Published in final edited form as:

Ann Neurol. 2015 December ; 78(6): 839–844. doi:10.1002/ana.24538.

Recognizing the Problem of Delayed Entry in Time-to-Event Studies: Better Late than Never for Clinical Neuroscientists

Rebecca A. Betensky¹ and Micha Mandel²

¹Harvard School of Public Health, Department of Biostatistics, Boston, Massachusetts, United States

²Department of Statistics, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem, 91905, Israel Jerusalem

Introduction

A common clinical study design in neurology and other fields involves measuring the time until a particular event occurs, in a cohort of human subjects who are enrolled based on a common diagnosis, exposure, or set of characteristics. For example, we may want to know the time until first seizure in patients who have suffered severe brain injury and are at risk for post-traumatic epilepsy. Or we may want to know the time until death in patients who have been given a diagnosis of leptomeningeal metastases. Ideally, we would have complete observations on a random sampling of all patients followed from the time origin of interest all the way through the event of interest, although this is rarely the case. Methods such as Kaplan-Meier survival analysis and Cox proportional hazards modeling are familiar to those in clinical neuroscience who conduct or interpret such time-to-event studies, and can take into account the well-recognized issues related to incomplete follow-up, such as when the available observations for some subjects end prior to occurrence of the event (known as “right-censoring”).

An important concern that is not as well acknowledged, however, is the problem of delayed entry. This arises in a clinical study when some subjects are enrolled after the natural time origin for the time to event measurement. The fundamental principle is graphically depicted in Figure 1, in which the times-to-event are shown as horizontal lines for each of eight subjects. If this set of eight subjects were randomly sampled from the target population, the analysis of time-to-event would use the observed times, and would adjust for right-censoring if follow-up ended prior to observation of the event (as for subjects 1, 3, 6). However, the subjects in this figure did not enter the study at time 0, but rather at some later time (depicted with an asterisk). Thus, the times-to-event are not a random sample from the target population, but rather are preferentially selected from the larger times-to-event of this population, since they are observed only if they exceed the times to study entry.

Corresponding author: Dr. Rebecca A Betensky, Harvard School of Public Health, Department of Biostatistics, Boston, Massachusetts, United States BETENSKY@hsph.harvard.edu. Co-author: Dr. Micha Mandel, Department of Statistics, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem, 91905, Israel Jerusalem, micha.mandel@mail.huji.ac.il.

This problem was addressed in a study of long-term health effects of the atomic bomb dropped on Hiroshima.¹ The government of the Hiroshima Prefecture gave out a booklet to atomic bomb survivors beginning in 1968 describing their potential long-term health problems, and urged them to register for the study. However, not everyone who was eligible registered in 1968, and registrants into the study were accepted through the end of follow-up in 1992. The survival times from the time of issuance of the booklets in 1968 for the participants in this study are not representative of all atomic bomb survivors from 1968. Instead they are longer than those of the reference population because the study participants had to have survived beyond 1968 to enter this study. Thus, everyone who entered after 1968 began the study with some positively valued survival time. Subjects who died in 1970 or 1980 or 1990 who had not previously entered the study are not included; i.e., *the shorter survival times are underrepresented*. Even more problematic is that those who entered the study after the initial registration period may have been influenced by their health status to enter (e.g., those who developed cancer may have been more likely to register late than those who remained healthy). While the ideal cohort for the analysis of long-term health effects would have been the entire population of survivors alive in 1968, the investigators extended accrual for several years beyond 1968 due to incomplete registration in 1968. As in this study, it is often not possible to fully accrue participants to a study at the time origin of interest, and so statistical methods are required to accommodate the imperfect sampling. Such problems in epidemiological and clinical studies are common.

