Immortal Time Bias in Observational Studies of Time-to-Event Outcomes: Assessing Effects of Postmastectomy Radiation Therapy Using the National Cancer Database

Parul Agarwal, PhD[1], Erin Moshier, MS[1], Meng Ru, MS[1], Nisha Ohri, MD[2],
Ronald Ennis, MD[3], Kenneth Rosenzweig, MD[2], and Madhu Mazumdar, PhD[1]

Abstract

The objectives of this study are to illustrate the effects of immortal time bias (ITB) using an oncology outcomes database and quantify through simulations the magnitude and direction of ITB when different analytical techniques are used. A cohort of 11 626 women who received neoadjuvant chemotherapy and underwent mastectomy with pathologically positive lymph nodes were accrued from the National Cancer Database (2004-2008). Standard Cox regression, time-dependent (TD), and landmark models were used to compare overall survival in patients who did or did not receive postmastectomy radiation therapy (PMRT). Simulation studies showing ways to reduce the effect of ITB indicate that TD exposures should be included as variables in hazard-based analyses. Standard Cox regression models comparing overall survival in patients who did and did not receive PMRT showed a significant treatment effect (hazard ratio [HR]: 0.93, 95% confidence interval [CI]: 0.88-0.99). Time-dependent and landmark methods estimated no treatment effect with HR: 0.97, 95% CI: 0.92 to 1.03 and HR: 0.98, 95% CI, 0.92 to 1.04, respectively. In our simulation studies, the standard Cox regression model significantly overestimated treatment effects when no effect was present. Estimates of TD models were closest to the true treatment effect. Landmark model results were highly dependent on landmark timing. Appropriate statistical approaches that account for ITB are critical to minimize bias when examining relationships between receipt of PMRT and survival.

Keywords

survival analysis, cox regression, immortal time bias, time-dependent, landmark, breast cancer, epidemiology, NCDB, observational, radiotherapy

Introduction

Immortal time bias (ITB), identified in epidemiology since the 1970s,[1,2] occurs when there is variation in timing of treatment initiation from cohort entry and time-to-treatment is misclassified or ignored. Immortal time bias can occur in observational studies when a cohort approach is followed during which outcomes cannot occur. In the drug effectiveness literature, various cohort designs may result in ITB.[3] Importantly, analyses can be flawed if ITB is not accounted for. To simplify, we generalize "treatment" to mean exposure, treatment, or response to exposure or treatment, as ITB can arise when the

[1] Department of Population Health Science and Policy, Institute for Healthcare Delivery Science, Tisch Cancer Institute (TCI), Icahn School of Medicine at Mount Sinai, New York, NY, USA
[2] Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[3] Department of Radiation Oncology, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA

Corresponding Author:
Madhu Mazumdar, Department of Population Health Science and Policy, Institute for Healthcare Delivery Science, Tisch Cancer Institute (TCI), Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
Email: madhu.mazumdar@mountsinai.org

timing to initiation of any of these states varies between patient groups.

Various studies have addressed ITB by applying statistical approaches such as time-dependent (TD) Cox regression, landmark analyses, prescription time-distribution matching, and the sequential Cox approach.[4-6] In a systematic review, over 40% of studies with a survival end point and time-varying treatment were susceptible to ITB.[7] Thus, it is important to identify and control ITB, because it may alter study conclusions by underestimating the hazard ratio (HR), a measure of an effect of an intervention on an outcome of interest over time. Consequently, researchers may falsely conclude that a treatment is effective in influencing outcomes. For example, a systematic review of studies using the Surveillance, Epidemiology, and End Results Program database comparing survival among patients who did and did not receive postoperative radiotherapy observed that ITB was not addressed or controlled for in most studies, which may have led to false conclusions.[8] A systematic review and meta-analysis on the effects of β-adrenergic receptor antagonists on cancer survival drew similar conclusions.[9]

Although systematic reviews identified ITB as an issue in the oncology literature, optimal methods for addressing ITB arising from variations in timing of initiation of radiation therapy are lacking. Immortal time bias may be more pronounced in the oncology literature due to greater availability and use of observational data to explore various research questions. Thus, it is important for clinical researchers to be aware of ITB and statistical methods to address it.

