ORIGINAL ARTICLE

Effect of different approaches to treatment of smoking as a potential confounder in a case-control study on occupational exposures

L Richiardi, F Forastiere, P Boffetta, L Simonato, F Merletti

Occup Environ Med 2005;62:101–104. doi: 10.1136/oem.2004.015941

See end of article for authors' affiliations

Correspondence to:
Dr Lorenzo Richiardi, Unit
of Cancer Epidemiology,
University of Turin, V
Santena 7, 10126 Torino,
Italy; lorenzo.richiardi@
unito.it

Accepted 17 August 2004

Aim: To evaluate the effect of different approaches to treatment of smoking as a potential confounder in an occupational study of lung cancer.

Methods: Data were used from a case-control study on 956 men with lung cancer and 1253 population controls recruited in two northern Italian areas during 1990–1992. The risk of lung cancer associated with 11 selected job titles and eight selected industrial activities was estimated using seven different methods to treat smoking history. To evaluate the confounding effect of smoking, odds ratios obtained using the first six models were compared with estimates from the seventh and most complex model, in which cumulative tobacco consumption and time since cessation were considered.

Results: Although crude odds ratios for some of the occupational categories were biased by up to 25%, such bias decreased to less than 10% when a simple model including smoking status (never, ex-, current) was used.

Conclusions: In occupational studies on lung cancer risk, information on smoking status may allow satisfactory control of the potential confounding effect of the habit.

The duration and intensity of smoking, age at initiation, time since cessation, and type of cigarette and of tobacco smoked all determine the health effects of smoking. Simultaneous consideration of several indicators of smoking, however, may lead to methodological difficulties, as there are complex combinations between different smoking related variables.¹ A recent study by Leffondré *et al* compared several approaches to the treatment of smoking related variables in estimating lung cancer risk and suggested a method that includes information on smoking status (ever/never), pack years (expressed as the product of intensity and duration), and time since cessation.²

Epidemiological studies aiming at investigating the effect of other exposures on diseases strongly linked to smoking require a proper smoking adjustment. However, it has been shown that the lack of adjustment for smoking introduces limited bias in occupational studies on lung cancer.^{3–7} This finding has relevant implications, as information on smoking habits is not available in several occupational studies, mainly those based on employment registries. The present study contributes to the topic by evaluating the effect of different methods of treating smoking as a potential confounder in a case-control study on occupational exposures and lung cancer risk.

METHODS

We used data from a population based case-control study on lung cancer that we conducted in two northern Italian areas (the city of Turin and the eastern part of the Veneto region) during 1990 to 1992, within the framework of a larger European study.⁸ The Italian study included 1171 cases with a primary diagnosis of lung cancer and 1569 population controls. Information on smoking habits, residential history, exposure to environmental tobacco smoke, and occupational history was obtained in a face to face interview. Study design and results on occupation for the Italian study have been described previously.⁹ The study was approved by the local ethics committee.

After exclusion of subjects with a poor recall of their smoking history, 1132 cases (956 men and 176 women) and 1553 controls (1253 men and 300 women) were retained. Analyses in the present study are limited to men.

For each subject, information was available on cumulative tobacco consumption, expressed as lifetime pack years smoked, age at initiation, and age at cessation. Ex-smokers were defined as smokers who had stopped smoking two years or more before the interview, whereas never smokers included all subjects who had smoked fewer than 400 cigarettes. Job titles and branches of industry for each occupational period that lasted at least six months were recorded and coded according to the international standard classification of occupations (ISCO)¹⁰ and the international standard industrial classification (ISIC),¹¹ respectively.

We used unconditional logistic regression to estimate the odds ratio (OR) of lung cancer, and the corresponding 95% confidence intervals (CI), for each ISCO and ISIC code with at least 50 exposed cases. Subjects were considered exposed to a specific ISCO or ISIC category if they were ever employed in the corresponding job or industry. Categories with fewer than 50 exposed cases were not considered, as most complex models to treat smoking related variables include several variables and therefore produce unstable estimates in the presence of sparse data.

