

* CHAPTER 18

Statistical evaluation of HSCT data

R.M. Szydło

1. Introduction

Stem cell transplantation (HSCT) is a widely accepted treatment modality, with both allogeneic and autologous HSCTs offering effective options for a number of diseases (eg. some leukaemias and severe aplastic anaemia) and curative potential for others (e.g. thalassaemia and CML). However, there is still much to be learnt, and the analysis of data generated from a stem cell transplant programme is not only fundamental to assessing the effectiveness of the treatment, but can provide invaluable information on the prognostic role of disease and patient factors. Thus, the appropriate analysis of such data is of paramount importance.

2. Outcomes

Patients who undergo a HSCT procedure require considerable support and supervision, which in turn, allows the treatment modality to be reviewed in a variety of ways. Key events are assessed at varying times post-HSCT and these can be used to calculate a number of outcomes defined below:

- Survival - the probability of survival irrespective of disease state
- Disease-free survival (DFS) - the probability of being alive and free of disease (in leukaemia this outcome could also be termed leukaemia-free survival (LFS))
- Graft vs. host disease (GvHD) - the probability of developing GvHD (the severity of disease being estimated would need to be clearly stated)
- Graft failure (GF) - the probability of primary graft failure
- Transplant related mortality (TRM) - the probability of dying without recurrence of disease
- Relapse - the probability of disease recurrence
- Progression-free survival (PFS) - the probability of being alive and with a disease stage not advanced of that at the time of transplantation
- Neutrophil engraftment - defined as the first of 3 consecutive days post HSCT where values above a specified level are achieved (e.g. $\geq 0.5 \times 10^9/L$)
- Platelet engraftment - defined as the first of 3 consecutive days post HSCT where values above a specified level are achieved (e.g. $\geq 50 \times 10^9/L$)

Probability curves describing these outcomes fall into two categories: Survival, DFS/LFS and PFS involve events with decreasing cumulative probabilities over time, whilst GvHD, TRM, GF and relapse involve events that result in increased cumulative probabilities over time.

3. Survival analysis

The outcomes outlined above require careful consideration before a statistical analysis can be considered. Each event of interest may occur at variable times post transplant,

so in statistical terms it has two components - whether it occurs at all and, if it does, the length of time from transplant to the event. However, inherent in many studies is the problem that the event of interest is seldom observed in all of the patients. Thus, a patient who has not yet had the event of interest at the time of analysis, or who is lost to follow-up, would be “censored” at the time of last contact. The inclusion of data that is censored precludes the use of simple statistical methods such as chi-squared analysis or rank methods and requires a statistical treatment known as survival analysis, which can be applied to a variety of end points.

3.1. Kaplan-Meier method

There are a number of methods for analysing survival data, and though these depend on the precision of the recorded time interval, are usually summarised as survival or Kaplan-Meier (1) curves which are derived from calculated tables commonly known as life tables (constructed on the basis of a series of conditional probabilities). The term life table is also frequently used to describe data where the results are grouped into time intervals, often of equal length, and this method of calculation is described as actuarial. In fact the terms “actual” and “actuarial” are often used mistakenly to describe survival probabilities generated by Kaplan-Meier methods.

Figure 1 shows typical survival data from fifteen consecutive patients transplanted at a single centre. Four have died and eleven were still alive at various time points post-transplant. If the data are rearranged in order of time, then a life-table can be calculated by the method of Kaplan-Meier as shown in Table 1.

Figure 1: Survival data from fifteen patients who received a haematopoietic stem cell transplant

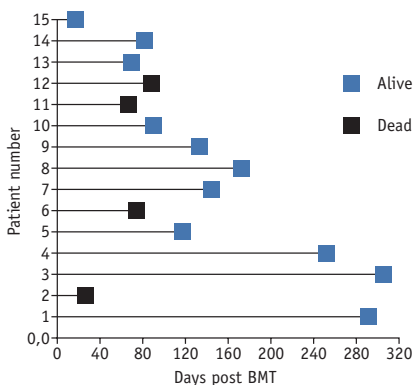


Table 1: Life table for fifteen patients who received an allogeneic haematopoietic stem cell transplant

Time (days)	Status (0=alive, 1=dead)	Number at risk	Probability of survival	Standard error
16*	0	15	1.00	
26	1	14	0.93	0.069
66	1	13	0.86	0.094
69*	0	12		
74	1	11	0.78	0.113
82*	0	10		
88	1	9	0.69	0.129
89*	0	8		
117*	0	7		
133*	0	6		
144*	0	5		
172*	0	4		
252*	0	3		
291*	0	2		
305*	0	1		

