Avoiding blunders involving 'immortal time'\(^1\)

[Draft Nov 12, 2012]

“As Groucho Marx once said ‘Getting older is no problem. You just have to live long enough.’”

Queen Elizabeth II, at her 80th birthday celebration in 2006

“… this award proves one thing: that if you stay in the business long enough and if you can get to be old enough, you get to be new again.”

George Burns, on receiving an Oscar, at age 80,\(^1\)

(Richard Burton died, a nominee 6 times, but sans Oscar, at 59. Burns lived to 100, so how much of the 41 years’ longevity difference should we credit to Burns’ winning the Oscar?)

Some time ago, while conducting research on U.S. presidents, I noticed that those who became president at earlier ages tended to die younger. This informal observation led me to scattered sources that provided occasional empirical parallels and some possibilities for the theoretical underpinning of what I have come to call the precocity-longevity hypothesis. Simply stated, the hypothesis is that those who reach career peaks earlier tend to have shorter lives.


“Statin use in type 2 diabetes mellitus is associated with a delay in starting insulin”

Yee et al. Diabet Med 2004

1. Introduction

For almost two centuries teachers have warned against errors, involving what is now called ‘immortal time,’ whereby life- or time-extensions that took place before an intervention or change of status, are nevertheless credited to it. Despite the warnings, and many examples of how to proceed correctly, this type of blunder continues to be made, and in a widening range of investigations. In some instances, consequences of the error are less serious, but in others the false evidence produced by this type of blunder has been seized on by social scientists, or used to promote greater use of pharmaceuticals, medical procedures and medical practices or to minimize occupational hazards.

In this article, we first use a recent example to illustrate this error. We discuss other names for it, how old it is and who has tried to warn against it. We suggest how to recognize it, and explore why it continues to trap researchers. Finally, we describe

---

\(^1\) James A. Hanley\(^a\) and Bethany J Foster\(^ab\), Department of (a) Epidemiology, Biostatistics and Occupational Health (b) Pediatrics; McGill University, 1020 Pine Avenue West, Montreal, Quebec, H3A 1A2, Canada.
Correspondence: james.hanley@mcgill.ca
statistical ways of dealing with denominators measured in units of time rather than in numbers of persons. Since the ‘extended’ Cox model is often used in this context, we describe it in some detail.

2. Example and commentary

Example

Most patients whose kidney transplants (allografts) have failed must return to long-term dialysis. But should the failed allograft be removed or left in? To learn whether its removal “affects survival”, researchers\{Ayus\} used the US Renal Data System to study “a large, representative cohort of [10,951] patients returning to dialysis after failed kidney transplant.” Some 1106, i.e. 32\%, of the 3451 in the “allograft nephrectomy” group, and 2679, i.e., 36\% of the 7500 in the “non-nephrectomy” group, i.e., 3785 in all, were identified as having died by the end of follow-up.

Compared with patients who did not undergo a nephrectomy, those who did had a lower-risk socio-demographic profile and lower comorbidity burden, but were more likely to have higher serum creatinine concentration and higher serum albumin and, “during follow-up, […] to have a hospitalization with a discharge diagnosis of fever, anemia, sepsis, urinary tract infection, or complication of the transplanted kidney”. To address this “possible treatment selection bias”, the researchers “constructed a propensity score for the likelihood of receiving allograft nephrectomy during follow-up using logistic regression (c statistic = 0.76) and included as covariates variables known to be associated with a clinical indication for nephrectomy and any other characteristics that differed between the two groups.” They then “performed multivariable extended Cox regression comparing the risk for death associated with receipt of allograft nephrectomy
during follow-up that adjusted for quartile of propensity score, variables previously shown to be associated with mortality after failed transplant, and any differences in any other characteristics between those who did and did not die during follow-up.”

In these analyses that adjusted “for the propensity score … and other potential confounders” they found that “receiving an allograft nephrectomy was associated with a 32% lower adjusted relative risk for all-cause death (adjusted hazard ratio 0.68; 95% confidence interval 0.63 to 0.74)

In their discussion, the researchers suggest that their finding of “improved survival” after allograft nephrectomy “challenges the traditional practice of retaining renal allografts after transplant failure” and that the (“surprising”) finding of significantly higher rates of repeat transplantation “argues against withholding transplant nephrectomy because of a presumed reduced chance of repeat transplantation.” The title of the article suggested causality. While they emphasized the large representative sample and the extensive and sophisticated multivariable analyses, they did caution that residual confounding and treatment selection bias may have influenced their findings, and remain limitations of this observational, registry study. Nonetheless, they suggested that these findings should prompt a randomized trial to evaluate the role of routine allograft nephrectomy compared with current management strategies as a possible strategy for improving outcomes among stable dialysis patients with a failed renal allograft.

The paper was accompanied by an editorial, which reiterated the concerns regarding residual confounding and selection bias, despite the statistical methods used to address possible treatment selection bias and adjust for characteristics that differed between the two groups. However, despite a vivid warning, in the same journal, that
would have been still fresh in their minds, the editorialists overlooked another aspect of
the analysis that has unequivocally distorted the comparison. The overlooked information
is to be found in the statements (italics ours) that

3451 received nephrectomy of the transplanted kidney during follow-up; the median time
between return to dialysis [the time zero in the Cox regression] and nephrectomy was 1.66 yr
(interquartile range 0.73 to 3.02 yr).

[Paragraph 1 of Results section]

and that

Overall, the mean follow-up was (only) 2.93 ± 2.26 yr. [Paragraph 3 of Results section]

From these and other statements in the report, it would appear that in their analyses
follow-up of both ‘groups’ began at the time of return to dialysis. The use of this time-
zero for the 3451 who had the failed allograft removed is not appropriate or even logical.
These patients could not benefit from its removal until after it had been removed; but, as
the median of 1.66 yr indicates, a large portion of their ‘follow-up’ was spent in the initial
‘failed graft still in place’ state – along with those who never underwent nephrectomy of
their failed allograft.

Since the 3451 patients who ultimately underwent a nephrectomy (the
“nephrectomy group”) had to survive long enough to do so (collectively, approximately
6,700 patient years, based on the reported quartiles of 0.73, 1.66 and 3.02), there were, by
definition, no deaths in these 6,700 pre-nephrectomy patient years. In modern parlance
these 6,700 patient years were “immortal.” There was no corresponding “immortality”
requirement for entry into the “non-nephrectomy group”. Indeed, all 10,951 patients
returning to dialysis after failed kidney transplant began follow-up in their initial “failed
graft in place” state. Some 7500 of these remained in that initial state until their death (for
some, death occurred quite soon, before removal could even be contemplated) or the end
of follow-up, while the other 3451 spent some of their follow-up time in that initial
(“failed graft in place”) state and then changed to the “failed graft no longer in place”,
i.e., post-nephrectomy state.