We started with the simplest model without smoking related variables and progressed through five models to the one suggested by Leffondré *et al*.² In addition, we modelled cumulative tobacco consumption using cubic b-spline regression with knots equally spaced at 10, 20, 30, and 40 pack years.¹² The seven models are described in table 1. All models also included age (in five year groups) and study area. For the present exercise we used results for the 11 job titles and the eight branches of industry associated with an odds ratio of at

Abbreviations: AIC, Akaike's information criterion; ISCO, international standard classification of occupations; ISIC, international standard industrial classification

Main message

- Under common circumstances, a simple model with information on the status of current smoker, ex-smoker, and never smoker at the time of interview in a case-control study allows satisfactory control of the confounding effect of smoking in the association between occupational factors and lung cancer risk.

least 1.40 or less than 0.70 in at least one of the seven models. Akaike's information criterion (AIC: $-2 [\log\text{-likelihood}] + 2 [\text{number of model parameters}]$) was used to compare the fit of the different models.¹³ As the measure of residual confounding, we calculated the ratio between, on the one hand, odds ratios for lung cancer risk obtained in models 1 to 6, and on the other hand, odds ratios obtained when model 7 was used.

We also evaluated the confounding effect of smoking on the association between duration of employment and risk of lung cancer. For this purpose, we estimated among controls the Spearman correlation coefficient between number of pack years ever smoked and the logarithm of the duration of employment, after adjusting for age, study area, and a parameter for ex-smokers. The very few subjects with missing data for duration of employment were excluded from the latter analysis.

RESULTS

As shown by decreasing AIC values from model 1 to model 7 (last column, table 1), the introduction of additional smoking related variables improved the fit of the model, indicating an increasing ability to accommodate the main effect of smoking on lung cancer risk. Similar results were obtained when each of the occupational variables was included in the models.

Figure 1 presents the residual confounding effect of smoking in the association between each ISCO and ISIC category and lung cancer risk estimated with models 1 to 7. Corresponding odds ratios and 95% confidence intervals obtained, respectively, in models 1, 3, and 7 are reported in table 2. As shown in fig 1, smoking had a moderate confounding effect on risk estimates, with a ratio between the odds ratios estimated in model 1 and those obtained from model 7 never higher than 1.25 or lower than 0.80. When smoking was introduced as a binary variable (ever/never smoker, model 2) part of the confounding effect was controlled and most of it was accommodated when model 3 was used. Further increases in the complexity of the models had a marginal improvement in the ability to accommodate confounding caused by smoking.

Among controls, the adjusted Spearman correlation coefficient (r) between pack years smoked and duration of employment never exceeded 0.20, with the exception of "salesmen, shop assistants, and related workers" ($r = 0.23$, $p = 0.05$, based on 87 subjects; ISCO code: 45), "street vendors, canvassers, and newsvendors" ($r = 0.47$, $p = 0.01$, based on 38 subjects; ISCO code: 452), and "manufacture of fabricated metal product—excluding machinery and equipment not elsewhere classified" ($r = 0.30$, $p = 0.07$, based on 50 subjects; ISIC code: 3819). For the latter three categories, the introduction of pack years in the model (models 4 to 7) had a marginal effect on the odds ratio estimates for duration of employment, which was introduced both as a continuous variable and as the logarithm of the continuous variable (data not shown).

DISCUSSION

We investigated lung cancer risk in association with job titles and industrial activities to evaluate the effect of different

Policy implications

- These results are relevant to future occupational studies, mainly those based on employment registries, where detailed information on smoking habits is not available.

methods of adjusting for the confounding effect of smoking on the risk of occupational exposures. We found limited changes in the risk estimates between a model including only smoking status, and more complex models in which intensity, duration, and time since cessation are considered. This is a consequence of the fact that the association between detailed aspects of the smoking habit (for example, time since cessation among ex-smokers) and being employed in at-risk occupations is weak. The effect of several methods to treat smoking as a confounder in the study of occupational factors has been evaluated previously in one case-control study of lung cancer among men,⁵ and in one case-control study of bladder cancer among women,¹⁴ using models that were partly different from those we used here. The conclusions of these previous studies are consistent with ours.

