Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2009;**38**:1197–1198
doi:10.1093/ije/dyp294

Commentary: Cornfield on cigarette smoking and lung cancer and how to assess causality

Marcel Zwahlen

Accepted 30 July 2009

It has now been 50 years—10 years before the first landing on the moon—that Cornfield and colleagues[1] tried to put to rest the discussion and arguments questioning the role of cigarette smoking in the aetiology of lung cancer. In a 40-page article in the *Journal of the National Cancer Institute*, several of the leading figures in the USA of the developing discipline called 'chronic disease epidemiology' summarized the current state of knowledge that implicated cigarette smoking as the leading factor contributing to the development of lung cancer. This review article did not adhere to current standards for a systematic review. It is nevertheless remarkable in many ways. It reviewed not only the epidemiological evidence ranging from descriptive to analytical epidemiology, but also included studies of carcinogenesis in laboratory animals. It examined, one by one, the arguments against the cigarette–lung cancer hypothesis.

A group of epidemiology students in Bern recently discussed this article as part of their training. They were surprised that half a century ago epidemiologists clearly recognized the various possible biases inherent in descriptive and observational epidemiology and the difficulties in interpreting such evidence. Those arguing against the cigarette–lung cancer hypothesis used several lines of argument.

- Descriptive epidemiology—showing dramatic increases in lung cancer incidence and mortality in the first half of the 20th century—provides an incorrect picture due to time trends in diagnostic accuracy, better recording and changes in the age structure of the populations in the USA and Europe.
- Case–control studies are plagued by selection and information bias, resulting in risk associations that may be spurious.
- Prospective studies can be flawed in ways that also produce spurious associations.
- The majority of smokers do not develop lung cancer, thus refuting a simplistic all-or-none mechanism of disease development.
- Results from animal studies are inconsistent.
- There may be an unmeasured characteristic 'X' that is the real cause of lung cancer but associated with becoming a smoker, thus producing a spurious smoking–lung cancer association.
- In the absence of evidence from controlled experiments in humans, confounding by such a factor 'X' cannot be ruled out.

In taking up these arguments, Cornfield and colleagues implicitly worked through a 'check list' to judge causality of risk factor–disease associations, which is reminiscent of the list put forward by Sir Austin Bradford Hill[2] in 1965, 6 years later. This seems to confirm a rule that for anything named after a person, there is someone else who had said or done it before (fortunately, I cannot remember to whom this rule should be accredited). Cornfield and colleagues, predating Hill, stressed that it is important to see a coherent pattern in population-based incidence rates, consistent findings across populations and study designs and a strong association on the relative risk scale, with a robust dose–response pattern.

Cornfield and colleagues could not truly refute all of the counter arguments. However, they organized their arguments with something we nowadays would recognize as the Bayes factor.[3,4] For each dimension of evidence, the likelihood ratio was informally discussed in which the likelihood of the data under the cigarette–lung cancer hypothesis was compared with the likelihood of the data under the alternative hypothesis. This approach was particularly clear when Cornfield and colleagues dealt with the criticism of Berkson, that the prospective studies on smokers and non-smokers could have given biased results because people who did not smoke were 'biologically self-protective'[5]: 'in this event, during the earlier period after selection, the death rate of the smokers in the study would be higher than the death rate of the non-smokers in the study, even if death rates were unrelated to smoking habits in the general population. If smoking is unrelated to death from lung cancer (or other causes), the death rate of the smokers would tend to equalize with that of the non-smokers as the study progressed'.[1] But this was not what was observed, rather 'the observed association between cigarette smoking and lung cancer was

Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland. E-mail: zwahlen@ispm.unibe.ch

stronger in the latter part than in the early part of the study'.[1]

