas. The type of cancer and the outlook for each depends on which function the cancer affects (Altieri et al. 2017). The **exocrine glands** produce juices, or enzymes, that enter the intestines and help digest fat, proteins, and carbohydrates. These involves most of the pancreas activities. The **endocrine glands** are small clusters of cells known as the islets of Langerhans. They release the hormones insulin and glucagon into the bloodstream. There, they manage blood sugar levels. When they are not working properly, the result is often diabetes.

There are two different types of pancreatic cancer, depending on whether it affects the exocrine or endocrine functions: (i) *exocrine pancreatic cancer*, (ii) *endocrine pancreatic cancer* (Amundadottir 2016). According to the International Cassidication of Diseases, Tenth edition (ICD-10), we refer to both types of pancreatic cancer as C25. They have different risk factors, causes, symptoms, diagnostic tests, treatments, and outlook. The first one is the most common type of the two. It could be benign, with the developement of cysts, or for the majority they occours as malignant. The latter are less widespread and takes different names as the type of hormon-producing cells where the cancer starts vary. Generally, pancreatic cancer happens when uncontrolled cell growth begins in a part of the pancreas. Tumors develop, and these interfere with the way the pancreas works. Pancreatic cancer often shows no symptoms until the later stages, for this reason it is frequently named as “silent disease”. Only when the tumor reaches advanced stages, the patient may start feeling pain in the upper abdomen, lose appetite, experiment nausea, vomit or diarrhea, significantly lose weight, feel tired and pale (Schmidt-Hansen, Berendse, and Hamilton 2016; Walter et al. 2016).

Up to now, according to scientists, it is still not clear the mechanism that induces uncontrolled cell growth in the pancreas (Chu, Goggins, and Fishman 2017), but many studies have tried to identify some possible risk factors. Among these, we find hereditary genetics, sex, age, Afro-American race, lifestyle habits (i.e. smoking, alcohol, sedentary, obesity, poor quality diet), exposure to environmental toxins (i.e. pesticides, chemical, dyes)(Korc et al. 2017; Midha, Chawla, and Garg 2016). It has also been demonstrated that pancreatic cancer is associated to many other diseases, such as diabetes mellitus, cirrhosis, stomach infections, chronic pancreatitis, periodontal disease, which may increase the risk of developing the analysed type of neoplasm (Andersen et al. 2017; Chang et al. 2016; Choi et al. 2018; Midha et al. 2016).

According to the American Cancer Society (American Cancer Society 2017), around 3% of all cancers in the United States are pancreatic cancers and it accounts for the 7% of all cancer deaths. Cancer of the pancreas remains one of the deadliest cancer types (Ilic and Ilic 2016). In 2018, they expect around 55,440 people to receive a diagnosis of pancreatic cancer. Based on the GLOBOCAN 2018 estimates (Bray et al. 2018), pancreatic cancer causes more than 432242 deaths every year, ranking as the seventh leading cause of cancer death in both sexes together. Globally, about 458,918 people had pancreatic cancer in 2018, making it the 11th most common cancer. The highest incidence and mortality rates of pancreatic cancer are found in developed countries. The overall five-year survival rate is about 6% (ranges from 2% to 9%), but this vary very small between developed and developing countries. There are no current screening recommendations for pancreatic cancer, so primary prevention is of utmost importance. A better understanding of the aetiology and identifying the risk factors is essential for the primary prevention of this disease. Pancreatic cancer patients have frequent hospitalizations for establishment of diagnosis, recovery from surgery and for complications related to cancer, surgery or chemotherapy. Hospital admissions are associated with decreased quality of life and increased cost of care (Boyd et al. 2012). It has been proved that recently, altough the number of admissions for pancreatic cancer has increased in the US, a decrease in the mortality trend has been observed (Bhandari et al. 2018).

Objective

The aim of the present study is to explore and investigate mortality among patients hospitalized for pancreatic cancer in the NY hospital dataset, detecting which may be the most relevant risk factors able to explain great part of such mortality. As second goal, I intend to analyse how different hospitals perform in terms of mortality rates, dealing with pancreatic cancer patients, examining how were their trends in the last six years.

