EDITOR'S CHOICE

Smoking and lung cancer: causality, Cornfield and an early observational meta-analysis

George Davey Smith

As has been noted before,[1] we receive a surprising number of papers that present data on smoking and lung cancer with a contrived air of originality and excitement. This is the first time that the link has been shown in this particular subgroup of our population! Well, the second perhaps, but we could adjust for more confounders. We have none of these papers in this issue of the *IJE*—or in fact in any issue, unless under the influence of influenza or a hangover we allow our guard to slip. However, we are lucky enough to publish the second article on smoking and lung cancer under our editorship that is indeed exciting and original. The first of these was our translation into English of Schairer and Schöniger's pioneering case–control study carried out in Germany in the early 1940s.[2] The second is a reprint to mark 50 years after the initial publication of Jerome Cornfield and colleagues' ground-breaking account of how causal inference could be applied to observational data on smoking and lung cancer,[3] with a series of admiring commentaries that place it in context and discuss how well its ideas have stood the test of time (pretty well is the general consensus).[4–7]

As a trained historian—Cornfield's (Figure 1) initial degree and graduate study were in history [8]—we hope he would appreciate the often historical focus of material in the *IJE*. It proved very difficult to select which of his important papers to reprint. A strong case could be made for his 1956 paper 'A statistical problem arising from retrospective studies',[9] generally noted for its discussion of how under certain assumptions the odds ratio is a fairly good approximation of the relative risk.[8] This paper also presented an early meta-analysis (although Cornfield did not call it this) of 14 case–control studies of smoking and lung cancer the data for which were attributed to a 1954 paper by Cornfield's ex-boss, Harold Dorn.[10] The Schairer and Schöniger case–control study[2] was included in this meta-analysis as study number 2, as was an earlier German study and several subsequent ones. The studies often said to be the pioneers—Wynder and Graham for North Americans[11] and Doll and Hill for Europeans[12]—were studies 6 and 8, respectively, indicating their position in the ordering by date of

E-mail: ije-editorial@bristol.ac.uk

Figure 1 Reproduced with permission of Johns Hopkins Bloomberg School of Public Health

publication. Cornfield pointed out that while 'methods exist for deciding whether the differences among the studies are significant, this is not a question of great interest. Rather we should like an interval estimate of the extent to which they do differ'.[9] Focusing on 10 studies which appeared to be attempting to estimate the same parameter (which lead to Wynder and Graham's study being excluded) he concluded that the relative risk lay between 5.0 and 7.2, pointing out that this would be considerably larger if the risk in just cigarette smokers could be examined. Reflecting one of the limitations of meta-analysis of published data, only the risk in smokers of any tobacco product vs non-smokers could be estimated.

Interestingly, through Cornfield's paper, the data on these 14 case–control studies of smoking and lung

cancer have been frequently utilized by methodologists developing different approaches to meta-analysis. Indeed, Schairer and Schöniger's pioneering study was more often included in such methodological work than it was directly cited, until the recent period of rekindled interest in it.[13,14]

Despite having so much to recommend it, the 1956 paper by Cornfield[9] contains pages consisting entirely of statistical equations. While this might appeal to that valued subgroup of the *IJE* readership that cannot maintain eye contact with other humans, we thought it could prove alienating to the less mathematically inclined. We therefore chose the 1959 paper, since the only mathematical formulae in it are usefully contained in appendices, for the delectation of aficionados.

