BJC

Misunderstandings in the misconception on the use of pack-years in analysis of smoking

J H Lubin*,[1] and N E Caporaso[2]

[1]Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Boulevard, Room 8105, Rockville, MD 20852, USA and [2]Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Boulevard, Room 7122, Rockville, MD 20852, USA

Sir,

In a recent editorial, Peto (2012) states that the 'single measure of lifetime cumulative dose (dose rate times duration)' is 'unnecessary and scientifically unhelpful', as there is 'long-standing evidence that cancer risk at a given cumulative dose sometimes varies substantially with the duration of exposure.' Further he also states 'Science advances by developing and testing plausible models, not by regression analysis of gross deviations from models that are clearly wrong'; 'Lung cancer risk is not proportional to pack-years'; and 'modeling of the variation in ERR (excess relative risk) per pack-year in relation to … smoking rate … is unlikely to be biologically informative.' He proffers two lung cancer-related examples: radon, where the ERR per working level month (WLM) increases with duration; and cigarette smoking. These diverse examples suggest that he intends his comments to apply universally to all exposures, and therefore: (i) cumulative exposure metrics are never useful for modelling risk; and (ii) variations of the disease and cumulative exposure association by duration (or equivalently exposure rate) are biologically uninformative. In our view, evidence strongly contradicts both statements.

Analyses by cumulative exposure and exposure rate provide a unique perspective on risk, potentially leading to enhanced mechanistic understanding, whereas, in contrast to common belief, parameter estimates from models in exposure duration and rate are not interpretable as 'separate' and 'independent' effects. The recommendation to abandon cumulative exposure-based metrics serves only to restrict flexibility in data analysis and risk modelling, and thereby limit inference on biological mechanisms. Cumulative exposure metrics have a long history of proven success in increasing our understanding of disease aetiology and formulating public health policy.

Cigarette smoking analyses typically start with computation of marginal relative risks (RRs), that is, unadjusted for other smoking-related variables, for three primary metrics: smoking duration, cigarettes smoked per day (CPD) and pack-years. As only pack-years estimate the total body burden of the presumed carcinogen, it is the single variable most relevant for characterising exposure and, thus, risk. Nevertheless, it is abundantly clear that pack-years alone does not fully describe smoking-related lung cancer risk (Doll and Peto, 1978; Lubin and Caporaso, 2006). Investigators therefore extend analysis to two variables, cross-classifying variables or adjusting one variable for the other. The selected variables may be smoking duration and CPD as in the Doll–Peto model (Doll and Peto, 1978), or pack-years and CPD as in our model (denoted as the L–C model; Lubin and Caporaso, 2006). As pack-years equal duration times CPD/20, any 'duration and CPD' model is transformed into a 'pack-year and CPD' model simply by replacing duration with pack-years/(CPD/20). Consequently, there is no practical difference in the choice and neither is intrinsically preferable for model building. The Doll–Peto model predicts that lung cancer rates increase with the fourth power of duration and the square of CPD. However, these predications are equally described as increasing with the fourth power of pack-years and the square of 1/CPD, that is, decreasing with CPD with pack-years fixed. This change alters only interpretation of parameters (see below), without affecting model fit. Furthermore, if 'aging *per se* is irrelevant' in a 'duration and CPD' model (Peto, 2012), then age is also irrelevant in a 'pack-years and CPD' model. Notably, the Doll–Peto model indeed varied with age when applied in both the American Cancer Society's Cancer Prevention Study I (CPS-I; Knoke *et al*, 2004) and II (CPS-II; Flanders *et al*, 2003). The real issue is not that 'duration and CPD' models are good and 'pack-years and CPD' models are bad, but rather the interpretability of parameters and consistency of models with observed data and with current understanding of biological mechanisms.

With 'duration and CPD' models, parameter interpretations are inherently ambiguous, as duration effects with CPD fixed necessarily embed pack-years effects (Lubin and Caporaso, 2006). In Peto's Table 1, predicted lung cancer rates for 20 CPD, current smokers increase with duration. However, there is no obligation to assign the cause of the increasing rates to increasing duration, rather it could equally be assigned to increasing pack-years.

*Correspondence: Dr JH Lubin; E-mail: lubinj@exchange.nih.gov

Published online 28 February 2013

Compared with a 70-year-old smoker, an 80-year-old 20 CPD smoker accrues not only 10 years additional duration but also 10 pack-years. Thus, it is no less reasonable to suppose that the increased lung cancer rate for the 80-year-old derives from the consumption of 73 000 additional cigarettes. Interpretation of CPD effects at fixed duration is likewise problematic. For a 30-year duration, risks at 20 and 30 CPD necessarily embed risks from 30 and 45 pack-years, respectively. RRs or absolute risks by duration and CPD are thus not interpretable as separate and 'independent' effects.

