



Missingness in the Setting of Competing Risks: from Missing Values to Missing Potential Outcomes

Bryan Lau^{1,2} · Catherine Lesko¹

Published online: 19 March 2018
© Springer International Publishing AG, part of Springer Nature 2018

Abstract

Purpose of Review The setting of competing risks in which there is an event that precludes the event of interest from occurring is prevalent in epidemiological research. Unless studying all-cause mortality, any study following up individuals is subject to having a competing risk should individuals die during time period that the study covers. While there are prior papers discussing the need for competing risk methods in epidemiologic research, we are not aware of any review that discusses issues of missing data in a competing risk setting.

Recent Findings We provide an overview of causal inference in competing risks as potential outcomes are missing, provide some strategies in dealing with missing (or misclassified) event type, and missing covariate data in competing risks. The strategies presented are specifically focused on those that may easily be implemented in standard statistical packages. There is ongoing work in terms of causal analyses, dealing with missing event type information, and missing covariate values specific to competing risk analyses.

Summary Competing events are common in epidemiologic research. While there has been a focus on why one should conduct a proper competing risk analysis, a perhaps unrecognized issue is in terms of missingness. Strategies exist to minimize the impact of missingness in analyses of competing risks.

Keywords Competing risks · Missing data · Causality · Imputation · Cause-specific hazard · Subdistribution hazard

Introduction

Epidemiologic research questions are often interested in estimating the time to some event. During the course of follow-up, should another event occur before the outcome of interest that precludes the outcome of interest from happening, the other event is termed a competing event. There has been increasing acknowledgement of the importance of conducting an appropriate analysis in the presence of competing risks in epidemiologic and medical research [1]. As shown in Fig. 1, there has been a rapid increase in the number of publications

mentioning competing risks with an approximate increase of 34% per year. However, it has been suggested that almost half of time-to-event studies in which the outcome may be precluded by a competing event overstated the risk of the event of interest by (inappropriately) censoring person-time after the occurrence of a competing event [2–5].

Missing data are also ubiquitous in epidemiologic research. In a causal inference setting, at least one potential outcome (i.e., outcome under a particular value of exposure) is always missing by definition [6, 7], and frequently, covariate information is missing. In situations where there are competing risks, the event time may be missing (i.e., censoring), but also the event type that occurred.

First, we briefly outline competing risks. Second, we review causal inference in competing risk settings. Third, we review several approaches for dealing with missing information on event type. Finally, we summarize methods for accounting for missing covariate information in the presence of competing risks; only recently have multiple imputation methods for time-to-event analyses with extensions to competing risk setting been described [8, 9, 10•]

This article is part of the Topical Collection on *Epidemiologic Methods*

✉ Bryan Lau
blau1@jhu.edu

¹ Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St, Baltimore, MD 21205, USA

² Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA

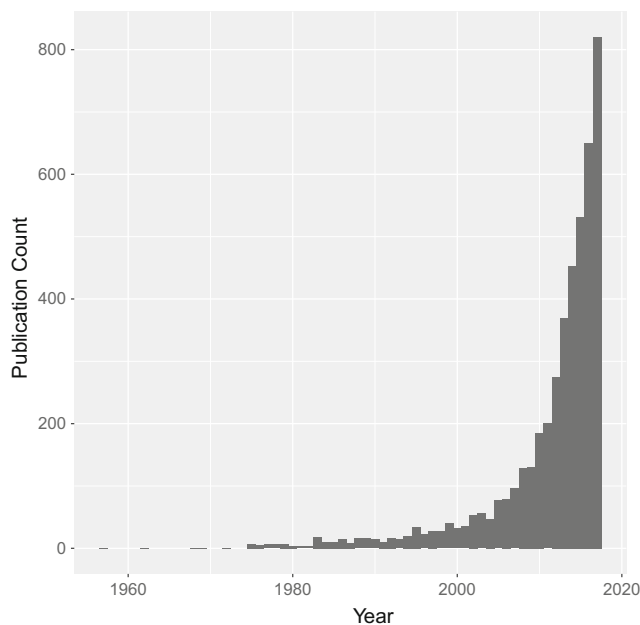


Fig. 1 The number of publications in PubMed mentioning a competing risk-related term by calendar year. The PubMed search was conducted searching on the following terms: “competing cause” OR “competing causes” OR “competing risk” OR “competing risks” OR “competing outcome” OR “competing outcomes” OR “competing endpoints” OR “subdistribution hazard” OR “cause-specific hazard”

