Should relative survival be used with lung cancer data?

SR Hinchliffe[*,1], MJ Rutherford[1], MJ Crowther[1], CP Nelson[1,2] and PC Lambert[1,3]

[1]Department of Health Sciences, Centre for Biostatistics and Genetic Epidemiology, 2nd Floor Adrian Building, University Road, University of Leicester, Leicester LE1 7RH, UK; [2]Department of Cardiovascular Sciences, Clinical Sciences Wing, Glenfield General Hospital, University of Leicester, Leicester LE3 9QP, UK; [3]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, PO Box 281, Stockholm SE-171 77, Sweden

BACKGROUND: Under certain assumptions, relative survival is a measure of net survival based on estimating the excess mortality in a study population when compared with the general population. Background mortality estimates are usually taken from national life tables that are broken down by age, sex and calendar year. A fundamental assumption of relative survival methods is that if a patient did not have the disease of interest then their probability of survival would be comparable to that of the general population. It is argued, as most lung cancer patients are smokers and therefore carry a higher risk of smoking-related mortalities, that they are not comparable to a population where the majority are likely to be non-smokers.

METHODS: We use data from the Finnish Cancer Registry to assess the impact that the non-comparability assumption has on the estimates of relative survival through the use of a sensitivity analysis.

RESULTS: Under realistic estimates of increased all-cause mortality for smokers compared with non-smokers, the bias in the estimates of relative survival caused by the non-comparability assumption is negligible.

CONCLUSION: Although the assumption of comparability underlying the relative survival method may not be reasonable, it does not have a concerning impact on the estimates of relative survival, as most lung cancer patients die within the first 2 years following diagnosis. This should serve to reassure critics of the use of relative survival when applied to lung cancer data.

British Journal of Cancer (2012) **106,** 1854–1859. doi:10.1038/bjc.2012.182   www.bjcancer.com
Published online 3 May 2012
© 2012 Cancer Research UK

Keywords: relative survival; expected survival; lung cancer; sensitivity analysis; smoking

Lung cancer is commonly known to be a disease that has strong associations with smoking (Doll and Hill, 1956; Korhonen *et al*, 2008; Papadopoulos *et al*, 2011). A report published by Peto *et al*, 2006 showed that, in Finland in the year 2000, 86% of lung cancer deaths in males and 60% of lung cancer deaths in females were deemed to be attributed to smoking. In addition to this, they showed that 12% of cardiovascular deaths in males and 3.6% of cardiovascular deaths in females were also deemed to be attributed to smoking. Figures were also reported for other types of cancer and other causes of death. Not only does smoking put you at a high risk of developing lung cancer and consequently dying from lung cancer (Doll and Hill, 1956; Papadopoulos *et al*, 2011), it also increases your chances of dying from many other diseases (Wolf *et al*, 1988), such as cardiovascular disease (Willett *et al*, 1987) and other less common forms of cancer (Moore, 1971; Fuchs *et al*, 1996).

This has led to heavy debate as to whether relative survival should be used as a method to analyse lung cancer data (Dickman and Adami, 2006; Sarfati *et al*, 2010). Relative survival is a method that compares the survival experience of a group of patients to the survival experience of the general population. The method is particularly advantageous, as it does not require an accurate cause-of-death information. Mortality estimates for the general population are usually taken from national life tables that are broken down by age, sex and calendar year. One of the key assumptions of relative survival is comparability – if the patient did not have the

cancer, then it is assumed that they would have the same survival experience as the general population. It is argued, as most lung cancer patients are smokers and therefore carry a higher risk of many other diseases, that they are not comparable to a population where the majority are likely to be non-smokers (Phillips *et al*, 2002). However, despite these potential problems, relative survival is still the usual method of analysis in population-based cancer studies.

This paper assesses the impact that the non-comparability has on the relative survival estimates through the use of a sensitivity analysis. Similar studies have been carried out previously to assess the impact that specific cancer deaths in the population mortality figures can have on the estimate of relative survival (Hinchliffe *et al*, 2011; Talbäck and Dickman, 2011).

