

Survival Analysis for Pancreatic Cancer Patients using Cox-Proportional Hazard (CPH) Model

Chris P. Tsokos

Received: 10 February 2021 Accepted: 3 March 2021 Published: 15 March 2021

Abstract

Pancreatic cancer is comparatively rare but extremely lethal. In the United States, pancreatic cancer is the 4th leading cause of cancer death, and in Europe, it is the 6th. Though Pancreatic cancer remains incurable if detected late, research into improving the therapeutic strategy has increased significantly in recent years. However, it is ambiguous if sustained improvements have been achieved by identifying the most prominent risk factors responsible for cancer. In this article, we studied the survival times of 677 pancreatic cancer patients with fifteen risk factors. The semi-parametric Cox proportional hazard (CPH) model was used to examine the covariate effect taking into account all of the statistically significant risk factors and their significant twoway interactions. A careful and rigorous assessment of the risk factors based on the AIC of the stepwise selection technique revealed seven risk factors, and ten interaction terms are statistically significantly contributing to the survival times. The final Cox-PH model was wellvalidated and satisfied all the key assumptions. The identified risk factors and their interactions are ranked according to the prognostic effect on the survival time based on the hazard ratio. We found the most contributing risk factor is the combined effect of patients with emphysema and cancer stage regional with a hazard ratio (HR) = 8.84.

Index terms— pancreatic cancer, cox-PH model, pancreatic survival function.

In the domain of the lethal carcinogenic diseases affecting humans, pancreatic cancer is one of the fatal cancers and continues to be a crucial unsolved health problem at the start of the 21st century. Because of the high fatality rates, pancreatic cancer incidence rates are almost equal to mortality rates (22). According to the current health science researchers, this disease causes approximately 30,000 deaths per year in the USA. (??1). It is the fourth principal reason for cancer death in the USA and leads to an estimated 227,000 deaths per year worldwide. The incidence and number of deaths caused by pancreatic tumors have been gradually increasing, even as incidence and mortality of other common cancers have been declining. Despite developments in detection and management of pancreatic cancer, only about 4% of patients will live five years after diagnosis, (??2). The normal pancreas consists of digestive enzyme-secreting acinar cells, bicarbonate-secreting ductal cells, centroacinar cells that are the geographical transition between acinar and ductal cells, hormone-secreting endocrine islets and relatively inactive stellate cells. The majority of malignant neoplasms of the pancreas are adenocarcinomas. Rare pancreatic neoplasms include neuroendocrine tumors (which can secrete hormones such as insulin or glucagon) and acinar carcinomas (which can release digestive enzymes into the circulation). Particularly, ductal adenocarcinoma is the most frequent kind of malignancy of the pancreas; this tumor (commonly referred to as pancreatic cancer) presents a substantial health problem, with an estimated 367,000 new cases diagnosed worldwide in 2015 and an associated 359,000 deaths in the same year(3)(4). After the detection of pancreatic cancer, doctors usually perform some additional tests to understand better if cancer has been spread or the spreading area of cancer. Different imaging tests, such as a PET scan, can help doctors identify the presence of cancerous growths. With these tests, doctors try to establish cancer's stage. Staging helps explicate how advanced the cancer is. It also assists doctors in deciding the treatment options. The following are the description of the stages used in our dataset according to the definition of the Surveillance, Epidemiology, and End Results (SEER) database.

Localized: There is no sign that the cancer has spread outside of the pancreas. 2. Regional: The cancer has spread from the pancreas to nearby structures or lymph nodes.

1 Distant:

The cancer has spread to distant parts of the body such as the lungs, liver or bones.

The following Figure 1 shows the different parts of the pancreas.

Although, in most cases, pancreatic cancer remains incurable, researchers have focused on how to improve the survival times of patients diagnosed with pancreatic cancer. A schematic diagram of the data used in our study with the description of risk factors is shown in Figure 2, below.

As the above Figure illustrates, we see that twelve out of fifteen risk factors are categorical, having two or more categories. Before we proceed with our main analysis, it is very important to investigate if there is any statistically significant difference between the survival times of male and female patients diagnosed with pancreatic cancer. If any significant differences are found, separate analyses for each gender should be performed. To answer this question, we used the non-parametric Wilcoxon rank-sum test with continuity correction and obtained a p-value of .47, indicating that there is not enough sample evidence to reject the following null hypothesis (H_0) at a 5% level of significance. H_0 : There is no statistically significant difference between the survival times of male and female patients. Thus we proceeded with our analysis and modeling by combining the male and female data together to constitute our sample size.

The Cox PH model, proposed by Sir David Cox, is a statistical method that can be used for survival time (time-to-event) outcomes on one or more risk factors and their interactions. In survival analysis, the Cox model has been widely recommended for semi-parametric modeling of the survival time relationship as a function of the risk factors. Kleinbaum & Klein (2010) gives a good introductory review of the background and methodology, and more detailed descriptions have been provided by Kalbeisch, and Prentice (2011) (2012). In this section, we give a brief review of the Cox proportional hazards model. An important aspect of the Cox PH model is the hazard function $h(t)$. It measures the rate of the event of occurrence (death) as a function of time t . We define the hazard function as follows; Let random variable T denotes the survival time with cumulative density function $F_T(t)$, given by where $f(t) = dF_T(t)/dt$ is the probability density function (pdf) of the random variable T . The survival function at time t is defined as: $F_T(t) = P(T \leq t) = \int_0^t f(t)dt$, $S(t) = P(T > t) = 1 - F_T(t) = \int_t^\infty f(t)dt$. (1)

$S(t)$ gives the probability that a specific individual would survive beyond time t . Since $S(t)$ is a probability, $0 \leq S(t) \leq 1$ and $S(0) = 1$, for $T \geq 0$ from (1) we have, $f(t) = dF_T(t)/dt = -dS(t)/dt$. (2)

For continuous survival data, the hazard function plays a very important role. It aims to quantify the instantaneous risks that an event will occur at time t . It is defined as the follows: $h(t) = \lim_{\Delta t \rightarrow 0} \frac{P\{t \leq T < t + \Delta t | T \geq t\}}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P\{t \leq T < t + \Delta t\}}{\Delta t} \frac{1}{S(t)} = f(t) S(t)$. (3)

Combining (2) and (3), we obtain, $h(t) = -d/dt \log\{S(t)\}$. (4)

Integrating both sides of equation (4) gives an expression for the survival function $S(t)$ in terms of the hazard function $h(t)$. That is, $S(t) = \exp \left\{ - \int_0^t h(u)du \right\}$. (5)

Now, from (3) and (5) we can express the pdf $f(t)$ as a function of $S(t)$ and $h(t)$ given by, $f(t) = h(t) \exp \left\{ - \int_0^t h(u)du \right\}$. (6)

From (3) the cumulative hazard function $H(t)$ can be expressed as: $H(t) = \int_0^t h(u)du = -\ln S(t)$. (7)

Now, suppose $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})$ are the realized values of the risk factor for the i th subject.

Then, the Cox PH model (not including time-dependent risk factors or non-proportional hazards) can be expressed in term of the hazard as: $h_i(t) = h_0(t) \exp \left\{ \sum_{j=1}^p \beta_j X_{ij} + \sum_{j,k=1}^p \gamma_{jk} X_{ij} X_{ik} \right\}$, $j, k = 1, 2, \dots, p$. (8)

In the above expression, $h_0(t)$ is called the baseline hazard which can be thought of as the hazard function for an individual for which all value of the risk factors are 0. β_j measures the impact of X_{ij} on $h_i(t)$. γ_{jk} is the interaction coefficient between j th and k th risk factor of the i th individual and Statistical Data Analysis and Survival Modeling measures the impact of $X_{ij} X_{ik}$ on $h_i(t)$. From (8), it is clear that the individual hazard is a function of the risk factors and their interactions and is connected through baseline hazard. From (8), we can write, $\ln h_i(t) - \ln h_k(t) = \sum_{j=1}^p \beta_j X_{ij} + \sum_{j,k=1}^p \gamma_{jk} X_{ij} X_{ik} - \sum_{j,k=1}^p \gamma_{jk} X_{ik} X_{kj}$, $j, k = 1, 2, \dots, p$. (9)

From the above expression we see that the ratio of log hazard of the i th and k th individual is constant over time. Thus, the name proportional in the Cox PH model. We interpret the hazard ratio (HR) in the following ways:

1. $HR = 1$; implies that there is no hazard effect. Thus, the risk factors have no relationship with the event probability, thus, no influence on the length of survival.

2. $HR > 1$ (i.e. equivalently $\beta_i > 0$), implies an increase in hazard. That is, the risk factors have a positive association with the event probability, thus, a negative association with the length of survival (bad prognostic factor).

3. $HR < 1$ (i.e. equivalently $\beta_i < 0$), implies a decrease in hazard. That is, the risk factors are negatively associated with the probability of the event, thus, positively associated with the length of survival (good prognostic factor).

A detailed description of the hazard ratio have been provided in (14) (15).

We now proceed to develop our most parsimonious statistical model using Cox PH. We initially started by

fitting the Cox-PH model to the survival times t as a function of all fifteen risk factors given in Figure 2 together with their two-way interactions. So, there were fifteen risk factors and 15

105 two-way interaction terms. We used a stepwise model selection procedure to select the best model with the minimum Akaike information criterion ($AIC = 2\ln(L) + 2k$, where L is the value of the maximum likelihood function of the model and k represents the number of estimated model parameters) (13). AIC gives an estimation of the relative amount of information missing in the model; hence, the smaller the AIC value, the better the quality of the model. It also deals with the risk associated with overfitting or under-fitting the model. One of the most important assumptions of the Cox PH is proportionality. Initially, all of the risk factors and two-way interactions except age satisfied the assumption. The range of the variable age was [50-90]. So, we divided the range into two categories, say [50,70), and [70,90). Now, we use stratification on the variable age. Stratification is one of the tools used by researchers when one of the risk factors does not satisfy the proportionality assumption. The stratification will produce hazard ratios for all other risk factors in the presence of two hazards intrinsic to the level of age. Since age violated the proportional hazards assumption, stratifying it will help meet the PH assumption and provide more valid estimates for all other risk factors. The stratified model allows the baseline hazard $h_0(t)$ to vary between strata but controls the effect of the risk factors to be the same for each stratum. For each subject in strata s , $s = 1, 2$, we have from (10),
$$h_i(t) = h_{0s}(t) \exp \sum_{j=1}^p \beta_j X_{ij} + \sum_{k=1}^p \beta_k X_{ik} \quad (10)$$

However, it is not possible to get an estimate of the risk factor (age) separately after stratification. The following Figure 3 illustrates the survival curve for the two age groups. The following Table 1 illustrates the count of each category of all the risk factors after stratification.