Competing Risks: a Brief Review

There are several introductions to competing risks in the epidemiologic and statistical literature [1, 11–13, 14••]. Nevertheless, for completeness, we review some central concepts here. For simplicity, we limit our discussion to two competing event types while noting that methods are easily extended to situations with more than two event types. Let $P(\cdot)$ denote the probability, let T denote the composite event time (that is, the time of the earliest of either the event of interest, any of the competing events, or censoring), and let J denote the event type where $j = \{0, 1, 2\}$ and $j = 0$ represent neither event having occurred (censoring).

Two different hazard functions have been defined in the presence of competing risks. The natural extension of standard time-to-event analyses to competing risk setting is the cause-specific hazard:
$$h_j(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{P(t < T \leq t + \Delta t, J = j | T > t)}{\Delta t} \right\}$$
 [15]. Note that the probability in the numerator of the cause-specific hazard is conditional on remaining free of all events (and censoring) until time t . The cause-specific hazard can be interpreted as the instantaneous rate of the j th event at time t , given the individual has survived to time t [5, 11, 12]. However, this hazard may not translate into the risk of the j th event, as the risk also depends on the cause-specific hazard for the competing event(s) [5, 11, 12]. If the cause-specific hazard of the competing event is high, the risk for the event of interest may actually be quite low, because individuals have the competing event before the event of interest can occur. The cause-specific

hazards act together to determine the timing of any event and the type of event [1, 12]. Therefore, by itself, the cause-specific relative hazard of the event of interest is insufficient for inference on the relationship between the exposure and the risk of the event [14••]. Nevertheless, the cause-specific relative hazard is a valid measure of association of the instantaneous rate and allows for direct assessment of the exposure and specific outcome on this scale.

The second hazard function that has been defined in the context of competing risks is the subdistribution hazard function: λ_j

$$(\lambda_j)(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{P[t < T \leq t + \Delta t, J = j | T \geq t \cup (T \leq t \cap J \neq j)]}{\Delta t} \right\}$$

[16]. In the subdistribution hazard, the probability in the numerator is conditional on remaining free of just the event of interest (and censoring). Alternatively stated, individuals who experience a competing event prior to time t remain in the risk sets after the competing event occurs. This may not seem intuitive, but stems from the idea of a cure model, in that individuals who experience the competing event have been “cured” as they cannot subsequently have the event of interest [14••, 16]. The appeal of this estimand is that an increase in the subdistribution hazard corresponds to an increase in the risk of the event, although the magnitude of the change will not be the same. Thus, the subdistribution hazard ratio reliably provides a qualitative description of the relationship between a variable and the risk of the outcome [14••].

The cumulative incidence is a natural estimand in the presence of competing events and is defined as $F_j^*(t) = P(T \leq t, J = j)$ where F_j^* is used to denote the probability that the j th event occurs by time t . We denote the cumulative incidence function (CIF) with a “*” to highlight that this is not a proper distribution that will integrate to 1 as $t \rightarrow \infty$ in the presence of a competing event. The CIF for the j th event is a function of the cause-specific hazard for the j th event as well as the cause-specific hazards for all other J events through the survival function, $S(t)$. The CIF can be written:

$$F_j^*(t) = \int_0^t S(u) h_j(u) du \tag{1}$$

where

$$S(t) = \exp\left(-\sum_{j=1}^J \int_0^t h_j(u) du\right)$$

As stated above, the CIF is directly related to the subdistribution hazard, and thus it can also be written:

$$F_j^*(t) = 1 - \exp\left(-\int_0^t \lambda_j(u) du\right) \tag{2}$$

Presenting both an estimate of the cause-specific and subdistribution hazard ratios or cause-specific hazard ratios

and corresponding CIFs provides a richer picture of the data and helps provide greater insights [17•]. Presenting the CIFs and absolute risk differences provides important information for public health and etiologic inference. CIFs are less frequently reported, perhaps due to a perceived difficulty generating adjusted estimates. Another estimand of use in the presence of competing risks is the restricted mean time to an event or differences in the restricted mean time to an event; restricted mean time is estimable as the area under the CIFs up to time t [18]. This may be interpreted as the expected time lost due to the event; for instance, the time lost due to AIDS-related mortality could be examined in the context of competing event of non-AIDS-related mortality. Difference in this expected time lost due to AIDS-related mortality could be examined by an exposure of interest [18].

