The analysis of survival data: the Kaplan–Meier method

Kitty J. Jager[1], Paul C. van Dijk[1,2], Carmine Zoccali[3] and Friedo W. Dekker[1,4]

[1]ERA–EDTA Registry, Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; [2]Department of Clinical Epidemiology, Biostatistics and Bio-informatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; [3]CNR–IBIM Clinical Epidemiology and Pathophysiology of Renal Diseases and Hypertension, Renal and Transplantation Unit, Ospedali Riuniti, Reggio Cal., Italy and [4]Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands

What is this patient's prognosis regarding graft rejection? Do patients using a particular drug live longer than those not using it? How does this co-morbidity affect access to transplantation? To answer this type of questions one needs to perform survival analysis. This paper focuses on the Kaplan–Meier method, the most popular method used for survival analysis. It makes it possible to calculate the incidence rate of events like recovery of renal function, myocardial infarction or death by using information from all subjects at risk for these events. It explains how the method works, how survival probabilities are calculated, survival data can be summarized and survival in groups can be compared using the logrank test for hypothesis testing. In addition, it provides some guidance regarding the presentation of survival plots. Finally, it discusses the limitations of the Kaplan–Meier method and refers to other methods that better serve additional purposes.

As one of the aims of physicians is to prevent and cure their patients from disabling and life-threatening diseases, a considerable part of the studies in the area of medicine deal with survival. Each month in every medical journal the technique of survival analysis is used both in prognostic and in therapeutic studies to estimate and interpret survival, to compare it between groups, and to assess the relationship of explanatory variables with survival time. Therefore, typical examples of publications using survival analysis are a report providing details on the prognosis of elderly transplant recipients regarding their estimated 5-year probability of graft rejection,[1] a publication on the difference in mortality between diabetic patients using or not using angiotensin-converting enzyme inhibitors[2] and an article describing the impact of comorbidity on access to transplantation in dialysis patients.[3]

The primary aim of survival analysis is the modeling and analysis of 'time-to-event' data; that is, data that have as an end point the time when an event occurs. In this respect, events are not limited to death but may include all kinds of 'positive' or 'negative' events like myocardial infarction, recovery of renal function, first renal transplant, graft failure, or time to discharge from hospital. In addition to these 'single' end points, an increasing number of studies examine the incidence of a combined or composite end point, which can merge a variety of outcomes in one group. For example, the 4-D study used 'death from cardiac causes, nonfatal myocardial infarction, and stroke' as their primary outcome,[4] whereas the CREATE trial chose to study the incidence of a 'composite of eight cardiovascular events'.[5] Using combined end points, one addresses event-free survival, an important criterion in therapy evaluation. It is customary to talk about survival analysis and survival data, regardless of the nature of the event. Still, by far the most frequently used event in survival analysis is overall mortality.

A clinical example of when questions related to survival are raised is the following. Suppose you consider starting the preparation of dialysis in the near future in a 70-year-old patient with ESRD with a myocardial infarction in his medical history. While preparing the discussion of this topic

Correspondence: Kitty J. Jager, ERA-EDTA Registry, Academic Medical Center, Department of Medical Informatics, J1b 115.1, PO Box 22700, Amsterdam 1100 DE, The Netherlands. E-mail: k.j.jager@amc.uva.nl

with your patient, you ask yourself questions like 'what is the probability of dialysis patients in this age category in my country to be alive for another 2 or 5 years?', 'how is the prognosis in this type of patients being affected by ischemic heart disease?' and 'what is the experience with the use of either hemodialysis or peritoneal dialysis in similar patient groups in terms of patient survival?'. A search through the literature in nephrology journals and renal registry reports will help you to answer these questions. All of them will have made use of survival analysis. This article provides a first introduction to those analysis techniques.