METHODS

Relative survival

Relative survival is a measure that estimates the survival from a particular disease in the absence of other causes of death. It can be written as the ratio of the observed survival in the study population to the expected survival in the general population (Ederer *et al*, 1961). More formally:

$$R(t) = \frac{S(t)}{S^*(t)} \qquad (1)$$

where $S(t)$ is the observed survival, $S^*(t)$ is the expected survival and t is the time from diagnosis (Lambert *et al*, 2010). When

*Correspondence: Dr SR Hinchliffe; E-mail: srh20@leicester.ac.uk

relative survival analysis is applied to a cohort of lung cancer patients, we are making a comparison of survival in lung cancer patients relative to survival in the general population. Because of the higher prevalence of smoking amongst lung cancer patients, the expected survival is likely to be too high. We adjust the expected survival via a sensitivity analysis to assess the impact on estimates of 1- and 5-year relative survival.

Sensitivity analysis

In Finland, it is required that all physicians, hospitals and other relevant institutions send notification to the Finnish Cancer Registry of all cancer cases that come to their attention. The Registry, therefore, has full population coverage for all cancer cases going back to 1953. Lung cancer data (ICD-O-3: C340-C349) were obtained from the Finnish Cancer Registry for patients diagnosed in the years 1995–2007, inclusive. Population mortality data for Finland, broken down by age, sex and calendar year, were obtained from the Human Mortality Database (2008). Patients under the age of 18 and anyone diagnosed through autopsy were excluded from the analyses. All relative survival analyses were carried out by the age groups 18–44, 45–59, 60–74, 75–84 and 85 +. To obtain up-to-date estimates of relative survival, a period analysis approach was adopted. The relative survival estimates were derived from data on the survival experience of patients in the 2005-2007 period (Brenner and Gefeller, 1996).

An initial relative survival analysis was carried out using the unadjusted population mortality data. The population mortality data was then modified to represent the scenario where 100% of the general population are assumed to be smokers. This creates a group that is more comparable to the cohort of lung cancer patients in which the vast majority are also smokers. The adjustment was made by considering the following quantities: the odds ratio for increased/decreased odds of dying from any cause for smokers compared with non-smokers denoted as θ, the probability of dying from any cause if you are a smoker denoted as p_s, the probability of dying from any cause if you are a non-smoker denoted as p_n, the total probability of dying from any cause in the general population denoted as p_t, and the proportion of daily smokers in the general population denoted as α. The above quantities are connected through the following equation

$$p_t = (1 - \alpha)p_n + \alpha p_s \qquad (2)$$

We developed an adjustment for p_n, which included all the terms described above. The formulae for this are given in the Appendix. It should be noted that p_t, p_n and p_s are yearly probabilities that will vary by age, sex and calendar year.

As we do not have information on the exact number of smokers in the population-mortality data file, it was assumed that the prevalence of smokers, α, was as shown in Table 1. These estimates were taken from a report of the 'Health in Finland' (Koskinen et al, 2006). The total probabilities of dying from any cause, p_t, were taken from the population-mortality data file. The odds ratio, θ, was set to 2, 3, 4 and 5 to demonstrate both plausible and extreme scenarios for the increased risk in overall mortality from smoking. This information was used to determine the probability of dying from any cause if you are a non-smoker, p_n, using the equations given in the Appendix. This value was subsequently used to estimate the probability of dying from any cause if you are a smoker, p_s.

Comparisons were made between the relative survival estimates derived using the total probability of dying, p_t, from the original unadjusted population mortality file and the relative survival estimates derived using the adjusted probabilities of dying from all causes for smokers, p_s.

A systematic review by Schane et al, 2010 reported an odds ratio of 1.6 (95% CI: 1.3 to 2.1) for the risk of all-cause mortality of light and intermittent male smokers compared with male non-smokers.

Table 1 Smoking prevalence in adults by gender (%; Koskinen et al, 2006)

Gender	Year	Percentage
Males	1975–1980	35
	1981–1985	33
	1986–1990	33
	1991–1995	30
	1996–2000	27
	2001–2008	26
Females	1975–1980	17
	1981–1985	16
	1986–1990	19
	1991–1995	19
	1996–2000	21
	2001–2008	18

To visualise the bias in the relative survival estimates when adjusting for a more realistic odds ratio, this odds ratio of 1.6 was taken as the 'estimated' value for θ for both genders and all age groups. This was done in addition to the adjustments made with odds ratios of 2, 3, 4 and 5.

