

Statistical modeling of complex disease histories in Bone Marrow Transplant.

Guidelines for proper use and interpretation of the Cox model for the European Group for Blood and Marrow Transplantation.

Simona Iacobelli
(iacobelli@bce.uniroma1.it)

EBMT Statistical Sub-Committee
GIMEMA (Italian Group for Hematological Diseases in Adults), Biostatistics Group

Acknowledgments

This work is a result of a research project financed by the European Group for Blood and Marrow Transplantation Research sub-committee.

The author would like to thank for their interest and help prof. van Houwelingen and dr. Ronald Brand of the Department of Medical Statistics of Leiden where the project was developed.

Many hints were given by the researches of the Chronic Leukemia Working Party, in particular dr. Curly Morris of the Hematology Department of the Belfast City Hospital who proposed the problem and provided the data for the study case developed in this report.

Finally, the author would like to thank the colleagues of the EBMT Statistical Sub-Committee for their comments and help in the final review of this document.

Contents

Introduction.....	4
1 Complex disease histories in BMT: methodological problems	5
2 Modeling and interpretation: the multi-state approach.....	8
2.1. Dealing with intermediate events.....	8
2.2. A comprehensive approach.....	10
2.3. Some examples and further remarks.....	13
3 Statistical analysis.....	17
3.1. Building the model.....	18
3.2. Summarizing the results.....	20
3.3. Estimation of summary probabilities	22
4 A study case: double transplantation in multiple myeloma.....	24
Suggested Readings	30

Introduction

The complexity of the story of the disease that develops after stem cell transplantation (SCT) requires a careful planning of the statistical analysis. Traditional methods like the Kaplan-Meier survival estimation or the standard Cox proportional hazards model are not sufficient to tackle multiple events and to assess the role of their occurrence and timing, problems which are often encountered in current research on SCT. On the other hand, certain features of the Cox model (in particular, the use of time-dependent covariates) can be applied to provide results that, if properly manipulated and interpreted (following the so-called multi-state modeling), can answer to many important clinical questions. Apart from some technical difficulties, the major challenge is that such an analysis must be planned integrating clinical issues with statistical (methodological) knowledge.

With this respect, while much statistical literature on the use of Cox model for complex disease histories is currently available, the applications of this methodology to real clinical studies are still very few. As a consequence, most researchers still ignore the existence and potentials of these methods, or even they are not fully aware of the limitations of the standard methods and of the possible pitfalls encountered if they are not properly applied. This in turn means that often clinicians “don’t formulate the correct questions”, and at the same time they miss the opportunity of extracting useful information from the data. On the other side, statisticians who work for medical research need to get more familiar with the multi-state approach for real clinical problems, and to overcome the technical difficulties in the practical application.

This work is meant to fill in some gap between clinical research and statistical methodology and to help researchers of the EBMT with the practical application of the multi-state approach. Using examples from current research on SCT, it provides intuitive explanations on the methodological problems and the correct statistical approaches, starting from the nature and use of time-dependent covariates in Cox model up to the general multi-state approach. Some examples of “typical” clinical questions in stem cells transplant research (involving e.g. graft-versus-host disease or donor lymphocyte infusion) will be “translated” into statistical terms, showing possible approaches to the analysis. One of this examples (regarding tandem transplantation and its timing) is developed in detail. Readers not interested in statistics could restrict to sections 1, 2 and 4. For those involved with the statistical analysis section 3 provides a brief introduction of the main elements of multi-state modeling and some suggestions of useful readings from current literature.

For the practical application of the multi-state approach, EBMT statisticians may contact the author (e-mail: iacobelli@bce.uniroma1.it) to view the R functions used for the implementation of the study case.

1 Complex disease histories in BMT: methodological problems

Treatment with (high-dose) chemotherapy followed by stem-cell recovery (here and in the following called for simplicity BMT, bone marrow transplantation) determines a complex course of the disease. Several different “events” can occur, which can be either further treatments (like subsequent transplantations or DLI) or outcomes, e.g. hematopoietic recovery, loss of engraftment, graft-versus-host disease (for allogeneic transplantation), relapse, death due to complications (such as infections), death due to relapse. Of course these events can be strongly related to each other. Treatments determine outcomes, and the other way round (DLI follows a relapse, for example), and some outcomes are mutually exclusive, that is, they cannot occur together, like death due to complications and relapse death (in statistical terms, they are called “competing risks”). As a further complication, not only the occurrence of an event can be a predictor for the risk of another event, but also its timing.

Modern studies in BMT research for the analysis of treatment efficacy and the investigation of prognostic factors require to take into account such complexity, but standard statistical methods like Kaplan-Meier curve and simple Cox models are inadequate for this purpose. This kind of problems will be now illustrated in the context of current research issues in BMT, focusing on typical “mistakes” in planning and interpretation; in section 2 proper approaches and methods for the analysis will be suggested.

Role (and timing) of 2nd SCT in multiple myeloma

Patients with multiple myeloma given autologous Stem Cells Transplantation often experience a relapse of their disease; after the relapse, it is possible to administrate another transplantation to attempt to re-induce remission. As an alternative to this “wait-and-see strategy”, in recent years it is becoming more and more common to administrate a second SCT in a short time after the first, while the patient is still in remission, with the purpose of reducing the risk of relapse. It is of strong interest nowadays to evaluate the effect of such an “elective” second SCT, and in particular to draw indications on the “optimal” time to 2nd transplantation.

Of course the only way to establish a *causal* relationship between a treatment (the elective 2nd transplantation) and the final outcome is to perform a randomized clinical trial, where the elective SCT is randomized at the time of first transplantation between two groups of patients comparable with respect to the prognostic factors, and to carry out the analysis on the basis of the intention to treat. Anyway, an investigation of data collected retrospectively is necessary at a previous step to provide support for hypotheses to test in a clinical trial, and this kind of analysis is more at risk of wrong planning and interpretation.

In the retrospective analysis, the typical mistake is to perform a direct comparison between patients who had an elective 2nd SCT and patients who didn't using a traditional method like the Kaplan-Meier curve or the estimation of the Hazard Ratio via a standard Cox model. In fact, in order to get an elective 2nd SCT patients had to survive relapse-free for a certain period; all those who died or relapsed soon after the 1st transplant belong to the second group, for which it is therefore likely to observe a lower survival, even in the case that the elective 2nd SCT actually increased the risk of death. What is not taken into account in such a procedure is that being transplanted for the second time while in remission is not a condition given at the beginning of the time period, but it is an *event* that occurred during the course of the disease.

Therefore, any comparison must be made conditionally on being alive relapse-free at a certain point in time, and the proper comparison is between patients who have *already* had an elective 2nd SCT and patients who didn't have it *yet*.

A second important point is that in order to provide a complete picture of the process of the disease with and without transplant we must take into account also the probability of *getting* the second SCT with each particular pattern of risk factors. For example, suppose that in a group of patients with certain characteristics we observe a much better outcome for those who got a 2nd transplant than for those who didn't get it; if given that pattern of covariates it was very difficult to get the elective second transplant due to a high risk of early relapse, then it is possible that the transplanted patients were actually (for some unknown characteristics) particularly protected against dismal outcomes, so that the advantage observed for this small group must be traced back not to the effect of transplantation, but to an auto-selection mechanism. In statistical terms, analyzing the probability of getting an elective 2nd transplantation means that this event must also be considered *as an outcome* of the process, and modeled separately (considering death in remission and relapse as competing risks). This further insight in the administration of transplantation could be an important contribution to the interpretation of the results, and it is anyway needed to give a complete picture of the process.

In conclusion, in this apparently simple example we enlightened two methodological problems: the need for a comparison *in time* between groups determined by the occurrence of an intermediate event of the process of the disease (the event is a predictor for the final outcome), and the need of a separate analysis of the occurrence of such event (the event is an outcome itself).

Role of acute and chronic graft-versus-host disease (GvHD)¹

The role played by acute and chronic GvHD in the process of the disease after stem cells transplantation is still object of investigation in BMT research. In some cases, there is need of enlarging the knowledge on the biological mechanisms of the disease, for example regarding the anti-tumor effect, therefore GvHD is the specific object of investigation. In other cases, researchers want to take into account GvHD as an adjustment risk factor when evaluating different treatment strategies or other prognostic factors.

From a methodological point of view, the manifestation of GvHD is an outcome of the process of the disease after transplantation that in turn must be considered as a predictor for other outcomes. The framework of the analyses involving GvHD is always a competing risks one: when analyzing GvHD as an outcome, because death for causes other than GvHD prevents the occurrence of GvHD; when including GvHD as a predictor for the final outcomes, because it is necessary to take into account at the same time its effect on treatment-related death and on relapse.