How big a distortion could the mis-allocation of these 6,700 patient years produce?
The article does not have sufficient information to re-create the analyses exactly. Figures
1 and 2 show a simpler hypothetical dataset which matches the reported summary
statistics quite closely, but was created assuming no variation in mortality rates over
years of follow-up. We set up the sham ‘intervention’ ‘retroactively’ so that it did not
affect (other than randomly) the mortality rates in the PY lived in each state.

Fig2A shows that even though the data were generated to produce the same
expected mortality rate of 11.8 per 100PY for the observed experience in the “initial” and
“post-‘intervention’” states, the inappropriate type of analysis used in the paper, applied
to these hypothetical data, would have resulted in a much lower rate (6.4) in the
“‘intervention’ group” and a much higher one (17.1) in the “’non-intervention’ group.”
The reason is that none of the 1031 deaths post-‘intervention’ could have occurred, and
none of them did occur, in the 6732 (immortal) pre-‘intervention’ PY that are included in
the denominator used in the calculation of the rate of 6.4; patients had to survive in order
to undergo the intervention. All patients who died before a nephrectomy was done were
remained (by default) in the non-nephrectomy group. It is equally important to recognize
that the 2759 deaths attributed to the non-nephrectomy “group” occurred not in only the
16096 PY experienced by those who never underwent nephrectomy, but rather in a much
larger denominator: 16096+6732 = 22828 PY lived in the “non-nephrectomy” state
(including the pre-intervention time of those who ultimately underwent nephrectomy).
The omission of the 6732 pre-intervention PY from the denominator led to the inflated rate of 17.1 deaths/100PY. Indeed it was because of these (misplaced) immortal 6732 PY they had already survived that the 3451 patients got to have the ‘intervention’. In other words, it may not have been that patients who underwent nephrectomy lived longer because of the ‘intervention’, but rather that they underwent the ‘intervention’ because they survived long enough to undergo it: an example of ‘reverse causality.’

In this admittedly over-simplified version of the data, with no covariates, the inappropriate analysis led to an apparent rate ratio of 6.4/17.1 = 0.37. The corresponding ‘reduction’ of 63%, and an “improved survival” of at least 2.5 years (areas under the first 11 years of the Kaplan Meier curves of 7.7 vs. 5.2 years), would have been interpreted as having been produced by the intervention; however, they are merely artifacts of the mis-allocation of the PY.

Fig2B shows an appropriate comparison of mortality rates in time-dependent states. With “each unit of person-time allocated to the state in which the death would have been assigned should it occur at that time,” the appropriate rates are (apart from random error) identical, as were the theoretical rates used to generate these hypothetical data. The theoretical rates were – unrealistically – taken as constant over the follow-up years. In reality, the PY in each year of follow-up time would be contributed by individuals who were almost one year older than the individuals who contributed PY the year before, and so the mortality rates in successive time-slices would also be successively higher. Thus, since the person-years in the “post-intervention” state are ‘older’ person years, a summary rate ratio computed using matched slices of follow-up time would be more appropriate than a crude rate ratio (see Fig 4).
One would also need to match the person-years on several patient-related factors. The necessarily now-more-individualized time-slices in the Poisson regression (or Mantel-Haenszel summary rate ratio) thus become smaller but more numerous, but the number of informative PT ‘segments’ remains limited by the total number of deaths. Cox’s use of risksets, each of which involves those alive in the moment before a death occurs, focuses only on these informative moments; it also focuses only on the (assumed to be) common mortality rate ratio, and avoids the modeling of the mortality rates themselves. We will return to this topic in section 5.

3. Teachings against such blunders

Warnings against on this error go back at least to the 1840s, when William Farr\(^\text{1885}\) reminded sanitarians and amateur epidemiologists that

```
Certain professions, stations, and ranks are only attained by persons advanced in years; [...] hence it requires no great amount of sagacity to perceive that “the mean age at death,” or the age at which the greatest number of deaths occurs, cannot be depended upon in investigating the influence of occupation, rank, and profession upon health and longevity.
```

Then, in an admirable style seldom matched in today’s teachings, he explained that

```
if it were found, upon an inquiry into the health of the officers of the army on full pay, that “the mean age at death” of “Cornets, Ensigns, and Second-Lieutenants” was 22 years; of “Lieutenants” 29 years; of “Captains” 37 years; of “Majors” 44 years; of “Lieutenant-Colonels” 48 years; of general Officers, ages still further-advanced --and that the ages [at death] of Curates, Rectors, and Bishops; of Barristers of seven years’ standing, leading Counsel and venerable Judges -- differed to an equal or greater extent,

then,
```

```
a strong case may no doubt be made out on behalf of those young, but early-dying Cornets, Curates, and Juvenile Barristers, whose “mean age at death” was under 30! It would be almost necessary to make them Generals, Bishops, and Judges--for the sake of their health.
```

Crediting the years of immortality required to reach it to the rank that *the person has reached by the time (s)he dies or follow-up ends* exaggerates any longevity-extending benefits of reaching this rank. Likewise, crediting the time *until* one receives a medical
intervention to the intervention exaggerates the life- or time-extending benefits of the intervention.

Whereas Farr spoke tongue in cheek, Bradford Hill\textsuperscript{[1984]} spelled out the reason for the longevity difference: “few men become bishops before they have passed middle life, while curates may die at any age from their twenties upwards.”

Breslow and Day\textsuperscript{[ref above1987]} use a diagram, and a simplified occupational epidemiology example, modeled on the classic blunder by Duck et al.\textsuperscript{[1975]}, to emphasize the correct allocation of person time, and the distortions produced by mis-allocation. In Figure 3 we illustrate the Duck et al. error and Wagoner et al.’s\textsuperscript{[1976]} re-allocation of the person years; we also repeat Breslow and Day’s succinct enunciation of the general principle.

