15. Moonesinghe SR, Mythen MG, Grocott MP. High-risk surgery: epidemiology and outcomes. *Anesthesia and Analgesia* 2011; **112**: 891–901.
16. Dripps RD, Lamont A, Eckenhoff JE. The role of anesthesia in surgical mortality. *Journal of the American Medical Association* 1961; **178**: 261–6.
17. Sutton R, Bann S, Brooks M, Sarin S. The Surgical Risk Scale as an improved tool for risk-adjusted analysis in comparative surgical audit. *British Journal of Surgery* 2002; **89**: 763–8.
18. Grocott MP, Pearse RM. Prognostic studies of perioperative risk: robust methodology is needed. *British Journal of Anaesthesia* 2010; **105**: 243–5.
19. Khuri SF, Henderson WG, Daley J, et al. Successful implementation of the Department of Veterans Affairs National Surgical Quality Improvement Program in the private sector: the Patient Safety in Surgery study. *Annals of Surgery* 2008; **248**: 329–36.
20. Barry MJ, Edgman-Levitan S. Shared decision making – pinnacle of patient-centered care. *New England Journal of Medicine* 2012; **366**: 780–1.
21. Pearse RM, Moreno RP, Bauer P, Pelosi P, Metnitz P, Spies C, Vallet B, Vincent JL, Hoeft A, Rhodes A; European Surgical Outcomes Study (EuSOS) group for the Trials groups of the European Society of Intensive Care Medicine and the European Society of Anaesthesiology. *Lancet* 2012; **380**: 1059–65.
22. Pandit JJ. The national strategy for academic anaesthesia. A personal view on its implications for our specialty. *British Journal of Anaesthesia* 2006; **96**: 411–4.
23. Donabedian A. Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly* 1966; **44**: 166–206.
24. Grocott MPW. Improving outcomes after surgery. *British Medical Journal* 2009; **339**: b5173.
25. Mythen M. Fit for surgery? *Anesthesia and Analgesia* 2011; **112**: 1002–4.

doi:10.1111/anae.12063

Editorial

Understanding immortal time bias in observational cohort studies

"*On a long enough time line, the survival rate for everyone drops to zero.*"

– Chuck Palahniuk (American author), *Flight Club*

Observational cohort studies are an important part of our research armamentarium, but are prone to biases. One bias – immortal time bias (alias survivor(ship) bias) – has been found to be quite prevalent in such studies [1, 2]. It is created when there exists a period of time during which the outcome of interest (e.g. death, stroke, delirium) for one of the cohorts cannot possibly occur [3].

Tea, anyone?

A ridiculous example of immortal time bias would be if seriously injured patients admitted to the intensive care unit (ICU) were offered tea and crumpets (T&C) on day 3. Subsequent analysis would likely find that T&C recipients not only have lower mortality, but fare better overall than non-recipients. T&C therapy is thus found to benefit ICU patients! Of course it is a ruse: all we have done with T&C is to select those patients destined to do well.

Numerical example

One hundred patients with angina are offered an experimental procedure called CABG. Fifty accept while 50 decline. The waiting time is 24 months, during which patients die at a rate of two per month. After 24 months, 48 patients have died. Of the remaining 52 patients, 26 undergo CABG at 24 months and 26 do not. Let us assume that CABG makes no difference, such that *all* the remaining 52 patients continue to die at two per month. Let us add up the person-months in each group. In the non-CABG group, there were the 48 who died during the first 24 months ($48 \times 12 = 576$ person-months), plus the 26 who died later ($26 \times (12 + 13) = 650$ person-months). In all, the non-CABG group of 74 enjoyed 1226 person-months, giving an average longevity of 1226/74 = 16.6 months. The CABG group had 26 patients who lived an average of $12 + 13 = 25$ months (650 person-months). Thus, for a therapy that confers no benefit or harm, we find a survival improvement of 8.4 months per person. The reason: those who died early were categorised in the non-CABG cohort.