The step-wise procedure produced seven out of fourteen significant risk factors and ten two-way interaction terms. There were some risk factors that did not contribute to the hazard individually, but, interacting with other risk factors, their effect was significant. Thus, we added those risk factors in our proposed model. That is why there are thirteen individual risk factors and ten interactions in the model (11). In the following model (11), we denote "Y" to indicate yes of a specific answer of a risk factor. That is, the specific category possesses the characteristic. For example, to answer the question "does the patient ever have diabetes?" the individual answers "yes." To describe any particular category of the risk factor stage, we use L, R, and D which are the first letters of Localized, Regional, and Distant. To describe male and female category of the variable Sex, we use the letters M and F, respectively. The most parsimonious model that we found after removing the insignificant ($p\text{-value} > 0.05$) term from the model is given as follows:
$$\ln h_i(t) / h_0(t) = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} + \beta_8 X_{8i} + \beta_9 X_{9i} + \beta_{10} X_{10i} + \beta_{11} X_{11i} + \beta_{12} X_{12i} + \beta_{13} X_{13i} + \beta_{14} X_{14i} + \beta_{15} X_{15i} + \beta_{16} X_{16i} + \beta_{17} X_{17i} + \beta_{18} X_{18i} + \beta_{19} X_{19i} + \beta_{20} X_{20i} + \beta_{21} X_{21i} + \beta_{22} X_{22i} + \beta_{23} X_{23i} + \beta_{24} X_{24i} + \beta_{25} X_{25i} + \beta_{26} X_{26i} + \beta_{27} X_{27i} + \beta_{28} X_{28i} + \beta_{29} X_{29i} + \beta_{30} X_{30i} + \beta_{31} X_{31i} + \beta_{32} X_{32i} + \beta_{33} X_{33i} + \beta_{34} X_{34i} + \beta_{35} X_{35i} + \beta_{36} X_{36i} + \beta_{37} X_{37i} + \beta_{38} X_{38i} + \beta_{39} X_{39i} + \beta_{40} X_{40i} + \beta_{41} X_{41i} + \beta_{42} X_{42i} + \beta_{43} X_{43i} + \beta_{44} X_{44i} + \beta_{45} X_{45i} + \beta_{46} X_{46i} + \beta_{47} X_{47i} + \beta_{48} X_{48i} + \beta_{49} X_{49i} + \beta_{50} X_{50i} + \beta_{51} X_{51i} + \beta_{52} X_{52i} + \beta_{53} X_{53i} + \beta_{54} X_{54i} + \beta_{55} X_{55i} + \beta_{56} X_{56i} + \beta_{57} X_{57i} + \beta_{58} X_{58i} + \beta_{59} X_{59i} + \beta_{60} X_{60i} + \beta_{61} X_{61i} + \beta_{62} X_{62i} + \beta_{63} X_{63i} + \beta_{64} X_{64i} + \beta_{65} X_{65i} + \beta_{66} X_{66i} + \beta_{67} X_{67i} + \beta_{68} X_{68i} + \beta_{69} X_{69i} + \beta_{70} X_{70i} + \beta_{71} X_{71i} + \beta_{72} X_{72i} + \beta_{73} X_{73i} + \beta_{74} X_{74i} + \beta_{75} X_{75i} + \beta_{76} X_{76i} + \beta_{77} X_{77i} + \beta_{78} X_{78i} + \beta_{79} X_{79i} + \beta_{80} X_{80i} + \beta_{81} X_{81i} + \beta_{82} X_{82i} + \beta_{83} X_{83i} + \beta_{84} X_{84i} + \beta_{85} X_{85i} + \beta_{86} X_{86i} + \beta_{87} X_{87i} + \beta_{88} X_{88i} + \beta_{89} X_{89i} + \beta_{90} X_{90i} + \beta_{91} X_{91i} + \beta_{92} X_{92i} + \beta_{93} X_{93i} + \beta_{94} X_{94i} + \beta_{95} X_{95i} + \beta_{96} X_{96i} + \beta_{97} X_{97i} + \beta_{98} X_{98i} + \beta_{99} X_{99i} + \beta_{100} X_{100i} + \beta_{101} X_{101i} + \beta_{102} X_{102i} + \beta_{103} X_{103i} + \beta_{104} X_{104i} + \beta_{105} X_{105i} + \beta_{106} X_{106i} + \beta_{107} X_{107i} + \beta_{108} X_{108i} + \beta_{109} X_{109i} + \beta_{110} X_{110i} + \beta_{111} X_{111i} + \beta_{112} X_{112i} + \beta_{113} X_{113i} + \beta_{114} X_{114i} + \beta_{115} X_{115i} + \beta_{116} X_{116i} + \beta_{117} X_{117i} + \beta_{118} X_{118i} + \beta_{119} X_{119i} + \beta_{120} X_{120i} + \beta_{121} X_{121i} + \beta_{122} X_{122i} + \beta_{123} X_{123i} + \beta_{124} X_{124i} + \beta_{125} X_{125i} + \beta_{126} X_{126i} + \beta_{127} X_{127i} + \beta_{128} X_{128i} + \beta_{129} X_{129i} + \beta_{130} X_{130i} + \beta_{131} X_{131i} + \beta_{132} X_{132i} + \beta_{133} X_{133i} + \beta_{134} X_{134i} + \beta_{135} X_{135i} + \beta_{136} X_{136i} + \beta_{137} X_{137i} + \beta_{138} X_{138i} + \beta_{139} X_{139i} + \beta_{140} X_{140i} + \beta_{141} X_{141i} + \beta_{142} X_{142i} + \beta_{143} X_{143i} + \beta_{144} X_{144i} + \beta_{145} X_{145i} + \beta_{146} X_{146i} + \beta_{147} X_{147i} + \beta_{148} X_{148i} + \beta_{149} X_{149i} + \beta_{150} X_{150i} + \beta_{151} X_{151i} + \beta_{152} X_{152i} + \beta_{153} X_{153i} + \beta_{154} X_{154i} + \beta_{155} X_{155i} + \beta_{156} X_{156i} + \beta_{157} X_{157i} + \beta_{158} X_{158i} + \beta_{159} X_{159i} + \beta_{160} X_{160i} + \beta_{161} X_{161i} + \beta_{162} X_{162i} + \beta_{163} X_{163i} + \beta_{164} X_{164i} + \beta_{165} X_{165i} + \beta_{166} X_{166i} + \beta_{167} X_{167i} + \beta_{168} X_{168i} + \beta_{169} X_{169i} + \beta_{170} X_{170i} + \beta_{171} X_{171i} + \beta_{172} X_{172i} + \beta_{173} X_{173i} + \beta_{174} X_{174i} + \beta_{175} X_{175i} + \beta_{176} X_{176i} + \beta_{177} X_{177i} + \beta_{178} X_{178i} + \beta_{179} X_{179i} + \beta_{180} X_{180i} + \beta_{181} X_{181i} + \beta_{182} X_{182i} + \beta_{183} X_{183i} + \beta_{184} X_{184i} + \beta_{185} X_{185i} + \beta_{186} X_{186i} + \beta_{187} X_{187i} + \beta_{188} X_{188i} + \beta_{189} X_{189i} + \beta_{190} X_{190i} + \beta_{191} X_{191i} + \beta_{192} X_{192i} + \beta_{193} X_{193i} + \beta_{194} X_{194i} + \beta_{195} X_{195i} + \beta_{196} X_{196i} + \beta_{197} X_{197i} + \beta_{198} X_{198i} + \beta_{199} X_{199i} + \beta_{200} X_{200i} + \beta_{201} X_{201i} + \beta_{202} X_{202i} + \beta_{203} X_{203i} + \beta_{204} X_{204i} + \beta_{205} X_{205i} + \beta_{206} X_{206i} + \beta_{207} X_{207i} + \beta_{208} X_{208i} + \beta_{209} X_{209i} + \beta_{210} X_{210i} + \beta_{211} X_{211i} + \beta_{212} X_{212i} + \beta_{213} X_{213i} + \beta_{214} X_{214i} + \beta_{215} X_{215i} + \beta_{216} X_{216i} + \beta_{217} X_{217i} + \beta_{218} X_{218i} + \beta_{219} X_{219i} + \beta_{220} X_{220i} + \beta_{221} X_{221i} + \beta_{222} X_{222i} + \beta_{223} X_{223i} + \beta_{224} X_{224i} + \beta_{225} X_{225i} + \beta_{226} X_{226i} + \beta_{227} X_{227i} + \beta_{228} X_{228i} + \beta_{229} X_{229i} + \beta_{230} X_{230i} + \beta_{231} X_{231i} + \beta_{232} X_{232i} + \beta_{233} X_{233i} + \beta_{234} X_{234i} + \beta_{235} X_{235i} + \beta_{236} X_{236i} + \beta_{237} X_{237i} + \beta_{238} X_{238i} + \beta_{239} X_{239i} + \beta_{240} X_{240i} + \beta_{241} X_{241i} + \beta_{242} X_{242i} + \beta_{243} X_{243i} + \beta_{244} X_{244i} + \beta_{245} X_{245i} + \beta_{246} X_{246i} + \beta_{247} X_{247i} + \beta_{248} X_{248i} + \beta_{249} X_{249i} + \beta_{250} X_{250i} + \beta_{251} X_{251i} + \beta_{252} X_{252i} + \beta_{253} X_{253i} + \beta_{254} X_{254i} + \beta_{255} X_{255i} + \beta_{256} X_{256i} + \beta_{257} X_{257i} + \beta_{258} X_{258i} + \beta_{259} X_{259i} + \beta_{260} X_{260i} + \beta_{261} X_{261i} + \beta_{262} X_{262i} + \beta_{263} X_{263i} + \beta_{264} X_{264i} + \beta_{265} X_{265i} + \beta_{266} X_{266i} + \beta_{267} X_{267i} + \beta_{268} X_{268i} + \beta_{269} X_{269i} + \beta_{270} X_{270i} + \beta_{271} X_{271i} + \beta_{272} X_{272i} + \beta_{273} X_{273i} + \beta_{274} X_{274i} + \beta_{275} X_{275i} + \beta_{276} X_{276i} + \beta_{277} X_{277i} + \beta_{278} X_{278i} + \beta_{279} X_{279i} + \beta_{280} X_{280i} + \beta_{281} X_{281i} + \beta_{282} X_{282i} + \beta_{283} X_{283i} + \beta_{284} X_{284i} + \beta_{285} X_{285i} + \beta_{286} X_{286i} + \beta_{287} X_{287i} + \beta_{288} X_{288i} + \beta_{289} X_{289i} + \beta_{290} X_{290i} + \beta_{291} X_{291i} + \beta_{292} X_{292i} + \beta_{293} X_{293i} + \beta_{294} X_{294i} + \beta_{295} X_{295i} + \beta_{296} X_{296i} + \beta_{297} X_{297i} + \beta_{298} X_{298i} + \beta_{299} X_{299i} + \beta_{300} X_{300i} + \beta_{301} X_{301i} + \beta_{302} X_{302i} + \beta_{303} X_{303i} + \beta_{304} X_{304i} + \beta_{305} X_{305i} + \beta_{306} X_{306i} + \beta_{307} X_{307i} + \beta_{308} X_{308i} + \beta_{309} X_{309i} + \beta_{310} X_{310i} + \beta_{311} X_{311i} + \beta_{312} X_{312i} + \beta_{313} X_{313i} + \beta_{314} X_{314i} + \beta_{315} X_{315i} + \beta_{316} X_{316i} + \beta_{317} X_{317i} + \beta_{318} X_{318i} + \beta_{319} X_{319i} + \beta_{320} X_{320i} + \beta_{321} X_{321i} + \beta_{322} X_{322i} + \beta_{323} X_{323i} + \beta_{324} X_{324i} + \beta_{325} X_{325i} + \beta_{326} X_{326i} + \beta_{327} X_{327i} + \beta_{328} X_{328i} + \beta_{329} X_{329i} + \beta_{330} X_{330i} + \beta_{331} X_{331i} + \beta_{332} X_{332i} + \beta_{333} X_{333i} + \beta_{334} X_{334i} + \beta_{335} X_{335i} + \beta_{336} X_{336i} + \beta_{337} X_{337i} + \beta_{338} X_{338i} + \beta_{339} X_{339i} + \beta_{340} X_{340i} + \beta_{341} X_{341i} + \beta_{342} X_{342i} + \beta_{343} X_{343i} + \beta_{344} X_{344i} + \beta_{345} X_{345i} + \beta_{346} X_{346i} + \beta_{347} X_{347i} + \beta_{348} X_{348i} + \beta_{349} X_{349i} + \beta_{350} X_{350i} + \beta_{351} X_{351i} + \beta_{352} X_{352i} + \beta_{353} X_{353i} + \beta_{354} X_{354i} + \beta_{355} X_{355i} + \beta_{356} X_{356i} + \beta_{357} X_{357i} + \beta_{358} X_{358i} + \beta_{359} X_{359i} + \beta_{360} X_{360i} + \beta_{361} X_{361i} + \beta_{362} X_{362i} + \beta_{363} X_{363i} + \beta_{364} X_{364i} + \beta_{365} X_{365i} + \beta_{366} X_{366i} + \beta_{367} X_{367i} + \beta_{368} X_{368i} + \beta_{369} X_{369i} + \beta_{370} X_{370i} + \beta_{371} X_{371i} + \beta_{372} X_{372i} + \beta_{373} X