For example, in a 2014 paper published in *Annals of Neurology*², the authors showed that carrying the APOE4 allele confers greater risk of Alzheimer's disease (AD) in women than in men. The study used data from the National Alzheimer's Coordinating Center (NACC) cumulative database (https://www.alz.washington.edu/WEB/researcher_home.html). This database contains longitudinal follow-up for subjects with AD and related disorders, as well as for cognitively normal subjects and those with mild cognitive impairment (MCI). It includes demographic, clinical, neuropsychological variables, in addition to neuropathological variables for an autopsy subcohort. The NACC database, which is publicly available, is a rich resource for the investigation of factors associated with the risk of progression from a well-defined time origin to some AD-related event of interest. However, there is often some post-baseline event that must occur prior to the milestone event of interest in order for subjects to be sampled into the study. For example, in the NACC analysis², the time origin was taken to be birth and the event of interest was first detection of MCI or possible or probable AD. Furthermore, subjects were included in the study sample only if they were deemed to be healthy controls (i.e., without a diagnosis of MCI or AD) at the time of their entry to the NACC database. This sampling feature introduced "delayed entry" or "left truncation" selection bias, which results in higher ages of onset of MCI or AD than in the population as a whole because individuals were sampled subject to the constraint that they exceeded the ages at NACC database entry. This feature must be accommodated in the analysis to avoid biased testing and estimation (other than a potential small and unavoidable bias in estimation due to lack of observation of the leftmost tail of the distribution). Importantly, the ages of entry to NACC vary among individuals; if they were all the same, a standard analysis would naturally account for the delayed entry. In this article, we describe the simple adjustment that is required for valid application of the Kaplan

Meier estimator³ for the survival probabilities and the Cox regression model⁴ for the hazard ratio, when a study includes some subjects with a delayed entry. We report results from simulation studies that illustrate the potential impact of ignoring delayed entry on survival function estimation, hazard ratio estimation, and type I error and power of hypothesis tests.

Time-to-event analysis and delayed entry

The most commonly used statistical method for estimation of the survivor function is the Kaplan Meier estimator³ and for estimation of the most familiar regression model for time-to-event data is the Cox proportional hazards model.⁴ Both of these methods are based on the hazard function, which is defined as the probability of experiencing the event in the next moment of time given that the event has not yet occurred. The hazard function is estimated through decomposition of the data into a sequence of “risk sets” defined at each event time. In the absence of delayed entry to the study, the risk set at a particular event time contains all subjects who are under observation (i.e., they have not been previously lost to the study or “censored,” such as by dropping out of the study) and they have not yet experienced the event. The assumption underlying this approach is that subjects are comparable, and thus comprise a risk set, only if they are currently in view and at risk for the event. This approach also assumes that any drop-out or censoring is independent of the event of interest. For example, in the NACC analysis², this approach is valid if subjects are lost to follow up due to cardiovascular or cancer death – assumed independent of their risk of AD -- but not if they are withdrawn because they were unable to drive to the hospital, which may be due to dementia.

This heuristic extends naturally to accommodate delayed entry by further constraining the definition of the risk set at a particular event time to include only subjects who have entered the study **prior** to that event time, and have not yet experienced the event of interest or been censored from the study by that time. For the data depicted in Figure 1, subject 7 entered the study at time 17 and experienced the event at time 30 is included in the risk set corresponding to the event that occurred at time 22, but is not included in the risk set defined by the event at time 15 because that subject was not yet under observation at time 15. Thus, while the incorrect estimate of the hazard for the event at time 15 that ignores delayed entry is 1/5 (since there is one event at time 15 and five subjects, numbers 4–8, at risk for the event at time 15), the correct estimate of the hazard that adjusts for delayed entry is 1/4, since subject 7 is not included in the risk set at time 15 due to entry at time 17. This simple example illustrates the simple adjustment for delayed entry that is incorporated in the Kaplan Meier estimator and Cox regression. This approach requires the assumption that the time to study entry is independent of the time to event in the source population. In the NACC analysis², this means assuming that subjects who enter NACC at older ages without MCI or AD are not at lower risk of AD than subjects who enter NACC at younger ages.

Simulation studies: delayed entry bias in Kaplan Meier analysis

To assess the impact of delayed entry on estimation of the survivor function, we simulated exponentially distributed times to event for 100 subjects with a mean survival of 5 years. The times to study entry were simulated from a normal distribution with mean of 4 years and

standard deviation of 2 years. We imposed delayed entry by retaining only those subjects for whom their times to event exceeded their times to study entry. For each simulated dataset, we calculated a Kaplan Meier estimate that did not adjust for delayed entry and a Kaplan Meier estimate that did adjust for delayed entry. We repeated this 5000 times and averaged the Kaplan Meier estimates. We also ran this simulation using a time to study entry distribution with a mean of 8 years. Figure 2 displays the true survivor function (solid line) and the averaged unadjusted estimates for mean delayed entry of 4 years (dashed line) and of 8 years (dotted line). It is clear that application of the unadjusted Kaplan Meier method in the presence of delayed entry produces a biased estimate, with an upward bias that depends on the study entry distribution. The true survivor function can be estimated from the observed data by applying the proper adjustment for the effect of delayed entry.³ This does not require knowledge of the distribution of the delayed entry.

