Statistical Methods in Medical Research 2002; **11**: 373–380

The ethics of consulting for the tobacco industry

DB Rubin Department of Statistics, Harvard University, Cambridge, MA, USA

This article describes how and why I became involved in consulting for the tobacco industry. I briefly discuss the four relatively distinct statistical topics that were the primary focus of my work, all of which have been central to my published academic research for over three decades: missing data; causal inference; adjustment for covariates in observational studies; and meta-analysis. To me, it is entirely appropriate to present the application of this academic work in a legal setting.

Introduction

For several years I have consulted for the tobacco industry in the context of their litigation with major plaintiffs (e.g., the Attorneys General of various states and the United States Federal Government). The four topics of my work have been: (1) to describe the proper way to handle missing values in large databases; (2) to formulate the statistically correct way to calculate the damages due to the alleged misconduct of the tobacco industry; (3) to describe how to conduct reliable adjustments for differences between smokers and nonsmokers on background variables when calculating the relative risks of various smoking behaviours; and (4) to summarize the evidence, based on three decades of studies, concerning what sorts of interventions work to curtail smoking, both its initiation and its cessation. I have never produced any specific numbers for damages in any of my expert reports, although I have written a variety of such reports, and have been deposed a number of times (see Appendix A). I testified in one trial—Minnesota (described in reference 1), primarily on two topics: the unreliability of *ad hoc* methods for handling missing data; and the unreliability of relying on regression adjustments to control for many background variables when comparing smokers and nonsmokers.

I have been invited to present my work on these four topics in many academic settings, including: American Statistical Association meetings; Columbia University—Statistics Department; UCLA—Biostatistics Department; Harvard University—Statistics and Health Care Policy Departments; Centers for Disease Control and Prevention; Wharton School—Statistics Department; University of California, Irvine—Statistics Group; and for the American Public Service Health Association's Lowell Reed Lecture. Moreover, I have published aspects of this work in a variety of places.[2–7] I have also used the tobacco example and my work on it in my classes on causal inference at Harvard University.

Address for correspondence: Department of Statistics, Harvard University, Cambridge, MA, 02138, USA. E-mail: rubin@stat.harvard.edu

 10.1191/0962280202sm297ra

Personally, I have experienced essentially no hostility at any of these presentations, although at times there has been substantial hostility toward the tobacco industry, which I do not combat. I am defending the importance of honest and competent statistics, that is all.

Some personal history

When I was first contacted by a tobacco lawyer, I was very reluctant to consult for them, for the standard 'politically correct' reasons, and I feared strong pressure to be dishonest, which there was none throughout. The original topic was simply to comment on the ways the plaintiffs' experts were handling missing data. On examination, their methods were, at best, three decades out of date, and, at worst, entirely silly (e.g., when missing 'marital status', call them 'married').

As I continued to read these initial reports I was appalled that hundreds of billions of dollars could be sought on the basis of such analyses. From a broader perspective, the logic underlying most of the analyses also seemed to me entirely confused. For example, misconduct seemed to play no role in nearly all calculations, and phrases such as 'caused by' or 'attributable to', were used nearly interchangeably and often apparently without thought. Should nearly a trillion dollars in damages be awarded on the basis of faulty logic and bad statistical analyses because we 'know' the defendant is evil and guilty? If the issue were assessing the tobacco industry a trillion dollar fine for lying about its products, I would be amazed but mute. But these reports were using statistical arguments to set the numbers—is it acceptable to use bad statistics to set numbers because we 'know' the defendant is guilty? What sort of precedent does that imply?

An experience from three and a half decades earlier weighed on me. After some evening party in graduate school, I had gone back to my car to find two 'punks' (looked 14) had just broken into it. Having enough alcohol in me to feel relatively fearless, I confronted them, and as we were arguing, a police car drove by and stopped to investigate. While discussions transpired, one of the kids let a switchblade fall to the street. The policeman picked it up, the kid denied having it, etc. Both kids were arrested and I drove home.

