Anabela Malho Guedes

Licenciada em Medicina

Competing Risks Modeling in Survival Analysis: Valuation in Peritoneal Dialysis

Dissertação para obtenção do Grau de Mestre

em Estatística para a Saúde

Orientador: Professora Doutora Sara Simões Dias,

Escola Superior de Saúde – Politécnico de Leiria,

Faculdade de Ciências Médicas – Universidade Nova de Lisboa

Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical

13 de Novembro de 2019

Competing Risks Modeling in Survival Analysis: Valuation in Peritoneal Dialysis
Copyright © Anabela Malho Guedes, Instituto de Higiene e Medicina Tropical e Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.
O Instituto de Higiene e Medicina Tropical Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Dedicado à minha Família,
à Mariana, ao Pedro, à Carolina e ao Tiago,
se as palavras não chegarem, que cheguem os exemplos...

This thesis is dedicated to my Family,
To Mariana, Pedro, Carolina e Tiago,
If words aren't enough, rely on the example...

Acknowledgements

I would like to express my very great appreciation to Professor Sara Dias who shared her knowledge and insights with me and had the sensitivity to keep me motivated and calm through the ups and down of this journey.

Suggestions made by Professor Laetitia Teixeira and Professor Luca Scrucca were of great assistance in my thesis progress. Thank you!

Also, I would like to thank the professors Maria do Rosário Oliveira Martins and Zulmira Hartz who have always believed in me and stimulated me to pursue higher goals.

My special thanks are extended to my family who have assisted and encouraged me, with endless patience.

Resumo

A análise de sobrevivência é uma pedra angular na pesquisa médica. A estimativa de Kaplan-Meier é o método estatístico mais amplamente utilizado, mas a presença de riscos competitivos viola o seu pressuposto fundamental que afirma que o mecanismo de censura é independente do tempo de sobrevida. Isto leva à sobrestimação da probabilidade cumulativa de falha de risco específico. Assim, nessas circunstâncias a estimativa da função de incidência cumulativa e análise de riscos competitivos são métodos preferíveis. O objetivo deste estudo foi comparar os métodos de análise de sobrevivência diferentes: estimativas de Kaplan-Meier e função de incidência cumulativa numa coorte de doentes em Diálise Peritoneal (DP).

Avaliou-se a sobrevivência de 123 pacientes incidentes em DP num hospital universitário. Foram estimadas as curvas de Kaplan-Meier, função de incidência cumulativa, risco específico e subdistribuição dos riscos.

O grupo *PD first* apresentou sobrevivência superior, quando comparado com os doentes transferidos para DP, a probabilidade de morte foi sobrestimada pelo método de Kaplan-Meier. O risco específico revelou como factores associados a maior sobrevivência a opção *PD first*, níveis de albumina e função renal residual superiores e menor índice de Charlson. Considerando a análise de riscos competitivos, na subdistribuição dos riscos a opção *PD first,* a função renal residual e índice de Charlson permaneceram significativos, mas não os níveis séricos de albumina.

É primordial reconhecer a presença de riscos competitivos e adoptar um método de risco competitivos em estudos com vários outcomes, como em estudos de Diálise Peritoneal, para estimar a incidência cumulativa e produzir resultados mais precisos.

Palavras-chave: Riscos competitivos, função de incidência cumulativa, modelo Fine-Gray, Diálise Peritoneal, modelo subdistribuição dos riscos, análise de sobrevivência

Abstract

Survival analysis is a cornerstone in medical research. For that purpose Kaplan-Meier estimate is the most widely used statistical test, but the presence of competing risks violates the fundamental assumption that states the censoring mechanism is independent of survival time. This leads to overestimation of the cumulative probability of cause-specific failure. Cumulative incidence estimate and competing risks analysis are preferred. The purpose of this study was to compare different survival analysis methods: Kaplan-Meier and cumulative incidence function estimates in a cohort of Peritoneal Dialysis (PD) patients.

The survival of 123 incident patients on PD in a university hospital was evaluated. Kaplan-Meier, cumulative incidence function, cause-specific and subdistribution hazards were performed.

PD first group had a better survival when compared with PD transfer patients, the probability of death was overestimated by the Kaplan-Meier method. The cause-specific hazard revealed factors associated with survival were PD first option, higher albumin levels, higher renal residual function and lower Charlson Index. Taking competing risks into account in the subdistribution hazard model PD first option, renal residual function and Charlson Index remained significant, but not albumin levels.

It is primordial to recognize the presence of competing risks in order to adopt a competing risk method in studies with multiple outcomes, as in Peritoneal Dialysis studies, to estimate cumulative incidence and yield more accurate results.

Keywords: Competing risks, Cumulative incidence function, Fine and Gray model, Peritoneal dialysis, Subdistribution hazard model, Survival analysis

Index

List of Tables

List of Figures

1. Introduction

Survival analysis is a cornerstone of health outcomes research. It measures the time span from some time origin until occurrence of the event of interest and then deals with modeling, estimating and testing for time to event data. Common examples include studies of treatment related effects to progression-free survival in cancer patients, time from infection by the HIV virus until the development of AIDS, or time to receiving a transplant in dialysis patient.

Censoring – when we fail to observe the event - may complicate analysis of survival data. Types of censoring include left censoring – when the event occurred before the observation time, right censoring – when the event occurs after the follow-up time, or interval censoring, when observation is not continual, but occurs at discrete times. Only the time between the beginning of the observation period and the occurrence of the event is known. Censored observations account not only for losses to follow-up but also deaths from other causes or other events if they hinder development of the endpoint under consideration (1).

In these disease/recovery processes, although only one type of event is designated as the event of interest, often more than one type of event plays a role. The other event types may prevent the event of interest from occurring. For instance, leukemia relapse or AIDS may be unobservable because the subject died before this diagnosis developed. This kind of situation is referred to as “competing risks” – an event that precludes a subject from being at risk for an outcome under investigation.

In survival analysis, two functions of time are of particular interest: the survival function and the hazard function. The Kaplan-Meier method can be used to estimate the survival function from the observed survival times with the only assumption that the censoring mechanism is independent of survival time. Treating the events of the competing causes as censored observations violates this fundamental assumption and can result in an over-estimation of the cumulative probability of cause-specific failure. The Cox proportional hazards model can still be used, but the interpretation of the results is different (2).

Emerging evidence now suggests that in the presence of competing risks the cumulative incidence function, a method that takes into account competing risks

occurrence, is the appropriate method use to estimate the probability of occurrence of the event of interest in the presence of other events.

This paper evaluates the statistical relevance of using the cumulative incidence estimate over the Kaplan-Meier method in survival analysis under right censoring. Chapter 2 reviews the hazard function estimate and the commonly used Kaplan-Meier approach. Chapter 3 explains competing risks settings, cumulative incidence function and subdistribution hazard. In Chapter 4 and Chapter 5 is presented a case study in Peritoneal Dialysis and the statistical methodology used. Results of comparison of the two types of estimates are provided in Chapter 6. Discussion and conclusions are held in Chapter 7 and Chapter 8, respectively.

2. The hazard function and Kaplan-Meier estimates

2.1 The hazard function

A central concept in survival analysis is the hazard function (also known as the failure rate, hazard rate, or force of mortality). The hazard function gives the instantaneous failure rate of an individual conditioned on the fact that the individual survived until a given time. That is,

(i)

$$\lambda\,(t) = \lim_{\Delta t \to 0} \frac{Prob\,(t \le T < t + \Delta t | T \ge t)}{\Delta t}$$

where Δt is a very small time interval. The hazard function is finite and nonnegative. However, these values do not correspond to probabilities and might be greater than 1.

The hazard function is related to the probability density function, $f(t)$, cumulative distribution function, $F(t)$, and survivor function, $S(t)$, as follows:

(ii)

$$\lambda\,(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)}$$

Thus, $\lambda(t)$ can be seen as the conditional probability that the event of interest occurs in the interval $[t, t + \Delta t]$, given that it has not occurred before time t. This function is particularly useful in determining the appropriate failure distributions when competing risks are present.

A quantity related to the hazard function is the probability of an individual surviving beyond time t, the *survival function*. The survival function, $S(t)$, is the exponential of the negative of the cumulative hazard function, i.e.

