Transplant International

REVIEW

An unjustified benefit: immortal time bias in the analysis of time-dependent events

Andreas Gleiss[1], Rainer Oberbauer[2] & Georg Heinze[1]

1 Section for Clinical Biometrics, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Vienna, Austria
2 Division of Nephrology and Dialysis, Department of Medicine, Medical University of Vienna, Vienna, Austria

Correspondence
Andreas Gleiss, Section for Clinical Biometrics
Center for Medical Statistics, Informatics, and Intelligent Systems
Medical University of Vienna
Spitalgasse 23
1090 Vienna, Austria.
Tel.: 0043 1 40400 66830;
fax: 0043 1 40400 66870;
e-mail:
andreas.gleiss@meduniwien.ac.at

SUMMARY

Immortal time bias is a problem arising from methodologically wrong analyses of time-dependent events in survival analyses. We illustrate the problem by analysis of a kidney transplantation study. Following patients from transplantation to death, groups defined by the occurrence or nonoccurrence of graft failure during follow-up seemingly had equal overall mortality. Such naive analysis assumes that patients were assigned to the two groups at time of transplantation, which actually are a consequence of occurrence of a time-dependent event later during follow-up. We introduce landmark analysis as the method of choice to avoid immortal time bias. Landmark analysis splits the follow-up time at a common, prespecified time point, the so-called landmark. Groups are then defined by time-dependent events having occurred before the landmark, and outcome events are only considered if occurring after the landmark. Landmark analysis can be easily implemented with common statistical software. In our kidney transplantation example, landmark analyses with landmarks set at 30 and 60 months clearly identified graft failure as a risk factor for overall mortality. We give further typical examples from transplantation research and discuss strengths and limitations of landmark analysis and other methods to address immortal time bias such as Cox regression with time-dependent covariables.

Transplant International 2018; 31: 125–130

Key words
immortal time bias, landmark analysis, survival analysis, time-dependent events

Received: 7 June 2017; Revision requested: 3 July 2017; Accepted: 6 October 2017

Immortal time bias

Already in 1972, a discussion on the life-prolonging effect of cardiac transplantation was supplemented by a "re-assessment" which claimed that two prominent transplantation studies were subject to a "probable selection bias" [1]. The author criticized that in these observational studies, all patients were "assigned to the nontransplant group by default" during the waiting time for transplantation. While the "sickest subjects" tend to die before a suitable donor becomes available, a large portion of "comparatively healthier patients" ends up in the transplant group. The author notes that this fact gives an "artificial boost" to the survival curve of the transplanted patients. In later decades, the "grace period" [1], which is experienced by patients in the transplant group due to having survived at least to the time of transplantation, was termed "immortal time." The aim of the present review is to bring this topic to the attention of transplant researchers again and to sensitize them to such study designs.

For demonstration, we consider a cohort of 633 patients with end-stage renal disease who were followed from renal

doi:10.1111/tri.13081

transplantation to death from any cause or end of observation (overall survival). The data were taken from a larger cohort of renal transplant patients [2]. Of the 633 patients, two-thirds were male (424 patients), 550 (86.9%) had a deceased donor, and nearly all of them (614 patients) were diabetic. Mean age at transplantation was 44.7 years (standard deviation, SD 14.2); mean donor age was 44.2 years (SD 15.2). Median overall survival amounts to 274 months (lower quartile 168); and 167 deaths were observed during a median follow-up time of 126.1 months (quartiles 63–190).

Here, we are interested in the influence of graft failure on overall survival since transplantation. A naïve Kaplan–Meier plot is shown in Fig. 1, treating patients experiencing functional graft failure during their individual observation period as "graft failure patients" and all others as "control patients." The two survival curves hardly differ. Applying a log-rank test, we would arrive at the conclusion that graft failure has no influence on overall survival ($P = 0.533$). This could be supported by a Cox regression analysis comparing these putative "groups" which results in a hazard ratio of 1.10 (95% confidence interval 0.81–1.50, $P = 0.534$). Adjusting this effect for donor type and donor age hardly changes this result (hazard ratio 1.08, $P = 0.636$), see first row of Table 1.

