Invited Commentary

Invited Commentary: Is It Time to Retire the "Pack-Years" Variable? Maybe Not!

Duncan C. Thomas*

* Correspondence to Dr. Duncan C. Thomas, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, 2001 N. Soto Street, SSB-202F, Los Angeles, CA 90089-9234 (e-mail: dthomas@usc.edu).

Cumulative exposure—the product of intensity and duration for a constant exposure rate or its integral over time if variable—has been widely used in epidemiologic analyses of extended exposures, for example, the "pack-years" variable for tobacco smoking. Although the effects of intensity and duration are known to differ for exposures like smoking and ionizing radiation and simple cumulative exposure does not explicitly allow for modification by other time-related variables, such as age at exposure or time since exposure, the cumulative exposure variable has the merit of simplicity and has been shown to be one of the best predictors for many exposure-response relationships. This commentary discusses recent refinements of the pack-years variable, as discussed in this issue of the *Journal* by Vlaanderen et al. (*Am J Epidemiol.* 2014;179(3):290–298), in the broader context of general exposure-time-response relationships.

cancer; cumulative exposure; exposure-time-response relationships; models of carcinogenesis; radiation; smoking; time-related modifiers; tobacco

Without doubt, the longer one smokes and the more cigarettes one smokes per day, the greater one's chances of getting lung cancer and other smoking-related diseases are, so it seems natural to relate risk to the product of the two (duration of smoking and intensity of smoking)—the "pack-years" variable that has been so widely used in epidemiologic analyses. But does risk really relate to both components in the same way? What about other temporal variables, such as age at starting or time since smoking cessation? It's widely known that risk declines markedly after quitting smoking (relative to continuing to smoke), although it may never return to the level of lifelong nonsmokers. In addition, absolute age-adjusted excess rates increase quadratically with the number of cigarettes smoked per day and as at least the fourth power of duration of smoking (1), although other studies have yielded somewhat different estimates (2). After division by the risk in nonsmokers, the dependence of relative risk on duration and time since quitting becomes more complex, but these observations would suggest that the pack-years variable might not be the best predictor of risk. For most carcinogens, there is an extended latent period between exposure and disease; therefore, recent contributions to total dose are unlikely to be relevant. Nevertheless, the pack-years variable has the

virtue of simplicity and has stood the test of time as a strong predictor of the risk of various smoking-related diseases (3). The problem of how to summarize the effects of extended and variable exposures is a universal one in epidemiology, and cumulative dose has been widely used in analyses of numerous other exposures.

Although it is generally recognized that no single index will adequately summarize a complex exposure-time-response relationship, it is useful for many purposes to have a simple risk index. Investigators commonly begin by examining the effects of intensity, duration, and cumulative exposure separately and then in various combinations. Cumulative exposure generally emerges as the most strongly associated variable, and for smoking and lung cancer, it is consistently monotonically related to risk. Although modification by time-related factors like intensity, duration, and time since quitting has been well established, major differences in predicted risk for the same pack-years of smoking require comparisons across extremes (e.g., 5 years of smoking 40 cigarettes per day vs. 40 years of smoking 5 cigarettes per day) that seldom exist in the population. For discovery of novel associations, cumulative exposure is probably the most powerful variable one could use. Indeed, I am not aware of any novel associations that were

only discovered with the use of a more complex exposure metric. Also, many epidemiologic analyses involve multiple risk factors, of which smoking is but one—perhaps not even the one of primary interest—and thus a complex model for smoking may be overkill. The number of pack-years is also probably the most appropriate variable for downstream analyses of, say, the modifying effect of cigarette formulation or ways of targeting screening or other public health interventions. However, it may fail to reveal subtler phenomena that could shed light on mechanisms.

