eAPPENDIX

Time-Window Bias in Case-Control Studies: Statins and Lung Cancer

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

MAGNITUDE OF THE BIAS

To assess the magnitude of this bias, we considered the situation where the statin prescriptions are distributed randomly over time for all subjects in the study population, implying no association between statins and lung cancer incidence, namely an odds ratio (OR) of 1. First, note that generally, if we have a follow-up time T distributed as exponential with mean m and a number of prescriptions X that are, conditional on T, distributed as Poisson with rate r, then

$$P(X=0|\ T=t) = e^{-r\,t}$$
$$f(t) = (e^{-t/m})/m, \text{ the density function of T,}$$

so that $P(X=0) = 1/(rm+1)$, which is used to compute OR.

To compute the expected proportion of non-users separately among the cases and controls, we assumed an exponential distribution, with mean m_1 years for the time to lung cancer diagnosis for the cases, and mean m_0 for the controls. We also assumed a Poisson distribution for the number of statin prescriptions per subject, conditional on the observation time, with a mean rate r prescriptions per person-year, common to cases and controls. As shown in eTable 1, the OR of lung cancer diagnosis associated with any statin exposure in such a case-control setting with no association would be

$$OR = (1-(1/(r\,m_1+1))\ (1/(r\,m_0+1)) / ((1-(1/(r\,m_0+1))\ (1/(r\,m_1+1)))$$
$$= m_1 / m_0$$

Thus, in the absence of an association between statins and lung cancer, namely an odds ratio of 1, the OR estimated by this approach is expected to be the ratio of the mean time to lung cancer diagnosis for the cases to the mean time of observation for the controls. Note that if the mean rate of prescriptions per person-year is different for cases (r_1) than for controls (r_0) then

$$OR = (1-(1/(r_1\,m_1+1))\ (1/(r_0\,m_0+1)) / ((1-(1/(r_0\,m_0+1))\ (1/(r_1\,m_1+1)))$$
$$= r_1\,m_1 / r_0\,m_0$$

This formula implies for example that a true OR of 2 for lung cancer incidence associated with statin exposure can be obscured and reduced to unity, or that a true OR of 1 can be artificially reduced to 0.5, by this biased approach if the mean time of follow-up for cases is half that of controls.

The correct analysis for such data must consider the time-dependent nature of the drug exposure. Thus, for a case-control design, the controls must be selected as a random sample either from all person-moments of the population, or from the risk set that re-computes the proportion of non-users up to the point in time that a case occurs for each case and control at that time.[1,2] Using this time-dependent approach, the probability of non-use of statins from cohort entry to the point at which the case occurs is $e^{-r\,t}$, where t represents the follow-up time at which the case occurred. This probability is thus identical for non-cases, since the value of r for the mean rate used to generate statin prescriptions is taken to be identical for the cases and non-cases. Consequently, the odds ratio at each time point that a lung cancer case occurs is one, so that the correct overall odds ratio from such an analysis would be as expected, namely unity.

eTable 2 provides data from the illustrative case-control study to support these formulae. First, the mean follow-up time is shorter for the cases (2.28 years) than for the time-independent controls (2.96 years), but not for the time-dependent controls (2.14 years). This suggests that in the absence of an association, the estimated odds ratio that would not account for time-dependency would be 2.28/2.96=0.77, a 23% bias. Second, the follow-up times are different between the unexposed cases and time-independent controls (2.18 versus 2.88 years) and the extent of unexposed person-time cumulated prior to the first statin prescription is differential between the exposed cases and time-independent exposed controls (0.97 versus 1.35 years). These differences are not seen with the time-dependent controls.

REFERENCES

(1) Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of Controls in Case-Control Studies: I. Principles. *Am J Epidemiol.* 1992;135:1019-1028.

(2) Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of Controls in Case-Control Studies: III. Design Options. *Am J Epidemiol.* 1992;135:1042-1050.

(3) 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;131:1282-8.

e**APPENDIX**

Time-Window Bias in Case-Control Studies: Statins and Lung Cancer

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

eTable 1

Odds ratios from the expected proportion of statin users among lung cancer cases and controls, assuming exponential distributions with mean m_1 for the time to lung cancer diagnosis for the cases, and mean m_0 for the observation time of controls. The number of statin prescriptions per subject is distributed as Poisson, conditional on the observation time, with mean rate r prescriptions per person-time equally for cases and controls

	Cases	**Controls**	**Odds ratio**
Time-independent control sampling			
Expected proportion of users (%)	$1-(1/(r\, m_1+1))$	$1-(1/(r\, m_0+1))$	m_1 / m_0
Expected proportion of non-users (%)	$1/(r\, m_1+1)$	$1/(r\, m_0+1)$	1 (Reference)
Time-dependent control sampling			
Expected probability of use at time[a] t	$1 - e^{-r\,t}$	$1 - e^{-r\,t}$	1
Expected probability of non-use at time t	$e^{-r\,t}$	$e^{-r\,t}$	1 (Reference)

[a] Time t represents the calendar time point at which a case occurred.

e**APPENDIX**

Time-Window Bias in Case-Control Studies: Statins and Lung Cancer

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

eTable 2

Comparison of observation and exposure times between the cases of lung cancer, the controls from time-independent approach used in the VA study,[3] and the controls from a time-dependent approach selection, using the cohort of 365,467 subjects identified from the GPRD.

	Cases	Controls	
		Time-independent sampling	Time-dependent sampling
All users			
Number	1,786	363,681	17,860
Mean time to index date (years)	2.28	2.96	2.14
Statin non-users			
Number	1,521	288,061	15,421
Mean time to index date (years)	2.18	2.88	2.07
Statin users	265	75,620	2,439
Mean time to index date (years)	2.82	3.28	2.58
Mean time to first statin prescription (unexposed)	0.97	1.35	0.98
Mean time from first statin prescription to index date	1.85	1.93	1.60