This analysis, however, might be severely biased as we compare those patients for whom a graft failure was observed at any time within their observation period versus patients observed without failure. In general, *immortal time bias* is a special kind of selection bias in observational studies of longitudinal data which occurs if information on the occurrence of a time-dependent event (such as graft failure) is falsely assumed to be known at baseline. In our example, some patients experienced graft failure after transplantation, but at the time point of transplantation, it was not yet known whether those patients would have graft failure at all and if yes, at which time point after transplantation. As this time dependence of the failure status is ignored, the resulting bias is also called *time-dependent bias* [3]; another term used in the literature is *survival bias* [4].

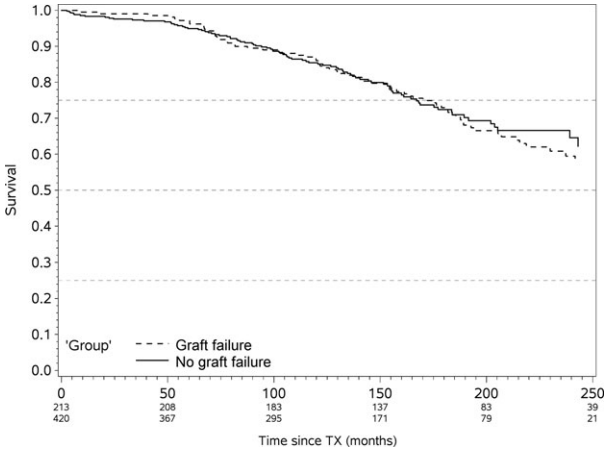

Figure 1 Naïve Kaplan–Meier estimates (numbers at risk per group indicated below time axis).

Let us consider five hypothetical patients from our study represented in Fig. 2. The first two patients remained free of graft failure during their observation period: Patient 1 died after 48 months, while patient 2 was censored after 108 months (e.g., lost to follow-up or due to study termination). Patients 3 and 4 had a graft failure after 48 and 12 months, respectively, and patient 5 died at month 24 before a (hypothesized) graft failure could be detected. A biased analysis ignoring the time dependence of graft failure would treat patients 3 and 4 as "patients with graft failure" and collect the remaining three patients in the "control" group (i.e., without failure).

Patient 3 would also have been counted in the "control group" if she had died during the first 4 years post-transplantation. A similar argument holds for the first year of patient 4 such that these two patients were effectively considered as "immortal" before graft failure; had they died during this period they would have been counted in the "control group" (just as patient 5 was). In addition, a biased analysis would attribute these 5 years of "immortal time" to the survival time of the "graft failure group" and add these years to the survival time after graft failure. On the other hand, patient 5 is classified as a "control patient" and his short survival time is credited to the "control" group since his early death prevented him from graft failure detection. In sum, these misclassifications produce an underestimation of a potentially harmful effect of graft failure on survival.

Landmark analysis

An easy way to prevent misinterpretation of patients and survival times and to correctly represent and model our survival data is a *landmark analysis*. First, we choose a clinically relevant *landmark*, that is, a point on the time axis at which we classify patients into those who had already experienced graft failure and those who were still free of graft failure up to that time. For the patients in Fig. 2, we set a landmark at 30 months and thus compare patient 4, the only patient with a graft failure before that time point, with those who were

Table 1. Estimates from various Cox regression models.

Model	Patients at risk	Unadjusted		Adjusted	
		HR (CI)	*P*	HR (CI)	*P*
Naïve model* (biased)	633	1.10 (0.81–1.50)	0.534	1.07 (0.79–1.47)	0.636
Landmark set at 30 months	600	1.66 (0.87–3.15)	0.124	1.90 (1.00–3.61)	0.049
Landmark set at 60 months	554	1.84 (1.19–2.86)	0.006	1.92 (1.23–2.99)	0.004

HR, hazard ratio; CI, 95% confidence interval; *P*, *P*-value; adjusted, for donor type and donor age.
*Comparing patients with observed graft failure with patients free of graft failure.