The effectiveness of postmastectomy radiation therapy (PMRT) is well established in the literature[10-12]; however, when results of studies are analyzed, variations in the timing of initiation of radiation therapy may lead to ITB. It is vital to address ITB in radiation oncology studies, since uncontrolled bias may affect estimates of therapeutic effectiveness of radiation on survival among patients with cancer and could generate spurious associations. Therefore, the present study used an oncology outcomes database and sought to quantify through simulations the magnitude and direction of ITB when different analytic techniques are used.

Methods

Study Design

A retrospective cohort study design was used to compare overall survival in patients who did and did not receive PMRT. Using parameter estimates obtained from the cohort (see subsequently), we conducted simulations to provide recommendations on including TD exposure in hazard-based analyses.

Because no patient, provider, or hospital identifiers are included in the analytic or simulation components of this study and no protected health information is present, institutional review board approval was not required. Informed consent was not required for the analytic sample or simulation study.

Data Source

Data were obtained for the years 2004 to 2008 from the National Cancer Database (NCDB), which contains outcomes from 1500 commission-accredited cancer programs in the United States and Puerto Rico. National Cancer Database contains information on all types of cancers and includes approximately 34 million records from cancer registries. The data set includes information on patients, tumor- and cancer-related characteristics, and outcomes but excludes direct patient identifiers. National Cancer Database has been used to conduct multiple clinical and quality improvement studies in the area of cancer research.

Analytical Sample

The analytical sample (N = 11 626) consisted of women who (1) were ≥18 years of age and diagnosed with invasive, non-metastatic breast cancer; (2) received neoadjuvant chemotherapy; (3) underwent mastectomy; (4) had macroscopic pathologically positive lymph nodes; and (5) were diagnosed between 2004 and 2008, to ensure a minimum follow-up of 5 years and limit censoring bias. We excluded women with (1) clinical or pathologic evidence of metastatic disease; (2) bilateral or inflammatory breast cancer; (3) exposure to neoadjuvant hormonal therapy, neoadjuvant radiation therapy, and intraoperative chemotherapy; and (4) no treatment at the reporting facility. Neoadjuvant chemotherapy was defined by an interval from initiation of chemotherapy to surgery of ≥80 days and ≤270 days.

Measures

Treatment versus comparator: PMRT versus No PMRT. Postmastectomy radiation therapy was defined as delivery of >4500 Gy of external beam radiation therapy to the chest wall with or without regional lymph node irradiation. The comparison group was those who received <45 Gy of external beam radiation therapy. A previously published manuscript provides greater detail regarding the definition of PMRT.[13]

Other measures. We did not consider the effects of confounders on the relationship between overall survival and treatment for this simulation study. This project was designed as an exercise for statistical demonstration and not a true clinical result.

Statistical Analysis

The standard Cox proportional hazards model assumes group membership is known at the time of diagnosis. We used this model to compare overall survival, as is often done in the literature, despite the fact that we compared PMRT (which is TD) and no PMRT. We then applied TD and landmark analyses to demonstrate how these more appropriate statistical methods can address potential ITB.

Cox proportional hazards models. The Cox proportional hazards model is the most frequently used analytical method in survival analysis. It models the hazard rate, which is the number of new events of disease per size of the population at risk, for a given unit of time. The hazard function is the likelihood of experiencing an event in the next instant for survivors to time t, adjusting for potential covariates. It can be denoted by $\lambda(t)$ and calculated from,

$$\lambda(t|\mathbf{Z}) = \lambda_0(t)\exp(\beta'\mathbf{Z}) = \lambda_0(t)\exp(\Sigma\beta_j Z_j)(\Sigma\beta_j Z_j),$$

where $\lambda_0(t)$ is the baseline hazard function and β_j is the parameter for the jth covariate (Z_j).