In the latter situation, as in the study of double transplantation seen before, an analysis which neglects the nature of *event* of acute and chronic GvHD is possibly biased and

¹ In very general terms, the graft-versus-host disease (GvHD) is an immunological reaction of the transplanted (grafted) immune cells against the recipient's tissues (of course, in the context of stem cells transplantation from external donor, or allogeneic transplantation). The acute form occurs in the first 100 days after transplant, the chronic form can occur by definition only after 100 days, as a persistence of the acute one, or as *de novo* GvHD. Diagnosis, treatment and prognosis are different for the two forms. GvHD plays an important role as risk factor for transplant-related mortality, but it is also thought to have a protective effect against relapse, depending on the disease being considered. Chronic GvHD is moreover particularly important in studies regarding the quality of life after transplantation and when investigating long-term outcomes.

incomplete. For example, consider the problem of evaluating the effect on survival of two different treatments A and B, where we want to take into account the occurrence of acute GvHD as an adjustment factor. First of all, it can't be included as a covariate in the same way as, say, the gender, which is a characteristic of the patient known since the beginning of the process: this would result in a classification of the patients based on future outcomes, possibly leading to seriously biased conclusions. Then, it must be evaluated separately as an outcome of the process: in fact if the treatment influences the occurrence of aGvHD, part of the effect of the treatment is captured by the covariate representing aGvHD.

The time dimension of the phenomenon can not be left aside from the analysis also in the case of chronic GvHD. First of all, any analysis involving cGvHD must be based (due to its definition) only on patients survivors at 100 days after transplant; then, its occurrence must be evaluated together with its timing. It is instead still very common to report only about the incidence of cGvHD; but a lower number of cases with cGvHD could be observed in a certain group of patients simply because in that group the survival is short, due for example to a high incidence of infection deaths. This kind of mistake often originates from the lack of data regarding the timing of chronic GvHD.

Further complex issues related to the *timing* of GvHD are current objects of investigation: researchers are interested in taking into account the time elapsed between the diagnosis and the occurrence of the two events, or between the acute and the chronic forms, or, in studies regarding cGvHD in particular, the time intervals between subsequent episodes and the duration of such episodes. These studies can actually be highly complicated, or even unfeasible; a careful planning of the analysis can be able to reduce the complexity of the problems and to keep at the same time the potential to focus on the clinical issue under investigation. We will see an example in section 2.

Current Leukemia-Free Survival in chronic leukemia

Patients who experience a relapse of chronic myelogenous leukemia (CML) after allogeneic SCT can gain a second, durable remission with donor lymphocyte infusion (DLI)². Therefore a relapse can be just a temporary state. This opportunity offered by DLI is a peculiarity of stem cells transplantation, and there is a great interest in taking it into account when comparing transplantation with alternative treatments, like Imatinib. The object of interest is the probability of being alive leukemia-free at any point in time, not restricted to first remission.

The methodological problem here is that this probability does not coincide with what is calculated by the standard method based on the Leukemia-free survival (LFS) curve, which is the probability of surviving and *never experiencing a relapse*. In fact, traditional "survival" methods were born to analyze the time to death, that is, to a single, "final" event; the standard LFS curve considers as a final event the combination "treatment-related death or relapse". In the new context of CML relapse is not a "final" event anymore, since it is possible to return to a leukemia-free state.

² DLI is an immunotherapy which exploits the anti-leukemia effect of the graft without causing severe GvHD. The administration of the treatment can follow different procedures, for example it can be given in one dose or in several (escalating) doses.

The traditional LFS curve clearly underestimates the probability of interest, therefore it is no more suitable to analyze the course of the disease after BMT³.

A new method must take into account that relapse is an outcome of the process that starts at transplantation, but it is also the starting point of another process that could give origin to DLI and then to the achievement of a (second) Complete Remission. In this second process, death is a competing risk both for DLI administration and for achievement of CR. All the events involved have also a time-dimension, and both occurrence and timing play a role in the processes. Neglecting any of these elements, as it was shown in the previous examples, could result in an analysis seriously biased. Also in this case, planning the analysis could be very complicated, in particular if the modalities of administration of DLI are a particular object of interest.

DLI studies

Studies are currently being carried out to indicate the best regimen; the analysis must therefore evaluate not only the dose of donor lymphocytes given, but also the role of the timing of the procedure, for example taking into account the time interval between relapse and DLI, and, in case of the escalating dose regimen, between subsequent infusions. We will see in the next section that all these complications can in principle be handled by a unique approach: the multi-state approach.

2 Modeling and interpretation: the multi-state approach

As it was illustrated in the previous section, traditional, simple statistical methods like Kaplan-Meier curve and basic Cox model are inadequate or insufficient to handle the complexity of the studies needed in the framework of BMT research. In this section, we will see how simple extensions and combinations of those methods (specifically, the Cox model) can be used for the statistical analysis, and how to interpret properly the results.

2.1. Dealing with intermediate events

The first methodological problem pointed out in section 1 is related to predictors with a time-dimension, that is, to events that occur during the course of the disease and that we need to take into account in an analysis of some final outcome. Such events are also called “intermediate” events (IE); examples are the 2nd transplantation and the occurrence of GvHD.

We cannot compare patients who for example had a 2nd transplant with those who didn't in the same way as we compare males with females. The mistake is defining the groups to be compared on the basis of a future outcome of the process. In fact, even though transplantation is a treatment, based on a decision by the treating physician, its administration is conditional on being alive (and relapse-free), and it could also depend on the prognosis for the patient at that time. Indeed, any kind of intermediate event must be considered, from a methodological point of view, as an outcome of the process.

Correct methods of analysis are the landmark analysis or the use of a Cox model with a time-dependent covariate for the intermediate event.

³ Notice that the estimation of the disease-free survival curve in a context where relapse is not a final event is needed also in studies regarding other diseases than CML, e.g. in myeloma, where patients can achieve a second remission after relapse.

Generally speaking, the **landmark analysis** for an outcome is restricted to patients alive and at risk of experiencing the event of interest at a certain point in time. For example, the analysis of relapse risk could be restricted to patients alive and in remission 1 year after the diagnosis. It can be used to handle the presence of an intermediate event, comparing patients who at that time have *already* had the IE with patients who didn't have it *yet*. This grouping of patients is conceptually equivalent to the grouping based on gender, because *having had* the IE is a characteristic of the individual, known at the time origin of the process. The restriction of the conclusions that it is possible to draw from such analysis is that they are valid for patients surviving at least up to that certain time x , and the comparison regards IE that occur in the first x months. The limitation of this approach is therefore that for a comprehensive understanding of the phenomenon we should repeat the analysis choosing different conditioning times.

The use of **time-dependent covariates in the Cox model** is with this respect more powerful. Time-dependent covariates are, generally speaking, variables whose value changes in time. In this framework, they can be used to represent the occurrence of the IE. For example⁴, we can define a covariate X that is 0 from the start of the clock (e.g. date of first transplantation) to the date of IE, then it switches to 1: the characteristic is absent before the occurrence of the IE, and it is present afterwards. The model for survival including X will return a hazard ratio (HR) for X , together with the hazard ratios for the other (time-fixed) covariates. The HR for X expresses the *change* in the evaluation of the hazard of death at the occurrence of the intermediate event, adjusted for the other characteristics of the patient (gender, conditioning regimen, etc), that remain the same. For example, in the study on tandem transplantation, an HR of X equal to 0.8 would mean that, in a group of patients with a certain pattern of characteristics, the hazard of the patients with a 2nd transplant is 20% lower than the hazard of those who *at the same time* are alive relapse-free without second transplantation⁵.

Whichever approach we use, it is important to stress that when **interpreting the results** we talk about a difference in the hazard (or, in other words, in the prognosis) and not about an *effect* of the intermediate event (avoiding statements like: “2nd transplantation reduces the risk of death by 20%”). In fact we ignore whether for example only patients with some unobserved characteristic related to a good prognosis got a second transplantation, in which case the variable X would reflect the effect of an auto-selection of patients with low hazard, independently on the real effect of 2nd transplantation (it could even be the case that it actually increased their risk of dying). The case of 2nd transplant is quite striking, but, depending on the clinical issues involved, a certain caution in the interpretation is required also when the event is not a treatment: while we can reasonably think that the occurrence of an infection actually increases the risk of death, for another kind of IE it is possible that its manifestation is just a consequence of an underlying, unobserved condition, so that it doesn't have an effect *on the outcome*, but it does have an effect *on our evaluation* of the risk of that outcome (that is, on the prognosis).

⁴ The inclusion of time-dependent covariates in a Cox model is technically possible in any statistical software; the intuitive meaning is explained here.

⁵ Notice that if the model includes a time-dependent covariate inference can be based only on hazard ratios, while it is not possible to draw a standard survival curve.

The approach to the treatment of intermediate events based on the Cox model is preferred to the landmark analysis because it is the building-block of a more general approach for the simultaneous modeling of all intermediate and final events involved in the course of the disease, therefore from now on we will focus on it.

In the Cox model, the underlying hypothesis of proportional hazards implies that the presence of the characteristic “IE occurred” affects the hazard function by a constant factor. If we expect that the change in the hazard could depend also on the timing of the intermediate event (e.g. elective 2nd transplant is expected to be more effective if performed earlier), we must include another time-dependent covariate to represent the timing, which switches from 0 to the proper value at the same time as X .

As regards the time-fixed variables, the HR of gender for example expresses the risk of being, say, male with respect to the risk of being female, given that all the other characteristics are the same, including X , that is, given that the patient is transplanted, or not transplanted: the effect of gender is supposed to be the same among transplanted and among non transplanted patients. If this assumption is too restrictive, it is possible to include an interaction between gender and X (which is also a time-dependent covariate), to express a change in the effect of gender at the occurrence of second transplant.