EVENTS VERSUS CENSORED DATA

The distinguishing feature of survival data is that at the end of the follow-up period, the event will probably not have occurred for all patients. For these patients, survival time is said to be censored. We do not know when or whether such a patient will experience the event of interest, but we know only that he or she has not done so by the end of the observation period.[6] Censoring may also occur for other reasons. A patient may be lost to follow-up during the study or may experience a 'competing' event as a result of which further follow-up is impossible.[6] For example, patients being followed for myocardial infarction on dialysis may die of a malignancy. Other cases of censoring are included in the following example on the survival of RRT patients.

Example 1. Survival time on RRT: events and censored observations

Incident RRT patients in the ERA-EDTA Registry were included in an analysis of patient survival on renal replacement therapy (RRT).[1] Like in most survival studies, patients were recruited over a period of time (1996–2000, the inclusion period) and they were observed up to a specific date (31 December 2005, the end of the follow-up period). During this period, the event of interest was 'death while on RRT', whereas censoring took place at recovery of renal function, loss to follow-up, and at 31 December 2005.

Figure 1 shows the times that eight of the patients from this cohort were at risk of death on RRT. Over the period, there were five events and three censored observations. Here, one can see the analogy with the concept of incidence rate that was explained in a previous paper.[7] In that article, it was outlined that incidence rate (synonym: hazard) is the ratio of the number of subjects developing disease (or other health outcome) to the time at risk for disease and that it is an instantaneous concept, like speed. In example 1, the investigators studied the incidence rate (hazard) of death, 'the speed of dying', as the number of deaths was related to the time at risk of death on RRT.

ASSUMPTIONS RELATED TO CENSORING

There are some assumptions related to the use of censoring.[8] The two most important ones will be discussed here. First, it is assumed that at any time, patients who are censored have the same survival prospects as those who continue to be followed. This assumption cannot easily be tested. In the survival of dialysis patients, for example, it is customary to censor the survival time of a patient at the time of transplantation, because at that time, the patient is no longer at risk of death on dialysis. However, we all know that dialysis patients who are placed on the waiting list for transplantation are healthier than those who are not.[9] Therefore, using censoring for transplantation while studying the survival on dialysis is needed, because of the change in treatment, but it is in this case probably that this first assumption for the use of censoring is not fully fulfilled. Secondly, survival probabilities are assumed to be the same for subjects recruited early and late in the study. This assumption may be tested, for example, by splitting a cohort of patients in those who were recruited early and those recruited late and checking if their survival curves are different.

Using survival data, investigators often wish to estimate the probability of a patient surviving for a given period like 1 or 2 years. In addition, they are also interested to compare the survival of different groups. The next paragraph will
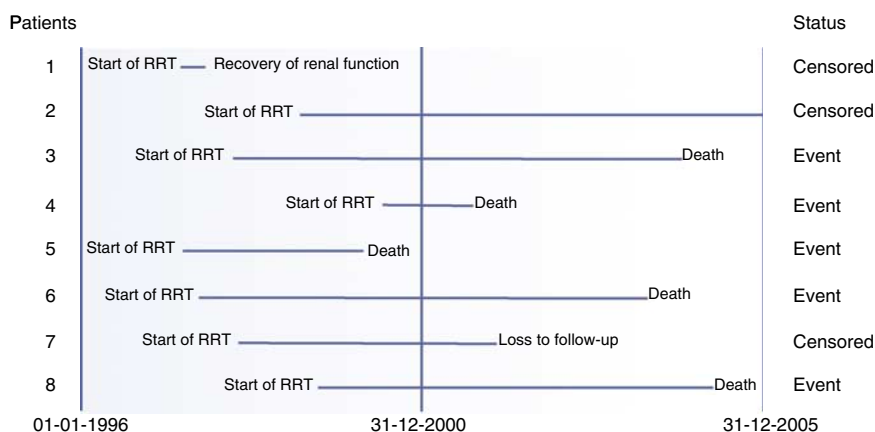

Figure 1 | Survival times of eight patients at risk of death on RRT. The inclusion period was 1996–2000, whereas follow-up was ended on 31 December 2005.