In contrast, a model in pack-years and CPD reformulates analysis in terms of the quantitative trend with pack-years and the modifying effects of CPD, or more precisely 'delivery rate' effects. Delivery rate effects describe the relative impact on the disease and pack-years association for a given pack-years delivered at higher exposure rate for shorter duration compared with lower exposure rate for longer duration. For 80 pack-years, the delivery rate effect measures the extent that smoking 2 packs/day for 40 years results in a larger, equal or smaller RR (or absolute risk) compared with smoking 4 packs/day for 20 years.

Specifically for adjustment variables (z), pack-years (d) and CPD (n), the L–C model posits a disease rate of $r(z, d, n) = r_o(z) \times RR(d, n)$, where $r_o(.)$ is the rate in never-smokers and $RR = 1 + \beta dg(n)$. The ERR/pack-year (β) represents the strength of association, whereas $g(.)$ describes delivery rate effects that may be fitted parametrically or with splines. For each n, RRs by pack-years increase linearly with slope $\beta\ g(n)$. This formulation emerged directly from observed RRs for pack-years and CPD, ensuring a good description of smoking-related risks. Questions concerning age, age at initiation, cessation and so on reflect potential effect modification, that is, variations of β and/or $g(.)$.

The L–C model predications compare favourably with other models. For CPS-I data, Knoke et al (2004) significantly improved the Doll–Peto model by including either age or age at smoking initiation. We compared the L–C model inserting Knoke's lung cancer rate model in never-smokers for $r_o(.)$ with Knoke's preferred duration/CPD/age model. Although L–C model parameters were estimated independently of CPS-I data, predicted smokers' rates were nearly identical (Figure 5 in Lubin and Caporaso, 2006). At age 60 years, predicted yearly lung cancer rates for 10, 20 and 30 CPD smokers were 0.0011, 0.0018 and 0.0025 for the duration/CPD/age model, respectively, and 0.0010, 0.0020 and 0.0027, respectively, for the L–C model.

Above 10–15 CPD, the L–C model specifies an inverse delivery rate effect, whereby smoking more CPD for shorter duration is less deleterious (per cigarette) than smoking fewer CPD for longer duration, a pattern consistent with 'reduced potency' (Lubin and Caporaso, 2006). The inverse delivery rate pattern occurs consistently across lung cancer studies and smoking-related cancer sites, including oesophagus, bladder, pancreas, kidney, oral cavity, larynx and pharynx (Lubin et al, 2007a, 2008, 2009, 2010, 2012). Thus, delivery rate represents an important modulator of risk, and its consistency suggests a general smoking-related phenomenon. Under 5–10 CPD, the L–C model describes a direct delivery rate effect, with increasing strength of association with increasing CPD; however, pack-year ranges are necessarily limited and effects are estimated with substantial uncertainty, and additional analyses are needed.

The inverse delivery rate may reflect smoking-related biological mechanisms, such as increased DNA repair, increased induction of detoxification enzymes or saturation of activation enzymes (Lewtas et al, 1997). Heavy smokers exhibited increased DNA repair capacities compared with light smokers (Wei et al, 2000; Shen et al, 2003; Spitz et al, 2003). Polycyclic aromatic hydrocarbons (PAHs) from incomplete tobacco combustion undergo metabolic activation to form DNA and protein adducts (Lewtas et al, 1997; Lutz, 1998; Phillips, 2002). Lewtas et al (1997) observed higher DNA adduct

levels in white blood cells per unit PAH exposure in environmentally exposed individuals than in high-exposed workers. More directly, nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) is a tobacco-specific carcinogen. Among smokers, ratios of urinary NNK metabolites to urinary cotinine declined with increasing cotinine, indicating reduced NNK uptake per unit cotinine with increasing cotinine (Lubin et al, 2007b). Finally, the N-acetyltransferase 2 (NAT2) enzyme detoxifies aromatic amines, a class of tobacco-related carcinogens, with slow acetylation phenotypes that have reduced detoxification capacity compared with rapid/intermediate phenotypes, and also have a well-described impact on both carcinogen-adduct levels and subsequent cancer. At low and moderate CPD, phenotypes exhibit similar bladder cancer risks, whereas at high CPD, rapid/intermediate acetylators exhibit reduced risks relative to slow acetylators (Gu et al, 2005; Lubin et al, 2007a).