Adding the interactions of the time-dependent variable representing the occurrence of the IE with all the other risk factors is conceptually equivalent to fit a separate model for the hazard after the intermediate event: this point will be illustrated in a while, and a further insight on “technical” issues will be given in section 3.

2.2. A comprehensive approach

It was argued in section 1 that for a comprehensive analysis of the process of the disease it is also necessary to analyze the occurrence of the intermediate event as an outcome. The following example will illustrate better this point.

Consider again (section 1) the comparison of the effect of two different treatments A and B on survival, adjusting for the occurrence of acute GvHD. Suppose that in the Cox model for survival, where aGvHD is included as a time-dependent covariate together with other factors, the HR of treatment A versus treatment B turns out to be equal to 0.5 (that is, the risk of death with treatment A is half the risk with B): we are tempted to conclude that A is better than B. But suppose that the HR of the variable representing aGvHD is 2, which means that the risk of dying is doubled at the occurrence of aGvHD. If treatment A is associated to a higher risk of occurrence of aGvHD, the conclusion that A is better than B could not hold. The analysis must therefore be completed by fitting another Cox model for the occurrence of the IE, adjusting for the other covariates.

The following step would be to summarize the results from both models. In the example above, if the HR of getting aGvHD with treatment A is 1.5, that is, the chance of aGvHD with A is 50% higher than with B, what is the global effect of treatment A on survival? It would be useful to be able to calculate e.g. the probability of dying with and without experiencing acute GvHD for both treatments, and then compare them.

The following schema illustrates the whole course of the disease represented as a process that runs through different states, from an initial state (indicated by 0) to a final state (2), with the possibility of passing through the intermediate state (1) caused by the occurrence of the IE. For example, 0 = patient at transplant, 1 = patient with acute GvHD, 2 = patient dead. Notice that transition 0→1 and transition 0→2 are

competing risks: they cannot occur both. Each transition is governed by a hazard – e.g. h_{01} is the hazard of occurrence of the IE, and h_{02} is the hazard of death without IE.

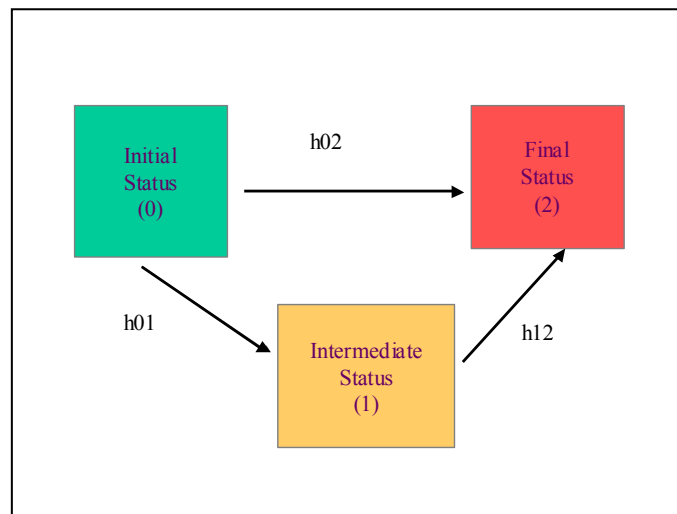


Figure 1

The idea at the basis of the multi-state approach is to fit a (Cox) model for each hazard involved, or, in other words, for each possible transition in the course of the disease:

- (1) A model for the IE = for h_{01}
- (2) A model for the final event without experiencing the IE, i.e. for h_{02}
- (3) A model for the final event given that the IE has occurred, i.e. for h_{12}

Actually, this is equivalent to the approach based on two models which was sketched above, involving a time-dependent covariates for the occurrence of the IE: instead of point (2) and (3) above, we could have:

- (2) A model for the final event, with the IE as time-dependent covariate since the latter model fits h_{02} and h_{12} simultaneously.

Therefore, with one intermediate event, we need either 2 or 3 models to describe the whole process.

This simple situation with only one intermediate event represented by Figure 1 is the most basic example of a so-called *multi-state model*. This case is also called “0-1-2” model, or “illness-death model” (being “illness” the intermediate state and “death” the final state), and it can satisfactorily represent many situations of interest in BMT research, for example:

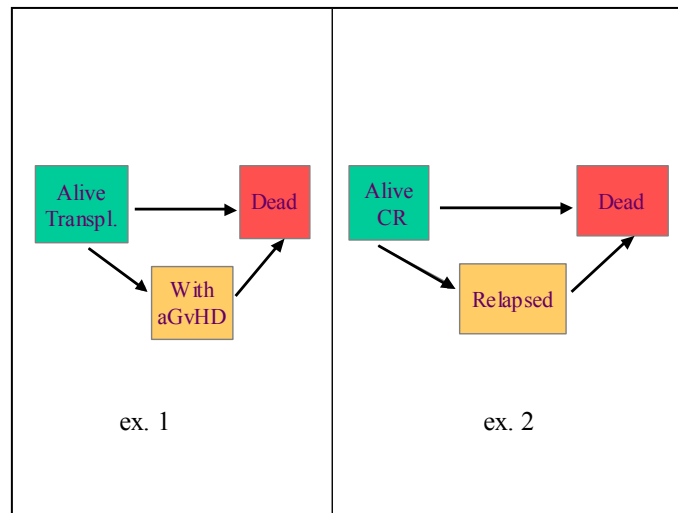


Figure 2

In this framework, objects of interest such as the probability of surviving given that the IE has occurred, or the probability of dying without experiencing the IE, are seen as probabilities of being in certain states at certain points in time conditional on the pattern of covariates known at the beginning of the process and on the previous course of the disease, which is called “history”, and includes possibly the timing of the IE. This allows in particular the evaluation of its role.

In the problem of evaluating treatments A and B in presence of aGvHD for example, we can calculate the probability of being alive given each treatment, which is the sum of $\Pr(\text{Alive without aGvHD})$ and $\Pr(\text{Alive having experienced aGvHD})$. In the double transplant study, we will consider the timing of 2nd SCT as a major issue of investigation. We will see more examples in a while, and a whole analysis for the double transplant problem will be developed in section 4.

As we have seen in section 1, in BMT research it is common to deal with more complex situations. In the presence of more than one event to be taken into account, the multi-state approach can be generalized.

It is clear that, having many events and transitions, a series of models must be fit, and the interpretation of the results can become a very hard task: it is therefore very useful, or perhaps necessary, to combine properly the results from all the models and elaborate summary probabilities, relative to the occurrence of an event, or of a series of events, within a certain time and given the history observed up to the time point when the prediction is made.

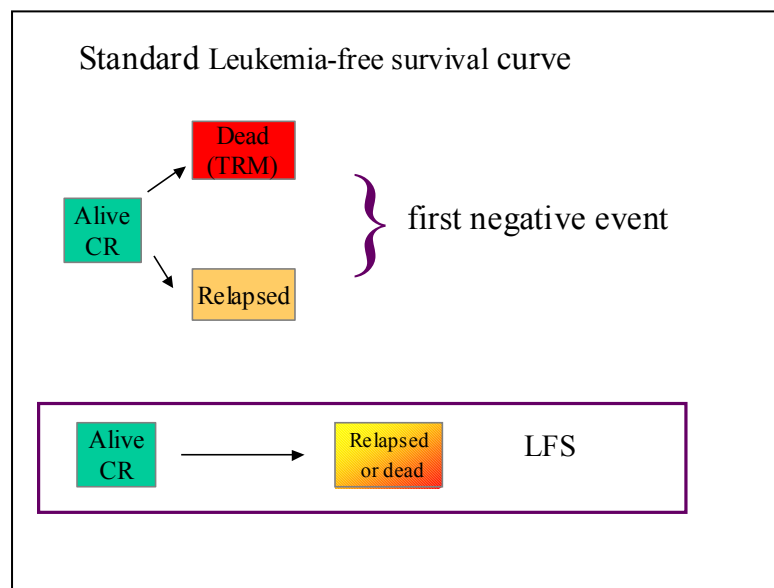
In section 2.3 a few examples of complex BMT situations will be schematized as multi-state processes, to show the flexibility of the approach and, at the same time, the potential complexity of the statistical modeling caused by purely clinical issues. The kind of summary of results needed depends anyway on the real object of investigation. One of the examples will be developed to “translate” the clinical questions to be answered (related to the role of timing of second transplant in multiple myeloma) into statistical terms. For further illustration of the kind of information that it is possible to get summarizing the results, it is very interesting for example the analysis of the recovery process after transplant, with relapse and death in remission as final events, and acute GvHD and platelet recovery as intermediate events, which is carried out in an early work of Klein and others⁽⁷⁾ focused on the interplay between

GvHD and leukemic relapse, and subsequently in a more general paper dedicated to multi-state models for BMT studies⁽¹⁰⁾.

Potentially, we can provide a complete *description* of the course of the disease after transplantation taking into account any possible issue – occurrence of minor outcomes, treatments, and the timing of these events. Such description must be interpreted as a picture of the course of the disease as it was observed in the population. In other terms, it provides an evaluation of the prognosis regarding each possible pattern of events during the disease. This kind of information is requested by patients and it is useful for clinicians to individuate high risk patients, formulate clinical hypotheses, and so on. Once more, it is important to stress that *no causal conclusion* can be drawn from an analysis on data collected retrospectively.

