

Are competing risks models appropriate to describe implant failure?

Adrian SAYERS^{1,2}, Jonathan T EVANS¹, Michael R WHITEHOUSE^{1,3}, and Ashley W BLOM^{1,3}

¹ Musculoskeletal Research Unit, Bristol Medical School, Southmead Hospital, Bristol, ² Population Health Sciences, Bristol Medical School, Bristol, ³ National Institute for Health Research Bristol Biomedical Research Centre, University of Bristol, UK

Correspondence: adrian.sayers@bristol.ac.uk

Submitted 2017-12-08. Accepted 2018-02-06.

Background and purpose — The use of competing risks models is widely advocated in the arthroplasty literature due to a perceived bias in comparison of simple Kaplan–Meier estimates. Proponents of competing risk models in the arthroplasty literature appear to be unaware of the subtle but important differences in interpretation of net and crude failure estimated by competing risk and Kaplan–Meier methods respectively.

Methods — Using a simple simulation we illustrate the differences between competing risks and Kaplan–Meier methods.

Results — Competing risk and Kaplan–Meier methods estimate different survival quantities, i.e., crude and net failure respectively. Estimates of crude failure estimated using competing risk methods will be less than net failure as estimated using Kaplan–Meier methods.

Interpretation — Kaplan–Meier methods are appropriate for describing implant failure, whereas crude survival estimated using competing risk methods estimates the risk of surgical revision as it depends on both implant failure and mortality. Both competing risk models and Kaplan–Meier methods are useful in arthroplasty, and both provide unbiased estimates of crude and net failure in the absence of any confounding or selection respectively. Surgeons and researchers should carefully consider whether the use of competing risks is always justified. Lower estimates of failure from competing risk models may be misleading to surgeons who are attempting to select the best implants with the lowest failure rates for their patients.

We have recently noticed a number of incidences in the arthroplasty literature of authors espousing the benefits of using competing risk models in preference to Kaplan–Meier (KM) estimates to describe the failure of implants due to a perception that the observed high mortality rates in elderly patients may lead to biased estimates using the KM method (Biau et

al. 2007, Fennema and Lubsen 2010, Keurentjes et al. 2012, Lacny et al. 2015, Porcher 2015, Wongworawat et al. 2015, Martin et al. 2016, Lampropoulou-Adamidou et al. 2017). This recent trend is somewhat worrying as we believe there is a fundamental misinterpretation of what Kaplan–Meier (KM) (Kaplan and Meier 1958) or competing risks (CR) (Coviello and Boggess 2004) models estimate, and under which circumstances each method may be preferable.

To correct this misunderstanding, we describe a simple simulation in a hypothetical situation with immortal patients, where no individuals are ever lost to follow-up. Figure 1 panel (a) illustrates this process using a line plot which illustrates when a patient becomes at risk and when a failure occurs and exits the study. In this situation, it is very easy to estimate implant survival at a time of interest, i.e., it is simply the proportion of those who fail. The numerator is the number of failures, and the denominator is the number of patients implanted. A simple proportion, KM estimates (Kaplan and Meier 1958), and the cumulative incidence function (CIF) (Coviello and Boggess 2004) from a CR model will give identical estimates. This scenario is the ideal scenario, as we need not concern ourselves with problems such as censoring (loss to follow-up or mortality), and we describe these estimates of failure as net failure, using the terminology of Lambert et al. (2010).

However, some researchers are under the misguided belief that this hypothetical situation is the only scenario in which the KM estimator is appropriate (Biau et al. 2007). The title of Kaplan and Meier's (1958) seminal work, "Nonparametric Estimation from Incomplete Observations," gives us a clue to why this is incorrect. The KM method was specifically developed to allow incomplete observations due to non-informative right censoring, i.e., individuals cease to be at risk of failure, but have not failed where the reason that they cease to be at risk is completely independent of the cause of failure.