Estimating the non-parametric CIFs under competing risk setting is fairly straightforward using the Aalen-Johansen estimator, $F_j^*(t) = \sum_{t_k} \left\{ \hat{S}(t_{k-1}) \frac{d_j(t_k)}{n_j(t_k)} \right\}$, where $\hat{S}(t_{k-1})$ is the estimate of the overall survival function just prior to time t_k , and $d_j(t_k)$ and $n_j(t_k)$ are the number j events and the number of individuals remaining in the risk set at time t_k , respectively. Inverse probability (IP) weighting may be used to standardize the CIFs [1, 19, 20]. IP weighting can also be used to standardize estimates from a cause-specific or subdistribution proportional hazards model.

Causal Inference in Competing Risk Settings: Missing the Potential Outcomes

The potential outcomes framework has become a prominent approach for conducting analyses that are trying to answer a causal scientific question. The potential outcome, usually denoted Y_i^a , is the outcome Y that would have been observed if, possibly contrary to fact, individual i was exposed to treatment $A = a$. For a binary exposure, each individual has two potential outcomes, one for each exposure level. However, at most, we can only observe the potential outcome under the realized (i.e., factual) exposure (additionally assuming treatment variation irrelevance [21–23]). The potential outcomes under all other exposure levels will be missing. We will suppress subscript i for the remainder of our discussion of potential outcomes in competing risk settings. A review of the entire causal inference literature is beyond the scope of this paper and we refer the readers to the following references [24–26].

Potential outcomes for competing risk settings have recently been defined [1, 27, 28, 29•]. Using the notation of Cole et al. [1, 28], let A represent exposure, let T^a represent the time of occurrence of any outcome (i.e., composite outcome) that would have been observed under exposure level $A = a$, and let J^a represent the event-type indicator under exposure level $A = a$ where $j = 1, 2$ for the case of two competing events. (While

we limit our discussion to two competing events, this is easily expanded to a setting with more competing outcomes.) The potential outcomes in a competing risk setting are then bivariate: (T^a, J^a) [1, 27].

The primary challenge of causal inference is that by definition, at least one potential outcome (i.e., outcome under a particular value of exposure) is always missing [6, 7]. Therefore, one can view bias in answering a causal scientific question as arising from improper imputation of the unobserved potential outcome [30]. These improper imputations are a result of lack of exchangeability [7] between those with and without the exposure, regardless of whether lack of exchangeability is due to confounding or selection bias.

Until recently, there has been little-to-no research on confounder control in competing risk settings. Informally, confounders are variables that could account for a lack of exchangeability between exposure groups. Epidemiologists have recently acknowledged advantages to identifying potential confounders using a directed acyclic graph [31]. However, to our knowledge, there are no established rules for drawing a causal diagram for the competing risk setting; when depicting research questions that involve competing risks, some investigators have (ad hoc) drawn a single directed acyclic graph with separate nodes for each outcome type [32, 33]. This depiction of causal mechanisms would lead most epidemiologists to identify only covariates on an open backdoor path between the exposure and outcome of interest as potential confounders. However, we have shown that estimates of the causal effect of the exposure on the event of interest will be biased if the adjustment set does not include covariates that are confounders of the exposure-competing event causal path (on a directed acyclic graph with separate nodes) [29•]. Some intuition for this finding is available in Eq. 1: the cumulative incidence is a function of all-cause-specific hazards. Failing to adjust for a covariate that changes the cause-specific hazard of the competing event and that is differentially distributed across exposure groups will result in residual confounding in the estimated cumulative incidence through confounding of the relationship between exposure and cause-specific hazard of the competing event. Given that the causal estimands using the CIF are biased when potential confounders of the exposure and competing event are not included, it reasons that estimands directly linked to the CIF, such as the subdistribution hazard ratio, would be biased. This is borne out in simulations [29•].

