Comparison of survival on CAPD and haemodialysis: statistical pitfalls

Wim Van Biesen¹, Raymond Vanholder¹, Dirk Debacquer², Guy De Backer² and Norbert Lameire¹

¹Renal Division, Department of Internal Medicine and ²Department of Social Medicine and Epidemiology, University Hospital Gent, Gent, Belgium

Introduction

Although the life expectancy of patients with end-stage renal disease (ESRD) has improved over the last decade, it is still below that of the general population.

Besides renal transplantation, an ESRD patient and his treating physician can basically choose between two different modalities: haemodialysis (HD) and peritoneal dialysis (PD). Since most centres can offer both PD and HD, the choice of an appropriate renal replacement therapy (RRT), both for the physician and for patient, is influenced by a number of medical and non-medical factors [1]. It is quite obvious that for the patients, total survival and quality of life are among the most decisive factors determining their decision. As many patients will change dialysis modality during their lifetime, we believe that outcome of different consecutive RRT modalities is more important than that of one single modality. Besides patient survival, technique failure and its underlying causes and the outcome after transfer from one modality to another should also be considered [2]. Many studies have compared outcomes of PD and HD, with seemingly conflicting results. However, these studies differ in methodology, and their interpretation is often difficult [3].

A basic understanding of the statistical methods used for survival analysis, and their pitfalls, is crucial for correct interpretation.

This comment does not intend to provide a course in survival statistics, nor is its aim to review the numerous studies comparing survival in PD and HD. It will focus on the ‘mathematical philosophy’ of
survival analysis from a clinical point of view, and will
discuss how statistical techniques can influence the
outlook of the results. The reasons that the final
interpretation of such an analysis may be misleading
are listed in Table 1.

**Statistical methods of survival analysis**

Analysis of survival data requires the use of special
statistical methods, since not all patients considered
will have died at the end of the observation period. In
these patients, the total time of survival is not known.
Therefore, the data of these patients cannot be entered
in ‘classic’ statistical methods [4]. For example, if a
patient is started on PD 36 months before the end of
the observation period, one can say is that the
patient lived longer than 36 months, but no data on
the real survival are available. For these patients it is
said that the information is ‘censored’, as they did not
reach the final outcome point. Survival analysis
techniques such as Kaplan–Meier or Cox regression
were developed to analyse this type of ‘censored’
information, and these techniques account for the ‘loss
of information’ due to censoring. It is easy to under-
stand that the results become less reliable as the ratio
of ‘censored/uncensored cases’ increases, as the fate of
the uncensored cases heavily influences the final verdict.
Consequently, the ‘event numbers’ (i.e. the real number
of patients at risk at a certain moment on which the
analysis is based) should always be provided in survival
curves. Also the reason why a patient was censored is
of importance. If the reason for censoring is not related
to the treatment that is analysed, then this censoring
is called ‘not-informative’. This occurs for example
when a patient is censored because the observation of
the study ended before he/she died, or when the patient
is lost for follow up because he/she moved to another
city. However, if a patient is lost to follow-up because
of non-compliance, then a relationship between the
treatment and the reason for non-compliance can
possibly exist. In this case, censoring is called ‘informa-
tive’, and such censoring is in principle not allowed.
Such a patient has to be withdrawn from the analysis,
or at least a separate analysis of this type of patient
has to be performed.

**Kaplan–Meier and Cox proportional hazards**

Calculation of a Kaplan-Meier curve is relatively
simple, and is provided in most statistical packages.
The core of the method is the fact that the survival
function \( S(t) \), being the probability to be alive at a
moment \( t \) can be expressed as the product of the
probabilities of survival to time point \( t(i−x) \), under
condition one has survived to time point \( t(i−x−1) \)
(notation \( S(t(i−x))/t(i−x−1) \)). At any time point,
\( S(t−1) \) can be estimated from the ratio of the number
of patients \( n \) who survive a certain time interval to
the total number of patients \( N \) who entered that
interval. This also implies that deviation from reality
is greatest in the longest survival times, as fewer
patients are then present on whom the estimations are
based. Kaplan–Meyer and life-table analyses give the
data ‘as real’, with no corrections made for underlying
covariates which are possibly related to survival. As
most ESRD patients have various comorbid condi-
tions, corrections for these confounding factors are
necessary.

