In their recently published paper, Leffondrè 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.

Leffondrè 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

\[ X = \ln \left( 1 - 0.5^{d/\tau} 0.5^{c/n} \right) \]

where \( d, c, n, \) and \( \tau \) denote duration of smoking (in years), time since cessation (in years), number of cigarettes smoked per day, and the biologic half-life of the carcinogen (in years), respectively. As one can see from figure 2 of Leffondrè et al.’s paper, the term should be logarithmically transformed, so a comprehensive smoking variable is given by \( X = \ln\left(1 - 0.5^{d/\tau} 0.5^{c/n} + 1\right) \).

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. Leffondrè 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 \( c^* = \max(c - \delta, 0) \) and \( d \) must be replaced by \( d^* = \max(d + c - \delta, 0) - c^* \). Figure 1 in Leffondrè 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 Leffondrè et al. (1). This comparison would show whether the model fit could be improved without impairing parameter interpretation and model stability.

REFERENCES


Kurt Hoffmann1 and Manuela M. Bergmann2

1 Department of Epidemiology, German Institute of Human Nutrition, Arthur-Scheunert-Allee 114–116, 14558 Bergholz-Rehbrücke, Germany
2 German Institute of Human Nutrition, Potsdam-Rehbrücke, Germany

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