Various authors have interpreted the effects of intensity, duration, and other time-related modifiers as suggesting that smoking has multiple effects in the carcinogenic process, for example, initiation and promotion (4). This is hardly surprising, given that tobacco smoke is a complex mixture of agents, including both specific carcinogens like benzo(a)-pyrene and nonspecific agents like reactive oxygen species. Other authors have attempted to model these hypotheses mechanistically using the Armitage-Doll multistage (5–7) and Moolgavkar-Knudson 2-mutation-clonal expansion (8) models. These are too complex for routine epidemiologic analyses, however, so others have examined whether standard tools like logistic or Cox regression could be used with some richer set of covariates (9–13). This is not straightforward, as simply adding terms like time since cessation or its interaction with pack-years to a logistic model could end up predicting a lower risk in former smokers than in never smokers after a long enough interval.

An important series of articles by Lubin et al. (14–16) laid the groundwork for relatively simple models based on data from 2 to 4 lung cancer case-control studies (one of these papers included 6 studies of other smoking-related cancers). Their basic modeling strategy used pack-years as the primary exposure variable but incorporated additional modifying factors for intensity to correct for the model misspecification. These modifiers were included as log-linear terms multiplying the main effect of pack-years, thereby forcing them to be positive, which eliminated the problem noted above. Their analyses were restricted to never, current, and former-smokers who had quit within 5 years. The contribution by Vlaanderen et al. (17) in the current issue extends this approach by adding flexible modeling of cigarettes per day and time since quitting (no longer limited to recent former smokers), and applies it to 15 new case-control studies in the SYNERGY (a pooled analysis of case-control studies on the joint effects of occupational carcinogens in the development of lung cancer) consortium. In this sense, it may be only an incremental refinement of previous work, but it does provide arguably the most detailed model to date for predicting the risk of lung cancer after any duration and intensity and at any ages of starting, stopping, and at risk.

Because the smoking and lung cancer link is so strong and there are numerous large studies available, these data offer an opportunity to explore the separate effects of dose rate and duration in greater detail than is possible for most other exposures. However, an important limitation of most smoking data is that little detail is available about changes in intensity over time or periods of attempted quitting and relapse. The real utility of an exposure-time-response model comes in settings in which the entire dose history is available, as in various occupational cohort studies. In this situation, one might take a principled approach, modeling risk at a given age as the sum of effects of all prior increments of exposure, assuming each contributes independently. A simple linear model for these incremental contributions would lead back to cumulative dose as the risk covariate, but these contributions can be modified in various ways—for example, as a nonlinear function of dose rate, perhaps weighted by some functions of age at exposure or time since exposure—leading to the more flexible models described in the Web Appendix of the article by Vlaanderen et al. The assumption of additivity of risk contributions can be tested by including additional terms, say, to assess whether subsequent exposures enhance the effects of earlier ones, as might be predicted under the multistage model if smoking affects 2 or more stages. For example, analyses of the joint effect of radon and smoking have suggested that later smoking promotes the initiating effect of radon exposure but not vice-versa (18, 19).

Lubin et al. (20) used such an approach to modeling the risk of lung cancer in various uranium and other underground miner cohorts exposed to radon progeny, with a sophisticated investigation of the nonlinear effects of duration, dose-rate, and time since exposure. In particular, they found that a long low dose-rate exposure was more hazardous than a short high dose-rate exposure for the same cumulative dose. This would imply that the effect of lifetime exposure to indoor radon at low concentrations would be substantially larger than linear extrapolation from the miner data would suggest (21), an expectation borne out by meta-analyses of the many case-control studies of residential radon (22). For smoking, the Doll and Peto model (1) and the Lubin et al. (23) and Vlaanderen et al. (17) analyses predicted similar behavior for tobacco smoking, again suggesting larger effects of long-term exposure to secondhand tobacco smoke than would be expected based on active smoker data, a prediction that has also been supported by numerous epidemiologic studies (24).

In contrast, for low linear energy transfer radiation (like x-rays), a fractionated or protracted exposure produces lower risk for the same cumulative dose, suggesting differences in mechanism. For example, high linear energy transfer radiation (like α particles from radon progeny) tends to produce double-strand breaks, which are less readily repaired than the single-strand breaks predominantly caused by low linear energy transfer, pairs of which if not repaired can lead to double-strand breaks. Longer exposure intervals would allow more time for repair to occur between multiple hits; for this reason, low linear energy transfer radiation may also produce a stronger quadratic component to the dose-rate effect (e.g., for leukemia) (25). Such contrasts are illustrative of how rich exposure-time-response models have the potential to shed light on mechanisms.

