EDE10-234
BRIEF REPORT (To be published with commentary by Hernández-Diaz)

Title: TIME-WINDOW BIAS IN CASE-CONTROL STUDIES: STATINS AND LUNG CANCER

Authors: Samy Suissa, Sophie Dell'Aniello, Sarah Vahey and Christel Renoux

Authors' institutions:
From the Center for Clinical Epidemiology, Lady Davis Research Institute, Jewish General Hospital, and the Departments of Epidemiology and Biostatistics and of Medicine, McGill University, Montreal, Canada.

Address for correspondence:
Samy Suissa
Centre for Clinical Epidemiology, Jewish General Hospital
3755 Cote Ste-Catherine
Montreal, Québec, Canada H3T 1E2
Tel: 514-843-1564 Fax: 514-843-1451
E-mail: samy.suissa@mcgill.ca

Running title: Time-window bias in case-control studies

Funding: Funded by grants from the Canadian Institutes of Health Research (CIHR) and the Canadian Foundation for Innovation (CFI).

Word count, abstract: 162
Word count, main text: 1460

No. Text pages: 10
No. Tables:
No. Figures: 1

Date submitted: 28 April 2010
Date accepted: 18 October 2010

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Editors' note: *A commentary on this article appears on page* ***xxx****.*

ABSTRACT

Time-related biases in cohort studies can produce illusory "beneficial" effects of medications due entirely to an artefact of the analytic design. We describe "time-window bias" in the context of a case-control study, reporting that statin use was associated with a 45% reduction in the incidence of lung cancer. This bias results from the use of time-windows of different lengths between cases and controls to define time-dependent exposures. We illustrate the bias using a population of 365,467 patients from the United Kingdom's General Practice Research Database, including 1786 incident cases of lung cancer during 1998-2004. The case-control approach used in the published study yielded a rate ratio of lung cancer incidence of 0.62 with statin use (95% confidence interval = 0.55-0.71). A case-control approach that properly accounts for time produces a rate ratio of 0.99 (0.85-1.16)—suggesting no benefit of statins on lung cancer risk. We show analytically that the magnitude of the bias is proportional to the ratio of the unequal time-window lengths.

Time-related biases have affected several observational studies reporting impressive results on the effectiveness of certain medications in reducing the incidence of major disease outcomes.[1-4] These biases have been described within cohort-study designs;[5] case-control studies have not been suspected of being susceptible to such biases. Recently a case-control study, conducted using an administrative health database, reported that statins were associated with a 45% reduction in the risk of lung cancer. Since this effect is so large and with such important implications given the poor prognosis of lung cancer, that alternative explanations must be investigated. We show that these effects are essentially due to a time-related bias that we call "time-window bias."

In this paper, we describe and quantify time-window bias in case-control studies, and illustrate its impact by replicating the published case-control study of statin use and lung cancer risk using data from the United Kingdom General Practice Research Database.

THE PUBLISHED CASE-CONTROL STUDY

The previously published study used the U.S. Veteran's Affairs (VA) database to identify 483,733 patients between October 1998 and June 2004.[6] All 7280 patients diagnosed with lung cancer during this period formed the case series, while the 476,453 remaining patients were taken as controls. Statin exposure was defined as any prescription for a statin during the observation period until "the data collection completion date." Thus, statin exposure was measured as any prescription for a statin prior to the date of lung cancer diagnosis for the cases, and prior to the end of the observation period for the controls. The analysis found a 45% reduction in the rate of lung cancer with any statin use

(adjusted odds ratio [OR] = 0.55 [95% confidence interval [CI] = 0.52-0.59]). With more than 4 years of statin use, the reduction was 77% (0.23 [0.20-0.26]).

DESCRIPTION OF THE BIAS

The source of time-window bias arises from the methods used to select controls and to measure their exposure. The study population was observed for 67 months, from 1 October 1998 until 1 June 2004. The observation period was necessarily less than 67 months for the cases occurring over the course of 1998-2004, which likely closer to 67 months for the controls. As a result, exposure assessment to statins—defined as any prescription for a statin during the observation period—was based on a shorter time-span for cases than controls. Sheerly on the grounds of time-length, we can expect that a subject with a shorter observation period was less likely to be exposed to statins than one observed for the entire 67-month span. Specifically, a lung-cancer case had less person-time to receive a prescription than one who did not have lung cancer for the entire 67 months. This can result in an over-representation of unexposed cases and a spurious appearance of benefit of the drug (Figure). The magnitude of this bias is derived analytically in the eAppendix (http://links.lww.com).