(iii)

$$S(t) = \exp\big(-H(t)\big) = \exp(-\int_0^t \lambda(u)du\,)$$

where $H(t)$ is the cumulative hazard function (3).

2.2 Kaplan-Meier estimate

Kaplan-Meier curve was designed in 1958 by Edward L. Kaplan and Paul Meier, when they published a collaborative paper on how to deal with time to event data. Kaplan-Meier survival analysis, or product limit estimator, is a non-parametric method that estimates survival probability beyond time t in right-censored data. It creates an estimator of survival function based on empirical data, taking censoring into account (4,5).

One basic assumption is that the censoring distribution and the event time distribution are independent. Then, at each point in time, the individuals who are censored can be represented by those who remain under observation. The Kaplan–Meier estimator assumes that the distribution is discrete instead of continuous, with the events only occurring at these observed time points (2).

To acknowledge the estimation by Kaplan-Meier, consider

- $0 < t_1 < t_2 < \ldots < t_N$ to be the ordered distinct time points at which events occur;

- $R(t)$ as the *risk set* (subjects in follow-up that have not reached their event) at time t;

- for each t_j define $R_j = R(t_j)$ to be the risk set at t_j and n_j the size of this risk set, the *number at risk*;

- for each t_j, define d_j as the number of observed events at t_j;

- events are assumed to occur only at the observed event times, subjects are considered "alive just before time t_j", often denoted as "alive beyond the previous time point t_{j-1}" or "alive at $t-$".

The probability of failure at a certain time is a conditional probability of having an event at that time, given that an individual has not had an event just prior to that time. This conditional probability of failing can be estimated as the probability of surviving just prior to that time multiplied by the number of patients with the event at that time, divided by the number of patients at risk. Under the assumption of independent censoring, subjects in the risk set are representative for all subjects alive at t_{j-1}, so $\lambda(t_j)$ can be estimated simply by the at risk sample proportion that fail at t_j, i.e. by

(iv)

$$\widehat{\lambda}\,(t_j) = \frac{d_j}{n_j}$$

Cumulative incidence of a failure is the sum of these conditional probabilities over time.

Survival probability at a certain time (t_j) is the product of the probability of surviving up to t_{j-1} and the conditional probability of surviving up to t_j given still alive beyond t_{j-1}:

(v)

$$\hat{S}\,(t_j) = \hat{S}\,(t_{j\_1})\left(1 - \widehat{\lambda}\,(t_j)\right) = \hat{S}\,(t_{j\_1})(1 - \frac{d_j}{n_j})\,)$$

The Kaplan-Meier estimate of survival probability is then the product of these conditional probabilities up until that time. The estimation by Kaplan-Meier is given by (2):

(vi)

$$\hat{S}\,(t) = \prod_{j:t_j \le t} (1 - \frac{d_j}{n_j})$$

In a practical example, supposing that 10 patients enter a clinical trial at different time points, are followed until a specified time, and have independent and distinct failure times. The event of interest is relapse. Subjects who fail from other causes or who are still alive at the specified time point are censored. Patients who are lost to follow up during the study period are considered alive and relapse-free at the last time seen alive and thus censored. Of note, ‘‘event’’ and ‘‘failure’’ are used interchangeably in the literature, and the event of interest might be death from any cause, relapse, treatment-related mortality, or other (6).

In order to calculate the Kaplan-Meier estimate of cumulative relapse-free survival probability, the observed failure times for all patients need to be ordered from the smallest to the largest, irrespective of the censoring status, and each failure time is paired with the information of censoring status. Let t_0 represent the time zero or the time of study entry, t_1 denote the smallest observed failure time, and t_{10} denote the largest observed failure time to the event. Then the Kaplan-Meier estimate of relapse-free survival probability, S (t), is

(vii)

$$S_{KM}^{rel}(t_0) = 1\ (because\ everyone\ is\ relapse\ free\ at\ study\ entry)$$

$$S_{KM}^{rel}(t_1) = S(t_0) * [\frac{no.of\ relapse - free\ patients\ at\ t_1}{no.of\ patients\ at\ risk\ just\ prior\ to\ t_1}]$$

$$S_{KM}^{rel}(t_2) = S(t_1) * [\frac{no.of\ relapse - free\ patients\ at\ t_2}{no.of\ patients\ at\ risk\ just\ prior\ to\ t_2}]$$

$$S_{KM}^{rel}(t_{10}) = S(t_9) * [\frac{no.of\ relapse - free\ patients\ at\ t_{10}}{no.of\ patients\ at\ risk\ just\ prior\ to\ t_{10}}]$$

The Kaplan-Meier estimate of incidence of relapse at a specified time point is then the probability of relapse-free survival just prior to that time, multiplied by the number of relapses at that time, divided by the number of patients at risk (that is, alive, relapse- free, and not lost to follow-up) just prior to that time (6). The Kaplan-Meier survival curve shows that it is a step function with jumps at the observed event times. The pattern of these jumps depends on the number of events at each event time t_j and on the pattern of the censored observations prior to t_j (shown as tick marks on the curve).

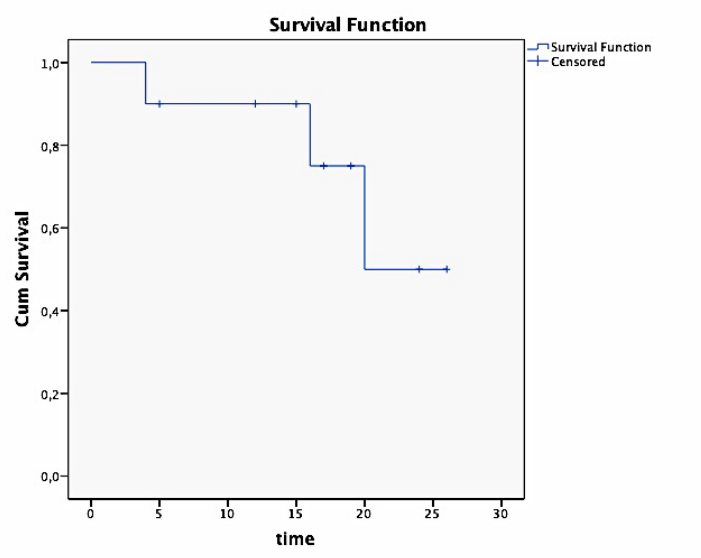

Figure 1 – Kaplan-Meier survival curve (illustrative example)

Outlining, for unadjusted survival analysis, commonly Kaplan–Meier analyses are applied, as it estimates the probability of survival up until a certain time point in the

presence of censored cases. For subjects whose data are censored, all information until their time of censoring is included in the analysis. Medical papers habitually present the complement of the Kaplan–Meier estimate [1−KM(t)], which gives the estimated probability of dying before time t. Cumulative incidence is then a sum of these conditional probabilities over time. Using the Kaplan-Meier method, denoted as CI^{rel}, is calculated as follows:

(viii)

$$CI_{KM}^{rel}(t_0) = 1 - S_{KM}^{rel}(t_0) = 0$$

$$CI_{KM}^{rel}(t_1) = 1 - S_{KM}^{rel}(t_1)$$

$$= CI_{KM}^{rel}(t_0) + S_{KM}^{rel}(t_0) * [\frac{no.of\ relapses\ at\ t_1}{no.of\ patients\ at\ risk\ jsut\ prior\ to\ t_1}]$$

$$CI_{KM}^{rel}(t_2) = 1 - S_{KM}^{rel}(t_2)$$

$$= CI_{KM}^{rel}(1_0) + S_{KM}^{rel}(t_1) * [\frac{no.of\ relapses\ at\ t_2}{no.of\ patients\ at\ risk\ jsut\ prior\ to\ t_2}]$$

$$CI_{KM}^{rel}(t_{10}) = 1 - S_{KM}^{rel}(t_{10})$$

$$= CI_{KM}^{rel}(t_9) + S_{KM}^{rel}(t_9) * [\frac{no.of\ relapses\ at\ t_{10}}{no.of\ patients\ at\ risk\ jsut\ prior\ to\ t_{10}}]$$

When is necessary to compare two or more groups and to determine whether the survival curves are statistically significantly different, log-rank test can be used.