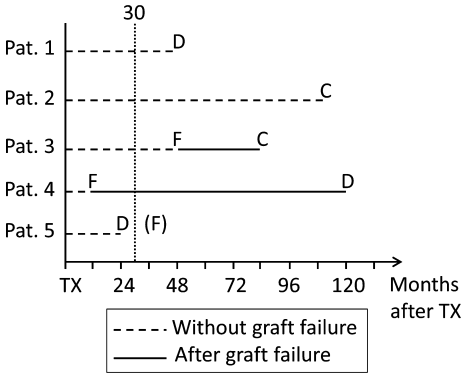

Figure 2 Hypothetical data of five patients (C = censored, D = died, F = graft failure).

still without failure at the landmark (patients 1–3). It is important to see that patient 3 has not had her graft failure at this time point. Furthermore, patient 5 is not at risk any more as he died 2 years after transplantation. Therefore, this patient is not included in the landmark analysis at 30 months.

The landmark analysis at 30 months applied to all 633 patients is shown in the upper panel of Fig. 3. The curves were produced using standard Kaplan–Meier estimates applied to the survival time starting from the landmark for the 600 patients still at risk at 30 months. Twenty-seven of these had a graft failure before the landmark time, that is, during their first 30 months after transplantation. Obviously, these patients exhibit a clear survival disadvantage compared to those who stayed free of graft failure between transplantation and the landmark. Our landmark analysis is performed from the perspective of the landmark time; that is, patients in the latter "group" may still experience graft failures after the landmark time. Our analysis ignores this possibility by intention, because we would like to express whether there is a difference in survival given the information at the landmark time without anticipating future events. We can use a standard Cox proportional hazards regression model to estimate the effect of graft failure up to

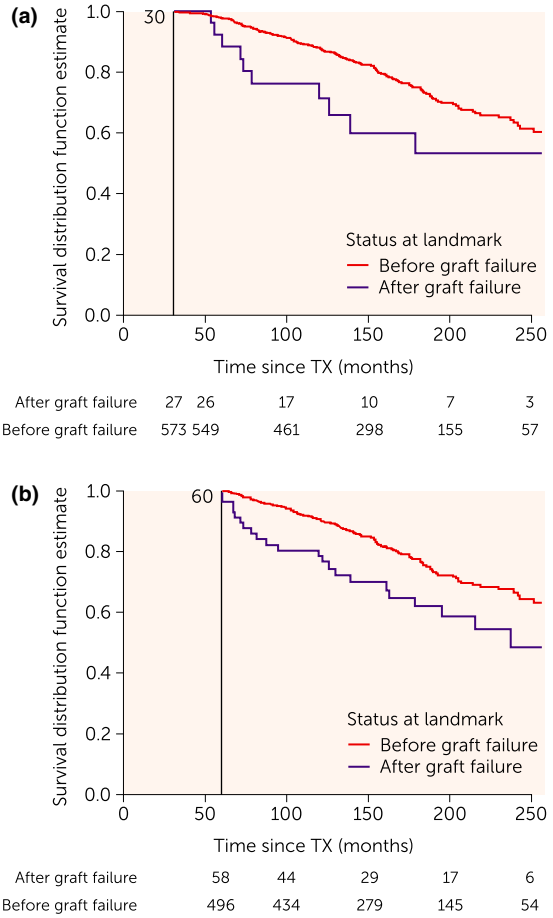

Figure 3 Kaplan–Meier type estimates for landmarks at 30 (upper panel) and 60 months (lower panel). Numbers at risk per group indicated below time axis.

month 30 on survival based on the 600 patients who lived up to this time point. This analysis is based on survival times starting from 30 months after transplantation and gives an adjusted hazard ratio equal to 1.90 (95% confidence interval 1.00–3.61, $P = 0.049$) demonstrating a significantly higher instantaneous risk of death (i.e., hazard) by 90% on average for the 27 patients who have had a graft failure during their first 30 months after transplantation compared to those 573 who were free of

graft failure up to that time. Repeating the analysis for the 554 patients (58 vs. 496) still at risk at the landmark set at 60 months gives a similar adjusted hazard ratio, see Table 1 and lower panel of Fig. 3.

The main advantage of the landmark method is its easy applicability: After carefully performing the necessary data operations for establishing the groups at the landmark and subtracting the landmark time from all survival and censoring times, standard statistical software can be used for graphical representation (using Kaplan–Meier procedures and re-shifting the time axis) and Cox regression modeling.