These advancements in (1) defining potential outcomes and (2) identifying bias when variables related to exposure and the cause-specific hazard of the competing event are not included in the adjustment set have furthered our understanding of causal questions in the competing risk setting. Identification of a set of rules for drawing directed acyclic graphs would help in assessing which variables are needed for d-separation to isolate the causal effect in question.

Missing Data on Event Type

A complication of the competing risk setting is that information on which event type occurred at the time of failure is often uncertain. For instance, in examining time to specific causes of death (e.g., HIV-related and non-HIV-related), the date of death may be known but cause of death on death certificates may be misclassified or missing. We present several analytic approaches that are valid if missingness (or misclassification) can be assumed to be missing at random (i.e., the probability of the missing event type only depends on the observed data [34]).

One approach when event type is misclassified would be to analyze the data using a Poisson-based model to obtain incidence rates for each event. Edwards et al. estimated the effect of occupational asbestos exposure on lung cancer death correcting for misclassification of event type using a Poisson model for two event types [35]. The likelihood function was modified to allow for inclusion of the sensitivity and specificity of the observed, but potentially misclassified, event type. To transform incidence rates into a CIF, the following formula may be used [36]:

$$F_j^*(t) = \frac{\alpha_j}{\alpha_1 + \alpha_2} [1 - \exp(-(\alpha_1 + \alpha_2)t)]$$

where α_j is the incidence rate for the $j = 1, 2$ event type. Note that the Poisson model and incidence rates for estimating the CIF assume constant rates over time although this assumption may be relaxed (for instance, by allowing for piecewise Poisson model).

Goetghebeur and Ryan showed that missing event type could be accounted for by modifying the partial likelihood of a Cox proportional hazards model by (1) modeling the event types jointly, (2) including a parameter for the ratio of the baseline hazards between event types (i.e., $\frac{h_{20}(t)}{h_{10}(t)} = \xi(t)$), and (3) including an additional term for those who have an event but unknown event type [37]. This partial likelihood links the underlying baseline hazards together in order to allow individuals who have an unknown event type to contribute to the analysis with proper contribution to event types based upon $\xi(t)$. If $\xi(t)$ is not known, then it can be estimated. Recently, this work was extended to allow for not only missing event type, but misclassification of the event type [38]. Finally, this approach has also been extended to situations in which the missing event type may depend on auxiliary variables (i.e., variables that are related to the missing event type and assumed to be collected on all individuals who have an event, but that are not being included in the final outcome model) [39]. This extension allows for a weaker missing at random assumption to be made. This may be useful if missingness in the event type is related to a marker of disease progression. For instance, Nevo et al. provide an example in

examining time to subtype of colorectal cancer (microsatellite instability or microsatellite stable) as the competing events, cancer subtype is often missing, and tumor location as an auxiliary variable is associated with microsatellite instability subtype [39]. R code to run these two extensions is available in the appendix of Van Rompaye et al. and available on request from Nevo et al. [38, 39].

Missing event type can also be multiply imputed to estimate either cause-specific or subdistribution proportional hazards ratios [40, 41]. To impute the missing event type, Lu and Tsiatis proposed modeling the probability of the event of interest given the event time, covariates, and auxiliary variables (\mathbf{Z}) using a logistic regression model, such that $P(J_i = 1 | J_i > 0, \mathbf{W}_i) = \frac{\exp(\beta^T \mathbf{W}_i)}{1 + \exp(\beta^T \mathbf{W}_i)}$, where $\mathbf{W}_i = (T_i, \mathbf{X}_i, \mathbf{Z}_i)$ and $J_i = 0$ indicate censored individuals. This model may include non-linear and interaction terms as appropriate. Using this model for imputation requires (i) randomly drawing β^* from $N(\hat{\beta}, \hat{Var}(\hat{\beta}))$, (ii) for the missing cases, compute the $\pi_i = P(J_i = 1 | \beta^*, \mathbf{W}_i)$, and (iii) replace the missing J_i with either $J_i = 1$ or $J_i = 2$ with probability π_i and $1 - \pi_i$, respectively [40, 41]. This is repeated multiple times, storing each imputed data set. Cause-specific or subdistribution hazard ratios are estimated within each imputed dataset and then combined across all imputed data sets using standard multiple imputation rules [42]. If there is also incomplete data in the covariates, the imputation for missing failure type and for missing covariates can be combined using an approach such as multiple imputation by chained equations (MICE, also known as fully conditional specification, FCS) [43, 44].