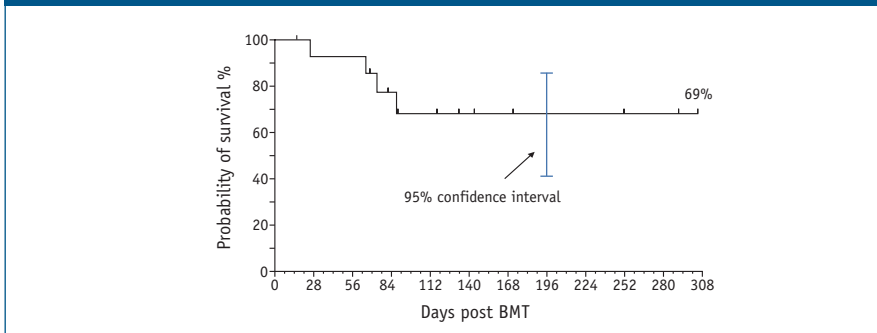
* censored observation

The data presented in Table 1 can be used to produce a survival curve, also known as a cumulative survival rate, or a survival function (Figure 2). Vertical tick marks on the curve represent censored individuals who make no contribution to the curve after that particular time point. The curve is an estimated probability of survival, and using appropriate methods to compute the standard error, 95% confidence intervals (95% CIs) can be calculated. In common with many analyses of small data sets, the standard error calculated from day 88 post BMT has yielded a large 95% CI, and so the survival curve must therefore be interpreted with some caution.

3.2. Cumulative incidence procedure

The following outcomes—relapse, TRM, GvHD and GF are subject to the problem of “competing risks” (for example, in the case of calculating a relapse probability, a patient who dies in remission cannot relapse), and so the most appropriate method of analysing such data is to produce a cumulative incidence curve (2). Although this methodology is not included in most commercial statistical packages, it is present in the statistical package NCSS (*Statistical analysis & data analysis software*), and

Figure 2: Probability of survival following a haematopoietic stem cell transplant (n=15)



macros are available to allow such curves to be calculated using the statistical packages SAS and R (3).

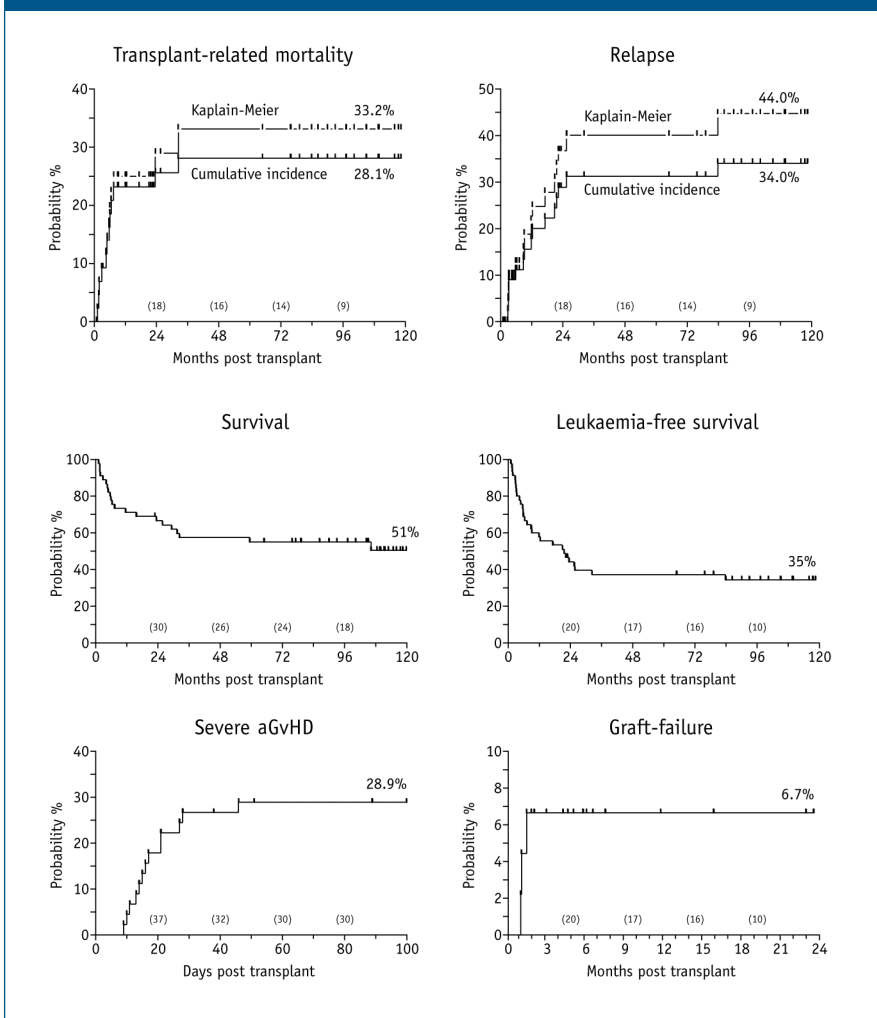
Although it is possible to use the Kaplan-Meier method for each of these outcomes, this is likely to produce an overestimate of the true probability as calculated by the cumulative incidence approach (4). This discrepancy will be largest where the event of interest occurs later after HSCT (TRM, relapse and chronic GvHD - see [Figure 3](#)) and may be negligible with early outcomes (acute GvHD and graft failure).

