

## Comment on 'The latency period of mesothelioma among a cohort of British asbestos workers (1978–2005)': methodological problems with case-only survival analysis

D Consonni<sup>\*1</sup>, F Barone-Adesi<sup>2</sup> and C Mensi<sup>1</sup>

<sup>1</sup>Epidemiology Unit, Department of Preventive Medicine, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Via San Barnaba 8, 20122 Milan, Italy and <sup>2</sup>Division of Population Health Sciences and Education, St George's, University of London, Cranmer Terrace, London SW17 0RE, UK

Sir,

We read with great interest the article by Frost (2013) on mesothelioma latency among British asbestos workers. The author examined data from the GB Asbestos Survey, a large prospective cohort established in 1971 to monitor the long-term health of asbestos workers. On the basis of analyses of latency (defined as the time from first occupational exposure to asbestos to death from mesothelioma) among mesothelioma deaths in the period 1978–2005, the author concluded that the study did not find sufficient evidence of an inverse relationship between the intensity of asbestos exposure and the length of the latency period. However, there are methodological problems that undermine these results. Even if the whole cohort of 94 960 subjects was available, only mesothelioma deaths (614 subjects) were included in the analysis (Frost, 2013). It is well known that comparing survival times only among cases is a flawed approach (Pike and Doll, 1965; Langholz *et al*, 1999; Consonni, 2013). In a cohort study, only a proportion of subjects develops the outcome of interest. The remaining subjects will be still alive at the end of the study or die due to a different cause of death. For these individuals, the survival times for the outcome of interest will not be observable. This phenomenon is known as 'censoring' in survival analysis (Cleves *et al*, 2010). In this situation, it could seem attractive to use only the available survival times from the cases to calculate mean (or median) survival time in the different groups (e.g., subjects exposed and not exposed to a risk factor) and compare them. The problem with this approach is that survival times of cases are not representatives of survival times of the whole cohort. This can be showed with a simple example. Let's imagine to conduct a clinical study over 40 subjects: 20 treated with a new drug and 20 treated with placebo. Note that in this example 'survival' is precisely a 'latency' (time since first exposure to the drug). Let us assume that all the subjects are followed until death occurs and that survival times of the subjects in the experimental group were 3, 4, 6, 6, 8, 9, 10, 13, 17, 28, 39, 41, 47, 49, 65, 68, 80, 96, 123, 140 months, and those in the placebo group were 1, 3, 3, 5, 6, 7, 11, 12, 13, 14, 16, 18, 22, 23, 25, 31, 31, 38, 92, 114 months. Because the survival times of all the subjects were observed, it would be possible to calculate mean and median survival times in the two groups (means: 43 and 24 months; medians: 33 and 15 months, respectively). However, in most epidemiological studies it is not possible to follow the subjects until they all die. Let us assume now that the study ends for practical reasons 48 months after the allocation of the treatment. Survival times larger than 48 months will then become unobservable. The survival times in the two groups available to the researchers would now be 3, 4, 6, 6, 8, 9, 10, 13, 17, 28, 39, 41, 47 months in the experimental group and 1, 3, 3, 5, 6, 7, 11, 12, 13, 14, 16, 18, 22, 23, 25, 31, 31, 38 months in the placebo group. The mean and median survival time in the two groups would appear now to be 18 and 16 months (means) and 10 and 13 months (medians), respectively, very different from the situation in which all survival times are observed and suggesting only a slight increase of survival in the experimental group compared with the placebo group. Therefore, if we used the approach adopted by Frost (2013) in her paper and calculate mean (or median) survival using only the available survival times from cases, we would obtain a different, incorrect result. Note that this phenomenon is quite general and goes beyond the fictitious example above. It happens because censoring 'cuts' the right tail of the distributions of the survival times in the two groups, making look the distributions more similar than they really are. It is noteworthy that even using a regression model like the one adopted by Frost (2013) will not solve the problem. Conversely, including all the subjects in the analysis and taking into account censoring

through standard techniques for survival analysis would produce unbiased results (Cleves *et al*, 2010). Analyses on the basis of data only from cases are particularly prone to bias when the proportion of censored survival times is high. Unfortunately this is the case of the study by Frost (2013), where more than 99% of the survival times for mesothelioma were not directly observable. This casts serious doubts on the validity of the study.