Materials

Data analysed in this study come from the long version of the 2015 New York Dataset of hospital inpatient discharges provided by the Statewide Planning and Research Cooperative System (SPARCS). It contains information about N=1867316 patients admitted in NY area hospitals in 2015. Each hospitalization is treated as an individual entry in the database and is coded with 263 diagnosis and 232 procedures. The dataset includes 37 variables, which provide details about demographic information, clinical characteristics of the patients and their hospital outcomes. In order to pursue the aim of this analysis, I considered as study population N=2389 individuals, who were identified selecting only patients with "cancer of pancreas" as diagnosis description. To access in-hospital mortality in pancreatic cancer patients, I selected among diseases individuals only those whose disposition was "expired" (N=325 subjects).

Methods

The first step aims to explore the study population, performing univariate and bivariate descriptive analyses of the patients included in the dataset.

After conducting the main descriptive statistics, I performed multivariate logistic regression models to determine the independent risk factors of mortality. I compared binomial regression model from the family of Generalized Linear Models (GLM) with Generalized Estimated Equations Models (GEE). The second type of models is an extension of the first one, which adds the information about the correlation structure that lies behind data. In this case, it is useful to take into consideration the correlation within hospitals that there could arise since different numbers of patients are admitted and treated in several hospitals. Among the different correlation matrix structures, the ones used in this analyses are the uncorrelated/identity matrix, which provides the same results of GLM because it assumes that only correlations on the main diagonal are 1, while all the other values are 0, and the exchangeable, which allows all the possible couples of considered hospitals to have the same correlation. The choice of a Binomial model was motivated by the fact that the outcome I am investigating on (in-hospital mortality for pancreatic cancer) is dichotomous, i.e. yes or no. The null hypothesis on which the models are based is that the regression coefficients are not statistically different from zero, which means that the considered covariate has not a relevant effect on the outcome. When a positive coefficient is obtained, this correspond to a possible risk factor, while when a negative coefficient is obtained, this indicated a possible protective effect. Once estimated two complete models, in which I inserted all the possible covariates explaining mortality, through the technique of the Backward Elimination, I started taking out one variable at a time from the model, beginning with the covariate which showed the lowest degree of significance (higher p.value). I repeat this operation until all the predictors included in the model resulted significant, with a p. value lower than 0.05. At this point, I compared the initial complete model with the reduced one using the log Likelihood Ratio Test (LRT) to evaluate which model to prefer. LRT provides the possibility to compare different nested models, taking into consideration both the complexity of the model and how it fits well the data:

$$\mathrm{LRT} = -2\ln(\hat{L})_{\mathrm{FULL}} - -2\ln(\hat{L})_{\mathrm{REDUCED}}$$

where $\hat{L}$ are the estimated likelihood functions of the complete and reduced models.

Then, direct and indirect standardization methods were applied in order to investigate mortality trends among different hospitals. The first method was applied standardizing the population of the different hospitals by age and severity of illness, while the indirect standardization was computed by age, race and gender. The first kind of standardization was useful to make proper comparisons both among several facilities and among hospitals within the area in which they are sited, because the same population was used as standard.

$$DMR = \frac{\sum_{ij} N_{ij} \frac{d_{ij}}{n_{ij}}}{N}$$

where N_{ij}/N describe the structure of the standard population, and it is applied to the number of deaths d_{ij} and number of individuals n_{ij} of the different hospitals one wants to compare.

On the other hand, indirect standardization does not allow direct comparisons among the analysed units (hospital); the only possible comparison is between each unit and the average. This is not permitted because adjusted rates are calculated used different populations, but the weighting factors are different. The application of indirect standardization methods lead to the computation of the following measure:

$$RAR_i = \alpha \frac{O_i}{E_i}$$

defined as Risk-Adjustment Rate and computed for each unit (hospital). α indicates the average rate in population, O_i the specific observed rate in hospital *i*, E_i the specific expected rate in hospital *i*.