2.3. Some examples and further remarks

Consider the problem of the calculation of the total **leukemia-free survival probability** introduced in section 1; the starting schema is on the right hand side of Figure 2 (ex. 2). The traditional LFS method does not take into account the course of the disease after Relapse:



The multi-state approach allows instead to “explode” the transition from “Relapsed” to “Dead”, including the possibility of returning to a Complete Remission state after relapse:

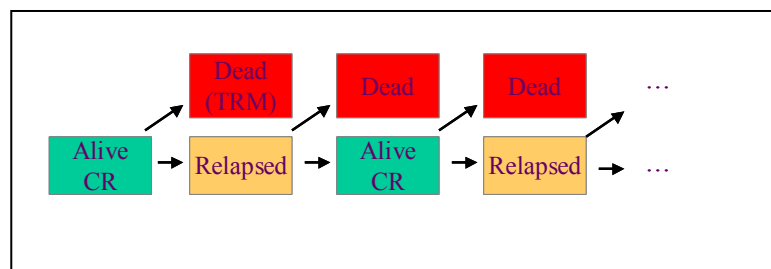


Figure 3

This part of the story of the disease can be further decomposed considering occurrence and effectiveness of Donor Lymphocyte Infusion:

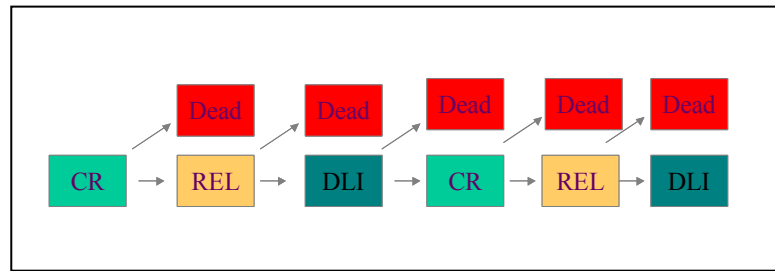


Figure 4

We can estimate the so-called “current” leukemia-free survival probability considering that the patient is alive relapse-free if neither relapse nor transplant-related death have occurred (traditional LFS probability) or if a series of events has occurred: {first relapse, administration of DLI, second CR achieved, and no further transition} or {the same sequence of events followed by second relapse, second DLI, etc}.

Clinical arguments should give hints for building the models for the transitions involved in the process. For example, the state “Dead” can either be considered as being the same in each phase of the process, or distinguishing between “Dead in Relapse” and “Dead in Remission”; this depends on whether the distinction is relevant, for the nature of the disease process and the purpose of the study, and it has a series of implications in choosing a suitable statistical model⁶.

Another interesting situation is when we consider the manifestations of **graft-versus-host disease in a survival study**. It is quite difficult to define a good schema, due to the definition of GvHD itself. The schema in figure 5 is reasonable, but not completely satisfying:

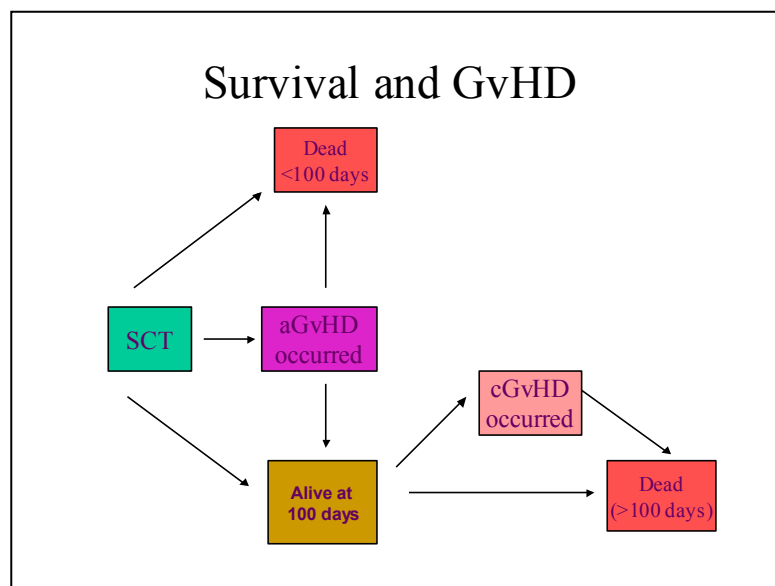


Figure 5

⁶ For a further insight on the estimation of current LFS based on multi-state modeling (and discussion of possible alternatives) see for example Klein et al. on *Statistics in Medicine*⁽¹¹⁾; for an application, see the paper on the *British Journal of Haematology*⁽⁸⁾.

In fact, the hazard of transition from state “aGvHD occurred” to state “alive at 100 days” is infinite at time 100 days, as well as the hazard from that state to “cGvHD occurred” in case of persistence of graft-versus-host disease beyond 100 days. This is not the only clinical issue that poses a series of questions regarding a proper statistical modeling.

For example, on which time-scale should the occurrence of chronic GvHD be measured? It seems reasonable to start the clock at 100 days instead that at transplantation, due to the definition of cGvHD. In a way, the process is split up in two: one part regarding the first 100 days, the other regarding the rest of the story of the disease. At 100 days, the information on occurrence and timing of acute GvHD is given as fixed covariates. This is an advantage because there can be biological or clinical reasons to believe that cGvHD is influenced by aGvHD, and anyway the latter must be taken into account since, by definition, when its duration goes beyond 100 days we have the chronic form. We could also wonder if it is necessary to consider the time elapsed between diagnosis and acute GvHD as a covariate.

The relations between acute GvHD and chronic GvHD and their timing could be influential also in the model for the final outcome. For example: does the effect of cGvHD depend on whether there was aGvHD before? Do the timing of aGvHD and of cGvHD influence survival?

Planning and carrying out an analysis in the framework multi-state modeling is not simple. First of all, it is necessary to give a multi-state structure to a clinical problem, which means taking into account all and only the relevant events. Moreover, in order to build the Cox models for each transition, it is important to individuate which elements of the previous history of the disease could have an influence on the outcome. As in the GvHD example, it could be complicated to choose suitable time-scales and to take into account the timing of the intermediate events. Planning the analysis requires a deep understanding of the problems in BMT studies as well as of the technical or methodological problems involved, therefore a close collaboration of clinicians and statisticians is particularly advisable.

Apart from the difficulties in building the multi-state structure and the models, there are problems also at the practical level. While the use of the Cox model to estimate the hazard for each possible transition is quite simple from a “technical” point of view (besides the difficulties mentioned, and given that there are enough data for the estimation – section 3), unfortunately the calculation of summary probabilities is not straightforward, and it is not yet available in the common statistical software, therefore the statistician has to implement his/her own software. For the simple case of a 0-1-2 model (only one IE) the R functions used for the application of section 4 are available upon request (although they may be of help only for expert users of R). In section 3.3 there are also indications for the estimation in more complicated models.

Notice anyway that often the simple 0-1-2 model is sufficient to analyze properly several clinical problems. For example, the study of survival taking into account GvHD (Figure 5) could be limited to the course of the disease after 100 days, which includes only 1 intermediate event. It is in fact sufficient to evaluate a series of clinical issues, with the only limitation of being applicable to patients surviving at

least 100 days; the process generating acute GvHD and its possible outcomes within 100 days can be analyzed separately, again using a 0-1-2 model:

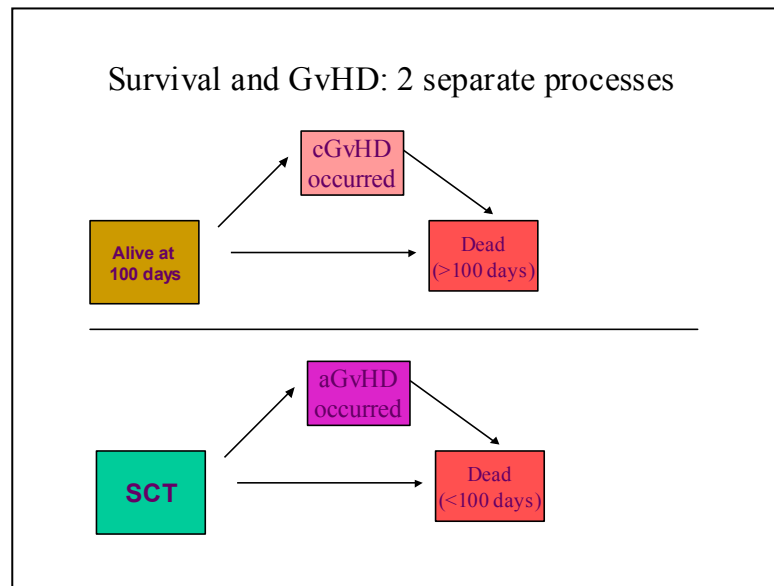


Figure 6

As a second example, consider the complex study of **tandem transplantation** in multiple myeloma. Until now, we have considered as an intermediate event the elective 2nd transplantation, but its definition (second transplant given in remission) implies that we must take into account relapse as another competing risk (besides death), and other possible patterns for the course of the disease:

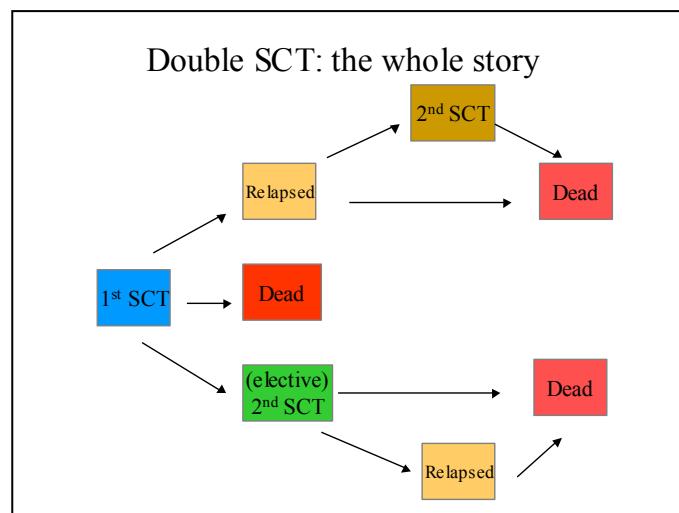


Figure 7

At first transplant, there are 3 possible outcomes: (transplant-related) death, relapse and elective 2nd transplantation. We are interested in what happens after the latter: the patient can relapse and/or die. Furthermore, a second transplant can also be given after the first relapse – and the final outcome is again death.