A couple of days later, I was called to testify in court against them. Missing classes, I dutifully went to court at 8 am, where the arresting officer wanted to meet with me to 'get our stories straight' before any testifying began. The 'straight' story was that the kids and I were fighting when the police arrived and the bad kid had the switchblade out and was attacking me, but the policeman saved me by taking the dagger away. Well no, it was not that way, I reacted. But, he said, these are real bad kids (grass smoking, glue sniffing, petty thieving, etc.), and with this 'assault with a deadly weapon' charge, we could put them away—get them out of our neighbourhoods where they are a bad influence on other kids, etc. No, I would not do it. I would tell the truth—get them for what they did do; I would not lie and get them for something they did not do just because they are evil. I was firm.

The case did not come before the judge that day. The next day I returned prepared to spend another day away from classes so that I could 'protect' the evil kids. The

policeman again tried to persuade me of the ethics of their position and I again refused to co-operate. The case was again not heard that day.

I never returned and I have always felt that I was being unethical not to be there to testify to what I knew to be true, even if it would be used to defend evil defendants.

How to handle missing data in public-use data sets

As mentioned at the start, the initial issue for my consulting was commenting on the methods used by the plaintiffs' experts to handle missing data in their analyses. These methods to me clearly used obsolete, inappropriate techniques for the imputation of missing data in publicly available data bases (e.g., NMES, the National Medical Expenditure Survey). I had been involved in a variety of projects for US federal agencies concerning multiple imputation of public use data sets for many years (e.g., the Census Bureau and the census,[8] the Internal Revenue Service,[9] the National Center for Health Statistics,[10] National Highway Traffic Safety Administration and its Fatal Accident Reporting System[11]). Although the agency then responsible for NMES (the Agency for Health Care Policy and Research) had expended an enormous amount of effort to impute missing medical costs, it was all singly imputed and much of it was based on a crude 'hot-deck', a method I had been criticizing for more than two decades.[12] Also, I considered what most 'experts' did with the other missing data in NMES to be even less principled than what the Agency had done.

Since I had been advocating the use of principled methods for the analysis of missing data for a quarter of a century (e.g., EM;[13] the standard terminology 'missing at random', etc., was defined in reference 14; I have been developing multiple imputation since 1977[15,16]), I considered it relatively unethical to remain silent on how missing data should be handled in any situation involving such databases and certainly in a situation where nearly a trillion dollars were being sought on the basis of analyses of such data. With respect to the NMES data set, my attempt to implement something substantially more valid than what had been done is described in reference 6. Also I understand that plaintiffs' expert Harrison has done some multiple imputation of NMES and found the results helpful.[17] This situation also stimulated a PhD thesis at Harvard to deal with some of the computational complexities of NMES, i.e., it led to the creation and study of 'nested' multiple imputations.[18]

A framework for estimating the causal effects of the tobacco industry's alleged misconduct

The second topic for my consulting was focused on causal inference, in the sense of trying to estimate the effect of the tobacco industry's misconduct, not the effect of the health risks of smoking, which have been 'known' for at least a century (e.g., cigarettes have been called 'coffin nails' since 1885 and 'cancer sticks' for years,[19] and Nazi Germany under Hitler had a vigorous campaign against cigarettes because of their health risks,[20]). My task was how to formulate and conduct this causal inference correctly, using concepts and terms (such as relative risks, attributable

risks, etc.) accepted by the plaintiffs' experts, the Surgeon General, etc., and already introduced to the court (judge and jury). This formalization was published in references 3 and 5 and was invited to be presented at the Centers for Disease Control and Prevention as well as at the American Association for Public Health. Since I had been working on problems of causal inference in the late 1960s in my PhD thesis work with my advisor, Bill Cochran, and have continued to do so to this day, I found it completely ethical to apply everything I knew to formulate this framework. In fact, Bill, who was a tremendous influence on me, taught me that statistics as a field should pursue truth with full vigour and not be dissuaded by political pressure: pretending that things are true that are false does not help solve problems. This framework specified the tasks that were needed to estimate the effect of the tobacco industry's alleged misconduct, including clarifying the necessary assumptions and the associated demands on data.

