Invited Commentary: Evolution of Epidemiologic Evidence on Magnetic Fields and Childhood Cancers

Charles Poole

The initial reports in this issue of two new studies of childhood brain cancer and 60-Hz magnetic fields (1, 2) illustrate a few general observations on evolving research in this area.

Epidemiologic assessment of magnetic field exposures is still primitive.

Magnetic field exposures come from many different sources. We should not be proud that we still study them one by one. Appliance-specific relative risks (one for blow dryers, one for curling irons, etc.) are the most glaring example. Time-activity data and magnetic flux density measurements are needed to quantify exposures from individual appliances. At a minimum, this information could be used to stabilize ensembles of appliance-specific relative risk estimates (3–6). Ideally, appliance exposures would be consolidated and combined with measures of exposure from power lines and other sources. Developing a program of combined-source magnetic field exposure assessment tailored for use in epidemiologic studies would be a challenging, but technically feasible, task. No methodological need is greater in this area of research.

Responding to a private exhortation to work toward a multiappliance exposure measure, an epidemiologist once expressed to me a reluctance to “combine apples and oranges.” The simile was ironic. Combining fruits is exactly what is done in the more mature field of nutritional epidemiology to measure consumption of fructose and other nutrients.

For power-line exposures, it has become clear that “wire codes,” i.e., magnetic field exposure measures based on characteristics of lines and their distances from residences, have substantial advantages over short-term flux density measurements in homes. The measurements, often made years after the case or control actually lived in the residence, are obviously limited by their temporal variability. Independently severe is the problem of missing data. Whether by design (1) or because of nonresponse (7), it is not uncommon for magnetic field measurements to be obtained for fewer than 50 percent of cases and controls. Relative risk estimates based on such small numbers of subjects are of highly limited value, whether viewed in isolation or in contrast to estimates based on wire codes.

The in-home measurements should not be abandoned entirely, however. They have considerable value when used in conjunction with wire codes rather than as alternatives to them. Specifically, the measurements can provide exposure scores for wire code categories in dose-response analyses (8). Conducting such an analysis forces one to confront the fact that not all wire code categories appear to differ in exposure. If any principle of dose-response analysis is beyond dispute, it is that response should not vary when dose does not vary. It is therefore necessary to combine some of the wire code categories. The categories to combine may change from locale to locale and over time within a locale.

In the report of their Los Angeles County study, Preston-Martin et al. (1) laudably focus on this issue. The authors present the case and control counts in each of the five wire code categories, but combine the second and third categories (“very low” and “ordinary low”) for data analysis. Only an anomalistically elevated relative risk estimate in the lowest (“underground”) category prevented them from combining the first three categories. Alternative analytic approaches could be defended, however. In the extreme, some of the more exposure-relevant measurements, e.g., median 24-hour child’s bedroom values of 0.5, 0.6, 0.4, 0.6, and 1.1 mG, might have supported a decision to combine all categories but the highest.

Unfortunately, the paper by Gurney et al. (2) does not provide quantitative information on magnetic flux density levels within wire code categories in western Washington State. Without such information, one cannot tell how strongly to interpret the resolutely null results the authors report for childhood brain cancer across the five wire code categories in that study.

Received for publication August 2, 1995, and in final form August 25, 1995.

Department of Epidemiology and Biostatistics, Boston University School of Public Health, Boston, MA.

Reprint requests to Dr. Charles Poole, Harvard Program in Refugee Trauma, 8 Story Street, Third Floor, Cambridge, MA 02138.
Control-selection biases are still potentially important, but are still poorly understood.

The principal biases pertain to the use of random digit dialing and to selecting controls “nonconcurrently,” i.e., after the cases were diagnosed (9). Both studies in this issue (1, 2) used random digit dialing. The problem with this control-selection method is primarily one of analytic control of variables such as parental income, education, and occupational class, since control groups selected in this way are suspected to be of spuriously high socioeconomic level (9–11). The magnitude and even the direction of any uncontrolled bias might vary from one study to another (9). Review of studies using random digit dialing should include careful assessment of the adequacy with which indicators of socioeconomic status are measured and controlled.

Bias due to nonconcurrent control selection is a somewhat more complicated matter. The Los Angeles study (1) was partially nonconcurrent; the cases were diagnosed in 1984–1991, but enrollment of controls did not begin until 1989. To the authors, this design feature explains the elevated relative risk estimate in the underground wire code category. Unfortunately, the report of the Washington State study (2) does not provide the basic information on the time lag, if any, between case diagnosis and control selection.

Two clarifying points may be helpful. The first is that bias would arise from nonconcurrent control selection only because potential controls in one or more wire code categories, e.g., the underground category in the Los Angeles study (1), died or moved away from the study area in greater proportions than did those in the other categories between the time of case diagnosis and the time of control selection. In-migration, e.g., of Latinos into Los Angeles County (1), during this interval is not of concern if children arriving after the diagnosis date are properly excluded.

