Published in final edited form as:

Surg Oncol. 2010 June ; 19(2): 52–54. doi:10.1016/j.suronc.2010.03.003.

Concepts in cancer survival analysis: research questions, data, and models

Ewout W. Steyerberg and

Dept of Public Health, Erasmus MC, Rotterdam, The Netherlands

Thomas A. Gerds

Dept of Biostatistics, University of Copenhagen, Copenhagen, Denmark

Ewout W. Steyerberg: E.Steyerberg@ErasmusMC.nl

The Editorial in this journal addresses the important point of analysis of cancer survival. Some issues may merit some further comments. First there is the issue of what can be estimated from the data as available. Different models have been proposed, with the Cox proportional hazard models being most popular in medicine. This model allows for the estimation of relative risk between groups, expressed in the ratio of hazards. Indeed, parametric models can also be used. The hazard ratio (HR) as estimated from e.g. a Weibull model will usually not differ much from the HR estimated from a standard Cox model ¹. But cure models bring about problems in their estimation, and more importantly, in their definition of ‘cure’. Cure status is often not properly defined in the sense that a medical doctor would be able to tell the difference between a cured and a non-cured patient at a specific time point during follow-up.

Second, there is the general issue of extrapolation beyond observed data. The Weibull model allows for that, and may sometimes in retrospect prove to be approximately correct, but this does not negate the fact that extrapolation has fundamental limitations. Obviously, if there are no data, we cannot draw conclusions that are based on data. The standard Cox model will naturally stop predictions after the last patient died or was censored.

Third, the research question is the most important issue in discussing approaches to a statistical analysis. If the research question comes down to an interest in relative effects, the hazard ratio from a Cox model addresses that well. Other questions are on absolute effect and predictions, e.g. differences in months lived or differences in numbers of deaths ². The Cox model then needs to consider the baseline hazard, and can only provide answers within the time frame observed.

From relative to absolute risk

In the example considered by the authors, the hazard ratio is 0.2, where HR denotes the ratio between the hazard rate in the treatment group and the hazard rate in the placebo group. The interpretation is that at any day in the follow-up period, the chances that the event occurs tomorrow are 80% higher in the placebo group. It can not be concluded from this number that the treatment would prevent $(1 - \text{HR}) = 80\%$ of the events in the placebo group. The difference in absolute event risk between the two groups crucially depends on the development of risk over time. To illustrate this issue we consider two different situations, where in both cases

HR=0.2, but the risk increase over time is rapid or slowly (Figure 1 and 2 respectively, simulated with $n=3000$ patients assuming Weibull models with different shapes). In Figure 1, the treatment would prevent 77% of the events till 12 years in the placebo group (8% instead of 33% event risk), in contrast to Figure 2, where the treatment would prevent only 45% of the events till 12 years in the placebo group (54% instead of 98% event risk). Only if we consider a short-term time point, e.g. 2 years, these reductions are close to 80% (79.8% and 79% respectively).

Competing risks, recurrence and death

Predictions from a standard survival analysis, for example the Kaplan-Meier curve or the Cox regression model, are only valid if there are no competing events that would take a person from being at risk for the event of interest, such as recurrence of disease. One might be interested in the effect of treatment on recurrence risk. Competing risks might e.g. include death from heart disease. Thus, besides the assumption regarding random censoring of patients as rightly noted by the authors, another crucial assumption is that every patient will eventually experience the event of interest and that competing risks cannot occur. A proper analysis in the presence of competing risks would certainly be beyond the ordinary Cox regression model^{3 4}.

To further analyze the role of intermediate events, such as recurrence, we might consider multistate models (Figure 3). These models allow for estimation of the relative and absolute risk of recurrence, and the estimation of the relative and absolute risk of death given that relapse has occurred⁵.

In sum, we agree that alternative models to the standard Cox proportional hazard model may have a place in the analysis of cancer survival⁶. We doubt that cure models are useful unless the cure state is defined from a medical perspective. Competing risk, and multistate models may be useful to address questions that go beyond standard questions on the relative effect of a treatment or other prognostic factors.

References

1. Harrell, FE. Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001.
2. Steyerberg, EW. Clinical prediction models : a practical approach to development, validation, and updating. New York: Springer; 2009.
3. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 1999;94:496–509.
4. Wolbers M, Koller MT, Witteman JC, Steyerberg EW. Prognostic models with competing risks: methods and application to coronary risk prediction. Epidemiology 2009;20(4):555–61. [PubMed: 19367167]
5. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat Med 2007;26(11):2389–430. [PubMed: 17031868]
6. Martinussen, T.; Scheike, TH. Dynamic regression models for survival data. New York, N.Y.: Springer; 2006.

Figure 1. Survival functions for slowly increasing risk. The hazard ratio (HR) is 0.2.

Figure 2. Survival functions for rapidly increasing risk. The hazard ratio (HR) is 0.2.

Figure 3.

Multistate model describing the possible course of a patient after being disease free, e.g. after surgery with curative intent. The arrows indicate transition probabilities, which depend on time. $a(t)$ and $c(t)$ indicated transition probabilities from disease-free to relapse and dead respectively, while $b(t)$ indicates the transition probability from disease-free to dead. With this model, we can e.g. compare the probability of dying in the initial state (disease-free) to the probability of dying in the relapse state, hypothesizing that $b(t) > c(t)$.