Although the term “immortal time” was used by George Hutchison in the 1970s, his Harvard colleague Alexander Walker\textsuperscript{[1991]} is the first we know of to have put the term in writing, in his 1991 textbook. Walker’s numerical examples all involve the correct allocation of such time, with no example given of the consequences of mis-allocation. The various editions of the Rothman and Greenland textbook\textsuperscript{[1998]} do have an example – albeit hypothetical – of the difference between incorrectly and correctly calculated rates based on two parallel groups of exposed and persons, and state the principle, “If a study has a criterion for a minimum amount of time before a subject is eligible to be in the study, the time during which the eligibility criterion is being met should be excluded from the calculation of incidence rates.” They also allude, without an example, to the more general situation where subjects change exposure categories. It is in this latter situation that most immortal-time blunders are made.
By using the term “immortal time” directly in the title of a 2003 article, Suissa\cite{2003} immortalized the term itself. Since then, more than a dozen articles and letters by him and his pharmaco-epidemiology colleagues have addressed the growing number of serious ‘immortal time’ errors in this field. Typically, cohort membership in these studies was defined at the time of diagnosis with, or hospitalization for, a medical condition. The blunders were created by dividing the patients into those who were dispensed a pharmacological agent at some time during follow-up (unlike in most clinical trials, not all received it immediately at entry to the cohort), and those who were not. Again, the problem lies in attributing to the ‘exposed’ group person-time during which patients were not actually exposed to the medication, and in failing to attribute this unexposed time to the ‘unexposed’ group. When exposed and unexposed time is correctly apportioned, the rate-lowering power of the agent disappears.

In several of their articles, Suissa use other real datasets to address the same question, and show the consequences of the mis-allocation. In the diabetes examples examined by Levesque\cite{2010} and Hoffmann\cite{2011} the minimum time requirements for statin and self-monitoring use increase the amount of immortal time and the size of the artifact. Our annotated bibliography gives several other examples (and collections of them) from several other commentators\cite{Rothman, Glesby, vanWaltvaren, Carieri}, of time-blunders in several other fields. The growing extent of this problem is of concern.

4. Recognizing and avoiding immortal-time blunders

Table 1 lists some ways to recognize immortal time, and to avoid the associated traps. We suspect that some of the blunders stem from the tendency – no matter the design – to
refer to ‘groups,’ as though – in a parallel-arm trial -- they were formed at entry and remained closed thereafter. Even when describing a crossover trial, authors mistakenly refer to the treatment group and the placebo group, rather than to the time when the (same) patients were in the treatment and placebo conditions or states. This tendency may reflect the fact that many questions of prognosis can only be studied experimentally by parallel group designs, and that (except in studying the short-term effects of alcohol and cell phone use while driving, or medication use or inactivity on blood clots) crossover designs (called split-plot designs in agriculture) are rare, and their statistical results harder to show graphically and in tables than are those that use independent ‘groups.’

Just as in the story of Solomon, it is appropriate that persons remain indivisible. But in epidemiology many denominators involve amounts of time (yes, contributed by persons, but time nonetheless), and time is divisible, just as any other (area- or volume-based) denominators that produce Poisson numerators are (the numerators are not divisible). The more comfortable biomedical researchers become in dividing an individual’s time (e.g., same person with different hearts in different time-intervals, see section 5), the less the risk of immortal time blunders.

It is our impression that epidemiologists are more comfortable ‘splitting time’ than many researchers in the social sciences, where correlations (rather than differences in and ratios of incidence rates), are the norm. By their nature, sociological correlations usually rely on intact and indivisible human beings, and intact lifetimes, and cannot easily handle the incomplete lives, and censored and truncated observations encountered more by epidemiologists. The “correlations between election age and death age for restricted subsamples based on election age percentile” {McCann 2003} is a good example of how much
flexibility is gained when data-analysts compare rates (dimension: time\(^{-1}\)) within the relevant time-windows rather than using each person’s total length of time as an indivisible ‘independent variable,’ or “setting time-zero” to some arbitrary birthday (e.g., 0, 50, 65 for Oscar-nominated performers). Moving up from being students of “times to event” or “survival” to being students of rates (or hazards, or the force of morbidity or mortality) has many benefits. It is not very meaningful to tell a 65-year old contemplating trip-cancellation insurance what the average or median “time to ‘have to cancel’ ” is; it would be more meaningful to say that the ‘have to cancel’ rate for 65 year olds travellers is 1 per 50 years of pre-trip waiting, and \((1/50) \times 1.08^x\) for persons aged 65+x. With this ‘rate function’, one can then calculate the real quantity of interest: the risk for a person of a given age for a given “time until the trip”.

Theories such as the just cited precocity-longevity hypothesis) are seductive, and have a certain plausibility. But some of this may be as a result of the framing. Sometimes a simple restatement of the ‘evidence’ can help uncover the fallacy: imagine if a Groucho Marx were to re-word it, using Ronald Reagan’s election and longevity as the example. And in any case (as we stated in our re-examination\(^{(Sylvestre 2006)}\) of the claimed 3.9 year longevity advantage for Oscar winners\(^{(Redelmeier 2001)}\), no matter how important or unimportant results would be if they were true, “readers and commentators should be doubly cautious whenever they encounter statistical results that seem too extreme to be true.”

5. Data analysis options, illustrated with ‘survival times after cardiac allografts’

The inset in panel A of Figure 4 shows information we extracted from Figure 1 of the
May 1969 report\(^{(Messmer 1969)}\), from Houston, on survival post heart-transplant.\(^{2}\) The report began: “The first cardiac allograft in man was done in December, 1967 (Barnard 1967). Since then, more than 100 cardiac transplants have been carried out all over the world in desperately ill patients. (...) The aim of the analysis described here is to answer the question: Does cardiac transplantation at the present stage of development prolong the life of patients with advanced therapy-resistant heart-disease?” The summary read:

During the period May 2, 1968, to March 1, 1969, fifteen patients underwent cardiac transplantation for end-stage heart-disease. Their survival-time is compared with that of forty-two potential recipients who did not receive allografts. Mean survival of the potential recipients was 74 days. The average for the transplant patients was 111 days (including 22 days waiting-time before operation).

In the authors’ opinion, “This difference does not justify wide clinical application of cardiac transplantation, but is an indication for its use in suitable cases where it may prolong life and relieve symptoms.”