To quantify an effect size the hazard ratio (HR) using Cox proportional hazards analysis is calculated. The HR is the ratio between the hazard of the event in the exposed group and the hazard of the event in the unexposed group. This hazard can vary over time, however the Cox model assumes that the HR between the two groups is constant over time. This is the proportional hazard assumption. In Cox regression analyses, it is also possible to adjust for (potential) confounders (7). The model has the form:

(ix)

$$\lambda_k(t) = \lambda_{0k}(t)\exp(Z\beta)$$

Designate the set of covariates by a vector Z (8).

2.3 The independence assumption

The Kaplan-Meier method is based on non-informative censoring (censored patients have the same probability of experiencing the event as patients who remain under follow-up). This method estimates the probability to survive up to until a certain point (time *t*) in the presence of censored cases, all information until their time of censoring is included in the analysis.

However in a competing risks setting, when more than one cause of failure is possible, other events can occur preventing the occurrence of the event of interest. Additionally each event has a hazard. Therefore, the number of failures from the competing risks will reduce the actual number of failures from the event of interest and consequently, influence the estimate of the probability of failure from this event (9).

In the presence of competing risks, treating the events of the competing causes as censored observations violates the Kaplan-Meier's independence assumption and can result in an over-estimation of the cumulative probability of cause-specific failure.

3. Competing Risks Approach

Competing risks occur when subjects can experience one or more events or outcomes which 'compete' with the outcome of interest. In those cases, the competing risk hinders the observation of the event of interest or modifies the chance that this event occurs. In these situations, comparing the cumulative incidence function gives an idea of the probability of occurrence of the event of interest.

3.1 Cumulative incidence function

Cumulative incidence function for a specific event, also known as the subdistribution function, is defined as the probability of failing from a given cause in the presence of competing events, given that a subject has survived or has already failed due to different causes. The cumulative incidence function for a specific event depends not only on the number of individuals who have experienced this type of event, but also on the number of individuals who have not experienced any other event (9).

Unlike in the application of the Kaplan–Meier method, competing events are not handled as regular censoring events without influence on the cumulative incidence function for the event of interest. Instead, the cumulative incidence, i.e. the probability of dying before time t, is lowered by the occurrence of the competing event and subjects experiencing the competing event are considered to be no longer at risk for the event of interest (7).

It can be calculated similarly as in the Kaplan-Meier estimate, the difference lies in the probability of an event-free survival just prior to a certain time. Using the previous example, with ten patients, the event of interest is relapse, and competing risk is death: in the Kaplan-Meier method the probability of an event-free survival is a relapse-free probability, whereas it is a relapse-free and death-free probability in the competing risk method (6).

(x)

$CI\ {}^{rel}_{CR}(t_0) = 0$ (because all subjects are alive and relapse-free at t_0)

$$CI\ {}^{rel}_{CR}(t_1) = CI\ {}^{rel}_{CR}(t_0) + S_{KM}^{rel,death}(t_0) * [\frac{no.of\ relapses\ at\ t_1}{no.of\ patients\ at\ risk\ just\ priot\ to\ t_1}]$$

$$CI\ {}^{rel}_{CR}(t_2) = CI\ {}^{rel}_{CR}(t_1) + S_{KM}^{rel,death}(t_1) * [\frac{no.of\ relapses\ at\ t_2}{no.of\ patients\ at\ risk\ just\ priot\ to\ t_2}]$$

$$CI\ {}^{rel}_{CR}(t_{10}) = CI\ {}^{rel}_{CR}(t_9) + S_{KM}^{rel,death}(t_9) * [\frac{no.of\ relapses\ at\ t_{10}}{no.of\ patients\ at\ risk\ just\ priot\ to\ t_{10}}]$$

Where $S_{KM}^{rel,death}(t)$ denotes the Kaplan-Meier estimate of survival probability for the joint events of relapse and death. In the competing risks method, patients dying are counted as events when calculating event-free survival, $S_{KM}^{rel,death}(t)$, and only patients who are truly alive are considered at risk for relapse.

It is now easy to understand that the relapse-free survival [S_{KM} (t)] in the Kaplan-Meier method is always greater than or equal to the relapse- and death-free survival [$S_{KM}^{rel,death}(t_9)$] in the competing risk method, meaning that the cumulative incidence using the Kaplan-Meier method is always greater than or equal to the cumulative incidence using the competing risk estimate. The extent of overestimation in the Kaplan-Meier method depends on the incidence rate levels of competing events. Furthermore, the cumulative incidence of death in the presence of relapse as a competing risk can be calculated similarly (6).

Importantly, the log-rank test that is used to test whether the survival functions are significantly different between groups when censoring is independent cannot be used in the competing risks context. Different tests based on cumulative incidence functions have been developed in the competing risks setting (7).

3.2 Fine and Gray model

Historically, the cumulative incidence function (also known as the subdistribution, the marginal probability function, the crude incidence, and the absolute cause-specific risk) has been first addressed by Gray in 1988. He proposed a class of tests for comparing the cumulative incidence curves of a particular type of failure

among different groups in the presence of competing risks. A decade later, Fine and Gray proposed to directly model the cumulative incidence function of a competing risks data by modeling the subdistribution hazard function (10,11).

The subdistribution hazard is defined as the hazard of failing from a given cause in the presence of competing events, given that a subject has survived or has already failed to different causes. We can write the subdistribution hazard for cause r as

(xi)

$$\lambda_r(t) = \lim_{\Delta t \to 0} \frac{Prob\,(t \leq T < t + \Delta t, R = r | T \geq t \cup (T \leq t \cap R \neq r))}{\Delta t}$$

$$= -\frac{d}{dt}\log\,(1 - I_r(t))$$

where Ir(t)=Prob (T≤t, R=r) is the cumulative incidence function for cause r (r=1,…,k).

Heuristically, the 'subdistribution hazard' λ_r can be thought of as the hazard for an individual who either fails from cause r or does not, and in the latter case has an infinite failure time for cause r. While this may seem unusual, indeed in the case of mutually exclusive event types, those who fail from one cause are invulnerable to failure from other causes (8). This is the fundamental difference between log-rank test and Gray's test. Supposing we have a failure cause 1 and a competing failure cause 2 for two treatment groups A and B and want to test if the groups differ with respect to occurrence of event 1. Then the first failure is a result of cause 1 in group A. The log-rank and Gray's test estimates at this failure time are identical. Then a failure resulting from cause 2 occurs in treatment group B, between the first and second failure times resulting from cause 1. For the log-rank test, that individual exits the risk set, whereas in Gray's test, with respect to event 1, such an individual remains in the risk set forever, indicated by the cause 1 failure time becoming T'=∞. Thus, the two test statistics will differ starting at the next cause 1 failure (12). Survival will be lower in competing risk analysis and higher (overestimated) in the Kaplan-Meier estimate.

Fine and Gray adopted a semiparametric proportional hazards model for the subdistribution hazard of cause r for a subject with covariate vector X as follows

(xii)

$$\lambda_r(t|X) = \lambda_{r0}(t)\exp\,(\beta_r^T X)$$

where $\lambda_{r0}(t)$ is the baseline subdistribution hazard of cause *r*, and β_r is the vector of

coefficients for the covariates. Estimation for this model follows the partial likelihood approach used in a standard Cox model in that covariates act to multiply the baseline subdistribution hazard in a time-independent manner (8,31). The Fine and Gray's model is based on the hazard of the subdistribution and provides a simple relationship between covariates and cumulative incidence. Modeling cumulative incidence functions for competing risks can be used to identify potential prognostic factors for a particular failure in the presence of competing risks, or to assess a prognostic factor of interest after adjusting for other potential risk factors in the model. However, this analysis cannot be fully interpreted without examining the effect of the covariates on the competing risks (6,9).

4. Case Study – Peritoneal Dialysis

In oncology and cardiovascular medicine, competing risks has been acknowledged for many years, whereas in nephrology, it has been recognized only recently. There are many situations in which competing risks play a role. For example, when studying the time until a peritonitis episode occurs in peritoneal dialysis (PD) patients, death, kidney transplantation and transfer to hemodialysis can be considered as competing risks because patients who experience one of these events are no longer at risk of developing PD- related peritonitis (7,9).