This schema is based on 9 transitions. Even if the use of Cox models with time-dependent covariates can reduce the number of models to be estimated from 9 to 3 (section 3), such an analysis could be highly complicated when we try to include all possible effects of occurrence and timing of each IE on the following events.

A very efficient simplification of the problem focused on the role of the elective second transplant is the following, where the final outcome is the relapse-free survival, and the only IE considered is 2nd SCT:

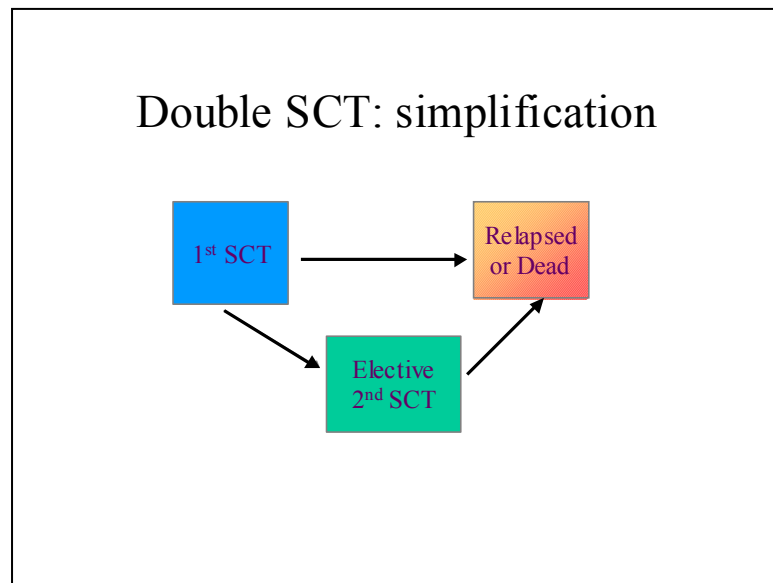


Figure 8

This approach to the study of double transplantation will be developed in section 4: the results of the Cox models will be showed and then some examples will be given on how to summarize those results to produce interesting graphics to enlighten the role of second transplant and its timing. Readers not interested in statistical modeling can skip the next section and proceed with the application.

3 Statistical analysis

Multi-state models appeared in the statistical literature in the first 1990s^(4,7). For a complete overview, which is far beyond the scope of this document, readers can refer for example to the textbook by Andersen, Borgan, Gill and Keiding⁽¹⁾, where the mathematical background is very solid, or the one by Klein and Moeschberger⁽⁹⁾, where the framework is more applied. Hougaard (1999)⁽⁵⁾ provides another (rather theoretical) review. Good introductory papers are for example Andersen et al. (2000)⁽²⁾ and many articles by John Klein and others, especially focused on BMT problems^(6,10,11). Volume 11, 2, of *Statistical Methods in Medical Research* (2002) is fully dedicated to multi-state models; in particular, the paper by Andersen and Keiding⁽³⁾ is an useful introduction to the alternatives to the multi-state models as they will be presented here (e.g. not based on the Cox proportional hazards model).

This section will sketch the main elements of the multi-state approach, focusing on those to take into account at the moment of planning an analysis and those involved in the presentation of the results. Alternatives for modeling will be illustrated in some detail for those not very familiar with the Cox model. Formulas for the estimation of summary probabilities will be provided for the simple “illness-death” model, represented in Figure 1, but they can be generalized.

3.1. Building the model

As we have seen in sections 2.2 and 2.3, a whole process involving k transitions determined by the occurrence of p events must in principle be described fitting k (Cox) models, one for each transition. A proper time scale must be chosen for each transition: we can adopt the calendar time scale, that measures time starting at the origin of the process, or we can reset the clock back at the occurrence of the last intermediate event.

The use of time-dependent covariates in the Cox model can reduce the number of models to estimate up to a minimum of p . The idea of this alternative approach (developed in particular by Klein) is in fact to model not the hazard functions of the transitions, but the hazard of occurrence of each possible event in the process, including the occurrence of other events prior to the one being modeled as a time-dependent covariates. As a result, a model for each transition is provided.

We can illustrate these issues in the simple framework of the “illness-death” model (Figure 1).

In a “illness-death” model we have $k=3$ transitions⁷ (01, 02 and 12) and $h=2$ events: the intermediate one (IE) and the final one (FE), say relapse and death, respectively. Indicate with t_{IE} the time to the intermediate event, and with t_{FE} the time to the final event; δ_{IE} and δ_{FE} are the indicators of occurrence of IE and FE respectively (=1 if the event occurred, =0 otherwise).

As a **first approach**, we can estimate separately the 3 hazard functions h_{01} , h_{02} and h_{12} , in the following way:

- h_{01} : response = (t_{IE}, δ_{IE}) ; if the FE occurs without IE, the observation is censored at t_{FE} , that is, the response is $(t_{FE}, 0)$.
- h_{02} : response = (t_{FE}, δ_{FE}) ; if the IE occurs before the FE, the observation is censored at t_{IE} , regardless of possible occurrence of FE afterwards;
- h_{12} : the model is fit restricting to the individuals who actually experienced the IE during the course of the disease. If we choose the calendar time scale, we have to fit a Cox model with left-truncation⁸ at the time to intermediate event: response = $(t_{IE}, t_{FE}, \delta_{FE})$. If we choose the clock-back approach, the response is $((t_{FE} - t_{IE}), \delta_{FE})$. In any case, the time of entry in state 1 (t_{IE}) can be included as a (fixed) covariate.

As a **second (alternative) approach**, we can estimate:

- One model for the IE: response = (t_{IE}, δ_{IE}) . This is the same as h_{01} above.
- One model for the FE with IE as time-dependent covariate; response = (t_{FE}, δ_{FE}) :

$$h^{FE}(t) = h_0^{FE}(t) \exp(\beta_{fixed}^{FE} x_{fixed} + \beta_{ie}^{FE} IE(t))$$

This model provides estimates of the hazard functions for transitions 02 and 12:

$$h_{02}(t) = h_0^{FE}(t) \exp(\beta_{fixed}^{FE} x_{fixed})$$

$$h_{12}(t) = h_0^{FE}(t) \exp(\beta_{fixed}^{FE} x_{fixed} + \beta_{ie}^{FE})$$

Time to IE can be added also as time-dependent covariate (following a certain functional form), as well as the interactions of $IE(t)$ with the time-fixed covariates, to express the change of their effects after IE:

$$h^{FE}(t) = h_0^{FE}(t) \exp(\beta_{fixed}^{FE} x_{fixed} + \beta_{ie}^{FE} IE(t) + \beta_{timeIE}^{FE} f(time_{IE}(t)) + \beta_{IE*x}^{FE} IE(t)x_{fixed})$$

⁷ Notation : for example, with 01 we indicate the transition 0→1, that is, from state 0 to state 1.

⁸ Left truncation, also called « delayed entry », corresponds to a conditioning. It is implemented in the framework of counting processes in many statistical packages, including SAS and S-Plus (or R), but not for example in SPSS.

Notice that the difference between this approach and the previous one is that the baseline hazard for transitions 02 and 12 is forced to be the same; in other words, the hazards h_{02} and h_{12} differ for a multiplicative effect $\exp(\beta_{IE})$, or $\exp(\beta_{ie}^{FE} + \beta_{timeIE}^{FE} + \beta_{IE*x}^{FE} x_{fixed})$ if time to IE and interactions of IE(t) with the fixed risk factors are present: they are proportional.

The approach based on the inclusion of (some of) the intermediate events as time-dependent covariates is of course particularly advantageous with more complex problems, such as the calculation of the current leukemia-free survival (figures 3 and 4), or the “complete” analysis of the double transplant problem of figure 7. In the literature, in the framework of BMT studies, they analyzed for example the recovery process after transplant^(7,10) or the role of acute and chronic GvHD⁽⁶⁾, in both cases considering two intermediate events and two final events (relapse and death in remission), in a schema based on 6 states and 12 transitions.