For Cornfield and colleagues, the most difficult challenge was the argument implicating a confounding variable 'X'. Here, Cornfield and colleagues had to combine arguments about the strength of the association (a relative risk of about nine when comparing smokers with non-smokers and one of a factor 60 when comparing heavy smokers with non-smokers), and how strongly associated this factor 'X' would need to be with smoking to produce these associations. As outlined in Appendix A of their article, boundaries can be derived on how prevalent the factor 'X' would need to be in smokers compared with non-smokers. These boundaries were then used to argue that no one could suggest a candidate for such a characteristic 'X' and the odds of finding one in the future seemed very small indeed. Here, Cornfield and colleagues clearly illustrate why epidemiologists examine relative risk measures when assessing whether an observed association is likely to be causal. Furthermore, we find here an early example of explicitly and quantitatively accounting for an unobserved additional variable. The area of explicitly modelling bias mechanisms or the influence of unobserved (often called latent) variables has gained importance in epidemiology in recent years.[6–8]

In a Bayesian inference framework of continuous learning from the available evidence,[9] one revises the prior odds of a hypothesis (compared with an alternative hypothesis) by multiplying it with the likelihood ratio of the data under the two hypotheses under consideration. When several independent sources of evidence are available, one can multiply the respective likelihood ratios to arrive at a final assessment, taking into account the totality of the evidence. It is hardly surprising that Jerome Cornfield (1912–79), who has been described as one of the most influential biostatisticians in the USA (with little formal training and no degree in statistics),[10] later contributed to the theory of Bayesian inference.[11–13]

No single study or source of evidence was able to put the debate on the association between cigarette smoking and lung cancer at rest. But, as our PhD students quickly realized, this seminal paper integrated the totality of evidence to demonstrate that the cigarette–lung cancer hypothesis was much more likely to be true than any of the alternative hypotheses put forward by the opponents. Of course, our students raised the question when the evidence becomes compelling enough to justify public health action. They were shocked when realizing how slow society can be in taking action. Cornfield and colleagues were well aware of the difficulties of changing a widespread habit that was supported by a powerful industry: '... if the findings had been made on a new agent, ... and on one which did not support a large industry, skilled in the arts of mass persuasion, the evidence for the hazardous nature of the agent would be generally regarded as beyond dispute'.[1] Switzerland is finally introducing smoking bans in public places that are more or less stringent, depending on the Canton. In the Canton of Bern, the smoking ban came into law on 1 July 2009, almost 50 years after the Cornfield paper. But people can still light up in some restaurants, which provide separate smoking areas.

Acknowledgements

I thank the following PhD students and scientists: Martin Adam, Matthias Egger, Martin Feller, Sereina Herzog, Eveline Nüesch, Olivia Keiser, Radek Panczak, Corina Rüegg and Ben Spycher, at the Institute of Social and Preventive Medicine in Bern, Switzerland, for discussing this historical paper during a lunch seminar on 2 July 2009, and providing helpful comments on drafts of this commentary.

Conflicts of interest: None declared.

References

[1] Cornfield J, Haenszel W, Hammond CE *et al*. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst* 1959;**22:**173–203. Reprinted in *Int J Epidemiol* 2009;**38**:1175–91.

[2] Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;**58:**295–300.

[3] Goodman SN. Toward evidence-based medical statistics. 2: the Bayes factor. *Ann Intern Med* 1999;**130:**1005–13.

[4] Goodman SN. Introduction to Bayesian methods I: measuring the strength of evidence. *Clin Trials* 2005;**2:**282–90.

[5] Berkson J. Smoking and lung cancer: some observations on two recent reports. *J Am Stat Assoc* 1958;**53:**28–38.

[6] Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003;**14:**451–8.

[7] Greenland S. Multiple-bias modelling for analysis of observational data. *J R Stat Soc Ser A* 2005;**168:**267–306.

[8] Turner RM, Spiegelhalter DJ, Smith GCS *et al*. Bias modelling in evidence synthesis. *J R Stat Soc Ser A* 2009; **172:**21–47.

[9] Berry DA. Bayesian clinical trials. *Nat Rev Drug Discov* 2006;**5:**27–36.

[10] Armitage P, Colton T. *Encyclopedia of Biostatistics*. West Sussex: John Wiley & Sons, 1998.

[11] Cornfield J. Sequential trials, sequential analysis and the likelihood principle. *Am Stat* 1966;**20:**18–23.

[12] Cornfield J. The Bayesian outlook and its application. *Biometrics* 1969;**25:**617–57.

[13] Cornfield J. Recent methodological contributions to clinical trials. *Am J Epidemiol* 1976;**104:**408–21.