Finally, an alternate analytic approach when some event types are missing is to decompose the joint distribution of the CIF into a mixture model [45–47, 48•]. That is, the CIF, $F_j^*(t) = P(T \leq t, J = j)$, by rules of conditional probability may be written as either $P(J = j)P(T \leq t | J = j)$ or $P(T \leq t)P(J = j | T \leq t)$. In the first case, when breaking the distribution into event times conditioned on event type, the likelihood function to be maximized may be written to include a term to allow individuals to contribute to the timing of both events [45, 49]. In the second case of vertical modeling, the likelihood can be factored into two parts [48•]. The first part of the likelihood is for the timing of events using the total hazard; this part ignores the cause of failure and all observations can contribute. The second part of the likelihood is for the event type given the survival time; only the failures with known event type contribute. Thus, the likelihood may be maximized separately using a model for overall survival (likelihood part one) and a logistic model (part two) with known cause [48•]. These likelihood functions could potentially be modified to allow for incorporation of sensitivity and specificity to allow for misclassification of event type similar to those of Edwards et al. [35, 50–52].

Missing Covariate Values

Missing values in covariates are ubiquitous in epidemiological research and multiple imputation has become a standard tool to deal with this issues [42]. It is recognized that inclusion of the outcome of interest in the imputation model is imperative [53]. However, in time-to-event analyses, inclusion of the outcome in the imputation model is more complicated as the data may include censoring (i.e., left, right, and interval censoring) and truncation (e.g., left truncation). In the setting of a single failure type, prior work compared including different combinations of an event indicator, the time of event or censoring, and the logarithm of the time of event or censoring [54–56]. Recently, the inclusion of the event indicator and the underlying baseline cumulative hazard has been promulgated as being less biased than inclusion of event or censoring time [8]. The authors proposed that the baseline cumulative hazard be estimated by the Nelson-Aalen estimator. Further improvements to the imputation could be achieved by inclusion of interaction terms between covariates and baseline cumulative hazard in the imputation model. A particular advantage of this approach is that it is invariant to monotonic transformation of the time axis and is approximately compatible with a proportional hazards model [8, 9, 10]. That is, when the outcome model (i.e., substantive model) is non-linear such as a proportional hazards model, the imputation model may impute values that are incompatible with the substantive model. A simple example of this from Bartlett et al., if the outcome Y is a function of covariate X and X^2 yet the imputation model for missing values of X is $X | Y$, then this imputation model is incompatible with the substantive model. This would result in a subset of data in which X has an imputed relationship that is linear in association [9, 57].

There has been even less research on multiple imputation for missing covariate values in the context of time-to-event outcomes when there are competing events. A natural extension would be to include the cumulative baseline cause-specific hazard and binary indicator variables for each event type. For competing risk outcomes, Bartlett et al. proposed an approach called substantive model compatible fully conditional specification (SMC-FCS) imputation [10]. However, this approach requires that the imputation model for missing values within covariate X not only be a function of the parameter φ for model $f(X|Z, \varphi)$ but also a function of parameter β for the outcome model $f(Y|X, Z, \beta)$. Exploiting the iterative nature of FCS algorithm [43, 44], both sets of parameters are estimated [9, 10].

We briefly note that it has been recommended that when an investigator is interested in a single event (e.g., death due to HIV-related causes), those all other competing events (e.g., death due to cancer and due to cardiovascular disease) are collapsed into a single competing event and then analyzed as a two-event situation [17]. While practical for the case of no

missing covariate data, this may result in inefficiency in imputing covariate(s) values when the relationship between the covariate and each of the “sub”-competing events may be different [10]. Nevertheless, a R package called “smcfc” is available for the imputation of data under a competing risks setting [58]. Whether or not this approach can be extended to the subdistribution proportional hazards model is still an open question [10].