4.1 Peritoneal Dialysis First

Chronic kidney disease is a lifelong disease. Patients with renal failure, especially younger ones, will likely need hemodialysis (HD), peritoneal dialysis (PD) and renal transplantation at different times over their lifetime (13). These renal replacement therapies (RRT) should be considered complementary. Starting patients on PD as their initial treatment modality seems appropriate for many reasons. Preservation of residual renal function is probably the most important one, as it relates directly with a survival advantage (13). One might consider other factors, as satisfaction with care, convenience of home therapy and lower healthcare cost (14).

Another compelling argument in favor of the PD first approach is the observation that since the last decade, infection-related complications are higher and appear to be increasing in HD patients, whereas such complications are steadily declining in PD patients over the past decade (14). Studies comparing mortality outcomes in HD *versus* PD patients have shown conflicting results. But, in contrast to a common belief that survival on PD is inferior as compared to HD, the converse is actually the case for most patient groups, particularly in the first few years of dialysis (15). As for hospitalization rate, when comparing HD to PD as initial RRT modality, HD patients are at twice as higher risk of hospitalization from septicemia than PD patients (14).

Comparisons between PD and HD outcomes regarding patient and technique survival are common with contradictory results (16-21). In PD, comparisons among

diabetic *versus* nondiabetic and anuric patients *versus* patients with residual renal function are frequent. However comparisons between patients undergoing PD as first technique versus on PD after temporary HD are scarce (22).

4.3 A different survival study – our cohort

Summarizing all the preceding information, in the analysis of a cohort of PD patients, if the event of interest is death, other events may occur (renal transplantation, transfer to HD, recover of residual renal function) and if one of these events occurs, the event of interest cannot be observed. The competing risk hinders the observation of the event of interest or modifies the chance of occurrence of this event. Statistical analysis and interpretation of competing risks data differ from survival analysis with only a single cause of failure. The vast majority of the studies address survival analysis in PD cohorts with Kaplan-Meier approach.

An observational study was proposed and conducted to study patient survival on peritoneal dialysis, where death on dialysis is the event of interest and kidney transplantation, transfer to hemodialysis or recovering of renal function are competing risk for death on dialysis, as follow:

a) *Hypothesis/question*

Can the survival advantage of PD first patients be attributed to the bias resulting from the misuse of Kaplan-Meier approach to estimate patient survival?

b) *Objective*

General objective

The purpose of this study was to compare different survival analysis methods: Kaplan-Meier and Gray estimates in a cohort of PD incident patients.

Specific objectives

- Comparison of regression models of risk factors for death: cause-specific hazard and subdistribution hazard.

5. Statistical methods

This study was based on human research and generates quantitative data, without combining with previous results of other studies. It is not randomized. According to the EQUATOR network, this study is planned to explore the relationship between exposure to risk or protective factors and outcomes, and should be constructed under STROBE guidelines.

5.1 Study design, setting and participants

This observational longitudinal cohort study was held at Peritoneal Dialysis Unit, Centro Hospitalar Universitário do Algarve. This one-centre study portrays the nephrology assistance to a population of 450,000 habitants. All incident peritoneal dialysis adult patients (older than 18 years of age) at our Center, between 1st of January of 2002 and 31st of December of 2016, were included. Exclusion criteria were previous hemodialysis treatment for more than 6 months, transferal between dialysis modalities before or previous renal transplantation. Two incident cohorts of patients were established: patients starting renal replacement therapy with PD (PD first) and patients switching to PD on the first 6 months of dialysis (PD transfer).

In the present study, the event of interest was death and the competing risk events were transfer to hemodialysis, renal transplantation and lost to follow-up (which included recovering of renal function and center transferal). Patients without any of these outcomes were censored at the date of their last recorded visit or at the end of the follow-up period (31st of December of 2017).

Of a total of 157 patients, only 123 patients were included.

Registry data collection and analysis was submitted to ethical appreciation and approved by the Hospital Ethics Committee.

5.2 Variables, data collection and measurements

Each consecutive patient admitted in this unit is systematically enrolled for the purpose of quality of care control, with regular monthly input of data related with

peritonitis events, catheter complications, hospitalizations, death, renal transplantation or transfer of modality with respective causes.

Data of all patients were collected at the time of peritoneal dialysis initiation. Patients follow 1-2 monthly regular visits. Baseline data included:

- Demographics (gender, date of birth),
- Diabetic status (previously diagnosed and medicated),
- Cardiovascular disease (defined as the presence of one or more manifestations of coronary artery disease, congestive heart failure, stroke, transient ischemic attack, and intermittent claudication),
- Charlson comorbidity index (calculated with the online calculator https://www.thecalculator.co/health/Charlson-Comorbidity-Index-(CCI)-Calculator-765.html),
- Initial modality of dialysis (hemodialysis or peritoneal dialysis, date of initiation, date of technique transfer),
- Biochemical data (hemoglobin (g/dL), albumin (g/dL), parathyroid hormone (pg/mL) on the first peritoneal equilibration test),
- First peritoneal equilibration test (PET) results (standard PET performed one month after PD initiation, if on stable PD prescription and without peritonitis or systemic infection, if not performed as soon as possible): residual renal function (mL/min, computed as the mean of urea and creatinine clearances), dialysis dosage (Kt/V), type of peritoneal transport (D/P creatinine).

Clinical data recorded during follow-up:

- Clinical follow-up (number of peritonitis, number and duration of hospitalizations),
- Patient outcome was defined as the earliest event among: death, transfer to hemodialysis, renal transplantation and lost to follow-up (which included recovering of renal function and center transferal);
- Causes of death were also obtained at the end of the study period for each cohort.

5.3 Statistical analysis

Descriptive statistics was done and results are presented as mean ± standard

deviation for continuous variables and as frequency and percentages (*n*, %) for categorical variables. For comparison between groups, the Mann-Whitney U test and chi-square test were used for continuous and categorical variables, respectively. A 5% significance level was considered.

SURVIVAL ANALYSIS

a) Kaplan–Meier survival and log-rank analysis

In this survival analysis, only death was considered a final event. Patients were censored at transplant, transfer to HD, lost to follow-up or end of study period. Survival on dialysis was examined using the Kaplan–Meier method and survival by cohort was compared using the log-rank test. Cox proportional hazard modelling was used to examine hazard ratios for death as outcome. Univariable Cox proportional hazards regression was applied to estimate the individual hazard for patient death, which are expressed as hazard ratios (HRs) with 95% confidence intervals (CIs). The significant variables ($p < 20\%$) in the univariable analysis were put into a final multivariable Cox model.

b) Cumulative incidence function and Gray's model

In the present study, the cumulative incidence for a specific event was calculated based simultaneously on the estimate of the overall survival function when all types of events are considered and on the hazard estimate of the specific event.

To analyze differences in the cumulative incidence between patient groups, Gray's test was used.

In the multivariable analysis, the hazard of the subdistribution was estimated. In the present study, to analyze PD first opportunity, the event of interest was death and the competing risk events were transfer to hemodialysis, renal transplantation and lost to follow-up. Permanence on peritoneal dialysis was censored.

The statistical analysis was performed using SPSS software for Windows (version 22.0; Chicago, USA) and R project (packages coxph and cmprsk).

6. Results

6.1 Sample

Of the 157 patients observed during this 15-year period, 34 patients were excluded due previous HD treatment for more than 6 months, transferal between dialysis modalities before or previous renal transplantation, according to exclusion criteria. A total of 123 patients were included, 31 (25.2%) were transferred from HD on the first 6 months of therapy (PD transfer group) and 92 (74.8%) started and continued with PD (PD first group).