In a Cox-regression model, corrections can be made
for different comorbidity factors. The Cox’s model
assumes that independent variables are related to sur-
 vival time by a multiplicative e
fect on the hazard
function \( H0(t) \), the latter being the underlying ‘basic’
hazard equal to all participants. Unfortunately, \( H0(t) \)
is unknown, and only the relative risk, being the ratio
of the hazards of two different subjects, can be calcu-
lated. This number is then the relative risk for subject
1 compared to subject 2. It should be emphasized that
curves that result from a Cox regression analysis are
only reflecting ‘predictive calculations’ rather than the
real situation. Results can, for example, be presented
with all comorbidity factors entered as absent, and the
resulting curve will show seemingly better results com-
pared to a curve where comorbid conditions are taken
at the mean (Figure 1). This is an example of presenta-
tion bias, and when comparing results of different
centres, one should be aware of this.

**Standardized mortality ratios and Poisson analysis**

Another method to analyse survival data is the use of
mortality rates [5]. The ‘standardized mortality ratio’
(SMR) is the ratio of the total observed mortality in
the investigated (treatment) group to the total number
of expected deaths in a group of patients. An SMR of
>1 indicates higher than expected mortality, while an
SMR <1 indicates lower than expected survival. A
potential pitfall here is that the SMR itself is prone to
random variation. Therefore, also confidence intervals
(CI, mostly 95%CI) of the SMR have to be provided.
The probability that the real value of the SMR will be
in this confidence interval is 95%. Another problem is

| Table 1. Reasons that survival statistics may be misleading |
|-----------------|-----------------|
| 1. The statistical approach used is not valid: because of shortcomings of the methodology itself; the inclusion of patients is not correct: no or incorrect randomization not all patients included (selection of patients) prevalent vs incident patient inclusion the study does not analyse what it purports to analyse presence of influencing factors not accounted for by the methodology. |
| 2. The presentation of the data is not correct, e.g. only positive results are reported. |
| 3. The interpretation of the data is not correct: overinterpretation of results incorrect interpretation of results. |
that the SMRs have to be calculated using the reference for a specific group of patients, as the expected morta-
ity also depends on patient characteristics such as age,
race, sex, diabetes, cardiovascular disease, or other comorbid conditions. Therefore, SMR analysis is par-
ticularly useful for analyses of large patient groups,
comparing one treatment group with another, or one centre with another.

Poisson analysis is another method for the calculation and comparison of adjusted survival rates. Variables denoting the number of events in a certain unit of time are distributed according to the Poisson distribution. It is assumed that the events occur randomly, independently of one another and with an average rate that remains unchanged over the observation period. For survival analysis, one can consider the number of patients at risk, multiply this with the days the patients were ‘at risk’ on the treatment, and divide this by the number of observed deaths. This results in a ‘mortality rate’. This mortality ratio can also be seen, just as in Cox regression analysis, as the composite regression of the influence of different risk factors, thus allowing for adjustment for comorbid conditions. The mortality ratio can then be compared with another population, or with a reference population. It is of note that, unlike with Cox regression, the use of Poisson analysis assumes that the form of the underlying risk distribution is known and constant during the observed period. As the hazard function of mortality in a dialysis patient is not constant over time, this premise means that Poisson analysis for survival in ESRD is only robust for observations made over shorter time periods.

**Incident vs prevalent patient inclusion**

One should also pay attention to the manner in which patients are included in the study. Patients can be included as ‘incident’, i.e. starting new on ESRD treatment, or as ‘prevalent’, i.e. patients who have been some time on ESRD treatment. In the latter case, patients who start RRT and die during the inclusion window are excluded from the study. Thus this analysis favours the method with the highest initial mortality. The use of incident or prevalent patients can give markedly different results, as was shown by Vonesh and Moran [6].

**Clinical relevance**

Another crucial point is the definition of what is being analysed. This largely determines the clinical relevance of the analysis. It is important to consider the period of time that is defined as ‘survival time’. In the first modality survival analysis, survival time is considered as time on the initial RRT modality. Only death is considered as a final event, and patients are censored at transfer to another dialysis modality, at transplantation, loss of follow-up, or at the end of observation.

The intention-to-treat survival analysis considers the sum of the time on HD and the time on PD. In this analysis death is considered as the final event, and patients are censored at the moment of transplantation, at loss of follow-up or at the end of the observation, but they are not censored at transfer from PD to HD or vice versa.