The landmark method needs some care in the interpretation of its results which are restricted to those patients who are still at risk at the landmark. Therefore, the landmark time needs to be chosen carefully, and the choice should be guided by clinical relevance. For a discussion of the choice of the landmark times, see [5]. Additional landmarks can be considered, for example, for sensitivity analysis. Furthermore, it is possible to build supermodels based on a sequence of landmark times to investigate how the hazard ratio changes with increasing landmark time (see [6]; for an application, see [2]). This idea can also be extended to recurrent events such as infections after kidney transplantation as exemplified by [7].

A limitation of the landmark approach is that patients are ignored who have died or been censored before the landmark. In this way we lose patients for the analysis and restrict our results to the remaining patients. This loss of study information implies a loss of statistical efficiency [4], the amount of which depends on size of risk sets at the specified landmark. Furthermore, it was noted that landmark analyses produce average results that are conditional on the set of "waiting times" (i.e., individual times until the time-dependent event) observed in the sample at hand. This precludes a "counterfactual interpretation" [8] which would be necessary for causal inference about the effect of the time-dependent event on survival. The results are more of a descriptive or predictive type, explaining how an occurred time-dependent event changes prognosis. If the time-dependent event is defined by the onset of a specific treatment, then usually causal inference is desired, but needs to be addressed by other methods such as marginal structural models [9].

Discussion

Immortal time bias is a phenomenon often encountered in leading clinical journals [10]. It has been identified in various medical subdisciplines such as pharmaco-epidemiology [11], kidney transplantation [12], or post-operative radiotherapy [13]. Furthermore, various letters published in medical journals draw attention to this kind of bias [14,15]. In extreme cases, the bias is visually identifiable in plots of survival curves when the group of patients who experienced the considered time-dependent event shows a close to horizontal curve at the beginning while the "control group" has an immediate and dramatic drop. In general, there is a risk for immortal time bias in studies of survival outcomes (or more general, time-to-event outcomes) where interest focuses on the effect of a time-dependent event on survival. The bias arises if the analysis uses statistical methods which were developed for comparing groups defined at baseline when in fact the group membership may change over time. It can be circumvented by choosing analysis methods which can deal with dynamic group definitions.

In the example data used for demonstrating immortal time bias, the outcome was overall survival after transplantation and the time-dependent event was graft failure. However, there are many other questions in transplantation research where this type of bias could arise. Some examples are listed in Table 2. Furthermore, immortal time bias is an issue in studies where patients are allocated to a particular intervention or start utilizing a therapy after they enter the study, for example, in oncologic studies of add-on therapies [16] or in pharmaco-epidemiology [11], or when treatment is defined as at least one prescription dispensed after hospital discharge or diagnosis [17]. For an overview of more general examples of immortal time bias, see [18].

It can be mathematically proven that analyses subject to immortal time bias result in biased effect estimates [3]: If the time-dependent event has a beneficial effect on survival (such as a treatment, e.g., ICD therapy in Table 2), a biased analysis will over-estimate this effect; if it has a harmful effect (such as graft failure, rejection or infection), then this effect will be underestimated; if it is unrelated to survival, then a biased analysis will pretend a beneficial effect. Several conditions in a study determine the magnitude of the bias. First, a longer time between inclusion of a patient and the time-dependent event will lead to more bias. For example, in the renal transplantation settings listed in Table 2, the time from transplantation to graft loss will usually be longer than the time to rejection. Second, the magnitude of bias can also depend on the evolution of mortality risk before and after the time-dependent event [4] which again depends on the severity of the considered time-dependent event for the patient's survival outcome. For example, consider dialysis as the inclusion criterion, and

Transplant International 2018; 31: 125–130

Table 2. Examples of studies in transplantation research where immortal time bias can arise.

Patient cohort	Survival outcome		Time-dependent event	Reference
	From	To		
Renal transplantation	Transplantation	All-cause death	Graft failure	Our example
Renal transplantation	Transplantation	Graft loss	Rejection	[19]
Hematopoietic malignancies	Remission	All-cause death	BMT	[20]
ALL	Remission	Failure	Stem cell transplantation	[8]
Heart failure	Hospital discharge	All-cause death	ICD therapy	[5]
Heart failure	Acceptance as transplant recipient	All-cause death	Heart transplantation	[1]

ALL, acute lymphoblastic leukemia; BMT, bone marrow transplantation; ICD, implantable cardioverter-defibrillator.

kidney transplantation as the time-dependent event. The mortality risk immediately after transplantation can be higher than under dialysis, but prognosis usually improves a few months thereafter. The impact of these conditions on the risk of bias has to be rated on a case-by-case basis, considering also aspects of study design like length of follow-up and inclusion criteria.