Landmark analysis. Landmark analysis was first introduced by Anderson et al[14] to address ITB in comparing survival between 2 groups, when the study group assignment (response vs no-response) cannot be determined until follow-up. In this method, a fixed time point is first selected as the landmark time. Patients who have experienced the event of interest, or are censored prior to the selected landmark time, are excluded from analysis; group membership is determined at the specific landmark time. In this approach, patients who initiate treatment after the landmark time are included in the no-treatment group, regardless of any treatment they receive thereafter. The standard Cox proportional hazards model described above was then applied to the landmark data, with the start time for analysis being the landmark time.[15]

Time-dependent Cox models. In TD Cox models, the standard Cox model is expanded by including a TD variable with group membership changing during the follow-up period. This method can also be used when other covariates (eg, biomarkers) change during follow-up. So, like standard Cox models, we can therefore estimate the hazard function for a TD Cox model as:

$$\lambda(t|\mathbf{Z}(t)) = \lambda_0(t)\exp\{\beta'\mathbf{Z}(t)\} = \lambda_0(t)\exp\{\Sigma\beta_j Z_j(t)\},$$

where the hazard at time t depends on the value of the TD covariate $Z(t)$.[16] A TD Cox model compares risk of an event between 2 treatment groups at each event time but also reevaluates which group each person belongs to based on their treatment assignment at that time.

The Weibull distribution. We examined the distribution of various types of event times (event refers to death, censored, or receiving PMRT) and found that the Weibull distribution was the best fit to all of these time-to-event outcomes. With a shape parameter (λ) and a scale parameter (β), it follows the probability density function as:

$$f(x; \lambda, \beta) = \begin{cases} \dfrac{\lambda}{\beta}\left(\dfrac{x}{\beta}\right)^{\lambda-1} e^{\left(-\frac{x}{\beta}\right)^{\wedge}\lambda}, & x \geq 0. \\ 0, & x < 0 \end{cases}$$

Previous literature has shown that Weibull distributions, along with the exponential and the Gompertz distribution, can be applied to generate appropriate survival times and also share the same assumption of proportional hazards with the Cox model.[17] We used the distribution parameters (the scale and shape of Weibull distributions) identified through curve fitting by Statistical Analysis Software (SAS) to generate survival time in simulations and computed HRs with Cox regression models.

Simulation approach. This simulation study was performed to explore the most appropriate statistical techniques to address ITB. Time-to-event was defined as the time from diagnosis of disease until death, censoring occurred during the period from diagnosis of disease to last follow-up of the patient, and time-to-treatment was defined as the time from diagnosis of disease to receipt of PMRT. Weibull distributions were assumed for times to event, censoring, and treatment initiation. A Weibull distribution is a continuous probability distribution used to describe any phase of an item's lifetime defined by its shape (λ) and scale (β) parameters. It is commonly used to simulate and model survival data since it can assume the characteristics of many different types of distributions, that is, it follows an exponential distribution when $\lambda = 1$.

A range of shape (λ) and scale (β) parameter values were assumed: (1) shape parameter, which represents the hazard rate as constant ($=1$), increasing (>1) or decreasing (<1), and (2) scale parameter, which represents the mean time-to-event. Different values were specified for the distributional parameters to yield different proportions of treatment and censoring. All 3 distributions were independent of each other. For each simulated patient, if the censoring time occurred before the event time, the event did not occur for that patient during the study. For each set of parameter values, 1000 samples of the size 10 000 were simulated. Various simulation scenarios were generated and analyzed by the standard, TD, and landmark Cox methods. Hazard ratios (PMRT vs no PMRT) were used to assess treatment effects.

For ease of illustration, we assumed that the treatment had no effect on survival (HR = 1). When the HR < 1, the treatment effect was overestimated, and when the HR > 1, the treatment effect was underestimated. For landmark analysis, we selected a 12-month landmark because most patients had received PMRT by then, and most patients were still alive and being actively followed. For each set of parameter values, we show the average HR among the 1000 simulated data sets and confidence interval coverage (CIC), calculated as the proportion of 95% confidence intervals that included the true HR of 1 among the 1000 simulations. Root mean squared error (RMSE) was also computed as the square root of the sum of the squared bias and the simulated variance for HR.[18] Confidence interval coverage and RMSE are standard performance metrics used in simulation studies to evaluate the accuracy and precision of the estimator. A higher CIC indicates better ability to capture the true value. RMSE measures the mean squared difference between the estimator and the parameter and evaluates the error made by the estimator; it can be decomposed into a sum of bias and variance where the bias (the difference between the true HR and average HR) measures the accuracy of the estimator and the variance measures the precision of the estimator. A smaller RMSE indicates a more precise and accurate estimator.