The inverse delivery rate pattern may also reflect dosimetric changes related to nicotine dependency, with heavier smokers inhaling less vigorously, leading to lower carcinogenic yields per cigarette. Although evidence supports such dosimetric changes (Patterson et al, 2003; US Department of Health and Human Services, 2010), in one lung cancer study inhalation did not confound pack-years variations with CPD (Lubin et al, 2007c). Also, sensitivity analyses using the relationship between urinary cotinine and CPD to 'correct' CPD estimates found that dosimetric changes could not fully explain delivery rate patterns (Lubin et al, 2007c).

Radon exposure also challenges Peto's assertions about cumulative exposure metrics. Multiple studies of underground miners demonstrate that lung cancer RRs by cumulative WLM increase linearly, and that the ERR/WLM decreases with working level (WL; National Research Council, 1999; Walsh et al, 2010). Moreover, miner-based model predictions correspond precisely to observed risks in residentially exposed populations, whereas in vivo studies, in vitro studies and radiobiological models provide a mechanistic basis for observed patterns (National Research Council, 1999). Radon and its decay products are α-particle emitters and a single α-particle can damage DNA. Radiobiological analysis predicts dose rate effects. At residential exposure levels, a cell nucleus incurs a <0.01 probability of 'seeing' even one α-particle per year, and hence cannot 'experience' a delivery rate effect. As multiple traversals are rare, doubling α-particles mainly doubles the number of cells traversed, that is, risks are approximately proportional to dose. At high exposures, multiple traversals are highly probable, yielding increased cell death, greater 'wasted dose' and a decreased exposure–response relationship. Miners' data exhibit both proportionality of excess RRs with WLM and ERR/WLM variations, with no delivery rate effects at low WLs and inverse delivery rate effects at high WLs (National Research Council, 1999). This concordance of epidemiology and radiobiology explains why expert committees and health policy agencies worldwide have long used this characterisation for predicting radon-associated lung cancer.

Parameters in cumulative exposure and exposure rate models are directly interpretable in terms of the disease and cumulative exposure relationship and the modulating effects of exposure delivery (high exposure rate for short duration or low rate for long duration). In contrast, interpretation of parameters in duration and exposure rate models is ambiguous due to imbedded cumulative exposure effects. More generally, increased understanding of biological mechanisms is best achieved when investigators analyse data carefully using the broadest range of tools. There is little rationale in arbitrarily labelling any class of exposure metrics as inherently invalid and off-limits, thereby restricting explanatory models. Prohibitions on exposure metrics or model formulations do not serve to advance science and should be rejected.

REFERENCES

Doll R, Peto R (1978) Cigarette-smoking and bronchial-carcinoma - dose and time relationships among regular smokers and lifelong non-smokers. *J Epidemiol Commun Health* **32**(4): 303–313.

Flanders WD, Lally CA, Zhu BP, Henley SJ, Thun MJ (2003) Lung cancer mortality in relation to age, duration of smoking, and daily cigarette consumption: results from Cancer Prevention Study II. *Cancer Res* **63**(19): 6556–6562.

Gu J, Liang D, Wang YF, Lu C, Wu XF (2005) Effects of N-acetyl transferase 1 and 2 polymorphisms on bladder cancer risk in Caucasians. *Mutat Res* **581**(1-2): 97–104.

Knoke JD, Shanks TG, Vaughn JW, Thun MJ, Burns DM (2004) 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* **13**(6): 949–957.

Lewtas J, Walsh D, Williams R, Dobias L (1997) Air pollution exposure DNA adduct dosimetry in humans and rodents: evidence for non-linearity at high doses. *Mutat Res* **378**(1-2): 51–63.

Lubin JH, Caporaso N (2006) Cigarette smoking and lung cancer: modeling total exposure and intensity. *Cancer Epidemiol Biomarkers Prev* **15**(3): 517–523.

Lubin JH, Caporaso N, Hatsukami DK, Joseph AM, Hecht SS (2007b) The association of a tobacco-specific biomarker and cigarette consumption and its dependence on host characteristics. *Cancer Epidemiol Biomarkers Prev* **16**: 1852–1857.

Lubin JH, Caporaso N, Wichmann HE, Schaffrath-Rosario A, Alavanja MCR (2007c) Cigarette smoking and lung cancer: modeling effect modification of total exposure and intensity. *Epidemiol* **18**: 639–648.

Lubin JH, Cook MB, Pandeya N, Vaughan TL, Abnet CC, Giffen C, Webb PM, Murray LJ, Casson AG, Risch HA, Ye W, Kamangar F, Bernstein L, Sharp L, Nyren O, Gammon MD, Corley DA, Wu AH, Brown LM, Chow WH, Ward MH, Freedman ND, Whiteman DC (2012) The importance of exposure rate on odds ratios by cigarette smoking and alcohol consumption for esophageal adenocarcinoma and squamous cell carcinoma in the Barrett's Esophagus and Esophageal Adenocarcinoma Consortium. *Cancer Epidemiol* **36**(3): 306–316.

