



Overview of the Epidemiology Methods and Applications: Strengths and Limitations of Observational Study Designs

GRAHAM A. COLDITZ

School of Medicine, Department of Surgery, Washington University School of Medicine, St. Louis, Missouri

The impact of study design on the results of medical research has long been an area of both substantial debate and a smaller body of empirical research. Examples come from many disciplines within clinical and public health research. Among the early major contributions in the 1970s was work by Mosteller and colleagues (Gilbert et al., 1997), who noted that innovations in surgery and anesthesia showed greater gains than standard therapy when nonrandomized, controlled trials were evaluated compared with the gains reported in randomized, controlled trials. More recently, we and others have evaluated the impact of design in medical and surgical research, and concluded that the mean gain comparing new therapies to established therapies was biased by study design in nonrandomized trials (Colditz et al., 1989; Miller et al., 1989). Benson and Hartz (2000) conducted a study in which they focused only on studies reported after 1985. On the basis of 136 reports of 19 diverse treatments, Benson and Hartz concluded that in only 2 of the 19 analyses did the combined data from the observational studies lie outside the 95% confidence interval for the combined data from the randomized trials. A similar study drew only on data reported from 1991 to 1995, which showed remarkably similar results among observational studies and randomized, controlled trials (Concato et al., 2000). These more recent data suggest that advancing the study design and analytic methods may reduce bias in some evaluations of medical and public health interventions. Such methods apply not only to the original studies, but also to the approaches that are taken to quantitatively combine results by using meta-analytic approaches such as random effects meta-regression, Bayesian meta-analysis, and the like (Normand, 1999). By focusing attention on thorough data analysis, design issues can be understood and their impact or bias can be estimated, on average, and then ideally accounted for in the interpretation of data.

Before discussing dietary data, let us first consider some of the more clearly delineated preventive exposures. Issues of study design have been addressed in terms of combining randomized trials and observational studies in evaluating preventive interventions such as *Bacillus Calmette-Guérin* vaccination (Colditz et al., 1994) and mammography screening (Desmissie et al., 1998). When one is interpreting the apparent heterogeneity in the results, it is important to step back and ask what is the relationship being evaluated under these different study designs? For example, a randomized, controlled trial uses the intention-to-treat analysis to preserve the merit of randomization. Such an analysis does not evaluate the exposure-disease relationship, but rather examines the impact of offering a new therapy versus an alternative therapy (regardless of adherence to the intervention, or control or placebo). On the other hand, a case-control study or a prospective cohort study will evaluate the impact of the screening test among those participants who were screened as compared with those who were never screened. In prevention studies, the design raises major issues of the timing of the exposure in the natural history of disease and also the adherence to therapy by healthy research volunteers. Case-control studies of preventive interventions such as screening mammography and prospective population-based studies of pap smears have capitalized on this variation in time since the last screen to evaluate the protective interval for a screening test (IARC Work Group, 1986). In contrast, a trial must choose a level of exposure, such as annual mammography screenings or colon screenings every 10 years with a colonoscopy, regardless of the evolving evidence on the duration of protection after a negative screening test.

Continuing with the mammography example, a detailed study by Demissie and colleagues (1998) combined data from seven randomized trials and six case-control studies that investigated the association between participation in breast cancer screening

programs and breast cancer mortality. The authors showed that if one assumes noncompliance with mammography (approaching 30%) and 20% of the control group is screened, then the benefit of mammography in terms of reduced mortality is comparable in randomized, controlled trials and epidemiologic studies after adjusting for nonadherence (Demissie et al., 1998). Thus, the different designs fundamentally measure different constructs of the impact of screening.

Zelen (1988) considered the challenges of primary prevention trials and addressed both compliance and models of carcinogenesis as major impediments to the use of randomized, controlled trials to evaluate cancer prevention strategies. It is important to contrast these issues in both treatment trials and prevention trials. In treatment trials, recently diagnosed patients, who are often in life-threatening situations, are typically offered the option to participate in a trial of a new therapy compared with standard therapy or placebo. Compliance or adherence to therapy is usually very high among these highly motivated patients and their outcomes are generally in the short- to mid-term. In contrast, prevention trials recruit large numbers of healthy participants, offer them a therapy, and then follow them over many years because the chronic diseases being prevented are relatively rare. With substantial noncompliance (often in the range of 20% to 40% over the duration of the trial), an intention-to-treat analysis is no longer unbiased, but rather gives a biased estimate of the effect, typically underestimating the magnitude of the association that is seen in observation studies in which those participants who have had exposure to a particular lifestyle component are compared with those without such an exposure.