The second point is that this hypothetical bias is empirically testable. Far from being an all-or-none phenomenon, nonconcurrent control selection has degrees. The more time that elapses between case diagnosis and control selection, the greater the bias will be. Thus, for the partially nonconcurrent Los Angeles study (1), the authors’ bias hypothesis predicts a specific shape for the time trend in relative risk estimates for the underground wire code category: a distinct ramp downward in 1984–1988 to the vicinity of the null value, followed by a relatively flat surface in 1989–1991. Flexible trend analysis methods (12) can be used, even with a limited number of subjects, to see how closely the results conform to this prediction. A simpler but similar analysis (13) firmly refuted, in my opinion, the hypothesis (7, 9, 14) that nonconcurrent control selection in a previous study (7) produced an upward gradient of leukemia relative risk estimates that closely tracked the median magnetic flux densities in the wire code categories (8).

Because of their efficiencies of cost and time, random digit dialing and nonconcurrent control selection are very popular methods in modern case-control research. With creditable candor, Preston-Martin et al. relate their original view that nonconcurrent control selection would be “justified in order to include a sufficient number of cases” (1, p. 111). The investigators’ subsequent experience should make others think twice about purchasing data quantity at the price of nettlesome concerns about data quality.

Acceptance of the null hypothesis for magnetic fields and childhood brain cancer would not strengthen the causal hypothesis for childhood leukemia and might weaken it.

Not long ago, a consensus seemed to be forming that, among children, brain cancer was the most likely cancer to be related to magnetic field exposure. In 1992, a panel called the epidemiologic evidence “less weak for brain cancer than for leukaemia” (15, p. 130). In the same year, another reviewer concluded that “[a] single finding in this extensive group of studies points to a possible causal association, namely, the increase in risk of central nervous system cancer in childhood...The leukemias may not be a fruitful area of study” (16, p. A-21).

Times have changed. The two studies in this issue (1, 2) and, in the aggregate (17), three recent Nordic studies (18–20) are liable to be pigeonholed as “negative” for brain cancer. With appeal to the ancestral criterion of specificity of effect, some observers might see the literature as beginning to point more specifically toward leukemia. Thus, by the peculiar logic of this causal criterion, “negative” evidence for brain cancer may become transmuted into “positive” evidence for leukemia.

The problem with this interpretation is that two key studies, by Wertheimer and Leeper (21) and by Savitz et al. (7), without which the case for a leukemia effect would not be much of a case at all, were “positive” for leukemia and for brain cancer. If one were to accept the null hypothesis for brain cancer on the basis of the more recent results, one would be obliged to conclude that both of these earlier studies were “false positives” for brain cancer but “true positives” for leukemia. Many possible explanations for “false-positive” results, e.g., certain biases or random control-sampling error, would have been nonspecific and thus would have affected the results for both diseases. As a consequence, without a tenable hypothesis of upward bias specific to brain cancer in the...
earlier studies, the recent “negative” results for brain cancer actually weaken the overall case for a leukemia effect.

Publication lies much closer to the beginning than to the end of peer review.

This fact of scientific life is not peculiar to research on childhood cancers and magnetic fields, nor even to epidemiology. Unfortunately, like other facts of life, it tends to be denied.

Who treats publication as the end of peer review? The usual suspects include journalists (22), of course, and their symbiotically codependent flacks at journals, funding sources, and universities (23), who all conspire to bestow Warhol’s predicted 15 minutes of fame onto newly published reports. The antijunk science movement (24) may be partly to blame for promoting peer-review publication as an imperative of quality, entitling one to defend scientific work not only on its merits but on its mere appearance in such a forum (25, 26).

As authors, however, we must shoulder much of the responsibility ourselves. Standing up to take part in the critical discussion of our published work is seldom pleasurable and seems like inefficient careerism. We all have new proposals to write, new data to collect, new papers to publish. Moreover, many of us seem to have a great deal of difficulty just saying “no comment” or something ambivalent to the sensationalizing press, when “no comment” or something ambivalent would be the most constructive or accurate thing to say. Thus, we too often fail to anticipate or resist the varied pressures toward taking a defensive stance about our published work.

In reality, the postpublication phase is the most important phase of peer review. Publication merely signifies that one’s work is deemed worthy of widespread critical scrutiny. Society would be far better served if it were persuaded to view publication not as a claim to knowledge, but simply as an open invitation to inspect one’s results and to consider one’s preliminary interpretation. Funding sources would do a great service by providing financial support for investigators to participate fully in the postpublication peer review of their own work.

The process is aided greatly when conscientious and self-critical authors like Preston-Martin et al. (1) leave relatively few important questions about methods and results unanswered in crafting their initial published reports. Nevertheless, the history of science, distant and near, reveals the wisdom in Bartley’s wistful plea that “it could only do good if every published manuscript were prominently marked ‘Damaged Goods’” (27, p. 450).

Until today, only the authors themselves and a small group of privileged observers have had the opportunity to develop preliminary ideas about the ultimate contributions these two newest studies (1, 2) will make to the epidemiology of magnetic fields and brain cancer in children. Today, the number of peer reviewers increases by orders of magnitude. As the most crucial phase of peer review gets under way, we shall learn more about, and from, these important investigations.

REFERENCES

19. Olsen JH, Nielsen A, Schulgen G. Residence near high volt-