Lubin JH, Gaudet MM, Olshan AF, Kelsey K, Boffeta P, Brennan P, Castellsague X, Chen C, Curado MP, Maso LD, Daudt AW, Fabianova E, Fernandez L, Wunsch-Filho V, Franceschi S, Herrero R, Koifman S, La Vecchia C, Lazarus P, Levi F, Lissowska J, Mates IN, Matos E, Menezes A, Morgenstern H, Muscat J, Neto JE, Rudnai P, Schwartz SM, Zaridze D, Shang B, Smith E, Sturgis EM, Szeszenia-Dabrowska N, Talamini R, Wei QY, Winn D, Zhang ZF, Hashibe M, Hayes RB (2010) Body mass index, cigarette smoking and alcohol consumption and cancers of the oral cavity, pharynx and larynx: modeling odds ratios in pooled case-control data. *Am J Epidemiol* **171**(12): 1250–1261.

Lubin JH, Kogevinas M, Silverman DT, Malats N, Garcia-Closas M, Tardon A, Hein DW, Garcia-Closas R, Serra C, Dosemeci M, Carrato A, Rothman N (2007a) Evidence for an intensity dependent interaction of NAT2 acetylation genotype and cigarette smoking in the Spanish Bladder Cancer Study. *Int J Epidemiol* **36**: 236–241.

Lubin JH, Purdue M, Kelsey KT, Zhang ZF, Winn DM, Wei QY, Talamini R, Szeszenia-Dabrowska N, Sturgis EM, Smith E, Shangina O, Schwartz SM, Rudnai P, Neto JE, Muscat J, Morgenstern H, Menezes A, Matos E, Mates IN, Lissowska J, Levi F, Lazarus P, La Vecchia C, Koifman S, Herrero R, Franceschi S, Wunsch-Filho V, Fernandez L, Fabianova E, Daudt AW, Dal Maso L, Curado MP, Chen C, Castellsague X, Brennan P, Boffeta P, Hashibe M, Hayes RB (2009) Total exposure and exposure rate effects for alcohol and smoking and risk of head and neck cancer: a pooled analysis of case-control studies. *Am J Epidemiol* **170**(8): 937–947.

Lubin JH, Virtamo J, Weinstein SJ, Albanes D (2008) Cigarette smoking and cancer: intensity patterns in the Alpha-Tocopherol Beta-Carotene Cancer Prevention Study in Finnish men. *Am J Epidemiol* **167**(8): 970–975.

Lutz WK (1998) Dose-response relationships in chemical carcinogenesis: superposition of different mechanisms of action, resulting in linear-nonlinear curves, practical thresholds, J-shapes. *Mutat Res* **405**(2): 117–124.

National Research Council (1999) *Health Effects of Exposure to Radon (BEIR VI)*. National Academies Press: Washington, DC.

Patterson F, Benowitz N, Shields P, Kaufmann V, Jepson C, Wileyto P, Kucharski S, Lerman C (2003) Individual differences in nicotine intake per cigarette. *Cancer Epidemiol Biomarkers Prev* **12**(5): 468–471.

Peto J (2012) That effects of smoking should be measured in pack-years: misconceptions 4. *Br J Cancer* **107**(3): 406–407.

Phillips DH (2002) Smoking-related DNA and protein adducts in human tissues. *Carcinogenesis* **23**(12): 1979–2004.

Shen HB, Spitz MR, Qiao YW, Guo ZZ, Wang LE, Bosken CH, Amos CI, Wei QY (2003) Smoking, DNA repair capacity and risk of nonsmall cell lung cancer. *Int J Cancer* **107**(1): 84–88.

Spitz MR, Wei QY, Dong Q, Amos CI, Wu XF (2003) Genetic susceptibility to lung cancer: The role of DNA damage and repair. *Cancer Epidemiol Biomarkers Prev* **12**(8): 689–698.

US Department of Health and Human Services (2010) *How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General*. ISBN 978-0-16-084078-4 US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Superintendent of Documents, US Government Printing Office: Washington, DC.

Walsh L, Tschense A, Schnelzer M, Dufey F, Grosche B, Kreuzer M (2010) The Influence of Radon Exposures on Lung Cancer Mortality in German Uranium Miners, 1946-2003. *Radiat Res* **173**(1): 79–90.

Wei QY, Cheng L, Amos CI, Wang LE, Guo ZZ, Hong WK, Spitz MR (2000) Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: a molecular epidemiologic study. *J Natl Cancer Inst* **92**(21): 1764–1772.