Landmarking is not the only way to deal with immortal time bias. Other statistical methods that were proposed to handle the bias, however, were shown to not (fully) eliminate the bias or need specific assumptions to work. As an alternative to naively calculating a "control" patient's survival time from inclusion, some authors imputed a starting time point ("time zero") for "control" patients randomly. This has been criticized as inappropriate by Zhou and co-workers [4]. The same authors proposed to assign a "time zero" for each "control patient" at random from the set of times observed in patients experiencing the time-dependent event and showed to effectively remove immortal time bias in this way. This method of *time-distribution matching* only sacrifices the data of "control" patients with an outcome event before their assigned "time zero" and usually leads to less information loss than the landmarking approach [4].

Bernasconi and colleagues [8] proposed to use the following strategy to construct survival curves contrasting patients with the time-dependent event and patients without. For patients with event, the "clock is started" at the time of the event; thus, the immortal time is removed from analysis. A survival curve describing the prognosis for patients remaining free of the event is constructed using all patients, but censoring those who experience the event at the time at which the event occurred. The assumption that this method needs to provide a fair comparison is that the time-dependent event must occur independently of a patient's prognosis. As an example where this assumption is approximately fulfilled, Bernasconi *et al.* consider a study of survival in patients with acute lymphoblastic leukemia, where initially all subjects are treated with chemotherapy until they receive a bone marrow transplant. This time-dependent event mainly depends on availability of a donor and not on the patient's survival prospects. However, this independence assumption is very often not fulfilled.

A frequently used alternative to the landmark approach is to dynamically update the "group definitions" when estimating survival curves or fitting Cox regression models [21]. Revisiting Fig. 2, there are five patients without graft failure at risk immediately after transplantation. After 12 months, these reduce to four while patient 4 is counted as "after failure." Thus, unlike the standard Kaplan–Meier procedure, this approach allows group sizes to increase over time. After 24 months, there are three and one patients still at risk with status "before failure" and "after failure," respectively, and after 48 months, there are one and two patients.

Unfortunately, even if possible with some software packages, Kaplan–Meier estimation should not be based on such dynamically updated risk sets ("extended Kaplan–Meier estimate" [22]), [23]. However, Cox regression, which compares hazards rather than survival probabilities, can be extended to incorporate *time-dependent covariables*, and corresponding options are available in the Cox regression procedures of SAS, R, or SPSS. This way of modeling has been shown to effectively remove immortal time bias from hazard ratio estimates [4]. The resulting hazard ratios quantify the immediate effect of the time-dependent event on the outcome. Returning to our example and modeling graft failure as a time-dependent covariable, we estimated an unadjusted hazard ratio of 2.98 (95% confidence interval 2.14–4.15, $P < 0.001$), suggesting that graft failure increases the all-cause mortality hazard to the threefold. A disadvantage of this

method is the abovementioned missing graphical description by survival curves. Furthermore, the hazard ratio of a time-dependent covariable can hardly be interpreted as its causal effect, because the time-dependent covariable is usually accompanied by time-dependent confounding. Sophisticated methods to deal with this problem were proposed, for example, by [9]. In randomized trials, confounding is usually absent. In addition, if randomization is performed at the time of a patient's inclusion into a study, even if it concerns a therapy to be started later during follow-up, analysis can fully ignore the time dependency of that therapy, because the group assignment is known at baseline.

Summarizing, we presented the landmark approach as a sensible and pragmatic solution to the problem of time-dependent events in transplantation research. Its simplicity, however, comes at the cost of a loss of information which can be considerable with late landmark times. Using time-distribution matching or a Cox model including a time-dependent covariable are alternatives, however, with their own limitations (cf. table 4 in [4]). Landmarking might then be used as a sensitivity analysis to gain a broader view of the impact of the considered time-dependent event on the outcome.

Conflict of interests

The authors have declared no conflicts of interest.

Funding

The authors have declared no funding.