As our commentators point out,[4–7] the 1959 paper is an extremely lucid exposition of many ideas that now seem to be the common sense of epidemiology. Other ideas remain to be fully assimilated—for example, the form of sensitivity analysis that was advanced probably remains under-utilized in current epidemiological practice. As with all observational data, even in the case of smoking and lung cancer it is possible to come up with non-causal explanations, however implausible these may be. Cornfield and colleagues discussed the 'constitutional hypothesis' advanced by the eminent geneticist and statistician RA Fisher, who proposed that cigarette smoking and lung cancer could both be influenced by a constitutional make-up, perhaps genetic in origin, which predisposes individuals to both of them. Cornfield and colleagues pointed out that it was unlikely that a randomized controlled trial with 30–60 years follow-up would ever be carried out to demonstrate that cigarette smoking caused lung cancer. Ironically, in writings elsewhere Fisher discussed the analogies between his pioneering work in randomized experiments and genetics, and stated that 'Genetics is indeed in a peculiarly favored condition in that providence has shielded the geneticist from many of the difficulties of a reliably controlled comparison. The different genotypes possible from the same mating have been beautifully randomized by the meiotic process. Generally speaking, the geneticist, even if he foolishly wanted to, could not introduce systematic errors into the comparison of genotypes, because for most of the relevant time he has not yet recognized them'.[15] Far from casting doubt on the cause and nature of the association between smoking and lung cancer, genetic variants related to smoking can provide essentially randomized evidence regarding smoking as a cause of lung cancer. When adequately powered genome-wide association studies were finally carried out on lung cancer they identified as their top hit (although one associated with a small increased relative risk of disease) a variant in the nicotinic receptor.[16–18] Elsewhere it has been shown[19] that this is related to differences in ability to quit smoking and to smoking behaviours, such as depth of inhalation.[20] Furthermore, the variant is associated with several other smoking-related diseases. The most parsimonious explanation is that this provides Mendelian randomization evidence[21] of a causal effect of smoking on these diseases.

Perhaps one of the advantages that Cornfield had was his lack of any sustained formal training in either epidemiology or biostatistics. As JBS Haldane—who recently graced the 'Reprints and Reflections' section of the *IJE*[22]—pointed out

'I consider it desirable that a man's or a woman's major research work should be on a subject in which he or she has not taken a degree. To get a degree one has to learn all the facts and theories in a somewhat parrot-like manner. One may also learn something much more important, namely how a branch of knowledge has been organised. And a piece of research directed by a good scientist should leave one with high standards of accuracy and integrity which one can transfer to other fields of science. It is rather hard to be highly original in a subject that one has learned with a view to obtaining first-class honours in an examination'.[23]

Perhaps the growth of formal epidemiology courses over recent decades is doing a disservice to the originality of thinking in the field.

As might be anticipated given its wide-ranging discussion of epidemiological fundamentals, the issues covered in the paper we reprint[3] are reflected in many of the topics covered in the current issue. For example, considerations of selection bias, among other biases, permeate the debate about whether or not there is a real increase in incidence of autism;[24–28] complex confounding is demonstrated in Mika Kivimäki and colleagues'[29] investigation of the contribution of psychosocial working environment to the link between socio-economic position and stroke and the entire panoply of considerations raised by Cornfield and colleagues apply to interrogating the association of biofuel exposure and environmental tobacco smoke with health outcomes among infants and children.[30,31]

Epidemiological investigations of cigarette smoking and lung cancer can be considered a major success of the discipline. However, after the associations are demonstrated the hard task can be seen to begin—how do we use this information to improve population health? With cigarette smoking it has been a long struggle, but in many places mixtures of social sanction, restriction on where smoking can occur and who cigarettes can be sold to and fiscal disincentives have produced dramatic reductions in cigarette smoking and health improvement. Policy in other areas is both less developed and perhaps more problematic to implement. A thoughtful discussion of how taxation policies could beneficially improve the health aspects of diet is given by Nnoaham and colleagues,[32] with

challenging implications for the regressive nature of such policies and their impact on social inequality. A mixture of data types are utlilized in this paper. How to reach adequate levels of causal certainty for basing policies on these kinds of data synthesis and modelling exercises is the kind of question that I would like to be able to ask Jerome Cornfield.

References

[1] Ebrahim S. Shock and awe: waking the dead. *Int J Epidemiol* 2007;**36:**701–02.

[2] Schairer E, Schöniger E. Lung cancer and tobacco consumption. *Int J Epidemiol* 2001;**30:**24–27.

[3] Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, ShimkinMB, Wynder EL. Smoking and lung cancer: recent evidence and a discussion of some questions. *Int J Epidemiol* 2009;**38:**1175–91.

[4] Cox DR. Commentary: Smoking and lung cancer: reflections on a pioneering paper. *Int J Epidemiol* 2009;**38:**1192–93.

[5] Vandenbroucke JP. Commentary: 'Smoking and lung cancer'—the embryogenesis of modern epidemiology. *Int J Epidemiol* 2009;**38:**1193–96.

