In their recently published paper, Leffondré et al. (1) compared different approaches to modeling smoking history as a risk factor for lung cancer. Regression models with several smoking-related variables such as intensity, duration, and recency of smoking (models 16, 17, and 18 in the paper (1)) were shown to be better than simple models using only smoking status or cigarette-years. The more characteristics of exposure were incorporated into the model, the better was the goodness of fit (measured by Akaike’s Information Criterion).

However, there are some methodological limitations involved if more than one smoking-related factor is chosen as an independent variable in Cox’s regression model. The hazard ratio that is associated with an increase of one unit in one particular smoking-related variable is assumed to be the same regardless of the values of the other smoking-related variables. This assumption is very restrictive. For example, the effect on cancer risk of 1 additional year of exposure to tobacco will probably decrease with increasing time elapsed since cessation of smoking. In general, duration of smoking and time since cessation of smoking interact, and their effects cannot be separated. In addition, the hazard ratio associated with an increase of one cigarette smoked per day is not constant; rather, it depends on duration and recency. Because the smoking-related variables are highly interrelated, regression parameters for single characteristics are meaningless and cannot be interpreted appropriately. The issue of misinterpreting results of regression analysis that incorporates terms for each characteristic of an exposure was comprehensively discussed by McKnight et al. (2). Additionally, regression models with more than one smoking-related independent variable are prone to multicollinearity and model instability. It is doubtful that a somewhat better goodness of fit for these models outweighs the disadvantage of uninterpretable and unstable parameter estimates.

Leffondré et al. (1) did not consider a third possible approach to modeling smoking history: the use of one comprehensive smoking variable which allows for intensity, duration, and recency of smoking, including their interactions. Such an approach would avoid the problem of multicollinearity and is promising for obtaining a good model fit. A possible comprehensive smoking variable could be derived as follows. Assuming that lung cancer can be caused by a carcinogenic substance contained in cigarettes and cigarette smoke, the level of the carcinogen in the human body can partly be attributed to smoking. Applying a one-compartment exponential elimination model, the smoking-related component of the carcinogen level is proportional to 

$$X = \ln\left(\left(1 - 0.5^{c}d\right)0.5^{n}n + 1\right)$$

The value of $X$ increases as intensity ($n$) or duration ($d$) increases, but it decreases as time since cessation ($c$) increases. The variable $X$ is a suitable measure for the lifelong risk incurred by smoking, and it can be calculated for each person. Especially, it allows one to compare the risks of ex-smokers with those of current smokers. Leffondré et al.‘s published tables (1) suggest that $\tau$ is approximately 10 years, but it can also be estimated by maximizing the model fit. Moreover, the comprehensive smoking variable can even incorporate a lag time parameter $\delta$ that reflects the time between causal action and disease detection. In this case, $c$ must be replaced by $c^\delta = \max(c - \delta, 0)$ and $d$ must be replaced by $d^\delta = \max(d + c - \delta, 0) - d^\delta$. Figure 1 in Leffondré et al.’s paper suggests that the lag time is approximately 1 year.

It would be interesting to compare a lung cancer model using the proposed comprehensive smoking variable with models explored by Leffondré et al. (1). This comparison would show whether the model fit could be improved without impairing parameter interpretation and model stability.

REFERENCES


The Authors Reply

We thank Drs. Hoffmann and Bergmann for their helpful observations (1). They comment on some aspects of our study (2) and suggest a novel approach to modeling smoking history.

Hoffman and Bergmann (1) suggest that we did not consider the difficulties involved in simultaneous modeling.
of several smoking-related variables, yet comparison of our models 16–18 clearly underlined the difficulties in simultaneously modeling age at initiation, duration, and/or time since cessation while adjusting for age (2). We demonstrated that including all of these variables (model 17) was not tenable because of multicollinearity, such that “interpretation of the resulting estimates is impossible” (2, p. 820). McKnight et al. (3), in a study cited in our paper, focused on difficulties specific to simultaneous modeling of two continuous exposure variables that were categorized and that both had an assigned value of zero for nonexposed subjects. They suggested a solution similar to our inclusion of a binary smoking status indicator in model 9 (2). However, our approach avoids the limitations of categorizing continuous variables and allows exploration of multicollinearity problems, which were ignored by McKnight et al.