RESULTS

Relative survival curves using odds ratios (θ) of 2, 3, 4 and 5 for increased odds of all-cause mortality for smokers compared with non-smokers are shown in Figures 1–4, respectively. Each figure compares the relative survival curve obtained using the unadjusted population mortality files to the relative survival curve that has been adjusted assuming that everyone in both the lung cancer cohort and population mortality file is a smoker. All four figures show that adjusting for a higher probability of death in smokers makes little, if any, difference in the 18–44 and 45–59 age groups, as the probability of death from other causes is low in these ages. There is also very little difference between the curves in the older three age groups until the odds ratio reaches 4 and 5, where the largest differences in the relative survival estimates are between 0.05 and 0.1.

Table 2 gives the percentage unit differences between the unadjusted 1-year and 5-year relative survival estimates and the 1-year and 5-year relative survival estimates adjusted using odds ratios of $\theta = 2$, 3, 4 and 5. It also includes a column showing the percentage unit differences when adjusting for the 'estimated' θ. The results show that by using unadjusted life tables, the relative survival estimates are slightly underestimated when compared with life tables that are adjusted using odds ratios of 2, 3, 4 and 5.

DISCUSSION

Although the assumption of comparability between the patient cohort and general population may be unreasonable for lung cancer, we have shown that correcting for this does not have a concerning impact on the relative survival estimates. In the younger age groups, the probability of dying from other causes is low; therefore, even a fairly large relative adjustment to this value will not have a large impact. It follows that the adjustment will therefore have little effect on the relative survival estimates.

Furthermore, for all age groups, the prognosis for lung cancer is poor, with the majority of patients dying within the first 2 years. If the majority of lung cancer patients are dying quickly from lung-cancer-related deaths, then the fact that these patients are also at an increased risk of death from other diseases will have little impact on the relative survival estimates. Patients do not have the
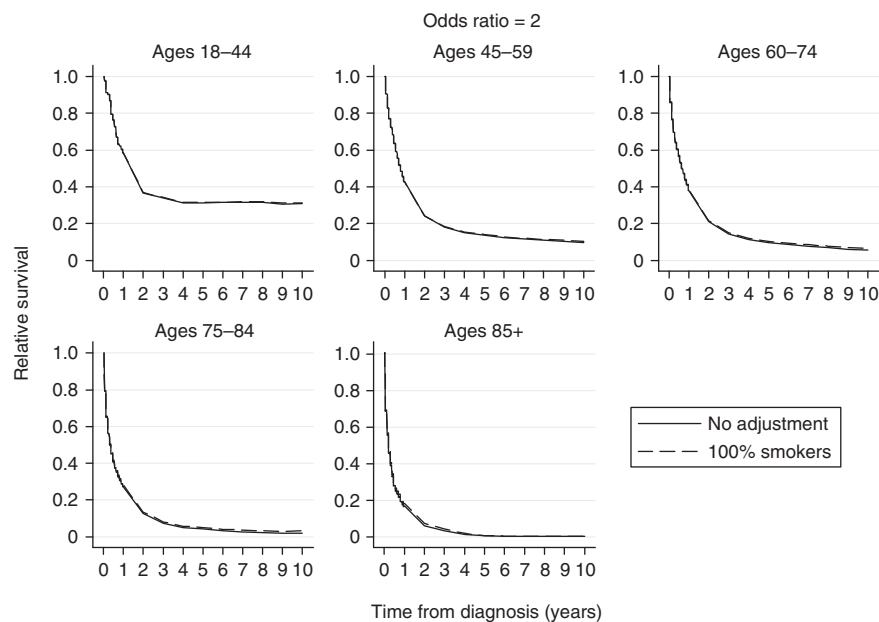
Epidemiology