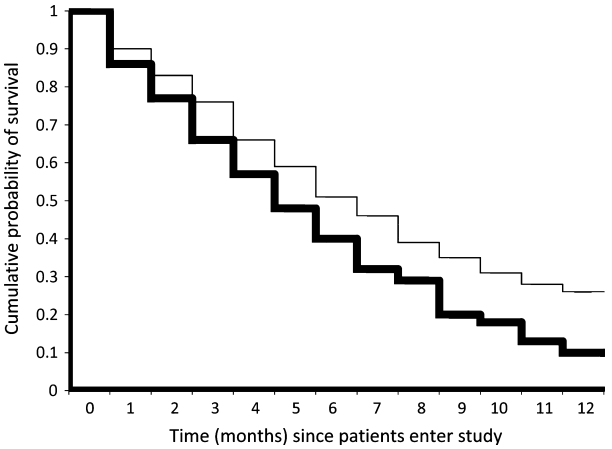

Figure 1 Typical Kaplan–Meier survival curves for two treatments, A (thick line) and B (thin line).

Eliminating immortal time bias

Immortal time bias in our CAGB study could be eliminated by comparing only those 52 patients surviving the waiting period. However, it is likely that not all 26 surgical candidates would undergo CABG immediately after the waiting period, and so there will still be some patients who die while waiting, and would be categorised into the non-CABG cohort.

Unfortunately, eliminating the waiting time data may create another form of bias. Let us assume that CABG *does* prolong life, but only in patients with severe disease. Since patients with severe disease are likely to die earlier, eliminating the waiting time leaves one with an overall group of less sick patients in whom any benefit of CAGB is less likely to be uncovered [2].

Another way to eliminate the bias is to compare patients who have received an intervention with a matched historic cohort. Unfortunately, secular trends and advances in management techniques over time may confound.

Survival curves

The way to deal with immortal time bias begins with using Kaplan-Meier survival curves (Fig. 1). The horizontal axis is time since a patient enters the study. The vertical axis is the cumulative probability of survival, and has no units. Consider patients exposed to treatment A or treatment B. Group A (thick line) starts with 100 patients. Now, if 86 survive the first month (14 deaths), the proportion surviving is 0.86. Our survival curve starts at time 0 always at the value 1.00 (pretending that 100% of the patients enter the study simultaneously). After one month the curve drops to 0.86. Let's say 77 survive the next month, so the probability of survival is 77/86, or 0.9. The probability of surviving the first two months is $0.86 \times 0.9 = 0.77$. If in the third month 11 people die, the survival probability is now 66/77, or 0.86. The probability of surviving all three months is $0.86 \times 0.9 \times 0.86 = 0.66$. We repeat the calculations for each subsequent month, and do the same for group B which has 60 patients (Fig. 1; thin line). Note that group B does not have to start with the same number of patients as group A. Starting at time 0 with 100% of the cohort, group B will also have a value of 1 on the y-axis at time 0. During the first month, if 90% of group B survive (6 deaths), the survival curve for group B drops to 0.9. The same process continues until the end of the follow-up period of 24 months.

We pause here to introduce the concept of censoring. In studies, patients exit a group for reasons other than death. In constructing survival curves, all patients who exit a group should be censored from further calculations. If four patients exit group A after the first month, the number of patients present during the second month becomes 100–14–4.

We go back to our curves for treatments A and B. To determine if the apparent superiority of B is significant, we employ the log-rank test. For each time period one notes the *total* number of deaths in both groups for that period. In the first month, we have 20 deaths. If the null hypothesis is true, we expect deaths in each month to be distributed proportionally to the number of patients in each group. Starting with the first month there are $100 + 60 = 160$ patients in total, so we expect that $(100/160) \times 20$ patients should come from group A, and $(60/160) \times 20$ patients should come from group B. We do this calculation for each month, and then for each group add up the total expected and actual deaths. One then does a chi-squared test

(for testing between proportions) with (for two treatments) one degree of freedom, to obtain a p value.

What does this approach do to our T&C study? Without survival analysis, the T&C cohort appears to have a survival advantage from the start up to the end of the immortal time period (3rd day). (Indeed, one should be wary whenever one sees a purported survival difference between two Kaplan-Meier curves in an observational cohort study occurring mainly in the beginning.) When such curves are drawn, it will become apparent beyond day 3 that this survival advantage has greatly diminished. However, it will also be found that the T&C cohort *still* does better beyond day 3. Indeed, in order to enjoy T&C, a person must be rather well. Thus there is a selection bias that can only be dealt with by using a proper study design, such as comparing T&C therapy against a similar sub-cohort of patients who are actually well enough to receive it, but for some reasons do not.