It is unclear why Hoffmann and Bergmann (1) are concerned that in our analyses, goodness of fit improves with the inclusion of additional variables. Akaike’s Information Criterion does not necessarily improve with an increasing number of covariates; rather, it corrects for this number (4). Indeed, our model 17 included one more variable than model 16 but yielded a worse Akaike’s Information Criterion (2).

Our paper focused on fairly simple and commonly used approaches to modeling smoking history. We never intended to explore all possible approaches. Specifically, we did not consider interactions between continuous smoking-related variables, partly because none of the recent studies we screened assessed such interactions and partly because of space limitations. For similar reasons, we did not consider more sophisticated approaches mentioned in our Discussion (2).

However, it may be of interest to investigate the advantages and limitations of the approach proposed by Hoffmann and Bergmann (1). Indeed, we believe that using their smoking indicator X may be especially interesting for testing the overall effect of smoking or adjusting for it. If the one-component model is consistent with the true (unknown) data structure, it might lead to a better goodness of fit than the use of, for example, separate variables for cigarette-years and time since cessation. However, there are some potential limitations of using X. First, the corresponding regression coefficient may be difficult to interpret. Moreover, the proposed formula of X implies, for example, a gradual leveling off of the effect of increasing smoking duration, which may not apply in some circumstances. Moreover, using X implies choosing a priori the values of half-time (τ) and lag (δ) parameters (1), both of which are likely to influence the results. However, in some studies there may not be sufficient prior knowledge to justify such choices. On the other hand, choosing these parameters a posteriori, as suggested by Hoffmann and Bergmann, may create some inferential problems, such as inflated type I error (5). Finally, some issues investigated in our paper (2) apply to X as well. Centering X and including the binary indicator of ever smoking in the model would help in interpreting the results of analyses that included never smokers. In addition, using X would not eliminate the problem of separating the effect of age at smoking initiation from other time-related smoking variables (2).

In summary, we think that the approach suggested by Hoffmann and Bergmann (1) may be of interest as a parsimonious representation of different aspects of smoking history. However, further investigation is needed to assess its potential advantages and limitations.

REFERENCES


Karen Leffondré1,2, Michal Abrahamowicz1,2, Jack Siemiatycki3,4, and Bernard Rachet4
1 Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada
2 Division of Clinical Epidemiology, The Montreal General Hospital, Montreal, Quebec, Canada
3 Département de Médecine Sociale et Préventive, Université de Montréal, Montreal, Quebec, Canada
4 Department of Epidemiology and Biostatistics, INRS-Institut Armand-Frappier, Université du Québec, Laval-des-Rapides, Quebec, Canada

In the October 15, 2002, issue of the Journal, Leroi et al. (1) presented an interesting paper suggesting that alcohol consumption is associated with less cognitive decline. However, their analysis was limited by the fact that the possible confounding effect of smoking on the association between alcohol and cognitive function was not reported.

In a prospective analysis, we examined the association of cigarette smoking and alcohol consumption at baseline with risk of poor cognitive function 13–18 years later (2). Smoking was associated with increased mortality in men and poorer cognitive function in women. After taking into account the effect of smoking, we found that in women, increasing consumption of alcohol predicted a decline in performance on two cognitive function tests and that consumption of two alcoholic drinks per day predicted decreased performance on the Buschke long-term recall task (2). However, the observed associations were weak, and no clear pattern was observed. Alcohol consumption was not associated with cognitive function in men.

Am J Epidemiol 2003;158:392–395

RE: “COGNITIVE FUNCTION AFTER 11.5 YEARS OF ALCOHOL USE: RELATION TO ALCOHOL USE”