Results and Discussion

***Tab. 1** Demographic characteristics of pancreatic cancer patients.*

Covariates	N	%
Age category		
0-17	5	0.2
18-29	5	0.2
30-49	145	6.0
50-69	1095	45.8
70 or older	1139	47.7
Sex		
Male	1155	48.4
Female	1233	51.6
Race		
White	1418	59.4
Black/African American	459	19.2
Multi-racial	33	1.4
Other race	479	20.0
Total	2389	100.0

***Tab. 2** Admission/Hospital characteristics of pancreatic cancer patients.*

Covariates	n	%
Severity of illness		
Minor	106	4.4
Moderate	589	29.1
Major	1311	54.9
Extreme	383	16.0
Primary Insurance Payer		
Medicare	1353	56.6
Medicade	327	13.7
Private Health Insurance	363	15.2
Blue Cross	257	10.8
Self-pay	24	1.0
Other	65	2.7
Hospital Area		
NY City	1550	64.9
Long Island	435	18.2
Hudson Valley	199	8.3
Capital/Adiron	189	7.9
Central NY	16	0.7
Main hospitals (n>100 patients)		
Memorial Hospital	255	10.7
NY Presb. Hospital-Col.	178	7.5
Calvary Hospital	128	5.4
North Shore Univ. Hospital	100	4.2
Length of stay	Mean=31 days (Range 1-65)	
Total	2389	100.0

Within the New York Hospital Inpatient Discharges of 2015, there were 2389 patients admitted with a diagnosis of pancreatic cancer. The demographic characteristics of the patients included in the study are shown in Table 1. The majority of patients were older than 50 (93%), female (52%) and white (60%). Females tend to have an older age distribution compared to men; in fact, while the first are for the 56% older than 70 years, men in the same age class account for only the 39%. Examining the bivariate distribution of race according to age groups, White are generally older than Black/African American. According to their admission/hospital characteristics, as seen in Table 2, patients were admitted for the majority with a high grade of severity (71%), Medicare (57%) as their primary insurance, in the NY city area (65%), and they were hospitalized in mean for about 30 days. In general, in the NY area they received admission in 116 different hospitals, but here in the table only the ones with the greatest volume (more than 100 patients) are reported, since they represent the population of the 25th percentile. 64 different procedures are applied to the patients, and, among these, oral gastrointestinal therapy (27%), no procedure (16%), non-oral gastrointestinal therapy (10%) are the most frequent.

Among the 2389 hospitalized patients, 325 (13.6%) died in hospital, maintaining the same proportion for both genders. From now on, the analyses will be focused on those 325 individuals and aim to explain mortality due to pancreatic cancer.

Moving to the multivariate analysis of mortality, two complete GLM and GEE models were computed, in order to investigate about the possible predictors of in-hospital pancreatic cancer mortality, as suggested from the literature by Bhandari (Bhandari et al. 2018). The models intended to evaluate the following covariates:

Model1 (complete): Dead ~ $\beta_0 + \beta_1$ age4 + β_2 age5 + β_3 gender + β_4 black + β_5 multi_racial + β_6 major_severity + β_7 extreme_severity + β_8 length_of_stay

After proceeding with the Backward Elimination technique, since in this present model several covariates were not significantly related to mortality, I stopped after eliminating respectively gender, major severity of illness and both the races Black/African American and Multi-racial.

Here follows the specification of the reduced model:

Model2 (reduced): Dead ~ $\beta_0 + \beta_1$ age4 + β_2 age5 + β_3 extreme_severity + β_4 length_of_stay

Tab. 3 *Predictors of in-hospital mortality in pancreatic cancer on multivariate analysis.* **(***REDUCED MODEL)*