Conclusion

Competing events are common in epidemiological research and awareness of the appropriate methods to account for their influence is increasing. Furthermore, missing data is also ubiquitous in epidemiologic research. While several other papers have focused on the interpretation of the cause-specific versus the subdistribution hazard ratio, there has been little focus on missingness in competing risk data. In this review, we sought to provide an introduction to competing risks and an introduction on missingness in a competing risks setting. However, the majority of the missing data has focused on the cause-specific hazards and future research on missingness as applied to the CIF and subdistribution hazard is needed.

Funding Information This work was supported by NIH grants U01 HL121812, U01 AA020793, P30 AI094189, and U24 OD023382.

Compliance with Ethical Standards

Conflict of Interest Bryan Lau reports grants from NIH, during the conduct of the study.

Catherine Lesko declares no conflicts of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance,
- Of major importance

1. Cole SR, Lau B, Eron JJ, Brookhart MA, Kitahata MM, Martin JN, et al. Estimation of the standardized risk difference and ratio in a competing risks framework: application to injection drug use and progression to AIDS after initiation of antiretroviral therapy. *Am J Epidemiol*. 2015;181:238–45.
2. Schumacher M, Ohneberg K, Beyersmann J. Competing risk bias was common in a prominent medical journal. *J Clin Epidemiol*. 2016;80:135–6.

3. van Walraven C, McAlister FA. Competing risk bias was common in Kaplan-Meier risk estimates published in prominent medical journals. *J Clin Epidemiol*. 2016;69:170–173.e8.
4. Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? *Stat Med*. 2012;31:1089–97.
5. Austin PC, Fine JP. Accounting for competing risks in randomized controlled trials: a review and recommendations for improvement. *Stat Med*. 2017;36:1203–9.
6. Holland PW. Statistics and causal inference. *J Am Stat Assoc*. 1986;81:945–60.
7. Westreich D, Edwards JK, Cole SR, Platt RW, Mumford SL, Schisterman EF. Imputation approaches for potential outcomes in causal inference. *Int J Epidemiol*. 2015;44:1731–7.
8. White IR, Royston P. Imputing missing covariate values for the Cox model. *Stat Med*. 2009;28:1982–98.
9. Bartlett JW, Seaman SR, White IR, Carpenter JR, Alzheimer's Disease Neuroimaging Initiative*. Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Stat Methods Med Res*. 2015;24:462–87.
10. Bartlett JW, Taylor JMG. Missing covariates in competing risks analysis. *Biostatistics*. 2016;17:751–63. **This paper provides details on imputing covariates in a manner that is compatible with outcome model. Reference 9 provides context for understanding this paper.**
11. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol*. 2009;170:244–56.
12. Allignol A, Schumacher M, Wanner C, Drechsler C, Beyersmann J. Understanding competing risks: a simulation point of view. *BMC Med Res Methodol*. 2011;11:86.
13. Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol*. 2012;41:861–70.
14. Austin PC, Fine JP. Practical recommendations for reporting fine-gray model analyses for competing risk data. *Stat Med*. 2017;36:4391–400. **This review provides further view on how to interpret competing risk estimands as well as recommendations for reporting analyses.**
15. Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978;34:541–54.
16. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496–509.
17. Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *J Clin Epidemiol*. 2013;66:648–53. **This study provides recommendations on reporting competing risk analyses.**
18. Andersen PK. Decomposition of number of life years lost according to causes of death. *Stat Med*. 2013;32:5278–85.
19. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Prog Biomed*. 2004;75:45–9.
20. Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med*. 2005;24:3089–110.
21. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? *Epidemiology*. 2009;20:3–5.
22. VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009;20:880–3.
23. VanderWeele TJ, Hernán MA. Causal inference under multiple versions of treatment. *J Causal Inference*. 2013;1:1–20.
24. Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol*. 2002;31:422–9.
25. Hernán MA. A definition of causal effect for epidemiological research. *J Epidemiol Community Health*. 2004;58:265–71.
26. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60:578–86.