The previous two analyses only consider the time on renal replacement, and exclude the time after trans-
plantation. These analyses are thus applicable for all patients, whether they are on the waiting list for transplantation or not. For the patients not on the waiting list, the intention-to-treat analysis gives their life expectancy. For patients on the waiting list, it shows their probability to survive until a renal graft becomes available. It is quite acceptable that patients on the waiting list for transplantation will in general
be in a better condition, compared to those not on the waiting list, which could be a possible confounding factor. This may explain, at least in part, the difference in survival on RRT between Europe and Japan, as in Europe, the ‘fittest’ patients drop out from the analysis at transplantation, leaving for further analysis only the patients with contraindications for transplantation. This may also be a confounding factor in the analysis between PD and HD.

In the total survival analysis, survival time is considered as the total time on PD, on HD, and after transplantation. Death is considered as the final event and patients are only censored at the end of observation.

Technique success is defined as the probability of having a patient alive on his initial modality. Death and change of modality are considered as final events, and patients are censored at the time of transplantation or at the end of follow-up. Another way of looking at ‘technique survival’ is to censor patients at their death, the underlying reasoning being that death is not the cause of technique failure, and that technique survival would have been longer if the patient had not died. In this analysis, only real ‘technique failures’ are considered as end-points. It is clear that the results of this type of analysis are far more ‘flattering’ compared to those where death is also considered as technique failure.

Most earlier survival analyses of ESRD patients used the first modality survival approach [7,8]. In our opinion, this ‘survival’ does not correspond with the clinical reality, as most ESRD patients are treated successively with different treatment modalities. Total survival time and intention-to-treat modality survival are of greater interest for the individual patient [3]. It is gratifying to see that the most recent papers on this topic used ‘intention-to-treat’ analysis [9]. However, to our knowledge, none of these papers has separately analysed the outcomes of patients transferred from one modality to another, thereby neglecting the potential importance of the sequence and timing of the different RRT modalities.

Another important confounding factor in survival analysis is, of course, the quality of the delivered care. Although this is difficult to measure, the experience with a technique can be estimated by the percentage of patients being treated on that modality. In this regard, it is striking that in the study by Bloembergen et al. [10], reporting ‘worse’ outcome for PD patients, only 13% of patients were treated with PD, while in the study by Fenton et al. [9], reporting better outcome for PD patients during the first 4 years, nearly half of the patients were treated with PD.

It should be appreciated that all statistical approaches have their shortcomings [11,12]: it is quite impossible to include all important comorbid conditions, and to exclude those that are less meaningful. It is very difficult to account for delivered dialysis dose, declining residual function, patient compliance or centre technical experience, although all these factors can have an important impact on outcome. Furthermore, an adequate quantification of the severity of comorbid conditions is also often difficult. Congestive heart failure or diabetes mellitus are very difficult to grade; for example, in most studies a patient with 3 years of diabetes is given the same risk as a patient with 20 years of diabetes, while the consequences of the disease will be greatly different. Mortality comparisons should thus be viewed with caution regarding the inclusion of comorbid conditions and the way their severity is scored. On the other hand, one should avoid correcting for factors that probably do not affect outcome in the observed time span, as this complicates the analysis and potentially increases random noise. Care should also be taken to correct for risk factors that are related to the treatment under consideration. Nutritional status for example might be a factor related to adequacy, so this covariate should not be corrected for in a prevalent patient analysis, as malnutrition may be the consequence of the treatment. In an incident approach, however, it might be included, as in this case it is a marker of nutritional status before dialysis treatment was started.

Another important point should be made concerning the exact meaning of relative risks. Relative risks can, from a clinical point of view, only be interpreted if also the real mortality risk is known. If the relative mortality risk of group A to group B is 2, this means that with a real mortality risk in group B of 1/10000, in group A it will be 2/10000. This can be a statistically significant difference, but from a clinical point of view it is often meaningless.

It is also of note that studies with a low number of patients are prone to false negative results (no statistically significant difference noted, whilst in reality there is one), and that in studies with large patient numbers, false positive results can emerge (statistically significant difference without clinical meaning).

In conclusion, the interpretation of papers analysing survival comparisons should be done with attention to the methodological biases, and their implications, realizing, however, that a ‘perfect’ comparison in a difficult field such as RRT is nearly impossible.

References