Most of the required tasks in this framework involved the estimation of actual world quantities: 'pots' of dollars for medical expenditures; the prevalence of different types of smoking behaviours; and the actual world relative dollar risks of smoking behaviours, all at a very detailed level (e.g., by age, sex, race, etc.). The need for this detail is for the purpose of justifying a key assumption, discussed relatively briefly in references 3 and 5. This assumption is described more completely in reference 21, but in a relatively unpolished form and I hope to have a more careful version of this exposition available as a *Journal of the Royal Statistical Society Series A* publication because I think the critical issues are generally not well understood, or at least well described, in the literature. A final task in this framework involved the need to 'estimate' the counterfactual world prevalences of smoking behaviours.

The way to estimate the actual world relative risks reliably was my third consulting topic and the review of literature relevant to the counterfactual world prevalences was my final topic.

Reliable adjustment for background factors when estimating relative risks

When estimating the relative risks of smoking, all of the plaintiffs' experts attempted to do some adjustment for background differences between smokers and nonsmokers, minimally for sex and at least coarse age, and sometimes as with Harrison,[22] for many factors. As indicated above, adjusting for many such factors is necessary to justify the key assumption, i.e., to believe the relevance of the relative risk calculations to the causal question of health care costs due to the alleged misconduct of the tobacco industry. When adjusting for these factors, all plaintiffs' experts relied on statistical models with linear assumptions at their core, linear assumptions that produce highly unreliable answers if the two groups were distributionally far apart because of the implicit extrapolation. This fact was the topic of my PhD thesis under William Cochran and it was also the reason for the development of matched sampling and more modern extensions such as propensity score methods. Cochran and I wrote about the unreliability of linear regression more than a quarter of a century ago,[23] and Paul Rosenbaum and I (and others, e.g., Neal Thomas) had been writing about the improvements in

the reliability of these adjustments from using matching and propensity scores ever since.[24–28]

Initial analyses showed that, indeed, in the data sets used by plaintiffs' experts (e.g., NMES), smokers and nonsmokers were too far apart to trust the models used by plaintiff's experts to adjust for the differences, according to the standards we have been writing about for nearly three decades. I had no reluctance saying so. I also had no reluctance going on to show that smokers and nonsmokers in NMES were not too far apart to make reliable adjustment impossible.[7]

Counterfactual world prevalences

My final consulting topic was the most speculative: the estimation of the prevalences of smoking (and associated health related outcomes, such as being overweight) in a counterfactual world without the tobacco industry's 'information-based' misconduct. That is, what is the summary of the evidence from the real world concerning the effect of information about the health risks of smoking on smoking initiation and smoking cessation?

Basically, the approach started with the examination of intervention studies in the last decades examining the effect of extra information on smoking initiation and smoking cessation. This was then used to estimate the effect of information-based misconduct by the tobacco industry in previous years. Fortunately, various Surgeon General Reports[29–32] summarized some of this literature, and reports from Centers for Disease Control (CDC)[33] summarized others. Also, reference lists from workers on smoking cessation produced hundreds more, which I reviewed.[34,35] Many health care workers clearly wanted to know what worked, and there were hundreds of carefully controlled randomized experiments studying this issue. The consistent picture that emerges is the following.

First, regarding smoking initiation, which usually takes place in high school: sadly, information about the health risks has no effect whatsoever (e.g., the Hutchinson 2000 15 year randomized experiment).[36] Second, regarding smoking cessation, such information, when delivered at the community level, has no effect on heavy smokers, and a slight effect, if at all (maybe 1% increase in quit rate per year) on light smokers.[37,38] Price matters, banning smoking in the workplace and restaurants reduces it, but information *per se* works only when delivered in a 'targeted' manner at the correct stage of preparation to quit, either by a health care professional (doctor) or a computerized expert system, to certain subgroups of people (e.g., pregnant women, men who have just had a heart attack), and then the effect is modest (typically a few percent increase in quitting rate).

Although I have written on meta-analysis,[39] I have not attempted a formal one of the smoking cessation literature, partially because it is so voluminous, but also because a variety of formal and informal meta-analyses already exist (e.g., in some of the references already cited). Also, in repeated personal conversations with researchers in this area, there seems to be no disagreement about the extremely limited effectiveness of information-based programmes. There is disagreement about the relevance of these results to the estimation of the alleged misinformation from the tobacco industry in the

past on smoking behaviour, but it is the only experimental evidence available on the effect of information on smoking behaviour.