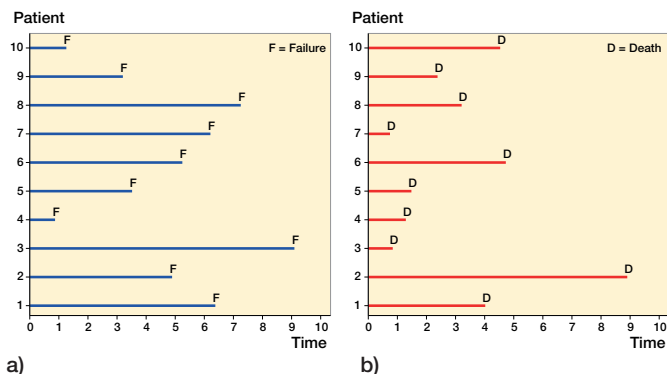


Figure 1. Panel (a) is a line plot that illustrates the time at risk of 10 patients entering a study following arthroplasty (time 0) and exiting the study after failure where the only possible mechanism of exiting the study is failure, i.e., no other cause of censoring occurs. Panel (b) is a line plot that illustrates a non-informative mortality profile of the same 10 patients entering a study following arthroplasty (time 0).

In arthroplasty failure studies, mortality is one possible cause of being censored. Figure 1 panel (b) illustrates a non-informative mortality profile of patients in Figure 1 panel (a).

In this more complex and alternate situation with mortal patients, the failure process is more difficult to estimate due to the presence of a mortality process. This additional process removes patients from the study and calculation of failure becomes more complex—see Figure 2 which overlays the failure and mortality processes.

Due to the complexity of this alternate situation with mortal patients, we are confronted with a choice of what to estimate. We can attempt to recover an estimate of net failure, which gives us an estimate of the failure of the implant, i.e., the failure estimate from the immortal cohort. Or, we can estimate crude failure, which represents the likely number of failures we see in practice, i.e., it is a composite of both the failure of the implants and the mortality process. The terminology used in this field is somewhat heterogeneous, therefore we use the terminology described by Lambert et al. (2010).

Standard methods of conducting survival analysis, i.e., KM or Cox regression focus on net failure, are based solely on the hazard profile of the cause of interest. Competing risk methods estimate crude failure and depend on both the hazard of the event of interest and the hazard of the competing event.

The differences in the KM estimate with immortal patients and mortal patients and the CIF (competing risks estimate) with mortal patients is presented in Figure 3. Here, we simply create 2 independent random uniform failure profiles between 0 and 10 years for 2 processes, (1) implant failure, and (2) mortality for 1,000 patients. Analysis of implant failure of immortal patients, ignoring the mortality process, can be considered the “truth,” and removing patients from the risk set due to a mortality event creates a mortal cohort, i.e., the observed. We expect the failure to be 100% at 10 years, and a straight line from 0 years to 10 years, i.e., a 45-degree line.

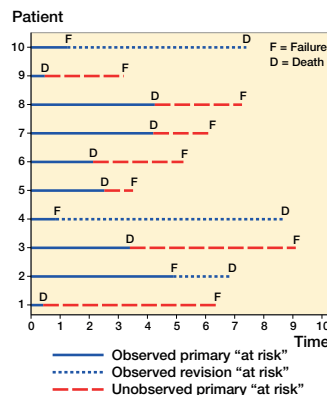


Figure 2. A line plot that illustrates the time at risk of 10 patients entering a study following arthroplasty (time 0) and the combination of a failure and mortality mechanism, i.e., mortal patients.

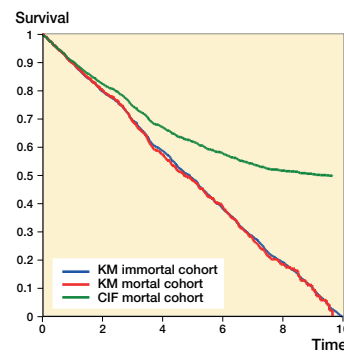


Figure 3. KM survival curves and the 1 minus the cumulative incidence function in mortal and immortal cohorts.