_{373i} + \beta_{374} X_{374i} + \beta_{375} X_{375i} + \beta_{376} X_{376i} + \beta_{377} X_{377i} + \beta_{378} X_{378i} + \beta_{379} X_{379i} + \beta_{380} X_{380i} + \beta_{381} X_{381i} + \beta_{382} X_{382i} + \beta_{383} X_{383i} + \beta_{384} X_{384i} + \beta_{385} X_{385i} + \beta_{386} X_{386i} + \beta_{387} X_{387i} + \beta_{388} X_{388i} + \beta_{389} X_{389i} + \beta_{390} X_{390i} + \beta_{391} X_{391i} + \beta_{392} X_{392i} + \beta_{393} X_{393i} + \beta_{394} X_{394i} + \beta_{395} X_{395i} + \beta_{396} X_{396i} + \beta_{397} X_{397i} + \beta_{398} X_{398i} + \beta_{399} X_{399i} + \beta_{400} X_{400i} + \beta_{401} X_{401i} + \beta_{402} X_{402i} + \beta_{403} X_{403i} + \beta_{404} X_{404i} + \beta_{405} X_{405i} + \beta_{406} X_{406i} + \beta_{407} X_{407i} + \beta_{408} X_{408i} + \beta_{409} X_{409i} + \beta_{410} X_{410i} + \beta_{411} X_{411i} + \beta_{412} X_{412i} + \beta_{413} X_{413i} + \beta_{414} X_{414i} + \beta_{415} X_{415i} + \beta_{416} X_{416i} + \beta_{417} X_{417i} + \beta_{418} X_{418i} + \beta_{419} X_{419i} + \beta_{420} X_{420i} + \beta_{421} X_{421i} + \beta_{422} X_{422i} + \beta_{423} X_{423i} + \beta_{424} X_{424i} + \beta_{425} X_{425i} + \beta_{426} X_{426i} + \beta_{427} X_{427i} + \beta_{428} X_{428i} + \beta_{429} X_{429i} + \beta_{430} X_{430i} + \beta_{431} X_{431i} + \beta_{432} X_{432i} + \beta_{433} X_{433i} + \beta_{434} X_{434i} + \beta_{435} X_{435i} + \beta_{436} X_{436i} + \beta_{437} X_{437i} + \beta_{438} X_{438i} + \beta_{439} X_{439i} + \beta_{440} X_{440i} + \beta_{441} X_{441i} + \beta_{442} X_{442i} + \beta_{443} X_{443i} + \beta_{444} X_{444i} + \beta_{445} X_{445i} + \beta_{446} X_{446i} + \beta_{447} X_{447i} + \beta_{448} X_{448i} + \beta_{449} X_{449i} + \beta_{450} X_{450i} + \beta_{451} X_{451i} + \beta_{452} X_{452i} + \beta_{453} X_{453i} + \beta_{454} X_{454i} + \beta_{455} X_{455i} + \beta_{456} X_{456i} + \beta_{457} X_{457i} + \beta_{458} X_{458i} + \beta_{459} X_{459i} + \beta_{460} X_{460i} + \beta_{461} X_{461i} + \beta_{462} X_{462i} + \beta_{463} X_{463i} + \beta_{464} X_{464i} + \beta_{465} X_{465i} + \beta_{466} X_{466i} + \beta_{467} X_{467i} + \beta_{468} X_{468i} + \beta_{469} X_{469i} + \beta_{470} X_{470i} + \beta_{471} X_{471i} + \beta_{472} X_{472i} + \beta_{473} X_{473i} + \beta_{474} X_{474i} + \beta_{475} X_{475i} + \beta_{476} X_{476i} + \beta_{477} X_{477i} + \beta_{478} X_{478i} + \beta_{479} X_{479i} + \beta_{480} X_{480i} + \beta_{481} X_{481i} + \beta_{482} X_{482i} + \beta_{483} X_{483i} + \beta_{484} X_{484i} + \beta_{485} X_{485i} + \beta_{486} X_{486i} + \beta_{487} X_{487i} + \beta_{488} X_{488i} + \beta_{489} X_{489i} + \beta_{490} X_{490i} + \beta_{491} X_{491i} + \beta_{492} X_{492i} + \beta_{493} X_{493i} + \beta_{494} X_{494i} + \beta_{495} X_{495i} + \beta_{496} X_{496i} + \beta_{497} X_{497i} + \beta_{498} X_{498i} + \beta_{499} X_{499i} + \beta_{500} X_{500i} + \beta_{501} X_{501i} + \beta_{502} X_{502i} + \beta_{503} X_{503i} + \beta_{504} X_{504i} + \beta_{505} X_{505i} + \beta_{506} X_{506i} + \beta_{507} X_{507i} + \beta_{508} X_{508i} + \beta_{509} X_{509i} + \beta_{510} X_{510i} + \beta_{511} X_{511i} + \beta_{512} X_{512i} + \beta_{513} X_{513i} + \beta_{514} X_{514i} + \beta_{515} X_{515i} + \beta_{516} X_{516i} + \beta_{517} X_{517i} + \beta_{518} X_{518i} + \beta_{519} X_{519i} + \beta_{520} X_{520i} + \beta_{521} X_{521i} + \beta_{522} X_{522i} + \beta_{523} X_{523i} + \beta_{524} X_{524i} + \beta_{525} X_{525i} + \beta_{526} X_{526i} + \beta_{527} X_{527i} + \beta_{528} X_{528i} + \beta_{529} X_{529i} + \beta_{530} X_{530i} + \beta_{531} X_{531i} + \beta_{532} X_{532i} + \beta_{533} X_{533i} + \beta_{534} X_{534i} + \beta_{535} X_{535i} + \beta_{536} X_{536i} + \beta_{537} X_{537i} + \beta_{538} X_{538i} + \beta_{539} X_{539i} + \beta_{540} X_{540i} + \beta_{541} X_{541i} + \beta_{542} X_{542i} + \beta_{543} X_{543i} + \beta_{544} X_{544i} + \beta_{545} X_{545i} + \beta_{546} X_{546i} + \beta_{547} X_{547i} + \beta_{548} X_{548i} + \beta_{549} X_{549i} + \beta_{550} X_{550i} + \beta_{551} X_{551i} + \beta_{552} X_{552i} + \beta_{553} X_{553i} + \beta_{554} X_{554i} + \beta_{555} X_{555i} + \beta_{556} X_{556i} + \beta_{557} X_{557i} + \beta_{558} X_{558i} + \beta_{559} X_{559i} + \beta_{560} X_{560i} + \beta_{561} X_{561i} + \beta_{562} X_{562i} + \beta_{563} X_{563i} + \beta_{564} X_{564i} + \beta_{565} X_{565i} + \beta_{566} X_{566i} + \beta_{567} X_{567i} + \beta_{568} X_{568i} + \beta_{569} X_{569i} + \beta_{570} X_{570i} + \beta_{571} X_{571i} + \beta_{572} X_{572i} + \beta_{573} X_{573i} + \beta_{574} X_{574i} + \beta_{575} X_{575i} + \beta_{576} X_{576i} + \beta_{577} X_{577i} + \beta_{578} X_{578i} + \beta_{579} X_{579i} + \beta_{580} X_{580i} + \beta_{581} X_{581i} + \beta_{582} X_{582i} + \beta_{583} X_{583i} + \beta_{584} X_{584i} + \beta_{585} X_{585i} + \beta_{586} X_{586i} + \beta_{587} X_{587i} + \beta_{588} X_{588i} + \beta_{589} X_{589i} + \beta_{590} X_{590i} + \beta_{591} X_{591i} + \beta_{592} X_{592i} + \beta_{593} X_{593i} + \beta_{594} X_{594i} + \beta_{595} X_{595i} + \beta_{596} X_{596i} + \beta_{597} X_{597i} + \beta_{598} X_{598i} + \beta_{599} X_{599i} + \beta_{600} X_{600i} + \beta_{601} X_{601i} + \beta_{602} X_{602i} + \beta_{603} X_{603i} + \beta_{604} X_{604i} + \beta_{605} X_{605i} + \beta_{606} X_{606i} + \beta_{607} X_{607i} + \beta_{608} X_{608i} + \beta_{609} X_{609i} + \beta_{610} X_{610i} + \beta_{611} X_{611i} + \beta_{612} X_{612i} + \beta_{613} X_{613i} + \beta_{614} X_{614i} + \beta_{615} X_{615i} + \beta_{616} X_{616i} + \beta_{617} X_{617i} + \beta_{618} X_{618i} + \beta_{619} X_{619i} + \beta_{620} X_{620i} + \beta_{621} X_{621i} + \beta_{622} X_{622i} + \beta_{623} X_{623i} + \beta_{624} X_{624i} + \beta_{625} X_{625i} + \beta_{626} X_{626i} + \beta_{627} X_{627i} + \beta_{628} X_{628i} + \beta_{629} X_{629i} + \beta_{630} X_{630i} + \beta_{631} X_{631i} + \beta_{632} X_{632i} + \beta_{633} X_{633i} + \beta_{634} X_{634i} + \beta_{635} X_{635i} + \beta_{636} X_{636i} + \beta_{637} X_{637i} + \beta_{638} X_{638i} + \beta_{639} X_{639i} + \beta_{640} X_{640i} + \beta_{641} X_{641i} + \beta_{642} X_{642i} + \beta_{643} X_{643i} + \beta_{644} X_{644i} + \beta_{645} X_{645i} + \beta_{646} X_{646i} + \beta_{647} X_{647i} + \beta_{648} X_{648i} + \beta_{649} X_{649i} + \beta_{650} X_{650i} + \beta_{651} X_{651i} + \beta_{652} X_{652i} + \beta_{653} X_{653i} + \beta_{654} X_{654i} + \beta_{655} X_{655i} + \beta_{656} X_{656i} + \beta_{657} X_{657i} + \beta_{658} X_{658i} + \beta_{659} X_{659i} + \beta_{660} X_{660i} + \beta_{661} X_{661i} + \beta_{662} X_{662i} + \beta_{663} X_{663i} + \beta_{664} X_{664i} + \beta_{665} X_{665i} + \beta_{666} X_{666i} + \beta_{667} X_{667i} + \beta_{668} X_{668i} + \beta_{669} X_{669i} + \beta_{670} X_{670i} + \beta_{671} X_{671i} + \beta_{672} X_{672i} + \beta_{673} X_{673i} + \beta_{674} X_{674i} + \beta_{675} X_{675i} + \beta_{676} X_{676i} + \beta_{677} X_{677i} + \beta_{678} X_{678i} + \beta_{679} X_{679i} + \beta_{680} X_{680i} + \beta_{681} X_{681i} + \beta_{682} X_{682i} + \beta_{683} X_{683i} + \beta_{684} X_{684i} + \beta_{685} X_{685i} + \beta_{686} X_{686i} + \beta_{687} X_{687i} + \beta_{688} X_{688i} + \beta_{689} X_{689i} + \beta_{690} X_{690i} + \beta_{691} X_{691i} + \beta_{692} X_{692i} + \beta_{693} X_{693i} + \beta_{694} X_{694i} + \beta_{695} X_{695i} + \beta_{696} X_{696i} + \beta_{697} X_{697i} + \beta_{698} X_{698i} + \beta_{699} X_{699i} + \beta_{700} X_{700i} + \beta_{701} X_{701i} + \beta_{702} X_{702i} + \beta_{703} X_{703i} + \beta_{704} X_{704i} + \beta_{705} X_{705i} + \beta_{706} X_{706i} + \beta_{707} X_{707i} + \beta_{708} X_{708i} + \beta_{709} X_{709i} + \beta_{710} X_{710i} + \beta_{711} X_{711i} + \beta_{712} X_{712i} + \beta_{713} X_{713i} + \beta_{714} X_{714i} + \beta_{715} X_{715i} + \beta_{716} X_{716i} + \beta_{717} X_{717i} + \beta_{718} X_{718i} + \beta_{719} X_{719i} + \beta_{720} X_{720i} + \beta_{721} X_{721i} + \beta_{722} X_{722i} + \beta_{723} X_{723i} + \beta_{724} X_{724i} + \beta_{725} X_{725i} + \beta_{726} X_{726i} + \beta_{727} X_{727i} + \beta_{728} X_{728i} + \beta_{729} X_{729i} + \beta_{730} X_{730i} + \beta_{731} X_{731i} + \beta_{732} X_{732i} + \beta_{733} X_{733i} + \beta_{734} X_{734i} + \beta_{735} X_{735i} + \beta_{736} X_{736i} + \beta_{737} X_{737i} + \beta_{738} X_{738i} + \beta_{739} X_{739i} + \beta_{740} X_{740i} + \beta_{741} X_{741i} + \beta_{742} X_{742i} + \beta_{743} X_{743i} + \beta_{744} X_{744i} + \beta_{745} X_{745i} + \beta_{746} X_{746i} + \beta_{747} X_{747i} + \beta_{748} X_{748i} + \beta_{749} X_{749i} + \beta_{750} X_{750i} + \beta_{751} X_{751i} + \beta_{752} X_{752i} + \beta_{753} X_{753i} + \beta_{754} X_{754i} + \beta_{755} X_{755i} + \beta_{756} X_{756i} + \beta_{757} X_{757i} + \beta_{758} X_{758i} + \beta_{759} X_{759i} + \beta_{760} X_{760i} + \beta_{761} X_{761i} + \beta_{762} X_{762i} + \beta_{763} X_{763i} + \beta_{764} X_{764i} + \beta_{765} X_{765i} + \beta_{766} X_{766i} + \beta_{767} X_{767i} + \beta_{768} X_{768i} + \beta_{769} X_{769i} + \beta_{770} X_{770i} + \beta_{771} X_{771i} + \beta_{772} X_{772i} + \beta_{773} X_{773i} + \beta_{774} X_{774i} + \beta_{775} X_{775i} + \beta_{776} X_{776i} + \beta_{777} X_{777i} + \beta_{778} X_{778i} + \beta_{779} X_{779i} + \beta_{780} X_{7$$