In our study the apparent overall confounding effect of smoking was modest, which is somewhat of a limitation of the exercise, although this finding is consistent with various previous studies showing that overall confounding by smoking in occupational studies on lung and other smoking related cancers is not likely to be more than about 20–25%.^{3–7 14 15} We cannot, however, rule out the presence of

Figure 1 Effect of seven different methods to model smoking history on odds ratio (OR) estimates for selected ISCO⁵ and ISIC⁶ categories (see table 1 for description of the models). The ratio between odds ratios obtained from each model and corresponding estimates from model 7 is reported on the y axis as a measure of residual confounding (residual confounding between 1.10 and 0.90 is considered negligible). Two Italian areas, 1990–1992. See table 2 for description of job titles and branches of industry corresponding to each ISCO and ISIC category reported in the figure.

Table 1 Description of the seven different methods to model smoking histories (two Italian areas, 1990–1992)

Model	Smoking habit	Pack years	Parameters*	AIC
1	–	–	0	3001.32
2	Ever, never	–	1	2780.61
3	Current, ex, never	–	2	2685.70
4	Current 20+ pyrs, 10–19 pyrs, 1–9 pyrs, ex, never	–	4	2628.51
5	Current, ex, never	Continuous	3	2615.99
6	Current, ex since 2–5 years, 6–10 years, 11–15 years, 16+ years, never	Continuous	6	2602.88
7	Current, ex since 2–5 years, 6–10 years, 11–15 years, 16+ years, never	Cubic b-splines. Knots at 10, 20, 30, 40 pyrs	13	2587.08

*All models also include age (five parameters) and study area (five parameters). AIC, Akaike's information criterion; pyrs, pack years.

residual confounding caused by non-differential misclassification of smoking status, as we did not validate our capacity to obtain reliable information on past smoking history through interview. There are reasons to believe that such misclassification was small. First, smoking was strongly associated with lung cancer in our study population,⁸ whereas misclassification would have most likely biased the estimates for the main effects of smoking downward; second, almost all interviews were done with the index cases and controls, with no surrogate information; and third, studies with re-interview of a sample of study subjects conducted in northern Italy showed good reliability of past smoking history.^{16, 17}

There is growing consensus that confounding assessment should be based on an understanding of the causal pathway linking the study variables.¹⁸ One reason for this approach is that the control of confounding may actually introduce bias, depending on the relation between the study variables, so that adjusted estimates may, under some circumstances, be more biased than crude ones. For instance, a factor that is caused by both the exposure and the disease should not be included in the models.¹⁹ In our example, however, smoking

was also a true confounder according to the path analysis theory.

It is likely that our results can be generalised to other places and periods in time with similar smoking and working conditions. Nevertheless, the prevalence of smokers in most populations changes over time,²⁰ new smoking related regulations may be introduced in the work places,²⁰ and older persons are replaced in the work force by younger persons, who may have different tobacco consumption, as smoking prevalence in each country is strongly related to the cohort of birth. For instance, in Italy, the proportion of male smokers peaked among the generation born in the 1920 to 1929.²¹ It might thus occur that some occupational groups have particular smoking habits that cannot be summarised by simple models. It should, however, be noted that recent developments in occupational epidemiology rely on the use of semiquantitative or quantitative measures of exposure to specific agents rather than on the use of job/industry information. In such studies, the confounding effect introduced by smoking is likely to be even smaller, as exposure to specific agents is less likely to be associated with a particular smoking habit.