One must be cautious in such interpretations, however. For example, most empirical and mechanistic models have assumed a homogeneous population, but if there is heterogeneity across individuals in the baseline hazard or sensitivity to smoking, then "survival of the fittest" will tend to distort the average relative risk (26); for example, in the Doll and Peto model, the relative risk in continuing smokers continued to increase with attained age, whereas in a heterogeneous population, it will begin to level off and even decline. Any

attempt to interpret such a phenomenon mechanically, without recognizing the differential survival effect, is bound to be misleading.

One practical difficulty is that the effects of age at starting, duration, time since quitting, and attained age are collinear—attained age being simply the sum of the first 3—so their modifying effects cannot be separated (27, 28). An adequate fit can be obtained using any subset of these variables, the choice generally being driven by parsimony and interpretability. For example, there tends to be less variability in age at starting than in the other variables, and one wouldn't want to analyze its effect without adjusting for one or more of the other variables, as their effects are too powerful to ignore.

There is also the question of scale: In some cases, it may be preferable to build models on an excess absolute rate scale, as Doll and Peto did (1) and then express these models in terms of relative risk as needed. This issue has been extensively discussed in the radiation field, where it has been shown that excess absolute risk models tend to allow more parsimonious modeling of time-related modifiers (29). For smoking, however, Lubin et al. (14–16) and Vlaanderen et al. (17) chose to model the effect of pack-years and its modifiers on an excess relative risk scale. It remains to be seen which scale yields more parsimonious or interpretable models for tobacco exposures.

In any event, the conclusion could be that the naïve use of pack-years (or cumulative dose) by itself is inadequate for any disease in which exposures are protracted over time and have a strong effect. Then again, it has served us well in countless studies, despite widespread acknowledgment of its potential limitations. To expect that any single covariate could entirely encapsulate the effects of a complex process extended over time is unrealistic. The article by Vlaanderen et al. (17) and others like it have pointed the way to relatively simple models that can further improve risk predictions by adding intensity, duration, time since exposure, or other time-related variables. However, for most purposes (discovery of novel effects, adjustment for confounding, exploration of other modifiers like cigarette formulation, prediction of risk for screening purposes, etc.), the simple pack-years variable should not be dismissed. In particular, powerful vested interests have a history of using criticisms of pack-years or other simple exposure indices as a way of obfuscating the evidence without providing any alternative metric. As Samet et al. point out, "models are just a means to the end—ending the epidemic of lung cancer deaths" (30, p. 650).

