HEI COMMUNICATIONS

Environmental Epidemiology Planning Project

Health Effects Institute

Number 3 March 1994

HEI HEALTH EFFECTS INSTITUTE

The Health Effects Institute, established in 1980, is an independent and unbiased source of information on the health effects of motor vehicle emissions. HEI studies all major pollutants, including regulated pollutants (such as carbon monoxide, ozone, nitrogen dioxide, and particulate materials), and unregulated pollutants (such as diesel engine exhaust, methanol, and aldehydes). To date, HEI has supported more than 120 projects at institutions in North America and Europe.

Typically, HEI receives half its funds from the Environmental Protection Agency and half from 28 manufacturers and marketers of motor vehicles and engines in the United States. Occasionally, grants from other public or private organizations support special projects such as this project. However, in all cases the Institute exercises complete autonomy in setting its research priorities and in disbursing its funds. An independent Board of Directors governs HEI. The Institute's Research Committee and the Review Committee serve complementary scientific purposes and draw distinguished scientists as members. The results of HEI-funded studies are made available as Research Reports, which contain both the investigator's report and the Review Committee's evaluation of the work's scientific and regulatory relevance.

Environmental Epidemiology Planning Project

HEI conducted the Environmental Epidemiology Planning Project in order to identify research needs and opportunities in selected areas of environmental epidemiology. Working groups in each selected area prepared documents composed of individually authored papers. The Planning Project documents were originally published in *Environmental Health Perspectives* (December 1993, vol. 102) a publication of the National Institute of Environmental Health Sciences.

HEI Epidemiology Planning Project Steering Committee	Leon Gordis *Chairman*, Curtis Harris, Millicent Higgins, Richard Remington (deceased), John Tukey, Arthur Upton, Mark Utell
Working Group Leaders	*Methodologic Issues* Kenneth J. Rothman; *Electric and Magnetic Fields* David A. Savitz; *Indoor Air and Other Complex Mixtures* Jonathan M. Samet and Frank Speizer; *Tropospheric Ozone* Ira Tager
Project Coordinator	Aaron J. Cohen
Contributing Staff	L. Virgi Hepner, Noreen S. Manzo, Kathleen M. Nauss, Andrew Sivak, Jane Warren
HEI Epidemiology Planning Project Sponsors	United States Environmental Protection Agency, American Petroleum Institute, Chemical Manufacturers Association, Engine Manufacturers Association, Electric Power Research Institute, Gas Research Institute, Motor Vehicle Manufacturers Association

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Agreement X-817463 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement should be inferred. The contents of this document, also, may not reflect the views and policies of the private sponsors of HEI, and no endorsement by them should be inferred. As indicated in the Introduction, this document has undergone peer review, although this review is not the same as the HEI review process used in reviewing HEI-supported research studies and Special Reports.

Environmental Health perspectives Supplements

Health Effects Institute Environmental Epidemiology Planning Project

September 1990 – September 1992
Cambridge, Massachusetts

Sponsored by

American Petroleum Institute
Chemical Manufacturers Association
Engine Manufacturers Association
Electrical Power Research Institute
Gas Research Institute
Motor Vehicle Manufacturers Association
United States Environmental Protection Agency

Organized by
The Health Effects Institute

Contents

Environmental Epidemiology Planning Project

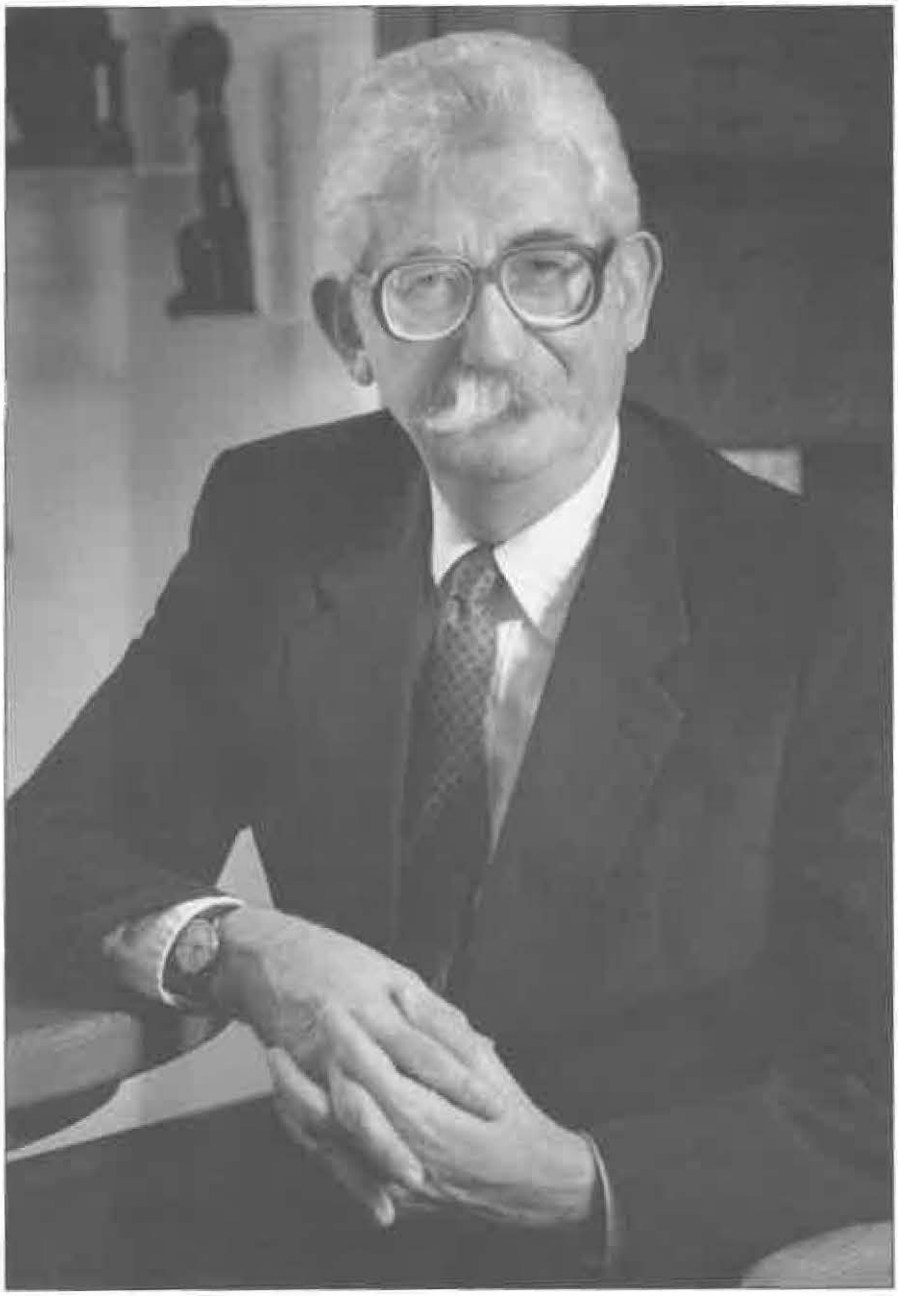
Methodologic Issues

Electric and Magnetic Fields

Indoor Air and Other Complex Mixtures

Tropospheric Ozone

Dedication



Dr. Richard DeKeraine Remington
August 2, 1931 – July 26, 1992
Chairman, Health Effects Institute
Research Committee
1989 to 1992

The Health Effects Institute dedicates the Epidemiology Planning Project documents to the memory of Dr. Richard D. Remington, a member of the Project Steering Committee and chairman of the HEI Research Committee from 1989 until June 1992.

Dr. Remington was committed to improving the lives of individuals through public health research, education, and music. His research concentrated on the epidemiology and control of cardiovascular diseases, in particular hypertension and stroke. He served in leadership roles in many local, state, and federal health agencies. As vice president for research and vice president for scientific councils, Remington developed the "Remington Plan" for reorganizing the American Heart Association. He was also a past president of the Association of Schools of Public Health and a member of The National Academy of Sciences' Institute of Medicine. At the University of Iowa, Dr. Remington served as director of the Institute for Health, Behavior and Environmental Policy, and he was honored for his contributions to education by being named the University of Iowa Foundation Distinguished Professor of Preventive Medicine and Environmental Health. Dr. Remington was also an accomplished musician who played tuba with the Alamo City Jass Band in San Antonio in the early 1960s and more recently with the Boll Weevil Jass Band in Ann Arbor.

At HEI, Dr. Remington provided great insight and forceful leadership for the Research Committee during a time when epidemiologic research was moving to the forefront of the institute's research agenda. He was the staunchest advocate for the role of epidemiology in environmental research in general and for the Epidemiology Planning Project in particular. His leadership in the institute helped bring quality science that benefits public health to the regulatory process. All those who worked with him at HEI are wiser and richer for the experience.

THE HEALTH EFFECTS INSTITUTE
ENVIRONMENTAL EPIDEMIOLOGY
PLANNING PROJECT
STEERING COMMITTEE

Introduction to the Health Effects Institute Environmental Epidemiology Planning Project Documents

Aaron J. Cohen[1*] and Leon Gordis[2]

[1]Health Effects Institute, 141 Portland Street, Cambridge, MA 02139; [2]Department of Epidemiology, The Johns Hopkins University, School of Hygiene and Public Health, 615 North Wolfe Street, Baltimore, MD 21205

— Environ Health Perspect 101(Suppl 4):15–17 (1993).

In recent years, the attention of the scientific community and of the public at large has focused more on environmental causes of human disease. In response, many branches of science have contributed to the study of the health effects of environmental pollutants. Epidemiology, the study of the distribution and determinants of human disease, has played a unique and critical role in this effort. Data from well-designed and carefully executed epidemiologic studies can measure the effects of pollutant exposures under the conditions most relevant to human experience and therefore can be especially informative about the causes of human disease. For this reason, epidemiologic data can make a unique contribution to regulatory decisions.

However, epidemiologic research on the health effects of environmental pollutants has proven difficult to conduct. In epidemiologic studies, the amount of exposure sustained by study subjects and the conditions under which that exposure occurs generally are beyond the direct control of the investigator. Human populations are exposed to multiple pollutants whose individual, let alone joint, effects are not known. Under these conditions, inaccurate measurement of exposure and the effects of extraneous factors on disease occurrence often present major threats to study validity. Epidemiologic research strategies and methods for improving exposure assessment and for measuring health effects under such real world conditions are still in their infancy; their maturation could provide the basis for substantial gains in knowledge about environmental causes of human disease, which, in turn, could provide a more scientifically sound basis for public health policy.

This Introduction was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990 – September 1992.

*Author to whom correspondence should be addressed.

In December 1989, the U.S. Environmental Protection Agency (EPA) asked the Health Effects Institute (HEI) to identify research needs and opportunities in environmental epidemiology. Founded in 1980, HEI is a nonprofit research institute that funds research on the health effects of automotive emissions with funds provided in equal amounts by the EPA and the automotive industry. HEI operates according to a mechanism designed to assure autonomy in the setting of research priorities and the dishursement of funds.

In response to the EPA's request, HEI conducted an environmental epidemiology planning project that brought together epidemiologists and other health and environmental scientists to address four selected areas of epidemiologic research. Three of these areas, electric and magnetic fields, indoor air pollution and other complex mixtures, and tropospheric ozone, involve environmental exposures of current scientific and regulatory interest. The fourth, methodologic issues, explores issues in the design and conduct of research that have implications for the study of health effects of many environmental agents.

The four general objectives of the planning project were *a*) to characterize the state of, and to identify gaps in, current knowledge in selected areas of environmental epidemiology and methodologic issues relevant to the design, conduct, analysis, and interpretation of environmental epidemiologic studies; *b*) to identify research needs and opportunities in the selected areas of environmental epidemiology; *c*) to communicate the results of the planning effort to the scientific and regulatory communities and the general public; and *d*) to aid HEI in the development of its own research programs in these research areas.

Working groups were assembled to address each of the four areas of research and were asked to produce working papers (Appendix A). Drafts of working papers were discussed at workshops held during the spring and summer of 1991, and revisions were made based upon these discussions. Revised papers were reviewed by two external reviewers and an internal HEI reviewer (Appendix B). The results of these efforts are contained in the four sets of collected papers that are published in this issue.

The project was designed to afford key researchers the opportunity to address important problems in environmental epidemiology while exploring new avenues for epidemiologic research. HEI did not ask the working groups to arrive at a consensus about their areas of research but rather, to quote University of North Carolina at Chapel Hill epidemiologist David Savitz, "to focus on the frontiers of existing knowledge and make recommendations about how to extend those frontiers" (*1*).

Hence, the papers do not necessarily reflect the views of HEI or the project's sponsors, which was intended. Neither was it our goal to achieve a uniformity of style or presentation; rather, we encouraged the working groups to define independently both the specific issues they would address and the manner in which they would address them. The resulting documents are a collection of papers that reflect each author's views as they emerged from collective discussion with members of their working group and the review process described above.

Each collection of papers offers thoughtful overviews, insightful critiques of current practice, and useful recommendations. The Working Group leaders have summarized the main conclusions and recommendations offered by their working groups in separate chapters of each document (*1–5*).

Several papers in the "Methodologic Issues" document address the critical problem

of measurement error in the characterization of exposure and argues for a reexamination of the potential contribution of ecologic or aggregate level studies in environmental epidemiology. Hatch and Thomas (*6*) critically discuss a variety of methods available to the epidemiologist to characterize environmental exposure and dose including pharmacokinetic and other models based on explicit biologic theories, sensitivity analyses, and study designs for increasing the precision of exposure measurement. Prentice and Thomas (*7*) review the statistical approaches available to account for measurement error. The articles by Morgenstern and Thomas (*8*) and Prentice and Thomas (*7*) argue for increased efforts in methods research on the theory, design, and conduct of aggregate level studies for understanding and reducing the acknowledged biases that impede the use of a potentially informative and efficient approach. Greenland (*9*) reviews the theoretical and practical issues that make the epidemiologic measurement of the effect of multiple exposures so difficult and concludes that a focus on the effects of the exposure mixture, rather than on the separate effects of its constituents, may be all that can be accomplished in most circumstances.

The current scientific interest in the health effects of electric and magnetic fields stems mostly from epidemiologic observations of increased rates of leukemias and central nervous system cancers among children exposed in the home and among certain occupational groups. The "Electric and Magnetic Fields" document offers comprehensive and critical reviews of existing knowledge about electric and magnetic fields (EMF) health effects in two areas that have, until now, received limited attention: adverse reproductive outcomes (*10*) and neurobehavioral effects (*11*). In addition, Kaune provides a comprehensive background discussion of the technical aspects of the measurement of fields (*12*) and then addresses the critical area of exposure characterization and measurement (*13*) in the context of epidemiologic research.

The collected papers in the "Indoor Air and Other Complex Mixtures" document considers the daunting task of epidemiologically studying the effects of simultaneous exposure to multiple pollutants. An important point addressed at length by Leaderer, et al. (*14*) is that improving the quality of exposure measurement would increase the informativeness of studies of multiple exposures. This echoes the views of Hatch and Thomas (*6*) and is emphasized by Greenland (*9*).

The health effects of tropospheric (ground level) ozone have been and remain an area of intense research activity for HEI and other organizations. Nevertheless, it is an area in which epidemiologic research on its long-term exposure effects is lacking and desperately needed. The collected papers in the "Tropospheric Ozone" document offers a perspective on future epidemiologic research, expounded in particular by Tager (*4*), that focuses on understanding the pathophysiologic processes and subclinical abnormalities that may constitute a relation between exposure and chronic disease. Balmes (*15*) critically reexamines the evidence, often viewed as conflicting, in search of a relation between the exacerbation of asthma (e.g., increased attack rates) and exposure to ozone, and he argues for more research on this subject.

It is noteworthy that because the planning project documents are diverse by design common threads run through them. As noted above, all four documents emphasize the need for methodologic advances in the measurement and characterization of environmental exposures for epidemiologic research. This common emphasis should not be surprising to epidemiologists, who have acknowledged that for a long time exposure assessment is the critical weakness in environmental epidemiology. Another recurring theme is the need to integrate observational epidemiologic research with experimental biologic and clinical research. In "Indoor Air and Other Complex Mixtures," Mauderly addresses the respective roles of toxicologic and epidemiologic research (*16*), Wilcosky examines the use of laboratory-derived markers of early disease (*17*), and McDonnell examines the incorporation of controlled human exposure studies into epidemiologic research protocols (*18*). The theme of integration of knowledge from experimental biology and epidemiology appears again in "Electric and Magnetic Fields," in which Stevens (*19*) proposes biologically based epidemiologic research on the carcinogenicity of electric and magnetic fields and in "Tropospheric Ozone," in which Devlin (*20*) discusses possible approaches to the development of biologic markers of exposure to ozone and early effects of this exposure. Balmes (*15*) proposes the incorporation of controlled human exposure experiments within observational study designs. In "Methodologic Issues," Hatch and Thomas (*6*) stress the need for the development of epidemiologically useful biologic markers of exposure and call the attention of biologists and epidemiologists to the necessary characteristics of such markers and potential pitfalls in their use.

The planning project was a cooperative venture between HEI and members of the environmental epidemiology research community. Project oversight was provided by a steering committee that included members of the HEI Research and Review Committees (Appendix C), several of whom served as observers and liaisons between the working groups and HEI staff. The EPA and a diverse group of private sector organizations (Appendix D) provided financial support for the project. Scientists recommended by the sponsoring organizations served as observers and liaisons between the sponsors and the working groups (Appendix A) and offered advice and commentary throughout the project.

The planning project documents already have made a valuable contribution to research planning at HEI. We hope that others will find them thought-provoking and useful for planning environmental health research. ehp

Appendices

Appendix A: Working Group Members, HEI Environmental Epidemiology Planning Project

Authors of the Working Group on Methodologic Issues. Kenneth J. Rothman, Working Group leader, editor, *Epidemiology*; Maureen O. Hatch, Columbia University School of Public Health; Hal Morgenstern, UCLA School of Public Health; Raymond Neutra, California Department of Health Services; Ross L. Prentice, Fred Hutchinson Cancer Research Center; Duncan Thomas, University of Southern California; and Dimitrios Trichopoulos, Harvard School of Public Health.

HEI Steering Committee Liaisons of the Working Group on Methodologic Issues. Richard Remington, University of Iowa and John Tukey, Princeton University.

Sponsor Observers and Liaisons of the Working Group on Methodologic Issues. John F. Aquavella, Monsanto Company and Gerhard K. Raabe, Mobil Oil Corporation.

Authors of the Working Group on Electric and Magnetic Fields. David A. Savitz, Working Group leader, University of North Carolina at Chapel Hill; William T. Kaune, EM Factors; Nigel Paneth, Michigan State University; Gary Shaw, March of Dimes, California Birth Defects Monitoring Program; Jack Siemiatycki, Institut Armand-Frappier; and Richard Stevens, Battelle Pacific Northwest Laboratories.

HEI Steering Committee Liaison of the Working Group on Electric and Magnetic Fields. Arthur Upton, New York University.

Sponsor Observers and Liaisons of the Working Group on Electric and Magnetic Fields. Donald A. Greschaw, Ford Motor Company and an alternate, Rebecca Calderon, U.S. Environmental Protection Agency.

Authors of the Working Group on Indoor Air and Other Complex Mixtures. Jonathan M. Samet, Working Group leader, University of New Mexico Cancer Center; Frank Speizer, Working Group leader, Harvard Medical School; Douglas Dockery, Harvard School of Public Health; Sander Greenland, UCLA School of Public Health; Brian Leaderer, Yale University; Paul Lioy, UMDNJ-Robert Wood Johnson Medical School; Joe Mauderly, Inhalation Toxicology Research Institute; William F. McDonnell, U.S. Environmental Protection Agency; Carl Shy, University of North Carolina at Chapel Hill; John Spengler, Harvard School of Public Health; Noel Weiss, University of Washington; and Timothy Wilcosky, Research Triangle Institute.

HEI Steering Committee Liaisons of the Working Group on Indoor Air and Other Complex Mixtures. Leon Gordis, Johns Hopkins University; Curtis Harris, National Cancer Institute; and Mark Utell, University of Rochester Medical Center.

Sponsor Observers and Liaisons of the Working Group on Indoor Air and Other Complex Mixtures. Irwin H. Billick, Gas Research Institute; Robert S. Dyer, U.S. Environmental Protection Agency; and alternates Neil C. Hawkins, Dow Chemical Company and Ronald E. Wyzga, Electrical Power Research Institute.

Authors of the Working Group on Tropospheric Ozone. Ira Tager, Working Group leader, Veterans Administration Medical Center; John Balmes, San Francisco General Hospital; David Bates, Vancouver, British Columbia, Canada; Robert Devlin, U.S. Environmental Protection Agency; Morton Lippmann, New York University; Alvaro Muñoz, Johns Hopkins School of Hygiene and Public Health; and Bart D. Ostro, California Department of Health Services.

HEI Steering Committee Liaison of the Working Group on Tropospheric Ozone. Millicent Higgins, National Heart, Lung and Blood Institute.

Sponsor Observers and Liaisons of the Working Group on Tropospheric Ozone. Jaroslav J. Vostal, General Motors Corporation and Barbara Divine, Texaco, Inc.

Appendix B: Reviewers

Methodologic Issues. John Bailar, McGill University; Lewis Kuller, University of Pittsburgh School of Public Health; and James H. Ware, Harvard School of Public Health.

Electric and Magnetic Fields. Gareth Green, Harvard School of Public Health; Charles Poole, Boston University School of Public Health; and Roy Shore, New York University Medical Center, Institute of Environmental Medicine.

Indoor Air and Other Complex Mixtures. Nathaniel Cobb, Centers for Disease Control-CEHIC; Ruth Etzel, Centers for Disease Control-CEHIC; Henry Falk, Centers for Disease Control-CEHIC; William E. Fayerweather, DuPont Company; Bernard Goldstein, Environmental and Community Medicine, UMDNJ, Robert Wood Johnson Medical School; David Mannino, Centers for Disease Control; and Roger McClellan, Chemical Industry Institute of Toxicology.

Tropospheric Ozone. Joseph D. Brain, Harvard School of Public Health; Patricia Buffler, University of California at Berkeley; and Roger Detels, UCLA School of Public Health.

Appendix C: Project Oversight

HEI Steering Committee. Leon Gordis, chairman, Johns Hopkins University, School of Public Health; Curtis Harris, National Cancer Institute; Millicent Higgins, National Heart, Lung and Blood Institute; Richard Remington, University of Iowa; John Tukey, Princeton University; Arthur Upton, New York University; and Mark Utell, University of Rochester Medical Center.

HEI Staff. Aaron J. Cohen, staff scientist; Noreen S. Manzo, administrative coordinator; Kathleen M. Nauss, director for scientific review and evaluation; Andrew Sivak, recent president (1989–1992); and Jane Warren, director of research.

Appendix D: Project Sponsors

The U.S. Environmental Protection Agency, The American Petroleum Institute, The Engine Manufacturers Association, The Motor Vehicles Manufacturers Association, The Chemical Manufacturers Association, The Electrical Power Research Institute, and The Gas Research Institute.

REFERENCES

1. Savitz DA. Health effects of electric and magnetic fields: overview of research recommendations. Environ Health Perspect 101(Suppl 4)71–72 (1993).
2. Rothman KJ. Methodologic frontiers in environmental epidemiology. Environ Health Perspect 101(Suppl 4):19–21 (1993).
3. Samet JM, Speizer FE. Introduction and recommendations: working group on indoor air and other complex mixtures. Environ Health Perspect 101(Suppl 4):143–147 (1993).
4. Tager I B. Introduction to the working group on tropospheric ozone, Health Effects Institute Environmental Epidemiology Planning Project. Environ Health Perspect 101(Suppl 4):205–207 (1993).
5. Tager I B. Summary of papers and research recommendations of the working group on tropospheric ozone, Health Effects Institute Environmental Epidemiology Planning Project. Environ Health Perspect 101(Suppl 4):237–239 (1993).
6. Hatch M, Thomas D. Measurement issues in environmental epidemiology. Environ Health Perspect 101(Suppl 4):49–57 (1993).
7. Prentice RL, Thomas D. Methodologic research needs in environmental epidemiology: data analysis. Environ Health Perspect 101(Suppl 4):39–48 (1993).
8. Morgenstern H, Thomas D. Principles of study design in environmental epidemiology. Environ Health Perspect 101(Suppl 4):23–38 (1993).
9. Greenland S. Basic problems in interaction assessment. Environ Health Perspect 101(Suppl 4):59–66 (1993).
10. Shaw G M, Croen L. Human adverse reproductive outcomes and electromagnetic field exposures: review of epidemiologic studies. Environ Health Perspect 101(Suppl 4):107–119 (1993).
11. Paneth N. Neurobehavioral effects of power-frequency electromagnetic fields. Environ Health Perspect 101(Suppl 4):101–106 (1993).
12. Kaune W T. Introduction to power-frequency electric and magnetic fields. Environ Health Perspect 101(Suppl 4):73–81 (1993).
13. Kaune W T. Assessing human exposure to power-frequency electric and magnetic fields. Environ Health Perspect 101(Suppl 4):121–133 (1993).
14. Samet JM, Speizer FE. Assessment of health effects in epidemiologic studies of air pollution. Environ Health Perspect 101(Suppl 4):149–154 (1993).
15. Balmes J R. The role of ozone exposure in the epidemiology of asthma. Environ Health Perspect 101(Suppl 4):219–224 (1993).
16. Mauderly J L. Toxicological approaches to complex mixtures. Environ Health Perspect 101(Suppl 4):155–165 (1993).
17. Wilcosky T C. Biological markers of intermediate outcomes in studies of indoor air and other complex mixtures. Environ Health Perspect 101(Suppl 4):193–197 (1993).
18. McDonnell W F. Utility of controlled human exposure studies for assessing the health effects of complex mixtures and indoor air pollutants. Environ Health Perspect 101(Suppl 4):199–203 (1993).
19. Stevens R G. Biologically based epidemiological studies of electric power and cancer. Environ Health Perspect 101(Suppl 4):93–100 (1993).
20. Devlin R B. Identification of subpopulations that are sensitive to ozone exposure: use of end points currently available and potential use of laboratory-based end points under development. Environ Health Perspect 101(Suppl 4):225–230 (1993).

Methodologic Frontiers in Environmental Epidemiology

Kenneth J. Rothman

School of Public Health, Boston University, Boston, MA 02215

Environmental epidemiology comprises the epidemiologic study of those environmental factors that are outside the immediate control of the individual. Exposures of interest to environmental epidemiologists include air pollution, water pollution, occupational exposure to physical and chemical agents, as well as psychosocial elements of environmental concern. The main methodologic problem in environmental epidemiology is exposure assessment, a problem that extends through all of epidemiologic research but looms as a towering obstacle in environmental epidemiology. One of the most promising developments in improving exposure assessment in environmental epidemiology is to find exposure biomarkers, which could serve as built-in dosimeters that reflect the biologic footprint left behind by environmental exposures. Beyond exposure assessment, epidemiologists studying environmental exposures face the difficulty of studying small effects that may be distorted by confounding that eludes easy control. This challenge may prompt reliance on new study designs, such as two-stage designs in which exposure and disease information are collected in the first stage, and covariate information is collected on a subset of subjects in state two. While the analytic methods already available for environmental epidemiology are powerful, analytic methods for ecologic studies need further development. This workshop outlines the range of methodologic issues that environmental epidemiologists must address so that their work meets the goals set by scientists and society at large. — Environ Health Perspect 101(Suppl 4):19–21 (1993).

Key Words: Environmental epidemiology, epidemiology study designs, exposure assessment

Introduction

The environment, for most epidemiologists, comprises everything that is not genetic; so diet, smoking, and even exercise are considered environmental factors. Environmental epidemiology, however, has a more restricted connotation, referring to those environmental factors that are outside the immediate control of the individual. Smoking, therefore, would not be a factor included in environmental epidemiology, but the effects of tobacco smoke put into the air by others would be. Other exposures of interest to environmental epidemiologists include air pollution, water pollution, and occupational exposure to physical and chemical agents.

The spread of infectious agents through water, foods, or other environmental media could be seen as part of environmental epidemiology, but this area has long been claimed by infectious disease epidemiologists and does not suffer from most of the methodologic problems facing environmental epidemiologists. Although there are areas of overlap between infectious disease and environmental epidemiology, such as the suspension of exotic pathogens in indoor air or the possibility of environmentally spread oncogenic viruses, environmental epidemiologists usually do not concern themselves with infectious agents.

This manuscript was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990 – September 1992.

Environmental epidemiology comprises the study of more than just physical and chemical agents, however. Rising health consciousness is a social phenomenon, and concern about the health of the environment itself, as well as its effect on us and other species, is a growing preoccupation among scientists and nonscientists alike. Psychosocial factors are increasingly important concerns in environmental epidemiology research: Studies of populations living near electric power lines or nuclear generating power plants can be neither conducted nor interpreted properly without a clear assessment of the role of the public's perception of environmental health risks. In some instances the psychologic reaction of the public may be a major component of the effect of an environmental influence; in others, the ability to conduct a study at all, and the way in which it should be conducted, are influenced profoundly by publicity and public response.

Why make a distinction between environmental exposures that can be controlled by the individual and those that are beyond his or her control? Those exposures that are beyond individual control are typically exposures that affect many individuals simultaneously and for which individual exposure may be difficult to measure. These conditions frequently lend themselves to what some epidemiologists call ecologic research, using aggregate rather than individual data. Those environmental studies that do have individual people as subjects often have distinctive methodologic features that derive from the nature of the exposure. It is as much these methodologic distinctions as the subject matter itself that warrant the use of a special term for environmental epidemiology. Furthermore, the most important research gaps in the area of environmental epidemiology may be methodologic problems.

Exposure Assessment

Atop the list of methodologic problems is the problem of exposure assessment, a problem that extends through all of epidemiologic research but is a towering obstacle in environmental epidemiology. Routine practice has been to use crude measures that are only tenuously related to the actual exposure experienced. Working in a plant, for example, has often been used as an indicator for occupational exposures that are varied in kind and intensity within the plant. Community-based sampling of air or water has been used commonly to approximate individual exposure in many studies. Indeed, in ecologic research, data may be aggregated over geographic units as large as continents. Any externally derived information as a proxy for individual exposure introduces measurement error that will affect the analysis. For exposures such as electromagnetic fields, which vary strikingly over short distances, measuring an individual's exposure by proxy measures is bound to result in substantial errors. For many exposures, a crucial part of the assessment includes the personal history. Such information is formidable to obtain

after the fact and can be obtained prospectively only with gargantuan effort. These problems in exposure assessment are compounded by the problems of low prevalence of putative high-risk exposures to the environmental agents and the low frequency of many of the outcomes of interest.

The long induction time likely to intervene between the presumed causal action of many environmental agents and the resulting appearance of disease aggravates the difficulties of exposure assessment. With a long time interval between exposure and disease, the investigator must either conduct a long, expensive prospective study or rely on retrospective measurement of the exposure information. Retrospective measurement is feasible for certain types of exposure, such as occupational exposures for which adequate employment records and industrial hygiene evaluations exist, or smoking for which the memory of the smoker usually contains a reasonable enough record of the exposure. For some exposures, such as ionizing radiation, medical records and employment information may give partial information on the amount and timing of exposure; but assessing the amount of exposure may involve considerable guesswork, making retrospective evaluations less informative. For certain unrecorded and imperceptible exposures, such as electromagnetic fields, retrospective evaluation can at best be highly indirect.

Better methods of assessing environmental exposures are a high priority for the future. One hope has been to find exposure biomarkers, which ideally might serve as built-in biologic dosimeters, to measure the biologic record of past exposure on the individual. An attraction of biomarkers is the theoretical concept that if a chronic exposure can affect disease risk, there must be a biological footprint somewhere in the organism that intermediates the causal action. The use of biomarkers can overcome measurement error that stems from an individual's incorrect recall or lack of awareness of an exposure. The use of biomarkers also can bypass exposure assessment errors arising from variation in individual absorption or metabolism of exposures by focusing on a later step in the causal chain. Chromosomal abnormalities among long-lived lymphocytes have been used in this way to assess the health effects of radiation in the studies of the Hiroshima and Nagasaki cohorts. Another example of this use of biomarkers is the possibility of using measurement of DNA adducts to assess the effects of tobacco smoke in target tissues, a method that may prove to be much more accurate than asking subjects about their smoking habits.

An additional approach to refining exposure measurement is to use multiple measures of exposure routinely until we find exposure measures that reflect the exposure as completely as the research problem demands. Replicate measures of exposure also can curb measurement uncertainty. The effect of residual uncertainty can be quantified by sensitivity analyses that explore the implications of errors in exposure assessment.

What are the priority areas for improving methods of exposure assessment in environmental epidemiology? The following areas are those that should command the highest attention [These recommendations are discussed in greater detail in the paper by Hatch and Thomas (*1*)]: *a*) development of dosimetric models using a combination of direct measurement, biological markers, and questionnaire data, and the development of new strategies for historical dose reconstruction of environmental exposures; *b*) development of sensitivity analysis and other approaches to estimating dose uncertainty, including methodology for validation substudies; and *c*) development of methods to measure covariates more accurately.

Study Design

The range of epidemiologic study designs comprises true experiments with randomized assignment of study subjects to intervention groups, as well as nonexperimental studies in which randomization cannot be relied upon to equalize the distorting effect of confounding factors related to both the exposure and the outcome. Randomized assignment of individuals into groups with different environmental exposures generally is impractical, if not unethical; community intervention trials for environmental exposures have been conducted, although seldom (if ever) with random assignment. Furthermore, the benefits of randomization are heavily diluted when the number of randomly assigned units is small, as when communities rather than individuals are randomized. Thus, environmental epidemiology consists nearly exclusively of nonexperimental epidemiology. Ideally, such studies use individuals as the unit of measurement; but often environmental data are available only for groups of individuals, and investigators turn to so-called ecologic studies to learn what they can.