Table 1. Hazard Ratios (HRs) for Overall Mortality (PMRT vs no PMRT).[a]

Method (# PMRT vs # no PMRT)	HR	95% CI	P Values
Standard Cox model (6726 vs 4900)	0.930	0.880-0.990	.0227
LM method,[b] N = 11 254 (6284 vs 4970)	0.978	0.920-1.040	.4838
TD Cox model, N = 11 626 (6726 vs 4900)	0.973	0.917-1.034	.3773

Abbreviations: CI, confidence interval; LM, landmark; PMRT, postmastectomy radiation therapy; TD, time-dependent.
[a]N = 11 626.
[b]12 month LM used.

Sensitivity analysis was performed on a smaller sample size of 5000 with 500 simulation iterations to assess robustness of results, since smaller samples and less iteration are more sensitive to variability.

All analyses were conducted using SAS Version 9.4 (SAS Institute, Cary, North Carolina) with simulation analyses adopted from the SAS macro presented by Jones et al.[5]

Results

Observational Study Results

From the NCDB data set with 11 626 patients, 6726 (58%) received PMRT and 4418 (38%) died. Table 1 shows the results from each model using the NCDB data. Significant differences were observed between treatment groups when the standard Cox model was used, indicating failure to address ITB. No significant differences were detected when landmark and TD Cox models were used to account for ITB. The findings on models adjusted for confounders can be found elsewhere.[13] Figure 1 shows the Kaplan Meier survival probabilities for PMRT and Non-PMRT groups with 12-months (clinically justified) LM and standard Cox models. Significant differences were observed between PMRT and Non-PMRT groups when standard Cox model was use, however ITB was ignored here. No significant differences were observed between PMRT and Non-PMRT groups when LM method was used, thereby accounting for ITB.

Simulation Study Results

We fitted the survival, censored, and treatment events from the NCDB data separately to Weibull distributions and used the estimated shape and scale parameters of those distributions for the following simulation study. Weibull distributions for shape and scale (months) had parameters of 1.81 and 48, 3.89 and 89, and 3.22 and 9, respectively. We then performed landmark, TD, and standard Cox analyses on the simulated data under different scenarios (Table 2). The sensitivity analysis yielded results consistent with those in Table 2.

The simulation results showed greatest bias using a Weibull distribution when a less than observed scale was assumed (scale = 10)—in other words, when the average time from
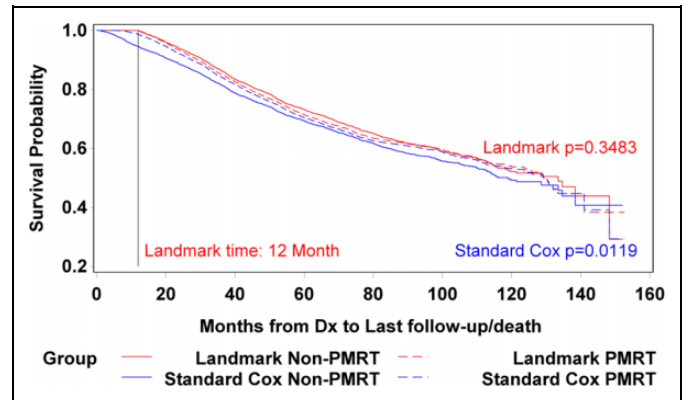

Figure 1. Kaplan-Meier survival probabilities for postmastectomy radiation therapy (PMRT) and non-PMRT groups with 12-month (clinically justified) LM and standard Cox models.

diagnosis to death was 10 months, indicating that in landmark analyses, more than half of the events would occur prior to the landmark time. This explains why the landmark Cox model had less accurate results than TD models, as CIC decreased and RMSE increased. When a scale greater than that observed was assumed (scale = 80)—that is, the average time from diagnosis to death was 80 months, indicating that most events would occur after the landmark time—estimates tended to be less biased as CIC increased and RMSE decreased.

If the hazard rate was assumed constant (shape = 1), with events equally likely to occur at any point in time, or decreasing (shape = 0.5), with events more likely to occur early during follow-up, the standard Cox model always overestimated the true treatment effect. Confidence interval coverage of 0 indicated that none of the CIs generated from the 1000 simulations included the true HR of 1. These results could be explained by the comparison made between event-free immortal time in treatment group and higher/constant risk time in the nontreatment group.