Figure 1 Comparison of relative survival curves with no adjustment made to the external population with relative survival curves, assuming external population consists of 100% smokers and that the odds of all-cause mortality is twice as high for smokers as compared with non-smokers.
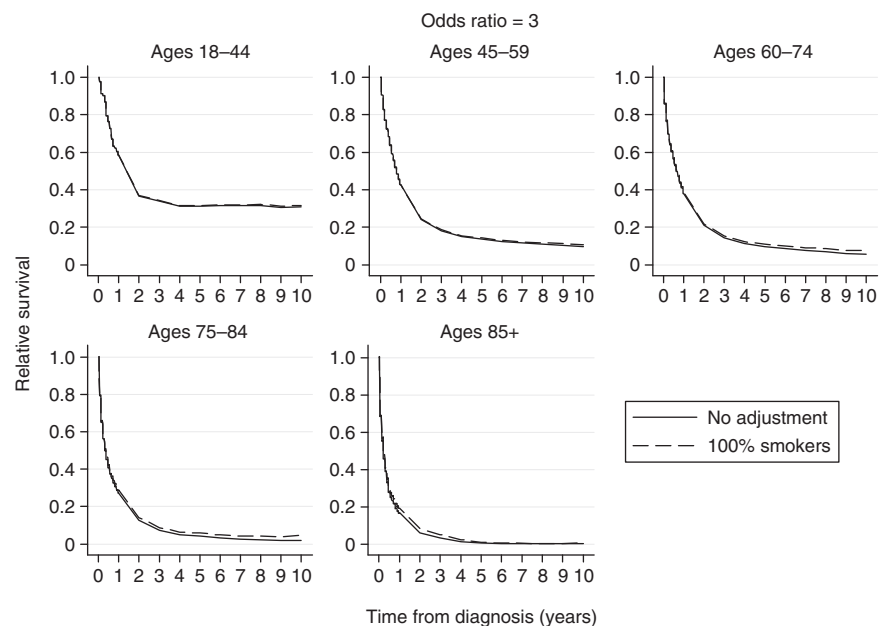

Figure 2 Comparison of relative survival curves with no adjustment made to the external population with relative survival curves, assuming external population consists of 100% smokers and that the odds of all-cause mortality is three times as high for smokers compared with non-smokers.

'opportunity' to die from other causes, because of the lethality associated with a diagnosis from lung cancer.

The performed sensitivity analysis made adjustments to the population mortality data to represent a scenario where 100% of the comparison population were smokers. This was done in an attempt to create a more comparable group to the lung cancer patient population. The true smoking figures amongst the lung cancer patient population will most likely not be 100%. Therefore, our adjustment was an extreme case. However, we have shown that the bias is relatively small regardless, and a more realistic proportion will only decrease this bias.

Although we have only considered lung cancer in this paper, we acknowledge that there are other cancer sites, such as bladder

cancer, and cancer of the oral cavity and pharynx, that have also been shown to be smoking-related. To carry out a similar sensitivity analysis for these cancer sites, an estimate of the prevalence of smoking within each cohort of cancer patients would be required. It would be unreasonable to assume that the proportion of smokers is anywhere near 100% in bladder and oral cancer cohorts. As these cancers have a better survival than lung cancer, it is likely that the lack of comparability of the life tables may have a larger impact on the relative survival estimates for these sites.

Unfortunately, information was not available on smoking status within the population mortality file. As a result, external information was used to obtain appropriate estimates for this
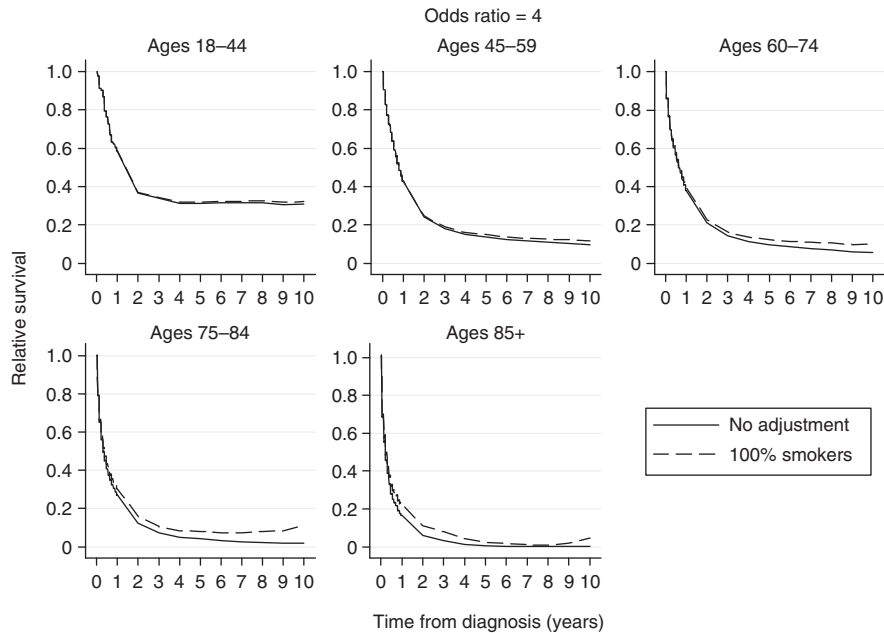
Epidemiology