The most basic epidemiologic study design, which includes experimental studies, is the cohort study. In a cohort study, a population is characterized as to its exposure to an agent of interest, and this population is then followed to measure the rate of occurrence of one or more types of disease events within variously defined exposure cohorts. Cohort studies may be entirely prospective, in which case they are expensive and usually last a long time, or they may be partially or completely retrospective, in which case they are shorter and cheaper but typically have to rely on data collected before the research plan was concocted. Case–control studies, although they have been described as backward cohort studies involving a comparison of exposure distributions in cases and controls, may be better conceptualized as streamlined cohort studies: They involve sampling the base population, or some facsimile of it, to learn the distribution of exposure within it, enabling the investigator to estimate the relative rate of disease occurrence within each exposure cohort. The sampling is usually a big cost-saver. It comes at a reasonable price—only relative rates of disease occurrence are calculable, unless the sampling fractions are known. If the sampling fractions are known, the case–control study can provide estimates of the absolute disease rates. Like cohort studies, case–control studies can be retrospective or prospective.

Ecologic studies differ from the basic cohort study in that individual exposure levels are not measured, or such exposure information, if it is measured, is not linked to disease occurrence at the individual level. The usual unit of statistical analysis is typically a geographic area, such as census tract, county, or state. For each group or region, we can estimate the distribution of individual exposures or at least the average exposure level, and we can estimate overall disease rates, but we do not have measurements of both exposure level and disease status that would allow one to estimate directly the joint distribution of the two variables. Therefore, it is impossible to get direct estimates of the rate of disease in exposed and unexposed populations from ecologic data; indirect estimates must be obtained. The indirect estimation of effects in ecologic studies and fundamental methodologic concerns, such as the control of confounding, are replete with methodologic complications that make ecologic studies a highly specialized methodologic area in epidemiology. The need to conduct such studies emanates primarily from the basic difficulty of obtaining high-quality data on environmental exposures and covariates.

The challenge posed by environmental epidemiology cannot be answered simply by conducting larger and more expensive studies; the special problems inherent in this area of research may call for new types of study

designs intended to address these problems. One example is the idea of conducting a two-stage study in which exposure and disease information are collected in the first stage, and covariate information is collected on a subset of subjects in the second stage. This study design should be useful when covariate information is expensive relative to information on exposure and disease. The results from stage one estimate a crude effect, and the information in stage two is used to estimate the effect adjusted for covariates. Covariate information is collected most efficiently in case–control studies, and therefore, we can look forward to seeing more two-stage studies in which the second stage of the investigation is a case–control study.

Another type of study that merits attention is one that focuses on intermediate steps in the causal path to disease. Such studies could give information about the relation between acute and chronic effects and provide some results much earlier than more traditional studies. Surveillance systems may be worthwhile so that selection and reporting biases can be avoided. As mentioned above, clearer understanding of the use and conduct of validation substudies is another important priority in study design. Theoretical work is needed on the validity of estimates from ecologic analyses to understand the relative importance of various assumptions and how departures from these assumptions affect the estimates. Understanding of the interaction of genes and environment will have to grow rapidly to keep pace with the information explosion about the genome. All these areas are fertile ground for more theoretical work on epidemiologic study designs.

Data Analysis

For studies on individuals with information on important confounders and little measurement error for the confounders, exposure, and outcome variables, the analytic methodology to assess exposure effects while controlling for confounding is reasonably well developed. Methods exist to control for confounding and to assess the exposure effect even when the exposures and confounding factors have complicated variations over time. Where analytic problems exist in environmental epidemiology research, it is usually the result of lack of information on confounding variables or measurement errors in confounders, exposure, or outcome variables. Such problems are the major sources of bias in environmental epidemiology research, although bias also arises from the same sources that affect all nonexperimental epidemiology, such as selection biases and information biases. Biases can arise in any study from the use of inappropriate mathematical models in an analysis; but this is a particularly important problem in ecologic studies, because they rely on aggregate data. The often-assumed linear relation between exposure and disease risk may not correspond to the biologic relation between exposure and disease. Ecologic studies also suffer from biases that distort the estimation of exposure effects because of heterogeneity of exposure status within population aggregates.

Measurement error usually has been taken into account by assuming a value for misclassification probabilities and recalculating effect estimates based on the assumed value, thus allowing a type of sensitivity analysis. Usually the misclassification probabilities are known from estimates based on limited data. A methodologic priority for data analysis is the development of methods to take account of uncertainty in the assumed values for misclassification probabilities, thus progressing from a sensitivity analysis to a more direct, corrected estimation of exposure effects that incorporates measurement error and the attached uncertainty.

Another important need is improved methods for the analysis of ecologic studies, especially with regard to controlling confounding. It would be useful to develop methods to control confounding in aggregate-data studies using information from surveys on individuals. Such approaches would call for corresponding innovation in data analysis.

Studies of multiple exposures face the formidable task of separating effects of interactions from variations in the induction periods and dose–response curves of different exposures. There is a need for analysis methods that simultaneously account for interactions, induction periods, and dose–response in a parsimonious fashion.

The difficulty and expense of epidemiologic research on environmental problems forces attention toward methods for aggregating results over a set of studies when appropriate. While many critics of meta-analysis rightly object to the pooling of inherently noncomparable work, no one argues that literature reviews are undesirable. It seems reasonable to review published work as objectively and quantitatively as possible. Meta-analysis should be thought of simply as a "quantitative literature review," as Greenland has called it (*2*). Meta-analyses should rely on the principle that the primary comparisons from which effect estimates are derived should be made within each study proper and then given appropriate statistical treatment, in terms of adjustment and weighting, to combine results across studies. Better methods are needed for adjusting the individual study-specific results to reduce bias before combining with other results, especially to take account of errors in exposure assessment that differ across studies.

Risk Assessment

Some people believe that we now live in a chemical soup that implacably erodes our health, while others believe that we have engineered an environment that protects us from most of the important health risks that otherwise would have been our fate. In either case, however, it is clear that assessing the risks of our technological world is becoming more complex.

The complexity is compounded by the intricacy of the public policy issues relating to environmental epidemiology, involving economic, political, and social concerns that must be taken into account along with the health consequences of environmental exposure. Perhaps the broadest and most important methodologic problem in environmental epidemiology is the problem of how environmental epidemiology should be used in relation to other sources of information to address these public policy issues. How many studies, and of what type, are needed before policy should be promulgated? What are the implications of publication bias (resulting from a failure to publish studies that do not show a relation between environmental exposures and health problems)? How should animal studies be weighed in relation to epidemiologic studies? What role should the public take in the conduct of research and risk assessment? The answers to these questions are important to us as citizens, but they are usually seen to be outside the scope of our work as scientists. This set of questions should be another priority for methodologic research.

REFERENCES

1. Hatch M and Thomas D. Measurement issues in environmental epidemiology. Environ Health Perspect 101(Suppl 4):49–57 (1993).
2. Greenland S. Basic problems in interaction assessment. Environ Health Perspect 101(Suppl 4):59–66 (1993).

Principles of Study Design in Environmental Epidemiology

Hal Morgenstern[1*] and Duncan Thomas[2]

[1]Department of Epidemiology, University of California at Los Angeles, School of Public Health, Los Angeles, CA 90024-1772; [2]Department of Preventive Medicine, University of Southern California, School of Medicine, Los Angeles, CA 90033-9987

This paper discusses the principles of study design and related methodologic issues in environmental epidemiology. Emphasis is given to studies aimed at evaluating causal hypotheses regarding exposures to suspected health hazards. Following background sections on the quantitative objectives and methods of population-based research, we present the major types of observational designs used in environmental epidemiology: first, the three basic designs involving the individual as the unit of analysis (i.e., cohort, cross-sectional, and case–control studies) and a brief discussion of genetic studies for assessing gene–environment interactions; second, various ecologic designs involving the group or region as the unit of analysis. Ecologic designs are given special emphasis in this paper because of our lack of resources or inability to accurately measure environmental exposures in large numbers of individuals. The paper concludes with a section highlighting current design issues in environmental epidemiology and several recommendations for future work. — Environ Health Perspect 101(Suppl 4):23–38 (1993).

Key Words: Study design, epidemiologic methods, environmental health, ecologic studies, aggregate studies, causal inference

Introduction

The purpose of this article is to discuss the principles of study design and related methodologic issues in environmental epidemiology. The focus is on studies aimed at evaluating causal hypotheses regarding exposures to suspected health hazards. Because the intended audience for this document includes scientists without formal training in epidemiology, parts of this article highlight basic principles of epidemiologic research. Nevertheless, we also try to summarize comprehensively the current state of the art and make recommendations for future developments in study design. For more extensive treatment of general research principles and methods in epidemiology, the interested reader should consult available textbooks in this area (*1–6*). More detailed examples of applications in environmental epidemiology may be found in several other books, such as those edited by Leaverton (*7*), Chiazze et al. (*8*), Goldsmith (*9*), and Kopfler and Craun (*10*).

Population Parameters

The major quantitative objectives of most epidemiologic studies are to estimate two types of population parameters: the frequency of disease occurrence in particular populations and the effect of a given exposure on disease occurrence in a particular population.

Measures of disease frequency involve the occurrence of new cases or deaths (incidence/mortality) or the presence of existing cases (prevalence). In both applications, the number of cases is expressed relative to the size of the population from which the cases are identified. With incidence measures, this denominator is the (base) population at risk (i.e., individuals who are eligible to become cases). Thus, the base population of a study (or study base) is the group of all individuals who, if they developed the disease, would become cases in the study (*3,11,12*).

Disease incidence, which is central to the process of causal inference, can be expressed as a cumulative measure (risk) or as a person-time measure (rate). The cumulative incidence (incidence proportion) or average risk in a base population is the probability of someone in that population developing the disease during a specified period, conditional on not dying first from another disease (*13*). The term cumulative incidence or cumulative incidence rate also is defined somewhat differently as the integral over the follow-up period of the hazard (rate) function (*14*). The incidence rate or instantaneous risk (hazard) is the limit of the average risk for a given period, per unit of time, as the duration of the period approaches zero. The average rate (incidence density) for a given period is estimated as the number of incident events divided by the amount of person-time experienced by the base population. For example, a rate of 0.001/year means that we would expect one new case to occur for every 1000 person-years of follow-up (e.g., 100 disease-free people followed for an average of 10 years).

Although there are many quantitative methods for expressing the magnitude of a statistical association between two variables (e.g., exposure status and disease occurrence), we are usually interested in a special class of such measures that reflect the net effect of the exposure on disease occurrence (i.e., causal parameters). In general, a causal parameter for a target population is a hypothetical contrast—in the form of a difference or ratio—between what the frequency of disease would be if everyone were exposed (at a given level) to what the frequency would be if everyone were unexposed (often called the reference level) (*15*). When this difference for a specific exposure is not zero (the ratio is not one), we say that the exposure is a risk factor for that disease in the target population. In practice, we estimate causal parameters indirectly by comparing disease frequency for an exposed group with disease frequency for an unexposed group. Epidemiologists typically estimate the risk or rate ratio (often called the relative risk) by comparing the exposed population with an unexposed population. The key assumption of this statistical approach is that the risk or rate observed for the unexposed group is the same (within confounder strata) as the risk or rate that would have been observed in the exposed group if that group had not been exposed

This manuscript was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990 – September 1992.

*Author to whom correspondence should be addressed.

This work was funded by the Health Effects Institute in Cambridge, MA. The authors would like to thank Dr. John Tukey, Dr. Sander Greenland, and other members of the HEI Methodology Working Group for their helpful comments.

(*16*). Thus, the (true) risk ratio may be interpreted as a causal parameter, which is the number of cases actually occurring in the exposed (target) population divided by the number of cases that would have occurred in the absence of exposure.

Certain measures of association, such as correlation coefficients and standardized regression coefficients, do not, in general, reflect any causal parameters. The reason is that the magnitude of these measures depends in part on the relative variances of the exposure and disease variables, which are influenced by the sampling strategy (i.e., noncausal parameters) (*17,18*). Another measure of association, the odds ratio, is used in certain types of epidemiologic studies (case–control designs) to estimate the risk or rate ratio indirectly when we cannot first estimate the incidence rate or risk in the exposed and unexposed populations (*1–6,19,20*).

Problems in Environmental Epidemiology

There are several general problems in environmental epidemiology that tend to limit causal inference and, therefore, shape design decisions.

Long Latent Periods. The interval between first exposure to an environmental risk factor (or the start of causal action of this factor) and disease detection (or symptom onset) may be many years or even decades. Such long latent periods are partly due to limitations of medical technology and incomplete surveillance for detecting disease; yet they are also due to a prolonged induction period in which years are needed for the disease process to begin (*5*). The term latent period also is used more specifically to indicate the hypothetical interval between disease initiation and detection (*5*). Refer also to Armenian and Lilienfeld (*21*) who discuss alternative definitions of latency. Unfortunately, long latent periods produce important practical constraints on our ability to estimate exposure effects. The investigator must either observe subjects for many years or rely on retrospective (historical) measurement of key variables. The latter alternative may be infeasible for certain types of exposures or in certain populations. Even when feasible, however, retrospective measurement usually increases the amount of error with which exposures are measured (see below). Furthermore, the level of most environmental exposures and many extraneous risk factors changes appreciably or unpredictably over time; long latent periods, therefore, seriously complicate our ability to estimate effects (*22*).

Errors of Exposure Measurement. A major challenge in environmental epidemiology is to measure accurately each individual's exposure to hypothesized risk factors (i.e., the biologically relevant dose [Thomas and Hatch, this issue]). This task is made very difficult by the lack of information about environmental sources of emission, the complex pattern of most long-term exposures, the individual's ignorance of previous opportunities for exposure, the lack of good biological indicators of exposure level, and the lack of sufficient resources to collect individual exposure data on large populations. The consequences of exposure mismeasurement are probable bias in the estimation of effect (see "Sources of Epidemiologic Bias") and possible loss of precision and power with which effects are estimated and tested (*23,24*). The problem and issues of exposure measurement are discussed more thoroughly by Hatch and Thomas in this issue.

Rare Diseases, Low-Level Exposures, and Small Effects. In most epidemiologic studies of environmental hazards, statistical objectives may be further compromised by the infrequent occurrence of the disease or outcome of interest, by the low prevalence or levels of environmental exposures in the general population, and by the search for small effects (for which the true rate ratio is between 0.5 and 2). A critical consequence of these features is usually substantial loss of precision and power with which effects are estimated and tested. In addition, it becomes more difficult for the investigator to separate the effect of the exposure of interest from the distorting effects of extraneous factors. Causal inference can then be seriously compromised.

Research Objectives and Strategies

Given the above problems, epidemiologists must carefully plan their studies, analyze their data, and interpret their findings. Inaccurate results reflect both random errors of estimation (chance) and systematic errors or bias. An epidemiologically unbiased or valid estimate of a causal parameter is one that is expected to represent perfectly (aside from chance) the true value of the parameter in the base population.

Sources of Epidemiologic Bias

A common framework for describing the validity of epidemiologic research is to consider three sources of bias in the estimation of effect: selection bias, information bias, and confounding (*2*). Despite the practical attractiveness of this framework, the three types of bias are not entirely separate concepts. The amount of confounding, for example, can depend on how subjects are selected.

Selection Bias. Selection bias means that the way in which subjects are selected into the study population or into the analysis (due to lost subjects or missing data) distorts the effect estimate. In general, this problem occurs when either disease status or exposure status influences the selection of subjects to a different extent in the groups being compared. Selection bias is most likely to be problematic when the investigator does not identify the base population from which study cases arose.

Information Bias. Information bias means that the nature or quality of measurement or data collection distorts the effect estimate. The primary source of information bias is error in measuring one or more variables. When exposure status or disease status is misclassified, bias usually occurs. If the probabilities of misclassification of each variable are the same for each category of the other variable (nondifferential misclassification) and if the errors for different variables are independent, the estimate of effect is usually biased toward the null value (indicating no effect). Possible exceptions to this principle of nondifferential misclassification leading to conservative effect estimates arise when the misclassified exposure variable is categorized into more than two groups (*25*). In other situations involving differential misclassification (unequal misclassification probabilities) or correlated measurement errors, the effect estimate may be biased in either direction. In many studies, therefore, the magnitude of misclassification bias is difficult to predict, especially when other biases are operating.

Confounding. Confounding refers to a lack of comparability between exposure groups (e.g., exposed versus unexposed) such that disease risk would be different even if the exposure were absent or the same in both populations (*16*). Thus, confounding is epidemiologic bias in the estimation of a causal parameter (see "Population Parameters"). Because there is no empirical method for directly observing the presence or magnitude of confounding, in practice we attempt to identify and control for manifestations of confounding. This is done by searching for differences between exposure groups in the distribution of extraneous risk factors for the disease, which are called confounders. Thus, a confounder is a risk factor (or proxy) that is associated with exposure status in the base population. A covariate meeting these criteria is not a confounder, however, if its association with the exposure is due entirely to

the effect of the exposure on the covariate; for example, the covariate might be an intermediate variable in the causal pathway between the exposure and disease. If the exposure and covariate are time-dependent variables, it is possible for that covariate to be both a confounder and an intermediate variable (see "Cohort Study").

The Need for Covariate Data

In addition to the exposure of interest, there is the need in virtually all epidemiologic studies to collect data on other known or possible risk factors for the disease. These covariates may be relevant to the exposure effect in three ways: *a*) as confounders, *b*) as intermediate variables, and *c*) as effect modifiers.

The effects of confounders must be controlled or removed analytically to obtain unbiased estimates of causal parameters. This control is usually achieved through stratification or model fitting. The assessment and control of intermediate variables can elucidate causal mechanisms that explain exposure effects (*26*). This approach often leads to new etiologic hypotheses and new intervention strategies for disease prevention.

When the exposure–effect measure varies across categories or levels of another factor, we call the second factor an effect modifier; this statistical phenomenon is called effect modification or an interaction effect. The assessment of effect modification is model-dependent, meaning that it depends on what (causal) parameter is used to measure the effect (*2–6*). For example, an extraneous risk factor that does not modify the risk ratio for the exposure will modify the risk difference. The assessment of effect modification is important for properly specifying the predictors in statistical models (*2,14*), for making inferences about possible biological (causal) interactions between exposures (e.g., synergy) (*5*), and for generalizing one's results to other populations (see "Cohort Study").

Types of Research

There are three general design strategies for conducting population research: *a*) experiments in which the investigators randomly assign (randomize) subjects to two or more treatment (exposure) groups; *b*) quasi-experiments in which the investigators make the assignments to treatment groups nonrandomly; and *c*) observational studies in which the investigators simply observe exposure (treatment) status in subjects without assignment (*2*). Although some epidemiologists classify the first two types as intervention studies, observational studies might also involve the evaluation of an intervention that was not implemented or controlled by the investigators. Social scientists often use the term quasi-experiment to mean any type of nonrandomized study (*27*).

Experiments. In a simple experiment, there are usually two treatment groups. One group is assigned to receive the new experimental intervention and the other (control) group is assigned to receive no intervention, a sham intervention (placebo), or another available intervention. Simple randomization of individuals to treatment groups implies that all possible allocation schemes of assigned subjects are equally likely (*28*). Following randomization, the investigator follows subjects for subsequent disease occurrence or change in outcome status. A comparison of risks between treatment groups provides an estimate of a causal parameter reflecting the treatment effect.

Because experiments are best suited ethically and practically to the study of health benefits, not hazards, experiments in environmental epidemiology would usually be limited to the study of preventive interventions. Furthermore, it is generally impossible or infeasible to randomize subjects individually. The only practical alternative, therefore, is to randomize by group, where the group might be a city, school, work site, etc. (*29*). The major limitation of group randomization is some within-group dependence (correlation) of the outcome variable, which reduces precision and power (*30,31*). Thus, the effective sample size falls between the number of randomized groups and the total number of subjects (see Prentice and Thomas, this issue).

As an example, consider the hypothesis that the intake of fluoride ions in drinking water has a protective effect on the occurrence of dental caries in children. An experiment might be conducted by randomly assigning many water districts (each with one fluoride-deficient water supply without treatment) either to implement sodium fluoride treatment under the control of the investigators or to continue its current policy of no treatment for the duration of follow-up. Assuming the hypothesis were true, we would expect the subsequent incidence rate of dental caries to be lower in the treated districts than in the untreated districts.

Randomization insures a valid comparison of subjects according to intended treatment, i.e., assigned treatment, but not according to treatment actually received (*16,28*). That is, randomization of a sufficient number of units (subjects or groups) provides some assurance that the assigned treatment groups are comparable with respect to inherent risk. This does not imply that there can be no confounding in a comparison of randomly assigned groups. Even with perfect adherence to treatment assignments and no loss to follow-up, assigned groups might have, by chance, different hypothetical risks in the absence of treatment. Nevertheless, such confounding, if it exists, is equally likely to be positive or negative; conventional confidence-interval estimates and p values reflect the possibility of this bias, which becomes smaller as the (effective) sample size increases (*28*). This protection against confounding afforded by randomization, however, does not apply to lack of adherence or loss to follow-up, both of which usually do not occur randomly. Furthermore, if some subjects cross over between treatments (e.g., residents of a fluoridated district obtain their water from nonfluoridated districts), a comparison of assigned groups will underestimate the true treatment effect even when the crossover is random (*32*). A comparison of compliers with noncompliers, on the other hand, is essentially observational and therefore prone to bias.

Quasi-Experiments. A quasi-experiment may be done similarly to an experiment by comparing two or more nonrandomized groups, or it may be done by comparing one or more groups over time, before versus after the intervention is initiated in at least one group. With the latter approach, the composition of each group may change over time so that subjects observed before the intervention are not the same subjects observed after the intervention.

Returning to the fluoride hypothesis, a quasi-experiment was done in the 1940s and 1950s by comparing two similar, nearby cities in New York State, both of which lacked fluoride treatment before 1945. Newburgh started sodium fluoride treatment in 1945 and continued throughout the 10-yr postintervention follow-up period; Kingston continued to use its fluoride-deficient water without treatment (*33*). The investigators found that the rate of decayed, missing, or filled (DMF) teeth in children, ages 6 to 12, decreased by almost 50% in Newburgh but increased slightly in Kingston.

Because subjects were not individually randomized in this study, it is possible that children in the treated group differed from children in the comparison group with respect to other risk factors for tooth decay, such as diet. Thus, the investigators' comparisons might have been confounded. Note, however, that randomization by city would not have reduced this possible bias

in the Newburgh–Kingston study, because the two assigned treatment groups would be equally noncomparable regardless of which city was assigned fluoride treatment.

Observational Studies. Unlike experiments and quasi-experiments, observational studies are commonly used to estimate the effects of exposures hypothesized to be harmful, fixed attributes (e.g., race and genotype), characteristics, behaviors or exposures over which the investigator has little or no control (e.g., weight, depression, and sunlight exposure), and other exposures for which manipulation or randomization would be unethical or infeasible. Observational studies are often conducted with secondary or retrospective data (instead of primary prospective data) and/or without following individual subjects for change in disease status. For example, the fluoride hypothesis could be tested by comparing the prevalence of decayed, missing, or filled teeth in children who live in areas supplied by fluoridated water with the corresponding prevalence in children who live in areas supplied by nonfluoridated water. Although such a study would be less expensive and easier to conduct than would the previous examples, there are additional methodologic problems that could lead to bias or misinterpretations.

The remainder of this article is devoted to an elaboration of observational study designs. In "Basic Observational Designs," we cover the basic designs in which data on disease status, exposure status, and all covariates are collected at the individual level; that is, the unit of analysis is the individual (or body part, such as the tooth or eye). In "Ecological Designs," we cover designs in which the unit of analysis is a group of individuals, such that information is missing on the joint distributions of key variables at the individual level.

Basic Observational Designs

Frequently, hypotheses about environmental risk factors for disease are derived from animal studies, clinical observations, reports of disease clusters, descriptive findings from population surveillance systems, and various types of exploratory studies (e.g., case series, mapping studies, and migrant studies). Formal testing of these hypotheses most often proceeds by conducting observational studies of the types described in this section.

Basic designs in epidemiology may be classified according to two dimensions: type of study population and type of sampling scheme (*34*). First, the study population is longitudinal, involving the detection of incident events during a follow-up period; or it is cross-sectional, involving the detection of prevalent cases at one time. Second, the sampling strategy involves complete selection of the entire population from which study cases are identified, or it involves incomplete or case–control sampling of a fraction (<100%) of the noncases in the population from which study cases are identified. Case–control sampling, therefore, implies stratification on disease status in the selection process. Combining these two dimensions results in four basic designs: longitudinal studies of a complete population (cohort studies); cross-sectional studies of a complete population (cross-sectional studies); longitudinal studies with case–control sampling (case–control studies with incident cases); and cross-sectional studies with case–control sampling (case–control studies with prevalent cases). In addition to these basic designs, we also discuss new developments in genetic studies for assessing gene–environment interactions (see "Genetic Studies").

Cohort Study

A cohort or follow-up study is a longitudinal design of a specified population in which exposure status is measured for all subjects at the start of follow-up (baseline) and possibly during follow-up. The entire study population—typically persons who are free of the index disease at baseline—are followed for detection of all incident cases or deaths of interest. Thus, the base population in this design is identical to the study population.

Cohort studies may be entirely prospective, meaning exposure status and disease occurrence are ascertained for the period during which the study is conducted, or they may be entirely retrospective (historical), meaning exposure status and disease occurrence are ascertained for a period before the study begins. Retrospective data are usually obtained from the subject's recall of past events or from abstracted records. Many cohort studies combine both data-collection procedures; e.g., the follow-up period for detecting the disease starts before the study and continues throughout the study period. Although retrospective studies are generally much less expensive and time-consuming, prospective studies can be designed to collect more appropriate, complete, and accurate data.

Example. Suppose we want to estimate the possible effect of prenatal exposure to passive smoke (not maternal smoking) on the risk of lower respiratory disease during the first 3 years of life. We might identify a large group of nonsmoking pregnant women and interview them just before delivery about their exposure to passive smoke during pregnancy and about other risk factors for the disease. The assessment of passive smoking would involve measuring exposure at home, work, and elsewhere with an attempt to quantify the number of smokers, cigarettes, and/or exposure time for each woman by trimester. Then each neonate would be followed by periodic examinations and parental reports of symptoms to his or her third birthday. By establishing a standard set of criteria for diagnosing new cases of lower respiratory disease and by categorizing the passive-smoke exposure into two or more categories, we can compare the 3-year risk of disease by exposure group. In this hypothetical example, the experience of each subject contributes to a single exposure group. Since subjects are not randomized to exposure groups, it is important to control analytically for other risk factors that are associated with exposure status in the study (base) population. For example, we might want to control for the child's exposure to passive smoke at home; if other family members smoked during the mother's pregnancy, they are also likely to have smoked during the child's first 3 years of life. On the other hand, we should probably not control for birth weight even if it is a risk factor for the disease, because prenatal smoking affects birth weight. Thus, provided low birth weight is a risk factor for lower respiratory disease during the first three years of life, low birth weight is likely to be an intermediate variable in the causal pathway between prenatal exposure to passive smoke and the disease.

Strengths of a Cohort Design. The prospective cohort study is the observational design that is most similar to an experiment. The major strengths of this design derive from the fact that disease occurs and is detected after subjects are selected and after exposure status is measured. Thus, we can usually be confident that the exposure preceded the disease (i.e., there is no temporal ambiguity). This feature is particularly important when disease can also influence exposure status (e.g., persons with asthma moving to drier, less-polluted areas). Well-designed retrospective cohort studies also lack temporal ambiguity of cause and effect.

Another major strength of the cohort design is the usual lack of selection bias that threatens other basic designs (*2*). Disease status cannot, in principle, influence the selection of subjects except, perhaps, in poorly designed retrospective cohort studies. Sometimes researchers, ignoring this principle, propose random sampling to reduce bias. In fact, random sampling in a cohort study, unlike random

assignment, does not prevent or necessarily reduce epidemiologic bias in effect estimation; i.e., random sampling generally does not improve comparability between exposure groups. It does, however, make the study population representative of a larger ,well-defined source population (sampling frame), which may make one's findings more generalizable. For example, suppose we initiated a prospective cohort study of lung cancer by mailing questionnaires to a random sample of 500,000 adults living in a given region served by population cancer registries. The questionnaire would request information on previous cancer diagnoses, exposure variables, and other risk factors for lung cancer. Following responses by 100,000 selected residents, the cancer registries would be used to identify all new cases of lung cancer diagnosed among respondents during the subsequent 5 years. Even though the 100,000 respondents will differ in many ways from the 400,000 nonrespondents, these differences will not cause epidemiologic bias in effect estimation. Nevertheless, the exposure effect observed for respondents (the base population) may not be generalizable to the population of nonrespondents. One possible reason for this lack of generalizability is that respondents and nonrespondents differ on the joint distribution of one or more effect modifiers.

As we will see in the next two sections, the same level of nonresponse in a cross-sectional or case–control study that we assumed in the above cohort example might seriously threaten the validity of effect estimation. Thus, unlike cohort (or randomized) studies, nonresponse in other basic designs can easily introduce selection bias because study cases have already occurred when subjects are selected. As noted in "Sources of Epidemiologic Bias," selection bias is most likely to be problematic when the investigator does not identify the base population from which study cases arose (as in cross-sectional studies and certain case–control studies).

Weaknesses of a Cohort Design. A potential weakness of cohort designs is the loss of subjects to follow-up due to death from other diseases, lack of participation, or migration. Unlike subject selection, loss to follow-up can easily bias effect estimation if attrition is associated with disease risk to a different extent for exposed and unexposed groups (*2,35*). Unfortunately, we can neither rule out nor confirm such bias by comparing lost subjects and followed subjects with respect to baseline characteristics (including risk factors) (*35*). At best, baseline similarities between lost and followed subjects only suggest that loss to follow-up is probably not a major threat to validity, especially if the attrition rate is low.

Perhaps the major practical limitation of a cohort design, especially prospective studies, is its inefficiency for studying rare outcome events, which is what most diseases are in nonclinical populations. Because exposure status and other covariates must be observed at the start of follow-up in the entire study population, a rare disease would mean that most subjects will remain noncases. Comparing a small number of cases with a large number of noncases is statistically and economically inefficient because of the diminishing marginal return from additional noncases. Assuming a fixed sample size, therefore, it is more efficient to study a disease with an expected risk of 30% than to study a disease with an expected risk of 1%; the former will result in more precision and power for estimating and testing the exposure effect. Moreover, substantial increases in the sample size to compensate for too few expected cases is often impractical or impossible, especially when the size of the exposed population available for study is limited.

Time-Dependent Exposures. In conventional analyses of cohort-study data, exposure status and other covariates are usually treated as fixed variables measured at baseline. Yet the instantaneous and cumulative level of most environmental exposures changes during the follow-up period. Consequently, the greater the change and the longer the follow-up, the less appropriate are conventional methods of analysis. A common solution to this problem is to measure average exposure, duration of exposure, or cumulative exposure before and during the follow-up period; then these variables are analyzed like the simple baseline exposure variable, as possible (fixed) predictors of disease occurrence. Unfortunately, this approach also has methodologic problems: *a*) if the follow-up period for detecting disease overlaps the period during which exposure change is measured, the temporal relationship of an exposure–disease association is ambiguous. We may not know whether exposure changes preceded disease occurrence or disease preceded changes in exposure level. *b*) If the levels of exposure and/or other risk factors change over time, the associations between the exposure and these covariates also can change; then the amount of confounding of the estimated exposure effect will change. The analytic method described above, therefore, will not, in general, eliminate confounding due to these risk factors (even when there is no misclassification). *c*) When an extraneous risk factor affects subsequent exposure status and is affected by previous exposure status, that risk factor can be a confounder and an intermediate variable simultaneously (*36,37*). For example, suppose we want to estimate the effect of exposure duration on mortality from a specific disease. If early symptoms of the disease lead to termination of exposure, then early symptoms, which is a risk factor for disease mortality, is both a confounder and an intermediate variable of the exposure–disease relationship. Consequently, standard methods of analysis will generally lead to a biased estimate of the exposure effect, whether or not one adjusts for the risk factor.

A statistical solution to the above problems was recently developed by Robins (*36,37*) who treats the prolonged or changing predictor variables as time-dependent covariates for which repeated observations are collected during the follow-up. The method involves estimating causal parameters for hypothetical exposure experiences of the study population (*15*). For example, we might want to compare the outcome risk for all subjects had they remained exposed throughout follow-up with these subjects had they remained unexposed, controlling for confounders at the start of each interval (time stratum).

Cross-Sectional Study

A cross-sectional design involves a single ascertainment of disease prevalence in a study population that is usually sampled randomly from a single source population. In this sense, the source population is that larger group of individuals who are designated by the investigator as being eligible for inclusion in the study. Generally, in a cross-sectional study, we do not know how long prevalent (existing) cases have had the disease, nor can we identify the base population (at risk) from which the study cases arose. Exposure data on time-dependent variables are usually measured retrospectively to allow for expected variations in disease latency (before detection) and duration of expression (after detection).

The statistical analysis of cross-sectional data typically resembles the analysis of cohort or case–control data. Instead of comparing disease risks for exposed and unexposed groups, we compare disease prevalences (P), as in a cohort study, or we compare the prevalence odds (P/(1-P)), as in a case–control study (see "Case–Control Study"). Under certain conditions or assumptions, the prevalence ratio or prevalence odds ratio is approximately equal to the ratio of incidence rates or risks (i.e., the causal parameter of interest) (*2,38*). For example, disease prevalence in a population

is a function of both incidence and the duration of disease. If the mean duration of disease (from onset to recovery or death) is known to be identical for exposed and unexposed cases, we can be more confident that the prevalence odds ratio approximates the incidence rate ratio.