Table 1. Summary demographics at peritoneal dialysis initiation

	Entire cohort	PD first	PD transfer	p
N (%)	123	92 (74.8%)	31 (25.2%)	
Time on PD (months)	23.0[11.0-42.0]	25.5[13.2-45.0]	18.0[5.0-38.0]	0.197*
Time on HD (months)	na	na	2.0[1.0-3.0]	na
Age (years)	54.1 ± 16.9	50.5 ± 16.6	55.4 ± 17.0	0.169*
Female gender, n (%)	52 (42.3%)	39 (31.7%)	13 (10.6%)	0.965**
Diabetes, n (%)	45 (36.6%)	36 (29.3%)	9 (7.3%)	0.313**
Cardiovascular disease, n (%)	62 (50.4%)	45 (36.5%)	17 (13.8%)	0.568**
Charlson Index	5.1 ± 2.4	5.1 ± 2.5	4.9 ± 2.3	0.739*
Hemoglobin (g/dL)	12.0 ± 1.6	12.0 ± 1.6	11.8 ± 1.9	0.704*
Albumin (g/dL)	3.7 ± 0.6	3.8 ± 0.6	3.7 ± 0.6	0.559*
Parathormone (pg/mL)	546.3±467.9	513.3±444.4	646.9±529.6	0.208*
RRF (ml/min)	7.0 ± 4.7	7.8 ± 4.8	4.7 ± 3.4	**0.005***
D/P creatinine	0.64 ± 0.12	0.62 ± 0.11	0.68 ± 0.12	**0.046***
Kt/V	2.82 ± 0.99	2.95 ± 1.01	2.36 ± 0.76	**0.010***
Number of hospitalizations	2.0[1.0-5.0]	2.0[1.0-5.0]	2.0[1.0-5.0]	0.642*
Days of hospitalization	19.0[6.0-40.0]	18.5[3.0-39.5]	19.0[11.0-46.0]	0.273*

na – not applicable; RRF – residual renal function; D/P creatinine – creatinine dialysate-plasma ratio; Kt/V – dialysis adequacy; *Mann-Whitney U test and **chi-square test were used for continuous and categorical variables, respectively.

Table 1 presents the demographic and clinical data at peritoneal dialysis initiation, and comparisons between those patients PD first *vs* PD transfer. There were no significant differences in demographics or in comorbidities between groups at the

start of PD therapy, including diabetes and cardiovascular disease. The median time on HD before transfer to PD was 2.0 months. The patients with early PD failure (less than 6 month PD therapy) were not excluded, consequently part of the population was not long enough on PD to be able to have the first peritoneal equilibration test (PET), which could, in part, explain the missing information (23 performed PET out of 31 patients in PD transfer group and 78 performed PET out of 92 patients in PD first group).

The clinical parameters evaluated on the first peritoneal equilibration test revealed higher residual renal function and kt/V in the PD first group. The D/P of creatinine was lower on this group. Simultaneous blood analysis revealed no differences between groups in the shown parameters. Annual peritonitis rate was 0.22 peritonitis/year for the entire cohort, with the PD first patients presenting with 0.20 peritonitis/year and the transfer group had 0.31 peritonitis/year. There were no differences in hospitalizations or total time of stay at the hospital.

6.2 Kaplan–Meier survival, log-rank analysis and Cox proportional hazard

The timeline considered for patient survival was the total time on dialysis, consequently for the PD transfer group the time on HD was summed up to the time on PD, considering the same need for dialysis as the beginning of follow-up. The overall mean expected survival was 71.6 months (95% CI: 63.6 to 79.6 months), with mean survival 57.8 months and 77.3 months for PD transfer and PD first, respectively. Of the 123 patients, 22 (17.9%) died. Figure 1 shows the associated Kaplan–Meier survival curve.

The cumulative survival rate during 8 years of follow-up was significantly different between PD transfer and PD first group (log-rank test, $p = 0.013$). The overall survival rates at 1, 3, and 5 years were 88.2%, 68.6%, and 38.1% in the PD transfer group, and 97.8%, 87.5%, and 73.2% in the PD first group. The main causes of death in the 22 patients who died were cardiovascular disease (40.9%) and infection (36.4%).

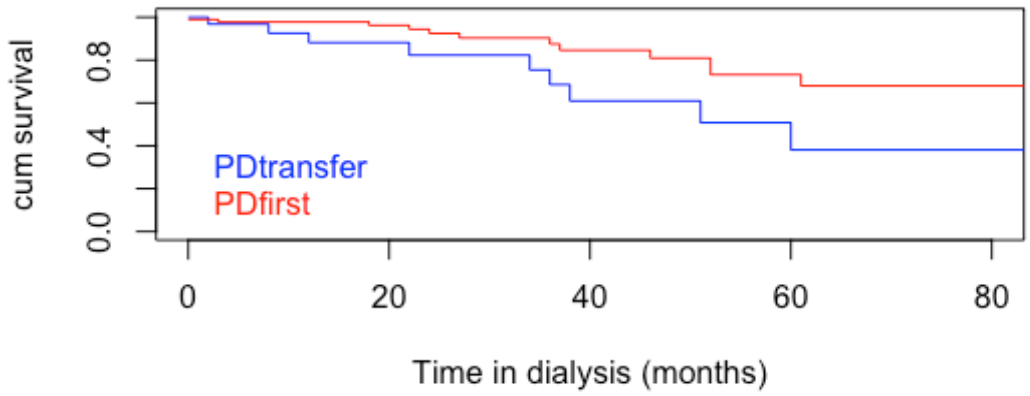

Figure 2 – Kaplan-Meier patient survival curve, considering PD transfer vs PD first (log-rank=0.012)

Univariable Cox proportional hazards analysis for patient survival was performed on each potential risk factor (table 2). Older age, lower RRF, PD transfer, diabetes, cardiovascular disease, lower albumin levels and higher Charlson Comorbidity Index were identified as significant risk factors ($p<0.05$). Table 3 shows the results of the multivariable Cox proportional hazards regression analysis, which evaluated factors that predicted death.

Table 2 - Univariable Cox proportional hazards analysis for patient survival

Characteristic	**HR**	**95% CI**	**p**
Age	1.043	1.011- 1.076	**0.008**
Gender	1.031	0.443-2.389	0.943
PD first	0.358	0.154 – 0.834	**0.017**
Total number peritonitis	0.941	0.683 – 1.295	0.708
Number of hospitalizations	1.013	0.886 – 1.157	0.853
Hospitalization days	1.004	0.997 – 1.012	0.251
RRF	0.833	0.723-0.961	**0.012**
Kt/V	0.803	0.471 – 1.367	0.419
D/P creatinine	1.029	0.019 – 55.551	0.989
Hemoglobin	1.028	0.802-1.318	0.825
Albumin	0.314	0.138 – 0.714	**0.006**
PTH	1.000	0.999-1.001	0.813
Cardiovascular disease	3.010	1.109 – 8.168	**0.030**
Diabetes	2.647	1.118 – 6.270	**0.027**
Charlson Index	1,317	1.132-1.532	**<0.001**

Charlson Index includes in its calculation age, presence of diabetes and cardiovascular events. To avoid redundancy and as the low number of events preclude adjusting for all the significant risks factors found above, age, presence of diabetes or cardiovascular disease were evaluated ultimately using only this comorbidity score.

Therefore after adjusting for Charlson Index, PD first RRF and albumin the model points out, as protecting variables, higher albumin (HR=0.170; CI 95% 0.052-0.550), higher residual renal function (HR=0.777; CI 95% 0.661-0.913) and PD first (HR=0.343; CI 95% 0.126-0.910). Higher Charlson Index predicted worse outcome (HR=1.462; CI 95% 1.159-1.843). PD as first dialysis therapy is associated with 65.7 % lower risk of death comparing with PD transfer. These results suggest that higher albumin, higher residual renal function, lower Charlson Index and PD first are independent predictors of survival in long-term PD patients.

Table 3 - Multivariable Cox proportional hazards regression analysis for patient survival

Characteristic	**HR**	**95% CI**	**p**
Charlson Index	1.462	1.159-1.843	**0.001**
PD first	0.343	0.126 – 0.910	**0.032**
RRF	0.777	0.661 – 0.913	**0.002**
Albumin	0.170	0.052 – 0.550	**0.003**

6.3 Cumulative incidence function, Gray's test, Fine and Gray model

Figure 3 summarizes cumulative incidence estimates for all possible outcomes taking competing risks into account.

