When observational studies can give wrong answers: the potential of immortal time bias

Dekkers, O.M.; Groenwold, R.H.H.

When observational studies can give wrong answers: the potential of immortal time bias

Olaf M Dekkers[ID][1,2] and **Rolf H H Groenwold**[1,3]

[1]Department of Clinical Epidemiology, [2]Department of Endocrinology, and [3]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands

Correspondence should be addressed to O M Dekkers
Email
o.m.dekkers@lumc.nl

Abstract

Immortal time bias should always be considered in an observational study if exposure status is determined based on a measurement or event that occurs after baseline. This bias can lead to an overestimation of an effect, but also to an underestimation, which is explained. Several approaches are illustrated that can be used to avoid immortal time bias in the analysis phase of the study; a time-dependent analysis to avoid immortal time bias optimizes the use of available information.

Suppose, researchers want to study whether getting children decreases mortality risk (through some endocrine mechanism). Based on register data, they select women with children and all other women without children and in the analysis, they start follow-up from birth. Likely, women with children will show a lower mortality risk than those without. Why? Because women who get pregnant and give birth to a child could not have died in the years prior to pregnancy, otherwise they would have been included in the other group. Obviously, this study is affected by the so-called immortal time bias.

Immortal time bias is a well-known bias in observational studies (1). The bias can lead to paradoxical and implausible results, such as the finding that patients with a second primary tumor after treatment for head or neck cancer have much better survival compared to patients without a second primary tumor (2). The present paper discusses the origin of immortal time bias applied to examples from endocrinology. It also provides an overview of approaches to avoid immortal time bias in the analysis stage of the study.

Immortal time bias: the basic concept

Think of an observational study, assessing the effect of radiotherapy (RT) after pituitary surgery to prevent disease recurrence in acromegaly. In an observational setting the dataset will reflect daily clinical practice (3), and the time between surgery and RT will differ between patients. In a naïve (and incorrect) statistical approach, follow-up starts at the time of surgery and disease recurrence is compared between patients with and without RT. As in this approach, the time between surgery and RT is not accounted for adequately, the estimated effect of RT will likely be biased.

To exemplify, we think of this study as consisting of 20 patients, with 10 years follow-up; 10 patients are treated with RT. Acromegaly recurs in 2 patients after RT and in 3 patients without RT. A naïve analysis suggests that RT reduces recurrence risk in acromegaly (20% vs 30%), with a risk ratio of 0.66 (we are not considering the imprecision of the risk estimate, and for the sake of argument, one might think of this study is 1000 times larger; also confounding is assumed not to be an issue). Likewise, the rate ratio is estimated to be 0.64. As we will

Published by Bioscientifica Ltd.

see below, both estimates are in fact biased due to the presence of immortal time.

Valid risk estimation requires valid classification of the number of events (numerator) but also of the persons or person-time at risk (denominator). When assessing or comparing the effect (or the risk) of a certain exposure, it should be taken into account whether the exposure changes over time and whether in the study exposure is defined based on information that becomes available after the start of follow-up (such as RT in the example above, or an attained lab value). For patients in the aforementioned study, there will be, to a varying degree, time between surgery and RT (see Fig. 1 for a graphical representation of the hypothetical study). For patients treated with RT, the time between surgery and RT is considered 'immortal' as in this period they could not have developed the outcome under study (recurrence). Had such a patient developed a

recurrence pre-RT, this patient would have been classified as a non-RT patient. If, for example, the second patient in the figure would have developed a recurrence before RT, this patient would have been classified as non-RT patient with a recurrence.

Similarly, immortal time bias can occur in studies that assess the effect of target lab values (such as IGF-1 below a certain threshold, or mitotane above a certain level). If these exposure values or categories occur after the start of follow-up and if these values are used to categorize patients at baseline, the naïve statistical approach can be biased.

It is worth noting that this immortal time bias does not apply to an intention-to-treat (ITT) analysis in a randomized controlled trial (RCT). In an ITT analysis, exposure status is defined at the start of follow-up even if the exposure is scheduled sometime after baseline. In an ITT analysis, exposure status can thus change over the course of follow-up for some patients, these changes, however, are neglected in the analysis. If we consider the RT example, a RCT could randomize between RT and no RT directly after surgery, which also defines the start of follow-up. In such a study, a patient randomized to RT with a recurrence before RT, would still be analyzed in the RT arm.

Figure 1

Representation of immortal time bias in a study of radiotherapy after pituitary surgery. Follow-up starts at the time of surgery (0 years of follow-up). Vertical marks indicate the start of radiotherapy. Solid dots represent events (recurrence of disease) and arrowheads end of follow-up. Dashed lines represent time without RT, solid lines time after RT. Classifying the full follow-up for the upper 10 patients as being exposed to radiotherapy (i.e. including the dashed observation time), introduces immortal time bias.

Potential solutions to avoid immortal time bias

Three analysis approaches can be used that avoid immortal time bias. It should be noted that the different approaches answer slightly different research questions and can give different numerical answers.