Table 1 | Cumulative survival table in 50 RRT patients

Time in days	Number at risk	Deaths	Withdrawn alive (censored)	Proportion surviving on this day	Cumulative survival[a]	Cumulative mortality (1−cumulative survival)
0	50	0	0	1.00	1.00	0
34	50	1	0	49/50=0.9800	0.9800	0.0200
35	49	1	0	48/49=0.9796	0.9600	0.0400
44	48	1	0	47/48=0.9792	0.9400	0.0600
57	47	0	1	1	0.9400	0.0600
75	46	1	0	45/46=0.9783	0.9196	0.0804
159	45	1	0	44/45=0.9778	0.8991	0.1009
192	44	0	1	1	0.8991	0.1009
199	43	1	0	42/43=0.9767	0.8782	0.1218
206	42	0	1	1	0.8782	0.1218
232	41	1	0	40/41=0.9756	0.8568	0.1432
316	40	1	0	39/40=0.9750	0.8354	0.1646
319	39	0	1	1	0.8354	0.1646
344	38	1	0	37/38=0.9737	0.8134	0.1866
363	37	1	0	36/37=0.9730	0.7914	0.2086
365	36	0	1	1	0.7914	0.2086
398	35	0	1	1	0.7914	0.2086
410	34	1	0	33/34=0.9706	0.7681	0.2319
491	33	1	0	32/33=0.9697	0.7449	0.2551
552	32	1	0	31/32=0.9688	0.7216	0.2784
575	31	0	1	1	0.7216	0.2784
628	30	0	1	1	0.7216	0.2784
728	29	1	0	28/29=0.9655	0.6967	0.3033
791	28	0	1	1	0.6967	0.3033
824	27	1	0	26/27=0.9630	0.6709	0.3291
896	26	0	1	1	0.6709	0.3291
954	25	1	0	24/25=0.9600	0.6441	0.3559
1115	24	0	1	1	0.6441	0.3559
1142	23	1	0	22/23=0.9565	0.6161	0.3839
1178	22	1	0	21/22=0.9545	0.5881	0.4119
1186	21	1	0	20/21=0.9524	0.5601	0.4399
1211	20	0	1	1	0.5601	0.4399
1323	19	0	1	1	0.5601	0.4399
1650	18	1	0	17/18=0.9444	0.5289	0.4711
1708	17	1	0	16/17=0.9412	0.4978	0.5022

[a]Cumulative survival is calculated: proportion surviving on this day × cumulative survival over the previous period.

elaborate on the method that is most frequently used to calculate survival probabilities, the Kaplan–Meier method.

DISPLAYING SURVIVAL DATA

A sample of 50 patients from a study on diabetes[10] is used to illustrate the application of the Kaplan–Meier method.

Example 2. Survival probability in RRT patients for ESRD due to diabetes mellitus and other causes

In a sample of 50 RRT patients taken from a study on diabetes mellitus,[10] survival time started running at the moment a patient was included in the study, in this case at the start of RRT. Patients were followed until death or censoring. The survival probability was calculated using the Kaplan–Meier method. Subsequently, the survival of patients with ESRD due to diabetes mellitus was compared with the survival of those with ESRD due to other causes.

Before analysis, the observed survival times were first sorted in ascending order, starting with the patient with the shortest survival time. This resulted in Table 1 showing the patient deaths and the censored observations over the first