Example. Suppose we want to estimate the possible effect of prenatal exposure to passive smoke (as in "Cohort Study") on birth weight, categorized for convenience into low (<2500 g) and normal. We identify all live births delivered in one hospital during a given period (the source population); then we take a random or quasi-random sample (e.g., every third birth). By obtaining exposure data retrospectively from mothers near the time of delivery, we can compare the prevalence of low birth weight for infants prenatally exposed and unexposed to passive smoke, controlling analytically for confounders (e.g., maternal age, maternal smoking, and prenatal care).

Even though births may be regarded as incident events, the infant's weight at birth is a prevalence measure, because we do not know the size of the base population. The causal parameter of interest is a hypothetical comparison of retarded development between fetuses exposed to passive smoke and those fetuses had they not been exposed. Not only can we not observe this hypothetical condition of exposed fetuses being unexposed, but we do not (or cannot) follow the base population; the prevalence of low birth weight is simply the end result of that hypothetical follow-up.

Strengths of a Cross-Sectional Design. Because there is no follow-up, cross-sectional studies are less time-consuming and costly than prospective cohort studies. It is also feasible to examine many exposures and diseases in the same study, which makes this design useful for screening new hypotheses. In addition to causal inference, cross-sectional studies are important descriptively in health administration, planning, and policy analysis; information on disease prevalence is often required to assess the need and demand for health services and to evaluate intervention programs in specific target populations (*2*).

Weaknesses of a Cross-Sectional Design. A major methodologic limitation of many cross-sectional studies for making causal inferences is temporal ambiguity of cause and effect. Because we usually do not know the duration of the disease in prevalent cases and because exposure status is measured at the same time as disease status, often we cannot determine that exposure (or a certain accumulation of exposure) preceded disease occurrence. One approach for minimizing this problem is to collect retrospective exposure data and information on previous medical diagnoses and the onset of symptoms associated with the disease under study. Not only may this approach be very uninformative for temporally linking exposure and disease, but it is also likely to worsen another potential problem, measurement error. Reliance on retrospective data increases the likelihood and magnitude of measurement errors, which generally leads to information bias. Furthermore, because all data are collected after disease has occurred, it is very possible for the error in measuring one variable to be related to the other variable (differential misclassification) or to error in measuring the other variable (correlated errors). Such possibilities are particularly likely in survey research and make potential information bias severe and unpredictable.

When cross-sectional studies are conducted without random sampling, they offer little opportunity for making statistical inferences about descriptive, population-specific parameters, e.g., the prevalence of a disease in a specified source population (*28*). The lack of random sampling may also worsen the potential problem of selection bias in effect estimation, which would be difficult to rule out a priori or to correct in the analysis. Even with random sampling, however, disease status or exposure status can influence the selection of subjects differentially by category of the other variable. For example, exposed cases may be less likely than others to be selected for study, perhaps because new exposed cases are less likely to survive than new unexposed cases (i.e., selective survival) or because exposed cases are less likely to enter the specified source population such as a hospital (i.e., Berkson's bias) (*2*). Similarly, selection bias can result from the differential participation of selected subjects (i.e., response bias).

Case–Control Study

Case–control studies are distinguished from other basic designs by their sampling strategy: The investigator selects only a fraction of noncases (controls) from the population from which the cases were identified (*2,3,5,34,39*). Sometimes this population is not the true (primary) base population (out of necessity or convenience), and occasionally controls are assembled without regard for the identification of cases. The design may be longitudinal, involving incident cases, or cross-sectional, involving prevalent cases. In both types, the investigator establishes the ratio of controls to cases, which does not depend directly on the frequency of disease in the population. As in cross-sectional studies, exposure data on time-dependent variables are generally measured retrospectively to account for expected variations in disease latency.

Estimation of Effect. Unless the crude disease rate or the size of the base population is known, we cannot estimate the risk or rate of the disease in the exposed and unexposed populations. Nevertheless, we can estimate the effect of the exposure on disease by calculating the exposure odds ratio, which computationally is similar to the prevalence odds ratio in a cross-sectional study (*2,3,19,20*). For this estimation of effect to be valid, however, the controls must be representative of the base population that gave rise to the study cases. In this context, representative means having a similar distribution on other disease risk factors and indicators of disease detection. The best method for making the controls representative in this way is to sample them randomly (with or without matching) from the base population (see below).

Matching. As in any observational study, the investigator should control analytically for confounders by stratification or model fitting. Intuitively, it would appear that one method for achieving this control is to match controls to cases on extraneous risk factors (i.e., making controls similar to cases on the joint distribution of these risk factors). In a case–control study, however, it is not the matching alone that controls for the confounding effects of the matching variables; rather, stratification in the analysis eliminates this bias (*1-6*). In fact, the net effect of matching in case–control studies (but not in cohort studies) is to introduce selection bias that must be controlled in the analysis. Thus, if the matching is ignored in the analysis, the effect estimate will usually be biased (*2,4,14*).

The potential advantage of matching in the selection of subjects is that it allows the investigator to control for confounders more efficiently than if matching is not used (*1–6*). Yet, in this regard, matching can be counterproductive if one matches in a case–control study on strong correlates of exposure in the base population that are not risk factors (or proxy risk factors) for the disease. This type of overmatching results in a decrease in statistical efficiency (i.e., less precision for a given number of cases and controls) (*1–6*). The conditions for overmatching, however, are very different in cohort studies in which unexposed subjects are matched to exposed subjects (*40*). Matching can also be economically counterproductive for achieving a certain minimal precision if it costs more to match

than to increase the sample size without matching (*41*).

Population-Based Case–Control Study. In a population-based or hybrid case–control study, controls (noncases) are sampled directly from the base population that gave rise to the cases (*39,42*). When this design involves the follow-up of a large dynamic population, such as residents of a state, identification of new cases is usually based on data collected through a population registry. The validity of effect estimation depends on the completeness and accuracy of case ascertainment and on careful description of the base population. When the design involves the follow-up of individuals in a fixed cohort (e.g., as a part of a clinical trial or cohort study), identification of new cases is done by exams, interviews, or questionnaires administered periodically to each individual in the cohort during the follow-up. This latter strategy is now called a nested case–control study but also has been called a synthetic case–control study (*43*).

There are three alternative methods for selecting controls in a longitudinal, population-based case–control study: *a)* In density sampling, controls are selected longitudinally throughout the follow-up. Typically, they are individually matched to cases on time of each case's diagnosis or identification and possibly other factors; i.e., each control is known to be at risk (disease-free) at the time its matched case was first identified as diseased. An advantage of time matching is that exposure status is measured at about the same time for all subjects in each matched set (*19*). *b)* In cumulative sampling, all controls are selected at the end of the follow-up period during which cases are identified. Both cumulative- and density-sampling methods can be used even when controls are not selected directly from the base population. *c)* In case–base or case–cohort sampling, all controls are selected from the fixed base population at the start of the follow-up (*42,44,45*). An advantage of this method is that one control group can be used to study multiple diseases, provided that prevalent cases of each disease are excluded from the analyses involving that disease. In both case–base and density sampling, it is possible for a selected control to subsequently develop the disease and become a case in the study.

Example. Suppose we want to estimate the possible effect of prenatal exposure to passive smoke (as in previous examples) on the risk of sudden infant death syndrome (SIDS). Using hospital records and birth certificate information, we identify a large number of live births occurring in a given region during a certain period. Then this base population is followed prospectively, using hospital records and/or a population registry to identify all infant deaths. For each diagnosed and confirmed case of SIDS, we randomly select two live controls matched to the case on age, race, and date of the case's death; thus, controls are density sampled from the follow-up experience (risk set) of the base population of live births. As soon as possible after case detection, we interview the mothers of all subjects to collect data on prenatal exposure to passive smoke and other covariates.

Proportional (Case–Control) Study. A proportional study is a special type of case–control study in which selected controls have developed or died from diseases other than the index disease under study (*2*). By definition, therefore, this is not a population-based design, since controls (especially deaths) may not be representative of the base population from which study cases arose. In a proportional morbidity study, both cases and controls are selected from a clinical population such as a hospital, clinic, physician's practice, or screening program. Controls are selected because they have other conditions or symptoms; thus, and they are likely to differ from the base population of cases in ways that affect disease occurrence or detection. This situation will usually occur when the exposure is a risk factor for those comparison diseases making up the control group. For example, we would obtain a severely biased estimate of the smoking effect in a hospital-based, case–control study if controls were selected from emphysema patients because smoking is a strong risk factor for emphysema.

Deaths comprise the entire population of a proportional mortality ratio (PMR) study. A group of deaths from the index disease (cases) is compared with a group of deaths from other diseases that might include selected comparison disease(s) or all other causes of death. Typically, all study deaths are identified retrospectively from the follow-up of a single population, such as persons living in a certain region or employed by a certain company during a given period. Although study deaths are incident events often identified from a defined base population, the outcome variable in this design is prevalence of disease at death; we do not have the proper denominator to estimate the disease-specific mortality rate in any (base) population. Furthermore, exposure data are not obtained for the base population but for study deaths only.

In the conventional proportional mortality study, comparison deaths are all other causes of death occurring in the population. The traditional method of analysis is to compute the PMR, which is the proportion of exposed deaths resulting from the index disease divided by the proportion of unexposed deaths resulting from the index disease (*6*). Alternatively, the data are analyzed as in a case–control study; the researcher computes the mortality odds ratio (*46,47*). An important advantage of the alternative approach is that the comparison (control) group might be selected to include only those diseases thought to be unrelated to exposure status. This design strategy, which also should be used in a proportional morbidity study, can help reduce selection bias by making the comparison group more representative of the base population. Another advantage is that it allows use of the many analytic techniques developed for case–control studies (*48,49*).

Strengths of a Case–Control Design. The major advantage of the case–control design over other basic designs is its efficiency for studying rare diseases, especially diseases with long latent periods. A greater proportion of study costs for collecting exposure and covariate data can be devoted to cases rather than expending most available resources on noncases. Thus, given a fixed sample size, case–control sampling in a study of a rare disease enhances the precision and power for estimating and testing the exposure effect. In addition, some case–control studies, particularly proportional mortality designs, tend to be relatively inexpensive and feasible because they can be based on readily available data sources.

Weaknesses of a Case–Control Design. A key issue in the design of case–control studies is the method and procedures for selecting controls. Ideally, we would like to make each study population-based, such that every new case occurrence in a well-defined base population is immediately identified by the investigators and controls are sampled randomly from the base population. In practice, however, this goal is not so easily accomplished, especially when the base is a large, dynamic population that cannot be examined periodically. Even population surveillance and registry systems, when they exist, are likely to be very incomplete for many diseases, such as prostate cancer, Alzheimer's disease, and ischemic heart disease. If exposed cases are more likely or less likely to be detected or reported than unexposed cases, the resulting effect estimate will be biased. In a cohort study, this detection problem would manifest as differential disease misclassification bias; but in a case–control study, the detection problem produces a form of selection bias that might involve no disease

misclassification in the total sample and, therefore, cannot be corrected after subject selection (*50*). To prevent such detection bias, the investigator might select controls who, purportedly, underwent the same degree of medical surveillance as did study cases (*51*) (e.g., persons screened for the disease or patients treated for other related conditions). Unfortunately, this approach could introduce another problem by selecting for the control group individuals with other exposure-related conditions (see the discussion of proportional morbidity studies). The end result might be, for example, to overcompensate for potential detection bias, producing net bias in the opposite direction. In general, in the absence of perfect population-based methods, investigators must select controls to reflect the expected magnitudes of various potential selection problems.

When there is relatively little variability of exposure in the base population, we expect imprecise estimation of the exposure effect, even if the exposure is a strong risk factor for the disease. Although such inefficiency is usually quite apparent in cohort studies, it may not be so apparent in case–control studies, especially when the investigator does not know the exposure distribution in the base population. For example, if environmental exposure levels are high throughout the region of the base population, a comparison of cases and controls would result in an unstable estimate of effect and low power. As in cohort studies, the problem is not one of bias. Limited variability of exposure is likely to occur when exposure status for individual subjects is measured ecologically by assigning to each subject the exposure level observed for the area in which that subject lives or works. Other problems accompanying ecologic measurement are discussed in "Ecologic Designs."

Two-Stage Designs. Just as cohort studies are inefficient for studying rare diseases, case–control studies are inefficient for studying rare exposures. When both disease and exposure are rare, therefore, any basic design might require a very large sample size to ensure adequate power. One solution to this problem is a two-stage design: stage 1 is a basic design in which data are collected on exposure and disease variables only; in stage 2, covariate data and possibly more refined exposure data (with less measurement error) are collected on separate random samples of exposed cases, exposed noncases, unexposed cases, and unexposed noncases, all of which are identified from stage 1 results (*52,53*). Sampling fractions for stage 2 are set larger for those exposure–disease groups that contain fewer subjects in stage 1. Thus, the investigator can obtain approximately equal numbers of the four exposure–disease groups in stage 2. Stage 1 results are used to estimate the crude (unadjusted) exposure effect, and stage 2 results are used to estimate the effect adjusted for covariates and possibly a more refined exposure effect. The analysis of stage 2 data considers the sampling fractions (*52–56*). The two-stage design is also advantageous when the cost of obtaining covariate data is large relative to the cost of obtaining exposure and disease data or when covariate data are missing on a majority of subjects (*52,54*).

Case–Crossover Design. A standard crossover design is an experiment or quasi-experiment in which each subject receives both the experimental and control treatments at different times (i.e., each subject serves as his or her own control) (*57*). Such designs are seldom used in environmental epidemiology because manipulation of treatment status (with or without randomization) is usually unethical or infeasible and because the outcome is usually a rare event. Recently, Maclure (*58*) proposed an observational analogue of the crossover study called the case–crossover design, which may be regarded as a special type of pairwise-matched, case–control study. This type of design can be used to estimate the possible transient effect of a brief exposure (e.g., coffee drinking) on the subsequent occurrence of a rare acute-onset disease (e.g., myocardial infarction) that is hypothesized to occur within a short time after exposure (i.e., during the effect period). All subjects are newly detected cases that serve as their own controls. That is, for each case, the observed odds of being exposed during the effect period (e.g., one hour before disease onset) is compared with the expected odds of being exposed during any random period of the same duration (assuming no exposure effect). The expected exposure odds is estimated from the subject's report of his usual exposure frequency before disease occurrence. For example, if a person drinks coffee twice each day and the effect period is 1 hr, the expected odds of exposure during any 1-hr period is 1/11. Thus, we would expect that, for every 12 cases who drank coffee twice each day, one case would have occurred by chance within 1 hr of exposure. Maclure recommends using standard methods of matched analysis for person-time (cohort) data to combine data from all cases; however, this approach needs further development to handle the temporal autocorrelation of outcome status (i.e., a case is either exposed during the effect period or unexposed, but it cannot be both). Although the case–crossover design has not yet been used to examine the possible short-term effect of an environmental exposure, this type of study is feasible if we can measure such exposures.

Genetic Study

The study of variation among individuals or groups in their sensitivity to environmental agents is one of the aims of environmental epidemiology. Such variation might be due to differences in host characteristics, including genetic factors, or to interactions with other exposures. A complete survey of the methods used to study the genetic determinants of disease would be beyond the scope of this report; instead, we will focus on the approaches that might be used to address the issue of gene–environment interactions.

Three basic types of information might shed light on the genetic component of such interactions: a classification of the subjects' genotypes at a major locus for disease susceptibility; some observable host characteristic (phenotype) that is genetically determined and linked with the genotype that was responsible for sensitivity; or family history as a surrogate for genetic (or shared environmental) influences. The choice of study design will depend upon which of these is sought.

The first is the most powerful approach, and its feasibility will grow as more and more genes are identified and assays for them become available. If the genotype is observable, it can be considered simply another risk factor and any of the basic design and analysis strategies used in epidemiology are applicable. For example, Caporaso et al. (*59*) reported a case–control study of lung cancer, in which the rate of metabolism of the antihypertensive drug debrizoquine was taken as a phenotypic marker for a gene in the cytochrome P450 system that is responsible for metabolism of carcinogens. It was shown that intermediate and high metabolizers were at higher risk of lung cancer overall and that there was an interaction between metabolic rate and exposure to occupational carcinogens and smoking. In this example, the genotype was not observed directly but inferred from the phenotype; but recent advances in molecular genetics, such as the use of restriction fragment length polymorphism, are making direct observation increasingly feasible.

Identifying host characteristics that interact with environmental exposures can be done in essentially the same way. A

familiar example might be skin color as a marker for sensitivity to sunlight in the production of melanoma. No extensions of standard epidemiologic methods would be needed to address this question. The only subtlety in this case arises when the gene determining the host characteristic is not the disease susceptibility locus but only linked to it (i.e., nearby on the same chromosome). A particular marker allele might be associated with the disease in one family and a different allele in another family; but in both families, the marker and the disease would be inherited together. This possibility requires family data and the techniques of linkage analysis. To date, such analyses have been applied only to the study of genetic effects without reference to the environmental covariates, but statistical techniques that would assess such combined effects recently have become available (*60*).

Before trying to identify a specific major gene that is related to sensitivity to environmental exposures, one should assess whether there is any evidence that such sensitivity has a genetic basis. This also requires collection of family data, but unlike the standard analyses aimed at examining the main effect of genetics, one would also want to examine interactions between family history and environmental exposures. Geneticists commonly assemble a small number of very large pedigrees, sometimes selected to maximize the chances that a major gene is operating in the families, and subject them to segregation analysis to study the mode of inheritance. Again, these analyses seldom account for environmental covariates and interactions, although such methods are now available. In contrast, epidemiologists begin with a large population-based series of cases and controls and restrict attention to the first- and sometimes second-degree family members. Their analyses usually are limited to a simple family history covariate (e.g., presence of an affected member, number of affected members, etc.) in standard multivariate risk-factor models, possibly but seldom including interactions with environmental covariates. Susser and Susser (*61*) have discussed two basic approaches to such data: in the case–control approach, cases are compared with controls in terms of their family histories; in the cohort approach, the incidence of disease in the exposed (case) families is compared with the incidence in unexposed (control) families. Either approach could easily be extended to incorporate environmental covariates and their interactions with family history. The only difference is that the cohort approach requires covariate data on all of the family members, whereas the case–control approach requires data only on the originally sampled cases and controls. Clearly, the cohort approach is more informative, but the conditions under which the additional effort warrants the gain in information about gene–environment interactions have not been investigated.

The two major design issues to be addressed in such studies are the method of ascertainment of families and the information to be collected on family members. The former has been discussed at great length in the genetics literature. The basic problem is that if families are ascertained through affected probands, families with multiple cases will tend to be overrepresented. Therefore, various corrections for ascertainment are applied in the standard methods of genetic analysis. The relevance of these approaches to the epidemiologic designs for gene–environment interactions requires further research. Often, only very limited information is collected about family history in epidemiologic studies. The minimal information should be an enumeration of all affected family members together with the sex and age of each family member at risk; as discussed above, information on major risk factors for all family members at risk may also be desirable. Because larger and older families are likely to have more familial cases, the presence or number of familial cases is not suitable as a family history covariate. Moreover, expressing the number affected as a proportion does not solve the problem because multiple cases in large families are more informative than single cases in small families. A more appropriate comparison is between the observed number of familial cases and the expected number based on the person-time at risk, in which the comparison is adjusted for age, sex, and other important risk factors (*62*).

Ecologic Designs

An ecologic or aggregate study is one in which exposure levels of individuals are not linked to disease occurrence of those individuals. The net result is that the unit of statistical analysis is usually the group, typically persons living in a geographic area such as a census tract, county, or state. For each group or region, therefore, we know the average exposure level or distribution and the disease rate, but we do not know the joint distribution of these two variables. Given a dichotomous exposure, for example, we would not know the numbers of exposed and unexposed cases in each group. Thus, we cannot estimate the exposure effect directly by comparing the disease rate for exposed and unexposed populations.

Ecologic designs are therefore incomplete (*2*) in the sense that they lack certain information ordinarily contained in the basic designs. As noted in "Problems in Environmental Epidemiology," the primary reason for this missing information in environmental epidemiology is our inability or lack of resources to accurately measure environmental exposures in large numbers of individuals. Thus, the widespread use of ecologic designs in environmental epidemiology reflects a fundamental problem of exposure measurement. In addition, ecologic studies represent an inexpensive design option for linking available data sets or record systems, even when exposures are measured at the individual level. The appeal of this alternative is that aggregate summaries of many exposures, including sociodemographic and other census variables, are often available for the same regions that are used to summarize morbidity and mortality data.

With the inclusion of covariate data in an ecologic study, the analysis may be only partly ecologic. This condition occurs when the joint distribution of two or more, but not all, variables is known within groups. For example, suppose we want to examine the possible effect of radon exposure on lung cancer incidence, controlling for age (the covariate). Although we might know the age distribution of all new cases and all persons at risk within each county (from tumor registry and census data), we would usually not know the within-county association between radon exposure and the other two variables. Sometimes data sets like this are analyzed with the individual as the unit of analysis, where each individual is assigned the average radon exposure level that was measured for the region in which he or she lives. Such ecologic measurement of exposure means that there is likely to be substantial error in measuring the individual's exposure to radon, which could result in information bias of effect estimation.

Types of Ecologic Studies

Ecologic studies may be classified into five design types that differ in several ways, including methods of subject selection and methods of analysis (*2,63*).

Exploratory Studies. In exploratory ecologic studies, we compare the rate of disease among many contiguous regions during the same period, or we compare the rate over time in one region. In neither approach are exposures to specific environmental factors measured (for individuals or groups). The purpose is to search for spatial or temporal patterns that might suggest

an environmental etiology or more specific etiologic hypotheses.

The simplest type of exploratory study of spatial patterns is a graphical comparison of relative rates across all regions (i.e., mapping study), possibly accompanied by a statistical test for the null hypothesis of no geographic clustering (*64*). In mapping studies, however, a simple comparison of estimated rates across regions is often complicated by two statistical issues. First, regions with smaller numbers of observed cases show greater variability in the estimated rate; thus, the most extreme rates tend to be estimated for those regions with the fewest cases. Second, nearby regions tend to have more similar rates than do distant regions (i.e., positive autocorrelation). A statistical method for dealing with both complications involves empirical Baye' estimation of rates using an autoregressive spatial model (*65*).

In certain exploratory studies of spatial patterns, regions are characterized in terms of general ecologic indicators such as degree of urbanization (urban versus rural), degree of industrialization (agricultural versus nonagricultural), population density, socioeconomic status, and ethnic diversity. The analysis of these data usually involves comparisons of regions grouped by one or more ecologic indicators. This approach resembles the statistical methods used in multiple-group studies (see "Interpretation of Results").

An exploratory ecologic study was conducted by Mahoney et al. (*66*), who compared age-standardized mortality ratios for cancers, by sex, among all cities and towns in New York State (exclusive of New York City) between 1978 and 1982. By grouping these regions by quintile of population density, they examined the associations between density and deaths from all cancer sites and selected sites, by sex. They found linear associations between increasing population density and total cancer mortality in both men and women. Because population density may reflect various risk factors for different cancers, the authors acknowledge that their findings are consistent with several alternative explanations.

An exploratory study of temporal patterns is generally done by comparing disease rates for a geographically defined population over a period of at least 20 years. A common statistical or graphical approach for analyzing such longitudinal data is cohort analysis (not to be confused with the analysis of data from a cohort study) (*2*). The objective of this approach is to estimate the separate effects of three time-related variables on disease occurrence: age, period (calendar time), and birth cohort (year of birth). Because of the linear dependency of these three variables, there is an inherent statistical limitation (identification problem) with the interpretation of cohort-analysis results. The problem is that each data set has alternative explanations with respect to the combination of age, period, and cohort effects. The only way to decide which interpretation should be accepted is to consider the findings in light of other (prior) knowledge of the disease and its determinants.

A cohort analysis was conducted by Lee et al. (*67*) on melanoma mortality among white males living in the United States between 1951 and 1975. They concluded that the apparent increase in the melanoma mortality rate during that period was due primarily to a cohort effect. That is, persons born in more recent years carried with them throughout their lives a higher mortality rate than did persons born earlier. In a subsequent review paper, Lee (*68*) speculates that the cohort effect might reflect the impact of changes in a major risk factor operating during youth, such as sunlight exposure or burning.

Space–Time Cluster Study. Space–time clustering refers to the interaction between place and time of disease occurrence, such that cases that occur close in space also occur close in time (*2*). Evidence of space–time clustering may suggest person-to-person transmission of an infectious agent or the effects of point-source exposures, depending on the disease and the cluster pattern. The analytic search for space–time clusters requires special statistical techniques that may or may not incorporate information on the base population and covariates (*69,70*). Although the unit of analysis for these methods is usually the individual, space–time cluster studies are classified here with ecologic designs because closeness in space and time is a proxy measure for environmental exposures—or at least the opportunity for exposures. Thus, use of place and time information is analogous to use of spatial or temporal indicators in the exploratory study.

Space–time cluster analyses may be used when members of a community perceive a cluster or excess number of cases of one or more diseases in their area. This activity is often motivated by the suspicion that the apparent cluster is caused by a specific environmental exposure, such as chemical waste, pesticides, or electromagnetic fields. When investigation begins, the first steps are to verify the diagnoses of all reported cases and identify any additional cases in the cluster area, which must be defined. In addition to space–time cluster analyses, the investigators will probably want to compare the disease rate in the cluster-area population with the rate in another population thought to be unexposed (retrospective cohort study), and they may conduct a population-based case–control study to identify risk factors for the disease.

Multiple-Group Study. In a multiple-group ecologic study, we assess the ecologic association between average exposure level or prevalence and the rate of disease among many groups or regions. This is the most frequently used ecologic design in environmental epidemiology. Studies are usually conducted by linking separate sources of data. For example, census and tumor-registry data might be combined to estimate cancer rates for all counties in a state; other state records or surveys might be used to estimate average exposure levels by county. Statistical methods for estimating exposure effects in multiple-group studies are discussed in "Interpretation of Results" and by Prentice and Thomas in this issue.

Hatch and Susser (*71*) conducted a multiple-group ecologic study to examine the association between background gamma radiation and childhood cancers between 1975 and 1985 in the region surrounding the Three Mile Island nuclear plant. Using data from a 1976 aerial survey, they estimated the average radiation level for each of 69 tracts in the study region. The results of their analyses showed a positive association between radiation level and the incidence of childhood cancers. The authors were cautious in making causal inferences, however, because the large effect observed for solid tumors, as well as leukemias, was not expected.

Time-Trend Study. In time-trend (or time-series) studies, we assess the ecologic association between change in average exposure level or prevalence and change in disease rate in one geographically defined population. The assessment may be done by simple graphical displays or by more formal statistical techniques (*72–75*). With either approach, however, the interpretation of findings is often complicated by two issues. First, changes in disease classification and diagnostic criteria can produce very misleading results. Second, the latency of the disease with respect to the exposure of interest may be long, variable across cases, and/or unknown to the investigator; thus, employing an arbitrary or empirically defined lag between the two trends can also produce very misleading results (*76*).

Darby and Doll (*77*) compared the trends of average annual absorbed doses of radiation fallout from weapons testing and childhood leukemia rates in three European countries between 1945 and 1985. Although the

leukemia rates varied over time in each country, they found no convincing evidence that these changes were attributed to changes in fallout radiation.

Mixed Study. The mixed ecologic design combines the basic features of the multiple-group study and the time-trend study. The objective is to assess the ecologic association between change in average exposure level or prevalence and change in disease rate among many groups. Thus, two types of comparisons are made simultaneously: change over time within groups and differences among groups.

For example, Crawford et al. (*78*) evaluated the hypothesis that hard drinking water (i.e., water containing more calcium and magnesium ions) is a protective risk factor for cardiovascular disease (CVD). They compared the absolute change in CVD mortality rate between 1948 and 1964, by age and sex, in 83 British towns. The towns were divided into three groups: *a*) five had experienced increases in water hardness; *b*) six had experienced decreases, and *c*) 72 had experienced little or no change in water hardness. In all sex–age groups, especially for men, the authors found an inverse association between trends in water hardness and CVD mortality. In middle-aged men, for example, the increase in CVD mortality was less in towns that made their water harder than in towns that made their water softer.

Interpretation of Results

Statistical analysis in a multiple-group study usually involves fitting the data to a mathematical model (see Prentice and Thomas, this issue). The outcome variable is a function of the disease rate in each group; predictors include the average exposure level or proportion exposed in each group plus other ecologic covariates, the effects of which the investigator wants to control. We show in "Control for Covariates" that these covariates need not be confounders (i.e., at the individual level within groups).

Results of the fitted model can be used to estimate the exposure effect, i.e., the same causal parameter we would like to have estimated had the study been conducted at the individual level (*63,79,80*). For example, suppose the exposure variable is the proportion exposed in each group and there are no covariates. Assuming a linear model, we can use weighted least-squares regression to estimate the slope (b) and intercept (a). The predicted disease rate in a group that is entirely exposed is then $a + b(1) = a + b$, and the predicted rate in a group that is entirely unexposed is $a + b(0) = a$; therefore, the estimated rate ratio is $(a + b)/a = 1 + b/a$. It is important to note that this estimation procedure implies extrapolating the results of the model to both extreme values of the exposure variable, either or both of which may lie well beyond the observed range. It is not surprising, therefore, that different model forms can lead to very different estimates of effect (*81*). In fact, certain model assumptions may lead to rate-ratio estimates that are negative and thus meaningless.

Ecologic Bias

The use of ecologic data to estimate causal parameters has a major methodologic limitation, called the ecological fallacy (*82*), aggregation bias (*83*), cross-level bias (*84*), and ecologic bias (*85,86*). Ecologic bias refers, in general, to the failure of ecologic estimates of effect to reflect the true effect at the individual level. Some of this bias may occur in individual-level studies of the same population, but some of it is due specifically to the aggregation of subjects into groups. More importantly, the magnitude of ecologic bias is likely to be more severe and less predictable than is individual-level bias in estimating the same effect (*63,81,86,87*). It is very possible, for example, that an ecologic analysis of a (true) positive risk factor would produce an apparently protective effect.

The underlying problem of ecologic bias may be regarded as a special form of information bias resulting from within-group heterogeneity of exposure status, which is not captured in the analysis. For example, a positive linear relationship between proportion exposed and disease rate does not necessarily mean that exposed individuals are at greater risk for the disease than are unexposed individuals; rather, unexposed individuals may be at greater risk in groups containing proportionally more exposed individuals. The implication of this latter explanation is that an individual's group affiliation has an effect on disease occurrence that reflects more than just the individual's exposure status.

A mathematical understanding of ecologic bias was first provided for correlation coefficients by Robinson (*88*) and later extended to regression coefficients by Duncan et al. (*89*). Nevertheless, the conditions for valid ecologic estimation and the relationship between ecologic bias and other methodologic issues are still not well understood. Because the results of ecologic analyses are often used to influence policy decisions, as well as to make causal inferences, it is important for researchers to appreciate the complexities of ecologic inference.

Sources of Ecologic Bias. Ecologic bias is often confused with confounding, perhaps because regional differences in disease rates can be due to variation in the distribution of extraneous risk factors across regions. To clarify the confusion between these two concepts, Greenland and Morgenstern (*86,87,90*) show that ecologic bias can arise from three different sources.

Within-Group Confounding (At the Individual Level). The exposure effect may be confounded within groups (as described for nonecologic studies in "Sources of Epidemiologic Bias"). Thus, if the within-group effect is equally confounded by the same unmeasured risk factors in every group, we can expect the ecologic estimate of effect to be biased as well. In general, ecologic estimates will be biased in this way if the net within-group bias across groups (due to uncontrolled confounders) is not zero. It is possible, therefore, for positive confounding in certain groups to cancel negative confounding in other groups.

The other two sources of ecologic bias are unique to this design and can be understood by considering group (or group affiliation) as a nominal predictor of disease occurrence at the individual level.

Confounding by Group. Ecologic bias can occur when the disease rate in the unexposed population varies across groups. Since average exposure level also typically varies across groups, group is a confounder of the exposure effect at the individual level. This set of conditions may occur if one or more unmeasured risk factors are differentially distributed across groups, even if these risk factors are unrelated to exposure status within groups and, therefore, are not confounders at the individual level.

Effect Modification by Group. Ecologic bias can also occur when the exposure effect varies across groups, i.e., when group modifies the effect of the exposure at the individual level. This condition may result from extraneous risk factors (effect modifiers) being differentially distributed across groups or by misspecification of the model form used to analyze the data. Ecologic bias of this type tends to be more severe when there is little variability in average exposure across groups (*85*), even when the effect modification is relatively weak and there is no confounding by group.

Taken together, the above principles imply that there will be no ecologic bias if the disease rate in the unexposed population and the exposure effect do not vary across groups and if there is no net confounding within groups. Unfortunately, it is very unlikely that all of these conditions

will be met in one ecologic study. Although small departures from these conditions may result in substantial bias (*81,86*), it is also possible that there will be little or no bias in certain studies when one or more of these conditions are not met.

If every group were completely exposed or unexposed, there would be no ecologic bias attributable to confounding or effect modification by group. Indeed, if all covariates were measured at the individual level, such a study would not be an ecologic design. Thus, to reduce ecologic bias, we should select regions that minimize within-region exposure variation and maximize between-region variation (*63,81*). One strategy for achieving these goals is to choose the smallest unit of analysis for which required data are available (e.g., census tracts or blocks). Unfortunately, certain data are seldom available at this level (e.g., personal behaviors and biomedical factors), and there is no guarantee that these smaller units are more homogeneous with respect to exposure status. Furthermore, use of smaller groups might increase the problem of migration between groups (see "Other Methodologic Problems").

Control for Covariates

In a study conducted entirely at the individual level, an extraneous risk factor produces bias (confounding) in effect estimation only if it is associated with exposure status in the base population (see "Sources of Epidemiologic Bias"). In a multiple-group ecologic study, however, an extraneous risk factor can produce ecologic bias even if it is not associated with exposure status within regions (at the individual level) (*86,87,90*). Such bias occurs typically because the ecologic association (across regions) between the exposure and risk factor produces confounding and/or modification of the exposure effect by group (see "Ecologic Bias"). Conversely, an extraneous risk factor that is a confounder within regions may not produce ecologic bias if the net within-group bias is zero (see "Ecologic Bias") or if the risk factor is ecologically uncorrelated with the exposure.