Table 2 Odds ratios and 95% confidence intervals for selected job titles and branches of industry obtained using three different methods to model smoking (models 1, 3, and 7 as described in table 1; two Italian areas, 1990–1992)

Item and code	Number of cases	Model* (OR, 95% CI)		
		1	3	7
Job title (ISCO code)				
Professional, technical, and related workers (0, 1)	91	0.5 (0.4 to 0.7)	0.6 (0.4 to 0.8)	0.6 (0.5 to 0.9)
Mail distribution clerks (37)	58	1.5 (1.0 to 2.2)	1.7 (1.1 to 2.5)	1.7 (1.1 to 2.7)
Salesmen, shop assistants, and related workers (45)	111	1.5 (1.1 to 2.0)	1.3 (1.0 to 1.8)	1.3 (1.0 to 1.8)
Metal processors (72)	93	1.4 (1.0 to 1.9)	1.4 (1.0 to 1.9)	1.4 (1.0 to 2.0)
Plumbers, welders, sheet metal, and structural metal preparers and erectors (87)	99	1.4 (1.1 to 1.9)	1.4 (1.0 to 1.9)	1.5 (1.1 to 2.0)
Blue collar workers (7, 8, 9)	757	1.8 (1.5 to 2.2)	1.6 (1.2 to 2.0)	1.6 (1.3 to 2.0)
Labourers (99)	201	1.5 (1.2 to 1.9)	1.4 (1.1 to 1.7)	1.3 (1.1 to 1.7)
Salesmen, shop assistant, and demonstrators (451)	65	1.5 (1.0 to 2.1)	1.2 (0.8 to 1.8)	1.2 (0.8 to 1.8)
Street vendors, canvassers, and news vendors (452)	50	1.5 (1.0 to 2.3)	1.5 (0.9 to 2.3)	1.5 (0.9 to 2.3)
Dockers and freight handlers (971)	91	1.5 (1.1 to 2.1)	1.4 (1.0 to 1.9)	1.3 (0.9 to 1.9)
Retail trade salesmen (4513)	54	1.6 (1.0 to 2.4)	1.3 (0.9 to 2.0)	1.3 (0.8 to 2.0)
Branches of industry (ISIC code)				
Financing, insurance, real estate, and business services (8)	58	0.7 (0.5 to 1.0)	0.7 (0.5 to 1.1)	0.7 (0.5 to 1.1)
Manufacture of non-metallic mineral products, excluding products of petroleum and coal (36)	58	1.5 (1.0 to 2.1)	1.3 (0.9 to 2.0)	1.3 (0.9 to 2.0)
Transport and storage (71)	148	0.8 (0.7 to 1.1)	0.8 (0.6 to 1.0)	0.7 (0.6 to 0.9)
Social and related community services (93)	54	0.5 (0.4 to 0.8)	0.6 (0.4 to 0.9)	0.6 (0.4 to 0.9)
Manufacture of fabricated metal products, excluding machinery and equipment (381)	138	1.7 (1.3 to 2.2)	1.6 (1.2 to 2.1)	1.6 (1.2 to 2.1)
Land transport (711)	100	0.8 (0.6 to 1.0)	0.7 (0.6 to 1.0)	0.7 (0.5 to 0.9)
Manufacture of fabricated metal products, excluding machinery and equipment , NEC (3819)	93	2.1 (1.5 to 2.9)	2.0 (1.4 to 2.8)	2.0 (1.4 to 2.9)
Freight transport by road (7114)	70	0.8 (0.6 to 1.1)	0.7 (0.5 to 1.0)	0.7 (0.5 to 0.9)

*See table 1 for description of the models.

CI, confidence interval; NEC, not specified elsewhere; ISCO, international standard classification of occupations¹⁰; ISIC, international standard industrial classification¹¹; OR, odds ratio.

In conclusion, our study suggests that under common circumstances information on the status of current, ex-, and never smoker at the time of interview in a case-control study allows satisfactory control of confounding effects caused by smoking in the association between occupational exposures and lung cancer risk. All possible models to control for smoking should be always evaluated when detailed information is available. However, epidemiological studies are often carried out using pre-existing data sources that may have limited information on possible confounders. Our study suggests that occupational studies with information on smoking status alone may still produce valid results.

ACKNOWLEDGEMENTS

This study was partially supported by the Italian Association for Cancer Research, the Italian Ministry of University and Scientific Technological Research (MURST), the Regione Piemonte-Ricerca Finalizzata, the National Research Council (Contract 91.00327.CT04), special project "Oncology", Compagnia di San Paolo/FIRMS. We thank S Massacesi, M Tedeschi, and M Artom for their contributions to the study.