4. Other methods for describing outcomes

Recent advances in statistical methodology have enabled LFS and DFS curves to be modified to take into account durable remissions achieved after relapse (5). The generation of current leukaemia-free survival (CLFS) curves does however require detailed follow-up data and the use of macros designed for the statistical software package SAS.

If the exact times to the onset of GvHD are not known, then simple proportions of grades of disease can be presented for those patients who survived long enough potentially to develop the disease (thus patients who died within 100 days post-HSCT are not eligible for chronic graft versus disease). Engraftment times can either be described with a median and range, or with cumulative probability curves. Comparisons between GvHD groups should be made using the chi-squared test or chi-squared trend test, whilst the Mann-Whitney or Kruskal-Wallis test are applicable for engraftment data. An example of data presentation from 45 consecutive patients is illustrated in [Figure 3](#).

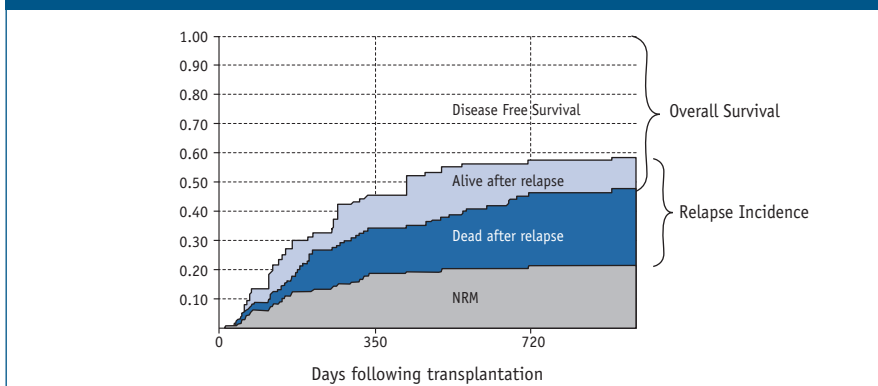
Figure 3: Probability curves of 6 different possible outcomes



5. Composite outcome diagrams

An interesting new way of graphically representing how outcome probabilities change with time post-HSCT has been developed by Ronald Brand and is included in a recent paper (Figure 4) (6). Cumulative incidences of relapse, and of non-relapse

Figure 4: Composite outcome diagram



death are estimated and simultaneously plotted with the Kaplan-Meier survival probability. The resulting diagram thus provides the 4 possible patient states after a HSCT - alive without relapse of disease (white), alive after relapse (light blue, an outcome not normally calculated), dead after relapse (dark blue) and non-relapse death (grey). In addition, the proportion of patients relapsing (relapse incidence) is provided by the sum of the horizontal and diagonal hashed groups, the interface between the alive and dead components represents overall survival, and the interface between alive with and without relapse – relapse-free survival. One is thus able to view in one diagram the relative importance of all these possible outcomes. This is especially useful for illustrating differences between groups identified from univariate or multivariate analyses as being of prognostic significance. Specialist software is not required to create such diagrams, as macros are available for the statistical package SPSS-14.