Simulation studies: delayed entry bias in Cox model analysis

To assess the impact of delayed entry on hazard ratio estimation and type I error and power based on fitting a Cox proportional hazards model, we simulated times to event and times to study entry for two groups of 100 subjects (Table 1). Each of six simulation scenarios is defined by study entry distributions and event time distributions for each group and is represented in a single row of Table 1. The first simulation scenario has equivalent time-to-event distributions for the two groups, but different study entry times. The study entry times are taken to be 0 years for group 1, which means all are entered in a single registration at day 0, and at 1.5 years for group 2. This would occur if the control group were recruited at the time origin (e.g., diagnosis) and the treated group were recruited at a subsequent follow-up visit. The times-to-event follow a Weibull distribution with a hazard function that is linear in time with a slope of 2. This time-to-event distribution, with an increasing hazard function, might apply, for example, to a group of 65 year old and older people where the event was onset of Alzheimer's disease and the risk of onset increases with age. This simulation scenario shows that failure to adjust for delayed study entry can lead to a biased hazard ratio estimate and an inflated type I error even when the time-to-event distributions are equivalent for the two groups (i.e., hazard ratio of 1). The unadjusted Cox model yields a biased hazard ratio of 5.01 (instead of the true hazard ratio of 1), incorrectly suggesting a highly elevated risk for group 2 versus group 1. The adjusted unbiased analysis yields an estimated hazard ratio of 1.0, which accurately conveys no difference in risk between the two groups. Correspondingly, the type I error (i.e., the false positive probability) of the unadjusted analysis is 1, while that of the adjusted analysis is 0.057, close to the nominal level of 0.05. This example demonstrates the potentially large impact of different entry time distributions on the unadjusted analysis, and the importance of accounting for delayed entry.

In the next five simulation scenarios, the entry times of subjects into the study are randomly generated from the same distribution for both groups, with approximately half of the subjects in each group entering late. (The study entry times for each group follow a Bernoulli distribution, i.e., a coin flip, taking on values of 0 years and 1.5 years with equal probability.) In practice, this might reflect two modes of recruitment to a study: e.g., one close to diagnosis, perhaps after recovery from surgery, and one after a greater time lag from diagnosis, perhaps during a regular follow-up visit. The two groups, however, have different

underlying risk, so that their time-to-event distributions are not equivalent. The event times still follow Weibull distributions with hazard functions that are linear in time, but in this case the slope is 2 for group 1 for all five scenarios, but the slope is $2/(\text{hazard ratio})$ for group 2, where the hazard ratios (defined as the hazard function for group 1 divided by the hazard function for group 2) are different for each of the five scenarios and range from 1.28 to 20.09. For example, for the scenario listed in the second row of Table 1, group 2 has a linear hazard function with slope of $2/1.28=1.56$. These five simulation scenarios show that failure to adjust for delayed study entry can lead to a biased hazard ratio estimate and inflated power even when the study entry distributions are equivalent for the two groups. It is seen in rows 2–6 of the table that the unadjusted analysis produces biased estimates of the hazard ratios, while the adjusted analysis is nearly unbiased. The unadjusted analysis has very high power (i.e., probability of statistical significance), though as seen in the first simulation scenario, the type I error can be hugely inflated and so assessments of statistical significance cannot be trusted.

Discussion

Delayed entry is common in studies that are based on longitudinal cohorts that use a time origin that occurred prior to study entry, such as birth, in a time to event analysis. This was the structure of the study sample used for analysis of an interaction between sex and the APOE4 allele in the risk of onset of MCI or AD.² This sampling feature necessitates proper analytic adjustment for unbiased estimation, such as with the Kaplan Meier estimator. Likewise, a two-group comparison of times-to-event and, more generally, a regression model for time-to-event data, such as the Cox model, also require adjustment.

The simple adjustment of the risk sets that form the basis for the Kaplan Meier and Cox model estimates is necessary but not always sufficient, as it is valid only under the condition that the event time and the study entry time are independent (given that the former necessarily exceeds the latter).⁵ This can and should be tested⁵ prior to application of any of these methods. As an example, if a treatment is initiated at the delayed study entry time, and this causes a change in the hazard for the event, the requisite independence does not hold. If independence does not hold, alternative methods that involve modeling of the dependence are required.^{6–9} In fact, the NACC analysis² applied the test for independence⁵ and detected dependence. They addressed this by including age-dependent terms in their Cox regression models.