ILLUSTRATION OF THE BIAS

To illustrate time-window bias, we used the General Practice Research Database, a computerized primary care database that contains medical information on more than 6 million patients registered in approximately 400 general practices in the UK.[7-9] To replicate the VA study, we identified all patients aged 50 to 90 years between 1 October 1998 and 1

June 2004, with entry defined as the later of the age or calendar-date criteria. Patients with a prior diagnosis of lung cancer or without smoking data were excluded. During the observation period, we identified all cases of lung cancer and obtained information on their statin prescriptions from entry until cancer diagnosis date. For other subjects, all statin prescriptions during the observation period were identified.

To analyze these data, we first used a straightforward full cohort analysis in which all person-days of follow-up were classified as non-exposed until the first statin prescription, and classified as exposed thereafter. The corresponding data analysis was based on Poisson regression to estimate the rate ratio (RR) of lung-cancer incidence associated with statin use. Second, we replicated the published case-control design, where controls were selected as all the non-cases and exposure for controls was defined as a statin prescription prior to end of observation (time-dependent sampling).[6] The corresponding data analysis was based on unconditional logistic regression to compute the odds ratio as an estimator of the rate ratio. Third, we used a random sample of person-moments of size 10 times the number of cases, selected from all person-moments (person-days) generated by the cohort (time-dependent sampling), according to the principle of incidence density sampling.[10,11] Exposure was defined as a statin prescription any time prior to the selected control person-moment. The analysis was also based on unconditional logistic regression.

The study population consisted of 365,467 patients, followed for a mean 3.0 years, during which 1786 incident cases of lung cancer were diagnosed (rate = 1.65 per 1000 per year). Table 1 provides the results of the full cohort analysis, which illustrates the extent of the potentially misclassified person-time corresponding to 11% (102,628/935,724) of all

unexposed person-years. The resulting crude rate ratio that properly accounts for this person-time is 1.11 (95% CI = 0.98-1.27), while the adjusted rate ratio is 1.02 (0.90-1.17).

Table 2 describes the characteristics of the lung cancer cases and controls, showing the well-known risk factors of male sex, older age, and smoking. The two control groups were similar in their characteristics at cohort entry. Table 3 displays the findings using the two case-control approaches. The time-independent approach in the VA study yielded a rate ratio of lung cancer incidence associated with statin use of 0.62 (95% CI = 0.55-0.71), suggesting a large protective effect. In contrast, the time-dependent approach using controls sampled from all person-moments produced a rate ratio of 0.99 (0.86-1.16), indicating no effect. The different time-window lengths and control sampling approaches led to the misclassification of 7% of the controls from unexposed to exposed.

DISCUSSION

The length of the time-window used to ascertain exposure is crucial in case-control studies of time-dependent exposures. We have confirmed that the strong protective effect of statins on lung cancer found in the VA study was spurious due to the longer time-window for measuring exposure in controls than in cases. The "protective" effect of statin use disappeared once time was properly accounted for by control selection.

Time-related biases such as the one due to "immortal" person-time have generally centered on cohort studies, mostly database studies of medication effects.[1,4,5] However, the majority of case-control studies in pharmacoepidemiology are conducted using existing computerized databases and thus one inherently within some form of cohort. Consequently, it is conceivable that similar time-related biases can also affect these case-

control studies. We show the importance of insuring an equal time-window to measure exposure for cases and controls. The spurious protective effect of statins on lung cancer incidence was introduced by selecting controls at the last available person-moment of follow-up, with statin exposure defined as any use prior to this date. As a result, the average time period available to measure exposure was longer for controls than for cases. Because of the time-dependent nature of the statin exposure, this difference led to an over-representation of exposed controls and an apparent protective effect of statins. Had controls been selected from the universe of all person-moments instead of the last one, the resulting exposure measurement for controls and cases would have been based on a more similar time span. Using a more proper method for control selection, statin use was no longer associated with a decreased risk of lung cancer.

This bias is not uncommon. It occurred in a recent study in which the case-control analysis used "the date of the end of the follow-up period for the controls" to find a 41% reduction in the risk of suicide associated with antiepileptic drugs in patients with epilepsy.[12] This bias will occur in case-control studies conducted from computerized health databases in which the time span of the available data easily lends itself to differential time windows. In contrast, most field-based case-control studies select controls around the same calendar time that the case occurs, thus avoiding differential time windows. However, some field-based case-control studies select controls from registries of patients, so that their time span may differ from that of the cases. Another case-control situation that may give rise to this bias would be drug exposures that are specific to a disease, so that disease duration itself may lead to differential time windows for exposure.