In addition to its subject-matter connection with the modern example in section 2, and its pioneering nature, the report is also statistically noteworthy. The raw data are presented graphically as individual lifelines, with the right hand end of each horizontal line indicated by a vertical line (indicating death) or a right arrow indicating still alive (censored) on March 1, 1969. Each lifeline was measured from the date the patient was considered as a potential recipient. One figure showed the lifelines for the 42 who were still on the waiting list, or had died; the other those for the patients given a cardiac allograft, with each of the 15 timelines divided into the time ‘waiting’, and the time ‘post-transplant’. Unlike the actuarial approach used in the Stanford report, the authors computed ‘mean’ survival, effectively treating the 17 censored observations (2/15 and

\(^{2}\) In the initial series from Stanford (Clark 1971) some 20 of 34 potential recipients underwent cardiac transplantation between July 1967 and March 1970; Survival was calculated “by the life table method. Survival for transplanted patients, from the time of operation; of non-transplanted patients from the time of selection for transplantation.” Updated versions of that dataset are now found in the R and Stata packages. We chose instead the smaller and earlier Houston series for its smaller size and earlier publication date, and more cautious outlook, and use of immortal time in the comparison.
15/42) as uncensored.

Most noteworthy, “to compare the survival-times of the potential recipients (“mean” 74 days) with those of the transplant patients,” the authors “added the waiting-time of the transplant patients from the moment they were considered as potential recipients to the post-operative survival time” to arrive at average for the transplant patients of 111 days. But, to their credit, the summary clearly states that the 111 “include 22 days waiting-time before operation.”

Beyond dealing properly with the censored observations, how should we deal with these data today? The usual (and usually appropriate) contemporary approach involves the use of the time-dependent covariates in a multivariable parametric or semi-parametric (hazard) regression model, with subjects switching exposure categories over time. Interestingly, the basic (“vanilla”) lifetable regression approach was only introduced by David Cox three years after the Houston report, and it took a few more years before it was ‘generalized’ or ‘extended’ to deal with exactly this type of ‘transient states’ data.

Nevertheless, it is quite instructive to begin with the classical approaches already widespread in 1969, in particular Mantel’s classic 1959 article. We will also use his generalization of lifetables to deal with transitions between ‘exposure’ states; indeed, this 1974 paper is the conceptual forerunner of what is now known as regression for ‘time-varying covariates.’

In the spirit of encouraging readers to be become students of rates, rather than of “mean” longevity or survival, Figure 4 relies entirely on mortality rates; patient-days denominators deal naturally with the censored observations. Figure 4A shows that the same contrast the authors made: the death rate in all of the 1,667 days lived by those who
ultimately received a transplant (0.78/100 PD) was slightly lower than that in all of the 3,111 days lived by those who did not (0.87/100 PD). But the 1,667 include the 311 “immortal” patient days lived in order to receive a transplant. Thus, a somewhat more appropriate contrast, shown in Fig4B, is between the rate of 0.96/100PD in the 1,667-311=1356 post-transplant days and that of 0.79/100PD in all of the 3,111+311=3,422 days lived as a potential recipient. The higher than unity rate ratio of 0.96/0.79 = 1.22 is in keeping with the visual impression obtained from Fig4B.

This (“crude”) contrast may be unfair, since even in the absence of a transplant, patients become sicker, and mortality rates higher, the longer they remain on the list; and, as is evident from Fig4B, the distribution of patient-days post-transplant is located 95 days post time-zero whereas that of the patient-days lived as a potential recipient-transplant is only 76. One way to minimize this is to “match” on the post time-zero time. The Mantel-Haenszel summary rate ratios using no, and then progressively finer time-matching, as well as the visual impression, indicate that the crude mortality rate ratio is indeed artificially low.

1972 also brought Nelder and Wedderburn’s\(^{(1972)}\) theoretical and computing advance in fitting generalized linear models, and paved the way for greater use of Poisson- and binomial-based regression models in epidemiology. Fig4B(2) shows Poisson-based rate ratios fitted using the numbers of deaths and patient days in time-slices ranging in width from 270 days (1 slice) to 1 day (270 slices). They are noteworthy for several reasons. First, they are testimony to the remarkable intuition of Mantel: Forty years later, Clayton\(^{(ref needed)}\) showed that his one-pass, statistically stable, summary ratio
in the case of ‘known-person-time’ denominators is the first iteration of the (iterative) Maximum Likelihood fit of the Poisson rate ratio model \{ref\}. Second, although for each slice the numbers of deaths were modeled as a pair of Poisson distributions whose parameters are coupled by a common rate ratio, the ratio was actually fit using in each slice the binomial distribution that results from \textit{conditioning} on the total number of deaths in the slice. It turns out, because of the special properties of the Poisson model, that the same rate ratio estimate is obtained if one fits an \textit{unconditional} Poisson model, with an indicator (“dummy”) variable for each slice. Third, as can be seen from the last column where time is sliced into single days, the rate ratio obtained from Cox’s proportional hazards model is identical, to two decimal places, to those obtained by both the conditional and unconditional approaches to fitting a Poisson model to the finely sliced data. Indeed, this is not that surprising: with time sliced so finely, no matter the ‘model’, those days in which there are no deaths provide no information on the rate ratio.

Time-slicing also allows us to relax the constant-over-time rate ratio (i.e., proportional hazards) assumption and to instead use Poisson (or, if we condition, binomial) regression to fit a rate-ratio \textit{function}. Despite this considerable flexibility, an important drawback to the traditional person-time regression models is that by definition the person-time in each state in each time-slice is an \textit{aggregate over several persons}. Thus, it is not possible to include in the model any person-level factors that differ among those in the same “cell”. For example, the 8 post-transplant persons contributing time to the data for day 106 had been transplanted for 30, 46, 48, 85, 102, 103, 104 and 106 days

\footnote{In so-called ‘case-control’ studies, the ‘controls’ only provide an \textit{estimate} of the ratio of the two underlying population-time denominators. In this example, the population-time denominators are known, and the conditional distribution of the number of events in the index category is the (single-parameter) binomial distribution, with a known offset. In a ‘case-control’ study, the conditional distribution is the more complex (single-parameter) non-central hypergeometric distribution.}
respectively. However, there is an easy way out of this: time-slice each lifeline and make a dataset with as many observations as there are person days (or person weeks, etc., depending on the level of time-granularity desired). It is easy to do-it-yourself in Stata, R and SAS, but these packages also have special functions that perform time-splitting. In our example, the file contained 4,778 observations (rows), one for each day for each patient. In addition to the day and patient identifiers, each row contained the patient’s state on that day (‘on waitlist’ or ‘post-transplant’) and if the latter, the number of days since the transplant, and if the former, the value 0. Each row also contained an indicator (0/1) outcome variable for whether the follow-up was terminated by death on that day.

For use with the extended (see below) Cox model, each row also contained the time the observation started and ended: e.g., in the case of day 106, the ‘start’ and ‘end’ times were 105.0 and 106.0.