Kaplan–Meier curves and log-rank analysis are powerful tools. However, further improvement can be achieved with a more sophisticated technique: 'Cox regression with treatment managed as a time-dependent covariate'.

Cox regression

With some imagination, one can see that survival curves are not straight lines, but look like exponential functions. Somewhat akin to the so-called 'least squares' fitting of data to a straight line, Cox regression fits the data to an exponential of a 'hazard function'. It is called a hazard function because it implies that something bad, such as death, may happen. Cox regression usually also has a 'treatment function', meaning something good, such as *increased* survival, may result, and sometimes other functions to represent, say, the sex or age of the patients. Cox regression can deal with patients who switch between the intervention and non-intervention groups during the study (i.e. the intervention is a time-dependent covariate). Possible lingering effects of the intervention or non-intervention's being carried over into the other cohort during the next time interval can also be dealt with using weighted mathematical functions to account for the average time each subject spends in each category. Although understanding Cox regression helps one to interpret observational cohort studies, merely detecting immortal time bias in these studies requires much less mathematical aptitude.

Clinical examples

The literature contains many examples of potentially erroneous conclusions resulting from immortal time bias. For example, two studies on cardiac transplantation from the 1970s, in which deaths that occurred while waiting for transplantation were categorised into the non-transplant cohort, exaggerated the advantage of transplantation because of this bias [4].

Ely et al. [5] found that ICU patients with delirium had longer stays in ICU. However, treating delirium in Ely et al.'s data as a time-dependent covariate, Shintani et al. found that delirium did not lead to a longer ICU stay [6]. In another ICU study, Jaber et al. [7] checked for cytomegalovirus (CMV) antigenaemia and found that antigenaemic patients had longer ICU stays. However, immortal time bias might have been present since patients discharged from the ICU before 20 ± 12 days of admission, the time when antigenaemia was detected, might have had less chance of antigenaemia detection and were thus more likely to have been categorised in the non-antigenaemic cohort. Thus, developing delirium or CMV antigenaemia does not necessarily lead to longer ICU stay, but longer ICU stay increases the chance of eventually having delirium or CMV antigenaemia documented at some point.

In another example, Walkey et al. [8] found that mortality was increased in severely septic patients with new-onset atrial fibrillation. However, since patients who died before fibrillation developed were categorised in the non-fibrillation group, that degree of association might have been underestimated [8].

Martin et al. [9] categorised diabetics as having self-monitored their blood glucose if they had done so for at least one year during a variable study period, and those who had failed to self-monitor for at least one year for various reasons (including death) as not having self-monitored. As was later pointed out, their conclusion that self-monitoring conferred an impressive 51% all-cause reduction in mortality over a mean follow-up of 6.5 years was rather implausible [10].