There are additional challenges for nutritional interventions, including the timing of diet as a preventive agent in the disease process and the range of nutrient intakes in the population. In retrospective case-control studies, recall bias of past diet is an additional issue with which to contend. Unlike smoking or screening tests in which the exposure is finite and can be completely stopped and started, one's diet, physical activity, and weight change cannot go to zero for prolonged periods and sustain life. The range of nutrient intake is a major issue when enrolling participants into prevention trials and observational studies. Health-conscious volunteers are more often identified and screened as eligible for a trial.

The epidemiology of diet and colon cancer has been extensively studied. For example, Cho et al. (2004) conducted a combined analysis of prospective dietary studies of calcium and vitamin D intake data from 10 cohorts. The dose-response relationship for calcium showed that the greatest benefit for increasing calcium intake was for those participants who had a reported daily intake below 1000 mg/d. Increasing the intake of those individuals with low intake to the level of 1000 mg/day would yield a 20% reduction in risk. Beyond this level of intake, there was little additional reduction in the risk of colon cancer. In the Women's Health Initiative, participants had a mean calcium intake of 1150 mg/d at baseline and increased this intake in the intervention arm to 2250 mg/d on average. This magnitude of increase was of limited association in the combined, prospective, cohort studies and was not related to risk in the randomized trial (Wactawski-Wende et al., 2006). Similar findings apply to the interpretation of the vitamin D intervention and highlight the role of dietary intake at randomization when evaluating dietary components.

Returning to the time frame of exposure in the carcinogenic process, the null randomized, controlled trials of fiber (Alberts et al., 2000) and fruit and vegetables (Schatzkin et al., 2000) for the prevention of polyp recurrence amply illustrate Zelen's concerns regarding the timing of the preventive intervention in the disease process. The extent of DNA damage accumulated across the colonic mucosa at the time of detecting the "eligibility polyp" was certainly not limited only to the removed polyp. Rather, these observations beg the question that at what stage in the disease process may fiber play a role in protecting against colon cancer? This contrasts with the richness of epidemiologic studies that can address exposure over the life course and relate such exposure to disease risk. Perhaps the best-known example is the radiation follow-up effects cohort in Japan in which a radiation dose was estimated for each woman who had been exposed to the effects of the atomic bombs on Hiroshima and Nagasaki and followed-up over 40 years. The results of this study showed a clear and strong relationship between the increased risk of breast cancer with higher exposure among those participants who were exposed before the age of 20 years (Land et al., 2003).

Retrospective assessment of diet after disease diagnosis has been demonstrated to introduce bias into the evaluation of exposure-disease relationships. For example, Giovannucci et al. (1993) evaluated retrospective recall of diet after breast cancer diagnosis within the ongoing Nurses' Health Study. In contrast with the prospective analysis in which no relationship between dietary fat and breast cancer was observed, the retrospective analysis yielded a positive relationship for total fat and saturated fat (Giovannucci et al., 1993). By comparing the top quintile versus the bottom quintile of reported intake, the retrospective assessment yielded odds ratios of 1.43 for total fat and 1.38 for saturated fat. Therefore, the magnitude of bias was sufficient to distort evaluation of the diet-disease relationship.

Prospective dietary data, as collected in numerous cohort studies that contribute to the pooled analysis of diet and chronic disease, have the added potential of allowing for consistent approaches to error correction across the participating cohorts (Smith-Warner et al., 2006). An additional advantage of prospective cohort studies is the potential to obtain unbiased recall from participants who are free from disease to "fill in" their dietary history from earlier in life. This approach, implemented within the Nurses' Health Studies, has allowed the assessment of the high school diet, and validation studies have shown that such adult recall provides a reasonably valid measure of the adolescent diet (Maruti et al., 2005; Maruti et al., 2006). Given the temporal constraints, such recall is necessary to evaluate dietary intake in early life in relationship to cancer end points.