One method to control for extraneous risk factors in ecologic studies is to include predictor terms for these risk factors in the model (e.g., the proportion of smokers or the mean family income in each region). Unfortunately, even when such covariate data are available for all regions, ecologic adjustment usually cannot be expected to remove completely the ecologic bias produced by these risk factors. In fact, it is possible for such ecologic adjustment to increase bias (*86*).

The general conditions under which the ecologic control for extraneous risk factors either increases or decreases bias have not been delineated. Yet, under certain restrictive conditions, ecologic control for covariates will produce unbiased estimates of the exposure effect, provided there are no other sources of bias (e.g., outcome misclassification). If the effects of the exposure and the covariate on disease rate are exactly additive within every region (i.e., the rate difference for each variable is constant across levels of the other variable) and if the rate conditional on both predictors is exactly the same in every region, ecologic regression of disease rate on the mean exposure and covariate levels (i.e., multiple linear regression) will lead to unbiased estimates of both effects (*83,84*). Under these conditions, group affiliation does not confound or modify the exposure effect at the individual level. However, as shown by Greenland (*81*), relatively minor deviations from perfect additivity (linearity) can lead to appreciable ecologic bias because ecologic rate ratios can be extremely sensitive to the choice of model form, in contrast to individual-level estimates. Furthermore, the two conditions noted above are only sufficient for no ecologic bias to occur; ecologic bias may be absent when either or both conditions are not met.

Richardson and Hémon (*91*) recently pointed out that there is another set of conditions for which ecologic control of covariates is possible. If *a*) the exposure and covariates are uncorrelated within regions, *b*) their effects on disease are multiplicative (i.e., the rate ratio for each variable is constant across levels of the other variable), and *c*) the rate conditional on both predictors is exactly the same in every region, then ecologic bias due to the covariates can be removed or largely reduced by including product terms in the linear model. Of course, such conditions are very difficult to verify in ecologic studies; if the exposure and covariates (other risk factors) are correlated within regions, the covariates will be confounders at the individual level and substantial ecologic bias can occur even with product terms in the model (*81*).

When the data are not entirely ecologic (see "Ecologic Designs"), rate standardization is another method often employed to adjust for extraneous risk factors in ecologic studies. For example, if the age distribution is known for cases and for the base population in every region, we can mutually standardize the rate in every region to the age distribution of a well-defined (standard) population (*5*); then we use the standardized rates as the outcome variable in the ecologic analysis. Unfortunately, this method does not always reduce ecologic bias due to the variables for which the rates are standardized; in fact, the result may be to increase bias appreciably (*86,92*). Standardization can be expected to reduce ecologic bias only if all variables in the model (i.e., disease and all predictors) are mutually standardized for those other confounders (e.g., age) not included as predictors in the regression model. This method is often not feasible, for example, when the investigator does not know the age distribution of exposed and unexposed populations within every region.

Other Methodologic Problems

In addition to ecologic bias and the related difficulties of controlling for extraneous risk factors, there are other methodologic problems with ecologic analysis, a few of which are addressed below.

Exposure Misclassification Bias. As noted in "Sources of Epidemiologic Bias," nondifferential misclassification of exposure status in individual-level studies nearly always results in bias toward the null value; e.g., the estimated rate ratio will be closer to one than is the true rate ratio. In multiple-group ecologic studies, however, this principle does not hold when the exposure variable is formed from the aggregated observations of all individuals in each region (e.g., the proportion exposed). Brenner et al. (*93*) have shown that nondifferential misclassification of a binary exposure within groups usually leads to overestimation of the rate ratio (away from the null value) in ecologic studies, which can be severe. This apparent contradiction between ecologic and individual-level studies can be understood by considering just two regions. Nondifferential exposure misclassification in both regions will produce an estimated difference in exposure prevalence that is smaller than the true difference. Consequently, the estimated regression coefficient (slope) for the exposure variable in a linear ecologic model will be overestimated, leading to overestimation of the rate ratio. Little is known about the impact in ecologic studies of within-group error in measuring continuous or multiple-category exposures.

Confounder Misclassification. In studies conducted at the individual level, misclassification of a confounder, if nondifferential with respect to exposure and disease, will usually reduce our ability to control for the confounder in the analysis (*94,95*). That is, adjustment will not completely eliminate the bias due to the confounder. In ecologic studies, however, nondifferential misclassification of a binary confounder within groups does not affect our ability to control for that confounder (*96*). Thus, sur-

prisingly, nondifferential misclassification of a confounder is less problematic in ecologic studies, provided there is no ecologic bias, than in individual-level studies.

Collinearity. It is probably more common in ecologic studies than in other studies for two or more predictors to be highly correlated across groups (*63,97,98*). This issue is particularly relevant with environmental factors, such as the associations between levels of different contaminants in air or drinking water or associations between different socioeconomic indicators. The implication of such collinearities is that it is very difficult, perhaps impossible, to separate these effects statistically; analyses yield model coefficients with very large variances and often severely distorted estimates of effect.

Temporal Ambiguity of Cause and Effect. Use of incidence data in a cohort study usually implies that disease occurrence did not precede exposure to the hypothesized risk factor. Yet, in multiple-group or time-trend ecologic studies use of incidence data provides no such assurance against this temporal ambiguity (*63*). This inferential problem is most troublesome when it is possible for disease to influence exposure status either at the individual level (see "Cohort Study") or at the ecologic level (e.g., interventions designed to reduce exposure levels in areas with high rates of disease).

The problem of temporal ambiguity in ecologic studies is further complicated by an unknown or variable latent period between exposure and disease occurrence. The investigator can only attempt to deal with this problem by establishing a specific lag period between observations of average exposure and disease rate. Even when the average latency is known, however, appropriate data may not be available to accommodate the desired lag.

Migration. Migration of individuals into or out of the base population can cause selection bias in any type of epidemiologic study, because migrants and nonmigrants may differ on both exposure prevalence and disease risk. Little is known about the magnitude of this bias or how it can be reduced in ecologic studies, especially when studying diseases with long latent periods. One approach might be to use larger geographic groups (e.g., states instead of counties as units of analysis) (*99*). Unfortunately, this approach is also likely to increase the potential for severe ecologic bias, because it makes the groups less homogeneous with respect to exposure (see "Ecologic Bias"). Another approach might be to incorporate available data on the distributions of residential durations within regions, but this approach needs more work to provide a reliable method of bias reduction.

Current Issues and Recommendations

A general goal of epidemiologic research is to obtain the most information about possible health effects with minimal and/or available resources. Given the difficulties in estimating effects of specific environmental exposures in human populations, this goal is not easily obtained and optimal research strategies are not readily identified. Below, we highlight several current methodologic issues in environmental epidemiology and make some recommendations for future work.

Study Design. No single design best meets the objectives of every epidemiologic study. In practice, study objectives are shaped by many factors—current knowledge, previous findings, institutional mandates, societal values, personal preferences, etc. Although a prospective cohort study might be expected, in general, to produce less bias than would a hospital-based (proportional) case–control study, the latter design might be a rational choice in certain situations. Even an ecologic design, despite its limitations, might be appropriate; it may be the only practical option at a given time.

The challenges of environmental epidemiology, therefore, cannot be solved simply by advocating the use of certain, more expensive study designs. In addition to committing more resources to the conduct of epidemiologic research, we need to develop new designs to meet specific objectives more efficiently. For example, in "Case–Control Study," we discussed the use of two-stage designs to investigate associations between rare diseases and rare exposures and to control for covariates that are relatively expensive to measure. New approaches are also needed to identify intermediate variables in observational designs, to evaluate interaction effects (effect modification) more efficiently, and to deal with the problems of nonparticipation, nonresponse, and noncompliance. Another need in environmental epidemiology is to understand better the relationship between acute biological changes and chronic health effects. For example, we might combine experimental and observational methods to determine the extent to which short-term changes in pulmonary function caused by exposure to air pollutants lead to chronic respiratory disease (*2*).

Bias Reduction. In nonrandomized studies, it is important for the investigator to deal with confounding in the analysis. This is achieved by identifying potential confounders in the design phase and measuring them accurately in the study population. The prevention of selection bias, however, is not so straightforward because it depends on identifying all cases that occur in a well-defined (base) population at risk. When new cases occur infrequently or when it is otherwise impractical to reexamine enough individuals to detect all new events, the prevention of selection bias depends on population surveillance and monitoring systems, such as population-based tumor registries and industrial surveillance. Although these systems may be expensive to implement and operate, they are often necessary to reduce the threat of selection bias.

Unfortunately, population-based systems may not be sufficient to prevent selection bias with diseases for which detection depends critically on care seeking, symptom reporting, and complex differential diagnoses. A key problem is that not all persons with an illness recognize their symptoms and seek medical attention. Thus, exposure effects observed for these diseases in epidemiologic studies might reflect the effects of the exposure on illness behavior as well as the effects on illness occurrence (*100*).

Another solution to incomplete or inadequate case detection is to control analytically for methodologic covariates that reflect differences in illness behavior. For example, we might measure the individual's tendency to seek medical care and treat this variable as a confounder. The measurement of this covariate should be independent of disease status; otherwise, covariate adjustment will probably lead to bias toward the null value. This approach needs further development and evaluation.

An alternative strategy for studying diseases that are difficult to detect in large populations, such as musculoskeletal conditions, is another type of two-stage design. In the first stage, a large population is surveyed cross-sectionally or longitudinally by questionnaires or interviews to identify persons with symptoms characteristic of the disease. The second stage involves case–control sampling of the population to compare persons with and without these symptoms (i.e., cases and controls). In this stage, subjects are given more definitive diagnostic tests to identify true cases of the disease. By comparing diagnostic test results between selected cases and controls, the investigators can assess the validity of their symptom-based criteria, suggest improvements in clinical diagnosis, and estimate exposure effects. The latter objective requires the development of statistical methods appropriate for the sampling strategy.

Quality of Measurement. As noted earlier (see "Problems in Environmental Epidemiology"), a major challenge in environmental epidemiology is to measure accurately each individual's exposure to suspected and known risk factors for the disease under study. In the absence of previously validated and inexpensive methods for measuring exposures and covariates in large groups, it has become common practice to use more than one method or source of information to measure these variables. Nevertheless, it usually is not clear how different methods or sources of information should be combined or what data should be combined to minimize measurement errors and estimation bias (*4,101,102*). We need more methodologic research in this area to provide guidelines for the measurement of specific exposures in particular types of populations. One approach that might be pursued with environmental exposures is to combine ecologic data with self-reported data on individual behaviors. For example, suppose we collect ecologic data on pesticide spraying and distribution throughout a large region. We could then obtain from subjects the location of their homes; the type and location of their work; their use of drinking water; and how often they swim, fish, and participate in other activities that would affect their exposure to pesticides in the region.

Frequently, an accurate method does exist for measuring an exposure, but the application of this method to all subjects in a population is prohibitively expensive or infeasible. In such cases, many investigators rely on less accurate methods for the total sample and use the more accurate method in a subsample of subjects. Assuming the accurate method is perfectly valid (i.e., the gold standard), the results of the validation substudy are used to quantify the amount of measurement error, which is then used in the total sample to correct for misclassification bias involving the imperfect measure of exposure. Some important issues need to be considered to make this approach advantageous. First, how many subjects and what proportion of the total sample should be included in the substudy (*103,104*)? Second, how should we correct for exposure misclassification in the analysis, especially when the accurate method may not be perfectly valid or when the subsample is not representative of the total sample (see also Prentice and Thomas, this issue)?

Ecologic Inference. Because of inherent problems of measurement, most epidemiologic studies of environmental exposures are at least partly ecologic. When all data, except a single exposure, are obtained at the individual level, however, the ecologic problem amounts to possible misclassification bias, which is well understood and often predictable. Yet, when the unit of analysis is the group, the resulting ecologic bias is far less predictable and can be relatively severe in magnitude, especially when other sources of bias are present. Thus, in general, ecologic analyses do not provide very accurate estimates of effect. To make ecologic findings more informative, therefore, we need more theoretical work to specify the conditions for which ecologic estimates can be expected to be reasonably valid. With this information, we might then collect additional data to check those key assumptions or to correct ecologic estimates. For example, by obtaining detailed individual-level data on the exposure and certain covariates in samples of selected groups, we might be able to determine the limits of ecologic bias in estimating the exposure effect (see "Control for Covariates" and Prentice and Thomas, this issue).

Essentially, ecologic bias (aside from within-group bias) occurs because group affiliation or the average exposure level of the group affects disease occurrence independently of exposure status at the individual level (see "Ecological Bias"). The structural effects of such ecologic variables, if they can be separated from other effects at the individual level, might be informative, rather than just a source of error. Thus, by including both ecologic and individual-level predictors (possibly of the same exposure) in the analysis, we might enhance our understanding of disease occurrence. This type of contextual or multilevel analysis has been used extensively in social science research (*105–108*) but rarely in epidemiologic research (*109*). In addition, if the effect of a risk factor is known from previous research, the results of an ecologic analysis involving that risk factor could be used to evaluate the potential or realized impact of a population intervention, which may not be completely estimable at the individual level (*63*). A more profound understanding of ecologic bias, therefore, could yield benefits to other public-health research.

Gene–Environment Interactions. Because both genetic and environmental factors contribute to the etiology of most diseases, we would typically expect factors of each type to confound and/or modify the effect of the other. We know, for example, that a combination of both environmental/personal factors and genetic susceptibilities are sufficient for the development of certain diseases. Yet standard methods of epidemiologic research and population genetics have not been well integrated (*110*). As indicated in "Ecological Bias," we need new methods for incorporating environmental variables in genetic (e.g., linkage) analyses of pedigree data. We also need to understand better the relationship between those parameters estimated in pedigree studies and the effect parameters estimated in standard epidemiologic studies; and we need to understand better how the estimates of gene–environment interactions in pedigree studies are biased by confounding, measurement error, and family selection (ascertainment). With this understanding, we can devise new methods to prevent or control bias. Analogously, the use of family data in standard epidemiologic designs (e.g., history of disease and/or its risk factors in relatives) requires further development in order to handle differences in family size and composition among subjects. With the recent advances in molecular genetics, the integration of epidemiology and population genetics is likely to become more important in the future.

Sample Size and Power. As noted in "Problems in Environmental Epidemiology," epidemiologic studies of environmental exposures often require large sample sizes to detect risk-factor effects with sufficiently small statistical error. To address this concern, researchers are usually expected to justify their proposed sample size by estimating the power of their study for testing one or more major hypotheses (i.e., the probability of detecting an association of at least a certain magnitude with a designated Type I error—alpha level typically set at 5%). This is a rather straightforward procedure when the power estimation is applied to two dichotomous variables (exposure and disease) (*1,4,111*). Yet all observational studies require more complicated analyses to make causal inferences — e.g., to deal with polytomous, continuous, or time-dependent exposures; covariate adjustment; the assessment of interaction effects; matching; and other special design features. Although methods of power estimation do exist for many of these complicating features, they require additional specifications (assumptions) about which the investigator is not likely to have adequate information. Further development of these methods would be useful, therefore, to identify techniques that are both practical and informative in specific situations, including ecologic studies for which sample size requirements have received little attention.

One parameter the investigator must specify to justify the proposed sample size is the magnitude of effect expected in the data or the minimum effect regarded as important to detect. In the absence of previous epidemiologic studies involving similar exposure levels, the expected effect is generally specified rather arbitrarily (e.g., a

rate ratio of 2). Sometimes, however, there are exposure–response findings from animal studies or occupational studies with higher exposure levels, which could be used to estimate the environmental exposure effect expected in the base population. This approach, which also requires further development, might allow research funds to be allocated more judiciously.

Data Analysis. Many of the recent developments and ideas for new study designs and data collection that were discussed in this article require parallel developments in statistical methods. For example, the analyst might have to deal with complex sampling strategies (as in two-stage designs); missing, misclassified, and/or aggregated data on relevant variables; time-dependent covariates; lag periods between first exposure and disease detection; incomplete case detection; and a limited sample size that severely restricts the number of covariates treated simultaneously. Several of these issues are covered further by Hatch and Thomas and by Prentice and Thomas in this issue.

REFERENCES

1. Schlesselman JJ. Case–control studies: design, conduct, analysis. New York: Oxford University Press, 1982.
2. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. Belmont, CA: Lifetime Learning Publications, 1982.
3. Miettinen OS. Theoretical epidemiology: principles of occurrence research in medicine. New York: John Wiley & Sons, 1985.
4. Kelsey JL, Thompson WD, Evans AS. Methods in observational epidemiology. New York: Oxford University Press, 1986.
5. Rothman KJ. Modern epidemiology. Boston, MA: Little, Brown and Co., 1986.
6. Checkoway H, Pearce N, Crawford-Brown DJ. Research methods in occupational epidemiology. New York: Oxford University Press, 1989.
7. Leaverton PE, ed. Environmental epidemiology. New York: Praiger, 1982.
8. Chiazze L Jr, Lundin FE, Watkins D, eds. Methods and issues in occupational and environmental epidemiology. Ann Arbor, MI: Ann Arbor Science Publishers, 1983.
9. Goldsmith JR, ed. Environmental epidemiology: epidemiological investigation of community environmental health problems. Boca Raton, FL: CRC Press, 1986.
10. Kopfler FC, Craun GF, eds. Environmental epidemiology. Chelsea, MI: Lewis Publishers, 1986.
11. Poole C. Would vs should in the definition of secondary study base (letter). J Clin Epidemiol 43:1016–1017 (1990).
12. Miettinen OS. The concept of secondary base (reply). J Clin Epidemiol 43:1017–1020 (1990).
13. Morgenstern H, Kleinbaum DG, Kupper LL. Measures of disease incidence used in epidemiologic research. Int J Epidemiol 9:97–104 (1980).
14. Breslow NE, Day NE. Statistical methods in cancer research, vol. 1. The analysis of case–control studies. Lyon: International Agency for Research on Cancer, 1980; 50–51.
15. Rubin DB. Bayesian inference for causal effects: the role of randomization. Ann Stat 6:34–58 (1978).
16. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol 15:412–418 (1986).
17. Greenland S, Schlesselman JJ, Criqui MH. The fallacy of employing standardized regression coefficients and correlations as measures of effect. Am J Epidemiol 123:203–208 (1986).
18. Greenland S, Maclure M, Schlesselman JJ, Poole C, Morgenstern H. Standardized regression coefficients: a further critique and review of some alternatives. Epidemiology (in press).
19. Greenland S, Thomas DC. On the need for the rare disease assumption in case–control studies. Am J Epidemiol 116:547–553 (1982).
20. Greenland S, Thomas DC, Morgenstern H. The rare-disease assumption revisited: a critique of estimators of relative risk for case–control studies. Am J Epidemiol 124:869–876 (1986).
21. Armenian HK, Lilienfeld AM. Incubation period of disease. Epidemiol Rev 5:1–15 (1983).
22. Thomas DC. Pitfalls in the analysis of exposure-time-response relationships. J Chron Dis 40 (Suppl 2):71S–78S (1987).
23. Shy CM, Kleinbaum DG, Morgenstern H. The effect of misclassification of exposure status in epidemiological studies of air pollution health effects. Bull N Y Acad Med 54:1155–1165 (1978).
24. Fleiss JL. Statistical factors in early detection of health effects. In: New and sensitive indicators of health impacts of environmental agents (Underhill DW, Radford EP, eds). Pittsburgh, PA: University of Pittsburgh, Center for Environmental Epidemiology, 1986; 9–16.
25. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? Am J Epidemiol 132:746–748 (1990).
26. Susser M, Stein Z. Third variable analysis: application to causal sequences among nutrient intake, maternal weight, birthweight, placental weight, and gestation. Stat Med 1:105–120 (1982).
27. Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. Boston, MA: Houghton Mifflin Co., 1979.
28. Greenland S. Randomization, statistics, and causal inference. Epidemiology 1:421–429 (1990).
29. Buck C, Donner A. The design of controlled experiments in the evaluation of nontherapeutic interventions. J Chron Dis 35:531–538 (1982).
30. Cornfield J. Randomization by group: a formal analysis. Am J Epidemiol 108:100–102 (1978).
31. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. Am J Epidemiol 114:906–914 (1981).
32. Zelen M. A new design for randomized clinical trials. N Engl J Med 300:1242–1245 (1979).
33. Ast DB, Schlesinger ER. The conclusion of a 10-year study of water fluoridation. Am J Public Health 46:265–271 (1956).
34. Greenland S, Morgenstern H. Classification schemes for epidemiologic research designs. J Clin Epidemiol 41:715–716 (1988).
35. Greenland S. Response and follow-up bias in cohort studies. Am J Epidemiol 106:184–187 (1977).
36. Robins J. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. J Chron Dis 40(Suppl 2):139S–161S (1987).
37. Robins J. The control of confounding by intermediate variables. Stat Med 8:679–701 (1989).
38. Newman SC. Odds ratio estimation in a steady-state population. J Clin Epidemiol 41:59–65 (1988).
39. Miettinen OS. The case–control study: valid selection of subjects. J Chron Dis 38:543–548 (1985).
40. Greenland S, Morgenstern H. Matching and efficiency in cohort studies. Am J Epidemiol 131:151–159 (1990).
41. Thompson WD, Kelsey JL, Walter SD. Cost and efficiency in the choice of matched and unmatched case–control study designs. Am J Epidemiol 116:840–851 (1982).
42. Greenland S. Adjustment of risk ratios in case-base studies (hybrid epidemiologic designs). Stat Med 5:579–584 (1986).
43. Mantel N. Synthetic retrospective studies and related topics. Biometrics 29:479–486 (1973).
44. Kupper LL, McMichael AJ, Spirtas R. A hybrid epidemiologic study design useful in estimating relative risk. J Am Stat Assoc 70:524–528 (1975).
45. Prentice RL. A case–cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 73:1–11 (1986).
46. Miettinen OS, Wang JD. An alternative to the proportionate mortality ratio. Am J Epidemiol 114:144–148 (1981).
47. Wang JD, Miettinen OS. The mortality odds ratio (MOR) in occupational mortality studies—selection of reference occupation(s) and reference cause(s) of death. Ann Acad Med 13(Suppl 2):312–316 (1984).
48. Butler WJ, Park RM. Use of the logistic regression model for the analysis of proportionate mortality data. Am J Epidemiol 125:515–523 (1987).

49. Robins JM, Blevins D. Analysis of proportionate mortality data using logistic regression models. Am J Epidemiol 125:524–535 (1987).
50. Greenland S, Neutra R. An analysis of detection bias and proposed corrections in the study of estrogens and endometrial cancer. J Chron Dis 34:433–438 (1981).
51. Feinstein AR. Methodologic problems and standards in case–control research. J Chron Dis 32:35–41 (1979).
52. White JE. A two-stage design for the study of the relationship between a rare exposure and a rare disease. Am J Epidemiol 115:119–128 (1982).
53. Walker AM. Anamorphic analysis: sampling and estimation for covariate effects when both exposure and disease are known. Biometrics 38:1025–1032 (1982).
54. Cain K, Breslow NE. Logistic regression analysis and efficient design for two-stage studies. Am J Epidemiol 128:1198–1206 (1988).
55. Breslow NE, Zhao LP. Logistic regression for stratified case–control studies. Biometrics 44:891–899 (1988).
56. Flanders WD, Greenland S. Analytic methods for two-stage case–control studies and other stratified designs. Stat Med 10:739–747 (1991).
57. Louis TA, Lavori PW, Bailar JC III, Polansky M. Crossover and self-controlled designs in clinical research. N Engl J Med 310:24–31 (1984).
58. Maclure M. The case–crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol 133:144–153 (1991).
59. Caporaso NE, Hayes RB, Dosemeci M, Hoover R, Ayesh R, Hetzel M, Idle J. Lung cancer risk, occupational exposure, and the debrisoquine metabolic phenotype. Cancer Res 49:3675–3679 (1989).
60. Mack W, Langholz B, Thomas DC. Survival models for familial aggregation of cancer. Environ Health Perspect 87:27–35 (1990).
61. Susser E, Susser M. Familial aggregation studies: a note on their epidemiologic properties. Am J Epidemiol 129:23–30 (1989).
62. Claus EB, Risch NJ, Thompson WD. Age at onset as an indicator of familial risk of breast cancer. Am J Epidemiol 131:961–972 (1990).
63. Morgenstern H. Uses of ecologic analysis in epidemiologic research. Am J Public Health 72:1336–1344 (1982).
64. Ohno Y, Aoki K, Aoki N. A test of significance for geographic clusters of disease. Int J Epidemiol 8:273–281 (1979).
65. Mollie A, Richardson S. Empirical Bayes estimates of cancer mortality rates using spatial models. Stat Med 10:95–112 (1991).
66. Mahoney MC, Labrie DS, Nascam PC, Wolfgang PE, Burnett WS. Population density and cancer mortality differentials in New York State, 1978–1982. Int J Epidemiol 19:483–490 (1990).
67. Lee JAH, Petersen GR, Stevens RG, Vesanen K. The influence of age, year of birth, and date on mortality from malignant melanoma in the populations of England and Wales, Canada, and the white population of the United States. Am J Epidemiol 110:734–739 (1979).
68. Lee JAH. Melanoma and exposure to sunlight. Epidemiol Rev 4:110–136 (1982).
69. Wallenstein S, Gould MS, Kleinman M. Use of the scan statistic to detect time–space clustering. Am J Epidemiol 130:1057–1064 (1989).
70. Roberson PK. Controlling for time-varying population distributions in disease clustering studies. Am J Epidemiol 132:S131–S135 (1990).
71. Hatch M, Susser M. Background gamma radiation and childhood cancers within 10 miles of a U.S. nuclear plant. Int J Epidemiol 19:546–552 (1990).
72. Ostrom CW Jr. Time series analysis: regression techniques, 2nd ed., quantitative applications in the social sciences, 07-009. Newbury Park, CA: Sage Publications, 1990.
73. McDowall D, McCleary R, Meidinger EE, Hay RA Jr. Interrupted time series analysis. Quantitative applications in the social sciences, 07–021. Newbury Park, CA: Sage Publications, 1980.
74. Sayrs LW. Pooled time series analysis. Quantitative applications in the social sciences, 07–070. Newbury Park, CA: Sage Publications, 1989.
75. Catalano R, Serxner S. Time series designs of potential interest to epidemiologists. Am J Epidemiol 126:724–731 (1987).
76. Gruchow HW, Rimm AA, Hoffmann RG. Alcohol consumption and ischemic heart disease mortality: Are time-series correlations meaningful? Am J Epidemiol 118:641–650 (1983).
77. Darby SC, Doll R. Fallout, radiation doses near Dounreay, and childhood leukaemia. Br Med J 294:603–607 (1987).
78. Crawford MD, Gardner MJ, Morris JN. Changes in water hardness and local death-rates. Lancet 2:327–329 (1971).
79. Goodman LA. Some alternatives to ecological correlation. Am J Sociol 64:610–625 (1959).
80. Beral V, Chilvers C, Fraser P. On the estimation of relative risk from vital statistical data. J Epidemiol Community Health 33:159–162 (1979).
81. Greenland S. Divergent biases in ecologic and individual-level studies. Stat Med 11:1209–1223 (1992).
82. Selvin HC. Durkheim's suicide and problems of empirical research. Am J Sociol 63:607–619 (1958).
83. Langbein LI, Lichtman AJ. Ecological inference, series 07-010. Beverly Hills, CA: Sage Publications, 1978.
84. Firebaugh G. A rule for inferring individual-level relationships from aggregate data. Am Sociol Rev 43:557–572 (1978).
85. Richardson S, Stucker I, Hémon D. Comparisons of relative risks obtained in ecological and individual studies: some methodological considerations. Int J Epidemiol 16:111–120 (1987).
86. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. Int J Epidemiol 18:269–274 (1989).
87. Greenland S, Morgenstern H. Neither within-region nor cross-regional independence of exposure and covariates prevents ecological bias (letter). Int J Epidemiol 20:816–817 (1991).
88. Robinson WS. Ecological correlations and the behavior of individuals. Am Sociol Rev 15:351–57 (1950).
89. Duncan OD, Cuzzort RP, Duncan B. Statistical geography: problems in analyzing areal data. Westport, CT: Greenwood Press, 1961.
90. Greenland S, Morgenstern H. Ecological bias and confounding (reply). Int J Epidemiol 19:766–767 (1990).
91. Richardson S, Hémon D. Ecological bias and confounding (letter). Int J Epidemiol 19:764–766 (1990).
92. Rosenbaum PR, Rubin DB. Difficulties with regression analyses of age-adjusted rates. Biometrics 40:437–443 (1984).
93. Brenner H, Savitz DA, Jöckel KH, Greenland S. The effects of nondifferential exposure misclassification in ecological studies. Am J Epidemiol 135:85–95 (1992).
94. Greenland S. The effect of misclassification in the presence of covariates. Am J Epidemiol 112:564–569 (1990).
95. Savitz DA, Baron AE. Estimating and correcting for confounder misclassification. Am J Epidemiol 129:1062–1071 (1989).
96. Brenner H, Savitz DA, Greenland S. The effects of nondifferential confounder misclassification in ecologic studies. Epidemiology 3:456–459 (1992).
97. Stavraky KM. The role of ecologic analysis in studies of the etiology of disease: a discussion with reference to large bowel cancer. J Chron Dis 29:435–444 (1976).
98. Connor MJ, Gillings D. An empiric study of ecological inference. Am J Public Health 74:555–559 (1984).
99. Polissar L. The effect of migration on comparison of disease rates in geographic studies in the United States. Am J Epidemiol 111:175–182 (1980).
100. Morgenstern H, Horwitz SM, Berkman LF. Connections between epidemiology and health services research: a review of psychosocial effects on childhood morbidity and pediatric medical care use. J Ambulatory Care Management 9:33–45 (1986).
101. Walter, SD, and Irwig, LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. J Clin Epidemiol 41:923–937 (1988).
102. Marshall RJ. Validation study methods for estimating exposure proportions and odds ratios with misclassified data. J Clin Epidemiol 43:941–947 (1990).
103. Greenland S. Statistical uncertainty due to misclassification: implications for validation substudies. J Clin Epidemiol 41:1167–1174 (1988).
104. Spiegelman D, Gray R. Cost-efficient study designs for binary response data with Gaussian covariate measurement error. Biometrics 47:851–870 (1991).
105. Boyd LH Jr, Gudmund RI. Contextual analysis: concepts and statistical techniques. Belmont, CA: Wadsworth Publishing Co., 1979.
106. Lincoln JR, Zeitz G. Organizational properties from aggregate data: separating individual and structural effects. Am Sociol Rev 45:391–408 (1980).
107. Aitkin M, Longford N. Statistical modeling issues in school effectiveness studies. J Roy Statist Soc (Series A)149(Part 1):1–43 (1986).
108. Iversen GR. Contextual analysis. Quantitative applications in the social sciences, 07–081. Newbury Park, CA: Sage Publications, 1991.
109. Humphreys K, Carr-Hill R. Area variations in health outcomes: artefact or ecology. Int J Epidemiol 20:251–258 (1991).
110. Khoury MJ, Beaty TH, Flanders WD. Epidemiologic approaches to the use of DNA markers in the search for disease susceptibility genes. Epidemiol Rev 12:41–55 (1990).
111. Morgenstern H, Winn DM. A method for determining the sampling ratio in epidemiologic studies. Stat Med 2:387–396 (1983).

Methodologic Research Needs in Environmental Epidemiology: Data Analysis

Ross L. Prentice[1*] and Duncan Thomas[2]

[1]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104; [2]Department of Preventive Medicine, School of Medicine, University of Southern California, 1420 San Pablo, PMB B-201, Los Angeles, CA 90033

A brief review is given of data analysis methods for the identification and quantification of associations between environmental exposures and health events of interest. Data analysis methods are outlined for each of the study designs mentioned, with an emphasis on topics in need of further research. Particularly noted are the need for improved methods for accommodating exposure assessment measurement errors in analytic epidemiologic studies and for improved methods for the conduct and analysis of aggregate data (ecologic) studies. — Environ Health Perspect 101 (Suppl 4):39–48 (1993).

Key Words: Aggregate data studies, analytic data studies, carcinogenic models, exposure measurement error, meta-analysis, relative risk regression, validation substudies

Introduction

Nearly all study of the health consequences of environmental and lifestyle exposures in human populations is purely observational. This means that the validity of the comparison of disease rates between more exposed and less or nonexposed persons is dependent on the assumption that disease rates in the two groups are comparable in the absence of such exposure. This comparability assumption can be weakened somewhat by the measurement and accommodation of other factors that are associated with disease risk and that have a different distribution in the compared exposure groups. If such confounding factors are accurately measured and adequately acknowledged in the data analysis, it is then sufficient that in the absence of the exposure of interest, the groups being compared have common disease rates conditional on the values of the confounding factors. Lack of validity (i.e., bias) in testing or estimation can be expected if there are unidentified confounding factors, if the recorded confounding factors are measured with error, or if the treatment of individual confounding factors is inadequate (e.g., linear allowance for confounders having effects that are substantially nonlinear). Bias also can be introduced if the exposure variables of interest or the health effects under study are not measured accurately. In practice these sources of bias can be reduced, but it is unlikely that they will be completely eliminated.

The sources of bias mentioned here are the principal reasons why epidemiologic cohort studies, among others, may yield inaccurate and conflicting results. Concern about residual, uncontrolled confounding can never be completely eliminated in any nonexperimental study. Hence, such studies are most reliable for the detection of moderate to large health effects (e.g., increase in disease incidence by a factor of two or more among highly exposed persons) that are unlikely to be qualitatively affected by modest confounding. There is also a strong role for the replication of results in diverse populations that are presumed to have different potentials for severe confounding. It is worth noting that experimental studies also have important practical limitations in the context of environmental epidemiology. Data analysis methods for cohort studies with accurate and complete assessment of exposure variables, confounding factors, and potential health consequences are well developed, as summarized in "Exposure-Response Estimate in Cohort Studies," below.