The mistake of restricting the analysis on cases without considering the population from which cases originated was recognised almost 50 years ago by Pike and Doll (1965). Despite that paper and the fact that scientific papers (Langholz *et al*, 1999), statistical (Colton, 1974) and epidemiological books (Rothman, 2012; Weiss, 2012) explicitly discourage the analysis of survival times among cases only, this problem occasionally re-appears in the literature (Consonni, 2013). Notably, this problem affected also studies based on case series from mesothelioma registries (Yeung *et al*, 1999; Neumann *et al*, 2001; Marinaccio *et al*, 2007; Bianchi and Bianchi, 2009).

In summary, the results reported in the paper by Frost (2013) are not interpretable because they were on the basis of analyses among mesothelioma deaths only, which are known to be incorrect. The only meaningful way to address the issue of mesothelioma latency using survival techniques is to include all subjects, and to treat those who do not die from mesothelioma during the follow-up period as 'censored' (Cleves *et al*, 2010). This approach implies the use of information from all the 94 960 workers in the GB cohort.

After all, occupational epidemiologists do not restrict analyses to dead workers while studying carcinogens and oncologists do not restrict analyses to dead patients while analysing cancer survival.

### REFERENCES

- Bianchi C, Bianchi T (2009) Malignant pleural mesothelioma in Italy. *Indian J Occup Environ Med* **13**: 80–83.
- Cleves M, Gould W, Gutierrez RG, Marchenko YV (2010) *An Introduction to Survival Analysis Using Stata*. Stata Press: Texas, USA.
- Colton T (1974) *Statistics in Medicine Chapter 12: Fallacies in numerical reasoning*. Little, Brown and Company: Boston, USA.
- Consonni D (2013) Something is missing: what's wrong with using age at diagnosis/death or latency among cases. *Epidemiol Prev* **37**: 85–88; article in Italian.
- Frost G (2013) The latency period of mesothelioma among a cohort of British asbestos workers (1978–2005). *Br J Cancer* **109**: 1965–1973.
- Langholz B, Thomas D, Xiang A, Stram D (1999) Latency analysis in epidemiologic studies of occupational exposures: application to the Colorado Plateau uranium miners cohort. *Am J Ind Med* **35**: 246–256.
- Marinaccio A, Binazzi A, Cauzillo G, Cavone D, De Zotti R, Ferrante P, Gennaro V, Gorini G, Menegozzo M, Mensi C, Merler E, Mirabelli D, Montanaro F, Musti M, Pannelli F, Romanelli A, Scarselli A, Tumino R. Italian Mesothelioma Register (ReNaM) Working Group (2007) Analysis of latency time and its determinants in asbestos related malignant mesothelioma cases of the Italian register. *Eur J Cancer* **43**: 2722–2728.
- Neumann V, Günther S, Müller K-M, Fischer M (2001) Malignant mesothelioma - German mesothelioma register 1987-1999. *Int Arch Occup Environ Health* **74**: 383–395.
- Pike MC, Doll R (1965) Age at onset of lung cancer: significance in relation to effect of smoking. *Lancet* **1**(7387): 665–668.
- Rothman KJ (2012) *Epidemiology. An Introduction*. Oxford University Press: New York, USA.
- Weiss NS (2012) *Exercises in Epidemiology. Applying Principles and Methods*. Oxford University Press: New York, USA.
- Yeung P, Rogers A, Johnson A (1999) Distribution of mesothelioma cases in different occupational groups and industries in Australia, 1979-1995. *Appl Occup Environ Hyg* **14**: 759–767.

\*Correspondence: Dr D Consonni; E-mail: Dario.Consonni@unimi.it  
Published online 29 April 2014

© 2014 Cancer Research UK. All rights reserved 0007–0920/14



<http://creativecommons.org/licenses/by-nc-sa/3.0/>