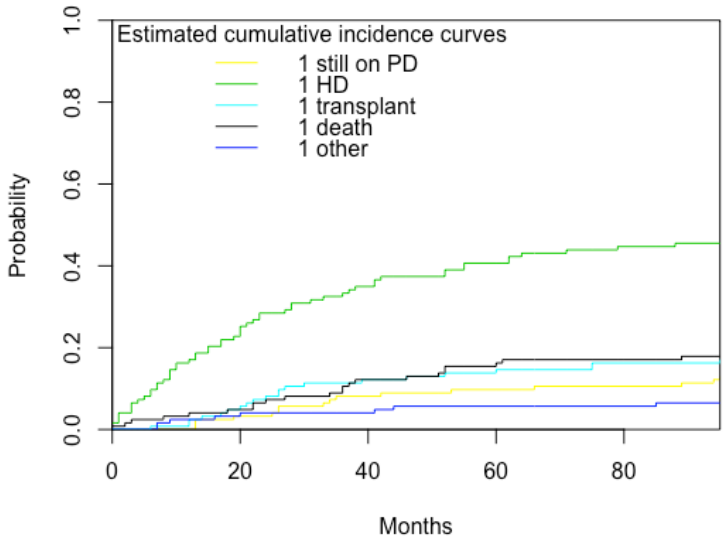

Figure 3 – Cumulative incidence curves, taking competing risks into account

It is possible to decide the time cohorts of interest, and the probabilities of experiencing death by 1, 3 and 5 years were 0.04, 0.11 and 0.16, respectively. Comparing theses values with 1- Kaplan-Meier estimates: 0.04, 0.19 and 0.36 by 1, 3 and 5 years, respectively, it is possible to acknowledge lower estimate of cumulative incidence by competing risks method. The magnitude of this difference increases with the follow-up.

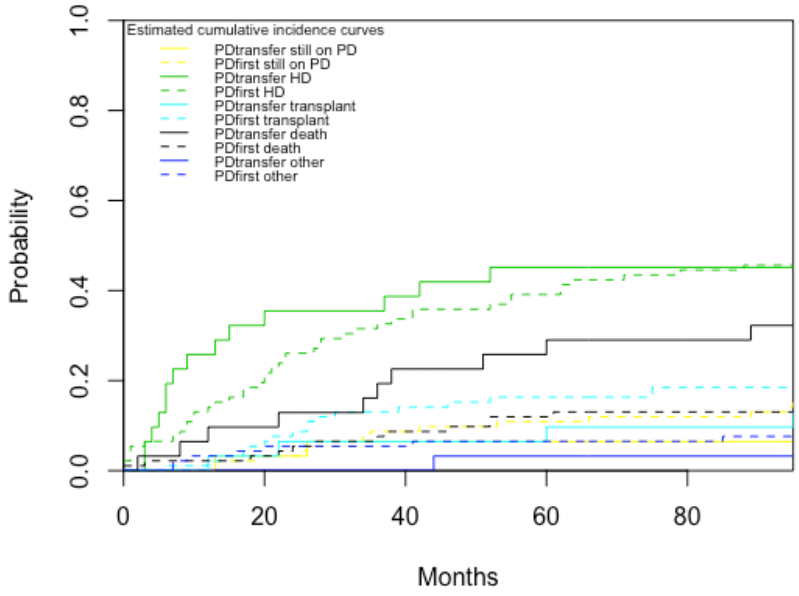

Figure 4 – Cumulative incidence curves considering PD transfer vs PD first (competing risks method)

Analyzing each outcome of the pair PD transfer vs PD first group, graphically depicted in figure 4, it is possible to perceive a higher distance between the black lines, which state the outcome death. The competing risk model analysis retrieves a p value for each pairs of lines and indeed it is the only statistically different, with p value of 0.016. Again, for comparative analysis with previous interpretation of Kaplan-Meier curve, choosing the event of interest death and the same time points:

In table 4 it is possible again to perceive the numeric difference between cumulative incidence and the complement of Kaplan-Meier, as well as the propensity to accentuate over time.

Table 4 – Comparison of Kaplan-Meier and Competing Risk estimates

	PD first			PD transfer		
	1 year	3 years	5 years	1 year	3 years	5 years
Kaplan- Meier	0.978	0.875	0.732	0.882	0.686	0.381
1-KM	**0.022**	**0.125**	**0.268**	**0.118**	**0.314**	**0.619**
Cumulative Incidence*	**0.022**	**0.076**	**0.120**	**0.097**	**0.194**	**0.290**

*Competing Risks analysis

Considering the unadjusted models (univariable) for the event of interest (death), it was found that the variables albumin, PD first, renal residual function and Charlson Index were significant in the cause specific-hazard model (Table 3) and identical results were attained in the subdistribution hazard model (Table 5). The subdistribution multivariable model found higher Charlson Index, PD transfer and lower renal residual function were statistically associated with death, but not albumin after adjustement. This was a different finding from the cause-specific hazard model. With the Fine and Gray competing risks analysis it is possible to infer PD as first therapy is associated with a 63.4% lower risk of death comparing with PD transfer (Multivariable Cox proportional hazards had stated a 65.7% survival advantage).

Table 5 – Fine and Gray model (hazard of the subdistribution model) for all possible events

		Unadjusted model			Adjusted model		
		sHR	95% CI	p	sHR	95% CI	p
Transfer to hemodialysis	Albumin	0.845	0.460-1.550	0.590	0.848	0.431-1.670	0.630
	Charlson Index	1.010	0.909-1.120	0.860	0.993	0.862-1.140	0.920
	Renal residual function	0.987	0.930-1.050	0.680	0.983	0.918-1.050	0.610
	PD first	0.925	0.492-1.740	0.810	1.511	0.662-0.648	0.340
Renal transplantation	Albumin	1.980	0.905-4.350	0.087	1.568	0.804-3.055	0.190
	Charlson Index	0.619	0.469-0.817	**<0.001**	0.628	0.460-0.856	**0.003**
	Renal residual function	1.020	0.933-1.110	0.690	1.058	0.947-1.183	0.320
	PD first	1.520	0.525-4.390	0.440	1.088	0.387-3.055	0.870
Death	Albumin	0.434	0.223-0.842	**0.014**	0.450	0.143-1.418	0.170
	Charlson Index	1.320	1.150-1.500	**<0.001**	1.411	1.135-1.755	**0.002**
	Renal residual function	0.816	0.698-0.954	**0.011**	0.797	0.680-0.934	**0.005**
	PD first	0.366	0.161-0.831	**0.016**	0.375	0.143-0.985	**0.047**
Lost to follow-up*	Albumin	1.920	0.448-8.200	0.380	1.638	0.611-4.390	0.330
	Charlson Index	0.969	0.766-1.230	0.800	0.865	0.678-1.100	0.240
	Renal residual function	1.140	1.020-1.290	**0.026**	1.166	1.016-1.340	**0.029**
	PD first	2.460	0.315-19.200	0.390	1.193	0.164-8.670	0.860

*included patients who recovered renal function or were transferred to other center

Analyzing the competing events, patients submitted to renal transplantation had a lower Charlson Index, as patients lost to follow-up, which included recovering of renal function had a higher residual renal function. None of these variables were risk factors for transfer to hemodialysis.

7. Discussion

The analysis of survival data plays a key role in medical research. Koller et al found that competing risks were present in a large majority of studies published in a sample of high-impact journals, but correct analysis was not performed (23). Failure to account correctly for competing events can result in adverse consequences, including overestimation of the probability of the occurrence of the event and misestimating the magnitude of relative effects of covariates on the incidence of the outcome.

When competing risks are present, there are three ways to analyze the data: (a) analysis of the event of interest ignoring competing risks, (b) analysis of joint events as a single end point, and (c) analysis of competing risks. The first approach, based on Kaplan-Meier method, is incorrect and will lead to inaccurate results as shown. The magnitude of the error can be considerable if the incidence of competing risks is high, or minimal if the incidence of competing risks is low. Nevertheless, one could not know the effect of competing risks a priori unless competing risk analysis is performed. The second approach is correct, but is too limited to address various important research questions. The combination of the second and third approaches is a comprehensive approach that addresses general as well as specific study questions (6).

In this study, patients starting renal replacing therapy with PD had a better survival than the patients who were transferred from HD, both statistical methods used drawn this conclusion. To decide which method for survival analysis should be used, it is important to know what kind of research question one aims to answer. Etiological research aims to investigate the causal relationship between risk factors or determinants and a given outcome. In this situation, cause-specific hazard model is generally more appropriate, since it quantifies the event rate among the ones at risk of developing the event of interest. In contrast, the focus is often on the direct assessment of actual risks and therefore for purposes of prediction or prognostic research and medical decision, which means subdistribution regression model for the cumulative incidence should be adopted (7,9,24).