In the double transplant problem as it was presented here, we have 9 transitions, but basically only 3 events, relapse, second transplantation and death, so in principle 3 models are sufficient to describe the whole process. A time-dependent covariate REL(t) for relapse is present in $h^{TRX}(t)$ and $h^{DEATH}(t)$, and another time-dependent covariate TRX(t) for the occurrence of 2nd SCT is present in $h^{REL}(t)$ and $h^{DEATH}(t)$. Then for example the transition to death for patients relapsed after elective 2nd transplantation is described by the hazard function obtained from the model for death with the two time-dependent covariates “switched on”:

$$h(t) = h_0^{DEATH}(t) \exp(\beta_{fixed}^{DEATH} x_{fixed} + \beta_{trx}^{DEATH} + \beta_{rel}^{DEATH})$$

An interaction between TRX(t) and REL(t) could take into account the change in the effects of one event when the other also occurs.

Anyway, in this study it is probably important to take into account also the time sequence of the occurrence of the events or, in other words, to distinguish between elective 2nd transplantation (given before first relapse) and transplant given after first relapse. In fact, without such distinction the hazard function above would also represent the risk of death of patients transplanted for the second time after relapse.

We could therefore consider the two types of transplantation as distinct events, represented respectively by covariates ELECT(t) and TRX(t). The transition to death of patients relapsed after elective 2nd transplant would therefore be described by:

$$h(t) = h_0^{DEATH}(t) \exp(\beta_{fixed}^{DEATH} x_{fixed} + \beta_{elect}^{DEATH} + \beta_{rel}^{DEATH})$$

while the hazard for the transition to death of patients transplanted after the first relapse would be:

$$h(t) = h_0^{DEATH}(t) \exp(\beta_{fixed}^{DEATH} x_{fixed} + \beta_{trx}^{DEATH} + \beta_{rel}^{DEATH})$$

The risk of relapse for patients transplanted for the second time in remission and still in remission would be:

$$h(t) = h_0^{REL}(t) \exp(\beta_{fixed}^{REL} x_{fixed} + \beta_{ELECT}^{REL})$$

and for patients not transplanted:

$$h(t) = h_0^{REL}(t) \exp(\beta_{fixed}^{REL} x_{fixed})$$

Also with this approach we would have a reduction of complexity: the number of models to be estimated would be 4 instead of 9.

Using time-dependent covariates to represent the occurrence of intermediate events allows to build parsimonious but still flexible models. Of course, this method relies on the assumption that different transition hazards are proportional⁹. It is advisable (or in some case necessary) when there are too few cases to model separately each transition. Compared to the other approach, it is easier to evaluate globally the results looking only at the estimated models since the “effects” of intermediate events on the hazard of other outcomes are expressed by single regression coefficients (or hazard ratios).

The approach based on the separate estimation of each transition hazard is on the other hand completely flexible, and in particular it is easier to introduce for each transition all and only the elements that we want to take into account (covariates, timing of events, etc).

As regards the choice of the **time scale**, there are some (rather theoretical) elements – very briefly summarized here – and some more practical suggestions to take into account.

Consider an illness-death model. The hazard of transition to death of diseased individuals (h_{12}) can depend both on t , the time since the origin, and on d , the time spent in state 1: being T_1 the time of entry in 1 (which is a random variable), $d=t-T_1$ (d is the clock-back time scale). Therefore, in general we have a hazard function with two arguments, $h_{12}(t,d)$.

Since using the Cox model it is possible to have only one baseline time scale, the other one must be introduced as a (time-dependent) covariate¹⁰. The baseline time-scale should be the one whose effect is not of interest (since there is no HR associated to it) or one whose effect is too complicated to be modeled as a time-dependent covariate. Moreover, it is also recommended to base the choice on biological/clinical arguments. Often the clock-back approach is more reasonable (e.g. for IE like relapse or 2nd SCT it is preferred to measure time since the event occurred) but it is also somewhat more difficult to globally interpret the results afterwards. Finally, notice that a model for $h_{12}(t)$ (conditioned on being in state 1) or for $h_{12}(d)$ can also include T_1 as a *fixed* covariate – which is another way to introduce a dependence on another time scale.

3.2. Summarizing the results

As it was said in section 2, the interpretation of the results is much easier and powerful if we can summarize properly the results from all the models fit to describe the process.

Indicating with $X(t)$ the state occupied at time t , we need to calculate the **transition probabilities** $\Pr(X(t)=j | X(s)=i, \mathcal{H}(s))$ that is, the probability of being in state j at time t given that the patient was in state i at time s and all other necessary information (the “history” of the disease up to time s , $\mathcal{H}(s)$ - including the fixed covariates) ($s < t$). In

⁹ If this assumption does not fit the data, it is possible to use the time-dependent covariate as a stratification variable ; refer for example to Klein and Moeschberger (1994).

¹⁰ If the function depends only on the time on study ($h_{12}(t,d)=h_{12}(t)$) the process is said to be *Markovian*, and it is somewhat easier to calculate summary results in this framework (cf Andersen et al 2000³). Using the Cox model, the Markovian assumption can be verified introducing in $h_{12}(t)$ a time-dependent variable representing the time spent in state 1. If this covariate is significant, the assumption does not hold (sometimes in the literature the model is denominated “semi-Markovian”). An example of a non Markovian model is the case with $h_{12}(t,d)=h_{12}(d)$, that is, the clock-back approach, were the dependence on t is eliminated.

other words, this probability represents the prognosis of being in a certain state (j) at a certain point in time given information on the present state and the pattern followed by the patient up to that state.

In a 0-1-2 multi-state model (Figure 1) the transition probabilities (suppressing in the notation the dependence on the fixed covariates) are:

$$- p_{00} = \Pr(X(t)=0 \mid X(s)=0) = \prod_{s \leq u \leq t} (1 - h_{01}(u) - h_{02}(u))$$

(F.Ia 1)

$$- p_{11} = \Pr(X(t)=1 \mid X(s)=1, \text{ entry in 1 (IE) at time } z) = \prod_{s \leq u \leq t} (1 - h_{12}(u; z)) \quad (z \leq s)$$

(F.Ia 2)

$$- p_{12} = 1 - p_{11}$$

$$- p_{01} = \Pr(X(t)=1 \mid X(s)=0) = \sum_{s \leq u \leq t} P_{00}(s, u^-) h_{01}(u) P_{11}(u^+, t; u)$$

(F.Ia 3)

- $p_{02} = \Pr(X(t)=2 \mid X(s)=0)$; it can be calculated as $1 - p_{00} - p_{01}$, and also as the sum of the following two, which regard the transition probability with and without experiencing the IE respectively:

$$- p_{02ie} = \Pr(X(t)=2, \text{ IE=yes} \mid X(s)=0) = \sum_{s \leq u \leq t} P_{00}(s, u^-) h_{01}(u) (1 - P_{11}(u^+, t; u))$$

(F.Ia 4)

$$- p_{02nie} = \Pr(X(t)=2, \text{ IE=no} \mid X(s)=0) = \sum_{s \leq u \leq t} P_{00}(s, u) h_{02}(u)$$

(F.Ia 5)

For example, if 0=transplant, 1=relapse and 2=death:

- p_{00} is the probability of being alive relapse-free from time s to time t after transplant;
- p_{11} is the probability of being alive (in relapse)¹¹ from time s to time t given that the patient had a relapse at time $z \leq s$;
- p_{12} is the probability of being dead at time t for a patient who had a relapse at time z and is still alive (in relapse) at time s ;
- p_{01} is the probability of being alive in relapse at time t given that the patient is alive relapse-free at time s ;
- p_{02} is the probability of being dead at time t for a patient who is alive relapse-free at time s ; for the same patient, p_{02nie} is the probability of being dead without relapse (at time t), and p_{02ie} is the probability of being dead with relapse (at time t).

Of course, in a model with more than one intermediate event, there are many transition probabilities and they can include as “history” all kind of information regarding the previous course of the disease (occurrence and timing of several IE). Their mathematical expressions are also more complicated. Examples can be found in the literature^(3,6,7,10,11).

¹¹ In this example, we don't consider explicitly the possibility of being in second remission ; « in relapse » could be intended as « after first relapse ».

The transition probabilities $\Pr(X(t)=j \mid X(s)=i, \mathcal{H}(s))$ can be calculated varying the elements t , s and $\mathcal{H}(\cdot)$ to get information on the course of the disease under different perspectives.

→ With fixed s and $\mathcal{H}(\cdot)$ and varying t we have curves similar to the standard “survival” ones, that is, we predict an outcome at several future points in time on the basis of the present knowledge on the patient and his/her history. E.g. we calculate the survival probability in time given that the patient is alive 3 months after transplant with platelet recovered and an episode of acute GvHD, and given all his/her characteristics known at transplant (age, gender, conditioning regimen, etc). This is the point of view of the prognosis.

→ With fixed t and $\mathcal{H}(\cdot)$ and varying s we evaluate how the prognosis of a certain event changes as the time of prediction changes. For example we evaluate the probability of death within 5 years after transplant at 6 months, at 1 year, at 2 years etc. after transplant, for the same kind of patient and the same history. This is called “the learning effect”, since it provides information on how much the prognosis is affected by knowing more and more about the patient’s history.