1. Time-dependent exposure analysis

A first statistical approach is to classify person-time and not persons (4). In such an approach follow-up time is split up into time periods according to actual exposure status, and all patients can contribute person-time to different exposure categories. For example, person-time at risk can be stratified in two periods for the RT treated patients: a period before RT and a period after RT. For the first person in Fig. 1, the first 2 years are classified as non-RT person-time, the last 3 years are classified as RT person time. It is obvious that counting the whole period as person-time at risk as if 'after RT', will bias the results, with an underestimation of recurrence risk for RT (denominator includes too much person-time), and an overestimation
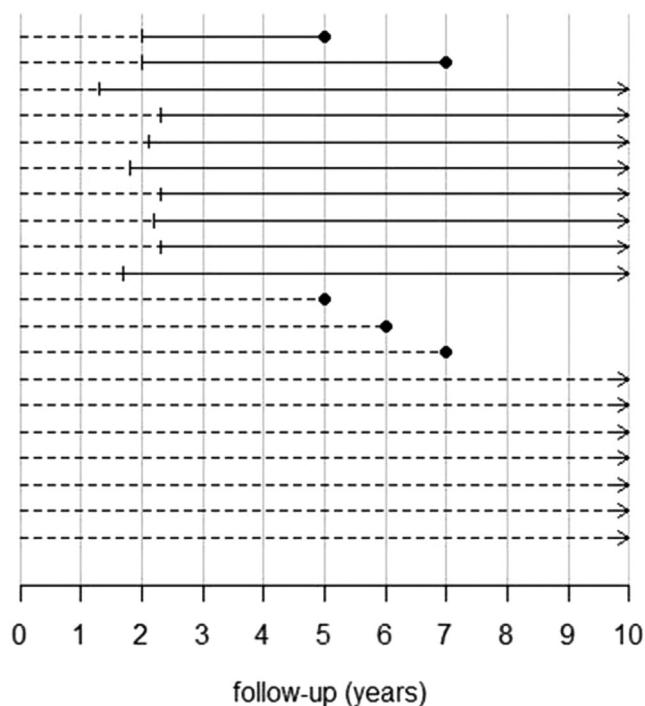
of recurrence risk for the non-RT group (denominator includes too little person-time).

Back to the example of 20 patients, the naïve statistical approach provides a (biased) risk ratio of 0.66. Let's assume the time between surgery and RT to be 2 years on average, and that recurrence of the disease occurs on average 6 years after surgery, irrespective of RT. A time-dependent approach, avoiding immortal time bias, would classify the average 2 years before RT as person-time at risk for the non-RT group (i.e. 20 person-years). Furthermore, the ten subjects who do not receive RT at all contribute 88 person-years (seven times 10 years for those in whom recurrence does not occur and three times 6 years for those in whom recurrence occurs). Based on these numbers the risk estimate is 3/108 person-years in the non-RT group. The total follow-up time in the RT group is 72 person-years (eight times 8 years for those in whom recurrence does not occur and two times 4 years for those in whom recurrence occurs) and the risk estimate is 2/72 person-years. This yields a relative risk of 1.0.

This approach is especially appealing for exposures with many levels or categories (Mitotane levels or IGF1 values) as all measurements can be taken into account. It should be emphasized that also confounders should be dealt with in a time-dependent way, which considerably adds to the complexity of the analysis, especially if persons can have many category-switches. Not all cohort studies have adequate information on time-dependent confounders, which hampers an optimal time-dependent approach.

2. Landmark approach

In a landmark approach, exposure status (e.g. RT yes/no) is determined for all patients at a certain predefined point in time (landmark). This landmark is the same for all patients and is also used as the start of follow-up. As the start of follow-up and determination of exposure by design now coincide, no immortal time is included. However, any change in exposure status after the landmark should then be ignored in this approach. For the RT example, the landmark could be set at, for example, 6 months after surgery, at which time-point it is determined for all patients whether they have been treated with RT or not, and this determines the two groups for comparison. Such analysis would then give an answer to the question of whether RT that is started in the first 6 months reduces recurrence risk compared to RT that is started later or not started at all. A landmark approach for the RT example, with the landmark set at 2.5 years, would classify ten subjects as being on RT, of whom in two the disease recurs. The other ten are classified as not being on RT, of whom three develop a disease recurrence. The resulting relative risk would be 0.67, which can be interpreted as the effect of RT among those who survive at least up to 2.5 years after surgery. The landmark analysis can also be used when two treatments are compared.

3. Matching on time

When matching on time, a patient who gets exposed is matched to a non-exposed patient who has been followed-up for the same amount of time. For example, if a patient receives RT 1 year after pituitary surgery, this patient is matched to a patient without RT 1 year after surgery and both are followed-up from that time onwards. This is similar to a landmark approach, be it that the landmark time points are the same for the whole cohort, whereas in case matching the time points are commonly defined by the timing of RT. However, the matching does not work (well) for exposures that change frequently over time (IGF1 levels); it is also inefficient if exposure occurs late after baseline.

Concluding remarks

Immortal bias should always be considered in observational studies if exposure status is determined based on a measurement or event that occurs after baseline. Several data analytic approaches can be used to avoid this specific bias. A time-dependent approach optimizes the use of available information, but the analysis can get complicated for exposures that change categories many times. Landmark approaches and matching are less efficient as they do not use all follow-time in the cohort. For conditions with a stable course (such as acromegaly) and an exposure that occurs reasonably short after the start of the cohort defining event (pituitary surgery for acromegaly), a landmark approach is a reasonable alternative to a time-dependent analysis.

References

1 Suissa S. Immortal time bias in pharmaco-epidemiology. *American Journal of Epidemiology* 2008 **167** 492–499. (https://doi.org/10.1093/aje/kwm324)

2 Rennemo E, Zatterstrom U & Boysen M. Impact of second primary tumors on survival in head and neck cancer: an analysis of 2,063 cases. *Laryngoscope* 2008 **118** 1350–1356. (https://doi.org/10.1097/MLG.0b013e318172ef9a)

3 Dekkers OM & Groenwold RHh. Study design: what's in a name? *Journal of Endocrinology* 2020 **183** E11–E13. (https://doi.org/10.1530/EJE-20-0873)

4 Lévesque LE, Hanley JA, Kezouh A & Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ* 2010 **340** b5087. (https://doi.org/10.1136/bmj.b5087)