cancer (X 5Y) is the least prognostic factor. We also provide the 95% confidence interval of the hazard ratios (HR) corresponding to the risk factors; that is, $P[UCL \leq HR \leq LCL] \geq 95\%$

where UCL and LCL are the upper and lower confidence limits and we are at least 95% confident that the hazard ratios will fall into the limits. The following Table 3 provides the three popular global tests of significance which our model is based on. As, the following table shows, our proposed model (??11) is highly significant based on all the three statistical tests. Assumptions of Cox ph Model and Validation of the Proposed Model

In order to apply the CPH model, we must verify that the following three key assumptions are satisfied, prior to its implementation. Failure to satisfy these assumptions will bring about inaccurate decisions about the subject matter. The data consists of times $T_1, T_2, \dots, T_n$ which are either observed survival times or censored times with censoring indicators $\delta_1, \delta_2, \dots, \delta_n$. $\delta_i = 1$ implies T_i is observed, and $\delta_i = 0$ implies T_i is censored. Suppose there are p fixed covariates/risk factors $Z_1, Z_2, \dots, Z_n$ and R_i be the risk set at time T_i denoted as $R_i = \{j : T_j \geq T_i\}$. Given the setup, the partial likelihood, proposed by Oakes (1975) is defined by: $L(\theta) = \prod_{i=1}^n \frac{\exp(\theta' Z_i)}{\sum_{j \in R_i} \exp(\theta' Z_j)}$. (14)

Let $\hat{\theta}$ be the usual estimator of θ that minimizes $L(\theta)$ in (13). Also, let $t(i)$ be the i th ordered observed survival time and $Z(i)$ and R_i the corresponding covariate vector and risk set. Then SCHOENFELD'S RESIDUALS are defined as follows: $r_i = Z(i)' \hat{\theta} - \sum_{j \in R_i} \frac{Z_j' \hat{\theta}}{\sum_{j \in R_i} \exp(\hat{\theta}' Z_j)}$. (15)

The following Figures 5 and 6 illustrate the plot of the scaled Schoenfeld residual against time for all risk factors and interaction terms used in the model (??11), respectively. It shows that there is no pattern as a function of time. Thus, the residuals are randomly scattered with no systematic departures from the horizontal fitted smoothing spline deep line (that is, the residuals are independent of times). A formal test for the PH assumption is given in Table 4. The covariates and the global test are non-statistically significant given by the large p -values. This is a further justification of the validity of the PH assumption for our proposed model. We have included all fourteen risk factors and ten interaction terms in the table. The number of terms in $M_i = \sum_{j=1}^p \hat{\theta}_j(t_i) \exp(\sum_{j=1}^p \hat{\theta}_j X_{ij} + \sum_{k=1}^q \hat{\theta}_{jk} X_{ij} X_{ik}), j = 1, \dots, p$,

where δ_i denotes the event indicator for i th observation, $\hat{\theta}_j(t_i)$ is the estimated cumulative hazard at the final follow-up time for the i th observation. Martingale residuals, M_i , have a skewed distribution. We have, $M_i = 1$ for for maximum possible values and $M_i = -3$. Testing influential observations and Outliers: Often influential observations can cause problems with modeling results. In order to check the influential observations, we visualized the dfbeta values. The dfbeta values estimates the of the i th -patient observation on the regression coefficients $\hat{\theta}_j$. A high value of dfbeta must be investigated carefully. Another method for checking influential observations is by assessing the deviance residuals (symmetric/normalized transformation of the Martingale residuals) plot. The deviance residual is defined by $d_i = \sin(M_i) / \sqrt{2} \sqrt{M_i' M_i} \log(\delta_i / M_i)$.