REFERENCES

1. Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. *J Epidemiol Community Health*. 1978;32(4): 303–313.
2. Knoke JD, Shanks TG, Vaughn JW, et al. Lung cancer mortality is related to age in addition to duration and intensity of cigarette smoking: an analysis of CPS-I data. *Cancer Epidemiol Biomarkers Prev*. 2004;13(6):949–957.
3. Lubin JH, Caporaso NE. Misunderstandings in the misconception on the use of pack-years in analysis of smoking. *Br J Cancer*. 2013;108(5):1218–1220.
4. Day NE, Brown CC. Multistage models and primary prevention of cancer. *J Natl Cancer Inst*. 1980;64(4):977–989.
5. Brown CC, Chu KC. Implications of the multistage theory of carcinogenesis applied to occupational arsenic exposure. *J Natl Cancer Inst*. 1983;70(3):455–463.
6. Freedman DA, Navidi WC. Ex-smokers and the multistage model for lung cancer. *Epidemiology*. 1990;1(1):21–29.
7. Whittemore AS. Effect of cigarette smoking in epidemiological studies of lung cancer. *Stat Med*. 1988;7(1-2):223–238.
8. Hazelton WD, Clements MS, Moolgavkar SH. Multistage carcinogenesis and lung cancer mortality in three cohorts. *Cancer Epidemiol Biomarkers Prev*. 2005;14(5):1171–1181.
9. Leffondre K, Abrahamowicz M, Xiao Y, et al. Modelling smoking history using a comprehensive smoking index: application to lung cancer. *Stat Med*. 2006;25(24):4132–4146.
10. Siemiatycki J. Synthesizing the lifetime history of smoking. *Cancer Epidemiol Biomarkers Prev*. 2005;14(10):2294–2295.
11. Leffondre K, Abrahamowicz M, Siemiatycki J, et al. Modeling smoking history: a comparison of different approaches. *Am J Epidemiol*. 2002;156(9):813–823.
12. Thurston SW, Liu G, Miller DP, et al. Modeling lung cancer risk in case-control studies using a new dose metric of smoking. *Cancer Epidemiol Biomarkers Prev*. 2005;14(10): 2296–2302.
13. Rachet B, Siemiatycki J, Abrahamowicz M, et al. A flexible modeling approach to estimating the component effects of smoking behavior on lung cancer. *J Clin Epidemiol*. 2004; 57(10):1076–1085.
14. Lubin JH, Caporaso N, Wichmann HE, et al. Cigarette smoking and lung cancer: modeling effect modification of total exposure and intensity. *Epidemiology*. 2007;18(5):639–648.
15. Lubin JH, Alavanja MC, Caporaso N, et al. Cigarette smoking and cancer risk: modeling total exposure and intensity. *Am J Epidemiol*. 2007;166(4):479–489.
16. Lubin JH, Caporaso NE. Cigarette smoking and lung cancer: modeling total exposure and intensity. *Cancer Epidemiol Biomarkers Prev*. 2006;15(3):517–523.
17. Vlaanderen J, Portengen L, Schüz S, et al. Effect modification of the association of cumulative exposure and cancer risk by intensity of exposure and time since exposure cessation: a flexible method applied to cigarette smoking and lung cancer in the SYNERGY Study. *Am J Epidemiol*. 2014;179(3): 290–298.
18. Moolgavkar SH, Luebeck EG, Krewski D, et al. Radon, cigarette smoke, and lung cancer: a re-analysis of the Colorado Plateau uranium miners' data. *Epidemiology*. 1993;4(3): 204–217.
19. Thomas D, Pogoda J, Langholz B, et al. Temporal modifiers of the radon-smoking interaction. *Health Phys*. 1994;66(3): 257–262.
20. Lubin JH, Boice JD Jr, Edling C, et al. Radon-exposed underground miners and inverse dose-rate (protraction enhancement) effects. *Health Phys*. 1995;69(4):494–500.

21. Lubin JH, Boice JD Jr, Edling C, et al. Lung cancer in radon-exposed miners and estimation of risk from indoor exposure. *J Natl Cancer Inst*. 1995;87(11): 817–827.
22. Krewski D, Lubin JH, Zielinski JM, et al. A combined analysis of North American case-control studies of residential radon and lung cancer. *J Toxicol Environ Health A*. 2006;69(7): 533–597.
23. Lubin JH. Estimating lung cancer risk with exposure to environmental tobacco smoke. *Environ Health Persp*. 1999; 107(suppl 6):879–883.
24. Surgeon General. The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General. Atlanta, GA: Centers for Disease Control and Prevention; 2006.
25. Kellerer AM, Rossi HH. A generalized formulation of dual radiation action. *Radiat Res*. 2012;178(2):AV204–AV213.
26. Thomas DC. *Statistical Methods in Environmental Epidemiology*. Oxford, UK: Oxford University Press; 2009.
27. Thomas DC. Pitfalls in the analysis of exposure-time-response relationships. *J Chron Dis*. 1987;40(suppl 2):71S–78S.
28. Thomas DC. Exposure-time-response relationships with applications to cancer epidemiology. *Annu Rev Public Health*. 1988;9:451–482.
29. Preston DL, Pierce DA, Shimizu Y, et al. Dose response and temporal patterns of radiation-associated solid cancer risks. *Health Phys*. 2003;85(1):43–46.
30. Samet JM, Thun MJ, de Gonzalez AB. Models of smoking and lung cancer risk: a means to an end. *Epidemiology*. 2007;18(5): 649–651.