→ With fixed t and s and varying the history $\mathcal{H}(\cdot)$ we have an insight on the effect of the events occurred and their timing in particular. For example we evaluate the probability of death within 5 years predicted 1 year after transplant first for a patient without chronic GvHD, then for a patient with chronic GvHD occurred at 4 months, then for a patient with chronic GvHD occurred at 12 months, and so on. This kind of information is particularly useful if the occurrence of the intermediate event is controlled by the physician (e.g. it is a treatment like second transplantation), but it can be useful anyway to suggest therapeutical strategies to prevent or induce the occurrence of events as GvHD.

3.3. Estimation of summary probabilities

The estimation of the summary probabilities in a multi-state approach simply requires the calculation of the formulas (reported in section 3 for the 0-1-2 model), replacing the hazard functions for each transition $h_{ij}(t)$ by their estimates. From the output of the Cox model:

$$h_{ij}(t) = h_{0ij}(t) \exp(\beta_{ij}x)$$

we obtain an estimated regression coefficient $\hat{\beta}_{ij}$ and a non-parametric estimate for the cumulative baseline hazard $\hat{H}_{0ij}(t)$ for each event-time¹² from which it is possible to derive the instantaneous baseline hazard $\hat{h}_{0ij}(t)$.

When the model is more complex than the 0-1-2 one, it could be difficult to write down the correct formulas and to implement the calculation. An alternative which is much simpler to implement, even for very complex schemas, is to estimate the transition probabilities simulating the outcomes for N (in the order of some thousands) patients. For the case of a 0-1-2 model, it was checked that this method produces estimates very similar to the ones obtained through direct calculation (already with $N=2000$; of course, the more complex is the schema, the higher the number of life-histories to be simulated).

¹² Depending on the software being used, the cumulative baseline hazard could not be provided if time-dependent covariates are included in the model.

If clock-back approach:

If $C < \min\{T01, T02\} \Rightarrow T_1 = \text{NA}, T = C, St = 0$

If $T02 < \min\{T01, C\} \Rightarrow T_1 = \text{NA}, T = T02, St = 2$

If $T01 < \min\{T02, C\} \Rightarrow \text{simulate } T12 \rightarrow \text{If } C < T01 + T12 \Rightarrow T_1 = T01, T = C, St = 1$
else $\Rightarrow T_1 = T01, T = T01 + T12, St = 2$

The procedure is repeated N times, to simulate life-histories for a population of N individuals with the same characteristics. Then, for the calculation of $\Pr(X(t)=j \mid X(s)=i, \mathcal{H}(s))$ it is sufficient to count how many patients were in state i at time s and, among those, how many were in state j at time t , and divide the second number over the first.

4 A study case: double transplantation in multiple myeloma

The study on the role of elective second transplant and its timing in multiple myeloma has been used throughout this work to illustrate methodological problems and possible approaches. We will now see how to develop an analysis based on the simplified schema in figure 8 and in particular how to elaborate useful summaries of the results. The transition probabilities introduced in section 3.2 are mentioned to explain what kind of summary is presented to answer to clinical questions, anyway in order to understand the potentials of the multi-state approach it is not necessary to know all the technical background.

The data used for this application were selected from the Myeloma Registry of the EBMT, and refer to 7452 patients transplanted for multiple myeloma, of whom 1725 had an elective second SCT. They have already been used to analyze relapse and treatment-related mortality (separately) in relation to elective 2nd transplant, and to analyze survival after first relapse⁽¹³⁾. The application proposed here is only for illustrative purposes; for an understanding of the clinical problem it is recommended to refer to the paper.

The schema chosen to handle the problem (Figure 8) implies that the major outcome considered for the analysis is the relapse-free survival, and that elective 2nd transplantation is the only intermediate event taken into account. The states of the process are therefore: 0=first transplantation, 1=elective second transplantation, 2=relapse or non-relapse death.

Notice that using this schema we will not consider the story of the disease after first relapse. In other words, we will evaluate the role of elective 2nd transplant in preventing/causing either relapse or non-relapse mortality, but we will not be able to assess if it has any influence on the prognosis in case a patient relapses. Therefore the comparison between the two treatment strategies, elective transplantation and wait-and-see procedure, is not exhaustive. This is the price we pay for being able to restrict to the simplest schema 0-1-2 instead of sticking to the complete one (Figure 7).

COX MODELS

Choosing the approach based on time-dependent covariates, we estimate two Cox models, one for time to elective SCT and one for time to relapse or death without

relapse including elective 2nd SCT and its timing as time dependent covariates (called TRX and T.TRX respectively). The analysis is adjusted for a series of risk factors among which age (with a linear effect) and stage (a categorical covariate with levels I, II and III) (the other risk factors are not reported).

Model for elective 2nd SCT → Transition 01:

variable	β	HR	p-value
Age	-0.01884	0.98	4.2e-11
Stage2	0.26185	1.30	5.0e-02
Stage3	0.48823	1.63	3.2e-05
...

Table 1

Model for Rel/Non-rel Death → Transitions 02 and 12:

variable	β	HR	p-value
TRX	-0.48411	0.62	3.6e-14 ←
T.TRX	0.02385	1.02	1.7e-04 ←
Age	0.01037	1.01	1.6e-05
Stage2	-0.02528	0.97	8.0e-01
Stage3	0.28630	1.33	7.2e-04
...

Table 2

Interpretation: after taking into account the pattern of risk factors included in the analysis, at the occurrence of 2nd SCT the hazard of Relapse or Non-Relapse Death was reduced by 40% (HR=0.62). Longer time to 2nd SCT was associated anyway to an increased HR (by 2% per month: HR=1.02). This means that the advantage for Relapse-Free Survival observed in patients at the moment they got the 2nd transplantation was reduced when the treatment occurred late after the first transplant, and in particular it was lost when 2nd SCT was performed later than (about) 24 months¹⁴.

SUMMARY OF THE RESULTS

The results are illustrated for a single representative patient characterized by median (for continuous variables) or most common (for discrete variables) risk factors (e.g. age equal to the median (54), stage III, etc). Of course for each pattern of risk factors of interest, it is possible to draw curves like the following, and therefore it is possible to evaluate the results for different kinds of patients, and make comparisons.

As a first summary, we can produce the **Relapse-free** survival curve corresponding to the estimated Cox model by representing the sum of two transition probabilities $p00 = \Pr(X(t)=0|X(0)=0)$ and $p01 = \Pr(X(t)=1|X(0)=0)$ for t running from 0 to the maximum follow-up time. In fact, in this context states 0 and 1 are both “alive relapse-free” states, therefore the curve represents the estimated proportion of patients still alive and relapse-free at each point in time.

¹⁴ $0.62 \cdot 1.02^{24} \cong 1$

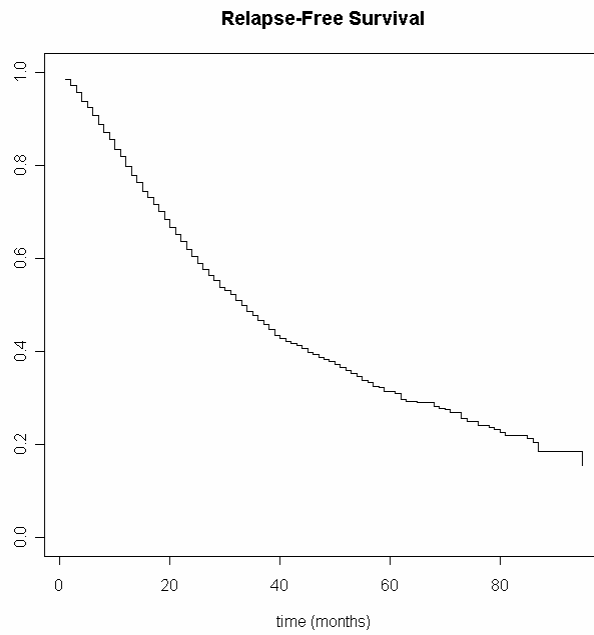


Figure 9

The major goal of this analysis is anyway to evaluate the **role of elective second transplantation and its timing**. A first way to focus on the latter issue is to consider whether among patients who got a 2nd SCT in remission there was any advantage if the transplantation was performed early.

We may represent the change in the prognosis of being alive relapse-free e.g. at 5 years (60 months) with respect to the time the patient gets the second transplant; the prediction is made, say, 1 and 2 years after first transplant.