In the above equation, M_i implies $d_i = 0$. The square root shrinks the large negative martingale residuals, while the logarithm transformation expands those residuals that are close to zero. The distribution of the residuals must approximately be symmetrical around mean zero and standard deviation of one. A very large/small/distant deviance residual values indicate influential observations or outliers. Figure 8 below implies that none of the observations is exceedingly influential individually, on average.

2 Results and Discussions

The following Figure 9 plots the deviance residual and the residual pattern looks fairly symmetrical around zero. The mean deviance residual for our model is .2 which is very small.

Given the risk posed by pancreatic cancer in the past few years, it is imperative to investigate the clinical diagnosis and enhance the therapeutic/treatment strategy of pancreatic cancer. The primary treatment for most types of pancreatic cancer is chemotherapy. Sometimes, with chemotherapy, specific therapy drugs are used. Usually, surgery and radiation therapy do not fall under crucial treatments VII.

3 Conclusion

for pancreatic cancer, but they might be used in exceptional circumstances. Also, the treatment approach for children with pancreatic cancer can be slightly different from that used for adults. Several research approaches and statistical methodologies (23) (24) have been developed to cure pancreatic cancer patients and boost their survival times. Chakraborty & Tsokos (to be published) performed data-driven research on pancreatic cancer patients by performing parametric analysis to improve the survival probabilities of patients of different cancer stages. In the present study, we initially investigated if there exists any statistically significant difference between the true mean survival times of the male and female pancreatic cancer patients using the Wilcoxon two-sample rank-sum test. The p -value (.47 > .05) of the test result suggests that there is no evidence of a significant difference between the true mean survival times of the males and the females. Hence, we proceed to perform to develop the Cox-PH (CPH) model with the combined information of male and female patients. While developing the CPH model, it is very important to justify the model assumptions. In the preliminary analysis, we found that all of the risk factors except age (X 1) did not satisfy the proportional hazard assumption. Thus, we introduced stratification in our model by dividing the covariate age into two groups. By doing stratification, we obtained more valid estimates of the other covariates, and the proportional hazard assumption was satisfied for all risk factors, including age. Performing stratification, we restrict the effect of the covariates to be the same

for each stratum. Our final developed Cox-PH model given by equation (11) identified all the significant risk factors along with all the significant interaction terms as contributing to the hazard. After building our model, we proceed to rank all significant individual risk factors and all possible significant interactions according to the hazard ratio, as shown in Table 2. From Table 2, we observe that X 6Y (patients having diabetes), X 4Y (patients taking ibuprofen regularly), X 2D (patients who are in stage distant (Cancer has spread to distant parts of the body)), X 9F (sex), and X 15Y (hypertension) are the most contributing risk factors individually to the survival of patients with a hazard ratio (HR) of 2.39, 1.83, 1.63, .7, and .7, respectively. For the risk factor X 6Y , $HR = 2.39$ indicates a strong association between the patients having diabetes and increased risk of death due to pancreatic cancer. Keeping the other covariates constant, being a diabetic patient has a 2.39-fold increase in the hazard of death; that is, 2.39-fold increased risk (or decreased survival). It is important to note that according to the American Cancer Society, one of the main risk factors of pancreatic cancer is diabetes which is supported by our study. Also, we have found that those who take ibuprofen regularly have an increased risk of 1.83-fold than those who do not take the medication on a regular basis. Also, being a female has approximately 30% less hazard than a male patient. Among the most significant interactions we have X 2R X 8Y , X 15Y X 12Y , X 12Y X 9F , X 3Y X 15Y , X 3Y X 9F , X 15Y X 9F , and X 2R X 11 with hazard ratio 8.84, 2.28, 1.98, 1.83, 1.5, 1.37, and 1.01 respectively. The most contributing risk factor is an interaction term (X 2R X 8Y) (patients with emphysema and cancer stage regional with $HR = 8.84$). However, they do not contribute significantly to survival. We see that X 15Y (hypertension) has a lower risk of survival ($HR = .79$). However, interacting with X 12Y (diverticulosis), it has a hazard ratio of 2.28. Also, interacting with X 3Y (person who uses Aspirin Regularly), it has a hazard ratio of 2.28. It is also important to note that X 3Y individually has lower risk (better survival) with $HR = .6$. Although X 12Y (diverticulosis) and X 9F (female) has a hazard ratio less than one, their combined effect remains significant with $HR = 1.98$.

In this study, we have estimated the survival probabilities of patients diagnosed with pancreatic cancer using the semi-parametric Cox proportional hazard (CPH) model. We believe the proposed Cox-PH model given by equation ??11) gives an accurate estimate of the survival probability of patients diagnosed with pancreatic cancer. The stratification of the age produced more reliable estimates of the risk factor included in the CPH model. We identified seven significant risk factors and ten significant interaction terms as contributing to the survival probability of patients diagnosed with pancreatic



Figure 1: Figure 1 :

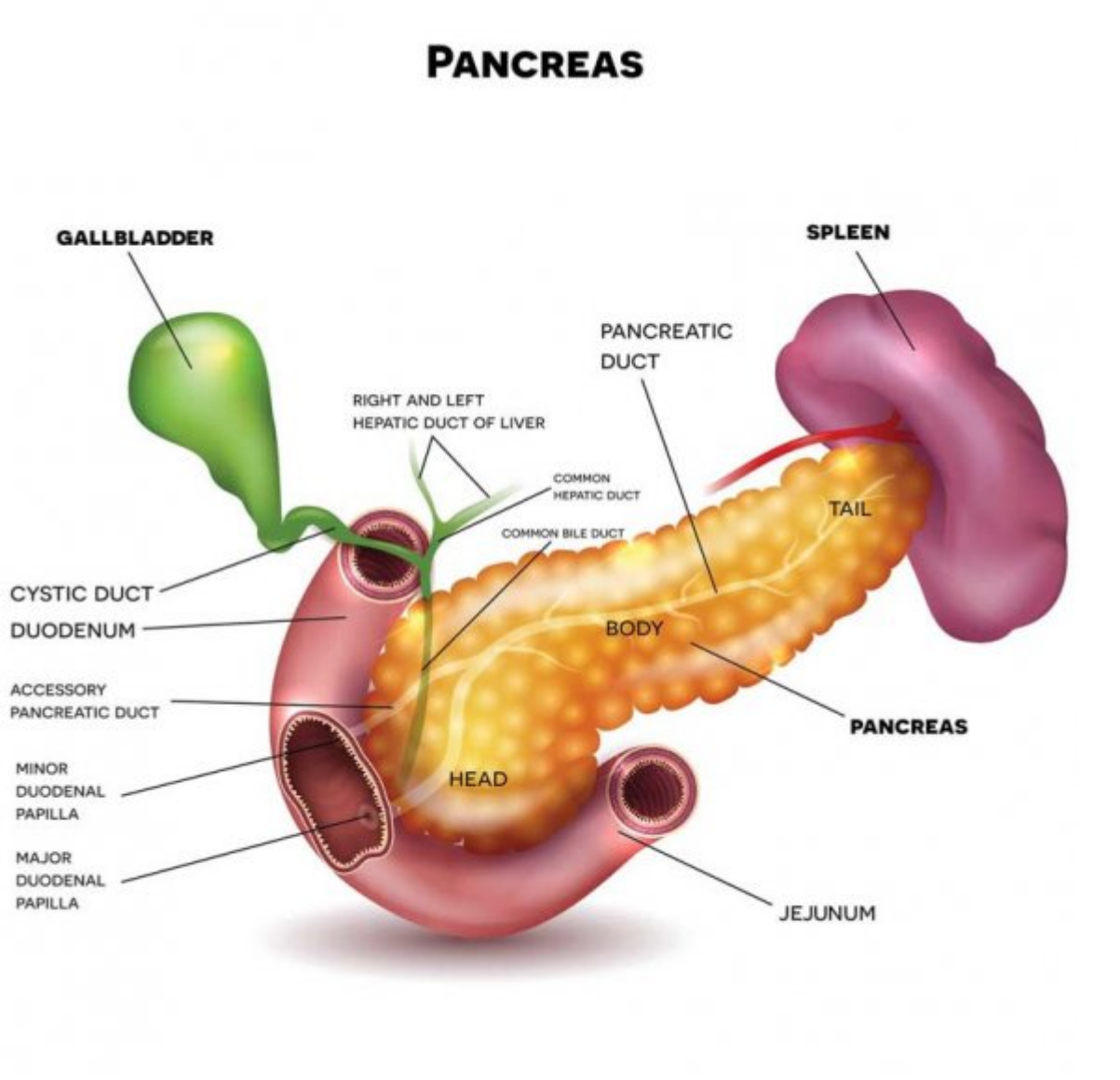


Figure 2:

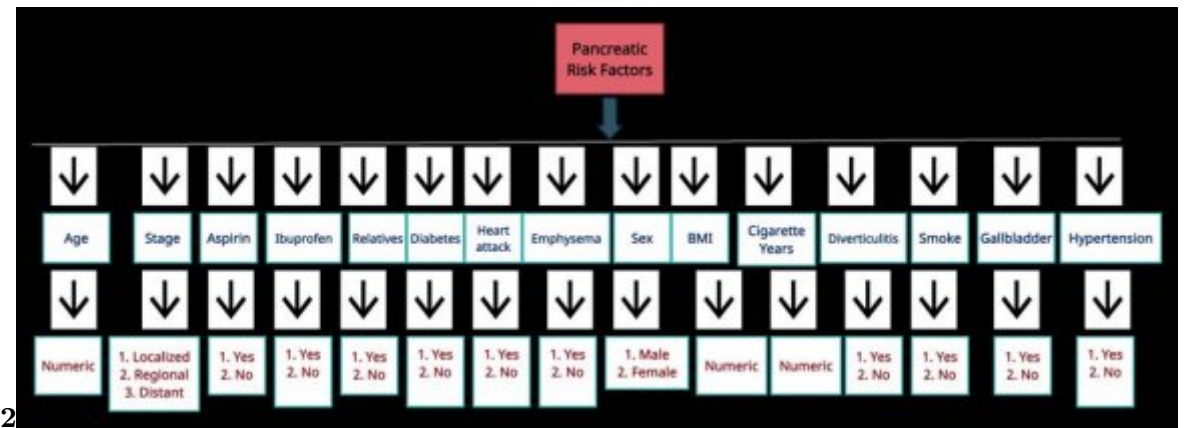
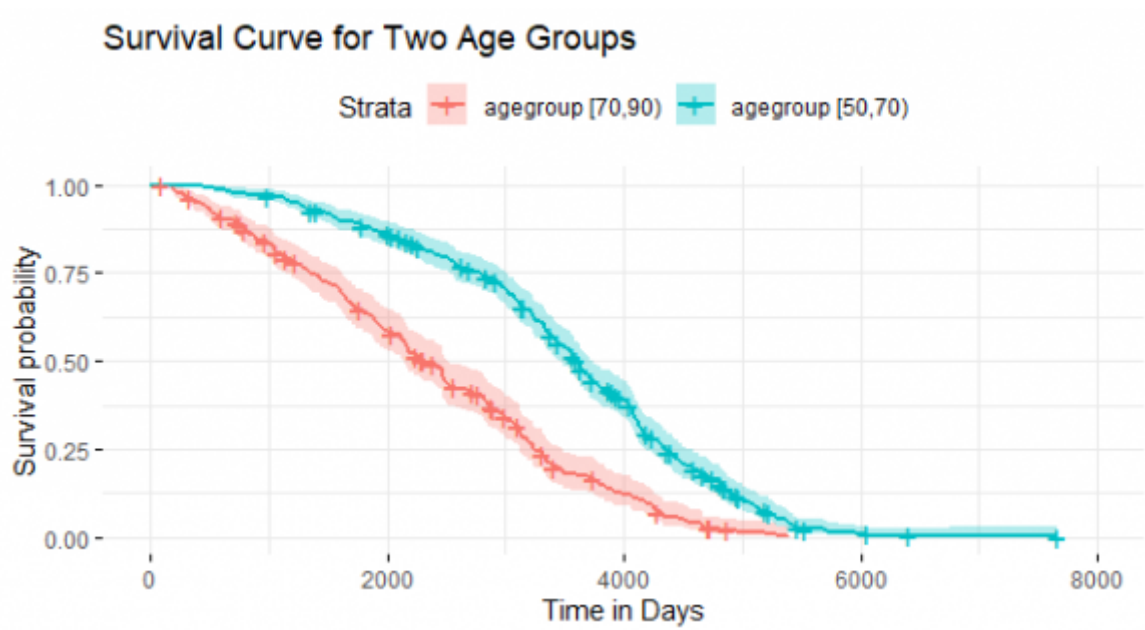
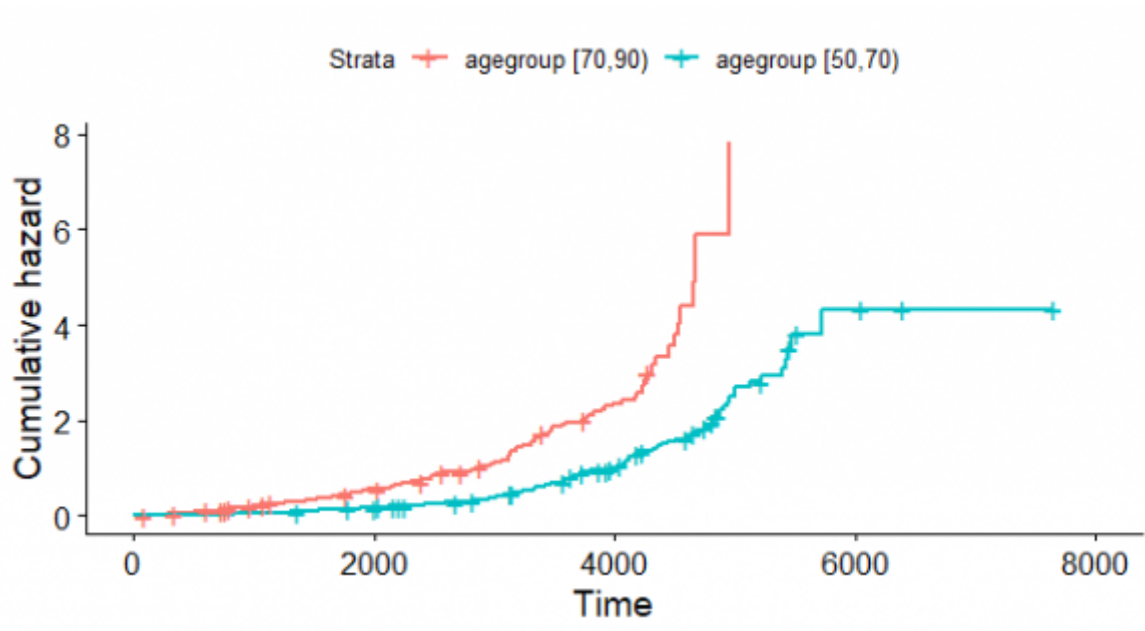


Figure 3: Figure 2 :



3

Figure 4: Figure 3 :



4

Figure 5: Figure 4 :

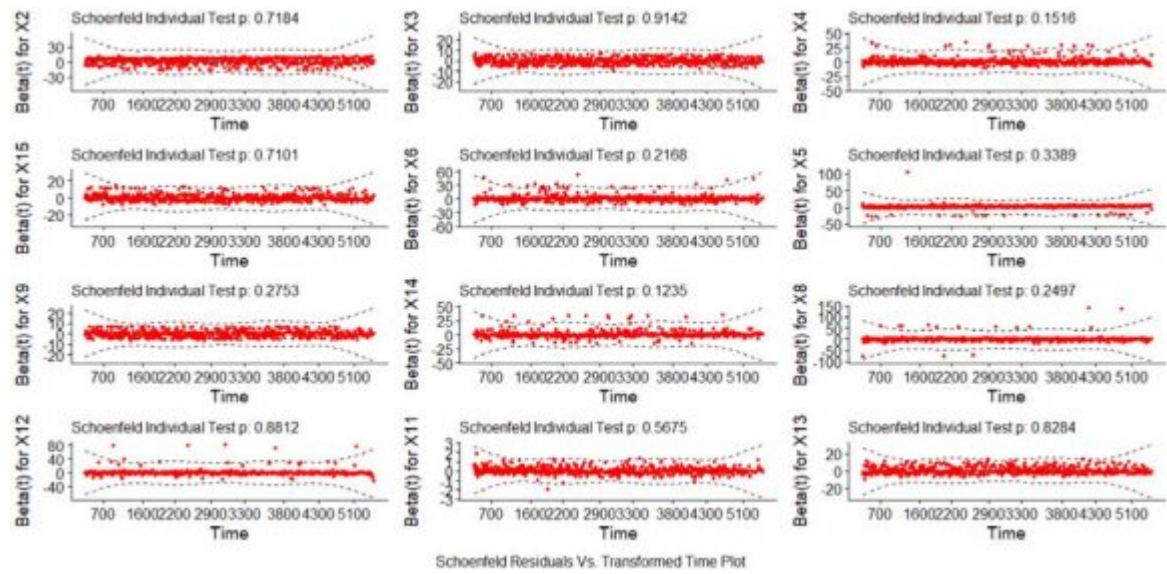


Figure 6:

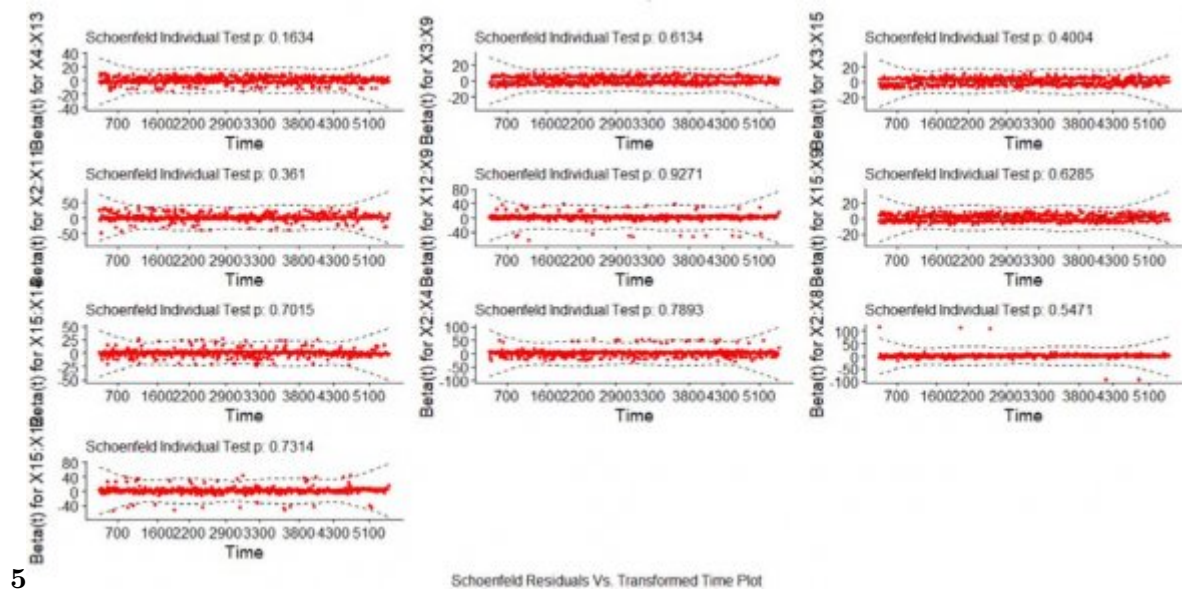


Figure 7: Figure 5 :