The ethics of testifying

Health care professionals as a group do not appear to be willing to testify voluntarily to affirm what they write in hundreds and hundreds of peer-reviewed publications, as to what works and does not work to curtail smoking. I am not saying that they should—I can understand why they do not wish to. However, I believe my testimony, to statistical views that I have been expressing for three decades, is ethical, and my testifying to the current conclusions, based on reviewing the literature, about my assessment of what has worked and has not worked to curtail smoking, is also ethical.

References

1 Rybak DC, Phelps D. *Smoked: the inside story of the Minnesota tobacco trial.* Minneapolis: MSP Books, 1998, pp 394–5.

2 Frangakis C, Rubin DB. What does it mean to estimate the causal effects of smoking? *Proceedings of the Section on Statistics in Epidemiology of the American Statistical Association*, 1998, pp 18–27.

3 Rubin DB. Statistical issues in the estimation of the causal effects of smoking due to the conduct of the tobacco industry. In: Gastwirth J, ed. *Statistical science in the courtroom.* New York: Springer-Verlag, 2000, pp 321–51.

4 Rubin DB. Statistical assumptions in the estimation of the causal effects of smoking due to the conduct of the tobacco industry. In: Blasius J, Hox J, de Leeuw E, Schmidt P, eds. *Proceedings of the Fifth International Conference on Logic and Methodology*, 6 October 2000, Cologne, Germany, 1–22.

5 Rubin DB. Estimating the causal effects of smoking. *Statistics in Medicine* 2001; **20**: 1395–414.

6 Rubin DB. Multiple imputation of NMES. *Proceedings of the International Conference on Quality in Official Statistics.* Stockholm, Sweden (May 14–15, 2001), 2002, CD 24.3.

7 Rubin DB. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services Outcome Research Methodology*, 2001; **2**: 169–88.

8 Clogg CC, Schenker N, Schultz B, Weidman L, Rubin DB. Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression. *Journal of the American Statistical Association* 1991; **86**: 68–78.

9 Czajka JC, Hirabayashi SM, Little RJA, Rubin DB. Projecting from advance data using propensity modeling. *Journal of Business and Economic Statistics* 1992; **10**: 117–31.

10 Ezzati-Rice T, Johnson W, Khare M, Little R, Rubin D, Schafer J. A simulation study to evaluate the performance of model-based multiple-imputations in NCHS health examination surveys. *Proceedings of Bureau of the Census Eleventh Annual Research Conference* 1995; 257–66.

11 US Department of Transportation, NHTSA. Multiple imputation of missing blood alcohol content (BAC) in FARS. Research note. National Highway Traffic Safety Administration, 1998.

12 Madow WG, Olkin I, Rubin DB. *Incomplete data in sample surveys (volume 2): theory and bibliographies.* New York: Academic Press, 1983.