	GLM Model		**GEE Model**	
Age Category	**OR (95% CI)**	**p.value**	**OR (95% CI)**	**p.value**
0-49	Ref.		Ref.	
50-69	1.87 (1.00, 3.57)	0.050	1.40 (0.94, 2.09)	0.096
70 or older	2.50 (1.31, 4.74)	0.005	1.61 (1.08, 2.41)	0.020
Severity				
Minor	Ref.		Ref.	
Major-Extreme	3.08 (2.36, 4.02)	0.000	3.94 (3.11, 4.99)	0.000
Length of stay	0.99 (0.98, 0.99)	0.000	0.993 (0.989, 0.998)	0.001

Table 3 compares the results of applying the reduced logistic GLM and GEE models. Looking at the ORs and considering each covariate effect independently from the others, we see that for both the models the two oldest classes of age resulted as statistically relevant risk factors, as well as presenting a highly severe stage of pancreatic cancer. As the classes of age increase, the risk of mortality gets very high, becoming 2 times and a half higher for the oldest patients when compared to 0-49 group individuals. The same direction is evident for the grade of illness severity, declared once the individual enters the hospital. For whom is classified as major/extremely ill, the risk of mortality increases hugely, with risks more than 3 times higher than who presents a minor severity of the disease. Length of stay is confirmed to be a protective factor, even if with a very reduced OR. For every added day a patient remain in a hospital, he/she slightly decreases his/her mortality risk due to pancreatic cancer. This could be reasonably explained by the fact that the more extreme cases of fragile patients at high mortality risk dye within the first days in the hospital, while who remain longer is in a more stable situation and doctors prefer to monitor and keep under observation in hospital before the discharge.

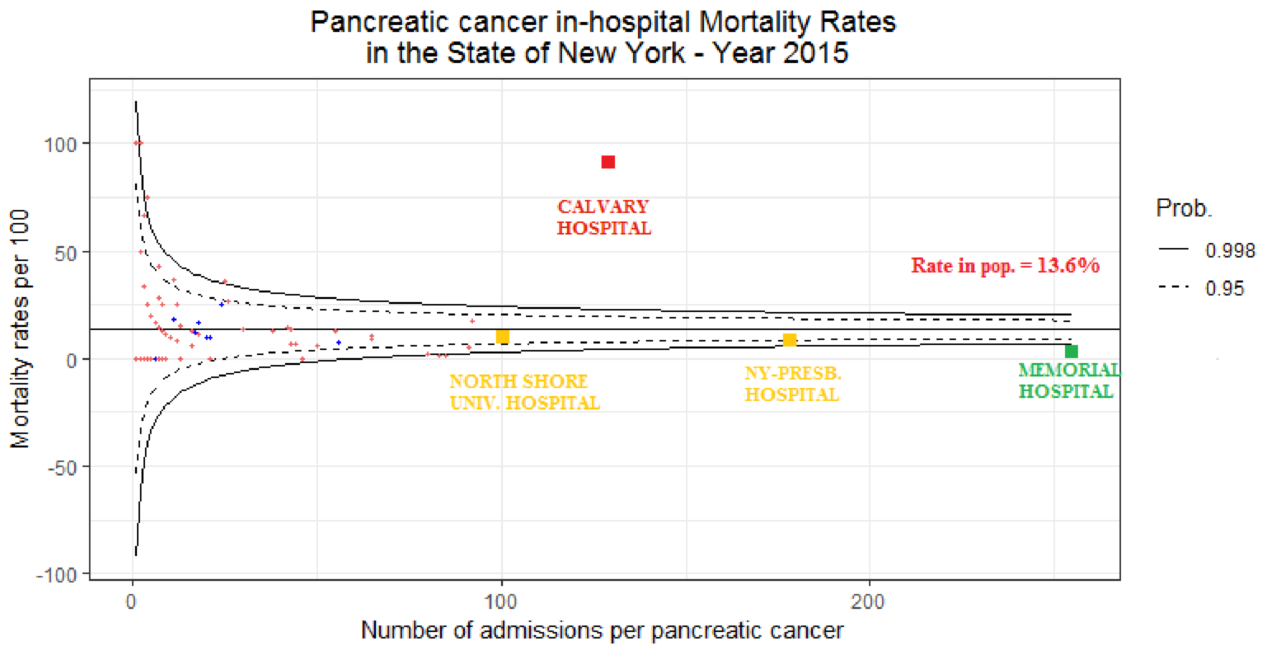
Since we are comparing two nested models, when dealing with model selection, in order to decide which of the two better fit my data, I obtained the following two log-likelihood function estimates: lnL1= -896.4942, lnL2= -898.84321. I conduced a log-likelihood ratio test, comparing the two obtained estimates, obtaining LRT=4.70 with a corresponding not significant p.value= 0.3197. Even if the two log-likelihood functions appear to be very similar and their differences are not so relevant, I would chose the second model, since it contains less covariate and less parameters, but it is able to capture all the relevant determinant of mortality risk in our situation. In terms of accuracy, both the model provide a probability of correct classification equal to 86.4% which is quite good; the drawback of both models is that they are able to estimate the cases who survives, but they encounter problems in the prediction of cases of death. Surely, covariates other than the ones considered in this study explain mortality with better predictive capability.