The ‘extended’ Cox model refers to the slicing of time so as to (i) fit the rate ratio not as a constant, but as a function of follow-up time and/or (ii) allow a subject’s ‘x’ variate(s) to take on different values over this time. In the usual Cox model, there is just one lifeline per subject; the start is assumed to be 0. Thus, the survival data are described by just two variables: one denoting the end time, and one indicating whether it is truly the end, or merely censored. In the extended version, three variables are required for each slice: its start, its end, and whether that end is the end, or merely censored. In addition, the record for each slice contains all of the relevant fixed and time-varying covariates for that person for that time-slice.

Returning briefly to the allograft nephrectomy study: Even though the use of a fixed-in-time variate that was coded 0 for patients who ‘never’ underwent nephrectomy,
and 1 for ‘ultimately’ did, is clearly inappropriate, there is no indication in the Ayus et al. report that it was used for purpose (ii); indeed, it would also appear that hospitalizations for anemia, abdominal pain, urinary obstruction, sepsis, urinary tract infection, malnutrition, or complication of transplanted kidney, occurring during the follow-up, were coded as present from the outset. The only indication that the authors split the individual timelines is in their statement “First, unadjusted rates (per 100 person-years) of death from any cause were calculated for periods with or without allograft nephrectomy, but this statement is in conflict with the unadjusted rates in their Figure 2.⁴ 

All other indications, such as the construction of a (one-time) propensity score for the likelihood of undergoing allograft nephrectomy during follow-up using logistic regression” suggest that if indeed the authors did use the extended Cox model, they did not use it to deal with the fact that patients changed states during follow-up.

Fig4B(3) shows what is possible with the ‘extended’ approach. The left panel shows three fitted rate ratio models involving time (t) measured from entry to the waitlist (a) the ‘constant over follow-up’ rate ratio model, already discussed: point estimate RateRatio[t] = 1.72 (b) the model RateRatio[t] = RateRatio[0] × {1 + b1 t} and (c) RateRatio[0] × {1 + b1 t + b2 t²}. [CHECK] The right panel shows three rate ratio models involving time measured from time of transplant. Other forms, such as the simultaneous use of both timescales, are also possible, but clearly, within this dataset with just 13 deaths in the post-transplant state, distinguishing between models is not feasible (the reader is referred to Crowley, Turnbull, and others for a fuller analysis of the (larger)

⁴ Their Figure 2 reports death rates of 32 and per 100 person-years; if these are correct, then based on the reported numbers of deaths, the total person years would only be just over 10,000. But this number is at variance with the overall sample size of 10,951 and the reported average follow-up of 2.93 years. Based on this, and on the reported numbers of deaths and numbers of persons, the reported ‘rates’ of 32 and 36 are probably percentages, based on ‘number of persons’ denominators.
Stanford experience). The fits in Fig4B(3) should be of interest those who continue to defend results derived from mis-allocated immortal-time, inappropriate Kaplan Meier curves, and fixed-in-time covariates, and who argue that models for time-varying covariates are too restrictive and inflexible. In fact, with sufficient data, and sensible biological forms for the supposed effect, the extended model offers considerable flexibility, and can handle various representations of each subject’s history as he proceeds through the risk sets. Thus, in an analysis of the longevity of Oscar winners, when comparing with those who have not won, it is quite feasible to allow 2 performers, who first won in their 20s and 60s respectively, to have post-win hazard ratio functions that reflect how early (or often) in life they won. Similar options apply in other applications, when e.g., examining the effect of past medications, treatments, or occupational exposures.

The computing storage/speed we have today allows us to fit models to datasets with many more rows than Mantel could when, in 1973, he was forced to develop the synthetic (sampling) approach in order to fit a logistic model to the Framingham data. However, a patient-days file for the nephrectomy data would contain approx 10 million rows (10,000 pts x an average of 1,000 days per patient). Alternatives to using a file of this size include coarsening time, storing each patient’s state transitions and covariate changes in compact arrays, and having the software compute ‘on the fly’ as it progresses through each riskset, or using a sampling approach, such as that described by Breslow and Day for so-called ‘nested case control’ studies, or by Hanley and Miettinen for what the latter terms the etiologic study. If it is considered important to match rather than model time, one can extract virtually all the rate information on the rate ratio by using the
case series and a matched base series which is sampled from those days when events occurred, and use conditional logistic regression; if it is considered important to model time, one can couple the case series with a randomly sampled base-series from all 4,778 patient days, and use unconditional logistic regression. Either way, the exponentiated regression coefficient can be directly interpreted as a rate (hazard) ratio. Incidentally, as Miettinen explains [http://bcooltv.mcgill.ca/ListRecordings.aspx?CourseID=6485; t=1h.14m.45s], there is no need to speak of an odds ratio.

**Concluding remarks**

Despite the warnings, and many examples of how to proceed correctly, blunders involving immortal-time continue to be made, and in a widening range of investigations. In some instances (as in the case of longevity of Oscar winners vs. nominees, or popes vs. artists), the consequences of the errors are less serious, but in others the consequences of heeding the false evidence produced by this type of blunder are serious and costly. And, as in the pharmaco-epidemiology work of Suissa and colleagues, the costs of correcting the record can be considerable. Moreover, the original stories in the lay press are not usually updated, even if subsequently contradicted. The Oscar longevity story in the Harvard Health Letter of March 2006 is still available – if one is willing to pay 5$ for the online access pdf – but a search within the Harvard Health Publications website does not mention subsequent findings. The articles by Forbes Magazine writer Matthew Herper (2003; Feb 26, 2011; Feb 26, 2012) and the incomplete link in the 2010 New York Times interview⁵ provide further evidence of the power of the original news story and the

---

resistance to correcting or even just updating the record.

This difficulty in correcting statistical errors is not limited to the lay press. Opportunities to correct blunders in medical journals that were missed by peer review are also limited. Indeed, when one of us (BF) wrote to the Editor of the Journal of the American Society of Nephrology to point out the strong possibility that the finding of “improved survival” following allograft nephrectomy was an artifact, she was told that the journal did not have a Letters to the Editor section, but that it would pass on the concerns to the authors. [Beth: can you supply the email response?]

Given H.R. Haldeman’s observation, “Once the toothpaste is out of the tube, it's hard to get it back in,” it therefore seems prudent, and scientifically responsible, to try to avoid immortal time errors from the outset. Researchers can do this by classifying person-moments into states, rather than incorrectly classifying persons who ultimately reach a certain state into ‘groups’ who have been in that state from the outset!