This clearly illustrates the CIF (competing risks estimate) is not the same as that of KM. It is a biased estimate of net failure, but an unbiased estimate of crude failure. Whilst the simulation is extreme, i.e., everyone fails and everyone dies, the results will hold in all circumstances that the censoring is non-informative. The degree to which the CIF is different from the KM profile depends on the mortality process. Prior to the first mortality event, KM and CIF are equal, and only following the first mortality event do they become unequal. In arthroplasty research differences between KM and CIF are likely to be more evident in series with long-term follow up, where mortality is inevitably higher, or in series with elderly or frail patients.

These differences are well known to those with a methodological interest in survival analysis. For example, Gooley et al. (1999) note that if one is interested in evaluating a cause-specific failure, the CIF may be misleading and inferences should be made from functions which are based solely on the hazard of failure from the cause of interest, i.e., use the KM estimator. Putter et al. (2007) similarly state that the “naive Kaplan–Meier estimator describes what would happen if the competing event could be prevented to occur, creating an imaginary world in which an individual remains at risk of failure from the event of interest,” i.e., an immortal patient cohort. Ranstam et al. (2011) describe this in an arthroplasty setting as the “implicit assumption that the patient will be alive until the implant fails.” Recently, we have similarly illustrated this result using a simulation study in the context of prosthesis benchmarking: we illustrate that KM provides unbiased estimates of net failure and provide nominal coverage, i.e., the confidence interval includes the true value on 95% of occasions (Sayers et al. 2017).

In as far as we currently know, the mortality process is independent of whether implants are revised or not, i.e., mortality satisfies the non-informative censoring assumption. Our

belief in this assumption is based on the observation that even when an implant or group of implants fail in a large number of patients, e.g., metal on metal, this is not associated with any increase in pathologies, in the short term, such as cancer that in turn may lead to an excess of mortality (Smith et al. 2012a, 2012b, 2012c). However, it is important these assumptions are checked periodically; an absence of evidence is not evidence of absence, and future information may require analyses to be modified to account for an informative censoring profile.

Simply, competing risk methods and non-competing risk methods estimate different quantities, and which quantity you should use depends on your application of interest. If you are interested in describing the failure of an implant, comparing the failure rate of a group of implants, looking for outliers, i.e., from a regulatory perspective, or attempting to select an implant for use that has the greatest longevity, you need estimates of net failure (KM). If you are interested in resource planning, health economics, or communicating with patients their likely chance of experiencing a revision, estimates of crude failure (CR) are more likely to be desirable.

Just because the estimate of net implant failure is higher than crude failure does not mean they are not correct or desirable in many circumstances in arthroplasty. However, it also important to remember that whilst KM and the CIF are statistically unbiased estimates for net and crude failure respectively, they are both equally likely to display bias in the presence of confounding factors and selection effects, and simply choosing the appropriate approach is not a panacea against this immutable problem.

Funding and conflict of interest

AS was supported by a MRC strategic skills fellowship: MRC Fellowship MR/L01226X/1. JTE was supported by the National Joint Registry of England, Wales, Northern Ireland and the Isle of Man and Royal College of Surgeons of England Fellowship.

This study was supported by the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health.

We have no competing interests to declare.

See also Editorial in the June 2018 issue of Acta Orthopaedica.

Acta thanks Nicole Pratt and other anonymous reviewers for help with peer review of this study.

AS, JTE, MRW, AWB conceived the manuscript, interpreted data from simulation, and approved the final version of the manuscript. AS wrote the first draft and performed the simulation. JTE and AS reviewed the literature.