__Tab. 4__ Number of cases, crude and direct standardized rates by age in 6 major hospitals, NY, 2015.

Hospital	N	Cases of death	Crude/ Observed Rate	Standardized Rate	Expected Rate	RAR (95% CI)
Memorial Hospital	255	8	0.031	0.019	0.124	0.035 (-0.001, 0.079)
NY Presb. Hospital-Col.	178	13	0.073	0.043	0.134	0.074 (0.024, 0.124)
Cavalry Hospital	128	114	0.891	0.675	0.201	0.603 (0.559, 0.647)
North Shore Univ. Hospital	100	11	0.110	0.094	0.141	0.106 (0.043, 0.170)
Rate in population (ȳ) 0.136						

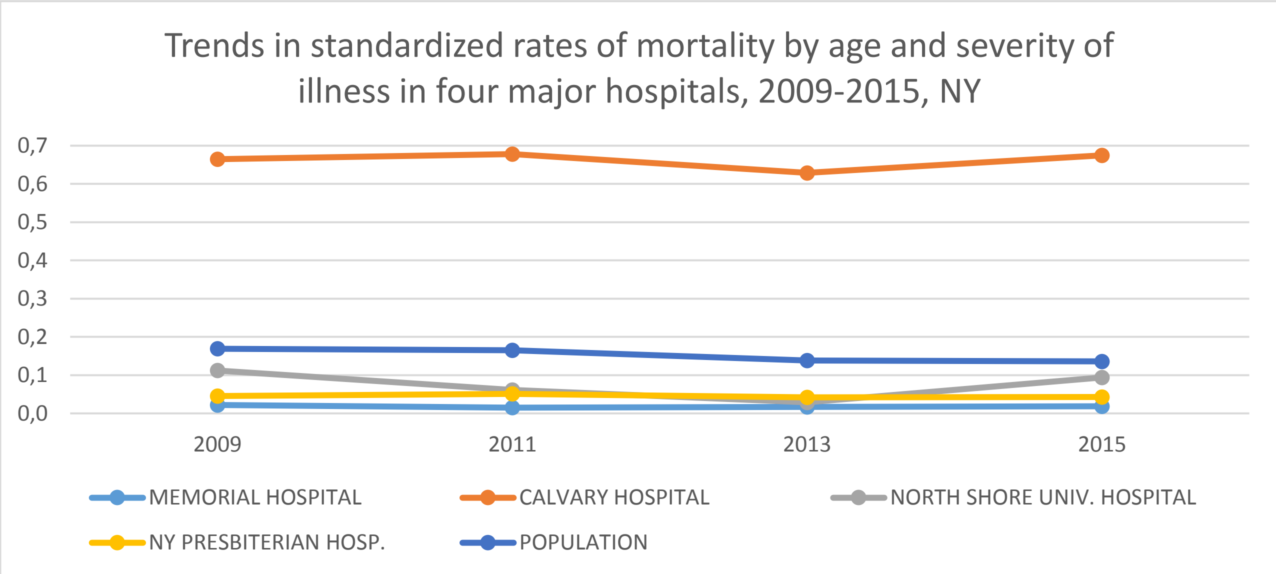
In tab. 3, we find the results about the four biggest hospitals and their relative mortality rates, after directly standardizing by age using the complete dataset as reference standard population and indirectly standardized by age, gender and race, obtaining RARs. It is immediately visible that Calvary Hospital is the worst performing one, even after standardized by age. It has a huge number of deaths, almost all admitted patients die in-hospital. This is not a strange result if we consider that such hospital is a hospice, specialized in palliative care, which focus on the end-of-life care. For this reason, such a high mortality rates is not attributable to the bad performance of this institute, but to the target patients that it uses to admits. Excluding this hospital from the comparison, we conclude that, among the other three, North Shore University Hospital, a teaching university centre, is the one which shows the most negative performances in terms of pancreatic cancer survival, with the 11% of pancreatic cancer patients dying, which becomes 9.5% after the standardization by age. When looking at its expected rate, it results to be higher than the actual observed one, and so we cannot conclude that it is a bad performing structure, considering the characteristic of its patients, because it is doing well than what is expected in term of survival for pancreatic cancer. On the other hands, Memorial Hospital appears to be the best performing one, since its standardized rate of mortality is really low; only 2 patients out of 100 dye in this structure. Even comparing the expected with the observed rate, it appears to perform much better than the expectations.