27. Bekaert M, Vansteelandt S, Mertens K. Adjusting for time-varying confounding in the subdistribution analysis of a competing risk. *Lifetime Data Anal*. 2010;16:45–70.
28. Cole SR, Hudgens MG, Brookhart MA, Westreich D. Risk. *Am J Epidemiol*. 2015;181:246–50.
29. Lesko CR, Lau B. Bias due to confounders for the exposure-competing risk relationship. *Epidemiol*. 2017;28:20–7. **First paper to illustrate that in a causal analysis, there is bias when not controlling for confounders of the exposure and competing event.**
30. Edwards JK, Cole SR, Westreich D. All your data are always missing: incorporating bias due to measurement error into the potential outcomes framework. *Int J Epidemiol*. 2015;44:1452–9.
31. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37–48.
32. Hernán MA, Schisterman EF, Hernández-Díaz S. Invited commentary: composite outcomes as an attempt to escape from selection bias and related paradoxes. *Am J Epidemiol*. 2014;179:368–70.
33. Kramer MS, Zhang X, Platt RW. Kramer et al. respond to “composite outcomes and paradoxes”. *Am J Epidemiol*. 2014;179:371–2.
34. Rubin DB. Inference and missing data. *Biometrika*. 1976;63:581–92.
35. Edwards JK, Cole SR, Chu H, Olshan AF, Richardson DB. Accounting for outcome misclassification in estimates of the effect of occupational asbestos exposure on lung cancer death. *Am J Epidemiol*. 2014;179:641–7.
36. Grambauer N, Schumacher M, Dettenkofer M, Beyersmann J. Incidence densities in a competing events analysis. *Am J Epidemiol*. 2010;172:1077–84.
37. Goetghebeur E, Ryan L. Analysis of competing risks survival data when some failure types are missing. *Biometrika*. 1995;82:821–33.
38. Van Rompaye B, Jaffar S, Goetghebeur E. Estimation with cox models: cause-specific survival analysis with misclassified cause of failure. *Epidemiol Camb Mass*. 2012;23:194–202.
39. Nevo D, Nishihara R, Ogino S, Wang M. The competing risks Cox model with auxiliary case covariates under weaker missing-at-random cause of failure. *Lifetime Data Anal*. 2017; In Press
40. Lu K, Tsiatis AA. Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics*. 2001;57:1191–7.
41. Bakoyannis G, Siannis F, Touloumi G. Modelling competing risks data with missing cause of failure. *Stat Med*. 2010;29:3172–85.
42. Rubin DB. Multiple imputation for nonresponse in surveys. John Wiley & Sons; 2004.
43. Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv Methodol* 2001;27:85–96.
44. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res*. 2007;16:219–42.
45. Lau B, Cole SR, Moore RD, Gange SJ. Evaluating competing adverse and beneficial outcomes using a mixture model. *Stat Med*. 2008;27:4313–27.
46. Nicolaie MA, van Houwelingen HC, Putter H. Vertical modeling: a pattern mixture approach for competing risks modeling. *Stat Med*. 2010;29:1190–205.
47. Lau B, Cole SR, Gange SJ. Parametric mixture models to evaluate and summarize hazard ratios in the presence of competing risks with time-dependent hazards and delayed entry. *Stat Med*. 2011;30:654–65.
48. Nicolaie MA, van Houwelingen HC, Putter H. Vertical modelling: analysis of competing risks data with missing causes of failure. *Stat*

- Methods Med Res. 2015;24:891–908. **This study provides information on how to conduct competing risk analyses when which event occurred may be missing for some observations**
49. Crowder MJ. Classical competing risks: CRC Press; 2001.
 50. Neuhaus JM. Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*. 1999;86:843–55.
 51. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective: CRC press; 2006.
 52. Lyles RH, Tang L, Superak HM, King CC, Celentano DD, Lo Y, et al. Validation data-based adjustments for outcome misclassification in logistic regression: an illustration. *Epidemiology*. 2011;22: 589–97.
 53. Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol*. 2006;59:1092–101.
 54. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med*. 1999;18:681–94.
 55. Clark TG, Altman DG. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. *J Clin Epidemiol*. 2003;56:28–37.
 56. Barzi F, Woodward M. Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. *Am J Epidemiol*. 2004;160:34–45.
 57. Seaman SR, Bartlett JW, White IR. Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC Med Res Methodol*. 2012;12:46.
 58. Bartlett J, Keogh R. smcfcs: Multiple imputation of covariates by substantive model compatible fully conditional specification [Internet]. 2017 [cited 2017 Dec 9]. Available from: <https://cran.r-project.org/web/packages/smcfcs/index.html>