- Biau D J, Latouche A, Porcher R. Competing events influence estimated survival probability: when is Kaplan–Meier analysis appropriate? *Clin Orthop Relat Res* 2007; 462: 229–33. doi: 10.1097/BL0.0b013e3180986753.
- Coviello V, Boggess M. Cumulative incidence estimation in the presence of competing risks. *Stata J* 2004; 4(2): 103–11.
- Fennema P, Lubsen J. Survival analysis in total joint replacement: an alternative method of accounting for the presence of competing risk. *J Bone Joint Surg Br* 2010; 92(5): 701–6. doi: 10.1302/0301-620X.92B5.23470.
- Gooley T A, Leisenring W, Crowley J, Storer B E. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med* 1999; 18(6): 695–706.
- Kaplan E L, Meier P. Nonparametric-estimation from incomplete observations. *J Am Stat Assoc* 1958; 53(282): 457–81. doi: 10.2307/2281868.
- Keurentjes J C, Fiocco M, Schreurs B W, Pijls B G, Nouta K A, Nelissen R G. Revision surgery is overestimated in hip replacement. *Bone Joint Res* 2012; 1(10): 258–62. doi: 10.1302/2046-3758.110.2000104.
- Lacny S, Wilson T, Clement F, Roberts D J, Faris P D, Ghali W A, Marshall D A. Kaplan–Meier survival analysis overestimates the risk of revision arthroplasty: a meta-analysis. *Clin Orthop Relat Res* 2015; 473(11): 3431–42. doi: 10.1007/s11999-015-4235-8.
- Lambert P C, Dickman P W, Nelson C P, Royston P. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Statistics in Medicine* 2010; 29(7–8): 885–95. doi: 10.1002/sim.3762.
- Lamproulou-Adamidou K, Karachalios TS, Hartofilakidis G. Overestimation of the risk of revision with Kaplan–Meier presenting the long-term outcome of total hip replacement in older patients. *Hip Int* 2017; [Epub ahead of print]. doi: 10.5301/hipint.5000575.
- Martin C T, Callaghan J J, Gao Y B, Pugely A J, Liu S S, Warth L C, Goetz D D. What can we learn from 20-year followup studies of hip replacement? *Clin Orthop Relat Res* 2016; 474(2): 402–7. doi: 10.1007/s11999-015-4260-7.
- Porcher R. CORR Insights((R)): Kaplan–Meier survival analysis overestimates the risk of revision arthroplasty: a meta-analysis. *Clin Orthop Relat Res* 2015; 473(11): 3443–5. doi: 10.1007/s11999-015-4291-0.
- Putter H, Fiocco M, Geskus R B. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med* 2007; 26(11): 2389–430. doi: 10.1002/sim.2712.
- Ranstam J, Karrholm J, Pulkkinen P, Makela K, Espehaug B, Pedersen A B, Mehnert F, Furnes O, NARA study group. Statistical analysis of arthroplasty data. II: Guidelines. *Acta Orthop* 2011; 82(3): 258–67. doi: 10.3109/17453674.2011.588863.
- Sayers A, Crowther M J, Judge A, Whitehouse M R, Blom A W. Determining the sample size required to establish whether a medical device is non-inferior to an external benchmark. *BMJ Open* 2017; 7(8): e015397. doi: 10.1136/bmjopen-2016-015397.
- Smith A J, Dieppe P, Howard P W, Blom A W, National Joint Registry for England and Wales. Failure rates of metal-on-metal hip resurfacings: analysis of data from the National Joint Registry for England and Wales. *Lancet* 2012a; 380(9855): 1759–66. doi: 10.1016/S0140-6736(12)60989-1.
- Smith A J, Dieppe P, Porter M, Blom A W, National Joint Registry of England and Wales. Risk of cancer in first seven years after metal-on-metal hip replacement compared with other bearings and general population: linkage study between the National Joint Registry of England and Wales and hospital episode statistics. *BMJ* 2012b; 344: e2383. doi: 10.1136/bmj.e2383.
- Smith A J, Dieppe P, Vernon K, Porter M, Blom A W, National Joint Registry of England and Wales. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. *Lancet* 2012c; 379(9822): 1199–204. doi: 10.1016/S0140-6736(12)60353-5.
- Wongworawat M D, Dobbs M B, Gebhardt M C, Gioe T J, Leopold S S, Manner P A, Rimnac C M, Porcher R. Editorial: Estimating survivorship in the face of competing risks. *Clin Orthop Relat Res* 2015; 473(4): 1173–6. doi: 10.1007/s11999-015-4182-4.