As the hazard rate increased above 1 (shape = 5), with events more likely to occur later during follow-up, the standard Cox model better estimated the true treatment effect. With an increased hazard rate and scale (shape = 5 and scale = 48, 80), the standard Cox model provided unbiased estimates, with mean HRs, CICs, and RMSEs comparable to those of the other models.

However, landmark and TD Cox models had more robust estimates of true treatment effects across all parameter simulation scenarios. The mean HR from all simulations and RMSE of estimates were consistent across all scenarios. When an appropriate landmark time was chosen, the estimates were stable regardless of the shape and scale. These provided unbiased estimates of treatment effect comparable to those from the TD Cox model. However, the performance of landmark analysis was subject to the choice of landmark time. When a longer landmark time was used (in all scenarios when scale = 10), more patients were excluded from the analysis because they experienced the event of interest or were misclassified as not

Table 2. Mean Hazard Ratios (HRs), Confidence Interval Coverage (CIC), and Root Mean Squared Error (RMSE) From Simulation Study.[a]

Survival Distribution (Weibull)				Landmark[b]		TD Cox		Standard Cox	
Shape	Scale (Month)	Exclusion Rate (%)	Event/Death Rate (%)	RMSE	Mean HR (CIC)	RMSE	Mean HR (CIC)	RMSE	Mean HR (CIC)
1.81 (increasing hazard)	48	3.0%	36.1%	0.033	1.00 (95.6)	0.032	1.00 (96.3)	0.097	0.91 (14.2)
	10	28.6%	13.2%	0.065	1.00 (94.8)	0.040	1.00 (95.2)	0.646	0.35 (0)
	80	1.2%	37.3%	0.031	1.00 (95.9)	0.031	1.00 (96.5)	0.048	0.96 (78.4)
0.50 (decreasing hazard)	48	15.0%	27.1%	0.034	1.00 (98.3)	0.033	1.00 (97.8)	0.516	0.48 (0)
	10	25.3%	17.0%	0.055	1.00 (95.9)	0.046	1.00 (96.2)	0.728	0.27 (0)
	80	12.2%	29.4%	0.031	1.00 (98.7)	0.031	1.00 (97.7)	0.447	0.57 (0)
1.00 (constant hazard)	48	8.4%	32.3%	0.034	1.00 (96.8)	0.032	1.00 (96.5)	0.294	0.71 (0)
	10	26.6%	15.6%	0.060	1.00 (95.0)	0.044	1.00 (95.5)	0.694	0.31 (0)
	80	5.3%	34.6%	0.031	1.00 (97.0)	0.030	1.00 (97.3)	0.198	0.80 (0)
5.00 (increasing hazard)	48	0.1%	38.0%	0.032	1.00 (95.5)	0.032	1.00 (96.4)	0.032	1.00 (96.2)
	10	34.9%	4.8%	0.113	1.01 (95.1)	0.035	1.00 (94.4)	0.533	0.47 (0)
	80	0.0%	38.0%	0.031	1.00 (95.3)	0.031	1.00 (96.5)	0.031	1.00 (96.5)
5.00[c]	48	1.5%	37.4%	0.032	1.00 (96.1)				
	10	38.4%	0%	[d]	[d]				
	80	0.5%	38.1%	0.031	1.00 (96.5)				

[a]Confidence interval coverage represents the percentage of 95% confidence intervals that included the true HR in 1000 simulations.
[b]Twelve-month landmark (LM) used unless otherwise specified.
[c]Twenty-four-month LM used.
[d]Not estimable due to 0 events; observational study parameters were shape = 1.81 and scale = 48.

treated before the selected landmark time. The statistical power of such analyses and the generalizability of the results decrease in these situations. In an extreme scenario (shape = 5, scale = 10, and landmark time = 24 months), all events occurred before the landmark and thus all were excluded, so that the treatment effect could not be estimated. Overall, landmark analysis, which would always exclude some early events and misclassify patients with later treatment, provided limited ability to analyze the effectiveness of treatment.