**Data and R/Stata codes:** The data and R/Stata codes for Figure 4 can be found in the Website [http://www.biostat.mcgill.ca/hanley/software](http://www.biostat.mcgill.ca/hanley/software)

**Funding:** This work was supported by the Natural Sciences and Engineering Research Council of Canada.
Annotated bibliography


Wagoner JK, Infante PF, Saracci R. Vinyl chloride and mortality? [Letter] Lancet. 1976;2:194-5. This letter, a response to Duck et al., is notable both for its tongue-in-cheek ‘possible interpretations’ of the SMRs of 112, 107 and 61 in the PY where exposure had been for <10, 10-15 and 15+ years respectively, and for its re-allocation of the PY giving new SMRs of 79, 137 and 353! See further details in Figure 3.


Rothman KJ. Longevity of jazz musicians: flawed analysis [Letter]. Am J Public Health. 1992;82:761. A letter in response to a retired professor of management, and jazz amateur (but sadly also a statistical amateur), whose data-analysis suggested that jazz musicians, despite their rough lifestyle, live at least as long as their peers. In “Premature Death in Jazz Musicians: Fact or Fiction?” [Spencer FJ. Am J Public Health. 1991 81(6):804-5] the longevity of their peers was measured by the life expectancy of those born the same year as they, although the musicians are, by definition, immortal until they became musicians and eminent enough to be included in the sample. The letter also provides also an interesting contrast with Farr’s teaching style. See Bellis et al. (Elvis to Eminem: quantifying the price of fame through early mortality of European and North American rock and pop stars. J Epidemiol Community Health 2007;61:896-901); Hanley et al. (How long did their hearts go on? A Titanic study. BMJ 2003; 327:1457); Abel et al. (The longevity of Baseball Hall of Famers compared to other players. Death Studies 2005;29:959-63); Redelmeier et al. (Death rates of medical school class presidents. Soc Sci Med. 2004 Jun;58(12):2537-43); Olshansky (Olshansky SJ. Aging of US Presidents. JAMA 2011; 306(21): 2328-2329) for more appropriate ways to carry out such longevity comparisons.

Glesby MJ, Hoover DR. Survivor treatment selection bias in observational studies: examples from the AIDS literature. Ann Intern Med. 1996;124:999- 1005. “Patients who live longer have more opportunities to select treatment; those who die earlier may be untreated by default” and their 3 words “survivor treatment selection” to describe the bias explain why some person-time is “immortal”.

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;57:672-82. They gave immortal time bias a slightly different name because they covered a slightly broader spectrum of situations. Their review surveyed articles containing survival analysis that may have incorrectly handled what they define as a “baseline immeasurable” time-dependent variable, i.e., one that could not be measured at baseline. They focused not just on the exposure of interest, but also other time-dependent covariates.

They describe an interesting study on whether patients having a follow-up visit with a physician who had received the discharge summary would have a lower rate of re-hospitalization. When analyzed as a fixed-in-time variable (ie as 2 ‘groups’, we found a large difference in readmission rates. But “this is a biased association, because patients who are readmitted to the hospital early after discharge do not have a chance to see such physicians and are placed in the “no-summary” group. When a (correct) time-dependent analysis is used, we found a much smaller rate difference. Ironically, in their examination of the article on the survival of Oscar nominees and winners, the only such covariate the reviewers could identify was the possibility that actors could change their names after baseline; they ignored the fact that some nominees who did not win the first time they were nominated did move to the winners category subsequently. So, they (inappropriately) “cleared” this article, declaring it to be free of any possible time-dependent bias.

Carriere MP, Serraino D. Longevity of popes and artists between the 13th and the 19th century. Int J Epidemiol. 2005 Dec;34(6):1435-6. Compared with the bishops in Farr’s example, an even more dramatic longevity advantage would apply to popes. Even though they were aware that longevity is a “necessary condition for being elected Pope,” their statistical approach did not fully address this constraint. Ideally, for each papacy-specific ‘longevity competition’, the time-clock should start when the pope is elected, and the competition should include the pope, and those artists born the same year as he, who were still alive when he was elected. However, for several papacies, such detailed matching is not possible. Instead, for each of the 1200–1599 papacies, their analysis effectively ‘started the clock’ at age 39—the age at which the youngest pope in that era was elected—by excluding artists who died before reaching that age. For the 1600–1900 papacies, it was started at age 38. A re-analysis (Hanley JA, Carriere MP, Serraino D. Statistical fallibility and the longevity of popes: William Farr meets Wilhelm Lexis. Int J Epidemiol. 2006 Jun;35(3):802-5. Epub 2006 Mar 16.) that used a papacy-specific time-clock for each papacy-
specific longevity competition reversed the original findings.


Redelmeier DA, Singh SM. Survival in Academy Award-winning actors and actresses. Ann Intern Med. 2001;134:955-962. Some of the widely reported almost-4-year longevity advantage over their ‘nominated but never won’ peers includes the immortal years between being nominated and winning. Moreover, use of date of birth as the primary time axis, even if doesn’t necessarily create any bias, does not provide a crisp causal contrast and shows researchers’ reluctance to divide up a person’s time. Despite the re-analysis by Sylvestre et al. and by others (refs) and the efforts of Forbes magazine, the continuing fascination with this finding, Harvard, NYTimes, this saga illustrates the persistence of errors and the difficulty in reversing flawed results

Source of data used in Section 5: see also: Clark, D.A., et al., "Cardiac Transplantation in Man. VI. Prognosis of Patients Selected for Cardiac Transplantation," Annals of Internal Medicine, 75 (July 1971), 15-21. [Stanford]

Mantel N, Byar DP. Evaluation of response-time data involving transient states—illustration using heart-transplant data. Journal of the American Statistical Association. 1974;69:81-6. In 1972, Gail had identified several biases in the first reports from Houston (15 transplanted, 41 not) and Stanford (). One was the fact that “patients in the [transplanted] group are guaranteed (by definition) to have survived at least until a donor was available, and his grace period has been implicitly added into [their] survival time”. Mantel was one of the first to suggest “statistical methodologies for avoiding (what he termed) this “time-to-treatment” bias, where “the survival of treated patients is compared with that of untreated controls results from a failure to make allowance for the fact that the treated patients must have at least survived from time of diagnosis to time of treatment, while no such requirement obtained for their untreated controls.” He introduced the idea of crossing over from one life table (“waiting for a transplant” state) to another (“post-transplant”) and make comparisons matched on day since entering the waitlist: “In principle there is no reason why the number of individuals at risk may not be increased by accessions of survivors from some other comparable study group […]. In the transient-state problem just such accessions do occur. Thus when a heart-transplant candidate receives his heart transplant, he becomes an accession into the transplanted group, though a loss from the untransplanted group.” Incidentally, Mantel’s choice of the word “guarantee” is not arbitrary: textbooks on survival data refer to a “guarantee time” such that the event of interest many not occur until a threshold time is attained. In oncology trials, a common error – usually referred to as “time-to-response” or “guarantee-time” bias -- is to attribute the longer survival of “responders” than “nonresponders” entirely to the therapy, and to ignore the fact that, by definition, responders have to live long enough for a response to be noted (Anderson ’83). Likewise, covariate autoantibodies. Finally, Cox time-dependent invented to deal with transplant data.