6. Comparison of survival curves

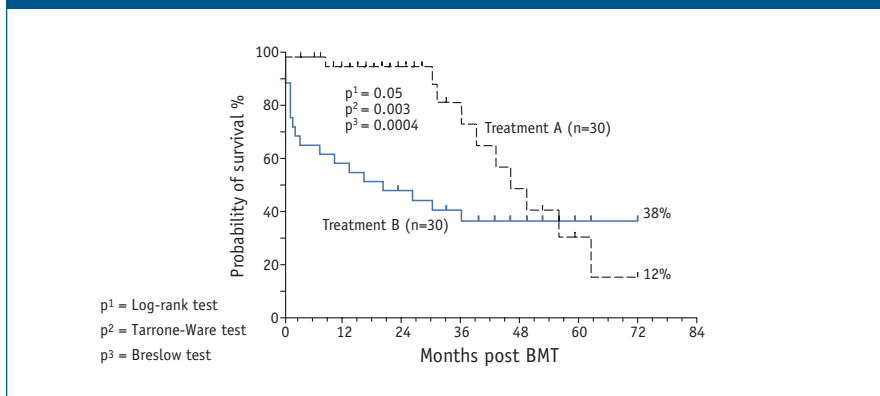
Survival curves provide a visual assessment for disease course and/or outcome of a particular treatment or disease course. In order to establish whether there is a survival advantage between, for example, two treatments, it is necessary to perform a statistical test to compare the two life tables. This is achieved using the logrank or Mantel-Cox test (7). In this test each observation is given an equal “weight”. However, in the transplantation setting, where there may be considerable early mortality, it may be more useful to “weight” early observations and in this context the Breslow test (8) may be more appropriate (this test is also less sensitive to late

events when few subjects may be present in the study). The Tarrone-Ware test (9) is a compromise between the Mantel-Cox and Breslow tests. An example of the relative merits of these tests is provided in Figure 5. Here survival data from two hypothetical groups are presented. The proportion of survivors could be compared using a standard chi-squared test (treatment A 20/30 vs. treatment B 12/30), yielding a significant result ($p=0.04$), and suggesting treatment A to be the superior. However, this analysis ignores the time to an event, and gives a misleading result as exemplified by Figure 5. There is a clear survival advantage in the first four years post BMT for treatment A, but for long-term survival, treatment B may be better. The curves presented in Figure 5 do however highlight another analytical problem. In order to perform a log-rank test, the groups being tested should run in parallel and not cross over. A more sophisticated approach for analysing such data should therefore be undertaken (4).

In addition to comparing treatments, the log-rank test can be used to compare selected sub-groups within one treatment or disease category e.g. males vs. females, patient age <30y vs. patient age ≥ 30 y, early vs. late stage of disease, etc. As with all statistical comparisons of subgroups, a more stringent criterion for significance needs to be predetermined in accordance with the number of tests to be performed.

If one or more prognostic variables are known, then a stratified log-rank analysis can be undertaken to look at the influence of these factors on outcome. Thus, for example, in acute myeloid leukaemia where disease sub-type is an important prognostic indicator, the effect of patient sex could be investigated with a stratified log-rank test using disease sub-type as the stratified variable.

Figure 5: Probability of survival following BMT for two hypothetical groups



7. Presentation of survival data

In quoting survival rates, probabilities at specific points in time should be indicated together with a confidence interval or standard error (e.g. the probability of survival at 3 years was 59% (95% CI's 44–73%)). Although 95% CI's can be calculated by taking the survival rate ± 1.96 *standard error, if the survival rate is close to 100% or 0%, this can lead to confidence intervals greater than 100% or less than 0%. To avoid this, asymmetrical confidence intervals can be calculated by the method of Rothman (10).

Median survival times can be derived from survival curves, and correspond to the time at which the survival proportion reaches 0.5, but this is not always possible (as demonstrated in Figure 2) if the survival curve reaches a plateau above this point. Guidelines for the presentation of results of transplantation data have been suggested by Klein (4) and Labopin (3). The presentation of survival curves from univariate analyses of prognostic factors should be viewed with extreme caution, as adequate control of other potential prognostic factors or biases cannot be guaranteed. A multivariate approach is therefore to be recommended.