Case–control studies in which exposure and confounding factors are assessed retrospectively are subject to all of the biases noted above, as well as to recall bias, which occurs when diseased individuals (cases) and disease-free individuals (controls) differentially recall their exposures, their confounders, or their health outcome. Aggregate data studies, referred to later in this paper as ecologic studies, attempt to relate the exposure and confounding factor experience of groups to their corresponding disease rates. Such studies may be subject to additional biases if the statistical model for the group disease rate does not equal the average of valid disease rate models for the individuals being aggregated.

Apparent disagreement between environmental epidemiologic studies can also arise, not from bias, but from lack of power combined with attention to point estimates rather than confidence limits. The ability to detect an association between the levels of an exposure variable or exposure history and the risk of a disease depends primarily on the observed number of disease events in the study sample, on the range of exposures in the sample, and on the strength of association between exposure and disease. The distribution of exposures in the study cohort or in the cohort from which cases and controls are selected for a case–control study also has important influences on study power. While random measurement error in (univariate) exposure assessment will not invalidate, under weak conditions, a test for the hypothesis that no association between exposure and disease exists, test power may be reduced considerably by such measurement errors. Also, estimates of dose–response parameters may be substantially distorted (usually biased downward), including the possibility of a loss of monotonicity of dose–response trends (*1*). Thus, the proper analysis and interpretation of environmental epidemiologic studies rely heavily on the investigator's assessment of the magnitude of both potential biases and study power in the absence of such biases. For practical reasons, the power of specific studies will often be rather low, and knowledge of disease mechanisms and measurement

This manuscript was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990 – September 1992.

*Author to whom correspondence should be addressed.

This work was supported partially by grant CA-53996 from the National Institutes of Health. The authors wish to thank Professor John Tukey, the members of the Environmental Working Group, and a reviewer for helpful suggestions.

properties will be too limited to place useful bounds on potential biases. Hence, there are important uses for formal tests of the equality of exposure–disease associations from two or more studies in differing populations and for techniques used in combining the results of several studies. This topic will be discussed in the section titled "Comparing and Combining the Results of Several Studies."

The following section describes statistical- and biological-based models that can serve as the basis for exposure–disease analyses.

Models for Disease Occurrence

The simplest cohort studies occur when exposure takes place in one instant, as in Japanese atomic bomb survivors, or is constant over the individual's lifetime, as in some animal inhalation experiments. However, most exposures, and most confounders, are complex functions of time and demand a more complicated mathematical description. Our discussion of descriptive disease occurrence models begins with the over-simplified case.

Let $\lambda_0(t)$ denote the instantaneous rate of occurrence of a study disease or other health-related event for subjects of age t who have not received the exposure of interest. This means that if N such persons, all at age t, were observed for a short time dt, the expected number of disease occurrences would be $N\lambda_0(t)dt$. If a person of age t received an exposure Z, the instantaneous occurrence rate would be altered from $\lambda_0(t)$ to $\lambda(t|Z)$, the (instantaneous) relative risk is $\lambda(t;Z) / \lambda_0(t)$. These rates are nonnegative and, provided neither is zero, one can take the logarithm of this relative risk. It is often convenient and useful to assume the logarithm of the relative risk to be a linear function of exposure and confounding factor measurements. This is equivalent to modeling the relative risk as an exponential function, $\exp(\chi\beta)$, where the vector $\chi = (\chi_1,\ldots,\chi_p)$, which replaces the more general Z, consists of carefully chosen (and usually incomplete) measures of exposure or confounding factors, with $\chi = (0,\ldots,0)$ corresponding to no exposure and standard values for confounders. The coefficients $(\beta_1,\ldots,\beta_p)$, regression coefficients that comprise the vector β (or, more precisely, its transpose β^T), then tell us about the impact of each χ_i on relative risk when the other χs are held fixed.

The result is a simple proportional hazards (or Cox) model

$$\lambda(t|Z) = \lambda_0(t)\exp(\chi\beta) \qquad [1]$$

which is used widely in the analysis of failure time data (*2*). In order to deal with complications inherent in most environmental epidemiologic studies, one must generalize this discussion and complicate the appearance of some formulae, but be careful not to change the essentials of the approach. Such generalization follows in the next subsection.

Descriptive Relative-Risk Models

As above, let $\lambda_0(t)$ denote the instantaneous rate of a study disease (or other health-related problem) at age t in the absence of the exposure of interest. A person of age t may have received exposures $z(u)$ at certain ages $u < t$. One can refer to $Z(t) = \{z(u),\ u < t\}$ as the person's exposure history up to age t. Furthermore, one can allow the vector $z(u)$ to include the values of confounding factors at age u, so that $Z(t)$ includes both exposure and confounding factor histories up to age t. The disease rate at age t is $\lambda\{t|Z(t)\}$, a function of this exposure and confounder history. The relative risk associated with history $Z(t)$ is then the ratio $\lambda\{t|Z(t)\}/\lambda_0(t)$. Because this ratio is nonnegative, it can be, and often is, modeled using an exponential function $\exp\{\chi(t)\beta\}$, where $\chi(t) = \{\chi_1(t),\ldots,\chi_p(t)\}$. This function consists of data-analyst-defined functions of $Z(t)$ and t, with $\chi(t) \equiv (0,\ldots,0)$ again corresponding to no exposure and standard confounder histories, while $\beta^T = (\beta_1,\ldots,\beta_p)$ is a corresponding vector of relative-risk parameters to be estimated.

This relative risk (RR) regression model

$$\lambda\{t|Z(t)\} = \lambda_0(t)\exp\{\chi(t)\beta\} \qquad [2]$$

also called the Cox-regression model or (inaccurately called) the proportional hazards model (*2–4*) or an approximation to these models, forms the basis for most descriptive analyses of environmental epidemiologic studies. As a simple example to illustrate the notation, consider the relationship between exposure to ionizing radiation and the rate of a certain cancer in the atomic bomb-exposed populations in Hiroshima and Nagasaki. One could define

$$z(u_0)^T = \{z_1(u_0), z_2(u_0)\}, \qquad [3]$$

as the gamma and neutron exposures for a person at age u_0 in 1945 when the exposure occurred and as $z(u) \equiv 0$ otherwise. A specification $\chi(t) \equiv z(u_0)$ then assumes a log-relative risk function that is linear in gamma and neutron exposure levels. The regression model can be relaxed to allow, for example, the relative risk to depend on age at exposure and time since exposure and to allow for nonlinear dependencies of the log-relative risk on gamma and neutron exposure.

As noted above, the histories of potential confounding factors can also be included in $Z(t)$, in which case $\chi(t)$ will include functions of both the exposures of interest and other factors, while product terms between the two will allow the relative risk associated with a given exposure history to depend on the value of other variables. This allowance is termed effect modification in epidemiologic parlance. Confounding factors may also be controlled by means of stratification rather than, or in addition to, regression modeling using the descriptive model

$$\lambda\{t|Z(t)\} = \lambda_{0s}(t)\exp\{\chi(t)\beta\} \qquad [4]$$

where the baseline rate $\lambda_{0s}(\cdot)$ is allowed to vary across a number of strata defined as functions of age (t) and confounding factor values.

Relative-risk forms other than exponential also may be considered in the above models. In particular, the linear form $1 + \chi(t)\beta$ often is felt to be theoretically and empiricaly more appropriate for certain carcinogenic exposures and has been used widely in radiation literature, sometimes with the addition of quadratic terms. Absolute rather than relative-risk models, such as

$$\lambda\{t|Z(t)\} = \lambda_{0s}(t) + \chi(t)\beta, \qquad [5]$$

also have been used in modeling radiation effects, although there is a consensus that it generally does not fit well without the addition of terms for the modifying effect of age at exposure and latency. It may also be useful for modeling certain rare diseases such as mesothelioma, for which the baseline rate in the absence of asbestos exposure is virtually zero. In all of these alternatives to the standard exponential relative-risk model, estimates of the relative-risk parameters and baseline rates are often found to have poor statistical properties. However, quite general programs that use likelihood-based methods to obtain appropriate confidence limits (*5*) are now available to fit a broad class of relative- and absolute-risk models with combinations of linear and exponential terms.

Suppose that the regression vector $\chi(t)$ in the above unstratified model consists only of functions of the exposure variable under study, and let $pr\{\chi(t)\}$ denote the probability density for value $\chi(t)$ in the source population of the modeled regression vector. In addition to estimating the relative-risk function, one may be interested in the fraction of the disease incidence at age t that may be attributed to

exposure. If the disease rate for all study subjects was reduced to the baseline rate $\lambda_0(t)$, then the overall incidence at age t would be reduced by the attributable proportion

$$AR(t) = \int \lambda_0(t) \left[\exp\{\chi(t)\beta\} - 1\right] pr\{\chi(t)\}\, d\chi(t) \Big/ \int \lambda_0(t) \exp\{\chi(t)\beta\}\, pr\{\chi(t)\}\, d\chi(t). \quad [6]$$

A similar expression can be written for the attributable proportion under the stratified relative-risk model.

In some applications of these relative-risk models, it is convenient to define the basic time variable t to be chronological time or time from entry into a certain cohort rather than age, which is accommodated through stratification or regression modeling. For example, in a cohort study with covariate information collected at specified points in chronological time, such a definition can help ensure comparability of the covariate (i.e., exposure and confounding) information on all study subjects at a given value of t.

There are distinct advantages in using hazard rates or instantaneous disease rates, $\lambda\{t|Z(t)\}$, in our formulae rather thanddisease rates over some specified age or time period, in part because the interpretation of these latter rates will depend on the duration of the age period or time period in question, which will vary inevitably from study to study. Nevertheless, in some studies one observes only whether disease occurs in a certain time period rather than the actual times or ages of disease occurrence. Let $D = 1$ denote disease occurrence during a prescribed disease ascertainment period for a study and $D = 0$ denote lack of occurrence. Ignoring issues such as competing risks and losses to follow-up, one may choose to model the disease probabilities $pr\{D = 1 | Z(t_0)\}$ by an exponential-form odds-ratio model in which

$$\frac{pr\{D = 1 | Z(t_0)\} \,/\, pr\{D = 1 | Z(t_0) = Z_0\}}{pr\{D = 0 | Z(t_0)\} \,/\, pr\{D = 0 | Z(t_0) = Z_0\}} = \exp\{\chi(t_0)\beta\}, \quad [7]$$

where $Z(t_0)$ denotes a subject's exposure and confounding factor history at age t_0 at the beginning of the ascertainment period, and Z_0 denotes the standard, or base, covariate history. This odds-ratio model can be rewritten as a logistic regression model

$$pr\{D = 1 | Z(t_0)\} = \exp\{a(Z_0) + \chi(t_0)\beta\} \Big/ \left[1 + \exp\{a(Z_0) + \chi(t_0)\beta\}\right], \quad [8]$$

where the function $a(Z_0)$ may, for example, be defined to take value a_s whenever the study subject falls in stratum s, which is defined as a function of potential confounding factor values at t_0.

The above relative-risk and odds-ratio models are purely descriptive models. Their application is intended as an aid for summarizing and displaying aspects of large, complex data sets. In some situations, such as a regulatory decision concerning the safe level of a certain exposure, it will be essential to bring to bear any available biologic or mechanistic knowledge on the inference problem. Such knowledge could be used, for example, to specify a form for the relative risk at age t as certain elements of $\chi(t_0)$ approach zero, where these elements capture the dosage, duration, or other aspects of the exposure in question. Similarly, knowledge or assumption about the pertinent biological mechanisms could be used to derive models for $\lambda\{t|Z(t)\}$ of forms other than those mentioned above. The next subsection overviews two classes of carcinogenesis models that have been proposed on mechanistic or biological grounds.

Mechanistic and Biologically Based Models

Efforts to describe a disease process in terms of deterministic or stochastic models have focused mostly on models for the spread of infectious diseases in a population and models for carcinogenesis. Some of the work on carcinogenesis models, as outlined below, may be pertinent to other diseases.

Much of the early work on mathematical models for cancer was reviewed in a classic paper by Armitage and Doll (*6*). Whittemore and Keller (*7*) also provide a comprehensive review. A major contribution of the Armitage and Doll paper is the use of the multistage model of carcinogenesis. This model is based on the assumptions that cancer results from a single cell line undergoing a series of discrete, heritable changes (e.g., point mutations, chromosomal breaks or translocations, or other types of copying errors) in a particular sequence, and the rates of such transitions do not depend explicitly on age, although they may be affected by exposure to carcinogens or by factors that modify the rate of cell division. As a consequence of these and some additional assumptions, it can be shown that the age-specific incidence rate is predicted to vary approximately as the $(k-1)^{st}$ power of age, where k is the number of transitions required (usually estimated to be about 5 to 7 for adult tumors). If a carcinogenic exposure occurs at a constant rate over time, the incidence will vary approximately as a polynomial function of dose rate of order equal to the number of dose-dependent transitions. If exposure is instantaneous or varies over time, the incidence rate will be modified by age at exposure and/or time since exposure, depending upon which stage(s) is dose-dependent.

Until recently, most of the empirical tests of these predictions have been done by fitting the model to aggregate data on population age-specific rates, or to broadly-grouped data on cohorts, stratified by dose, age at exposure, or time since exposure. A problem with this approach is the difficulty of separating the effects of dose rate, age at first exposure, duration of exposure, time since last exposure, and attained age, all of which influence the predictions of the model. Simple comparisons of one factor without controlling the other factors can be misleading. This is less of a problem when animal bioassay data are used, as these are usually limited to constant, lifetime, dose regimens. However, such data are not informative about whether the carcinogen acts at an early or late stage. Nevertheless, the approach has been used for risk-assessment purposes by many regulatory agencies. The default approach advocated by the U.S. Environmental Protection Agency (EPA) and others involves fitting the multistage model to available epidemiologic or toxicologic data and using an upper confidence limit on the estimated slope coefficient (scaled for species differences in weight and life span) to compute the lifetime excess risk in humans. The scientific and statistical validity of this approach is controversial (*8,9*).

With the development of general relative-risk models ("Descriptive Relative-Risk Models" above), it has become possible to test the multistage and other models by fitting them directly to data on individuals. This offers great advantages for dealing with time-dependent exposures, which are the most informative about the stage at which a carcinogen acts. This approach has been applied to data on occupational exposures to asbestos (*10*), arsenic (*11*), and benzene (*12*); on the atomic bomb survivors (*13*); and on smoking (*14*), with varying results. The three occupational applications all were consistent with a single stage of action (relatively late for asbestos and arsenic, early for benzene), while the radiation and smoking data both showed signs of two stages being affected.

The multistage model has several important limitations, including its inability to account

for leukemia and childhood cancers, the genetics of cancer, and the distinction between mechanisms of initiation and promotion. It also has been criticized for its need for as many as 5 to 7 stages to account for the steep age dependence, when only two or three have been established in experimental systems. Moolgavkar and Knudson (*15*) have proposed an alternative model that addresses these issues. This model assumes that two mutational events are required and the cell lines that have experienced the first event may be at a competitive advantage (proliferation) or disadvantage (repair) relative to normal cells. Carcinogens might act by affecting either mutation rates or proliferation rates. Major gene effects are accounted for by assuming that individuals who inherit the gene begin life with all cells in the intermediate stage. This model has been successful in fitting epidemiologic data on smoking (*15,16*), breast cancer (*17*), and radon (*18*). In the latter example, data from an experimental study of rats exposed to radon were fitted to the model and radon was found to have an effect on both the mutation and proliferation rates. However, the interpretation of this result is complicated by the authors' use, for both of these dependencies, of a power function dose–response relationship with a very low exponent rather than a simple linear dose–response. Thomas (*19*) has proposed a variant of this model that adds an additional stage to the process to try to explain the difference in the modifying effect of dose rate and the duration of exposure for different types of radiation; so far, no attempt has been made to test this model.

With the rapid growth in our understanding of the fundamental biology of cancer, further development of methods to validate these mechanistic ideas and, where appropriate, to incorporate them into the analysis of epidemiologic data would be worthwhile. Most of the models that have been considered seriously are sufficiently general that some parameter values can be found to provide an adequate fit to epidemiologic data sets. Thus, these models are not easily falsified as a class, and it is unlikely that one could choose among them on purely statistical grounds. Instead, their utility lies in the types of comparisons that can be made within the context of a particular model—whether a carcinogen acts at an early or a late stage in the multistage model or as an initiator or a promoter in the two-stage model, for example. Their real value, therefore, lies in their ability to organize a complex set of hypotheses into a unified framework and to suggest empirical tests, in populations of humans or animals, of mechanistic ideas suggested by observations at the cellular level. Research efforts to identify and measure the assumed biological entities on the pathway to cancer cell formation seem particularly well motivated.

Exposure–Response Estimation in Cohort Studies

Relative-Risk and Odds-Ratio Estimation

Consider the unstratified relative-risk regression model of "Relative-Risk Models." A cohort study involves the selection of a sample of individuals from the population under study, succeeded by a follow-up to observe disease occurrence. The relative-risk parameter β can be estimated by maximizing a partial likelihood function $L(\beta)$ that is a product over all disease occurrence times (ages) that appear in the sample of the ratio of the relative risk for the subject developing disease to the sum of the relative risks for all subjects at risk at that time (*20*). The corresponding likelihood function under the above stratified relative-risk model is simply the product over strata of the stratum-specific likelihood functions. Note that this estimation procedure is quite general in that exposure variables, confounding variables, and stratum assignments each can vary with follow-up time. The principal assumption underlying this estimation procedure requires the set of subjects at risk for disease at any follow-up time to be representative of the base population, conditional on the covariate history and stratum assignment. This assumption will be satisfied, for example, if study subjects are sampled randomly and independently from the study population, and if rates of censoring (e.g., losses to follow-up) at a given follow-up time depend most on the covariate histories and stratum assignments at that time. Also, under weak conditions, $L(\beta)$ can be manipulated as if it were an ordinary likelihood function for asymptotic inference on β (*21,22*). Various computer programs are available now for the estimation of β, and, therefore, also of the relative-risk process $\exp\{\chi(t)\beta\}$.

The score statistic $U(\beta_0)$, defined as the value at $\beta = \beta_0$ of the derivative with respect to β of $\log L(\beta_0)$, can be used to test $\beta = \beta_0$. If $\beta_0 = 0$ and $\chi(t)$ consists only of indicator variables to distinguish exposure groups, then $U(\beta_0) = U(0)$ is known as the log rank statistic. Other choices of $\chi(t)$ yield other familiar, censored data test statistics, including generalizations of the Wilcoxon statistic.

Suppose now that there is no possibility of early censorship in the cohort study throughout the follow-up. The odds-ratio parameter β in the logistic regression model of "Relative-Risk Models," along with the location parameters $\alpha_s = \alpha(Z_0)$, can be estimated by a likelihood function $L(\beta)$ that is simply the product over all study subjects of the logistic regression probabilities $pr(D = 1 \mid Z(t_0))$ for subjects developing disease, and one minus such probabilities for other study subjects. Computer programs are widely available for inference on β from this likelihood function. If there are few disease events in stratum s, it is preferable to eliminate α_s by conditioning on the number of such events prior to applying standard likelihood procedures for the estimation of β (*23*).

The likelihood functions just described may seem esoteric to readers not having a statistical background. The main point to note, however, is that estimation of relative-risk and odds-ratio parameters in the very flexible models of "Relative-Risk Models" is now routine, and suitable software is available. Of course, the odds-ratio parameter will approach the relative-risk parameter if the disease acquisition period dt becomes short. This occurs because the odds of disease,

$$pr\{D = 1 \mid Z(t)\} \,/\, \left[1 - pr\{D = 1 \mid Z(t)\}\right] \qquad [9]$$

then typically approaches $\lambda\{t \mid Z(t_0)\}dt$, from which the exponential-form odds ratio approaches a corresponding exponential-form relative risk with identical regression parameter.

Estimation of the relative-risk regression parameter β may be computationally demanding if there are many distinct disease incidence times and if the regression vector and stratum assignment depend on time. However, if each $\lambda_{0s}(t)$ is defined to be constant over a partition of the time axis and $\chi(t)$ is restricted to be constant within the elements of this partition, then β can be estimated in a computationally simple fashion using Poisson regression methods. See Preston et al. (*24*) for application of such methods to radiation dose–response estimation from the Hiroshima and Nagasaki cohorts.

Particular care is required if these estimation procedures are applied to cohorts having few cases or if most cases occur within a small portion of the overall range of exposures. Asymptotic formulae for interval estimation on β may then be inaccurate and more specialized procedures (e.g., resampling methods) may be required. In fact, there has been little study of cohort data configurations under which such asymptotic formulae will provide adequate approximations.

Kalbfleisch and Prentice (*3*) and Cox and Oakes (*4*) provide detailed accounts of the theory and application of relative-risk regression models.

Disease Rate Estimation and Graphical Models

Denote by

$$\Lambda_{0s}(t) = \int_{t_{0s}}^{t} \lambda_{0s}(u)du, \qquad [10]$$

the cumulative baseline disease rate in stratum s in the stratified model of "Relative-Risk Models" over the range of ages $t \geq t_{0s}$ represented in the cohort. A simple nonparametric estimator of $\Lambda_{0s}(t)$ can then be defined as the sum over all disease occurrence times in stratum s of the ratio of the number of stratum s failures to the sum of the relative risks for all subjects at risk in stratum s at that time, with all relative risks evaluated at that β which maximizes $L(\beta)$.

As with ordinary regression methods, model-checking procedures are important to the application of relative-risk and odds-ratio models. Such procedures naturally focus on the assumed relative-risk process, $\exp\{\chi(t)\beta\}$, because other aspects of the model essentially are nonparametric. For example, the postulated relative-risk function can be generalized by adding well-selected additional elements to $\chi(t)$ and testing the hypothesis that corresponding coefficients equal zero. Computationally feasible methods also have been developed for approximating the influence of each study subject or each age group on β-estimation, in order to highlight questionable data points and to highlight vulnerabilities of the inference to model assumptions (*25*). Graphical procedures particularly are useful. In addition to the usual types of plots of influence (i.e., sensitivity) values and residuals, plots of separate estimates of $\Lambda_{0s}(t)$ for subsets of the cohort can provide useful visual checks on proportionality and other relative-risk assumptions (*3*).

The fact that the baseline rates $\lambda_{0s}(\cdot)$are unrestricted is an important source of robustness in respect to β-estimation. Specifically, relative-risk estimation is unlikely to be affected much if the intensity of ascertainment of disease events in the cohort varies somewhat across time or among strata. Similarly, location shifts in the modeled regression vector $\chi(t)$ across different values of t would not affect β-estimation in the exponential-form relative-risk model. However, more general measurement error in the ascertainment of $\chi(t)$ may have a profound effect on relative-risk estimation.

Measurement Error in Exposure Variables and Confounding Factors

Epidemiologists have long recognized that errors in the measurement of the study variables, including misclassification in the case of categorical variables, can lead to biased tests and estimates of the associations under study. Measurement error in the exposure histories or confounding factor histories may be of particular importance in environmental epidemiologic applications. Unfortunately, the methodology for avoiding bias due to measurement error is still at a rudimentary stage of development.

Consider the unstratified relative-risk regression model of "Relative-Risk Models" and suppose that rather than the covariate history $Z(t)$ one observes an estimate $W(t)$. The disease rate function at age (or chronological time) t, given the observed covariate history $W(t)$ can then be written (*26*)

$$\lambda\{t;W(t)\} = \lambda_0(t)E\left[\exp\{\chi(t)\beta\} \mid W(t)\right], \qquad [11]$$

where the expectation also is conditional on lack of disease occurrence or censorship prior to t. In fact, this induced relative-risk model also requires

$$\lambda\{t;Z(t),W(t)\} = \lambda\{t;Z(t)\} \qquad [12]$$

so that the $W(t)$ is unrelated to disease risk, given the true covariate history $Z(t)$. Unfortunately, the expectation in $\lambda\{t;W(t)\}$ generally depends on the baseline rates $\lambda_0(u)$, $u \leq t$, which complicates the estimation. However, in cohort studies in which the cumulative probability of disease occurrence is small, this dependence usually can be ignored and estimation of β can be based on a likelihood function in the form described above upon specifying a measurement error distribution for $\chi(t)$ given $W(t)$, from which $\lambda\{t;W(t)\}$ can be calculated.

Specification of the distribution of $\chi(t)$ given $W(t)$ would seem to be a hazardous undertaking unless there is a subsample in which both $Z(t)$ and $W(t)$ are available. In the presence of such a validation sample, simultaneous inference on relative-risk parameters and measurement error distribution parameters is possible (*27*), though further development is necessary before such estimation can be viewed as routine. More difficult issues arise if a true validation sample is not available. A reliability sample, in which separate estimates $W_1(t)$ and $W_2(t)$ of $Z(t)$ are obtained on a cohort subsample at two (or more) points in time, permits insight into some aspects of measurement error distribution, but additional strong assumptions are required for the estimation of β.

Even if the exposures under study are precisely estimated and pertinent confounding factors are identified, severe confounding may occur if confounding factor histories are measured with error (*28*), as is obvious if one considers an extreme situation in which measurement error produces a totally useless confounding factor estimate. This bias is likely to be more acute if the exposure and confounding factor values appearing in $X(t)$ are highly correlated.

A hypothetical cohort study of prenatal exposure to passive smoking in relation to the risk of lower respiratory disease during the first 3 years of life provided illustration in Morgenstern and Thomas, this volume. Any elevation in the odds of lower respiratory disease among more heavily exposed neonates may be severely attenuated by inaccuracies in exposure assessment in such a study. An analysis that controls for passive smoke exposure during the first 3 years of life, an exposure that would often be highly correlated with prenatal exposure, may be dominated by measurement error and be totally unreliable. A more practical illustration of the impact of measurement error is seen in the analysis of the mortality rates of various cancers in relation to gamma and neutron exposures in the Hiroshima and Nagasaki cohorts. Individual exposure estimates were constructed based on each study subject's location and shielding information as early as 1960. These estimates have continued to be refined in succeeding years through the use of improved models for the yields of the two bombs and more sophisticated models for the formation, transmission, and attenuation of gamma and neutron radiation. Many of the analyses of these cohorts simply combine gamma and neutron exposures into a single total dose estimate. The corresponding cancer mortality analyses have been affected somewhat by the changes in total dose estimates from one dosimetry system to the next (e.g., in the magnitude of elevated relative risks and the apparent shape of the dose-response curves), whereas analyses that attempt to estimate simultaneously the effect of gamma and neutron exposures on relative risk have been completely changed by dose estimate modifications. This illustrates the difficulty of reliably estimating exposure–disease associations when there are two or more exposure variables that are each measured with error (random or systematic) or, analogously, when there are exposure and confounding variables each measured with error. Very similar issues arise in epidemiologic studies of nonenvironmental factors; for example, they arise in attempts to separate the effects of fat and calories on cancer risk in nutritional epidemiology, or to separate the effects of

types of fat by degree of saturation on cancer risk in nutritional epidemiology (*29*).

Some recent work has concentrated on developing methods to adjust associations for the effects of measurement errors when their distributions are known. A very general framework for attacking this problem has been outlined by Clayton (*30*), who specifies the problem in terms of component models: the disease model describes the dependence of disease risk on true exposures and other factors; the measurement error model describes the relationship between true and measured exposures and any modifying factors; and the exposure model describes the population distribution of true exposures. These three models are combined in a maximum likelihood framework, and approaches to estimating the parameters of the disease model are described. Unfortunately, the approach is mathematically intractable in its general form, but useful progress has been made in some special cases. For categorical variables, Greenland and Kleinbaum (*31*) described a method based on applying the inverse of a matrix of known misclassification rates to the subject counts by measured exposure and disease classifications. Hui and Walter (*32*) have considered the case in which replicate measurements of exposure are available, and they use a form of log-linear model for the resulting four-way contingency table (counting true exposure as an unobserved dimension). For continuous variables, Prentice (*26*), Pierce et al. (*33*), Whittemore (*34*), Sposto et al. (*35*), and others have discussed approaches that replace the measured doses with empirical Bayes estimates of the true dose and use these in standard analyses. For a general review of these approaches, see Armstrong (*36*) and Thomas et al. (*37*). Another recent development involves combining nonparametric density estimation techniques with a computational device known as Gibbs sampling to overcome the tractability problems in the Clayton approach and avoid the need for parametric assumptions about the distribution of true doses. This method has been applied to data on studies of leukemia and thyroid disease in Utah residents downwind of the Nevada Test Site (*38,39*). These approaches are in an early stage of development, but they offer the prospect of removing the bias due to misclassification, correcting the shapes of dose–response curves, adjusting for covariates, and examining interaction effects, all while allowing for the additional uncertainties due to uncertainties in exposure estimates. Further developments along these lines are highly desirable.

Most of the literature on correcting for measurement errors has assumed that the misclassification rates were known and were constant across subjects. In practice, only estimates of these error distributions are available, either from earlier validation studies, from replicate measurements, from gold standard measurements on a subset of the subjects, or from theoretical uncertainty analysis. Methods need to be developed to account for uncertainties in the estimates of these misclassification rates (*40*). As a design issue, the optimal allocation of resources between high-quality measurements on a subset and larger numbers of approximate measurements should be considered (*41,42*). A unique aspect of the Utah fallout studies is the availability of individual-specific uncertainty estimates based on elaborate sensitivity analyses of the exposure pathways. This has allowed subjects with more precise exposure estimates to be given heavier weight in the analysis. Whether such efforts are warranted in terms of improved precision needs to be considered.

In summary, covariate measurement errors can bias severely the results of environmental epidemiologic studies. Improved analytic methods for accommodating random, nondifferential covariate measurement errors are required. Such methodologic developments might naturally focus on the potential for obtaining a true validation sample, on validation study design, and on the incorporation of validation study data in the overall estimation procedure (*27*).

Exposure–Response Estimation Under Case–Control and Other Sampling Procedures

Relative-risk and odds-ratio estimation often can be carried out more economically by sampling only subjects developing the study disease (the cases) or a random sample thereof, along with a suitably matched sample of subjects without disease (the controls). Typically covariate histories $Z(t)$, where t is the age (time) of case or control ascertainment, then have to be obtained retrospectively.

Consider the stratified relative-risk model of "Relative-Risk Models" and suppose that each case has one or more randomly selected controls that are matched on age at ascertainment (t) and stratum (s). Given the covariate histories $\{Z_1(t),\ldots, Z_m(t)\}$ for a case and its (m–1) age- and stratum-matched controls, the probability that exposure history $Z_1(t)$ corresponds to the case is simply the relative risk at t for the case divided by the sum of such relative risks for the m-matched subjects (including the case). Hence, the relative-risk parameter β can be estimated by maximizing the likelihood function $L(\beta)$, which is formed by multiplying these ratios for all matched case–control sets (*43*). To avoid strict matching on (t,s), relaxations of this sampling scheme are possible.

Similarly, the odds-ratio parameter β in the logistic regression model of "Relative-Risk Models" can be estimated under case–control sampling by maximizing the resulting logistic regression likelihood function by acting as though a prospective study had been conducted, though the estimates of αs no longer reflect disease incidence probabilities (*23*). In fact, the baseline rates $\lambda_{0s}(\bullet)$ and α_s in the relative-risk and odds-ratio regression models of "Relative-Risk Models" cannot be identified from case–control data in the absence of additional information on case and control sampling fractions.

In general, relative-risk and odds-ratio parameter estimates from case–control studies will be subject to the same biases as cohort studies. They also may be subject to recall bias if exposures or other covariate histories are differentially recalled by cases and controls or if they involve measurements that are affected by disease occurrence or its sequelae. There are often various practical steps that can be taken to minimize bias in ascertaining the covariate histories $Z(t)$ (e.g., interviewers blinded to case or control status), but usually it is not possible to identify residual recall bias because the requirement to obtain prediagnosis and postdiagnosis covariate histories on a sufficient sample of cases would often eliminate much of the efficiency of the case–control design.

As with the cohort study design, nondifferential measurement errors lead to the expectation

$$E\left[\exp\left\{\chi(t)\right\} \mid W(t)\right], \qquad [13]$$

where $\chi(t)$ is the true and $W(t)$ is the measured regression vector at age t, as the identifiable relative-risk function under age- and stratum-matched case–control sampling. To the extent that a representative validation sample can be ascertained retrospectively, there will be a potential to conduct valid relative-risk estimation from this type of study without making further assumptions.

A case–cohort (case–base) sampling procedure can also be considered as a means of reducing the cost or simplifying the logistics of a cohort study. With this design, covariate histories $Z(t)$ are assembled only

for cases and a (stratified) random sample of the study cohort. This sampling procedure has advantages if several end points (diseases) are to be studied in relation to an exposure. Also, the subcohort may be used to monitor exposures and other variables during the study's follow-up. However, estimation may be less efficient than estimation based on a case–control study with a comparable number of study subjects if cases and subcohort members are not well matched (*44,45*), and recall bias typically will be an issue. Prentice (*46*) has developed a procedure for estimating the relative-risk and odds-ratio parameters from case–cohort samples, and, in contrast to case–control sampling, baseline rates also can be estimated without external information. Comparisons and refinements of these sampling procedures are worthwhile research activities. Note also that the use of so-called two-stage designs (*47,48*) can lead to further valuable efficiency gains in some case-control study applications.

Exposure–Response Estimation in Aggregate Data (Ecologic) Studies

As discussed previously, sometimes it will be economical and convenient to examine an exposure–disease association by relating the disease rates among several groups of individuals to aspects of the exposure experience of each group. Such studies can be referred to as aggregate data studies since they involve the disease rates and exposures for the aggregate, rather than for individuals. These studies also are commonly referred to as ecological studies since groups having differing exposure histories are sometimes defined on an ecologic or geographic basis.