Discussion

Through the use of observational oncology data, we demonstrate how statistical approaches that do not control for time-varying factors may provide spurious findings. Furthermore, our simulation approaches demonstrate the importance of using TD Cox models to estimate treatment effects compared to standard Cox models and landmark analyses. We also demonstrate appropriate methods to minimize ITB in radiation oncology research.

Significant differences existed between patients in the NCDB data set who did and did not have PMRT when the standard Cox model was used, that is, when ITB was not controlled. However, these significant differences were not observed when using TD and 12-month landmark models, that is, when ITB was controlled. Our simulation approach confirmed the findings from the observational data set. The standard Cox model significantly overestimated the treatment effects when no effect was present. Estimates of TD models were closest to the true treatment effect. The landmark model results were highly dependent on landmark timing. A longer landmark time excluded more patients from analysis, thereby

limiting statistical power; a shorter LM time provided an incomplete picture of treatment effects when large immediate effects were expected.

Whether standard Cox produces unbiased estimates of treatment effect can be determined by how quickly the hazard increases, the distribution of immortal time, and the proportion of patients in the treatment group. With an increasing hazard for both time-to-event and treatment (shape > 1), if the mean time-to-event (scale = 48, 80) is longer than the mean time-to-treatment (scale = 9), standard Cox can produce unbiased estimates of treatment effect. However, when the mean time-to-event (scale = 10) is a similar range of the mean time-to-treatment (scale = 9), standard Cox overestimates treatment effect as in all other scenarios.

Our findings show that it is critical to account for ITB when examining how PMRT affects survival in circumstances where nonexperimental study designs are applied. Uncontrolled ITB may create spurious results when using observational data sets and thereby hamper the ability to make evidence-based conclusions in clinical scenarios. Moreover, controlling for ITB does not require any additional data collection and can be accounted for with appropriate application of analytical techniques.

Various studies were conducted in the past that compared analytic approaches to address ITB without simulation.[19-21] Jones et al used a similar simulation approach to account for ITB.[5] However, they did not discuss various simulation scenarios, as we have explained. For example, in the Jones et al's study, the authors varied shape parameters to simulate various scenarios[5]; however, they did not present any data varying the scale parameter. Varying mean time-to-event (scale) is also important because in landmark analysis it may affect the

exclusion and event rate, thereby affecting the power and performance of these models. Our study builds on, and extends, this prior work on ITB. We evaluate the accuracy and precision of our estimator using CIC and RMSE, which is commonly done in simulation studies and is missing from other studies. Without these metrics, it is challenging to objectively compare competing methods in various scenarios, which is the aim of our study.

Through this study, we aim to bring attention to ITB and demonstrate the importance of addressing it in clinical cancer research. However, our study has several limitations. Our study used data from a clinical cancer database where the outcome is survival time, a "censored data" variable; therefore, generalizations must be treated with caution. For example, one study of acute pancreatitis asserted that when mortality is treated as a binary variable, ITB may be more extreme and may affect precision differently than in studies focused on other outcomes.[22] More studies are required to examine whether similar statistical approaches can account for ITB when considering other types of outcomes. Finally, our sample had no missing data. Other studies will be needed to know how ITB behaves when covariates included in the models have missing data that were addressed by multiple imputation methods.

To conclude, appropriate statistical approaches that account for ITB are critical to measure unbiased relationships between TD treatments such as PMRT and survival. Immortal time bias is a complex phenomenon, and many factors (eg, missing data, types of outcome) need to be considered while choosing the appropriate statistical approach to control it. Although ITB was identified some time ago, many research studies do not control for it in the oncology literature. Observational cohort studies should be critically evaluated to identify such biases.

Authors' Note

M.M. conceived the study. P.A., E.M., M.R., and M.M. performed literature search and further refined the study. E.M. performed statistical analysis of analytic sample. M.R. performed statistical analysis and programming for the simulation study with guidance from P.A., E.M., and M.M. P.A. wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript. Because no patient, provider, or hospital identifiers are included in the analytic or simulation components of this study and no protected health information is present, institutional review board approval was not required. Informed consent was not required for the analytic sample or simulation study.