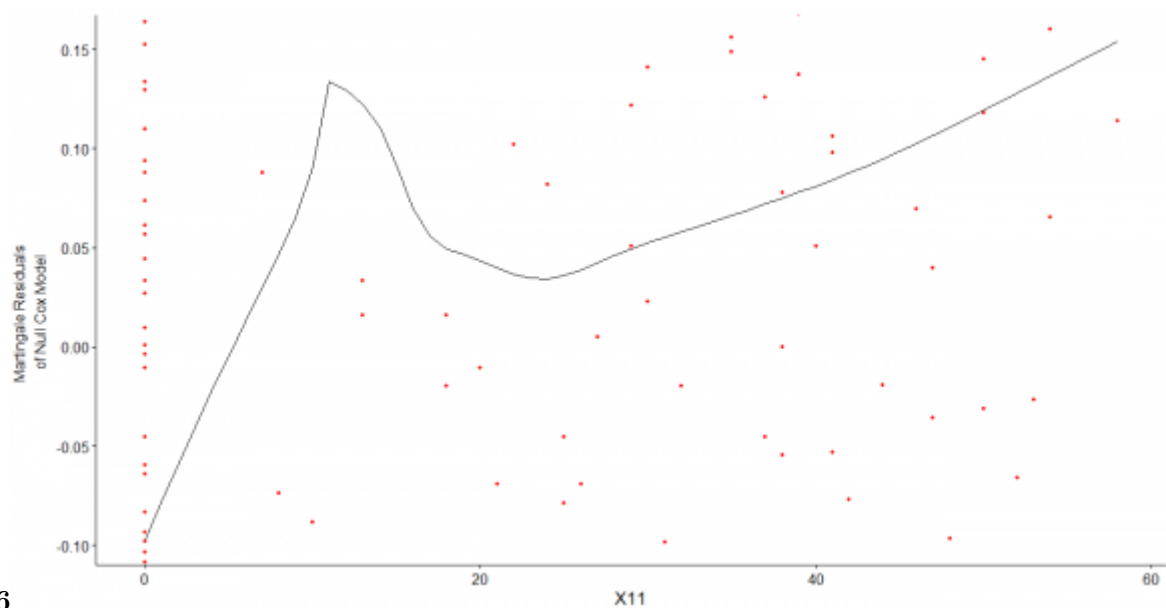


Figure 8: Figure 6 :

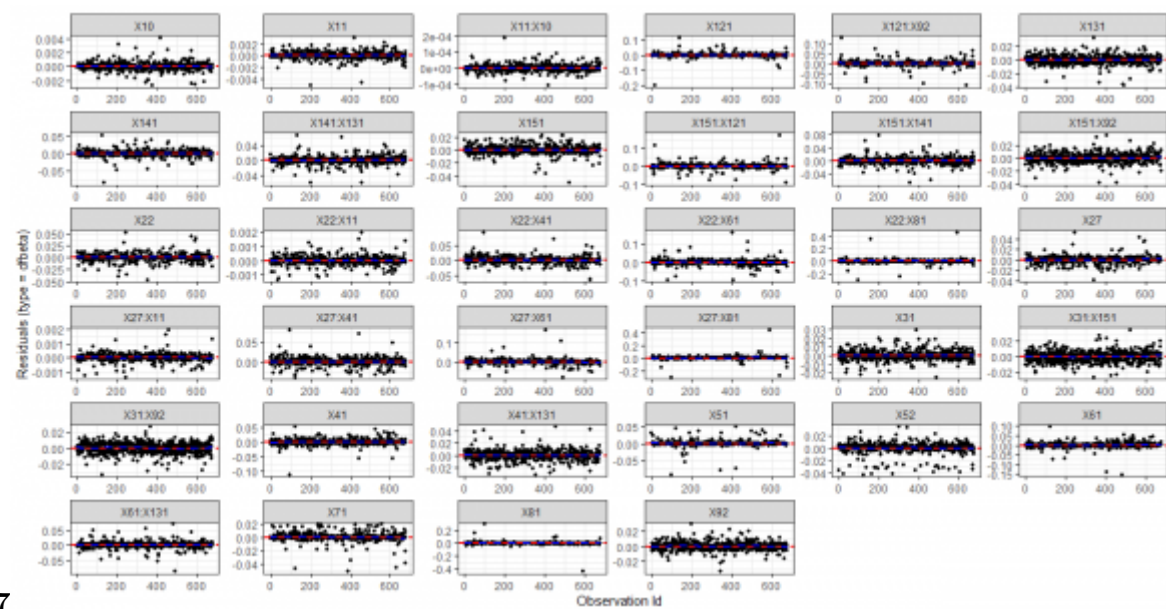
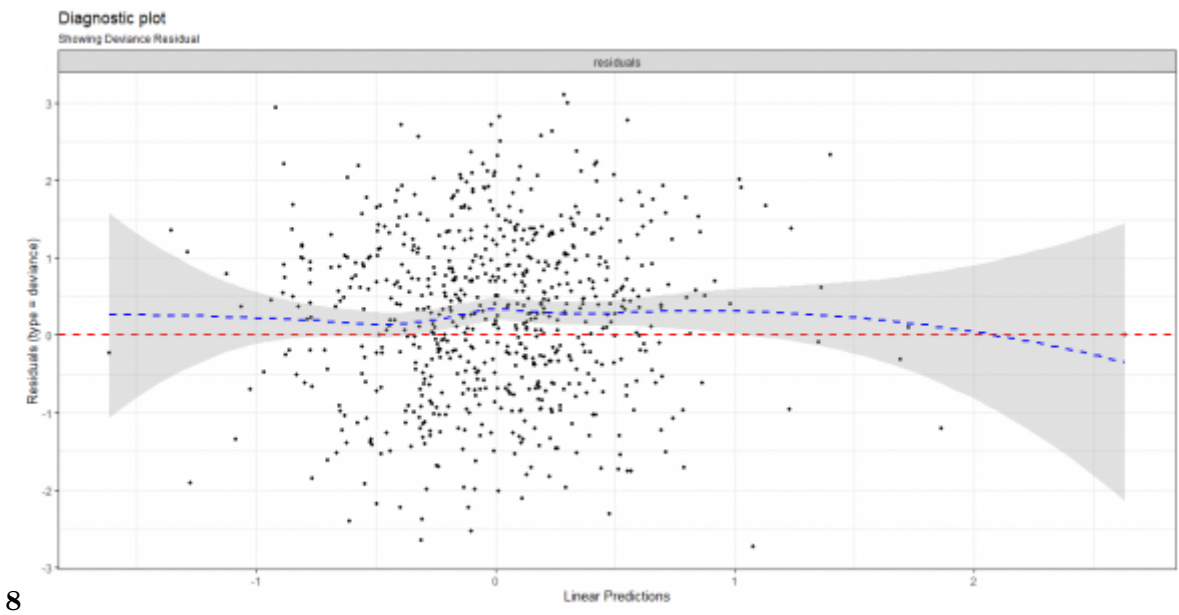


Figure 9: Figure 7 :



1

group [70,90).		
Risk Factors		Count
Stage	Localized 135	
	Regional 178	
	Distant	364
Aspirin	Yes No	333 344
Ibuprofen	Yes No	168 509
Relatives	Yes No	650 27
Diabetes	Yes No	83 594
Heart attack	Yes No	84 593
Emphysema	Yes No	19 658
Sex	Male Female	388 289
BMI		677
Cigarette Years		677
Diverticulosis	Yes No	41 636
Smoke	Yes No	404 273
Gallbladder	Yes No	98 579
Hypertension	Yes No	256 421

Figure 11: Table 1 :

	Rank	Risk	Factors	coeff(?)	HR [exp(?)]	[S.E(?)]	Lower 95%	Upper 95%	
	1		X 2R X 8Y	2.18	8.84	.96	1.32	59.1	
	2		X 6Y	.87	2.39	.33	1.2	4.6	
	3		X 15Y X 12Y	.8	2.28	.38	1.07	4.87	
	4		X 12Y X 9F	.68	1.98	.39	.92	4.25	
	5		X 4Y	.61	1.834	.25	1.27	2.62	
	6		X 3Y X 15Y	.6	1.831	.18	1.11	3.02	
	7		X 2D	.5	1.63	.17	1.16	2.3	
	8		X 3Y X 9F	.41	1.5	.18	1.06	2.13	
	9		X 15Y X 9F	.32	1.37	.18	.96	1.96	
Volume	10	11	12	13	14	X 2R X 11 X 9F	0.01	1.01	.7
XXI	15	16	17			X 15Y X 15Y X	-.35	.69	.63
Issue III						14Y X 13Y X 4Y	-.37	.63	.6
Version I						X 3Y X 2R X 4Y	-.47	.59	.55
						X 5Y	.46	-.53	.2
							-.52	-.6	

D D D D
) F
(
Medical
Research
Global
Journal
of

© 2021 Global
Journals

Figure 12: Table 2 :

Test	Test Statistics	Value	df	p-value
Likelihood Ratio	96.6	34	7*10 ⁸	
Wald Test	100.8	34	2*10 ⁸	
Score (log-rank) Test	109.9	34	6*10 ¹⁰	

Volume XXI Issue III Version I
D D D D) F
(
Medical Research
Global Journal of
© 2021 Global Journals

Figure 13:

Figure 14: Table 3 :

1

Figure 15: 1 .