Denote by $\lambda_{ki}(t)$ the age- and sex-specific disease rate in the k^{th} group during (chronological) time period t. A multiple group study involves the analysis of estimates of $\lambda_{ki}(t)$, $k = 1,\ldots,K$ during a fixed time period; a time trend study involves estimates of $\lambda_{ki}(t)$, $t = 1,\ldots,T$ in a single population, while a mixed study involves estimates of $\lambda_{ki}(t)$ at several values of both k and t. An exponential-form relative-risk model for $\lambda_{ki}(t)$ can be written, in the notation of "Relative-Risk Models," as

$$\lambda_{ki}(t) = \lambda_{k0}(t)\exp\{\chi_{ki}(t)\beta\}, \quad [14]$$

from which the average disease rate $\lambda_k(t)$ for the $n_k(t)$ individuals in group k during time period t is

$$\lambda_k(t) = \lambda_{k0}(t)\Big[\sum_{i=1}^{n_k(t)}\exp\{\chi_{ki}(t)\beta\}/n_k(t)\Big] = \lambda_{k0}(t)\exp\{\bar{\chi}_k(t)\beta\}$$

$$\Big[\sum_{i=1}^{n_k(t)}\exp\{d_{ki}(t)\beta\}/n_k(t)\Big], \quad [15]$$

where

$$\bar{\chi}_k(t) = n_k^{-1}(t)\sum\chi_{ki}(t) \quad [16]$$

and

$$d_{ki}(t) = \chi_{ki}(t) - \bar{\chi}_k(t) \quad [17]$$

Let $y_k(t)$ denote the observed age- and sex-specific disease incidence rate in group k during time period t, as may be available from a disease register or other admininstrative source. From the above expression for $\lambda_k(t)$, one expects a regression of $\log y_k(t)$ on $\bar{\chi}_k(t)$ for various values of k or t (or both) to yield biased extimates of the relative-risk parameter β, because of the influences of the residuals $d_{ki}(t)$, even if the logarithms of the baseline rates $\lambda_{k0}(t)$ can be regarded as independent random variables with a common mean. This specification bias will be small if the $d_{ki}(t)$ values are small, that is, if the exposure and other regression variables have little variation within groups. Such bias presumably can be reduced by extending the regression equation to include averages of squares and of higher powers of the $d_{ki}(t)$ terms, though there does not appear to have been specific study of this approach. A closely related approach would replace the exponential-form relative-risk model by a linear-form model, so that

$$\lambda_{ki}(t) = \lambda_{k0}(t)\{1 + \chi_{ki}(t)\beta\} \quad [18]$$

and

$$\lambda_k(t) = \lambda_{k0}(t)\{1 + \bar{\chi}_k(t)\beta\} \quad [19]$$

from which the regression of $y_k(t)$ on $X_k(t)$, under certain random-effects assumptions on the baseline rates $\{\lambda_{k0}(t)\}$, will yield valid estimates of the linear relative-risk parameters (*49*). Note, however, that an exponential-form relative-risk model often might be more parsimonious than a linear-form model in environmental epidemiologic applications so that the regression vector in a linear relative-risk model may need to be quite lengthy and involve, for example, the average of product terms between exposure and potential confounding factors in order to adequately describe the data. In a multigroup study, it may be sensible to assume the $\lambda_0(t)$ terms are independent random variables with a common mean for $k = 1,\ldots,K$, thought it often may be useful to allow for the possibility of correlation among groups in a similar geographic area. In time-trend and mixed studies, however, it will typically be essential to model, or otherwise accommodate, the correlation structure among $\lambda_{k0}(t)$, $t = 1,\ldots,T$ at any fixed k. Inadequate modeling of the $\{\lambda_{k0}(t)\}$ may lead to aggregation bias.

These types of data analysis methods have received very little attention in the scientific literature and constitute an important gap in the collection of methods pertinent to environmental epidemiologic applications.

Aggregate data studies involving the simple linear regression of disease rates or the logarithm of disease rates on average exposures and average values of potential confounding factors can often be conducted quickly and cheaply and can play a useful role in hypothesis generation. It is obvious, however, that more comprehensive data sources and more sophisticated data analyses typically will be required if aggregate data studies are to contribute reliably to the identification and estimation of exposure-disease associations. Better data could come from randomly sampling each of the compared groups in order to obtain estimates, $X_k(t)$ of acceptable precision for use in a linear relative-risk model or to obtain estimates of the average of $\exp\{X_{ki}(t)\beta\}$, $i = 1,\ldots,n_k(t)$ for use in an exponential relative-risk model. Random measurement error in the ascertainment of individual exposure and confounding factors could impact substantially survey design. Better data analyses may arise from the application of so-called marginal methods (*50,51*) to mean and covariance models for the set of $y_k(t)$ or $\log y_k(t)$ values being analyzed.

Most effort to date concerning aggregate data studies has been directed to identifying the biases that may arise from aggregation, confounding, and other sources (*52,53*). It seems timely to direct a major effort to the development of procedures to prevent (or greatly reduce) such biases and, hence, to evaluate whether aggregate data studies can play a more fundamental and useful role in environmental epidemiologic studies and in epidemioligic research more generally.

Comparing and Combining the Results of Several Studies

Studies of a certain exposure–disease association may, for a variety of practical reasons, be lacking in power, and they may be subject to biases that can differ according to the population under study, the type of study design, and the rigor of the investigation. It follows that tests of agreement among the results of various studies and the formal combining of results from pertinent studies can play an important role in an overall exposure–disease association assessment.

Under ideal conditions, each of the types of studies described above can yield a valid estimate $\hat{\beta}$ of the logarithm of the relative risk associated with a specified exposure history, as well as an estimate $\hat{\sigma}^2$ of its variance. The

logarithm is used here, because its estimate is likely to adhere more closely to a normal distribution (with mean β) than the estimate $e^{\hat{\beta}}$ of the relative risk itself. Suppose m-independent studies yield (scalar) log-relative risk estimates of $\hat{\beta}_1,\ldots,\hat{\beta}_2$ with corresponding variance estimates $\hat{\sigma}_1,\ldots,\hat{\sigma}_2$

$$\tilde{\beta} = \sum \hat{\sigma}_i^{-2} \beta_i / \sum \hat{\sigma}_i^{-2} \qquad [20]$$

estimates a weighted mean of β_i's, which reduces to a common β if all β_i's are identical. To obtain the most stable estimate of this common mean, one can follow developments arising from Cochran's (*54*) introduction of partial weighting, thereby avoiding weights $\hat{\sigma}_i^{-2} \neq \sigma_i^{-2}$, which may be too small.

If all the β_is are the same and the $\hat{\beta}_i$ are independent and normally distributed, then

$$X^2 = \sum_{i=1}^{m} \hat{\sigma}_i (\hat{\beta}_i - \tilde{\beta})^2 \qquad [21]$$

will have a chi-square distribution with $m-1$ degrees of freedom, thereby giving a simple test of "all $\beta_i = \beta$" (assuming each $\hat{\beta}_i$ is distributed normally). If the β_i's are not identical, then a t-procedure can be used to set confidence limits for the weighted mean

$$\hat{\beta} = \sum \hat{\sigma}_i^{-2} \beta_i / \sum \hat{\sigma}_i^{-2} \qquad [22]$$

Confidence limits on $\hat{\beta}$ are approximately

$$\tilde{\beta} \pm t_\nu \left(\sum \hat{\sigma}_i^{-2} \right)^{-\frac{1}{2}} \qquad [23]$$

where t_ν is a critical value of t on ν (somewhat less than m) degrees of freedom. These limits are often conservative, particularly when the $\hat{\beta}_i$ follow longer-tailed distributions.

There are various reasons why the chi-square test described may provide evidence of heterogeneity of the relative-risk estimates from the m studies. For example, studies of the same type (e.g., m-cohort studies) may have differentially controlled for confounding or may have defined and measured exposure differently. Studies of different types (e.g., m-cohort, case–control and aggregate studies) have different sources of potential bias, for example, recall bias for case–control studies and aggregation bias in ecologic studies. Hence, it may be useful first to contrast and combine studies of the same type and then to examine whether the summary estimates of β from each study type are heterogenous. In respect to studies of the same type, the overview, or metanalysis, may be strengthened by analyzing the raw data from each study in a uniform format, which would maximize their comparability in terms of confounding control and exposure modeling. A fundamental principle of such analyses is that the parameter estimate $\hat{\beta}$ is based only on the combination of within-study information, as is the case for the heterogeneity test and the log-relative risk estimate described above.

Measurement error in exposure and in covariate assessment may be a particularly important source of heterogeneity among relative-risk estimates. For example, random measurement error may attenuate severely or otherwise distort relative-risk estimates in a cohort or case–control study if, for example, exposure assessment is based on data provided by individual interviews (e.g., location and shielding information in the Hiroshima and Nagasaki cohorts), but such attenuation may not be an issue in an aggregate data study if the desired averages (see "Exposure–Response Estimation in Aggregate Data Studies") can be estimated precisely. In this circumstance, some effort to deattenuate the analytic study relative-risk estimates, or to attenuate equally the aggregate data relative-risk estimates, is essential prior to the comparison of these estimates. See Prentice and Sheppard (*55*) for a recent attempt to study the consistency of international disease rate, time-trend, case–control and cohort studies in the dietary carcinogenesis area. Note also that $\tilde{\beta}$ will be biased as an estimator of β if the available log-relative risk estimates $\hat{\beta}_1,\ldots,\hat{\beta}_m$ are a biased sample of estimates from existing studies, which may arise if there is so-called publication bias in which relative-risk estimates that are significantly different from unity are more likely to be reported in the scientific literature. See Yusuf et al. (*56*) for a discussion of some issues in the conduct of such metanalyses.

Other Data Analysis Topics

The above presentation emphasized time to disease endpoints and corresponding relative-risk and odds-ratio models. In some areas of environmental epidemiologic research (e.g., respiratory epidemiology or neuroepidemiology), important endpoints are continuous. Much of the corresponding data analysis methodology is well established and does not need to be discussed here. However, methods for handling measurement error with continuous data (*57*) also require much additional development. Recent advances in the methods for analysis of longitudinal data (*50*) for discrete or continuous data are also quite relevant to the analysis of certain types of environmental epidemiologic data.

Preceding sections also have not addressed the simultaneous analysis of two or more endpoints. For example, in respiratory epidemiology, there may be several measures of lung function, and a data analysis goal may be to summarize exposure effects over several correlated measures of change in lung function. The estimating equation approaches mentioned above (*50,51*) provide an approach to such problems with discrete or continuous outcomes, but work could be done to compare these methods to univariate methods based on some summary endpoint. Methods for the analysis of correlated failure time data currently are not well established, though much statistical research is underway presently. See, for example, Clayton and Cuzick (*58*), Wei et al. (*59*), and Prentice and Cai (*60*) for recent contributions. Correlated failure–time methods also are required for the investigation of genetic factors or gene–environment interactions under certain types of study designs. For example, in a pedigree cohort study, it typically will be essential to allow for dependence between the disease occurrence times of family members when studying environmental exposure effects in relation to genetic indicators of susceptibility.

Morgenstern and Thomas, in this volume, mention certain designs other than those discussed thus far in this article, as well as the use of biomarker endpoints. Corresponding data analysis issues and methods will be mentioned only briefly here.

It was noted that experimental designs are practical occasionally in environmental epidemiologic research. The relative-risk and odds-ratio regression methods described above apply equally well for the comparison of disease incidence (or mortality) rates between randomization groups in individually randomized designs. However, a group-randomized design (e.g., with community as the unit of randomization) is more likely to be feasible, in which case it is essential to acknowledge the possibility of correlation among the responses (e.g., disease incidence times) of subjects in the same randomization group, which require the use of the type of correlated failure-time methods mentioned above.

In the discussion of ecologic designs it was noted that descriptive studies of the clustering of disease (e.g., in space or time) can play a useful role in the generation of environmental health hypotheses. These types of studies also have specialized data-analytic issues and methods. Statistical analysis has little to offer in the event of an isolated cluster discovered by ad hoc methods. Clusters within which the disease counts substantially

exceed expected counts perhaps are best addressed by direct fieldwork to identify a putative cause. On the other hand, hypotheses of a general tendency to cluster can be addressed statistically by using methods that compare the number of cases in certain neighborhoods of each case to the expected number of cases, while also taking account of population density. Local neighborhood tests also are available with case–control sampling. See Rothman (*61*) and other papers in this volume for discussions of disease-clustering methods.

The design chapter (in this issue) also emphasizes cross-sectional studies for the estimation of prevalence rates. The logistic regression methods outlined in "Relative-Risk and Odds-Ratio Estimation" may be used to relate prevalence probabilities to retrospectively obtained exposure and confounding factor histories. Of course, such prevalence probabilities reflect aspects of both disease incidence and disease duration, and therefore, may be difficult to interpret. Keiding (*62*) provides a comprehensive discussion of the relationships between prevalence probabilities, incidence rates, and disease durations and of the possibility of deriving estimates of age-specific incidence from cross-sectional studies.

As discussed previously, biomarkers may serve usefully as exposure indicators or as early indicators of disease (see Hatch and Thomas, this volume). An example of a biomarker as an intermediate endpoint is seen in chromosomal abnormalities in the radiation-exposed cohorts of Hiroshima and Nagasaki. The rates of such abnormalities among long-lived lymphocytes (usually 100 cells examined for each subject) have played a useful role in assessing the health effects of radiation exposure in these populations. The correlation among the chromosomal events in cells from the same study subject has a strong influence on dose–response analyses in this application (*35,63*). Recent advances in the ability to study the cellular and molecular mechanisms involved when responding to exposure and disease pathogenesis will lead inevitably to greater use of biomarkers and biological measurement in environmental epidemiologic studies. Hence, data analysis methods that incorporate such measurements in a biologically meaningful fashion are required. Suitable methods for dose–response analysis with biomarker endpoints will vary according to the type of endpoint(s) involved. Recent estimating equation approaches (*50,51*) often may be useful for such analyses. Circumstances under which a biomarker endpoint can substitute for disease occurrence and yield valid dose-response tests and estimates is also of considerable interest. See Prentice (*64*) for the introduction and discussion of such criteria.

Finally, it seems worth noting that the interpretation of relative-risk estimates from a study may depend on prior knowledge and on study goals. For example, if such estimation takes place in the context of a study specifically designed to confirm a particular association, the corresponding tests and confidence intervals are more appropriately taken at face value than if the relative risk is estimated in a purely exploratory context wherein various other exposures also are examined in relation to disease risk. In this latter situation, formal methods may be used to acknowledge the multiple hypotheses being examined, but precise statistical methods for doing so in a general way are not available. (So-called Bonferroni methods are available widely and may be precise enough.) Also, one is often neither in a purely exploratory nor a purely confirmatory mode in data analysis.

Summary Recommendations

Perhaps the single most important data analysis research need in environmental epidemiology concerns the development of improved methods to accommodate measurement errors in exposure assessment. Efforts aimed at the design and use of validation studies would be particularly useful, as would studies to document the scope and magnitude of measurement error influences.

A second important need concerns improved methods for the conduct and analysis of aggregate data (ecologic) studies. The development of strategies for controlling potential confounding, particularly by using individual surveys in multigroup studies, along with corresponding innovative data analysis methods, will be important. Empirical studies that illustrate various analytic and aggregate data analyses of real data sets also would be valuable.

Other pertinent topics for data analysis research include the development of improved methods for meta-analyses when studies of different types with differing potential for measurement error biases are available, the development of flexible data analysis methods, and the study of properties of analyses based on biomarker indicators of exposure or biomarker end points. Studies that evaluate and compare strategies for the control of confounding also merit continuing attention in environmental epidemiology as in other observational research areas. Further work on biologically based mathematical models for cancer and for other disease also would be well motivated.

REFERENCES

1. Dosemeci M, Wacholder S, Lubin J H. Does nondifferential misclassification of exposure always bias a true effect toward the null value? Am J Epidemiol 132:746–748 (1990).
2. Cox DR. Regression models and life tables (with discussion). J R Statist Soc B 34:187–220 (1972).
3. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York: Wiley, 1980.
4. Cox DR, Oakes D. Analysis of survival data. London: Chapman and Hall, 1984.
5. Preston DL, Lubin JH, Pierce DA. Epicure: risk regression and data analysis software. Seattle, Washington: HiroSort International Corp, 1991.
6. Armitage P, Doll R. Stochastic models for carcinogenesis. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol 4 (Neyman J, ed). Berkeley: University of California Press, 1961; 19–38.
7. Whittemore AS, Keller JB. Quantitative theories of carcinogenesis. SIAM Rev 20:1–30 (1978).
8. Freedman D, Navidi W. Multistage models for carcinogenesis. Environ Health Perspect 81:161–188 (1989).
9. Freedman DA, Zeisel H. From mouse to man: the quantitative assessment of cancer risks (with discussion). Statistical Science 3:3–56 (1988).
10. Thomas DC. Statistical methods for analyzing effects of temporal patterns of exposure on cancer risks. Scand J Work Environ Health 9:353–366 (1983).
11. Brown CC, Chu KC. A new method for the analysis of cohort studies: implications of the multistage theory of carcinogenesis applied to occupational arsenic exposure. Environ Health Perspect 50:293–308 (1983).
12. Crump KS, Allen BC, Howe RB, Crockett PW. Time factors in quantitative risk assessment. J Chron Dis 40(2):101–111 (1987).
13. Thomas DC. A model for dose rate and duration of exposure effects in radiation carcinogenesis. Environ Health Perspect 87:163–171 (1990).
14. Brown CC, Chu KC. Use of multistage models to infer stage affected by carcinogenic exposure: example of lung cancer and cigarette smoking. J Chron Dis 40(2):171–179 (1987).

15. Moolgavkar SH, Knudson A. Mutation and cancer: a model for human carcinogenesis. J Natl Cancer Inst 66:1037–1052 (1981).
16. Moolgavkar SH, Dewanji A, Luebeck J. Cigarette smoking and lung cancer: a reanalysis of the British doctor's data. J Natl Cancer Inst 81:415–420 (1989).
17. Moolgavkar SH, Day NE, Stevens RG. Two-stage model for carcinogenesis: epidemiology of breast cancer in females. J Natl Cancer Inst 65:550–569 (1980).
18. Moolgavkar SH, Cross FT, Luebeck G, Dagle GE. A two-mutation model for radon induced lung tumors in rats. Radiat Res 121:28–37 (1990).
19. Thomas DC. A model for dose rate and duration of exposure effects in radiation carcinogenesis. Environ Health Perspect 87:163–171 (1990).
20. Cox DR. Partial likelihood. Biometrika 62:269–276 (1975).
21. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. Ann Stat 10:1100–1120 (1982).
22. Prentice RL, Self SG. Asymptotic distribution theory for Cox-type regression models with general relative-risk form. Ann Statist 11:804–813 (1983).
23. Breslow NE, Day NE. Statistical methods in cancer research. In: The analysis of case-control studies, vol 1. Lyon: International Agency for Research on Cancer, 1980.
24. Preston DL, Kato H, Kopecky KJ, Fujita S. Studies of the mortality of the A-bomb survivors: cancer mortality 1950–82. Radiat Res 111:151–178 (1987).
25. Barlow WE, Prentice RL. Residuals for relative risk regression. Biometrika 75:65–74 (1988).
26. Prentice RL. Covariate measurement errors and parameter estimation in Cox's failure time regression model. Biometrika 69:331–342 (1982).
27. Pepe MS, Fleming T. A nonparametric method for dealing with mismeasured covariate data. J Am Stat Assoc 86:108–113 (1991).
28. Kupper LL. Effects of the use of unreliable surrogate variables on the validity of epidemiologic research studies. Am J Epidemiol 120:643–648 (1984).
29. Prentice RL, Sheppard L. Dietary fat and cancer: rejoinder and discussion of research strategies. Cancer Causes and Control 2:53–58 (1991).
30. Clayton DG. Models for the analysis of cohort and case–control studies with inaccurately measured exposures. In: Statistical models for longitudinal studies of health (Dwyer JH, Lippert P, Feinleib M, Hoffmeister H, eds). Oxford: Oxford University Press, 1992; 301–331.
31. Greenland S, Kleinbaum DG. Correcting for misclassification in two-way tables and matched-pair studies. Int J Epidemiol 12:93–97 (1983).
32. Hui SSL, Walter SD. Estimating the error rates of diagnostic tests. Biometrics 36:167–171 (1980).
33. Pierce DA, Stram D, Vaeth M. Allowing for random errors in radiation dose estimates for the atomic bomb survivors. Radiat Res 126:36–42 (1990).
34. Whittemore AS. Errors-in-variables regression using Stein estimates. Am Stat 43:226–228 (1989).
35. Sposto R, Stram DO, Awa AA. An estimate of the magnitude of random errors in the DS86 dosimetry from data on chromosome aberrations and severe epilation. Radiat Res 128:157–169 (1991).
36. Armstrong BG. The effect of measurement error on relative risk regressions. Am J Epidemiol 132:1176–1184 (1991).
37. Thomas D, Stram D, Dwyer J. Exposure measurement error: influence on exposure–disease relationships and methods of correction. Annu Rev Public Health 14:69–93 (1993).
38. Stevens W, Thomas DC, Lyon JL, Tills JE, Kerber R, Simon SL, Lloyd RD, Elghany NA, Preston-Martin S. Leukemia in Utah and radioactive fallout from the Nevada Test Site: a case–control study. JAMA 264:585–591 (1990).
39. Thomas DC, Gauderman J, Kerber R. A nonparametric Monte Carlo approach to adjustment for covariate measurement errors in regression (tech rep #16, USC Department of Preventive Medicine). Biometrics, (in press).
40. Greenland S. Variance estimation for epidemiologic effect estimates under misclassification. Stat Med 7:745–757 (1988).
41. Greenland S. Statistical uncertainty due to misclassification: implications for validation substudies. J Clin Epidemiol 4,1:1167–1174 (1988).
42. Spiegelman D, Gray R. Cost–efficient study designs for binary response data with Gaussian covariate measurement error. Biometrics 47:851–869 (1991).
43. Prentice RL, Breslow NE. Retrospective studies and failure time models. Biometrika 65:153-158 (1978).
44. Langholz B, Thomas DC. Nested case–control and case–cohort methods of sampling from a cohort: a critical comparison. Am J Epidemiol 31:169–176 (1990).
45. Langholz B, Thomas DC. Efficiency of cohort sampling designs: some surprising results. Biometrics 47:1553–1571 (1991).
46. Prentice RL. A case–cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 73:1–11 (1986).
47. White JE. A two-stage design for the study of the relationship between a rare exposure and a rare disease. Am J Epidemiol 115:119–128 (1982).
48. Breslow NE, Cain KC. Logistic regression for two-stage, case–control data. Biometrika 74:11–20 (1988).
49. Prentice RL, Sheppard L. Validity of international, time-trend and migrant studies of dietary factors and disease risk. Prev Med 18:167–179 (1989).
50. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 73:13–22 (1986).
51. Prentice RL, Zhao LP. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. Biometrics 47:825–839 (1991).
52. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. Int J Epidemiol 18:269–274 (1989).
53. Piantadosi S, Byar DP, Green SB. The ecological fallacy. Am J Epidemiol 127:893–904(1988).
54. Cochran WG. Some methods for strengthening the common χ^2 tests. Biometrics 10:417–451 (1954).
55. Prentice RL, Sheppard L. Dietary fat and cancer: consistency of the epidemiologic data and disease prevention that may follow from a practical reduction in fat consumption. Cancer Causes and Control 1:87–97 (1990).
56. Yusuf S, Simon R, Ellenberg S. Proceedings of a workshop on methodologic issues in overviews of randomized clinical trials. Statist Med 6:217–409 (1987).
57. Fuller W. Measurement error models. New York: John Wiley and Sons, 1987.
58. Clayton D, Cuzick J. Multivariate generalizations of the proportional hazards model (with discussion). J R Statist Soc A 148:82–117 (1985).
59. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J Am Stat Assoc 84:1065–1073 (1989).
60. Prentice RL, Cai J. Covariance and survivor function estimation using censored multivariate failure time data. Biometrika 79:495–512 (1992).
61. Rothman KJ. A sobering start for the cluster buster's conference. Amer J Epid 132:S6–S13 (1990).
62. Keiding N. Age–specific incidence and prevalence: a statistical perspective (with discussion). J R Statist Soc A 154:371–412 (1991).
63. Otake M, Prentice RL. The analysis of chromosomally aberrant cells based on a beta–binomial distribution. Radiat Res 98:456–470 (1984).
64. Prentice RL. Surrogate end points in clinical trials: discussion, definition and operational criteria. Stat Med 8:431–440 (1989).

Measurement Issues in Environmental Epidemiology

Maureen Hatch[1*] and Duncan Thomas[2]

[1]Columbia University School of Public Health, 600 West 168th Street, New York, NY 10032; [2]University of Southern California, School of Medicine, Department of Preventive Medicine, 1420 San Pablo Street, Los Angeles, CA 90033

This paper deals with the area of environmental epidemiology involving measurement of exposure and dose, health outcomes, and important confounding and modifying variables (including genotype and psychosocial factors). Using examples, we illustrate strategies for increasing the accuracy of exposure and dose measurement that include dosimetry algorithms, pharmacokinetic models, biologic markers, and use of multiple measures. Some limitations of these methods are described and suggestions are made about where formal evaluation might be helpful. We go on to discuss methods for assessing the inaccuracies in exposure or dose measurements, including sensitivity analysis and validation studies. In relation to measurement of health outcomes, we discuss some definitional issues and cover, among other topics, biologic effect markers and other early indicators of disease. Because measurement error in covariates is also important, we consider the problems in measurement of common confounders and effect modifiers. Finally, we cite some general methodologic research needs. — Environ Health Perspect 101(Suppl 4):49–57 (1993).

Key Words: Biologic markers, dose, environment, exposure, mathematical modeling, measurement error, psychosocial factors, sensitivity analysis, susceptibility

Measuring Environmental Exposure and Dose

Concepts

Environmental exposures can occur as a result of contact with a variety of elements (air, water, soil) that, in turn, influence the pathways for exposure (inhalation, ingestion, dermal). Individuals' interactions with these elements are complex, and therefore it is not surprising that exposure assessment and dose estimation are formidable challenges to those investigating the health effects of environmental agents.

The concepts of exposure and dose have been elaborated in a series of recent publications issued by the Board on Environmental Studies and Toxicology of the National Academy of Sciences (*1,2*). The term exposure refers to the concentration of an agent at the boundary between an individual and the environment as well as the duration of contact between the two, but dose refers to the amount actually deposited or absorbed in the body over a given time period. Although internal dose is the ideal measure from the scientific standpoint, regulation can deal only with external exposures, and therefore one may want to measure both exposure and dose.

This manuscript was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990 – September 1992.

*Author to whom correspondence should be addressed.

This work was supported in part by grant R01-HD24659 from the National Institute for Child Health and Human Development.

Individuals' exposures may be modified by factors such as activity patterns, which determine encounters with various sources of exposure; bioavailability of the agent in time and place; and the rate at which exposure occurs (e.g., a relatively constant rate versus a variable rate). From a given exposure, a person's resultant dose will depend on host characteristics, such as age, sex, and metabolism. It also will reflect the susceptibility of target tissue at the time of exposure; any shielding provided by the body (e.g., the placenta, the blood–brain barrier) or modulation by buildings that attenuate exposure to electric fields and gamma radiation but can be a source of exposure to radon; and the effect of concurrent exposures, such as cigarette smoking or medications. In addition, only particular components of the dose may be relevant to health effects. For calculating dose–response relationships, this biologically effective dose is what ought to be quantified. But in many instances it may be difficult to define what the biologically effective dose is, much less measure it. In any event, the definition is time-dependent and subject to change along with the state of scientific knowledge, just as measurement capabilities change with new technology. Epidemiologists undoubtedly need to prepare for a new generation of studies in which measurement of variables will involve data at the level of the gene. A commitment of resources, such as talent and funding, could improve the state of the art in exposure and dose assessment and potentially yield better estimation of exposure–response relationships and more effective measures of environmental protection.

In the past, the methods used to assign exposures in environmental health studies were quite crude, and to some extent they still are (e.g, pesticide usage patterns, residence near a point source of pollution). Even in studies where disease has been ascertained at the individual level, exposure measures may be ecologic in nature and based on average levels for a group. When the group is defined in geographic terms, exposure levels might be estimated from values recorded by environmental sampling in a subject's general vicinity. However, recent research has shown that correlations sometimes are weak between readings from area monitors and subjects' exposures measured using personal monitors (*3*), which are presumed to relate more closely to the true dose. Discrepancies between readings from personal and areawide samples can result from heterogeneity of exposures, from poor placement of samplers (e.g., air monitors at elevations well above the breathing zone), or from failure to take account of human activity patterns and other sources of exposure.

Exposure monitoring systems can be and are being improved, however. Newer approaches include sampling the microenvironments where exposure principally occurs, including indoor environments (e.g., bedrooms and living rooms in studies of radon and electric and magnetic fields), as well as total exposure monitoring in which all potentially relevant microenvironments are

sampled (*4,5*). The latter approach is particularly important for ubiquitous compounds like the polycyclic aromatic hydrocarbons. To some extent, personal exposure monitoring is also beginning to be incorporated into environmental health studies. In addition to these attempts to improve externally derived measures of exposure, efforts are being made to estimate internal dose using strategies like empirical dosimetric modeling, pharmacokinetic modeling, and biologic markers.

Such efforts are important. The failure to assign individual exposure and dose accurately leads to measurement errors with consequent effects on measures of association (and, ultimately, risk assessments) that will differ depending on whether the error is random or systematic and whether the unit of analysis is the individual or the group. Systematic error in exposure measurement can introduce bias either toward or away from the null. Random error tends to bias results toward the null, although exceptions to the rule can be found in unusual circumstances (*6*). For ecologic studies in which exposure is a binary variable derived from combinations of individual observations, the rule stating random error generally biases results toward the null may not hold (*7*).

Given the consequences of error in estimating exposure, it is important to try to increase accuracy of measurement at the design stage of a study. How, then, does an investigator decide when the use of a surrogate exposure measure (i.e., an error-prone measure) is acceptable, and when it is not? Rosner et al. have shown (*8*) that for correlations between surrogate and true measures of exposure less than 0.8, the odds ratios estimated by logistic regression will differ markedly for the surrogate and the true exposure measure, while much less bias will occur when correlations between the two measures are 0.8 or greater. *In vivo* tibia lead levels measured by X-ray fluorescence have been proposed as a good surrogate for cumulative blood lead levels on the basis of a correlation coefficient of 0.84 (*9*). For dietary exposures, however, the correlation between food frequency questionnaires and less error-prone methods (food records, measurements in food or biological samples) is only around 0.5 (*10*); yet food frequency questionnaires continue to be applied in large-scale studies, only occasionally with correction of risk estimates for error in measurement. On the other hand, the failure to find a correlation (actual coefficients not given) between current adipose tissue or serum dioxin levels and surrogate measures of past exposure to Agent Orange in Vietnam (*11,12*) affected a decision not to conduct further research using exposure surrogates based on troop location and herbicide spraying records. These examples underscore the need to be explicit about criteria for acceptable surrogate measures, as well as the need to take error into account when surrogates are used, even while emphasizing the development of better approaches to exposure-dose assessment.

In the following section, we describe methods designed to reduce error in exposure measurement insofar as is currently possible (approaches such as dosimetric modeling, pharmacokinetic modeling, biologic markers, and use of multiple measures), as well as approaches to assessing the residual uncertainties in the estimated dose. Even the best of the current methods will not yield a measure that is completely error-free, and it is therefore important to recognize and characterize the residual error in measurement so that it can be considered in analysis of the data.

Measurement Approaches

Exposure or Dose Modeling

Estimating a subject's exposure to an environmental agent involves combining information about possible sources of exposure (usually obtained from the subject, from some other respondent, or from records) with an assessment of the likely degree of exposure from each source.

When an exposure under study is environmental, there may be multiple pathways by which a person might be exposed and it can be important to consider all elements and all routes. For example, residents downwind of the Nevada Test Site could have been exposed to external gamma radiation from the passing fallout cloud itself, from ingesting contaminated milk or vegetables, or, in the case of infants, from in utero exposures or breast-feeding. For each of these pathways, several different radionuclides might need to be considered. After eliminating pathways that would be expected to make a negligible contribution to the total dose, one can estimate the likely dose rate per unit of exposure to each pathway. In the fallout example, this involved consideration of *a*) source term, the amount and type of radionuclide released; *b*) the environmental transport, dispersion from the source to sites of deposition; *c*) rate of radioactive decay and environmental dispersion of the radionuclides; *d*) farm management practices leading to contamination of dairy cattle or vegetables; *e*) estimates of the uptake of radionuclides by vegetables and milk; *f*) distribution of milk and vegetables to consumers; and *g*) uptake by the target organ from ingested radionuclides. To calculate an individual's dose, this information was then combined with extensive questionnaire data on breast-feeding and maternal and individual consumption of milk and vegetables at various ages. For some subjects, modifications were needed to allow for homegrown vegetables or backyard cows or goats. For subjects with incomplete exposure information, distributions of default values specific to their particular circumstances (age, sex, location, etc.) were developed. Similar calculations were performed for each of over 100 nuclear tests, and the results then were summed to produce estimates of each subject's total dose (*13*).

The process described above is far more complex than has been the norm in environmental epidemiology, but it represents the current state of the art in environmental dose assessment. Less refined, but perhaps less costly, approaches to exposure-dose modeling (often for households or geographic areas rather than for individuals) have been based on Gaussian-dispersion modeling of airborne emissions (*14–16*), hydrogeologic modeling of waterborne exposures (*17*), and isopleth modeling of soil contaminants (*18*). Assuming that dosimetry models are reasonably accurate, such approaches should decrease bias arising from measurement error and increase precision. Assessment of the validity of dosimetry models should be made whenever possible. For example, an environmental dispersion model of emissions at the time of the accident at the Three Mile Island nuclear plant was validated by the readings from off-site thermoluminescent dosimeters.