8. Proportional hazards regression analysis

The log-rank test enables the survival experience of two or more groups to be compared but in order to investigate a number of possible prognostic variables simultaneously, a regression method introduced by Cox and known as proportional hazards regression analysis has to be employed (11). The special nature of survival data as outlined previously, makes the use of usual regression methods (e.g. linear regression, logistic regression) inappropriate. The use of the Cox model allows the identification of prognostic factors that are related to the outcome. In addition, the search for variables of unknown prognostic significance can be performed after adjusting for variables of known prognostic significance. Thus for example, in chronic myeloid leukaemia where the stage of disease at transplant is a major factor in survival, the influence of other factors would be investigated having taken into account disease stage. This approach also allows for the generation of survival curves for a given factor that are adjusted for the influence of other factors. Such curves are likely to be much more informative than simple univariate analysis curves.

The use of the Cox model does require a sound statistical knowledge, as there are many potential difficulties with the method both in application, and interpretation of results. Several reviews of the subject have been published and are recommended (3, 4).

9. Conclusions

The analysis and presentation of survival data can provide important information on the effectiveness of transplantation in treating a particular disease, and with sufficient numbers of patients, subtle differences between patient groups can be identified. The now routine availability of computers and statistical software enables the analysis of complex data sets to be carried out with relative ease, but the importance of statistical advice at all stages of data analysis should not be underestimated.

References

1. Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. *J Am Stat Assoc* 1958; 53: 457-481.
2. Gooley TA, Leisenring W, Crowley JA, Storer BE. Estimation of failure probabilities in the presence of competing risks: New representations of old estimators. *Statistics in Medicine* 1999; 18: 695-706.
3. Labopin M, Iacobelli S. Statistical guidelines for EBMT. 2003; http://www.ebmt.org/1What-is-EBMT/Op_Manual/OPMAN_StatGuidelines_oct2003.pdf
4. Klein JP, Rizzo JD, Zhang M-J, Keiding N. Statistical methods for the analysis and presentation of the results of bone marrow transplants. Part I: Unadjusted analysis. *Bone Marrow Transplant* 2001; 28: 909-915. Part II: Regression modeling. *Bone Marrow Transplant* 2001; 28: 1001-1011.
5. Craddock C, Szydlo RM, Klein JP, et al. Estimating leukemia-free survival after allografting for chronic myeloid leukemia: a new method that takes into account patients who relapse and are restored to complete remission. *Blood* 2000; 96: 86-90.
6. Lim ZY, Ingram W, Brand R, et al. Clonal gammopathies following alemtuzumab-based reduced intensity conditioning haematopoietic stem cell transplantation: Association with chronic graft-versus-host disease and improved overall survival. *Bone Marrow Transplant*, Epub 2007, Aug 20: 747-752.
7. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 1966; 50: 163-170.
8. Breslow N. A generalised Kruskal-Wallis test for comparing k samples subject to unequal patterns of censorship. *Biometrika* 1974; 57: 579-594.
9. Tarone RE, Ware J. On distribution-free tests for equality of survival distributions. *Biometrika* 1977; 64: 156-160.
10. Rothman KJ. Estimation of confidence limits for the cumulative probability of survival in life table analysis. *J Chron Dis* 1978; 31: 557-560.
11. Cox DR. Regression models and life tables. *Journal of the Royal Statistical Society* 1972; 34, Series B: 187-220.

Mutiple Choice Questionnaire

To find the correct answer, go to <http://www.esh.org/ebmt-handbook2008answers.htm>

1. Who are censored patients?

- a) Those who have experienced the event of interest early after SCT.....
- b) Those who have not experienced the event of interest and are alive
- c) Those who have experienced the event of interest but are lost to follow-up.....
- d) Those who failed to engraft.....

2. What is the correct test for comparing two Kaplan-Meier survival curves?

- a) Chi-squared test
- b) Mann-Whitney test
- c) T-test
- d) Log-rank test.....

3. What is the cumulative incidence procedure used for?

- a) To calculate survival curves where there is a competing risk to the event of interest
- b) To compare survival curves where the event of interest occurs early after SCT
- c) To compare survival curves where the event of interest occurs late after SCT.....
- d) To identify censored patients.....

4. What proportional hazards regression analysis is used for which of the following:

- a) To calculate survival curves with more than one event of interest.....
- b) To help identify prognostic factors.....
- c) To identify patients lost to follow-up.....
- d) To censor events of interest that are competing risks.....

5. How do cumulative incidence curves differ from Kaplan-Meier curves:

- a) Only when the event of interest is time independent
- b) By taking into account competing risks
- c) By providing the correct estimate of the competing risk
- d) Because they don't take into account censored events