REFERENCES

1. Gail MH. Does cardiac transplantation prolong life? A reassessment *Ann Intern Med* 1972; **76**: 815.
2. Haller MC, Kammer M, Kainz A, Baer HJ, Heinze G, Oberbauer R. Steroid withdrawal after renal transplantation: a retrospective cohort study. *BMC Med* 2017; **15**: 8.
3. Beyersmann J, Gastmeier P, Wolkewitz M, Schumacher M. An easy mathematical proof showed that time-dependent bias inevitably leads to biased effect estimation. *J Clin Epidemiol* 2008; **61**: 1216.
4. Zhou Z, Rahme E, Abrahamowicz M, Pilote L. Survival bias associated with time-to-treatment initiation in drug effectiveness evaluation: a comparison of methods. *Am J Epidemiol* 2005; **162**: 1016.
5. Mi X, Hammill BG, Curtis LH, Greiner MA, Setoguchi S. Impact of immortal person-time and time scale in comparative effectiveness research for medical devices: a case for implantable cardioverter-defibrillators. *J Clin Epidemiol* 2013; **66**: S138.
6. van Houwelingen HC, Putter H. *Dynamic Prediction in Clinical Survival Analysis*. New York, NY: CRC, Chapman & Hall, 2012.
7. Musoro JZ, Struijk GH, Geskus RB, Ten BI, Zwinderman AH. Dynamic prediction of recurrent events data by landmarking with application to a follow-up study of patients after kidney transplant. *Stat Methods Med Res* 2016. https://doi.org/10.1177/0962280216643563. [Epub ahead of print]
8. Bernasconi DP, Rebora P, Iacobelli S, Valsecchi MG, Antolini L. Survival probabilities with time-dependent treatment indicator: quantities and non-parametric estimators. *Stat Med* 2016; **35**: 1032.
9. Hernán MÁ, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000; **11**: 561.
10. Van Walraven C, Davis D, Forster AJ, Wells GA. Time-dependent bias was common in survival analyses published in leading clinical journals. *J Clin Epidemiol* 2004; **57**: 672.
11. Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf* 2007; **16**: 241.
12. Kim SJ. Immortal time bias in cohort studies of kidney transplant recipients, letter to the editor. *Am J Transplant* 2010; **10**: 190.
13. Park HS, Gross CP, Makarov DV, Yu JB. Immortal time bias: a frequently unrecognized threat to validity in the evaluation of postoperative radiotherapy. *Int J Radiat Oncol Biol Phys* 2012; **83**: 1365.
14. Lash TL, Cole SR. Immortal person-time in studies of cancer outcomes. *J Clin Oncol* 2009; **27**: e55.
15. Lange T, Keiding N. Skin cancer as a marker of sun exposure: a case of serious immortality bias. *Int J Epidemiol* 2014; **43**: 971.
16. Gleiss A, Frass M, Gaertner K. Re-analysis of survival data of cancer patients utilizing additive homeopathy. *Complement Ther Med* 2016; **27**: 65.
17. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ (Online)* 2010; **340**: 907.
18. Wolkewitz M, Allignol A, Harbarth S, De Angelis G, Schumacher M, Beyersmann J. Time-dependent study entries and exposures in cohort studies can easily be sources of different and avoidable types of bias. *J Clin Epidemiol* 2012; **65**: 1171.
19. Loupy A, Vernerey D, Tinel C, *et al.* Subclinical rejection phenotypes at 1 year post-transplant and outcome of kidney allografts. *J Am Soc Nephrol* 2015; **26**: 1721.
20. Gray R, Wheatley K. How to avoid bias when comparing bone marrow transplantation with chemotherapy. *Bone Marrow Transpl* 1991; **7**: 9.
21. Mantel N, Byar DP. Evaluation of response-time data involving transient states: an illustration using heart-transplant data. *J Am Stat Assoc* 1974; **69**: 81.
22. Snapinn SM, Jiang Q, Iglewicz B. Illustrating the impact of a time-varying covariate with an extended Kaplan-Meier estimator. *Am Stat* 2005; **59**: 301.
23. Beyersmann J, Gerds T, Snapinn SM, Jiang Q, Iglewicz B. Illustrating the impact of a time-varying covariate with an extended Kaplan-Meier estimator. *Am Stat* 2005; **59**: 301: Comment by Beyersmann, Gerds, and Schumacher and response. Am Stat. 2006;60(3):295–6.