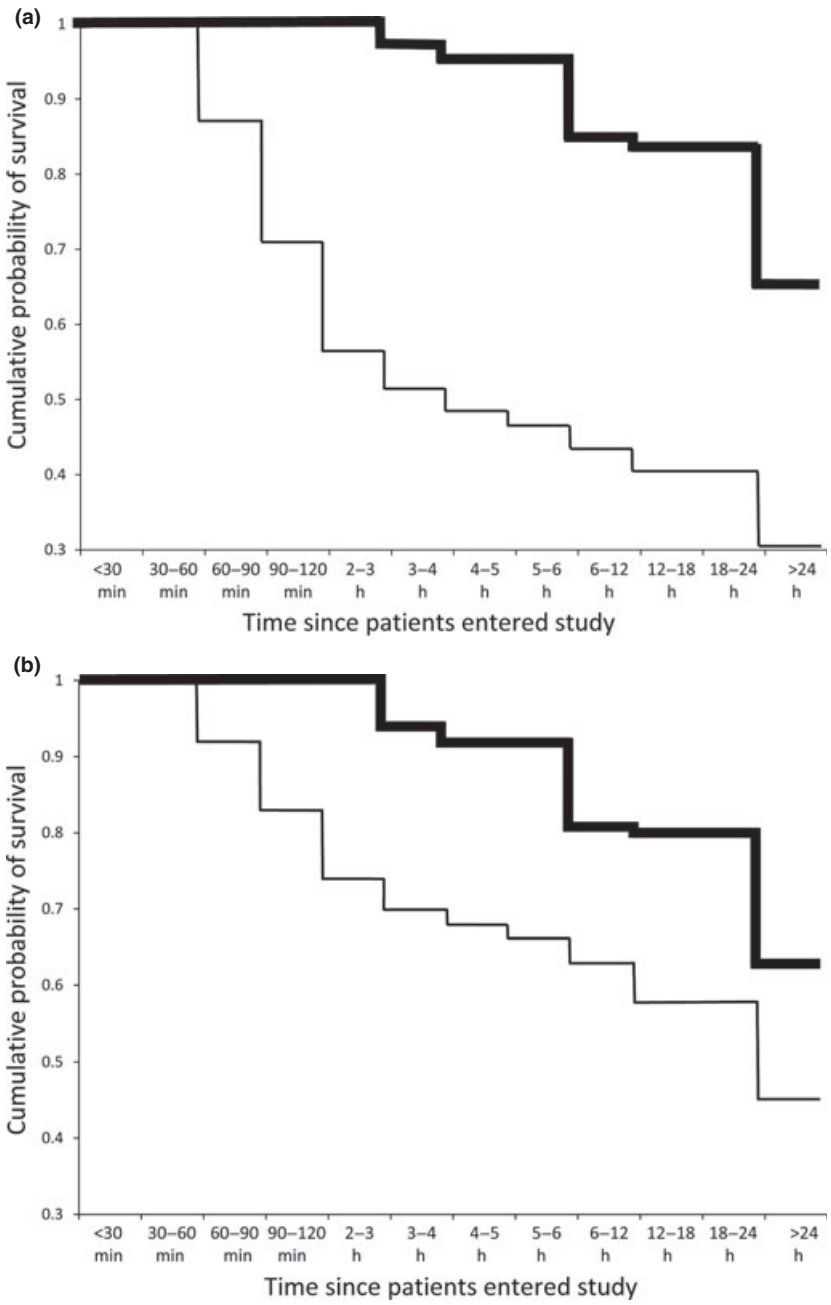

Figure 2 Kaplan–Meier survival curves of trauma patients [12] requiring massive transfusion, for high (thick) and low (thin) fresh frozen plasma: packed red blood cell ratio; a) ratio not treated as a time-dependent covariate ($p < 0.001$); b) ratio treated as a time-dependent covariate ($p = 0.17$).

Finally, in trauma requiring massive transfusion, fresh frozen plasma (FFP) typically is minimally given during the first few hours of resuscitation whereas packed red blood cells (PRBC) are usually abundantly available. As more FFP becomes available, the FFP:PRBC ratio approaches ~1:1 as resuscitation continues, *if* the patient is still alive, typically hours later. By categorising early-deaths into the 'non-1:1' cohort and those surviving at least long enough to reach '1:1' into the '1:1' cohort, many authors have found that '1:1' improves survival [2, 11]. However, with the FFP:PRBC ratio of ~1:1 treated as a time-dependent covariate using Cox regression modeling, Snyder et al. [12] found in their study that a survival advantage, suggested when analysed with the flawed methodology described above, vanished. In other words, patients did not necessarily die because they did not receive enough FFP, but rather, they did not receive enough FFP because they died. We have re-analysed Snyder et al.'s data and plotted two pairs of survival curves to illustrate how treating '1:1' as a time-dependent covariate circumvents the problem of immortal time bias (Fig. 2).