1708 days. By then, considering time in many small intervals, it becomes possible to calculate the probability of surviving a given day. At the start of the study, all 50 patients were alive, so the proportion surviving and the cumulative survival (synonym: cumulative proportion surviving) both were 1.00. When the first patient died on day 34 after the start of RRT, the proportion surviving on that day was 49/50 = 0.9800 = 98%. To calculate the cumulative survival, this proportion surviving of 0.9800 was multiplied by the 1.0 cumulative survival from the previous step resulting in a cumulative survival dropping that day to 0.9800. Then, when the second patient died at day 35, the proportion surviving on that day was 48/49 = 0.9796. To obtain the cumulative survival at day 35, again, this proportion was multiplied by the 0.9800 cumulative survival from the previous step, which resulted in a cumulative survival dropping that day to 0.9600. Along the same lines, cumulative survival on day 44 dropped to 0.9400. On day 57, however, a patient was withdrawn alive from the study, because his follow-up time was censored at the end of the study period. The proportion surviving that day was 47/47 = 1.00, as this patient did not die but was
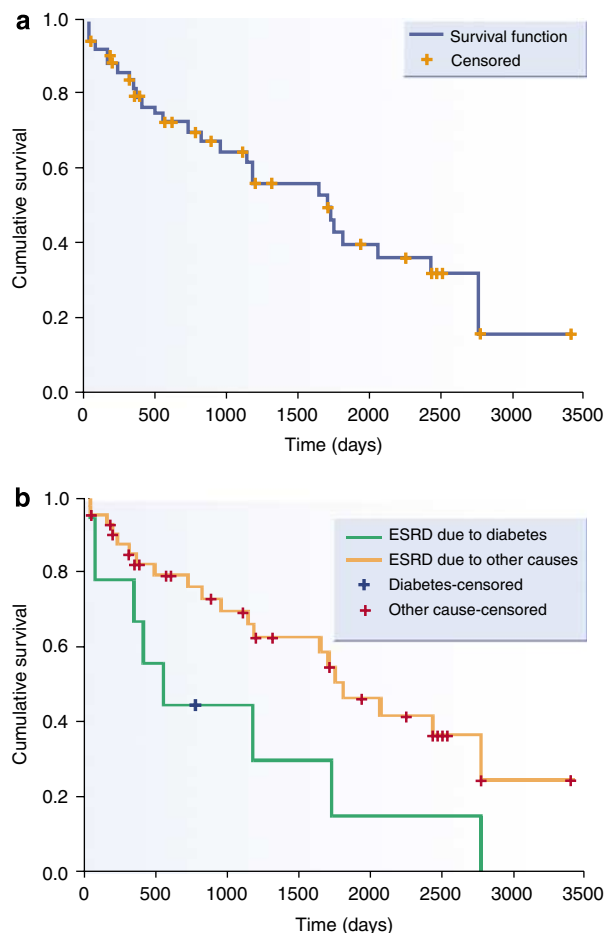
Figure 2 | **Cumulative survival in RRT patients.** (**a**) overall and (**b**) by cause of ESRD.

withdrawn alive from the study. As a result, the cumulative survival did not drop that day but remained unchanged at 0.9400. This example shows that cumulative survival is a probability of surviving the next period multiplied by the probability of having survived the previous period. Secondly, the example shows that all subjects at risk (also those not experiencing the event during the observation period) can contribute survival time to the denominator of the incidence rate. Finally, it demonstrates that by censoring, one is able to reduce the number of persons alive without affecting the cumulative survival. The same data were used to construct the survival curve of these 50 patients, which is shown in the upper panel of Figure 2. The figure displays visually that cumulative survival drops with every death, whereas it remains unchanged with every censored observation (indicated by the plus signs).

REPORTING SURVIVAL BASICS
Survival can be summarized by quoting survival probabilities. The 1-, 2-, and 5-year cumulative survival of the patients in example 2 were 79, 70, and 40%. An alternative way to summarize survival is to quote the median survival. The median survival is that point in time from the time of

inclusion when the cumulative survival drops below 50%. It is not related to the number of deaths or to the number of subjects that is still at risk of death. In example 2, including 50 patients, the median survival time was 1708 days, whereas the total number of deaths up on that day was 22 and the number of patients still at risk of death was 16. In clinical trials, median survival time is a way to measure the effectiveness of a treatment.