In the presence of competing risks outcomes, the Kaplan-Meier core assumption of non-informative censoring does not hold, as the presence of competing risk events results in informative censoring (2). Many authors have shown that the complement of

Kaplan-Meier and cumulative incidence function can result in different estimates when competing risks are present (7,9,25). The bias resultant of this approach is especially great when the hazard of the competing risks is large (25). In general, the greater the percentage of competing events is, the greater the potential for bias in treating competing events as censoring events. Careful attention to the scientific objectives and methodology of the analysis is crucial when absolute percentages of competing events surpass10% (24).

It is shown in this example, the actual probability of death is wrongly overestimated using Kaplan-Meier method and the longer the duration of follow-up the larger the difference between the estimated by these two methods. Using the more appropriate methodology we were able to find that PD first patients had a 0.022, 0.076 and 0.120 probability of death by 1, 3 and 5 years, respectively; lower than PD transfer group with 0.097, 0.194 and 0.290 concerning the same time points. Cumulative incidence function is useful for evaluating the effect of a covariate on the probability of the occurrence of an event of interest over a period of time. It is the best measure for absolute risk calculations and risk prediction. In the analysis of competing risks data, cumulative incidence function should always be reported for all events (interest and competing events) (6). There was no other statistical significant association with the remainder risks.

Fine and Gray developed a technique to evaluate multivariable models in the presence of competing risks and just as in the standard survival analysis, analysis of competing risks is incomplete without regression analysis. The subdistribution multivariable model used found that PD first group had a survival advantage, even after adjusting for albumin levels, Charlson morbidity Index and residual renal function. It is important to present results for all causes and for both cause-specific hazard functions and subdistribution hazard functions. Such a practice enables a more complete understanding not only of the effects of prognostic factors, but also of the absolute risks of the different outcomes in the study sample. In our study the analysis of the other outcomes showed that patients with lower Charlson Index had a higher probability of receiving a renal transplant, and patients with higher renal residual function were more prone to “lost to follow-up” competing risk (which includes recovering renal function).

The presence of competing risks influences the number of events observed, as

the occurrence of a competing risk (renal transplantation, transfer to hemodialysis or other) preclude the occurrence of the outcome of interest. If patients with a certain characteristic were more likely to have a competing risk, the event of interest could not be observed and therefore the effect of this covariate would be diminished (9). This may explain why the variable albumin was found to be statistically significant in the cause-specific hazard model but not in the subdistribution model.

Competing risks are prevalent in Nephrology (7), and in the area of peritoneal dialysis, in recent years, several other studies have included competing risks in their survival analysis (9, 26-29). Besides portraying an increasing awareness of the presence of competing risks in this field of medical research, there is a shift in the profile of the authors of these publications. Early studies were conducted or assisted by statisticians, whilst medical doctors write more recent publications. The problematic issue with statistical software has somehow been surpassed. For SPSS, a macro is available to perform the cumulative incidence competing risks method. Software SAS and STATA also have resources for this analysis, pshreg procedure and stcrreg comand, respectively (7). The freely available software R has the package cmprsk available for download, and useful scripts for R competing risks analysis were made available by Professor Scrucca (30,31) and are simple to follow. The program code performed in this paper was adapted from the script used by Professor Laetitia Teixeira (9), is written in Appendix A. Nevertheless recognition of a competing risk paradigm is the essential step to adopt correct survival analysis methodology.

8. Conclusions

The important first step for the analysis of competing risks data is the recognition that competing risks are present. The competing risks assessment should include a calculation of cumulative incidence of an event of interest and of the different types of competing risks, a proper test for cumulative incidence curves of an event, and competing risks regression analyses (6). The Kaplan–Meier method for unadjusted survival analysis can handle only one-outcome and yields unreliable results for the estimation of survival probability in the presence of competing risks (7). Researchers need to define clearly the research objective, as the cause-specific hazard function is more appropriate for etiologic research, and the subdistribution hazard function is more appropriate prognosis predicting or estimating incidence. In some settings, both models should be estimated for each of the competing risks to permit a fully understanding of the effect of covariates on the incidence and rate of occurrence of each outcome (24), as complimentary methods.

In conclusion, a competing risk method should be adopted in studies with multiple outcomes, as in Peritoneal Dialysis studies, to estimate cumulative incidence and yield more accurate results.

APPENDIX A

PROGRAM CODE

```
library(splines)
library(survival)
library(cmprsk)
library(lattice)
library(MASS)

# Select the directory
setwd("~/Desktop/Ficheiros em R")
PD<-read.table(file="PDfirst2.csv",header=T, sep= ";", dec=",")
summary (PD)

##definition of variables as categoric
attach(PD)
obito=factor (obito, levels=c(0,1), labels =c("alive", "dead"))
PD_first=factor (PD_first, levels=c(0,1), labels=c ("PDtransfer","PDfirst"))
outcome=factor (outcome, levels=c(1,2,3,4,5), labels=c("still on PD", "HD","transplant","death", "other"))

#obtain cumulative incidence estimates
g=cuminc(tempototaldialise,outcome)

#prints each estimated cumulative incidence curve and estimated variances as a vector of times.
print(g)

#change point of cumulative incidence estimates
timepoints(g, times=c(6, 12, 24, 36, 48, 60))

#plot CIF taking competing risks into account
plot.cuminc(g, ylab="CIF", xlab="Months", xlim=c(0,80),
        col = c(4, 3, 9, 2, 14), curvlab=c("still on PD", "HD", "Transplant", "death", "other"))

#To obtain and compare cumulative incidence estimates by PD_first
h=cuminc(tempototaldialise,outcome, PD_first)
h

#change point of cumulative incidence estimates
timepoints(h, times=c(6, 12, 24, 36, 48, 60))

#plot CIF taking competing risks into account
plot.cuminc(h, ylab="CIF", xlab="Months", xlim=c(0,80),
        col = c(4,4,3,3,9, 9, 2,2,14,14), curvlab=c("still on PD", "HD", "Transplant", "death", "other"))


# Unadjusted Fine and Gray Models (4 events)

# Model 3a (failcode4=death)
FG3a1=crr(PD$tempototaldialise, PD$outcome, as.factor(PD$PD_first), cencode=1, failcode=4)
summary(FG3a1)
FG3a2=crr(PD$tempototaldialise, PD$outcome, PD$albumina, cencode=1, failcode=4)
summary(FG3a2)
FG3a3=crr(PD$tempototaldialise, PD$outcome, PD$indiceCharlson, cencode=1, failcode=4)
summary(FG3a3)
FG3a4=crr(PD$tempototaldialise, PD$outcome, PD$FRR, cencode=1, failcode=4)
summary(FG3a4)

# Model 3b (failcode2=Hemodialysis)
FG3b1=crr(PD$tempototaldialise, PD$outcome, as.factor(PD$PD_first), cencode=1, failcode=2)
summary(FG3b1)
```

```
FG3b2=crr(PD$tempototaldialise, PD$outcome, PD$albumina, cencode=1, failcode=2)
summary(FG3b2)
FG3b3=crr(PD$tempototaldialise, PD$outcome, PD$indiceCharlson, cencode=1, failcode=2)
summary(FG3b3)
FG3b4=crr(PD$tempototaldialise, PD$outcome, PD$FRR, cencode=1, failcode=2)
summary(FG3b4)

# Model 3c (failcode3=transplant)
FG3c1=crr(PD$tempototaldialise, PD$outcome, as.factor(PD$PD_first), cencode=1, failcode=3)
summary(FG3c1)
FG3c2=crr(PD$tempototaldialise, PD$outcome, PD$albumina, cencode=1, failcode=3)
summary(FG3c2)
FG3c3=crr(PD$tempototaldialise, PD$outcome,PD$indiceCharlson, cencode=1, failcode=3)
summary(FG3c3)
FG3c4=crr(PD$tempototaldialise, PD$outcome,PD$FRR, cencode=1, failcode=3)
summary(FG3c4)

# Model 3d (failcode5=other)
FG3c1=crr(PD$tempototaldialise, PD$outcome, as.factor(PD$PD_first), cencode=1, failcode=5)
summary(FG3c1)
FG3c2=crr(PD$tempototaldialise, PD$outcome, PD$albumina, cencode=1, failcode=5)
summary(FG3c2)
FG3c3=crr(PD$tempototaldialise, PD$outcome, PD$indiceCharlson, cencode=1, failcode=5)
summary(FG3c3)
FG3c4=crr(PD$tempototaldialise, PD$outcome, PD$FRR, cencode=1, failcode=5)
summary(FG3c4)

#Adjusted Fine and Gray models
# Model 5, including all covariates
# Define set of covariates
#PD2 is the same database, but without the lines with missing values
#(which would not included in the model, but were generating an error in the model)
PD2<-read.table(file="PDfirst3.csv",header=T, sep= ";", dec=",")
summary (PD2)
cov=cbind(PD2$albumina, PD2$indiceCharlson, PD2$FRR,as.factor(PD2$PD_first))

#Model 5a (failcode4=death)
FG5a=crr(PD2$tempototaldialise, PD2$outcome, cov, failcode=4, cencode=1)
summary(FG5a)

#Model 5b (failcode2=hemodialysis)
FG5b=crr(PD2$tempototaldialise, PD2$outcome, cov, failcode=2, cencode=1)
summary(FG5b)

#Model 5c(failcode3=transplant)
FG5c=crr(PD2$tempototaldialise, PD2$outcome, cov, failcode=3, cencode=1)
summary(FG5c)

#Model 5d(failcode5=other)
FG5c=crr(PD2$tempototaldialise, PD2$outcome, cov, failcode=5, cencode=1)
summary(FG5c)
```