Acknowledgments

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

References

1. Messmer BJ, Nora JJ, Leachman RD, Cooley DA. Survival-times after cardiac allografts. *Lancet*. 1969;1(7602):954-956.

2. Clark DA, Stinson EB, Griepp RB, Schroeder JS, Shumway NE, Harrison DC. Cardiac transplantation in man. VI. Prognosis of patients selected for cardiac transplantation. *Ann Intern Med*. 1971;75(1):15-21.

3. Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol*. 2008;167(4):492-499.

4. Karim ME, Gustafson P, Petkau J, et al. Comparison of statistical approaches for dealing with immortal time bias in drug effectiveness studies. *Am J Epidemiol*. 2016;184(4):325-335.

5. Jones M, Fowler R. Immortal time bias in observational studies of time-to-event outcomes. *J Crit Care*. 2016;36:195-199.

6. Mi X, Hammill BG, Curtis LH, Lai EC, Setoguchi S. Use of the landmark method to address immortal person-time bias in comparative effectiveness research: a simulation study. *Statist Med*. 2016;35(26):4824-4836.

7. van Walraven C, Davis D, Forster AJ, Wells GA. Time-dependent bias was common in survival analyses published in leading clinical journals. *J Clin Epidemiol*. 2004;57(7):672-682.

8. Park HS, Gross CP, Makarov DV, Yu JB. Immortal time bias: a frequently unrecognized threat to validity in the evaluation of postoperative radiotherapy. *Int J Radiat Oncol Biol Phys*. 2012; 83(5):1365-1373.

9. Weberpals J, Jansen L, Carr PR, Hoffmeister M, Brenner H. Beta blockers and cancer prognosis—the role of immortal time bias: a systematic review and meta-analysis. *Cancer Treat Rev*. 2016; 47:1-11.

10. Fisher B, Brown A, Mamounas E, et al. Effect of preoperative chemotherapy on local-regional disease in women with operable breast cancer: findings from National Surgical Adjuvant Breast and Bowel Project B-18. *J Clin Oncol*. 1997;15(7):2483-2493.

11. Overgaard M, Hansen PS, Overgaard J, et al. Postoperative radiotherapy in high-risk premenopausal women with breast cancer who receive adjuvant chemotherapy. Danish Breast Cancer Cooperative Group 82b Trial. *N Engl J Med*. 1997;337(14):949-955.

12. Overgaard M, Jensen MB, Overgaard J, et al. Postoperative radiotherapy in high-risk postmenopausal breast-cancer patients given adjuvant tamoxifen: Danish Breast Cancer Cooperative Group DBCG 82c randomised trial. *lancet*. 1999;353(9165):1641-1648.

13. Ohri N, Moshier E, Ho A, et al. Postmastectomy radiation in breast cancer patients with pathologically positive lymph nodes after neoadjuvant chemotherapy: usage rates and survival trends. *Int J Radiat Oncol Biol Phys*. 2017;99(3):549-559.

14. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol*. 1983;1(11):710-719.

15. Dafni U.Landmark analysis at the 25-year landmark point. *Circ Cardiovasc Qual Outcomes*. 2011;4(3):363-371.

16. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. *Annu Rev Public Health*. 1999;20:145-157.

17. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models by Ralf Bender, Thomas Augustin and Maria Blettner, *Statistics in Medicine* 2005; 24: 1713-1723. *Stat Med*. 2006;25(11):1978-1979.

18. Naimi AI, Cole SR, Westreich DJ, Richardson DB. A comparison of methods to estimate the hazard ratio under conditions of time-varying confounding and nonpositivity. *Epidemiology*. 2011; 22(5):718-723.

19. Matok I, Azoulay L, Yin H, Suissa S. Immortal time bias in observational studies of drug effects in pregnancy. *Birth Defects Res A Clin Mol Teratol*. 2014;100(9):658-662.

20. Daniel S, Koren G, Lunenfeld E, Levy A. Immortal time bias in drug safety cohort studies: spontaneous abortion following non-steroidal anti-inflammatory drug exposure. *Am J Obstet Gynecol*. 2015;212(3):307.e1-e6.

21. Giobbie-Hurder A, Gelber RD, Regan MM. Challenges of guarantee-time bias. *J Clin Oncol*. 2013;31(23):2963-2969.

22. Liang C, Seeger JD, Dore DD. Implications of immortal person-time when outcomes are nonfatal. *Ann Epidemiol*. 2016;26(3):212-217.