13 Dempster AP, Laird N, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 1977; **39**: 1–38.

14 Rubin DB. Inference and missing data. *Biometrika* 1976; **63**: 581–92.

15 Rubin DB. The design of a general and flexible system for handling non-response in sample surveys. Unpublished manuscript, 1977.
16 Rubin DB. Multiple imputations in sample surveys—a phenomenological Bayesian approach to nonresponse. *The Proceedings of the Survey Research Methods Section of the American Statistical Association* 1978: 20–34. Also In: *Imputation and editing of faulty or missing survey data.* Washington DC: US Department of Commerce, 1978: 1–23.
17 Harrison GW. Communication at public presentation. Harvard School of Public Health, December 14, 2001.
18 Shen Z. Nested Multiple Imputation. Cambridge, MA: Harvard University Department of Statistics, PhD thesis, 2000.
19 http://detnews.com/2000/religion/0004/24/04220016.htm
20 Proctor RN. Bitter pill. *The Sciences* 1999; **39**: 14–19.
21 Rubin DB. Statistical inference for causal effects in epidemiological studies via potential outcomes. *Proceedings of the XL Scientific Meeting of the Italian Statistical Society* 2000, 419–30.
22 Harrison GW. Expert report, April 27, 1998: Health care expenditures attributable to smoking in Oklahoma. The State of Oklahoma, ex.rel., *et al.*, plaintiffs, versus Reynolds Tobacco Co., *et al.*, defendants 1998. Case no. CJ-96-1499-L, District Court of Cleveland County, Oklahoma.
23 Cochran WG, Rubin DB. Controlling bias in observational studies: a review. *Sankhya-A* 1973; **35**: 417–46.
24 Rubin DB. Matching to remove bias in observational studies. *Biometrics* 1973; **29**: 159–83.
25 Rubin DB. The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics* 1973; **29**: 184–203.
26 Rosenbaum P, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**: 41–55.
27 Rubin DB, Thomas N. Affinely invariant matching methods with ellipsoidal distributions. *Annals of Statistics* 1992; **20**: 1079–93.
28 Rubin DB, Thomas N. Characterizing the effect of matching using linear propensity score methods with normal covariates. *Biometrika* 1992; **79**: 797–809.
29 US Department of Health and Human Services, Public Health Service, Office of the Surgeon General, Publication No. 1103. Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service 1964. Washington DC: US Government Printing Office.
30 US Department of Health, Education, and Welfare, Public Health Service, Office of the Surgeon General, Publication No. 79-50066. Smoking and Health: A Report of the Surgeon General 1979. Washington DC: US Government Printing Office.
31 US Department of Health and Human Services, Public Health Service, Office of the Surgeon General. Preventing Tobacco Use Among Young People: A Report of the Surgeon General, Centers for Disease Control and Prevention 1994. Washington DC: US Government Printing Office.
32 US Department of Health and Human Services. Reducing tobacco use: a report of the Surgeon General. Atlanta, GA: 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, 2000.
33 Fiore *et al.* Smoking cessation. Clinical Practice Guidelines, No. 18, USDHHS, Public Health Service, Agency for Health Care Policy and Research, Centers for Disease Control and Prevention, AHCPR Publication No. 96-0692, April 1996.
34 Strecher VJ. Computer-tailored smoking cessation materials: a review and discussion. *Patient Education and Counseling* 1999; **36**: 107–17.
35 Velicer WF *et al.* An expert system intervention for smoking cessation. *Addictive Behaviors* 1993; **18**: 269–90.
36 Peterson AV, Kealey KA, Mann SL, Marek PM, Sarason IG. Hutchinson smoking prevention project: long-term randomized trial in school-based tobacco use prevention—results on smoking. *Journal of the National Cancer Institute* 2000; **92**: 1979–91.
37 COMMIT Research Group. Community Intervention Trial for Smoking Cessation (COMMIT): I. cohort results from a four-year community intervention 1995. *American Journal of Public Health* 1995; **85**: 183–92.

38 COMMIT Research Group. Community Intervention Trial for Smoking Cessation (COMMIT): II. changes in adult cigarette smoking prevalence. *American Journal of Public Health* 1995; **85**: 193–200.

39 Rubin DB. A new perspective on meta analysis. In: Watcher KW, Straf ML, eds. *The future of meta analysis.* Washington DC: Russell Sage/NAS, 1990, pp 155–65.

Appendix A

Testimony of Donald B Rubin re tobacco litigation

Trial

1 1, May 1998. *State of Minnesota v. Philip Morris, Inc.*, et al.

Deposition

1 9, July 1997. *Florida v. The America Tobacco Co.*, et al.
2 25 and 29, September 1997. *Texas v. The American Tobacco Co.*, et al.
3 6 and 7, October 1997 and April 13, 1998. *Minnesota v. Philip Morris, Inc.*, et al.
4 19, August 1998. *Washington v. The American Tobacco Co.*, et al.
5 28, September 1998. *Northwest Laborers–Employers Health & Security Trust Fund*, et al. *v. Philip Morris, Inc.*, et al.
6 27, January 1999. *Iron Workers Local No. 17 Insurance Fund v. Philip Morris, Inc.*, et al.
7 2, July 1999. *Northwest Laborers–Employers Health & Security Trust Fund*, et al. v. *Philip Morris, Inc.*, et al.
8 25, June 2000. *Blue Cross and Blue Shield of New Jersey, Inc.*, et al. *v. Philip Morris, Inc.*, et al.
9 20, May 2002. *Devin Daniels* et al. *v. Philip Morris, Inc.*, et al.
10 17 and 18, July 2000. *United States of America v. Philip Morris, Inc.*, et al.