Clayton DG. For now, until I find the paper, Clayton and Hills section 15.3 page 143-145


Turnbull BW, Brown BW, Hu M. Survivorship analysis of heart-transplant data. Journal of the American Statistical Association. 1974;69:74-80. *Their permutation test motivated us to simulate, by lottery, an innocuous intervention (a prize that Oscar winners did not even know they had received, a sham intervention for those in the failed allograft study, ...). It is a useful technique for quantifying how much of the longevity advantage is an artifact (see Sylvestre at al.).*


Leibovici L. Effects of remote, retroactive intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial. BMJ. 2001;323: 1450-1. [PMID: 11751349]. *Reverse time scale, since ‘God does not necessarily work linearly’. There have been more than 90 responses to this article.*
Table: Ways to recognize immortal time

<table>
<thead>
<tr>
<th>Suggestion</th>
<th>Remarks / Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distinguish state from trait</td>
<td>A trait (e.g., blood group) is usually forever; people and objects move between states (on/off phone, intoxicated/not; on/off medication; failed allograft in place/removed)</td>
</tr>
<tr>
<td>Distinguish dynamic from closed population</td>
<td>Membership in a closed population (cohort) is initiated by an event (transition from a state) and is forever; in a dynamic one, it is for the duration of a state. Dynamic populations are the only option for studying transient exposures with rapid effects (e.g., cell phone/alcohol use vis a vis the rate of motor vehicle accidents).</td>
</tr>
<tr>
<td>Focus on person-time in index and reference categories, rather than on people in exposed and unexposed 'groups'</td>
<td>These refer to exposure categories, not to people per se; a person’s time may be divided between exposure categories; unless people remain in one category, it is misleading to refer to them as a ‘group’.</td>
</tr>
<tr>
<td>If authors used the term ‘group’, ask …</td>
<td>When and how did persons enter a ‘group’? Does being in or moving to a group have a time-related requirement? Is the classification a fixed one based on the status at time zero, or later? Is it sufficient to classify a person just once, or do we need to classify the “person-moments,” that is, the person at different times?</td>
</tr>
<tr>
<td>Sketch individual timelines</td>
<td>If there are two timescales, a Lexis diagram can help; use different notation for the time portion of the timeline where the event-rate of interest might be affected, and the portion where it can not (see Figures).</td>
</tr>
<tr>
<td>Measure the apparent longevity- or time-extending benefits of inert agents/interventions</td>
<td>After the fact, use a lottery to assign sham (and never actually delivered) interventions, but with same timing as the one under study. Or use actually-received agents with same timing.</td>
</tr>
<tr>
<td>Imagine this agent/intervention were being tested within a randomized trial.</td>
<td>How, and when after entry, would the agent be assigned? administered? How would event rates be computed? How would Farr have tested his ‘early-promotion’ suggestion?</td>
</tr>
<tr>
<td>Run time backwards</td>
<td>Amounts of person-time in the index or reference state are the areas under the time-curves of the numbers in these states.</td>
</tr>
</tbody>
</table>
Figure Legends

Figure 1
Hypothetical lifelines constructed to have an average mortality rate of 3785 deaths in (10,951 × 2.93 = 32,086) patient years (PY), i.e., 11.8 per 100PY (as in the nephrectomy actual study), but with no variation over years of follow-up, and no difference (other than random) between the states of the failed graft (“in place” or “removed”). We constructed the data by distributing the numbers of new cohort entries in a smooth decreasing pattern over the 11 calendar years, and applying the death rate of 11.8 per 100PY to the various resulting lengths of available follow-up, until the total number of deaths matched the reported 3785 and the number of PY of follow-up matched the reported 32,086. The 10,951 hypothetical lifelines (3785 completed, 7166 censored) were then ordered from shortest to longest. Finally, starting from the day of return to dialysis and working forward, each follow-up day a number of persons were chosen randomly from among those who had not already been selected, were still alive, and under observation on that day. These persons were designated to undergo sham interventions, the timings of which (3471 in all) were set so that the median and quartiles of the delay between return to dialysis and the intervention matched those in the article. The selections, made by computer in 2012, were made blindly, in a retroactive Leibovici-style lottery, applied in a forward direction, beginning in January 1994, to lifelines that had already run up to December 2004. Just as in Leibocini, and in Turnbull et al, the interventions could not have affected the comparative mortality rates. Shown are 30 such lifelines selected systematically from these 10,951 ordered hypothetical ones, with a completed lifeline indicated by a single straight line, and a censored one by a pair of lines forming an arrowhead. The timing of the sham intervention is indicated by an x, and the post-intervention patient years by red rather than grey boundary lines.
Hypothetical lifelines constructed to have an average mortality rate of 3785 deaths in (10,951 x 2.93 = 32,086) patient years (PY), i.e., 11.8 per 100PY (as in the actual nephrectomy study), but with no variation over years of follow-up, and no difference (other than random) between the states of the failed graft ('in place' or 'removed').

We constructed the data by distributing the numbers of new cohort entries in a smooth decreasing pattern over the 11 calendar years, and applying the death rate of 11.8 per 100PY to the various resulting lengths of available follow-up, until the total number of deaths matched the reported 3785 and the number of PY of follow-up matched the reported 32,086. The 10,951 hypothetical lifelines (3785 completed, 7166 censored) were then ordered from shortest to longest.

Finally, starting from the day of return to dialysis and working forward, each follow-up day a number of persons were chosen randomly from among those had not already been selected, were still alive, and being followed that day. These persons were designated to undergo a ‘partial nominectomy’ whereby, within the database, patients’ names were replaced by their initials. The timings of these sham interventions (3471 in all) were set so that the median and quartiles of the delay between return to dialysis and the intervention matched those in the article.