APPENDIX B

ETHICS APPROVAL

Exma. Senhora

Dra. Anabela Malho Guedes

Faro, 27 de Junho de 2019

Assunto: Estudo sobre a Avaliação das Consequências da Estratégia Peritoneal Dialysis First

O estudo acima referido tem todo o interesse sob os pontos de vista clínico e científico.

Os direitos dos doentes estão devidamente salvaguardados.

Deste modo autorizamos a sua realização.

A Comissão de Ética

HOSPITAL DE FARO E.P.E.
Comissões Técnicas

References

1. Pepe MS. Inference for Events With Dependent Risks in Multiple Endpoint Studies. J Am Stat Assoc 1991; 86(415): 770-778.
2. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: Competing risks and multi-state models. Statist Med 2007; 26:2389-2430.
3. Kleinbau DG, Klein M. Survival Analysis. Statistics for Biology and Health. 2nd edition. Springer, 2005.
4. Etikan I, Abubakar S, Alkassim R.The Kaplan Meier Estimate in Survival Analysis. Biom Biostat Int J 2017; 5(2): 00128. DOI: 10.15406/bbij.2017.05.00128
5. Carvalho MS, Andreozzi VL, Codeço CT, Barbosa MTS, Shimakura SE. Análise de Sobrevivência: teoria e aplicações em saúde, 2.ª edição, 2011.
6. Kim HT. Cumulative incidence in competing risks data and competing risks regression analysis. Clin Cancer Res 2007; 13(2) January 15, 2007.
7. Noordzij M, Leffonfré K, van Stralen KJ, Zocalli C, Dekker FW, Jager KJ. When do we need competing risks methods for survival analysis in nephrology? Nephrol Dial Transplant 2013; 28: 2670–2677.
8. Dignam JJ, Zhang Q, Kocherginsky MN. The use and interpretation of competing risks regression models. Clin Cancer Res 2012; 18(8): 2301–2308.
9. Teixeira L, Rodrigues A, Carvalho MJ, Cabrita A, Mendonça D. Modelling competing risks in nephrology research: an example in peritoneal dialysis. BMC Nephrology 2013, 14:110.
10. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. Ann Stat 1988; 16:1141-1154.
11. Fine JP, Gray RJ: A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 1999; 94:496-509.
12. Dignam JJ, Kocherginsky MN. Choice and interpretation of statistical tests used when competing risks are present. J Clin Oncol 2008; 26:4027-4034.
13. Wang AY, Wang M, Woo J, et al. Inflammation, residual kidney function, and cardiac hypertrophy are interrelated and combine adversely to enhance mortality and cardiovascular death risk of peritoneal dialysis patients. *J Am Soc Nephrol* 2004;15:2186-2194.
14. Chaudhary K, Sangha H, Khanna R. Peritoneal Dialyis First: Rationale. *Clin J Am Soc Nephrol* 2011; 6: 447–456.
15. Dalal P, Sangha H, Chaudhary K. In Peritoneal Dialysis, Is There Sufficient Evidence to Make "PD First" Therapy? *Int J Nephrol* 2011; 2011: 239515.
16. Fenton SS, Schaubel DE, Desmeules M et al. Hemodialysis versus peritoneal dialysis: a comparison of adjusted mortality rates. *Am J Kidney Dis* 1997; 30: 334–342.
17. Jaar BG, Coresh J, Plantinga LC et al. Comparing the risk for death with peritoneal dialysis and hemodialysis in a national cohort of patients with chronic kidney disease. *Ann Intern Med* 2005; 143: 174–183.
18. Mehrotra R, Chiu YW, Kalantar-Zadeh K et al. Similar outcomes with hemodialysis and peritoneal dialysis in patients with end-stage renal disease. *Arch Intern Med* 2011; 171: 110–118.
19. Weinhandl ED, Foley RN, Gilbertson DT, Arneson TJ, Snyder JJ, Collins AJ. Propensity-matched mortality comparison of incident hemodialysis and peritoneal dialysis patients. *J Am Soc Nephrol* 2010; 21(3): 499–506.
20. Kumar VA, Sidell MA, Jones JP, Vonesh EF. Survival of propensity matched incident peritoneal and hemodialysis patients in a United States health care system. *Kidney Int* 2014; 86, 1016–1022
21. Yeates K, Zhu N, Vonesh E et al. Hemodialysis and peritoneal dialysis are associated with similar outcomes for end-stage renal disease treatment in Canada. *Nephrol Dial Transplant* 2012; 27: 3568–3575
22. Barone RJ, Cámpora MI, Gimenez NS, Ramirez L, Panese SA, Santopietro M. Peritoneal Dialysis as a First versus Second Option after Previous Haemodialysis: A Very Long-Term

Assessment. *Int J Nephrol* 2014: 693670.
23. Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? *Stat Med*. 2012; 31:1089–1097. doi: 10.1002/sim.4384.
24. Austin P, Lee D, Fine J. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation*. 2016;133:601–609.
25. Lim HJ, Zhang X, Dyck R, Osgood N. Methods of competing risks analysis of end-stage renal disease and mortality among people with diabetes. BMC Med Res Methodol 2010; 10:97.
26. Evans DW, Ryckelynck JP, Fabre E, Verger C. Peritonitis-free survival in peritoneal dialysis: an update taking competing risks into account. Nephrol Dial Transpl 2010, 25(7):2315–2322.
27. Beuscart JB, Pagniez D, Boulanger E, Lessore De Sainte Foy C, Salleron J, Frimat L, Duhamel A. Overestimation of the probability of death on peritoneal dialysis by the Kaplan-Meier method: advantages of a competing risks approach. BMC Nephrol 2012, 13:31.
28. Lambie M, Teece L, Johnson DW, Petrie M, Mactier R, Solis-Trapal I et al. Estimating risk of encapsulating peritoneal sclerosis accounting for the competing risk of death.Nephrol Dial Transplant 2019; Feb 28. pii: gfz034. doi: 10.1093/ndt/gfz034.
29. Chen HL, Tarng DC, Huang LH. Risk factors associated with outcomes of peritoneal dialysis in Taiwan: An analysis using a competing risk model. *Medicine (Baltimore)*. 2019;98(6):e14385.
30. Scrucca L, Santucci A, Aversa F. Competing risks analysis using R: an easy guide for clinicians. Bone Marrow Transplantation 2007; 40:381-387.
31. Scrucca L, Santucci A, Aversa F. Regression modeling of competing risks using R: an in depth guide for clinicians. Bone Marrow Transpl 2010; 45:1388-1395.