Environmental Research 140 (2015) 275–281

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Disease fatality and bias in survival cohorts

Vaughn Barry*, Mitchel Klein, Andrea Winquist, Lyndsey A. Darrow, Kyle Steenland

Departments of Environmental Health and Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Road, Atlanta, GA 30322 USA

ARTICLE INFO

Article history:
Received 23 October 2014
Received in revised form
19 March 2015
Accepted 31 March 2015

Keywords:
Bias
Fatality
Simulation
Survivor cohort

ABSTRACT

Objectives: Simulate how the effect of exposure on disease occurrence and fatality influences the presence and magnitude of bias in survivor cohorts, motivated by an actual survivor cohort under study.
Methods: We simulated a cohort of 50,000 subjects exposed to a disease-causing exposure over time and followed forty years, where disease incidence was the outcome of interest. We simulated this 'inception' cohort under different assumptions about the effect of exposure on disease occurrence and fatality after disease occurrence. We then created a corresponding 'survivor' (or 'cross-sectional') cohort, where cohort enrollment took place at a specific date after exposure began in the inception cohort; subjects dying prior to that enrollment date were excluded. The disease of interest caused all deaths in our simulations, but was not always fatal. In the survivor cohort, person-time at risk began before enrollment for all subjects who did not die prior to enrollment. We compared exposure–disease associations in each inception cohort to those in corresponding survivor cohorts to determine how different assumptions impacted bias in the survivor cohorts. All subjects in both inception and survivor cohorts were considered equally susceptible to the effect of exposure in causing disease. We used Cox proportional hazards regression to calculate effect measures.
Results: There was no bias in survivor cohort estimates when case fatality among diseased subjects was independent of exposure. This was true even when the disease was highly fatal and more highly exposed subjects were more likely to develop disease and die. Assuming a positive exposure–response in the inception cohort, survivor cohort rate ratios were biased downwards when case fatality was greater with higher exposure.
Conclusions: Survivor cohort effect estimates for fatal outcomes are not always biased, although precision can decrease.

1. Introduction

When eligibility for a study is dependent on survival to a certain time point (the enrollment date) after the exposure of interest begins, and hence after the risk for the disease of interest begins (called a 'survivor' or 'cross-sectional' cohort), measures of effect describing the relationship between exposure and disease onset may be subject to selection bias, and may therefore differ from these measures in the underlying target cohort of interest (the 'inception' cohort) (Hudson et al., 2005; Delgado-Rodríguez et al., 2004; Rothman et al., 2008). This is generally not an issue when the incidence of a non-fatal disease is the outcome of interest, since few people who develop disease would be lost in the survivor cohort. However, when disease of interest is sometimes fatal, those who die of the disease prior to the enrollment date will be excluded from survivor cohorts. Selection bias could occur if those who would have been eligible for study in the underlying inception cohort, but die from the disease of interest before enrollment in a 'survivor' cohort, have a different exposure–disease relationship than subjects that survive long enough to enroll in the survivor cohort. It is natural to assume that the presence and magnitude of this bias may be a function of the specific survival pattern associated with the disease outcome of interest. If exposure causes disease and those with higher exposure are more likely to get disease and then die, and hence fail to enroll in the survivor cohort, it might be assumed that exposure–disease relationships may be biased in the survivor cohort versus the original underlying inception cohort. Yet this may not always be the case. Disease fatality and the specific relationships between the exposure, disease, and death may all be factors that can determine whether bias is induced or averted in studies where participation is dependent on survival. Our focus is on examining the presence and magnitude of bias associated with diseases with different survival patterns in survivor cohorts.

It should be noted that our 'survivor' cohort is not the same as a typical 'left-truncated cohort', which is usually thought of as a

* Corresponding author. Fax: +1 404 727 8744.
E-mail address: vbarry@emory.edu (V. Barry).

http://dx.doi.org/10.1016/j.envres.2015.03.039
0013-9351/

survivor cohort where no person-time at risk or observed events prior to enrollment can be included (Applebaum et al., 2011; Schisterman et al., 2013). For example, in mortality studies where death from a certain cause is the outcome of interest, all such deaths prior to enrollment will be missed, and correspondingly all person-time at risk due to a given exposure prior to enrollment must also be excluded, to avoid including 'immortal' person time in the denominator of rates, biasing such rates downward (Rothman et al., 2008). In our survivor cohort, disease incidence (whether it is eventually fatal or not) is the outcome of interest, person-time at risk begins before enrollment for survivors who reach the enrollment date, and the incident disease outcome of interest can occur before enrollment. Our survivor cohort, is however, *partly* 'left-truncated', in that people who die of the disease of interest prior to enrollment are missing from the cohort. However, there is no 'immortal' person-time in our survivor cohort.