(Common) sense and sensibility

Even good studies can be unfairly dismissed once such bias has been uncovered in related studies. In our trauma example, unless one has followed the literature carefully, one may assume that *all* such studies are flawed. However, we have found after reviewing 26 observational studies of the effect of a FFP:PRBC ratio of ~1:1 in massive hemorrhage that 15 of them were free of immortal time bias, of which 10 actually showed '1:1' to be beneficial [2]. We now eagerly await the results of randomised controlled trials on the subject. Meanwhile, we have in place a '1:1' policy for serious trauma patients requiring massive blood transfusion at the Prince of Wales Hospital, having analysed all '1:1'

studies and reasoned that massive loss of blood needs replacement with its equivalent (roughly equal number of units of PRBC, FFP and platelets) [2, 13, 14]. One must always apply common sense. Even if the claim that delirium prolongs ICU stay is diminished because of potential bias, the negative implications of ICU delirium on hospital stay and mortality cannot be ignored. In the diabetes example, even though the study might have been tainted by the possibility of bias, the merits of vigilant self-monitoring of blood glucose have not. Fortunately, the merits of heart transplantation were not discarded despite early flawed studies. As for tea and crumpets: well, such a fine tradition is always therapeutic, anytime, anywhere.

Competing interests

No external funding and no competing interests declared.

A. M.-H. Ho
Professor
Department of Anaesthesia and Intensive Care
The Chinese University of Hong Kong
Shatin, Hong Kong
Email: hoamh@yahoo.com
P. W. Dion
Staff Anaesthesiologist
Department of Anaesthesia
St. Catharines General Hospital
St. Catharines
Ontario, Canada
C. S. H. Ng
Associate Consultant
Department of Surgery
Prince of Wales Hospital
Shatin, Hong Kong
M. K. Karmakar
Associate Professor
Department of Anaesthesia and Intensive Care
The Chinese University of Hong Kong
Shatin, Hong Kong

References

1. van Walraven C, Davis D, Forster AJ, et al. Time-dependent bias was common in survival analyses published in leading clinical journals. *Journal of Clinical Epidemiology* 2004; **57**: 672–82.
2. Ho AMH, Dion PW, Critchley LAH, et al. Prevalence of survivor bias in observational studies on the plasma:erythrocyte ratios for trauma patients requiring massive transfusion. *Anesthesiology* 2012; **116**: 716–28.
3. Suissa S. Immortal time bias in pharmacoepidemioogy. *American Journal of Epidemiology* 2008; **167**: 492–9.
4. Gail MH. Does cardiac transplantation prolong life? A reassessment. *Annals of Internal Medicine* 1972; **76**: 815–7.
5. Ely EW, Shintani A, Truman B, et al. Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. *Journal of the American Medical Association* 2004; **291**: 1753–62.
6. Shintani AK, Girard TD, Arbogast PG, et al. Immortal time bias in critical care research: application of time-varying Cox regression for observational cohort studies. *Critical Care Medicine* 2009; **37**: 2939–45.
7. Jaber S, Chanques G, Borry J, et al. Cytomegalovirus infection in critically ill patients. Associated factors and consequences. *Chest* 2005; **127**: 233–41.
8. Walkey AJ, Wiener RS, Ghobrial JM. Incident stroke and mortality associated with new-onset atrial fibrillation in patients hospitalized with severe sepsis. *Journal of the American Medical Association* 2011; **306**: 2248–55.
9. Martin S, Schneider B, Heinemann L, et al. Self-monitoring of blood glucose in type 2 diabetes and long-term outcome: an epidemiological cohort study. *Diabetologia* 2006; **49**: 271–8.
10. Hoffmann F, Andersohn F. Immortal time bias and survival in patients who self-monitor blood glucose in the Retrlective Study: self-monitoring of blood glucose and outcome in patients with type 2 diabetes (ROSSO). *Diabetologia* 2011; **54**: 308–11.
11. Ho AMH, Dion PW, Yeung JHH, et al. Simulation of survivorship bias in observational studies on plasma to red blood cell ratios in massive transfusion in trauma. *British Journal of Surgery* 2012; **99**(Suppl 1): 132–9.
12. Snyder CW, Weinberg JA, McGwin G Jr, et al. The relationship of blood product ratio to mortality: survival benefit or survival bias? *Journal of Trauma* 2009; **66**: 358–62.
13. Rainer TH, Ho AMH, Yeung JHH, et al. Early risk stratification of patients with major trauma requiring massive blood transfusion. *Resuscitation* 2011; **82**: 724–9.
14. Ho AMH, Dion PW, Ng CSH, et al. Fresh frozen plasma transfusion strategy during major trauma resuscitation: common sense and sensibility. *Resuscitation* 2010; **81**: 1079–81.

doi:10.1111/anae.12120