Authors' affiliations

L Richiardi, F Merletti, Unit of Cancer Epidemiology, CeRMS and Centre for Oncological Prevention, University of Turin, Italy

F Forastiere, Department of Epidemiology, Rome E Health Authority, Rome, Italy

P Boffetta, International Agency for Research on Cancer, Lyon, France

L Simonato, Department of Environmental Medicine and Public Health, University of Padua, Italy

Competing interests: None declared

REFERENCES

- McKnight B, Cook LS, Weiss NS. Logistic regression analysis for more than one characteristic of exposure. *Am J Epidemiol* 1999;**149**:984-92.
- Leffondrè K, Abrahamowicz M, Siemiatycki J, et al. Modelling smoking history: a comparison of different approaches. *Am J Epidemiol* 2002;**156**:813-23.
- Axelsson O. Aspects of confounding in occupation health epidemiology. *Scand J Work Environ health* 1978;**4**:85-9.
- Blair A, Hoar SK, Walrath J. Comparison of crude and smoking-adjusted standardized mortality ratios. *J Occup Med* 1985;**27**:881-4.
- Siemiatycki J, Wacholder S, Dewar R, et al. Degree of confounding bias related to smoking, ethnic group, and socioeconomic status in estimates of the associations between occupation and cancer. *J Occup Med* 1988;**30**:617-25.
- Flanders WD, Khoury MJ. Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. *Epidemiology* 1990;**1**:239-46.
- Haldorsen T, Andersen A, Boffetta P. Smoking-adjusted incidence of lung cancer by occupation among Norwegian men. *Cancer Causes Control* 2004;**15**:139-47.
- Simonato L, Agudo A, Ahrens W, et al. Lung cancer and cigarette smoking in Europe: an update of risk estimates and an assessment of inter-country heterogeneity. *Int J Cancer* 2001;**91**:876-87.
- Richiardi L, Boffetta P, Simonato L, et al. Occupational risk factors for lung cancer in men and women: a population-based case-control study in Italy. *Cancer Causes Control* 2004;**15**:285-94.
- International Labour Office. International Standard Classification of Occupations. Geneva: International Labour Office, 1968.
- International Labour Office. International Standard Industrial Classification. United Nations Publications (ST/STAT/M. 4/Rev 2/ Add 1, Sales No: E, 71.XVII.8). New York: Publishing Service, United Nations, 1971.
- Greenland S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology* 1995;**6**:356-65.
- Akaike H. A new look at the statistical model identification. *IEEE Trans Automatic Control* 1974;**19**:716-23.
- Mannetje A, Kogevinas M, Chang-Claude J, et al. Smoking as a confounder in case-control studies of occupational bladder cancer in women. *Am J Ind Med* 1999;**36**:75-82.
- Kriebel D, Zeka A, Eisen EA, et al. Quantitative evaluation of the effects of uncontrolled confounding by alcohol and tobacco in occupational cancer studies. *Int J Epidemiol* 2004;**33**:1040-5.
- D'Avanzo B, La Vecchia C, Katsouyanni K, et al. Reliability of information on cigarette smoking and beverage consumption provided by hospital controls. *Epidemiology* 1996;**7**:312-15.
- Donato F, Boffetta P, Fazioli R, et al. Reliability of data on smoking habit and coffee drinking collected by personal interview in a hospital-based case-control study. *Eur J Epidemiol* 1998;**14**:259-67.
- Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;**10**:37-48.
- Hernan MA, Hernandez-Diaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* 2002;**155**:176-84.
- US Department of Health and Human Services. Reducing tobacco use: a report of the Surgeon General. Atlanta: US Department of Health and Human Services, Center for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2000.
- La Vecchia C, Decarli A, Pagano R. Prevalence of cigarette smoking among subsequent cohorts of Italian males and females. *Prev Med* 1986;**15**:606-13.