In general, investigators use median survival rather than mean survival. There are a number of reasons for this. The first one is that samples of survival times are mostly highly skewed and in those cases the median is generally a better measure of central location than the mean. A second reason is that in survival analysis, one makes use of censoring. As we explained earlier, one does not know when or whether such a person will experience the event of interest, but one knows only that he or she has not done so by the end of the observation period. This complicates the calculation of the mean. Finally, even in cases where there is no censoring, to calculate a mean, one would need to wait until all persons reached the event of interest, and this may require quite a long period. For these reasons, it is simpler to use median survival, as this is completely defined once the survival curve descends to 50%.

COMPARING SURVIVAL BETWEEN GROUPS
When comparing survival between different groups, one could compare the cumulative survival at some specific time. The weakness of this approach is, however, that it does not provide a comparison of the total survival experience of the two groups, but rather gives a comparison at one arbitrary time point. To prevent this problem, one may use the logrank test, the most popular method of comparing the survival of groups, which takes the whole follow-up period into account.[11] The log-rank test addresses the hypothesis that there are no differences between the populations being studied in the probability of an event at any time point. The test is based on the same assumptions as the Kaplan–Meier method.

The lower panel of Figure 2 that was derived from example 2 displays the survival curves of the patients with ESRD due to diabetes mellitus and those with ESRD due to other causes. These curves provide a means of assessing visually whether survival was different for these subgroups. When the log-rank test was used to test formally whether the difference was statistically significant, it showed a *P*-value of 0.0425; therefore, also in this small sample of 50 patients, the difference was statistically significant.

SURVIVAL PLOTS
Usually, survival plots are presented as cumulative survival displaying the proportion of patients free of the event declining over time. Sometimes, however, they are presented as cumulative mortality showing the proportion of subjects experiencing the event by time. The data from a third study[12] are used to show this difference.
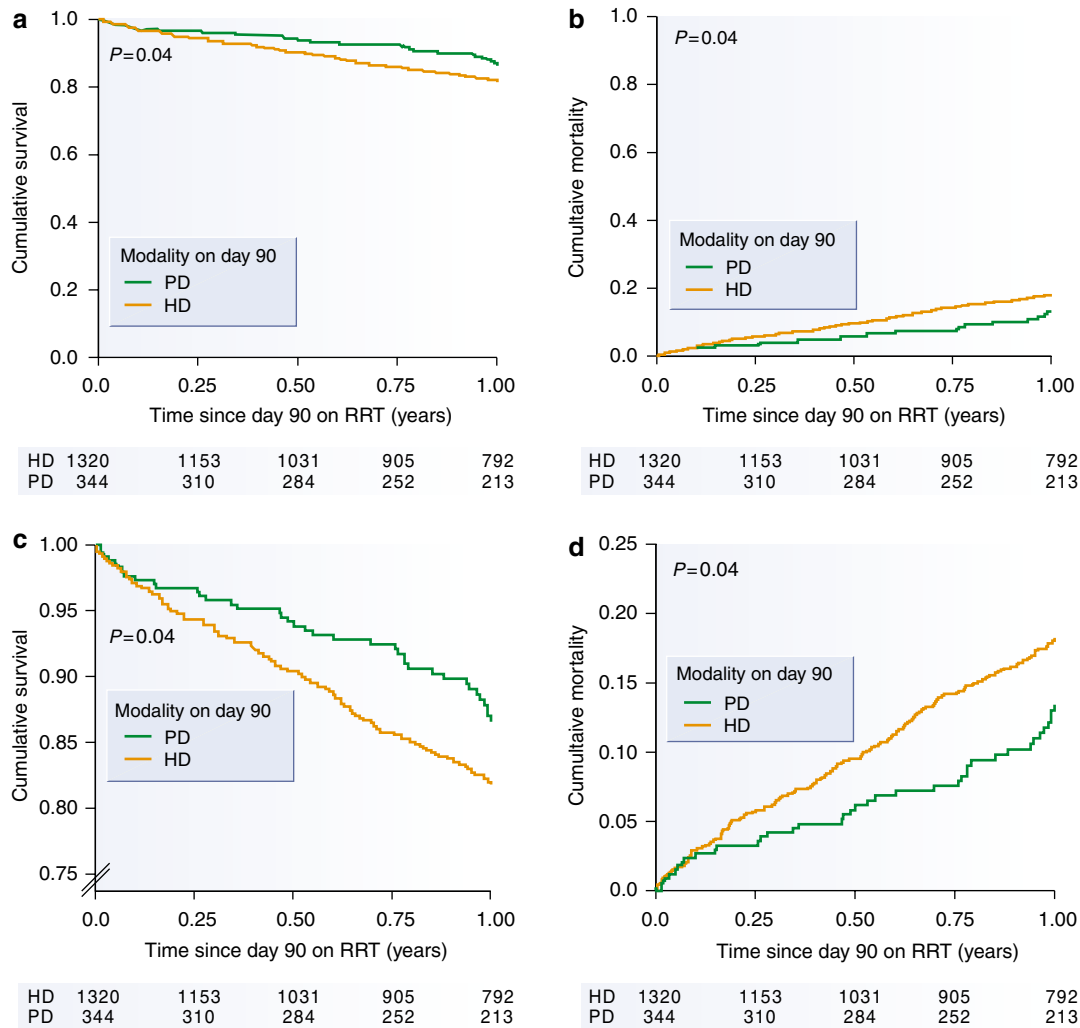
Figure 3 | Survival plots by treatment modality. (**a**) Cumulative survival (*y* axis 0–1); (**b**) cumulative mortality (*y* axis 0–1); (**c**) cumulative survival (*y* axis 0.75–1.0); (**d**) cumulative mortality (*y* axis 0–0.25). The numbers below the figures denote the number of patients 'at risk' in each group.