To this purpose we may calculate transition probabilities of the kind p11 changing the information regarding the timing of 2nd transplant:

- First graph: $\Pr(X(60)=1 \mid X(12)=1, \text{ entry in 1 (2}^{\text{nd}} \text{ SCT) at time } z)$ for z running from 0 to 12;
- Second graph: $\Pr(X(60)=1 \mid X(24)=1, \text{ entry in 1 (2}^{\text{nd}} \text{ SCT) at time } z)$ for z running from 0 to 24:

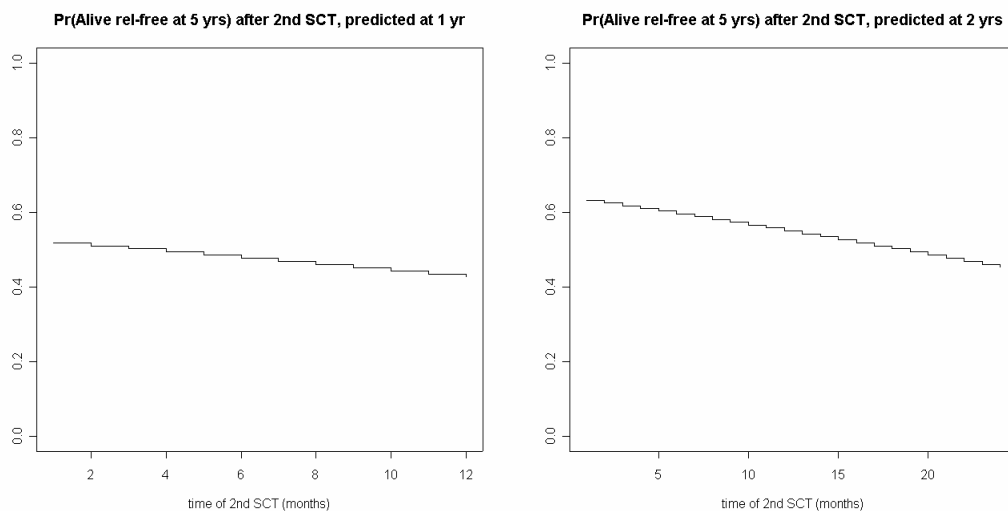


Figure 10

The first plot suggests that for patients transplanted during the 1st year and still alive at 12 months, the prognosis is not much affected by the time to second transplantation. In the second plot, relative to patients alive at 2 years with elective 2nd SCT the advantage associated to early transplantation is more marked: the relapse-free survival probability at 5 years reduces from 0.63 to 0.45 if the elective 2nd transplantation is made at month 24 instead of month 1.

To investigate further the role of time of 2nd transplant we must not restrict anyway to patients already transplanted. After the first transplantation, clinicians have to choose between the elective second transplant and the wait-and-see strategy, that is, to reserve the possibility of performing another transplant for the case that the patient relapses. It is therefore useful to compare the prognosis of patients who get transplantation at a certain moment in time with the prognosis of those patients who are not (yet) transplanted at that time. We can also consider the prognosis of patients already transplanted.

The next graph represents the predicted relapse-free survival probability at 5 years in different situations (“histories”), with respect to the time of prediction (on the horizontal axis, time s). The different situations are:

- patients without 2nd SCT at the time of prediction, who could be transplanted later on (red curve); this is $\Pr(X(60)=0 \text{ or } 1 \mid X(s)=0)$ that is $p_{00}+p_{01}$;
- patients who get a 2nd SCT at the time of prediction (black curve): $\Pr(X(60)=1 \mid X(s)=1, 2^{\text{nd}} \text{ SCT at time } s)$ (p_{11} with $z=s$ =abscissa);
- patients who got 2nd SCT at 6 months (green curve): $\Pr(X(60)=1 \mid X(s)=1, 2^{\text{nd}} \text{ SCT at time } 6)$ (the prediction starts therefore at month 6; p_{11} with $z=6$);
- patients who got 2nd SCT at 12 months (blue curve): $\Pr(X(60)=1 \mid X(s)=1, 2^{\text{nd}} \text{ SCT at time } 12)$ (the prediction starts therefore at month 12; p_{11} with $z=12$);

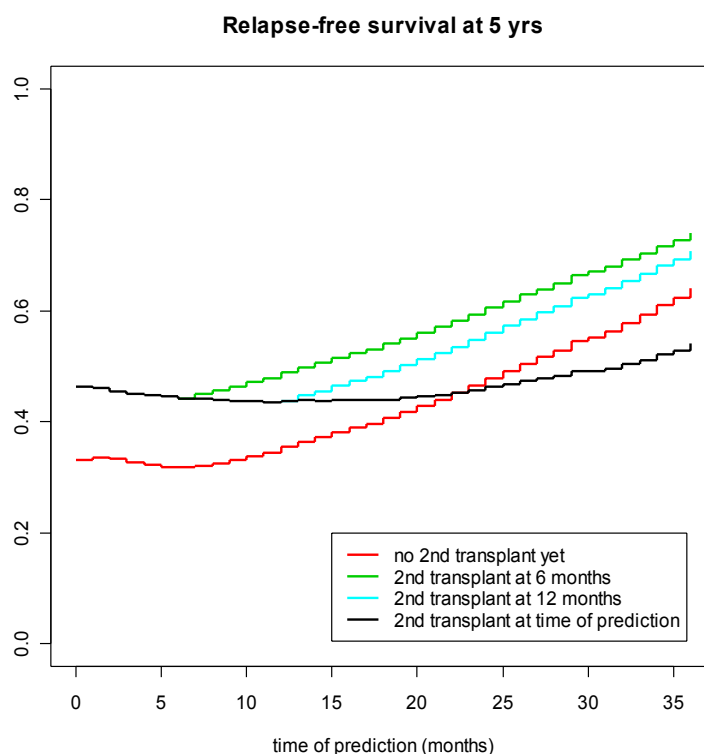


Figure 11

For example, a patient not yet transplanted (red curve) at 10 months has a predicted probability of being alive relapse-free at 5 years around 30%; if he/she is still alive relapse-free and not (yet) transplanted at month 36, the probability is about 60%. If a patient is transplanted at 6 months (green curve), the prognosis at month 36 gives a 70% chance of being alive relapse-free 2 years later. And so on.

The comparison between the red curve and the black curve shows the difference in the prognosis between a patient not transplanted (yet) and a patients who gets 2nd transplantation at that moment in time. There is an advantage for the latter one, but only within the first 2 years, while later on his/her prognosis is even worse.

The green and the blue curves represent the prognosis of patient already transplanted (at 6 months and at 12 months respectively). They show how early elective 2nd transplantation is effective in preventing relapse or non-relapse death.

Looking at this graph, we could suggest that there is an advantage in relapse-free survival with second transplantation performed before 2 years – actually, “the sooner, the better” – while it should be avoided afterwards. Notice that this is the same conclusion reached looking at the results of the regression model (Table 2), anyway this representation of the results is more powerful and more complete, and it allows to provide estimates of the relapse-free survival probability.

As it was already stressed before, it is important to remind that this is an *hypothesis*, not a conclusion: since the allocation of 2nd SCT was not randomized with equal probability, we cannot exclude that there was a selection of patients with better prognosis to have an early second transplantation, while for patients with poor prognosis it was decided not to proceed to a second transplant, or to delay it. Late

transplantations could have been performed for patients about to relapse or die, and it could have been effective, in the sense that without late transplantation the relapse-free survival would have been even worse.

This example shows how the results from the Cox models can be combined to illustrate several interesting aspects of the course of the disease in relation to the occurrence and timing of different events. The final product of the multi-state approach is a rich, comprehensive picture of all the possible developments of the process of the disease as they were observed in the population. Being a description, this kind of information is mostly useful for prediction purposes, or for proposing hypotheses, while in no way this approach can actually detect causal relations between events from observational, retrospective studies; this should always be stressed when reporting results, warning against possible sources of bias and misinterpretation.

Suggested Readings

- 1 Andersen PK, Borgan O, Gill RD, Keiding N (1993): *Statistical models based on counting processes*, Springer-Verlag New York
- 2 Andersen PK, Esbjerg S, Sorensen T (2000): "Multi-state models for bleeding episodes and mortality in liver cirrhosis", *Stat. in Medicine*, 19, 587-599
- 3 Andersen PK, Keiding N (2002): "Multi-state models for event history analysis", *Stat. Meth. Med. Research*, 11, 2, 91-116
- 4 Dabrowska DM, Sun G, Horowitz MM (1994): "Cox regression in a Markov renewal model: an application to the analysis of bone marrow transplant data", *J. Am. Stat. Ass.*, 89, 867-877
- 5 Hougaard P (1999): "Multi-state models: a review", *Lifetime Data An.*, 5, 239-264
- 6 Keiding N, Klein JP, Horowitz MM (2001): "Multi-state models and outcome prediction in bone marrow transplantation", *Stat. in Medicine*, 20, 12, 1871-1885
- 7 Klein JP, Keiding N, Copelan EA (1994): "Plotting summary predictions in multistate survival models: probabilities of relapse and death in remission for bone marrow transplantation patients", *Stat. in Medicine*, 13, 2315-2332
- 8 Klein JP, Keiding N, Shu Y, Szydlo RM, Goldman JM (2000): "Summary curves for patients transplanted for chronic myeloid leukaemia salvaged by a donor lymphocyte infusion: the current leukemia-free survival curve", *Brit. J. Haem.*, 109, 148-152
- 9 Klein, J.P. - Moeschberger, M.L.(1994): *Survival analysis: technics for censored and truncated data*, Springer, New York
- 10 Klein JP, Shu Y (2002): "Multi-state models for bone marrow transplantation studies", *Stat. Meth. Med. Research*, 11, 2, 117-140
- 11 Klein JP, Szydlo RM, Craddock C, Goldman JM (2000): "Estimation of current leukemia-free survival following donor lymphocyte infusion therapy for patients with leukemia who relapse after allografting: application of a multistate model", *Stat. in Medicine*, 19, 3005-3016
- 12 Labopin M, Iacobelli S (2003): *Statistical guidelines for EBMT* (available on the EBMT web site)
- 13 Morris C, Iacobelli S, Brand R, Bjorkstrand B, Drake M, Niederwieser D, Gahrton G (2004): "Benefit and timing of second transplantations in multiple myeloma: clinical findings and methodological limitations in a European Group for Blood and Marrow Transplantation registry study", *J Clin Oncol.*, 22(9),1674-81