The selections, made by computer in 2012, were made blindly, in a retroactive lottery, applied in a forward direction, beginning in January 1994, to lifelines that had already run up to December 2004. Just as in Leibocini, and in Turnbull et al., we never communicated this list of selections to the patients or their physicians; moreover, since the interventions were limited to the computer file, they could not have affected the comparative mortality rates.

Timelines that end in a straight edge were ended by death
Timelines that end in an angled edge were censored
Changes from the 'initial to the 'post-nephrectomy' status are denoted by an 'x'
Figure 2

Mortality rates and rate ratios produced by the (A) mis- and (B) proper allocation of pre-‘intervention’ patient years. As explained in Figure 1, the hypothetical data for the 10,951 patients were constructed to have an average mortality rate of 3785 deaths in \((10,951 \times 2.93 = 32,086)\) patient years (PY) i.e., 11.8 deaths per 100PY (as in the actual study), but with no variation over years of follow-up, or between states (“no, or pre-‘intervention’” and “post-intervention”). Indeed, the selection of those who changed states was made at random, and retroactively.

In B, upper panel the number being followed at any time is < 3451 because some who had received the ‘intervention’ were already dead before the last ones received it.
**Immortal PY**

**Definition of...**
The experience of study subjects that is event free by definition  
Walker, p161

**Correct Handling of...**
The correct assignment of each increment in person-time of follow-up is to...  
THAT SAME EXPOSURE CATEGORY TO WHICH A DEATH WOULD BE ASSIGNED SHOULD IT OCCUR AT THAT TIME  
Breslow & Day, Vol II, page 83 on 'Allocation of person-time to time-dependent exposure categories'
Figure 3
Incorrect (at termination of follow-up, top panel) and correct (time-dependent, i.e., as follow-up time progresses, bottom panel) allocation of increments of follow-up time in the Duck et al. study. The lengths of the horizontal timelines in the top panel represent the total durations the individuals were exposed to vinyl chloride; 1538, 246 and 336 men worked with polyvinyl chloride for durations ranging from 0-10 years, 10-15 and 15+ years respectively (only a sample of timelines is shown). Duck et al. correctly calculated the observed numbers of deaths (83, 28 and 25) by assigning each death to the year-bracket (exposure category) in which it occurred. But their expected numbers of deaths (74.0, 26.9 and 41.3), and thus their 3 SMRs, were inappropriately based on numbers of man-years (13697, 3,271 and 6,084) which they obtained by assigning each man’s total (ie ultimate) duration of exposure to these 3 brackets. Wagoner, Infante and Saracci correctly explained, “If in a follow-up study one selects, for example, a subgroup of workers by the fact that they have achieved at least 15 years’ exposure, then none of these workers could have died [in today’s terminology, they are ‘immortal’] before the 15th anniversary, so information on risk of dying can only come from the number of man-years at risk and the number of deaths after 15 years. Of course these same men, provided they are properly regrouped together with those dying or coming to the end of the follow-up between, for example, 10 and 14 years can provide similar information for this time-interval-and so on for all previous time-intervals.” Wagoner, Infante and Saracci’s correct re-allocation (regrouping) of PYs, and dramatically different pattern of re-calculated SMRs, are shown in the lower panel (see also the small worked example, motivated by this example, in Breslow and Day). What makes this example more striking than many others is that the compared (exposure) states are directly defined by amounts of time.
Inappropriate assignment of (immortal) PY to ULTIMATE exposure categories:
- 2460 immortal PY in '0-10' category assigned to '10-15' category
- 3360 immortal PY in '0-10' category assigned to '15-20' category
- 1680 immortal PY in '10-15' category assigned to '15-20' category

Duck, Carter, and Coombes; Lancet 1975, ii, 1197.

The CORRECT assignment of each increment in person-time of follow-up is to...
THAT SAME EXPOSURE CATEGORY TO WHICH A DEATH WOULD BE ASSIGNED SHOULD IT OCCUR AT THAT TIME

Breslow & Day, Vol II, page 83
‘Allocation of person-time to time-dependent exposure categories’
Figure 4

A. Incorrect (after the fact) allocation of follow-up time in the Houston heart-transplant series, leading to misleading mortality rates and rate ratio. The daily amounts of patient time (days) in the transplanted (n=15, top) and non-transplanted (n=42, bottom) ‘groups’ are enclosed within polygons. Individual deaths (13 and 27) are denoted by dots within these. The 311 ‘immortal’ pre-transplant patient days (PDs) are included in the denominator of the mortality rate for the transplant ‘group’ whereas only those PDs in the shaded area should be; thus, the death rate of 13/1667PD = 0.78/100PD is too low. So readers can do their own analyses, the inset contains the relevant raw data for the 57 patients. The records are ordered from top to bottom within a block, and from left to right block, by how long they waited for an allograft, or if they died before they could receive one, how long they been waiting. Transplants are indicated by the letter “t” and (again) deaths by dots. The 15 patients in the allograft “group” waited an average of 22 days (IQR 2-38) before an allograft became available; thus, when they changed their status from “on waitlist” to “post-transplant”, they had been “immortal” for a total of 311 days.

B. Correct (as time progresses) allocation of follow-up time. The daily amounts of patient time in the transplanted (1356PD, top) and non-transplanted (3422PD, bottom) ‘states’ are enclosed within polygons; individual deaths (13 and 27) are denoted by dots within these. The higher death rate of 0.96/100PD in the transplanted state reflects the smaller denominator, while the higher rate in the non-transplanted state (0.79/100PD) reflects the larger denominator. Shown in B(1) are the time-matched Mantel-Haenszel numerators and denominators, using 30-day time-slices or ‘strata.’ The quotient of their sums, i.e., (2.5+1.5+ … + 0)/(3.0+1.0+ … + 0) yields a summary (overall) mortality-ratio of 9.3/5.5 = 1.70. The corresponding M-H rate ratios for different window-widths, ranging from 270 (‘crude’) on the left, to the finest time-stratification, with ‘1 day wide’ windows, on the right, are shown in inset B(2). Also shown for the same window-widths are the corresponding rate ratios fitted by (both conditional and unconditional) Poisson regression. The fitting of a rate ratio by the Cox PH model involves 31 risksets -- as many ultra-fine strata as there are distinct times of death. Inset B(3) shows rate-ratio functions fitted by the extended Cox model using a dataset of 4,778 observations, one for each day for each patient.