[6] Greenhouse JB. Commentary: Cornfield, epidemiology and causality. *Int J Epidemiol* 2009;**38:**1199–1201.

[7] Zwahlen M. Commentary: Cornfield on cigarette smoking and lung cancer and how to assess causality. *Int J Epidemiol* 2009;**38:**1197–98.

[8] Greenhouse SW, Greenhouse J. Jerome Cornfield. In: Armitage P, Colton T (eds). *The Encyclopedia of Biostatistics*. New York: John Wiley, 1998.

[9] Cornfield J. A statistical problem arising from retrospective studies. *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics* 1956;**4:**135–48.

[10] Dorn HF. The relationship of cancer of the lung and the use of tobacco. *Am Stat* 1954;**8:**7–13.

[11] Wynder EL, Graham EA. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. *J Am Med Assoc* 1950;**143:**329–36.

[12] Doll R, Hill AB. Smoking and carcinoma of the lung. Preliminary report. *Br Med J* 1950:739–48.

[13] Davey Smith G, Ströbele SA, Egger M. Smoking and health promotion in Nazi Germany. *J Epidemiol Community Health* 1994;**48:**220–23.

[14] Proctor R. *The Nazi war on tobacco: ideology, evidence, and possible cancer consequences*. The Johns Hopkins University Press, 1997.

[15] Fisher RA. Statistical methods in genetics. *Heredity* 1952;**6:**1–12.

[16] Thorgeirsson TE, Geller F, Sulem P *et al*. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008;**452:**638–42.

[17] Amos CI, Wu X, Broderick P *et al*. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 2008;**40:**616–22.

[18] Hung RJ, McKay JD, Gaborieau V *et al*. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008;**452:**633–37.

[19] Freathy R, Ring SM, Shields B *et al*. A common genetic variant in the 15q24 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNB4) is associated with a reduced ability of women to quit smoking in pregnancy. *Hum Mol Genet* 2009;**18:**2922–27.

[20] Le Marchand LC, Derby KS, Murphy SE *et al*. Smokers with the CHRNA lung cancer-associated variants are exposed to higher levels of nicotine equivalents and a carcinogenic tobacco-specific nitrosamine. *Cancer Res* 2008;**68:**22.

[21] Davey Smith G, Ebrahim S. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32:**1–22.

[22] Haldane JBS. A defense of beanbag genetics. *Int J Epidemiol* 2008;**37:**435–42.

[23] Haldane JBS. *Science and life*. London: Pemberton Publishing, 1969.

[24] Charman T, Pickles A, Chandler S *et al*. Commentary: Effects of diagnostic thresholds and research versus service and administrative diagnosis on autism prevention. *Int J Epidemiol*.

[25] Rutter M. Commentary: Fact and artefact in the secular increase in the rate of autism. *Int J Epidemiol*.

[26] Hertz-Picciotto. Commentary: Diagnostic change and the increased prevalence of autism. *Int J Epidemiol*.

[27] King M, Bearman P. Diagnostic change and the increased prevalence of autism. *Int J Epidemiol*.

[28] Bresnahan M. Li G, Susser E. Editorial: hidden in plain sight. *Int J Epidemiol*.

[29] Kivimäki M, Gimeno D, Ferrie JE *et al*. Socioeconomic position, pyschosocial work environment and cerebrovascular disease among women: the Finnish public sector study. *Int J Epidemiol*.

[30] Tielsch JM, Katz J, Thulasiraj RD *et al*. Exposure to indoor biomass fuel and tobacco smoke and risk of adverse reproductive outcomes, mortality, respiratory morbidity and growth among newborn infants in south India. *Int J Epidemiol* 2009;**38:**1351–63.

[31] Kyu HH, Georgiades K, Boyle MH. Maternal smoking, biofuel smoke exposure and child height-for-age in seven developing countries. *Int J Epidemiol* 2009;**38:**1342–50.

[32] Nnoaham KE, Sacks G, Rayner M, Mytton O, Gray A. Modelling income group differences in the health and economic impacts of targeted food taxes and subsidies. *Int J Epidemiol* 2009;**38:**1324–33.