Figure 1



In the funnel plot (Fig.1), we see represented as dots all the NY hospital standardized mortality rates by age and severity of the illness of the NY. The four highlighted with coloured squares are the four biggest hospitals in terms of volume, on which this analysis is concentrated. At first look, it is immediately visible how Calvary hospital performs so critically in terms of mortality rates, and it is an outlier in negative sense. On contrast, Memorial hospital results to have great performance, even outside the 99.8% control limits, appearing as the only example of positive outlier. The other two considered hospitals place within the control limits, which means that they behave not so differently compared to the average mortality trend of the population. All the other dots on the left correspond to many other hospitals, but with a very reduced number of patient so that their estimated mortality rates are not stable and reliable, because based on little number of cases and populations. When comparing the four hospitals with the areas in which they are placed, North Shore University hospital is the only one located in Long Island Area, which has an overall mortality rates of 0.035, and so the hospital we are considering seems to perform more negatively, with a standardized rate of 0.094. The other three structures are situated in the NY city area, whose mortality rate is 0.113; as a consequence, Calvary Hospital again presents negative differences with its area while the remaining two perform better than other structures in their area.

Figure 2



After having comparing these main hospitals within their areas, I look back of 6 years and I analysed their mortality trend since 2009 to 2015. In fig.2, the illustration of the standardized mortality rates compared to the average population mortality rate among pancreatic cancer affected patients. During the 6 years, mortality trend has been quite stable, with little variations. In population, it seems to be slightly decreasing, moving from 0.17 to 0.13 Calvary hospital is the exception which always has its own performances due to the kind of structure it is. The other biggest hospitals seems to have positive results when compared to the NY average of all hospitals.

Conclusions

The present work found as possible predictors of in-hospital mortality among pancreatic patients the oldest classes of ages and the highest grade of severity of the disease, i.e. the older and more severe an individual, the higher his/her risk of dying in the hospital due to the neoplasms. I explored and found also some variability when comparing performances of various hospitals in terms of mortality rates. Since our study includes only the hospitalized patients, the results might be little different from that from the outpatient setting, where the patients tend to be less severely ill. A lack of this analysis is that it focuses only on hospital mortality, while many patients decide to go to hospice or remain at home during the last phase of their life, so that the system does not take into account. Another issue could be related to the structure of the dataset, which contains de-identified records; it is not guarantee that each row correspond to a different subject, because do not have information about readmission of the same patient, who need again hospital cares after having been discharged.

References

Altieri, Barbara, William B. Grant, Silvia Della Casa, Francesco Orio, Alfredo Pontecorvi, Annamaria Colao, Gerardo Sarno, and Giovanna Muscogiuri. 2017. "Vitamin D and Pancreas: The Role of Sunshine Vitamin in the Pathogenesis of Diabetes Mellitus and Pancreatic Cancer." *Critical Reviews in Food Science and Nutrition* 57(16):3472–88.