This work was motivated by a specific situation that involved emissions of a chemical (perfluorooctanoic acid or 'PFOA') from a chemical plant over a fifty year period (C8 Health Project, 2012). The chemical contaminated the local drinking water supply of several nearby communities beginning in the 1950's and continuing through the year 2000. In the year 2005, a series of community health studies was initiated to determine whether exposure to the chemical had caused adverse health and disease in community residents (Frisbee et al., 2009; Winquist et al., 2013). Most of the participants in the cohort had to be alive in 2005 to enroll in these studies (about 10% of the cohort were workers at the chemical plant and they did not have this restriction). One common assumption about a survivor cohort of this type is that if there is a true positive exposure–disease association, any measure of effect assessing the relationship between the chemical exposure and a highly fatal disease will be negatively biased, because those who developed fatal diseases were 1) more likely to have had higher exposures, and 2) more likely to die before 2005, and thus less likely to be included as participants in the studies. Here we examine whether this assumption holds.

One of the cancer types in which we were interested in our cohort was kidney cancer, since one prior occupational mortality study found evidence of a positive PFOA-response for this disease (Steenland and Woskie, 2012). We found a moderately significant relationship between the chemical and kidney cancer incidence in our survivor cohort of 32,254 residents living near the chemical plant (Barry et al., 2013). There were 113 kidney cancer cases in our study. Because the estimated five year survival rate after kidney cancer diagnosis in the U.S. population is 72.4%, there were most certainly residents who developed kidney cancer and died before enrollment in 2005 (Howlader et al., 2013). The issue is whether those who did not survive until enrollment had a different PFOA-kidney cancer relationship compared to those that survived, creating a selection bias in our survivor cohort. We wondered if our estimates of the PFOA-kidney cancer association were negatively biased in our survivor cohort, and this was the motivation for us to design the simulation which is the subject of this paper.

We used simulated data to explore how assumptions regarding the disease fatality and the effect of exposure on disease fatality in an inception cohort influence the presence and magnitude of bias in survivor cohorts. Note that we assumed that all exposed subjects were equally susceptible to the effects of exposure in causing disease, hence the problem we address is different from that of the depletion of a sub-set of susceptibles among the exposed (the 'frailty' effect discussed by Hernan) (Hernán, 2010). Note also that we ignore the issue of loss-to-follow-up for other reasons than death from the disease of interest (eg, migration in and out of the study area) or death from other causes (such deaths are not included in our simulation, where everyone who dies, dies from the disease of interest). Here we assume that other reasons for loss-to-follow-up in a survivor cohort are ignorable, ie, will not bias observed exposure–response relationships.

Our overall strategy was to simulate a group of subjects with different exposure levels and then follow them through time to see who developed disease and who died from that disease (i.e. the inception cohort). We assumed a "true" effect of exposure on disease in this inception cohort. Next, we created a subset of the inception cohort (i.e. the survivor cohort) which survived until a given point in time. We then calculated the estimated effect of exposure on disease in the survivor cohort and compared it to the inception cohort estimate. We examined how different assumptions about disease fatality impacted bias in the survivor cohort estimates. For simplicity we assumed no other causes of death, other than from the disease of interest.

2. Materials and methods

2.1. Inception Cohort: exposure

We simulated a prospective cohort of 50,000 subjects. We assumed that each subject was followed from age 20 to age 60 years (40 years per subject) or death with no loss to follow up. Each subject was assigned an age at which he or she began experiencing exposure during the 40-year period. Additionally, each subject was exposed for a specific number of years (i.e. cumulative exposure). We assumed that exposure intensity was constant over time and at a rate of 1 unit exposure per year. Because intensity was constant, duration and cumulative dose were equivalent. This design was roughly based on the idea of an exposure being present in a given region over a 40 year period and is equivalent to enrolling 50,000 twenty-year old subjects in to a study in a given year. We imagined that subjects could move in and out of the region at different ages (but were followed whether they were in or out, ie, their person-time started at age 20 until disease or end-of-follow-up, whichever came first) and for different numbers of years and thus be exposed for different amounts of time. This scenario roughly corresponded to our study of PFOA in the mid-Ohio valley.

We generated age at first exposure for each subject by drawing ages randomly from a uniform distribution between 20 and 40. Thus, age at first exposure was equally likely to occur anytime between age 20 and 40 years with an average of 30 years (range = 20–40 years). We assumed that the cumulative amount of exposure each subject had would follow a normal distribution with a mean of 20 and standard deviation of 5, left truncated at 0. Consequently, on average, each subject was exposed to 20 units of exposure over their lifetime but any given subject could have been exposed from 1 to 40 units of exposure during their lifetime. We chose these numbers because they gave us a cohort with a wide-range of exposure levels, similar to the original cohort of residents living near the chemical plant.

Note that using these assumptions, some subjects (~8%) had an assigned cumulative exposure that was too large given the age at which they were first exposed. For example, a subject with an assigned cumulative exposure of 25 units who began the exposure at age 39 years will have had only 21 units at the age of 60 years (which is the end of follow-up). This meant that the right tail of the normal distribution curve that generated cumulative exposure was partially truncated. After taking this into account, on average each subject was exposed to 20 units of exposure during their lifetime and any given subject could have been exposed to anywhere from 1 to 37 units of exposure.