The increasing availability of large computerized health databases represents a unique opportunity to study drug effects but it also presents important methodological challenges. Among them, the failure to properly take time into account at the design stage of a case-control study can directly affect exposure measurement and produce spurious associations. We show that time-window bias can occur in case-control studies if time is not properly considered in the selection of controls, and can create an artificial appearance of drug benefit.

Figure.

Description of six typical cases and controls with exposure defined as prescriptions dispensed prior to the end of observation period.

Reference List

(1) Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf.* 2007;16(3):241-249.

(2) Lash TL, Cole SR. Immortal person-time in studies of cancer outcomes. *J Clin Oncol.* 2009;27(23):e55-e56.

(3) Gardarsdottir H, Egberts TC, Stolker JJ, Heerdink ER. Duration of antidepressant drug treatment and its influence on risk of relapse/recurrence: immortal and neglected time bias. *Am J Epidemiol.* 2009;170(3):280-285.

(4) Levesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *Br Med J.* 2010;340:b5087.

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

(6) Khurana V, Bejjanki HR, Caldito G, Owens MW. Statins reduce the risk of lung cancer in humans: a large case-control study of US veterans. *Chest.* 2007;131(5):1282-1288.

(7) Garcia Rodriguez LA, Perez GS. Use of the UK General Practice Research Database for pharmacoepidemiology. *Br J Clin Pharmacol.* 1998;45(5):419-425.

(8) Lawrenson R, Williams T, Farmer R. Clinical information for research; the use of general practice databases. *J Public Health Med.* 1999;21(3):299-304.

(9) Jick SS, Kaye JA, Vasilakis-Scaramozza C et al. Validity of the general practice research database. *Pharmacotherapy.* 2003;23(5):686-689.

(10) Miettinen OS. Estimability and estimation in case-referent studies. *Am J Epidemiol.* 1976;103(2):226-235.

(11) Miettinen OS. Etiologic research: needed revisions of concepts and principles. *Scand J Work Environ Health.* 1999;25(6):484-490.

(12) Arana A, Wentworth CE, Ayuso-Mateos JL, Cellai F FM. Suicide-related events in patients treated with antiepileptic drugs. *N Engl J Med.* 2010;363(6):542-551.

Table 1

Full cohort time-dependent analysis of lung cancer risk associated with statin use using a cohort of 365,467 subjects from General Practice Research Database, of which 1,786 developed lung cancer during follow-up.

	No. Person-years	**No. Incident lung cancer cases**	**Rate[a]**	**Rate ratio (95% CI)**
Statin use	146,253	265	1.81	1.11 (0.98-1.27)
No statin use [b]	935,724	1521	1.63	1.00
During entire follow-up	833,096	1521		
Prior to first statin prescription	102,628	0		

[a] Incidence rate of lung cancer per 1000/year.

[b] Reference category.

Table 2
Characteristics of cases of lung cancer and controls within the cohort of 365,467 subjects identified from the GPRD, with controls selected using two approaches: the time-independent approach employed in the VA study,[6] and a time-dependent approach based on incidence density sampling of person-moments.

	Cases	**Controls**	
		Time-independent sampling	**Time-dependent sampling**
Number	(n = 1786)	(n = 365,467)	(n = 17,860)
Male sex; %	61	46	45
Age in years; mean	69	63	63
Body weight status; %			
Obese	11	4	4
Non obese	55	17	15
Unknown	34	79	81
Smoker; %	85	52	51
Diabetes; %	6	7	7

Table 3
Comparison between the two approaches of control selection, namely the time-independent technique used in the VA study[6] and the time-dependent technique based on incidence density sampling, in estimating the rate ratio (RR) of lung cancer associated with statin use using a case-control design within the cohort of 365,467 subjects identified from the General Practice Research Database.

Sampling of controls	Cases	Controls	Crude RR (95% CI)	Adjusted[a] RR (95% CI)
Time-independent				
Number	1786	363,681		
Statin use; %	14.8	20.8	0.66 (0.[illegible]-0.76)	0.62 (0.55-0.71)
No statin use; %[b]	85.2	79.2	1.00	1.00
Time-window length (years); mean	2.28	2.9[illegible]		
Time-dependent				
Number	[illegible]786	17,860		
Statin use; %	14.8	13.6	1.10 (0.96-1.26)	0.99 (0.85-1.16)
No statin use; %[b]	85.2	86.4	1.00	1.00
Time-window length (years); mean	2.28	2.14		

[a] Adjusted for sex, age, body mass index, diabetes, and smoking status
[b] Reference category

Figure