American Cancer Society. 2017. "Key Statistics for Pancreatic Cancer." Retrieved May 25, 2019 (https://www.cancer.org/cancer/pancreatic-cancer/about/key-statistics.html).

Amundadottir, Laufey T. 2016. "Pancreatic Cancer Genetics." *International Journal of Biological Sciences* 12(3):314–25.

Andersen, Dana K., Murray Korc, Gloria M. Petersen, Guido Eibl, Donghui Li, Michael R. Rickels, Suresh T. Chari, and James L. Abbruzzese. 2017. "Diabetes, Pancreatogenic Diabetes, and Pancreatic Cancer." *Diabetes* 66(5):1103–10.

Bhandari, Sanjay, Mubeen Khan Mohammed Abdul, Will Hollabaugh, Kanav Sharma, Douglas B. Evans, and Nalini Guda. 2018. "Decreased Trend in Hospital Mortality from Pancreatic Cancer despite Increase in Number of Hospital Admissions" edited by C. M. Isales. *PLOS ONE* 13(7):e0199909.

Boyd, Casey A., Daniel W. Branch, Kristin M. Sheffield, Yimei Han, Yong-Fang Kuo, James S. Goodwin, and Taylor S. Riall. 2012. "Hospital and Medical Care Days in Pancreatic Cancer." *Annals of Surgical Oncology* 19(8):2435–42.

Bray, Freddie, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. 2018. "Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries." *CA: A Cancer Journal for Clinicians* 68(6):394–424.

Chang, Jeffrey S., Chia-Rung Tsai, Li-Tzong Chen, and Yan-Shen Shan. 2016. "Investigating the Association Between Periodontal Disease and Risk of Pancreatic Cancer." *Pancreas* 45(1):134–41.

Choi, Kun Kug, Seung Hyuk Kim, Kyung Don Yoo, Hyo Jin Kim, Ji In Park, Subin Hwang, Ho Jun Chin, and Ho Suk Ku. 2018. "Variation in the Rate of Well-Controlled Status of Chronic Disease by Income Level in Korea: 2010 to 2015." *Medicine* 97(34):e12059.

Chu, Linda C., Michael G. Goggins, and Elliot K. Fishman. 2017. "Diagnosis and Detection of Pancreatic Cancer." *The Cancer Journal* 23(6):333–42.

Ilic, Milena and Irena Ilic. 2016. "Epidemiology of Pancreatic Cancer." *World Journal of Gastroenterology* 22(44):9694.

Korc, Murray, Christie Y. Jeon, Mouad Edderkaoui, Stephen J. Pandol, Maxim S. Petrov, and Consortium for the Study of Chronic Pancreatitis, Diabetes, and Pancreatic Cancer (CPDPC). 2017. "Tobacco and Alcohol as Risk Factors for Pancreatic Cancer." *Best Practice & Research Clinical Gastroenterology* 31(5):529–36.

Midha, Shallu, Saurabh Chawla, and Pramod Kumar Garg. 2016. "Modifiable and Non-Modifiable Risk Factors for Pancreatic Cancer: A Review." *Cancer Letters* 381(1):269–77.

Schmidt-Hansen, Mia, Sabine Berendse, and William Hamilton. 2016. "Symptoms of Pancreatic Cancer in Primary Care." *Pancreas* 45(6):814–18.

Walter, Fiona M., Katie Mills, Silvia C. Mendonça, Gary A. Abel, Bristi Basu, Nick Carroll, Sue Ballard, John Lancaster, William Hamilton, Greg P. Rubin, and Jon D. Emery. 2016. "Symptoms and Patient Factors Associated with Diagnostic Intervals for Pancreatic Cancer (SYMPTOM Pancreatic Study): A

Prospective Cohort Study." *The Lancet Gastroenterology & Hepatology* 1(4):298–306.