4

[Note: is greater than Table2since we have included all of the fourteen individual risk factors used in our analysis in]

Figure 16: Table 4

4

Volume XXI Issue III Version I
D D D D) F
(
Medical Research
Global Journal of

Figure 17: Table 4 .

4

Figure 18: Table 4 :

.1 Acknowledgement

The authors are thankful to the National Cancer Institute (NIH) for making the data available.

.2 Funding Not Applicable

.3 Ethics approval and consent to participate Not Applicable

.4 Competing interests

There are no competing interests to declare.

.5 References Références Referencias

cancer, as described in Table ?? . We also ranked those risk factors and their interactions based on the hazard ratio. There have not enough studies been done in the literature that incorporates the significant interaction effect of two risk factors. Interaction effects play a major role as a prognostic factor in addition to the individual risk factors in the CPH model. We found some of the risk factors used in our study individually have hazard less than one, but by combining with some other risk factor, the hazard was more than 1.5, and the combined effect was significant. Our final proposed Cox-PH model is of very high quality, robust, and efficient, given by the fact that it satisfies all the major assumptions described in Section 5. The stepwise model selection procedure was utilized to carefully assess and select the risk factors and the interaction term based on their statistical significance to the survival probability. Depending on the survival analysis of the survival times based on the CPH model of the pancreatic cancer patients, we recommend the following. 1. Besides the survival time of patients, if any additional details regarding some of the potential risk factors are known, then use of the Cox proportional hazard (CPH) model can reflect a better picture of covariate effect on survival via hazard ratio.

2. Before implementing the developed CPH model, one should be careful about the fact that the CPH model assumptions are satisfied. In our present analysis, we justified the key assumptions of the CPH model.

3. The significant two-way interaction effects of the risk factors in the CPH model should not be excluded because they can significantly influence the prediction accuracy of the model and survival rate of pancreatic cancer patients, which might lead to serious clinical and therapeutic/treatment issues.

4. The ranking of the individual and interacting risk factors can be wisely used in pancreatic cancer research to improve the treatment options.

[Yong and Tsokos ()] , X Yong , C Tsokos . *PROBABILISTIC COMPARISON OF SURVIVAL ANALYSIS MODELS USING SIMULATION AND CANCER DATA* 2009.

[Akaike ()] ‘A new look at the statistical model identification’. H Akaike . *IEEE Trans. Autom Control* 1974. 19 p. .

[Chakraborty et al. (2021)] ‘A Real Data-Driven Analytical Model to Predict Happiness’. Aditya Chakraborty , & Chris , P Tsokos . *Sch J Phys Math Stat* 2021 Mar. 8 (3) p. .

[Perera and Tsokos ()] ‘A Statistical Model with Non-Linear Effects and Non-Proportional Hazards for Breast Cancer Survival Analysis’. M Perera , C Tsokos . *Advances in Breast Cancer Research* 2018. 7 p. .

[Aspirin (Categorical) (X 3) : Does the person use Aspirin Regularly?] *Aspirin (Categorical) (X 3) : Does the person use Aspirin Regularly?*,

[Asano et al. (2014)] ‘Assessing the prediction accuracy of cure in the Cox proportional hazards cure model: an application to breast cancer data’. J Asano , A Hirakawa , C Hamada . 10.1002/pst.1630. 25044997. *Pharm Stat* 2014 Nov-Dec; 13. 2014 Jul 16. (6) p. .

[Diverticulosis] *Categorical) (X 12) : Did the participant ever have diverticulitis or diverticulosis? 13. Smoke (Categorical) (X 13) : Has the patient ever smoked cigarettes regularly for six months or longer?*, Diverticulosis .

[Ibuprofen] *Categorical) (X 4) : Does the person use Ibuprofen Regularly?*, Ibuprofen .

[Emphysema] *Categorical) (X 8) : Did the patient ever have emphysema?*, Emphysema .

[Cigarette Years (numeric) (X 11) : The total number of years the patient smoked] *Cigarette Years (numeric) (X 11) : The total number of years the patient smoked*,

[Du et al. ()] ‘Comparison of the flexible parametric survival model and Cox model in estimating Markov transition probabilities using real-world data’. X Du , M Li , P Zhu , J Wang , L Hou , J Li . 10.1371/journal.pone.0200807. <https://doi.org/10.1371/journal.pone.0200807> *PLoS ONE* 2018. 13 (8) p. e0200807.

[Current Body Mass Index (BMI) at Baseline (In lb/in2)] *Current Body Mass Index (BMI) at Baseline (In lb/in2)*,

[Diabetes (Categorical) (X 6) : Did the patient ever have diabetes?] *Diabetes (Categorical) (X 6) : Did the patient ever have diabetes?*,

- [Michaud (2004)] ‘Epidemiology of pancreatic cancer’. D S Michaud . 15238885. *Minerva Chir* 2004 Apr. 59 (2) p. .
- [Ferlay ()] *GLOBOCAN 2012: cancer incidence and mortality worldwide: IARC CancerBase No. 11. International Agency for Research on Cancer*, J Ferlay . <http://globocan.iarc.fr> 2013.
- [Heart attack (Categorical) (X 7): Did the participant ever have coronary heart disease or a heart attack?]
Heart attack (Categorical) (X 7): Did the participant ever have coronary heart disease or a heart attack?,
- [L Douglas Case et al. ()] ‘Interpreting Measures of Treatment Effect in Cancer Clinical Trials’. Gretchen L Douglas Case , Electra D Kimmick , Kurt Paskett , Robert Lohman , Tucker . *The Oncologist* 2002. 7 p. .
- [Therneau et al. ()] ‘Martingale-Based Residuals for Survival Models’. T Therneau , P Grambsch , T Fleming . 10.2307/2336057. *Biometrika* 1990. 77 (1) p. .
- [Winnett ()] ‘Miscellanea. A note on scaled Schoenfeld residuals for the proportional hazards model’. A Winnett . 10.1093/biomet/88.2.565. *Biometrika* 2001. 88 (2) p. .
- [Fernandez ()] ‘Non-parametric maximum likelihood estimation of censored regression models’. L Fernandez . 10.1016/0304-4076(86)90011-4. *Journal of Econometrics* 1986. 32 (1) p. .
- [Numeric) (X 1): Age of diagnosis of the patient Age] ‘Numeric) (X 1): Age of diagnosis of the patient’. *Age*
- [Sashegyi and Ferry ()] ‘On the Interpretation of the Hazard Ratio and Communication of Survival Benefit’. A Sashegyi , D Ferry . 10.1634/theoncologist.2016-0198. <https://doi.org/10.1634/theoncologist.2016-0198> *The oncologist* 2017. 22 (4) p. .
- [Kleeff et al. ()] ‘Pancreatic cancer’. J Kleeff , M Korc , M Apte . 10.1038/nrdp.2016.22. <https://doi.org/10.1038/nrdp.2016.22> *Nat Rev Dis Primers* 2016. 2 p. 16022.
- [Li et al. ()] *Pancreatic cancer. The Lancet*, D Li , K Xie , R Wolff , J L Abbruzzese . 10.1016/S0140-6736(04)15841-8. 2004. 363 p. .
- [Vincent et al. ()] *Pancreatic cancer. The Lancet*, A Vincent , J Herman , R Schulick , R H Hruban , M Goggins . 10.1016/S0140-6736(10)62307-0. 2011. 378 p. .
- [Chakraborty and Tsokos ()] ‘Parametric and Non-Parametric Survival Analysis of Patients with Acute Myeloid Leukemia (AML)’. A Chakraborty , C P Tsokos . 10.4236/ojapps.2021.111009. <https://doi.org/10.4236/ojapps.2021.111009> *Open Journal of Applied Sciences* 2021. 11 p. .
- [Grambsch and Therneau ()] ‘Proportional hazards tests and diagnostics based on weighted residuals’. P M Grambsch , T M Therneau . 10.1093/biomet/81.3.515. *Biometrika* 1994. 81 (3) p. .
- [Cox ()] ‘Regression Models and Life-Tables’. David R Cox . *Journal of the Royal Statistical Society, Series B* 1972. 2985181. JSTOR. 34 (2) p. .
- [Relatives (Categorical) (X 5): The number of first-degree relatives with pancreatic cancer] *Relatives (Categorical) (X 5): The number of first-degree relatives with pancreatic cancer*,
- [Sex of the individual] *Sex of the individual*, (X 9)
- [Stage (Categorical) (X 2): Pancreatic Cancer Stages, categorized as a) localized, b) regional, and c) distant]
Stage (Categorical) (X 2): Pancreatic Cancer Stages, categorized as a) localized, b) regional, and c) distant,
- [Mamudu et al. ()] ‘Survival Analysis of Multiple Myeloma Cancer (MMC) Using the Cox-Proportional Hazard Model’. L Mamudu , C P Tsokos , O E Oluwaseun . *Med Clin Res* 2020. 5 (7) p. 147.
- [Kleinbaum and Klein ()] ‘The Cox Proportional Hazards Model and Its Characteristics’. D G Kleinbaum , M Klein . 10.1007/978-1-4419-6646-9_3. https://doi.org/10.1007/978-1-4419-6646-9_3 *Survival Analysis. Statistics for Biology and Health* 2012. Springer.
- [Kalbeisch and Prentice ()] *The Statistical Analysis of Failure Time Data*, J D Kalbeisch , R L Prentice . 1980. 1980. New York: John Wiley Sons. p. .
- [Theory of Partial Likelihood Ann. Statist (1986)] ‘Theory of Partial Likelihood’. 10.1214/aos/1176349844. <https://doi.org/10.1214/aos/1176349844> *Ann. Statist* March, 1986. 14 (1) p. .
- [Tom and Guidelines ()] Brody Tom , ; Guidelines . *Clinical Trials: Study Design, Endpoints and Biomarkers, Drug Safety*, 2011. 2011. Academic Press. p. .