Neurologists who are performing long-term cohort studies need to include these considerations in their study design and analyses. Those who are not familiar with these issues are well advised to seek advice from a biostatistician. Readers of their work, including reviewers for the *Annals of Neurology*, should be concerned when they read such studies whether the possibility of a delayed entry effect has been considered by the authors and, if necessary, accommodated by appropriate statistical correction.

Acknowledgments

Potential Conflicts of Interest

Rebecca Betensky has received funding from the National Institute of Health for research on problems of delayed entry. She also is compensated for her work as the Statistical Editor for *Annals of Neurology*. Micha Mandel has received funding from the Israel Science Foundation for research on related problems.

References

1. Matsuura M, Eguchi S. Modeling late entry bias in survival analysis. *Biometrics*. 2005; 61:559–566. [PubMed: 16011705]
2. Altmann A, Tian L, Henderson VW, Greicius MD for the ADNI Investigators. Sex modifies the APOE-related risk of developing Alzheimer's disease. *Annals of Neurology*. 2014; 75:563–573. [PubMed: 24623176]
3. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assn.* 1958; 53:457–481.
4. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B*. 1972; 34:187–220.
5. Tsai WY. Testing the assumption of independence of truncation time and failure time. *Biometrika*. 1990; 77:169–177.
6. Efron B, Petrosian V. Survival analysis of the gamma-ray burst data. *Journal of the American Statistical Association*. 1994; 89:452–462.
7. Chaieb LL, Rivest LP, Abdous B. Estimating survival under a dependent truncation. *Biometrika*. 2006; 93:655–669.
8. Gail MH, Graubard B, Williamson DE, Flegal KM. Comments on 'Choice of time scale and its effect on significance of predictors in longitudinal studies'. *Stat Med*. 2009; 28:1315–1317. [PubMed: 19288532]
9. Jones MP, Crowley J. A general class of nonparametric tests for survival analysis. *Biometrics*. 1989; 45:157–170. [PubMed: 2655728]

Figure 1.

Depiction of times to study entry and times to event for eight subjects. Each horizontal line represents a different subject, identified with a number (1–8) as marked on the vertical axis. Time 0 represents the origin of interest (e.g., 1968 for the Hiroshima study and birth for the NACC Alzheimer's study). The asterisk on each horizontal line marks the time (measured from the origin) of study entry. The right endpoint of each horizontal line marks the total follow-up (measured from the origin) for each subject. Solid circles indicate that the event occurred, while outlined circles indicate that the event had not occurred at the end of follow-up (i.e., censored observations).

Figure 2.

Depiction of true and estimated survivor functions for our simulation study: the true survivor function is shown with a solid line and the averaged unadjusted Kaplan Meier estimate for a mean delayed entry of 4 years is shown with a dashed line and for a mean delayed entry of 8 years with a dotted line.

Table 1

Simulation results (10,000 repetitions; $n=100^J$ per group) for Weibull Cox model analysis with and without adjustment for delayed entry

group 1		group 2			hazard ratios ^{<i>J</i>}		type I error/power	
mean event time ²	entry time ³	mean event time	entry time	true ⁴	unadjusted estimates	adjusted estimates	unadjusted	adjusted
0.89	0	0.89	1.5	1.00	5.01	1.04	1	0.057
0.89	0.5 or 2.5	1.00	0.5 or 2.5	1.28	1.31	1.29	0.463	0.420
0.89	0.5 or 2.5	1.14	0.5 or 2.5	1.65	1.73	1.65	0.971	0.938
0.89	0.5 or 2.5	1.46	0.5 or 2.5	2.72	3.19	2.74	1	1
0.89	0.5 or 2.5	2.41	0.5 or 2.5	7.39	11.33	7.56	1	1
0.89	0.5 or 2.5	3.97	0.5 or 2.5	20.09	36.23	21.17	1	1

^{*J*} Group 1 versus group 2

² Event times follow Cox proportional hazards Weibull distributions

³ Entry times are 0 for group 1 and 1.5 for group 2 when hazard ratio=1; and 0.5 with probability 0.5 and 2.5 with probability 0.5 when hazard ratio>1

⁴ True hazard ratio based on simulation model