Odds ratio = 4

Figure 3 Comparison of relative survival curves with no adjustment made to the external population with relative survival curves, assuming external population consists of 100% smokers and that the odds of all-cause mortality is four times as high for smokers compared to non-smokers.
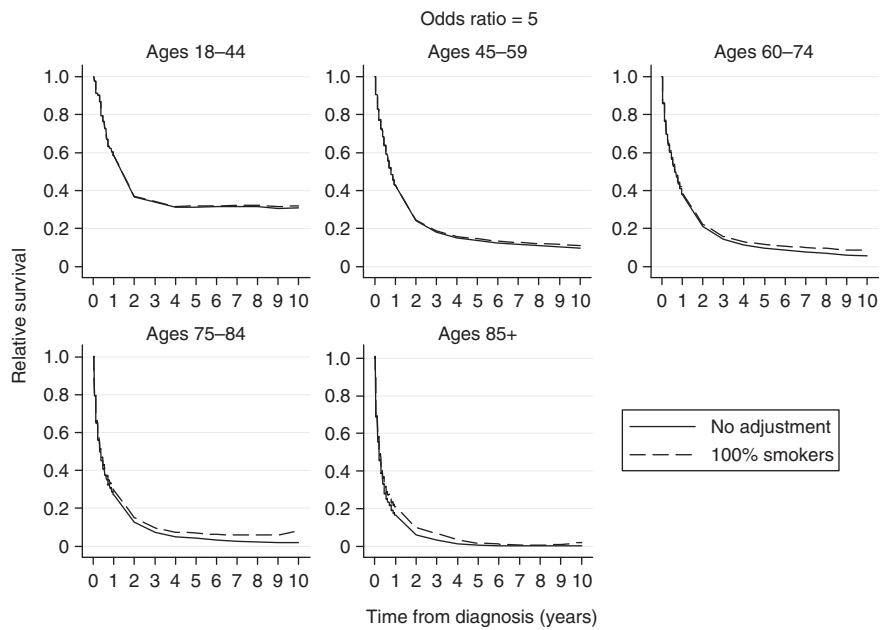
Odds ratio = 5

Figure 4 Comparison of relative survival curves with no adjustment made to the external population with relative survival curves assuming external population consists of 100% smokers and that the odds of all-cause mortality is five times as high for smokers compared with non-smokers.

(Table 1; Koskinen *et al*, 2006). These estimates were not stratified by age group. Should the proportion of smokers be larger in any of the age groups, then the bias in the relative survival estimates would most likely increase. This is particularly true for the oldest age group.

If smoking status had been available, then it would be preferable to create separate life tables for smokers and non-smokers. However, difficulty lies in making a strict definition of a 'smoker'. People's smoking status varies over time, as does the level of cigarette consumption. Both of these factors are likely to have an impact on the general health status and prognosis from lung cancer, and so, would also ideally be incorporated into the life table.

We have focussed on the potential bias in the relative survival estimates, as this is the measure most commonly reported. However, if there was interest in comparing groups in terms of the excess mortality, then there may also be bias in the excess mortality-rate ratio. Had smoking status been available, then a comparison could have been made using both smoking-adjusted and -unadjusted life tables. Using the general population life tables, we would expect that the excess mortality-rate ratio for smoking status would be downwardly biased, as the excess mortality rate for smokers would be underestimated and the mortality rate for non-smokers would be overestimated.