Example 3. Survival probability in patients treated with hemodialysis and peritoneal dialysis

In a sample of almost 1700 dialysis patients taken from a study on the effect of comorbidity on survival,[12] survival probabilities from day 90 after the start of RRT were calculated for patients on hemodialysis and peritoneal dialysis.

Figure 3 shows cumulative survival and cumulative mortality plots containing the same information, but using different ranges for the *y* axis. We will use these to discuss the recommendations of Pocock *et al.*[13] on the publication of survival plots. These authors have a preference for survival plots being presented upward as cumulative mortality, especially for studies with a low event rate (lower than 30%), as in this way without using the full range of 0–1 for the *y* axis (Figure 3d), it is possible to provide more detail without the need for a break in the *y* axis (Figure 3c). Secondly, they state that the number of participants that is still at risk should be listed under the time axis and that plots should only be extended through the period of follow-up

achieved by a reasonable proportion (10–20%) of participants. Furthermore, they recommend plots always to include some measure of statistical uncertainty. For Kaplan–Meier curves, this may be the *P*-value derived from the log-rank test, whereas for Cox regression, hazard ratios may be presented together with their confidence intervals. Therefore, according to Pocock *et al.*, Figure 3d would be the best way to present the data in example 3.

Another matter to consider while inspecting survival plots is how likely the outcomes are over time. The incidence rate of an event may be high early during follow-up with a decrease later or vice versa. In Figure 4, the follow-up in example 3 was extended to 3 years. It shows that after 2.5 years, the difference in survival between the modalities became smaller to disappear at 3 years. Therefore, in survival plots, it is important to look at time-dependent effects, which will be extensively discussed in a future article in this series. To be able to do so, it is important to have follow-up data of sufficient duration on an adequate sample size.
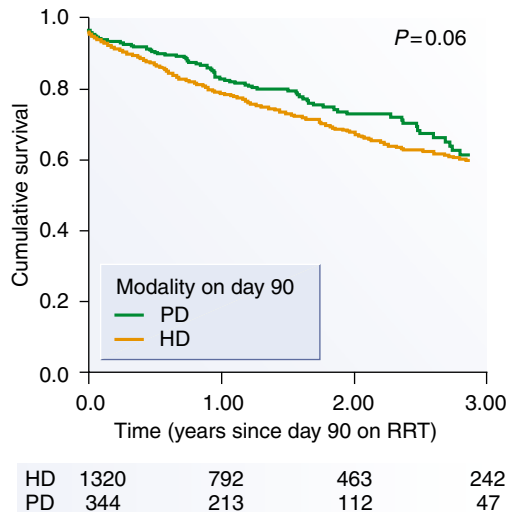
HD	1320	792	463	242
PD	344	213	112	47