Overall, prospective cohort studies offer several advantages for the study of the relationship between diet and disease, particularly when validation studies allow for the consideration of error correction (Rosner et al., 1989; Rosner and Gore, 2001). By using repeated measures, one can address the timing of exposure in the disease process to determine the mechanism of disease and prevention. The dose of exposure varies according to current population practices, and changes in diet over time can be related to disease risk, just as the time since cessation of smoking can be related to disease mortality when one has repeated measures of cigarette smoking (Kenfield et al., 2008). Although trials for prevention of chronic disease have merit, their interpretation must be cautioned by the level of participation and the potential that they are addressing an intervention at the wrong point in the carcinogenic process.

REFERENCES

- Alberts, D. S., Martinez, M. E., Roe, D. J., Guillén-Rodríguez, J. M., Marshall, J. R., van Leeuwen, J. B., Reid, M. E., Ritenbaugh, C., Vargas, P. A., Bhattacharyya, A. B., Earnest, D. L., and Sampliner, R. E. (2000). Lack of effect of a high-fiber cereal supplement on the recurrence of colorectal adenomas. Phoenix Colon Cancer Prevention Physicians' Network. *N. Engl. J. Med.* **342**: 1156–1162.
- Benson, K., and Hartz, A. J. (2000). A comparison of observational studies and randomized, controlled trials. *N. Engl. J. Med.* **342**: 1878–1886.
- Cho E., Smith-Warner S.A., Spiegelman D., Beeson, W. L., van den Brandt, P. A., Colditz, G. A., Folsom, A. R., Fraser, G. E., Freudenheim, J. L., Giovannucci, E., Goldbohm, A., Graham, S., Miller, A. B., Pietinen, P., Potter, J. D., Rohan, T. E., Terry, P., Toniolo, P., Virtanen, M. J., Willett, W. C., Wolk, A., Wu, K., Yuan, S.-S., Zeleniuch-Jacquotte, A., and Hunter, D. J. (2004). Dairy foods, calcium, and colorectal cancer: a pooled analysis of 10 cohort studies. *J. Natl. Cancer Inst.* **96**: 1015–1022.
- Colditz, G., Miller, J., and Mosteller, F. (1989). How study design affects outcomes in comparisons of therapy. I. Medical. *Stat. Med.* **8**: 441–454.
- Colditz, G. A., Brewer, T. F., Berkey, C. S., Wilson, M. E., Burdick, E., Fineberg, H. V., and Mosteller, F. (1994). The efficacy of BCG vaccination in the prevention of tuberculosis: meta-analysis of the published literature. *JAMA.* **271**: 698–702.
- Concato, J., Shah, N., and Horwitz, R. I. (2000). Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N. Engl. J. Med.* **342**: 1887–1892.
- Demissie, K., Mills, O. F., and Rhoads, G. G. (1998). Empirical comparison of the results of randomized controlled trials and case-control studies in evaluating the effectiveness of screening mammography. *J. Clin. Epidemiol.* **51**: 81–91.
- Gilbert, J., McPeck, B., and Mosteller, F. (1977). Progress in surgery and anesthesia: benefits and risks of innovative therapy. **In**: Cost Risks Benefits of Surgery. Bunker, J., Barnes, B., and Mosteller, F., Eds., Oxford University Press, Oxford. UK.
- Giovannucci, E., Stampfer, M. J., Colditz, G. A., Manson, J. E., Rosner, B. A., Longnecker, M., Speizer, F. E., and Willett, W. C. (1993). A comparison of prospective and retrospective assessments of diet in the study of breast cancer. *Am. J. Epidemiol.* **137**: 502–511.
- IARC Work Group. (1986). Screening for squamous cervical cancer: duration of low risk after negative results of cervical cytology and its implication for screening policies. IARC Working Group on evaluation of cervical cancer screening programmes. *Br. Med. J. (Clin. Res. Ed.)*. **293**: 659–664.