The value of θ that was chosen as the 'estimated' odds ratio was taken from a systematic review that was carried out to identify

Epidemiology

Table 2 Percentage unit difference in 1-year and 5-year relative survival estimates between values with no adjustment and 2, 3, 4, 5, and 'estimated' (1.6) adjustments

| Age (years) | Odds ratio (θ) | | | | | | | | | |
| | 2 | | 3 | | 4 | | 5 | | 'Estimated' | |
	1 year	5 years	1 year	5 years	1 year	5 years	1 year	5 years	1 year	5 years
18–44	0.06	0.20	0.10	0.30	0.20	0.60	0.15	0.40	0.0004	0.10
45–49	0.17	0.30	0.29	0.50	0.59	1.10	0.44	0.80	0.11	0.20
60–74	0.42	0.70	0.70	1.10	1.45	2.40	1.07	1.80	0.27	0.40
75–84	0.77	0.70	1.32	1.30	2.72	3.20	2.06	2.30	0.50	0.50
85 +	0.84	0.10	1.48	0.30	3.12	1.00	2.20	0.60	0.54	0.08

studies on the health outcomes associated with light and intermittent smoking. The value of 1.6 was calculated using data on males only, but we used this value to represent all ages and both genders in our sensitivity analysis. Although this value may be overestimated or underestimated for some subgroups of patients, given that even with an odds ratio of 5, the difference between the curves is still reasonably small, we can conclude that in practice, we don't have to be too concerned about the level of bias that may be introduced into the relative survival estimates by the assumption addressed in this paper.

The method described in this paper only makes adjustments for the assumption of comparability between the observed and expected populations. Other assumptions, such as independence between the mortality associated with the disease of interest and the mortality associated with other causes, are presumed to be reasonable.

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

Brenner H, Gefeller O (1996) An alternative approach to monitoring cancer patient survival. *Cancer* **78**(9): 2004–2010

Dickman PW, Adami H-O (2006) Interpreting trends in cancer patient survival. *J Intern Med* **260**(2): 103–117

Doll R, Hill AB (1956) Lung cancer and other causes of death in relation to smoking: a second report on the mortality of British doctors. *Br Med J* **2**: 1071–1081

Ederer F, Axtell L, Cutler S (1961) The relative survival rate: a statistical methodology. *Natl Cancer Inst Monogr* **6**: 101–121

Fuchs CS, Colditz GA, Stampfer MJ, Giovannucci EL, Hunter DJ, Rimm EB, Willett WC, Speizer FE (1996) A prospective study of cigarette smoking and the risk of pancreatic cancer. *Arch Intern Med* **156**(19): 2255–2260

Hinchliffe SR, Dickman PW, Lambert PC (2011) Adjusting for the proportion of cancer deaths in the general population when using relative survival: a sensitivity analysis. *Cancer Epidemiol* **36**: 148–152

Human Mortality Database (2008) University of California and Rostock: Max Planck Institute for Demographic Research: Berkley

Korhonen T, Broms U, Levälahti E, Koskenvuo M, Kaprio J (2008) Characteristics and health consequences of intermittent smoking: Long-term follow-up among Finnish adult twins. *Nicotine Tob Res* **11**(2): 148–155

Koskinen S, Aromaa A, Huttunen J, Teperi J (eds) (2006) *Health in Finland*. KTL, Stakes and Ministry of Social Affairs

Lambert PC, Dickman PW, Nelson CP, Royston P (2010) Estimating the crude probability of death due to cancer and other causes using relative survival models. *Stat Med* **29**(7-8): 885–895

Moore C (1971) Cigarette smoking and cancer of the mouth, pharynx, and larynx. *J Am Med Assoc* **218**(4): 553–558

Papadopoulos A, Guida F, Cénée S, Cyr D, Schmaus A, Radoï L, Paget-Bailly S, Carton M, Tarnaud C, Menvielle G, Delafosse P, Molinié F, Luce D, Stücker I (2011) Cigarette smoking and lung cancer in women: Results of the French ICARE case-control study. *Lung Cancer* **74**: 369–377

Peto R, Lopez A, Boreham J, Thun M, Heath JC (2006) *Mortality From Smoking in Developed Countries 1950–2000*, 2nd edn. Oxford University Press: Oxford