Figure 4 | Survival plot by treatment modality. The numbers below the figures denote the number of patients 'at risk' in each group.

WHAT THE KAPLAN–MEIER METHOD AND THE LOG-RANK TEST CAN AND CANNOT DO

The Kaplan–Meier method is the most popular method used for survival analysis. Together with the log-rank test, it may provide us with an opportunity to estimate survival probabilities and to compare survival between groups. Most of the time, however, one would like to do more than that. In example 3, where the survival on hemodialysis and on peritoneal dialysis were compared, one would have liked to be informed on the size of any potential difference. In addition, to make a more fair comparison between the two dialysis modalities, one would have liked to adjust for age and other confounding variables. However, as the log-rank test is purely a significance test, it cannot provide an estimate of the size of the difference between groups and a related confidence interval. Secondly, the Kaplan–Meier method and the log-rank test can only study the effect of one factor at the

time, and therefore they cannot be used for multivariate analysis. For these purposes, one may use a regression technique like the Cox proportional hazards model, which will be described in the next article in this series.

DISCLOSURE
All the authors declared no competing interests.

SUPPLEMENTARY MATERIAL
Supplementary material is available.

REFERENCES
1. ERA-EDTA Registry: ERA-EDTA Registry 2005 Annual Report Academic Medical Center, Department of Medical Informatics, Amsterdam, The Netherlands, June 2007.
2. Heart Outcomes Prevention Evaluation (HOPE) Study Investigators. Effects of ramipril on cardiovascular and microvascular outcomes in people with diabetes mellitus: results of the HOPE study and MICRO-HOPE substudy. *Lancet* 2000; **355**: 253–259.
3. Stel VS, van Dijk PC, van Manen JG *et al*. Prevalence of co-morbidity in different European RRT populations and its effect on access to renal transplantation. *Nephrol Dial Transplant* 2005; **20**: 2803–2811.
4. Wanner C, Krane V, Marz W *et al*. Atorvastatin in patients with type 2 diabetes mellitus undergoing hemodialysis. *N Engl J Med* 2005; **353**: 238–248.
5. Drüeke TB, Locatelli F, Clyne N *et al*. Normalization of hemoglobin level in patients with chronic kidney disease and anemia. *N Engl J Med* 2006; **355**: 2071–2084.
6. Altman DG, Bland JM. Time to event (survival) data. *BMJ* 1998; **317**: 468–469.
7. Jager KJ, Zoccali C, Kramar R *et al*. Measuring disease occurrence. *Kidney Int* 2007; **72**: 412–415.
8. Bland JM, Altman DG. Survival probabilities (the Kaplan–Meier method). *BMJ* 1998; **317**: 1572.
9. Wolfe RA, Ashby VB, Milford EL *et al*. Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *New Engl J Med* 1999; **341**: 1725–1730.
10. Van Dijk PC, Jager KJ, Stengel B *et al*. Renal replacement therapy for diabetic end-stage renal disease: data from 10 registries in Europe (1991–2000). *Kidney Int* 2005; **67**: 1489–1499.
11. Bland JM, Altman DG. The logrank test. *BMJ* 2004; **328**: 1073.
12. van Manen JG, van Dijk PC, Stel VS *et al*. Confounding effect of comorbidity in survival studies in patients on renal replacement therapy. *Nephrol Dial Transplant* 2007; **22**: 187–195.
13. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet* 2002; **359**: 1686–1689.