- Kenfield, S. A., Stampfer, M. J., Rosner, B. A., and Colditz, G. A. (2008). Smoking and smoking cessation in relation to mortality in women. *JAMA.* **299**: 2037–2047.
- Land, C. E., Tokunaga, M., Koyama, K., Soda, M., Preston, D. L., Nishimori, I., Tokuoka, S. (2003). Incidence of female breast cancer among atomic bomb survivors, Hiroshima and Nagasaki, 1950–1990. *Radiat. Res.* **160**: 707–717.
- Maruti, S. S., Feskanich, D., Colditz, G. A., Frazier, A. L., Sampson, L. A., Michels, K. B., Hunter, D. J., Spiegelman, D., and Willett, W. C. (2005). Adult recall of adolescent diet: reproducibility and comparison with maternal reporting. *Am. J. Epidemiol.* **161**: 89–97.
- Maruti, S. S., Feskanich, D., Rockett, H. R., Colditz, G. A., Sampson, L. A., and Willett, W. C. (2006). Validation of adolescent diet recalled by adults. *Epidemiology.* **17**: 226–229.
- Miller, J., Colditz, G., and Mosteller, F. (1989). How study design affects outcomes in comparisons of therapy. II. Surgical. *Stat. Med.* **8**: 455–466.
- Normand, S. (1999). Meta-analysis: formulating, evaluating, combining and reporting. *Stat. Med.* **18**: 321–359.
- Rosner, B., Willett, W. C., and Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat. Med.* **8**: 1051–1069.
- Rosner, B., and Gore, R. (2001). Measurement error correction in nutritional epidemiology based on individual foods, with application to the relation of diet to breast cancer. *Am. J. Epidemiol.* **154**: 827–835.
- Schatzkin, A., Lanza, E., Corle, D., Lance, P., Iber, F., Caan, B., Shike, M., Weissfeld, J., Burt, R., Cooper, M. R., Kikendall, J. W., and Cahill, J. (2000). Lack of effect of a low-fat, high-fiber diet on the recurrence of colorectal adenomas. Polyp Prevention Trial Study Group. *N. Engl. J. Med.* **342**: 1149–1155.
- Smith-Warner, S. A., Spiegelman, D., Ritz, J., Albanes, D., Beeson, W. L., Bernstein, L., Berrino, F., van den Brandt, P. A., Buring, J. E., Cho, E., Colditz, G. A., Folsom, A. R., Freudenheim, J. L., Giovannucci, E., Goldbohm, R. A., Graham, S., Harnack, L., Horn-Ross, P. L., Krogh, V., Leitzmann, M. F., McCullough, M. L., Miller, A. B., Rodriguez, C., Rohan, T. E., Schatzkin, A., Shore, R., Virtanen, M., Willett, W. C., Wolk, A., Zeleniuch-Jacquotte, A., Zhang, S. M., and Hunter, D. J. (2006). Methods for pooling results of epidemiologic studies: the Pooling Project of Prospective Studies of Diet and Cancer. *Am. J. Epidemiol.* **163**: 1053–1064.
- Wactawski-Wende, J., Kotchen, J. M., Anderson, G. L., Assaf, A. R., Brunner, R. L., O'Sullivan, M. J., Margolis, K. L., Ockene, J. K., Phillips, L., Pottern, L., Prentice, R. L., Robbins, J., Rohan, T. E., Sarto, G. E., Sharma, S., Stefanick, M. L., Van Horn, L., Wallace, R. B., Whitlock, E., Bassford, T., Beresford, S. A., Black, H. R., Bonds, D. E., Brzyski, R. G., Caan, B., Chlebowski, R. T., Cochrane, B., Garland, C., Gass, M., Hays, J., Heiss, G., Hendrix, S. L., Howard, B. V., Hsia, J., Hubbell, F. A., Jackson, R. D., Johnson, K. C., Judd, H., Kooperberg, C. L., Kuller, L. H., LaCroix, A. Z., Lane, D. S., Langer, R. D., Lasser, N. L., Lewis, C. E., Limacher, M. C., Manson, J. E. (2006). Calcium plus vitamin D supplementation and the risk of colorectal cancer. *N. Engl. J. Med.* **354**: 684–696.
- Zelen, M. (1988). Are primary cancer prevention trials feasible? *J. Natl. Cancer Inst.* **80**: 1442–1444.