Phillips N, Coldman A, McBride ML (2002) Estimating cancer prevalence using mixture models for cancer survival. *Stat Med* **21**(9): 1257–1270

Sarfati D, Blakely T, Pearce N (2010) Measuring cancer survival in populations: relative survival vs cancer-specific survival. *Int J Epidemiol* **39**(2): 598–610

Schane RE, Ling PM, Glantz SA (2010) Health effects of light and intermittent smoking: a review. *Circulation* **121**(13): 1518–1522

Talbäck M, Dickman PW (2011) Estimating expected survival probabilities for relative survival analysis. Exploring the impact of including cancer patient mortality in the calculations. *Eur J Cancer* **47**(17): 2626–2632

Willett WC, Green A, Stampfer MJ, Speizer FE, Colditz GA, Rosner B, Monson RR, Stason W, Hennekens CH (1987) Relative and absolute excess risks of coronary heart disease among women who smoke cigarettes. *N Engl J Med* **317**(21): 1303–1309

Wolf PA, D'Agostino RB, Kannel WB, Bonita R, Belanger AJ (1988) Cigarette smoking as a risk factor for stroke. *J Am Med Assoc* **259**(7): 1025–1029

APPENDIX

To carry out the sensitivity analysis, we need to partition the total probability of dying from any cause in the general population into the probabilities for smokers and non-smokers separately.

If we consider the odds ratio, θ, which compares the odds of dying from any cause if you are a smoker to the odds of dying from any cause if you are a non-smoker. By re-arranging the formulae for an odds ratio, we can write in terms of the probability of dying from any cause if you are a smoker (p_s):

$$p_s = \frac{\left(\frac{p_n}{1-p_n}\right)\theta}{\left(\frac{p_n}{1-p_n}\right)\theta + 1} \qquad (3)$$

We now have the probability of dying from any cause if you are a smoker (p_s), as a function of both the odds ratio, θ, and the probability of dying from any cause if you are a non-smoker (p_n).

We also know that the total probability of dying from any cause (p_t) can be written as a function of p_s and p_n, if we can quantify the proportion of smokers in the general population (α):

$$p_t = (1-\alpha)p_n + \alpha p_s \qquad (4)$$

By substituting equation (3) into equation (4), we can write the total probability of dying from any cause, p_t, in terms of the odds ratio, θ, the proportion of smokers in the general population, α, and the probability of dying from any cause if you are a

non-smoker, p_n, as follows:

$$p_t = (1-\alpha)p_n \frac{\alpha\theta p_n}{(1-p_n)\left(\frac{\theta p_n}{1-p_n}+1\right)} \qquad (5)$$

We can re-arrange equation (5) as follows:

$$p_t = p_n - \alpha p_n + \frac{\alpha\theta p_n}{\theta p_n + 1 - p_n} \qquad (6)$$

$$p_t = \frac{\theta p_n^2 - \alpha\theta p_n^2 + p_n - \alpha p_n - p_n^2 + \alpha p_n^2 + \alpha\theta p_n}{\theta p_n + 1 - p_n} \qquad (7)$$

$$\theta p_n^2 - \alpha\theta p_n^2 + p_n - \alpha p_n - p_n^2 + \alpha p_n^2 + \alpha\theta p_n - \theta p_t p_n - p_t + p_t p_n = 0 \qquad (8)$$

$$p_n^2((1-\theta)(\alpha-1)) + p_n(1+(p_t-\alpha)(1-\theta)) - p_t = 0 \qquad (9)$$

The equation is now in the format with which the quadratic formula can be used to solve equation (9):

$$p_n = \frac{-(1+(p_t-\alpha)(1-\theta)) + \sqrt{(1+(p_t-\alpha)(1-\theta))^2 + 4p_t((1-\theta)(\alpha-1))}}{2((1-\theta)(\alpha-1))}$$

$$(10)$$

Now that we can calculate the probability of dying from any cause if you are a non-smoker, p_n, using equation (3), we can also calculate the probability of dying from any cause if you are a smoker (p_s).

The population mortality file can now be adjusted, so that rather than using the total probability of dying from any cause (p_t) as we would have done previously, we now use the probability of dying from any cause if you are a smoker (p_s). This now assumes that 100% of the population are smokers.

Epidemiology