Dosimetric modeling methods are likely to be used more frequently in future environmental health studies. A question is whether the effort required both in terms of the information that must be collected from study subjects and/or by environmental sampling and the effort involved in development of the dosimetric model itself are warranted by the gain in precision or reduction in bias of the exposure estimates. Information on this point could be obtained by comparing the point and interval estimates of associations observed using gold standard dose estimates with those that would be obtained using cruder methods. Such comparisons could be made in existing data sets. Understanding when the gains from dosimetric modeling are substantial and when they are only marginal would be useful in establishing methodologic standards of practice.

Some other issues related to dosimetry are exemplified by studies of cancer and electric and magnetic fields (EMFs). The initial hypothesis about EMFs was derived from observations showing apparent excesses of leukemia (and some other cancers) both in children living near electric power lines that would be expected to generate high magnetic fields (*19*) and in certain classes of electrical workers (*20*). In both the residential and occupational settings, it has been difficult to establish whether the magnetic fields are the responsible agent. While subsequent studies have demonstrated that certain electrical wiring configurations and certain categories of electrical work are associated with higher than average fields, so far no convincing associations have been found between leukemia risk and individuals' exposure to electric or magnetic fields determined by area measurements. No studies using personal dosimetry have yet been reported.

Four possible explanations are suggested for the failure to establish a clear association between cancer and measured field strengths. First, it may be due to their extreme variability in space and time. Any necessarily short-term measurement (24 hr or a week in a small number of locations) is a poor surrogate for lifetime dose; under this explanation, household wiring classifications and job titles may be more stable measures of long-term exposure. Second, the failure to detect an association with measured fields may reflect a failure to measure the biologically relevant parameter (e.g., peaks, transients, resonance between static and oscillating fields rather than the time-weighted average). Studies of reproductive outcomes, where the period of exposure is much shorter than for cancer and where there may be a particular time window of vulnerability, could help indicate whether the discrepancy in associations with wire codes and measured fields is due to their capturing different time frames or different dimensions of EMFs. A third explanation for the associations of cancer with wiring configurations, but not with measured fields, relates to selection bias (lower selection probabilities for controls living near wiring with high current configurations). Fourth, the surrogate exposure measures (wire codes, job titles) may be confounded by other correlated risk factors. This controversy is still far from resolved, but consideration of selection bias and possible confounders together with careful assessment of all potentially salient aspects of electric and magnetic fields and of the variability of the different measurements should shed light on the issue.

The EMF example underscores the need for making multiple measures of exposure. In particular, it argues for continuing to include surrogate measures along with gold standard measures in studies of health effects until the relations between the surrogate and criterion measures are well understood and there is certainty about the true gold standard (i.e., until the correct biologic mechanism is known). Substituting an incorrect gold standard for a surrogate measure can actually increase measurement error. One analytic approach to using multiple measures that has been proposed as a means of increasing validity is to restrict analysis to subjects who are classified as exposed or unexposed by two different, if imperfect, exposure measures (*21*). This clearly risks some loss in power since subjects with discordant results on the two measures are excluded from analysis. Another proposed approach is to estimate the misclassification probabilities for each measure and from them to estimate the prevalence of exposure (*22*).

Some mention of personal monitors should also be made. While these do not provide a measure of resulting body burden, as biologic markers are meant to do, personal monitors may measure the intensity of an individual's total exposure to airborne agents better than fixed-site area monitors. This is not always the case, however, particularly in studies of long-term exposures or where areawide concentrations are fairly uniform. The TEAM study (TotalExposure Assessment Methodology) conducted by the U.S. Environmental Protection Agency (EPA) found that personal air monitors were acceptable to subjects from 7 to 85 years of age (*23*). Investigators studying effects of exposure to EMFs and indoor air pollutants on children are anxious to develop personal monitors that can be used with children under age seven, including toddlers. At present, personal monitors for EMFs are in the form of wristbands and may not be suitable for very young children. Technology for personal exposure monitoring is still evolving, but it will rarely be feasible to apply personal exposure monitoring to all subjects and all relevant time periods. Therefore, methodologic approaches are needed for combining collected exposure data with personal samplers and environmental monitors.

Pharmacokinetic Modeling

Pharmacokinetic modeling is an approach to dosimetry that incorporates information about the internal pharmacologic processes that ensue once an agent reaches the portal(s) of entry into an individual's body (*24*). These include uptake into the circulation; distribution within the body; and metabolism, storage, and elimination. These models can be simple, involving only one body compartment, or complex, involving multiple body compartments. In either case, compartmental rate relationships are used in the model's equations to estimate concentrations at critical tissues. Such models are also useful as guides to temporally relevant and efficient ambient sampling (*24*). Pharmacokinetic modeling of exposure and dose may be viewed as a counterpart to biologically based disease models.

Biologic Markers

Because of the difficulty of obtaining accurate and unbiased exposure information from study subjects and the difficulty of estimating the doses that such exposures might produce, there has been great interest in the development of biologic markers. These may be defined as "cellular, biochemical, or molecular alterations that are measurable in biological media, such as human tissue, cells, or fluids" (*25*). If used appropriately, biologic markers allow for considerable improvement in measurement of dose. First, they may obviate the errors arising from subjects' lack of knowledge, memory failure, biased recall, or deliberate misinformation (*26*). Second, even when subject reports of exposure are accurate, individuals may vary considerably in uptake and handling of a material; the error introduced by such individual variation can be reduced or removed by using markers that provide an estimate of the dose to a particular individual. Third, some markers can be used to detect biological interactions between the exposure of interest and critical tissues; DNA adducts are an example of this type of marker. In studying environmental tobacco smoke, for instance, one can—in addition to asking about maternal smoking during pregnancy—actually measure smoking-related DNA adducts in placentae (*27*) and, where the fetus is lost, in critical organs such as fetal lung or liver (*28*). Another advantage of biologic markers is that generally they give a quantitative, or at least semiquantitative, estimate of dose. They also can serve as the gold standard for other information sources, thus providing a basis for error allowance procedures in studies that rely on less accurate exposure measures due to the cost of the marker.

Other Biologic Dosimeters

Certain signs or symptoms can also be viewed as biologic dosimeters. For example,

in the cohort of atomic bomb survivors, it has been reported that subjects with a history of epilation have a 2.5-fold steeper dose–response curve for leukemia than those without (*29*). This can be interpreted either as an indicator of their greater radiosensitivity or as an indicator of misestimation of their doses, perhaps as a result of differences in shielding not accounted for by available dosimetry data.

To be useful in environmental epidemiology studies, a biologic exposure marker should be clearly better than anamnestic data or environmental measures; should allow for differentiation between exposure levels; should be applicable on a large scale; or if too costly for large-scale use, should at least be acceptable to subjects in a validation substudy. Before markers are used in epidemiologic research, their sensitivity and specificity should be known from both the laboratory and epidemiologic perspectives; reproducibility of results within and between laboratories must also be known; and, very importantly, the particular time frame they reflect and during which they can be measured *in vivo* must be established (*25*) so that they provide interpretable data regarding time and dose.

At present, few exposure markers satisfy these requirements. Some markers may provide a record of cumulative exposure (e.g., bone lead measurement, mercury or cocaine measurements in hair), but most can assess only relatively recent exposures. Studies of biologic markers that use a case–control design and a cross-sectional marker of exposure can be difficult to interpret because of ambiguity about the temporal sequence of the marker and the disease [e.g., whether selenium levels in breast cancer cases are cause or consequence (*30*)]. Indeed, such studies can be misleading. Vineis and Caporaso (*31*) have described how a case–control study nested in a cohort allowed Wald and his colleagues (*32*) to make use of the time between initial collection of specimens from members of the cohort and subsequent onset of cancer to clarify the time order in the relationship with blood retinol. Although analysis considering only the early cases of cancer suggested that blood retinol might be protective, ultimately it was apparent that some metabolic change associated with the disease was acting to reduce retinol levels, rather than vice versa. In addition to such problems in interpretation, biological measurements are often costly to perform. Furthermore, the need to obtain specimens can reduce the cooperation of subjects and introduce the potential for selection bias to occur through initial refusal or later attrition, although these problems are probably not insurmountable if they are anticipated and addressed.

Use of Multiple Measures

When the biological basis of an association is poorly understood, it can be very helpful to have various types of exposure measurements available. Or, as mentioned previously in connection with personal exposure monitoring, it may be necessary to rely on another source of exposure information for portions of the study period. The obvious approach is to analyze each type of measurement separately, but there may be merit in combining them into an index, if only to reduce measurement error. Complications can arise if all measurements are not available on the same subjects. Any associations observed might be due to differences in the measurements or to differences in the subgroups of subjects for whom the measurements are available. In a study of childhood leukemia and electric and magnetic fields, London et al. (*33*) reported the results separately for various summaries of 24-hr bedroom dosimetry, spot measurements at various locations, and wiring configurations. However, drawing on all of these data, they also developed regression models for magnetic fields at various locations based on attributes of the wiring and used the values predicted by these models as the time-weighted average fields for all houses lived in. Thus, predicted values were used both to replace existing measurements and to impute missing values. The rationale behind the approach is to avoid the loss of information and possible selection bias associated with restricting analysis to subjects with data for all measurements made (*34*). One alternative is to retain measurements where they exist and to impute only the missing values, leaving open the possibility of stratifying on data quality in the analysis. Other approaches undoubtedly can be devised, and it would be desirable to compare their validity using data sets in which exposure–response relationships are well understood and where more than one measure of exposure exists.

Other Issues in Measurement of Exposure

Taking Account of Critical Periods for Exposure

A principal problem in environmental epidemiology has been that the inaccuracy in measurement generally (although not always) operates in the direction of overestimating exposure and therefore underestimates risk or perhaps misses health effects altogether. For example, when assigning the same level of exposure to all 1000 residents living within five miles of a toxic dump site when only 100, say, were truly exposed and the other 900 were either unexposed or exposed at very low levels, one would be certain to calculate an observed relative risk for exposure that would be lower than the true risk. Hence the importance of increasing the accuracy of exposure definitions and measurement is obvious. Rothman and Poole have pointed out (*35*) that it is also important to use information on critical periods for exposure, either in the design phase of a study, in the analysis phase, or in both. For example, in a study of Down's syndrome, parental exposures occurring after the fertilization period are presumably irrelevant to the outcome; in fact, there is mounting evidence that most cases of Down's are traceable to errors at the time of the first meiotic division in the maternal germ cell (*36*). By removing all exposures that are not of biologic consequence from the estimate of association, one can expect the magnitude of the estimated association to increase. Moreover, information on known critical periods might be used to test whether an association appears to be spurious. If an association were found not only during the critical period but also for exposure during noncritical periods, then the association might be due to recall bias, or it could be reflecting autocorrelations in exposure status. Multivariate analysis of the effects of exposure in various critical and noncritical periods could, in principle, overcome this problem, provided there are enough exposed subjects with different temporal patterns of exposure to be informative.

Taking Account of Migration In and Out of Exposed Areas

The problem of in- and out-migration is frequently raised as an issue in interpreting results of studies that define exposure in terms of time and place. Although several studies have considered the effects of population migration on the validity and precision of estimated associations between exposure and disease (*37*) and have described when and in what direction bias is likely to arise, these issues are still not understood well. Perhaps more simulations or empirical demonstrations are needed to improve the general level of comprehension about the effects of population mobility on geographic studies. In the case of specific studies, it would help to know something about duration of residence or at least age-specific duration patterns in an area. One recent suggestion is to estimate by various means the fraction (f) of time spent by a subject in a particular place and to assign for

the remaining fraction $(1-f)$ the average exposure for some total referent area (*38*).

Assessing Past Exposure

A major problem in many environmental health studies is the difficulty of estimating past exposures when only present-day measurements are available. Often, some data on subjects' past exposures can be obtained by questionnaire or review of existing records. For example, in occupational studies, payroll records are used to assemble a job history. The use of records from years past to establish exposure status has the important advantage of obviating recall bias, although it may introduce its own problems (e.g., missing records or less specificity in records from early years). Estimating the actual historical exposure levels is more difficult than simply classifying exposure status, and it often involves a large degree of judgment. Clearly, the more historical data there are on variation in exposure levels over time and place, the better. Study of such patterns of variation can suggest models for predicting exposures at times for which no measurements are available. For example, in a study of salivary tumors and dental X-rays, Preston-Martin et al. (*39*) reviewed 58 studies that described doses from various procedures at various times and, while taking into account the dates of introduction of new technologies, used regression analysis to develop models for the expected dose as a function of calendar year. In occupational settings, the subjective experience of long-service workers has been used to compare current exposures with those in the distant past. Similar strategies (i.e., tracking technological developments, use of knowledgeable informants) need to be applied in the assessment of past environmental exposures. For example, in a case–control study of colorectal cancer and water chlorination among women teachers in New York State, Lawrence et al. (*40*) used current water sampling in conjunction with records from water treatment plants covering the previous 20 years in a mathematical model to estimate cumulative exposures to chloroform in drinking water at home and at work.

Uses of Existing Environmental Databases

One limitation on assessing past environmental exposures is that reviews of existing data bases at the national and state level srepeatedly have found them to be inadequate for epidemiologic purposes because of insufficient data points to assess variability, lack of a standardized Quality Assessment/Quality Control protocol, incomplete geographic coverage, and missing information (*41*). Efforts are underway to modify the major air and water data bases to make them more useful for future environmental health studies.

Existing environmental data banks could also be used to define strata within which to conduct sample surveys. Surveys of individuals within these ecological exposure groupings would help document human activity patterns and could indicate the distribution of exposure and important confounding or effect-modifying variables in each stratum. Potentially, such stratified-sample surveys might provide the basis for constructing an environment–exposure matrix similar to the job–exposure matrices used in occupational studies. Such exposure matrices are generally assumed to have a "Berkson error" structure (*42*), in which the average of the true doses for all subjects in an exposure assignment group is equal to the assigned value. As a consequence, if the true dose–response is linear, the estimated slope of a linear relationship will not be biased toward the null.

Estimating Dose Uncertainties

A major concern among environmental epidemiologists is the influence of errors in exposure estimates on associations with disease and methods of dealing with such errors. The best cure for this problem is to avoid measurement error in the first place. When this is not feasible (and it often may not be, particularly in investigating common source exposures such as toxic dump sites), it is helpful to be able to quantify the direction and magnitude of the errors. This can be done in a number of ways, including *a*) validation studies on a subset of the study sample or a pilot sample to compare the measurements to be made in the field with a gold standard, *b*) replication of measurements to assess within-subject variability, *c*) multiple types of measurements to assess validity, and *d*) sensitivity analysis to estimate the influence of various unknowns or uncertain parameters on the estimated doses. The goal might be either to describe the distribution of exposure errors across the population (or subgroups there of) or to obtain an estimate of the precision of each subject's exposure assignment.

Because a gold-standard assay is often not feasible for use in the field (because of cost, time, acceptability, etc.), validation studies usually must be limited to a relatively small number of subjects. The resulting estimates of error distributions may be imprecise (*43*), although this will be less of a problem if the data are treated as continuous and if parameters for sensitivity and specificity do not have to be estimated (*8*). Nonetheless, sample sizes for validation studies that are needed to insure good estimates of the error rates in field measurements should be calculated carefully. Other considerations are to insure that the measurement error process in the sample used for validating the field measure is similar to that in the target population for the full study and to avoid selection bias in the validation study, which might arise if requirements associated with use of the gold standard measure are very demanding and participation rates are consequently low. In the New Jersey case–control study of radon and lung cancer among women, in-home radon measurements were obtained for only 40% of the houses targeted, and smoking rates differed among those with measured and unmeasured homes, raising the possibility of selection bias (*44*). If data on disease are collected on validity study participants, potential selection bias can be examined by testing for heterogeneity in the risk estimates.

Replicate measurements are useful for describing repeatability (*45*) but cannot assess other components of error, such as subjects' tendency to consistently overreport or underreport exposures. Having different types of measurements available may be more useful in estimating misclassification probabilities, even if none of the measures is error-free. See, for instance, Hui and Walter's maximum likelihood method for estimating error rates with two independent assessments of exposure (*22*).

Sensitivity analyses can take a number of forms. The basic idea is to consider a range of plausible values for each of the unknowns in the exposure assignment process. If there are only a few unknowns, one might consider each of them and evaluate their influence on either the individual exposure assignments or the final dose–response relation. If there are many, one can estimate the distribution of assigned doses, either analytically or by Monte Carlo simulation. The latter approach was used in the studies around the Nevada Test Site because of the complexity of the dosimetry algorithm. Components of uncertainty that were considered include the source term, environmental transport, farming practices and distribution, and default values for individuals' missing data. A series of sensitivity analyses were also carried out on a mathematical model that estimated the relative geographic distribution of exposure to accident emissions at Three Mile Island by examining variations in modeling assumptions for their effect on the base case (*46*). Parameters considered were the source term, the degree of plume rise, wind shifts, and residual error weighting. In addition, a Bayesian analysis was used to quantify uncertainty about the time-release pattern.

Measuring Outcome of Environmental Exposures

Definitional Issues

As strong effects of environmental exposure have been identified and dealt with, environmental epidemiology increasingly has become a search for weaker associations. It is all the more important, therefore, to improve measurement of outcome through careful definition and avoidance or reduction of error (*35*). In defining study end points, the aim should be to specify the health outcome of interest as precisely as possible in order to avoid further dilution of a weak association through inclusion of irrelevant cases. In fact, it may be desirable to consider subgroups of disease that are etiologically homogeneous and that are believed to be responsive to the exposure of interest on the basis of theory or prior observations (e.g., certain histopathologic types of lung cancer and radon; leukemia types and subtypes with ionizing radiation and EMFs). This can present something of a dilemma, however, because statistical power for examining subgroups is likely to be low unless the difference in effect size among subgroups is sufficient to offset the reduced sample size.

The virtues of lumping versus splitting frequently come up for discussion in the context of studies of congenital anomalies. It is unlikely that an exposure would affect all types of congenital defects. With maternal cocaine use during pregnancy, for example, defects involving vascular disruption seem to be implicated. However, a biological basis for positing subgroups of interest is often lacking; empirical Bayesian approaches may be useful in helping to formulate relevant subgroupings. In any event, the numbers in particular case groups are likely to be small for all but a few categories. If sufficiently large series cannot feasibly be accrued in a single study, multisite (even multinational) projects may need to be mounted, or more reliance may need to be placed on meta-analyses combining results from several studies. Which of these strategies to pursue should be discussed by groups of investigators studying the same exposure, and their potential funding sources.

Disease outcomes in environmental epidemiology can be measured on a continuous scale or categorically as incident or prevalent cases or as deaths. Incidence data are usually preferable for investigating etiology since prevalence or mortality data may be influenced by factors affecting duration of disease and survival as well as those relating to cause. However, incidence data are often less easily accessed than mortality data, and they can be subject to artifactual variations in ascertainment—as a result of screening programs, for example. Whether incidence or mortality is the more reliable indicator of health status and in what age groups it is reliable have been discussed extensively but not resolved. See, for example, the recent papers by Doll (*47*) and by Davis et al. (*48*) about cancer time trends. It might be helpful to have a set of recommended approaches for trend analysis that were developed by a group of dispassionate methodologists. For etiologic studies, incidence data seem conceptually superior; when mortality data are used, consideration needs to be given to accounting for influences on survival since these might correlate with exposure.

In some areas of research, such as reproduction and development, different outcomes can occur depending on the timing and dose of exposure. In such circumstances, it may be important to examine several end points. Extending population-based registration systems to cover more outcomes than cancer and birth defects and to cover more geographic areas potentially could be useful for environmental studies in several respects: in identification of cases, in validation of self-reported information, and in ascertaining disease status of migrants.

Biologic Effect Markers and Other Early Indicators of Disease

Biologic effect markers potentially have a number of advantages as study end points, particularly if they are strongly prognostic of disease in ways not explained by available exposure information—for example, by reflecting susceptibility or the action of cofactors (*26*). While some effect markers are actually subclinical events (e.g., biochemical tests of occult pregnancy loss), often markers of effect correlate only weakly with disease. Serum alpha-fetoprotein is a useful marker for liver cancer as well as a prenatal marker for neural tube defects. Markers that are not as clearly predictive of risk, particularly at the individual level, can lead to problems of interpretation and to needless anxiety for those individuals found to have elevated levels. The premature application of a poorly standardized cytological assay on a group of already concerned residents at Love Canal is a case in point. Calls have been made repeatedly to carry out longitudinal studies, in experimental animals and humans, that will measure the positive predictive value of such markers before applying them in field studies; but these have been largely ignored. The Scandinavian countries, however, have mounted a collaborative prospective study of cancer in a cohort of 3190 individuals who have been tested for sister chromatid exchanges (SCEs), structural chromosome aberrations, or both. A report based on a 13-year follow up of 800 subjects in the Finnish portion of the data (*49*) found a moderate, statistically significant positive association between cancer risk and chromosome aberrations (SMR = 2.65; 95% CI 1.2, 5.0); there was a positive trend (SMR = 2.06; 95% CI 0.8, 4.2) for SCEs. Additional prospective studies of this kind are needed to establish the relationships between markers and disease in order to assure their appropriate use and interpretation. In addition, determining when a marker could serve as the basis for preventive health measures directed at a distal end point such as cancer is an important issue; see Prentice (*50*) for a useful discussion of this and a proposed operational criterion for surrogate response variables.

Other potential advantages of biologic effect markers are their use in classifying disease more precisely and in suggesting mechanisms of action, such as those relating to susceptible subpopulations. For example, biologic markers that distinguish slow from fast acetylators have indicated that the enzyme *N*-acetyltransferase plays an important role in bladder cancers induced by exposure to aromatic amines (*51,52*). Methodologic needs in the area of effect markers include attention to sources of variability, both biological and laboratory-related, and to logistical issues, such as how to achieve reasonable participation rates when the effect marker requires a demanding regimen. Three current studies of early pregnancy loss illustrate this latter problem. Two of the studies ask participants for daily urine samples. The third study uses a modified specimen collection scheme requiring urine samples only twice monthly, at the beginning of menses. Preliminary data indicate higher response rates for the study with the simplified collection protocol. Whether the variability in enrollment is due to the differing demands on study subjects or to other variable aspects of the three studies (such as the perceived salience of the topic in the target population) is not known. Systematic research is needed to determine how to achieve cooperation in studies that use biologic markers and how to provide for calculating or estimating the extent and magnitude of selection bias.

Subclinical End Points

What role should physiologic changes (e.g., nerve conduction velocity, T-cell subsets, sperm count) have in environmental health assessments? It has been argued that functional alterations and nonspecific symptoms are likely to be more frequent consequences of low-level environmental exposures than frank disease (*53*). However, baseline rates and normal ranges for such end points may be lacking. Objective methods of assessment to remove the potential for biased recall may be at an early stage of development, and interpretation of results in terms of risk to groups and to individuals frequently is problematic, particularly as assay improvement allows for discriminating function more and more minutely. These methodologic limitations can be addressed—semen evaluation is a case in point (although the clinical significance of altered semen quality is still not clear-cut)—however, substantial time and effort will be required.

Measuring Confounders and Effect Modifiers

Effect on Risk Estimates If Inadequately Controlled

A confounding variable is one that, if not controlled appropriately, will tend to distort the exposure–disease association. For example, when studying whether household exposure to radon is a cause of lung cancer, one should be concerned about the possible confounding effect of smoking. Smoking is clearly a major risk factor for lung cancer. If houses with high radon levels are more likely to be inhabited by smokers, then this would produce an apparent relationship between radon and lung cancer even if there were no causal effect. The converse also could happen; if smokers tended to live in low-radon houses, then one might fail to find an association between radon and lung cancer if it really were present.

The strategies commonly used by epidemiologists to control confounding include restriction (e.g., to nonsmokers), matching, or statistical adjustment. All of these approaches presume that the confounding variable has been correctly measured. Greenland (*54*) has pointed out that errors in measurement of a confounding variable will tend to cause partial loss of an ability to eliminate confounding bias; for example, if the true odds ratio (adjusted for the true confounder) is 2.0 and the crude odds ratio (unadjusted) is 4.0, then the odds ratio adjusted for an incorrectly or crudely measured confounder might be 3.0. This intermediate outcome can only be counted upon in a case in which the errors in measuring the confounder are random (unrelated to exposure or disease status); in other cases, the adjusted odds ratio could be further from the truth than the unadjusted odds ratio. Kupper (*55*) has shown that an inaccurate surrogate confounder can produce seriously misleading inferences.

A factor like smoking, in addition to being a confounder, could also act as an effect modifier—that is, a variable that modifies the strength of the association between exposure and disease. A major question in the radon literature is whether the joint effects of smoking and radon exposure are multiplicative, additive, or some intermediate possibility. If they act additively, for example, then radon exposure would produce the same additional risk of lung cancer in smokers and nonsmokers; but because lung cancer is rare in nonsmokers, it would follow that radon exposure might account for a much larger proportion of lung cancers in that group. Conversely, if the two exposures act multiplicatively, the proportional increase in lung cancer rates due to radon exposure would be the same in smokers and nonsmokers; but because of the higher rates in smokers, the absolute increase would be larger in smokers. This issue therefore has important risk assessment and public health policy implications. Again, Greenland (*54*) has shown that errors in measurement of a covariate can distort its modifying effect and possibly introduce an apparent interaction where none exists. Diet and cooking habits in relation to aflatoxin exposure, and showering habits in relation to radon are additional examples of potentially important confounding or effect-modifying variables in environmental epidemiology.

Approaches to Measuring Common Confounders and Modifiers

The implications of the previous section are that careful measurement of strong confounders or modifiers should be given as much attention as the exposure and disease variables. It follows that some of the same approaches discussed in the sections on measurement of exposure and disease, such as use of multiple measures and biologic markers, will pertain here as well.

Continuing with the example of smoking, it is not sufficient simply to classify subjects by their present status as current, former, or never smokers. As long as smoking is a risk factor for the disease under study, one usually tries to obtain information on at least the ages at starting and stopping and the average daily amount of smoking. These data can be used to compute pack-years (the product of amount and duration), which is a stronger predictor of lung cancer risk than current status. In some other cases, however, such a product term may actually increase error. Better yet, nonlinear multivariate models could be used to allow for the joint effects of age at starting, duration and intensity of smoking, and time since quitting. Other modifying factors might include changes in level of smoking over time, use of filter cigarettes, and depth of inhalation. However, incorporating multiple modifying factors into an analysis needs to be done with considerable thought to produce models that are biologically plausible. Routine inclusion of interaction terms in a multiple logistic regression analysis can produce models in which ex-smokers eventually become at lower risk than never smokers, or light smokers have the same dependence on duration or age at start as heavy smokers. Use of general risk models based on biologically plausible theories is an attractive alternative.

Even the most complete smoking history is still likely to be misclassified, and the errors might well be related to the exposure or disease variables under study. In an occupational study of radon exposure and lung cancer, for example, miners with lung cancer might preferentially underreport their smoking histories to avoid prejudicing a compensation claim. For these reasons, there has been great interest in developing unbiased methods of assessing potential confounders. Biological measures, such as urinary cotinine for smoking or 4-aminobiphenyl-DNA adducts, are very attractive for this purpose. Other approaches were discussed above, in the section on exposure measurement. The disadvantage of most of these methods is that they measure only recent exposure and lifetime exposure will still be misclassified. The development of methods for combining information from different types of measurements could be very useful. Also discussed previously in the exposure measurement section, and equally relevant here, is the need to assess and allow for measurement error in confounders and effect modifiers whenever possible. Therefore, consideration should be given to mounting validation substudies to quantify measurement error in important covariates.

Susceptibility

Variation within a population in sensitivity to an exposure of interest can be substantial. Khoury et al. (*56*) estimated the proportion of susceptible individuals in the population for cigarette-induced cancers at

several sites; the proportions varied from <1% for oral and esophageal cancer up to 13% for cancer of the lung. Bias in risk estimates will arise if individuals with similar exposures but different susceptibilities are treated the same. There are a number of epidemiologic designs for assessing sensitivity to environmental exposures. As a measurement problem, the central issue is whether the marker for sensitivity being examined is a measurement of the genotype itself, some host characteristic, or family history.

The ability to classify genotypes directly has profound implications for identifying sensitive individuals. The obvious difficulty is that there are millions of genetic loci, for which only a relatively small number have probes available and only a few might be relevant to any particular disease. Thus, some prior knowledge that a locus has a role in the disease process is essential before embarking on a search for interactions with possible environmental exposures. Even so, the information for identifying genetically susceptible individuals may involve invasive and costly tests.

Recognition of phenotypically distinguishable subgroups of the population that have different baseline risks of disease or sensitivities to environmental exposures can therefore be very useful for public health protection. The measurement issues that arise here are essentially no different from those for any other effect modifier, as discussed above.

For family history as a marker of susceptibility to a disease, the basic minimal information that needs to be collected is the identification of the family members with the disease and the number, ages, and relationships of family members at risk. This information should be collected systematically for all first-degree relatives (parents, siblings, and offspring), and possibly for all second-degree relatives. As the objective is to examine family history as a marker of sensitivity to an environmental exposure, every effort should be made to obtain exposure information on all relatives, not just the affected ones.

Psychosocial Stress as Confounder, Effect Modifier, and Mediator

The psychosocial stress that may be associated with exposure to a perceived environmental hazard can potentially confound, mediate, or modify any associations between the exposure and disease. Stress might operate indirectly and cause exposed individuals to alter risk behaviors. Stress also could have an artifactual association with the end point of concern because of changes in care seeking, diagnostic practices, or self-reported health states. Alternatively, concern about environmental exposures could cause adverse outcomes other than those potentially associated with the perceived hazard. For example, studies around the Three Mile Island and Chernobyl nuclear plants indicate that the perception of danger can increase distress levels or clinical states like anxiety and depression (*57,58*), irrespective of whether radiation-induced increases in cancer actually occur.

The issue of stress as a confounder, effect modifier, mediator, indicator of some methodologic bias—or even as an exposure or outcome—needs to be explicitly addressed in future environmental epidemiologic research conducted on sensitized populations. Some relevant methodology has been developed in studies of communities near toxic wastes to distinguish between biologic effects of exposure to hazardous substances at such sites and either symptoms of stress or altered symptom reporting (*59,60*). These preliminary efforts include use of a scale to measure hypochondriasis and stratified analysis of self-reported symptoms to take account of subjects' perception about the source of pollution. Environmental epidemiologists need to learn when and how to address the issue of psychosocial stress in order to clarify interpretation of health effects studies and to estimate the importance of stress in its own right. Consideration should be given to measuring perceived stress and physiologic indicators of stress as well as to collecting data on methodological covariates such as motivation to participate, interest in receiving health care, and beliefs about the exposure in question as a cause of adverse health effects.

Methodologic Needs and Recommendations

The aspect of study design that involves measurement of variables is critical, especially in fields like environmental epidemiology where the risks from exposure are likely to be small, difficult to detect, and perhaps not clinically significant, yet may be of public health importance. Methodologic research in this area should emphasize the further development and application of dosimetric modeling. Existing data sets representing a range of research problems within environmental epidemiology could be used to assess the gains from dosimetry algorithms compared with cruder, more conventional methods of exposure assessment.

Dosimetry models invariably will use a combination of questionnaire data, environmental measurements, and biologic markers; this underscores the need for development and refinement of methods for handling multiple measures. Biologic markers themselves, as measures of exposure, effect, or susceptibility, are an area where additional methodologic development would be desirable.

A second important aspect of methodologic research relates to sensitivity analyses and other approaches for estimating the uncertainty in measurement of exposure and dose. Included in this category would be validation studies to compare a gold standard with a more error-prone exposure measurement in order to allow for correction of bias in the analysis stage of research. Consideration needs to be given to the costs and benefits of investigating measurement error in the primary study or in a substudy (which could be carried out internally or externally in relation to the primary study). A final area that deserves attention is measurement error in covariates, which can be as important as measurement error in the exposure or outcome variables.

REFERENCES

1. National Research Council. Biologic markers in reproductive toxicology. Washington DC: National Academy Press, 1989.
2. National Research Council. Human exposure assessment for airborne pollutants: advances and opportunities. Washington, DC: National Academy Press, 1991.
3. Wallace LA, Ziegenfus RC. Comparison of carboxyhemoglobin concentrations in adult nonsmokers with ambient carbon monoxide levels. J Air Pollution Control Assoc 35:944–949 (1985).
4. Lioy PJ. Exposure analysis and assessment for low-risk cancer agents. Int J Epidemiol 19:S53–S61 (1990).
5. Wallace L Pellizzari E, Hartwell TD, Perritt R, Sparacino C, Whitmore R, Sheldon L, Zelon H. The TEAM (total exposure assessment methodology) study: personal exposures to toxic substances in air, drinking, and breath of 400 residents of NJ, NC, and ND. Environ Res 43:290–307 (1987).
6. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? Am J Epidemiol 132:746–748 (1990).
7. Brenner H, Savitz DA, Jöckel KH, Greenland S. The effects of nondifferential exposure misclassification in ecological studies. Am J Epidemiol 135: 85–95 (1992).

8. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. Am J Epidemiol 132:734–45 (1990).
9. Somervaille LJ, Chettle DR, Scott MC, Tennant DR, McKiernan MJ, Skilbeck A, Trethowan WN. In vivo tibia lead measurements as an index of cumulative exposure in occupationally exposed subjects. Br J Ind Med 45:174–181 (1988).
10. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 122:51–65 (1985).
11. Kang HK, Watanabe KK, Breen J, Remmers J, Conomos MG, Stanley J, Flicker M. Dioxins and dibenzofurans in adipose tissue of US Vietnam veterans and controls. Am J Pub Health 81:344–349 (1991).
12. Centers for Disease Control Veterans' Health Studies. Serum 2,3,7,8-tetrachlorodibenzo-p-dioxin levels in U.S. Army Vietnam-era veterans. J Am Med Assoc 260:1249–54 (1988).
13. Stevens W, Thomas DC, Lyon JL, Till J, Kerber RA, Simon SL, Lloyd RD, Elghany NA and Preston-Martin S. Leukemia in Utah and radioactive fallout from the Nevada Test Site: a case–control study. J Am Med Assoc 264:585–591 (1990).
14. Hatch M, Beyea J, Nieves J, Susser W. Cancer near the Three Mile Island nuclear plant: radiation emissions. Am J Epidemiol 132:397–412 (1990).
15. Brown LM, Pattern LM, Blot WJ. Lung cancer in relation to environmental pollutants emitted from industrial sources. Environ. Res. 34:250–261 (1984).
16. Eifler CW, Stinnett SS, Buffler PA. Simple models for community exposure from point sources of pollution. Environ Res 25:139–146 (1981)
17. Wrensch M, Swan S, Murphy PJ, Liscomb J. Hydrogeologic assessment of exposure to solvent-contaminated drinking water: pregnancy outcomes in relation to exposure. Arch Environ Health 45:210–216 (1990).
18. Crump KS, Ng TH, Cuddihy RG. Cancer incidence patterns in the Denver metropolitan area in relation to the Rocky Flats plant. Am J Epidemiol 126:127–135 (1987).
19. Wertheimer N, Leeper F. Electrical wiring configuration and childhood cancer. Am J Epidemiol 109:273–284 (1979).
20. Milham S. Mortality from leukemia in workers exposed to electrical and magnetic fields. N Engl J Med 307:249 (1982).
21. Marshall JR, Graham S. Use of dual responses to increase validity of case-control studies. J Chron Dis 37:125–136 (1984).
22. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. Biometrics 36:167–171 (1980).
23. Wallace LA. The total exposure assessment methodology (TEAM) study. Washington DC: U.S. Environmental Protection Agency, 1987.
24. Smith TJ. Exposure assessment for occupational epidemiology. Am J Ind Med 12:249–268 (1987).
25. Hulka B, Wilcosky T, Griffith J. Biological markers in epidemiology. New York: Oxford University Press, 1990.
26. Stein Z, Hatch M. Biological markers in reproductive epidemiology: prospects and precautions. Environ Health Perspect 74:67–75 (1987).
27. Everson RB, Randerath E, Santella RM, Cefalo RC, Avitts TM, Randerath K. Detection of smoking-related covalent DNA adducts in human placenta. Science 231:54–57 (1986).
28. Hatch M, Warburton D, Santella R. Polycyclic aromatic hydrocarbon-DNA adducts in spontaneously aborted fetal tissue. Carcinogenesis 11:1673–1675 (1990).
29. Sposto R, Stram DO, Awa AA. An investigation of random errors in the DS86 dosimetry using data on chromosome aberrations and severe epilation (RERF TR 7–90). Hiroshima: Radiation Effects Research Foundation, 1990.
30. Meyer F, Verreault R. Erythrocyte selenium and breast cancer risk: brief report. Am J Epidemiol 125:917–919 (1987).
31. Vineis P, Caporaso N. Applications of biochemical epidemiology in the study of human carcinogenesis. Tumorigenesis 74:19–26 (1988).
32. Wald N, Boreham J, Bailey A. Serum retinol and subsequent risk of cancer. Br J Cancer 54:957–961 (1986).
33. London SK, Thomas DC, Bowman JD, Sobel E, Peters JM. Exposure to residential electric and magnetic fields and risk of childhood leukemia. Am J Epidemiol 134:923–937 (1991).
34. Rubin J. Multiple imputation for non-response in surveys. New York: John Wiley, 1987.
35. Rothman K, Poole C. A strengthening program for weak associations. Int J Epidemiol 17 (Suppl):955–959 (1988).
36. Antonarakis SE, and the Down's Syndrome Collaborative Group. Paternal origin of the extra chromosome in trisomy 21 as indicated by analysis of DNA polymorphisms. N Engl J Med 324:872–876 (1991).
37. Polissar L. The effect of migration on comparison of disease rates in geographic studies in the United States. Am J Epidemiol 111:175–182 (1980).
38. Cohen BL. Ecological versus case-control studies for testing a linear-no threshold dose–response relationship. Intl J Epidemiol 19:680–684 (1980).
39. Preston-Martin S, Thomas DC, White SC, Cohen D. Prior exposure to medical and dental x-rays related to tumors of the parotid gland. J Natl Cancer Inst 80:943–949 (1988).
40. Lawrence CE, Taylor PR, Trock BJ, Reilly AA. Trihalomethanes in drinking water and human colorectal cancer. J Natl Cancer Inst 72:563–68 (1984).
41. Armentrout DW. Overview of EPA major air data bases. In: Environmental epidemiology (Kopfler FC, Craun GF, eds). Chelsea, MI: Lewis Publishers, 1986; 149.
42. Armstrong BG. The effects of measurement errors on relative risk regressions. Am J Epidemiol 132:1176–1184 (1990).
43. Greenland S. Statistical uncertainty due to misclassification: implications for validation substudies. J Clin Epidemiol 41:1167–1174 (1988).
44. Schoenberg J, Klotz J, Wilcox HB, Nicholls GP, Gil-del-Real MT, Stemhagen A, Mason TJ. Case–control study of residential radon and lung cancer among New Jersey women. Cancer Res 50:6520–6524 (1990).
45. Brunekreef B, Noy D, Clausing P. Variability of exposure measurements in environmental epidemiology. Am J Epidemiol 125:892–898 (1987).
46. Beyea J, DeCicco J. Reestimating the noble gas releases from the Three Mile Island accident. Health Phys (submitted).
47. Doll R. Are we winning the fight against cancer? An epidemiological assessment. Eur J Cancer 26:500–508 (1990).
48. Davis DL, Hoel D, Lopez A. International trends in cancer mortality in France, West Germany, Italy, Japan, England and Wales, and the USA. Lancet 336:474–481 (1990).
49. Sorsa M, Ojajarvi A, Salomaa S. Cytogenetic surveillance of workers exposed to genotoxic chemicals: preliminary experiences from a prospective cancer study in a cytogenetic cohort. Teratog Carcinog Mutagen 10:215–221 (1990).
50. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 8:431–440 (1989).
51. Cartwright R, Ahmad RA, Barham-Hall D, Glashan RW, Rogers HJ, Higgins E, Kahn MA. Role of N-acetyltransferase phenotypes in bladder carcinogenesis: a pharmacogenetic epidemiologic approach to bladder cancer. Lancet 2:842–845 (1982).
52. Lewalter J, Korallus V. Blood protein conjugates and acetylation of aromatic amines. Int Arch Occup Environ Health 56:179–196 (1985).
53. Phillips AM, Silbergeld EK. Health effects studies of exposure from hazardous waste sites—where are we today? Am J Ind Med 8:1–7 (1985).
54. Greenland S. The effect of misclassification in the presence of covariates. Am J Epidemiol 112:564–569 (1980).
55. Kupper LL. Effects of the use of unreliable surrogate variables on the validity of epidemiology research studies. Am J Epidemiol 120:643–648 (1984).
56. Khoury MJ, Flanders WD, Greenland S, Adams MJ. On the measurement of susceptibility in epidemiologic studies. Am J Epidemiol 129:183–190 (1989).
57. Dohrenwend BP, Dohrenwend BS, Kasl SV, Warheit GJ, Bartlett GS, Chisholm RF, Goldsteen RL, Goldsteen K, Martin JL. Report of the Public Health and Safety Task Group on Behavioral Effects. In: Staff reports to the President's Commission on the Accident at Three Mile Island Public Health and Safety Task Force. No 052–003–00732–1. Washington, DC: U.S. Government Printing Office, 1979; 257–308.
58. Dickman S. Chernobyl effects not as bad as feared. Nature 351:335 (1991).
59. Ozonoff D, Colten ME, Cupples A, Heeren T, Schatzkin A, Mangione T, Dresner M, Colton T. Health problems reported by residents of a neighborhood contaminated by a hazardous waste facility. Am J Ind Med 11:581–597 (1987).
60. Rohl LH, Vernon SW, Weir FW, Pier SM, Sullivan G, Reed LJ. Community exposure to hazardous waste disposal sites: assessing reporting bias. Am J Epidemiol 122:418–433 (1985).

Basic Problems in Interaction Assessment

Sander Greenland

Department of Epidemiology, UCLA School of Public Health, Los Angeles, CA 90024-1772

This paper reviews problems with the definition and estimation of interactions in epidemiologic studies. Methods for modeling interactions and dose–response also are reviewed, and references to more detailed literature are provided. Concepts are illustrated in the context of evaluating the joint effects of household radon exposure and environmental tobacco smoke. — Environ Health Perspect 101(Suppl 4):59–66 (1993).

Key Words: Epidemiologic methods, interaction, relative risk, risk assessment

Introduction

In any study of the health effects of exposure mixtures, it is natural to ask whether or not observed effects are due to interactions of the mixtures' components; for example, one may inquire whether or not the effect of one component is modified by the effect of another (effect modification). Four major problems in addressing such a question are: *a*) The term "interaction" has no single definition but requires precise definition in order to be studied; *b*) even when it is precisely defined, there is often little study power to detect interaction; *c*) interactions are inevitably confounded with dose–response and latency relationships; and *d*) measurement errors, even if independent and nondifferential (random), may severely distort interaction assessment. This paper reviews these four problems and provides references to detailed literature on the problems. Definitions of the central concepts are reviewed first in order to provide a basis for precise problem statements. Next, the problems are described and illustrated in the context of evaluating effects of household radon exposure and environmental tobacco smoke (passive smoking). Finally, methods for dealing with the problems are reviewed.

Issues concerning mechanisms of interaction are not addressed here. As recently discussed by Thompson (*1*), epidemiologic data are limited inherently in their ability to discriminate among such mechanisms, because different mechanisms may predict identical patterns of disease. This problem is a logical one and persists even if the problems discussed here are eliminated.

This manuscript was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990 – September 1992.

I would like to thank W. Douglas Thompson and Irva Hertz-Picciotto for their helpful comments on this paper.

Definitions

Main Effects and Causal Effects

A source of ambiguity in the study of interactions (and indeed in the study of any effects) is the existence of multiple definitions of the term effect. Two major definitions exist. Ironically, both seem to have developed from the pioneering work on experimental design conducted by Fisher, Neyman, and others during the period between the first and second world wars.

The first definition, the parametric definition, is by far the most common today: An effect is a coefficient of a study exposure in a generalized linear model for the outcome of interest. [A generalized linear model is simply a linear model for some transformation of the expected outcome (*2*).] As an example , consider a log-linear (multiplicative) model for the rate R (in cases per person-year) of lung cancer in a cohort of married nonsmokers, given a certain exposure to spousal tobacco smoke x and radon level z, within a stratum k defined by some combination of age, sex, and (possibly) other determinants of lung cancer:

$$\log_e R_{kxz} = \alpha_k + \beta x + \gamma z \qquad [1a]$$

or, equivalently,

$$R_{kxz} = \exp(\alpha_k + \beta x + \gamma z). \qquad [1b]$$

Here, $k = 1, 2, 3,...$ simply indexes the various strata created by subdividing the cohort into subcohorts homogeneous on age, sex, etc., and α_k represents the log rate among stratum-k subjects who have no smoke or radon exposure ($x = 0$ and $z = 0$).

The coefficients β and γ of x and z traditionally are called the main effects of smoke and radon. This term suggests that β and γ represent some sort of causal action of smoke and radon on lung cancer rates. Such an interpretation could, however, be misleading. For example, the magnitude of β and γ would be affected by a failure to stratify on some cause of lung cancer that is distributed differentially across levels of radon and smoke exposure. For example, if asbestos exposure were associated with radon and smoke exposure in this cohort but the stratification did not include asbestos, one would say that the causal effect of smoke and radon was confounded with the asbestos effect, or that there was confounding by asbestos in the above model and in effect estimates derived from the model.

The parametric definition arose in the context of randomized experiments in agricultural research. Given randomization, the definition is not very misleading. If subjects had been randomized to the various smoke and radon levels, one would not expect smoke or radon to be associated with any potential confounder such as asbestos. Unfortunately, a causal interpretation of β and γ requires (among other things) absence of confounding; given the difficulty of assuring the latter condition, references to β and γ as main effects should be regarded as traditional rather than careful usage.

The difficulty with the parametric definition stems from the fact that model 1 describes many different subcohorts of the same cohort (one subcohort for every level of smoke and radon in the total cohort). That is, model 1 is a descriptive model with no causal or biological content. It only describes how the rate varies across subcohorts exposed to different levels of smoke and radon; it does not describe how any of the subcohorts would respond if smoke or radon levels were altered (unless, fortuitously, there is no confounding within the analysis strata). If, say, x is measured in "pack-decades smoked by spouse," a coefficient of $\beta = 0.182$ only says that the rate in subcohorts (strata) with one pack-decade of spousal smoking is on average $\exp(0.182) = 1.2$ times higher than in subcohorts with no spousal smoking; it does not say that this elevated rate is caused by the environmental tobacco smoke.

The second major definition of effect, the counterfactual definition, attempts to deal explicitly with the preceding problem. A causal effect is defined as a contrast of the outcome of a single subject under two different exposure conditions. Consider a married man in our cohort of nonsmokers. Suppose this man would have contracted lung cancer at age 85 if he had no smoke or radon exposure. However, since his marriage at age 20, he has been living with his wife who smokes a pack a day in a house that produces 1 WLM/year of radon-progeny exposure; these exposures result in his developing lung cancer at age 55. Thus, the causal effect of his actual smoke–radon exposure (relative to no exposure) on his incidence time is −30 years; that is, he contracted lung cancer (became an incident case of lung cancer) 30 years sooner than he would have in the absence of both exposures. Note that the condition of no exposure is counterfactual: It refers to what would have happened if, contrary to fact, the man had not been exposed to either smoke or radon.

Counterfactual models for causal effects extend at least as far back as the 1920s but have only seen extensive development over the last few decades (*3*). Modern development began in philosophy literature (*4*) and in educational statistics (*5*); another line of development was introduced into epidemiology by Rothman (*6*). In the ensuing decade, epidemiologists have employed counterfactual models to define biological interactions (*7–9*), confounding (*10*), and intermediate effects (*11,12*).

The present discussion ignores the problem of competing risks, that is, outcome events that remove a subject from risk of the outcome of interest. For lung cancer, all such competing risks are deaths from other causes, such as a car crash. Proper conceptualization of competing risks in a causal framework is somewhat controversial (*13,14*). To avoid complexities in presentation, the remaining discussion assumes that within levels of radon, smoking, age, sex, and other controlled covariates, competing risks occur independently of lung cancer. This assumption allows one to interpret all lung cancer incidence times as expected times, given no competing risks occur. Nevertheless, in any application, the assumption would need to be evaluated critically against any available background knowledge.

Statistical Interaction

In the theory surrounding generalized linear modeling, one commonly sees interactions defined as the coefficients of exposure products in the model. ("Product" here refers to multiplication, not the result of a chemical reaction.) Continuing our smoke–radon example, consider the model

$$R_{kxz} = \exp(\alpha_k + \beta x + \gamma z + \delta xz). \quad [2]$$

In the context of this model, the interaction of smoke and radon usually refers to the coefficient δ of the product xz of smoke and radon level; often, the entire product term δxz is called an interaction term. If model 1 is correct, it is said that no exposure interactions or nonlinearities are present on the log-linear or multiplicative scale.

Such usage of the term interaction has been criticized on several grounds (*15–17*). One criticism is that such usage is algebraic, divorced from any consideration of what constitutes interaction or synergy on the biological level. Another criticism is that such usage renders the presence or absence of interaction entirely dependent on the form of the statistical model one chooses; for the same data, interaction may appear to be present when using one model but absent when using another.

To illustrate the last point, suppose the lung cancer rates in our cohort follow the no-interaction log-linear model [1] with $\beta = 0.182$ per pack-decade spousal use and $\gamma = 0.693$ per 100 working-level months (WLM) radon-progeny exposure. Then the expected rates in stratum k will be $R_{k00} = \exp(\alpha_k)$ among subjects with no exposure,

$$R_{k10} = \exp(\alpha_k + 0.182) = 1.2\exp(\alpha_k) \quad [3]$$

among subjects with one pack-decade of spousal-smoke exposure but no radon-progeny exposure,

$$R_{k01} = \exp(\alpha_k + 0.693) = 2.0\exp(\alpha_k) \quad [4]$$

among subjects with no spousal-smoke exposure but 100 WLM radon-progeny exposure, and

$$R_{k11} = \exp(\alpha_k + 0.182 + 0.693) = 2.4\exp(\alpha_k) \quad [5]$$

among subjects with one pack-decade of spousal-smoke exposure and 100 WLM radon-progeny exposure. When expressing these four rates in the format of a linear excess-rate-ratio model

$$R_{kxz} = (1 + \beta^* x + \gamma^* z + \delta^* xz)\exp(\alpha_k), \quad [6]$$

one finds that

$$R_{k10} = 1.2\exp(\alpha_k) = (1 + \beta^*)\exp(\alpha_k), \quad [7]$$

$$R_{k01} = 2.0\exp(\alpha_k) = (1 + \gamma^*)\exp(\alpha_k), \quad [8]$$

and

$$R_{k11} = 2.4\exp(\alpha_k) = (1 + \beta^* + \gamma^* + \delta^*)\exp(\alpha_k). \quad [9]$$

The rate among the unexposed, $\exp(\alpha_k)$, cancels out of these three equations; this results in three simple linear equations with solutions $\beta^* = 0.2$, $\gamma^* = 1.0$, and $\delta^* = 1.2$. In other words, although no interaction is present when the rates are expressed in a log-linear model (i.e., $\delta = 0$), interaction is present when the rate ratios are expressed in a linear model (i.e., $\delta^* \neq 0$).

Causal Interactions

A different concept of interaction arises under the counterfactual model of effects. Consider again the man who developed lung cancer at age 55 after living 35 years with a wife who smoked a pack a day, in a house that produced 1 WLM/year of radon-progeny exposure. It was assumed that this man would have survived to develop lung cancer at age 85 only if all smoke particles and radon progeny in his household air had been removed (e.g., filtered) from the air he breathed.

Now ask whether or not the lung tumor he developed (at age 55) would have occurred later (if at all) if all the smoke particles but none of the radon progeny had been removed from the air. If the answer is yes, one says that spousal smoke advanced the incidence time of the subject's lung cancer. Also ask whether or not the tumor would have occurred later (if at all) if none of the smoke particles but all the radon progeny had been removed. If the answer to this question is yes, one may say that the radon advanced the incidence time. If the answer to both questions is yes, so that both exposures contributed to the advance in incidence time, one may say that the factors exhibited cooperative interaction (causal coaction, or synergism) in advancing the subject's incidence time.

To extend the example, suppose the subject would have developed lung cancer at age 70 if only the smoke particles had been removed and at age 65 if only the radon progeny had been removed. The advance in incidence time from 65 in the presence of smoke alone to age 55 in the presence of both exposures represents a portion of the total advance (of 30 years) that required the presence of both exposures to occur. Thus, the portion of the advance from 65 to 55 may be called the interaction effect or coaction of the two exposures.

Coaction is a special case of a more general concept of causal interaction or interdependence of causal effects, which

formalizes (in the counterfactual framework) concepts such as synergy, antagonism, and competitive action. Greenland and Poole (*9*) review this counterfactual theory and derive its connection to the sufficient-component causal theory of Rothman. Under the counterfactual theory, an instance of synergism between two factors is defined as an instance of disease in an individual that would not have occurred (by a specified time) if either or both factors had been absent. The connection to the above example is that lung cancer would not have occurred by age 55 if either or both factors (35 pack-years of spousal smoke exposure and 35 WLM of radon exposure) had been absent.

Note that the preceding counterfactual concept of synergism does not correspond to mechanism-based concepts of interaction [for example, see (*1*)]. Certain mechanisms do, however, predict response patterns consistent with this concept when interaction is present.

Connections among Definitions of Effects and Interaction

The definition of coaction just given bears no resemblance to the statistical definition of interaction; in particular, the concept of coaction is connected only indirectly to the statistical model for the rates. In terms of incidence time, the definition of coaction conflicts with the common definition of synergy as a total effect greater than the sum of the separate effects: In the example, the advance of lung-cancer time when both exposures are present (30 years) is less than the sum of the advance when only radon is present (85 − 70 = 15 years) and the advance when only smoke is present (85 − 65 = 20 years). Nevertheless, there is a connection among these concepts when the problem is reformulated in terms of incidence proportions (i.e., average risks of disease).

As an illustration of this connection, consider the subcohort of male nonsmokers whose exposure histories (up to the time they contract lung cancer) are, say, x = 1 pack/day spousal cigarette use and z = 1 WLM/year radon-progeny exposure, starting at age 20. Let $R_{xz}(t)$ be the actual proportion of this subcohort contracting lung cancer by age t. Define the three counterfactual proportions $R_{x0}(t)$ = proportion of the subcohort contracting lung cancer by age t if only the radon progeny had been removed from the environment; $R_{0z}(t)$ = proportion of the subcohort contracting lung cancer by age t if only the tobacco smoke had been removed; and $R_{00}(t)$ = proportion of the subcohort contracting lung cancer by age t if both the radon progeny and the smoke had been removed. From the four proportions just defined, one can compute two average-risk differences as measures of the effects radon and smoke would have had in the absence of the other,

$$RD_{x0}(t) = R_{x0}(t) - R_{00}(t) \text{ (radon)} \qquad [10]$$

and

$$RD_{0z}(t) = R_{0z}(t) - R_{00}(t) \text{ (smoke)} \qquad [11]$$

which are entirely counterfactual, and a difference that measures their actual combined effect,

$$RD_{xz}(t) = R_{xz}(t) - R_{00}(t). \qquad [12]$$

It can be shown that superadditivity of the differences,

$$RD_{xz}(t) > RD_{x0}(t) + RD_{0z}(t) \qquad [13]$$

can occur only if, in some subjects, radon and smoke causally interact in some of the cohort members; that is, only if coaction has occurred in some members (*8,9*). Note, however, that the converse is not true: Coaction may take place without superadditivity occurring (*8,9*).

It follows that an upper one-sided test of the additivity condition

$$RD_{xz}(t) = RD_{x0}(t) + RD_{0z}(t) \qquad [14]$$

may be regarded as a test for the occurrence of coaction. Various forms of this conclusion, and tests of additivity (model 14) as a test for synergism, can be found in the pharmacology literature as far back as the 1920s (*18*). The idea did not seem to attract notice in the epidemiologic literature until the 1970s; see Rothman (*15*), Koopman (*7*), and Miettinen (*8*) for some early formulations. Inequality 13 conforms to the common notion of synergy as a combined effect exceeding the sum of separate effects; note, however, that the effect referred to here is the effect of the exposures on an entire, homogeneously exposed subcohort. In contrast, the effect referred to in the definition of coaction refers to a single subject.

Inequality 13 also conforms to the definition of statistical interaction if one adopts an additive model for the average risks. To see this, define

$$\alpha(t) = R_{00}(t),\ \beta(x,t) = RD_{x0}(t),\ \gamma(z,t) = RD_{0z}(t), \qquad [15]$$

and

$$\delta(x,z,t) = \left[RD_{xz}(t) - RD_{x0}(t) - RD_{0z}(t)\right]. \qquad [16]$$

Then, with a little algebra, we see that inequality 13 implies

$$R_{xz}(t) = \alpha(t) + \beta(x,t) + \gamma(z,t) + \delta(x,z,t)$$

with

$$\delta(x,z,t) > 0. \qquad [17]$$

Thus, as before, superadditivity of effects (in particular, an additive-risk model with two causal exposures and a positive product term) implies the presence of interaction. Although the counterfactual and statistical definitions do not otherwise coincide, the superadditive case is, fortunately, the one of primary concern in the study of environmental and occupational hazards, for it is this case that is usually of most public-health concern (*16,17*).

The counterfactual proportions $R_{x0}(t)$, $R_{0z}(t)$, and $R_{00}(t)$ used for empirical testing of additivity would ordinarily be estimated from comparison groups that are subject to the various combinations of exposure. For example, $R_{00}(t)$ would be estimated from a subcohort with no (or negligible) smoke and radon exposure. This estimate must be adjusted for possible confounding.

In observational epidemiology, adequate adjustment may be difficult or impossible to achieve. There are usually too few subjects to allow simultaneous stratification on all important adjustment variables (confounders) and detailed comparison of exposure groups (although this problem generally is dealt with by using statistical models to estimate the average risks). More intractably, some important confounders may be impractical to measure accurately or to measure at all, and thus may remain uncontrolled. Problems arising from confounder mismeasurement are well recognized in the epidemiologic literature, however (*19–21*), and will not be a point of focus here. Instead, later sections will discuss the implications of exposure measurement problems for the assessment of interaction.

Some Problems in Interaction Assessment

The Power and Precision Problem

In epidemiologic settings, the power to detect statistical interactions is typically an order of magnitude less than the power to detect main effects; see Greenland (*22*) and Breslow and Day (*23*) for examples. Similarly, the variance of the interaction estimate will be an order of magnitude greater than the variance of the main-effects estimate under a no-interaction model.

An intuition for these results may be obtained by comparing variance formulas for estimates of main effect and interaction when both exposures x and z are dichotomous with levels 1 (exposed) and 0 (unexposed). Here we consider the basic linear-risk model

$$R_{kxz} = \alpha_k + \beta x + \gamma x + \delta xz \qquad [18]$$

which may be viewed as a special case of model 17. If there is only one stratum and δ is assumed to be zero (no interaction), the usual estimates of β will have a variance approximately equal to $V_1 V_0/(V_1 + V_0)$ where V_1 and V_0 are the variances of the estimates of $R_{k11} - R_{k01}$ and $R_{k10} - R_{k00}$. In contrast, the usual estimates of δ will have a variance equal to $V_1 + V_0$. The ratio of the latter variance to the first is $(V_1 + V_0)^2/V_1 V_0$, which equals 4 if $V_1 = V_0$ and will be larger if $V_1 \neq V_0$. Thus, in this simple case, the precision of the interaction estimate will be no more than a quarter that of the usual main-effect estimate. An identical result is obtained if one considers a log-linear rate model such as model 1 (*23*).

Situations involving continuous exposure measurements are considerably more complex, but nevertheless reveal that considerably larger study sizes are needed to study interactions than are required to detect effects (*24*). We will return to this issue in the discussion of designs for the study of interactions.

Confounding of Interaction and Dose–Response

In common epidemiologic usage, dose–response refers to the changes in risk produced by changes in a single exposure, whereas interaction refers to changes in risk produced by two or more exposures. Thomas (*25*) has pointed out that a major problem in the assessment of both dose–response and interaction is their tendency to confound one another, as well as their tendency to confound and be confounded with latency estimates. For example, consider the full quadratic generalization of model 18 to continuous exposures,

$$R_{kxz} = \alpha_k + \beta_1 x + \beta_2 x^2 + \gamma_1 z + \gamma_2 z^2 + \delta xz. \qquad [19]$$

In practice, x and z may be centered (that is, have their sample means subtracted off their observed values) to minimize correlation among the coefficient estimates. Even if this is done, however, the quadratic dose–response terms x^2 and z^2 will usually be highly correlated with the interaction (product) term xz; consequently, if β_2 and γ_2 are nonzero, x^2 and z^2 will act as confounders for xz, so that a biased estimate of δ will result if x^2 or z^2 is omitted from the model. In a symmetric fashion, omission of xz will bias the β_2 and γ_2 estimates if δ is nonzero.

More generally, failure to adequately model dose–response and latency can lead to bias in interaction estimates and vice-versa. Perhaps a more illuminating way to view this problem is to recognize that dose–response, latency, and interaction assessment are actually facets of a single task, namely assessment of the shape of the joint time-dependent dose–response surface relating incidence to both exposures. For example, model 19 specifies that this surface is quadratic; without specific prior knowledge about combined smoke and radon effects, there would be no basis for omitting any term from the model (unless the data clearly indicated a term was negligible).

Of course, model 19 is fairly restrictive as is its log-linear analogue (obtained by replacing R_{xz} with $\log_e R_{xz}$), and does not encompass the possibility of transforming x and z to improve model accuracy and to model latency. Some alternative modeling approaches will be discussed below. The present point is that dose–response and interaction should be viewed in a unified fashion if one wishes to avoid higher-order confounding.

Measurement Errors

In ordinary language, a measurement error is simply the act of recording an incorrect value for some variable on some subject. Statistical theory is concerned with the distribution of these errors in the study population and the relationship between true and measured values. For example, one may ask a number of questions involving the measured and true values for environmental tobacco smoke, such as: *a*) What is the distribution of true values x among subjects with measured values x_m? *b*) What is the distribution of measured values x_m among subjects with true values x? *c*) Do the errors in the measured values x_m vary systematically across levels of the true values x of smoke? (If not, the smoke errors are said to be additively homogeneous.) *d*) Do the errors in the measured values x_m vary systematically across levels of other variables? (If so, the errors are said to be differential; if not, the errors are said to be nondifferential.) *e*) Are the errors in the measured values x_m of smoke associated with the errors in the measured values x_m of radon? (If not, the errors in the two variables are said to be independent of each other.)

An analogous list can be made for the errors in measuring lung cancer incidence time. Traditionally, however, disease outcomes have been treated as dichotomies (diseased/not diseased), and errors in disease measurement have been treated as diagnostic errors, which are evaluated in terms of sensitivity (probability of true positive among cases) and specificity (probability of true negative among noncases).

The above listing does not exhaust the possibilities, and hence it may be clear that the topic of measurement error, and all its possible effects, can become exceedingly complex. It should not be surprising then that most studies on the topic are limited in scope and usually make several simplifying assumptions. Most commonly, errors are assumed to be independent and nondifferential, so that the answers to questions *d* and *e* and the analogous questions for disease are negative. One rationale for such an assumption in methodologic studies is that if some bias arises from well-behaved (independent nondifferential) errors, the same sort of bias or worse should be expected if the errors are not well behaved. Although this rationale is not valid universally (*26*), investigators often attempt to ensure that these errors will be independent and nondifferential, and so such errors are worth studying in detail.

Nevertheless, it should be recognized that optimistic conclusions based on assuming independent nondifferential errors cannot be extended to dependent or differential errors, and that the errors actually occurring in a study can become differential under ordinary circumstances. Consider, for example, exposure measurements over time. Such measurements often are based on historical records or, worse, subject memory. In such situations, exposure measurements for the more distant past may be less accurate than measurements for more recent exposure; if so, accuracy of cumulative exposure measurement will vary with any variable correlated with calendar time, such as another exposure. Even if the intrinsic accuracy of the exposure measurements do not vary over time, the degree of bias produced by measurement errors may still vary over time (*27*). Similar problems will arise if accuracy of outcome measurement (e.g., disease diagnosis) varies over time.

The Impact of Measurement Errors

The impact of measurement errors on main-effect estimates has been studied extensively, especially for situations involving independent nondifferential error. One well-known result is that independent nondifferential errors in the classification of a dichotomous exposure and covariate cannot produce bias away from the null value

of the exposure effect; for example, any bias produced by such error in the estimate of β in model 1 will be towards zero. This result, while useful, is often stated without mention of the assumptions of independent errors and dichotomous exposure.

Unfortunately, violations of either assumption can result in bias away from the null; Dosemeci et al. (*28*) show that independent nondifferential classification error can produce bias away from the null if the exposure has as few as three levels. It is not clear, however, how often such bias occurs in practice, and there are a number of special error models under which the estimated coefficients in linear or log-linear models can only be biased towards the null. For example, this is so under the classical model, in which the measured value x_m is given by $x_m = x + \varepsilon_x$, where x is the true value, ε_x is the x error, and x and ε_x are normally distributed with ε_x independent of all other variables (including x)—that is, the error is independent, additively homogeneous, nondifferential, and normal. Although these conditions are restrictive, the result extends to various cases involving nonnormal exposures and errors. Extension to multiplicative errors, with $x_m = x \cdot \varepsilon_x$ and x and ε_x strictly positive, follows by using $\log(x_m) = \log(x_t) + \log(\varepsilon_x)$ in place of x_m as the regressor variable. These and other results for special models are reviewed by Armstrong (*29*). Lubin et al. (*25*) specifically consider models for radon measurement to evaluate the impact of measurement errors in studies of tobacco smoke, radon, and lung cancer.

The impact of measurement errors on interaction estimates has been studied less thoroughly. Independent nondifferential classification errors can produce spurious appearances of interaction and can mask true interactions, depending on other features of situation (*19*). More generally, the interaction coefficient δ in models 17 and 18 may be biased towards or away from the null by independent nondifferential errors in the study covariates (regressors); errors in disease classification may further aggravate such biases, thus distorting the entire shape of the dose–response surface. These results easily extend to situations involving arbitrary polytomous or continuous exposures (Appendix). Nevertheless, there are a number of special cases in which nondifferential independent error will not affect the validity of tests for interaction, and may rarely or never produce bias away from the null; for example, if the true values were distributed jointly and normally and if the errors were independent, additively homogeneous, nondifferential, and normal (that is, if $x_m = x + \varepsilon_x$ and $z_m = z + \varepsilon_z$, where x, z are bivariate normal and the errors ε_x, ε_z are normal and independent of x, z, and each other), or if the errors were independent, nondifferential, and x and z were not associated with each other (Appendix).

The distortion of dose–response and interaction estimates produced by measurement error depends heavily on the particulars of the study distribution of exposures and errors. Thus, rather than rely on any general (and possibly misleading) conclusions, it may be best to evaluate the effects of measurement error on a study-specific basis, using methods of the sort discussed in the next section. In the particular case of environmental tobacco smoke and radon, measurement errors may render the study of interactions infeasible due to attenuated power (*24*); a similar conclusion may apply to most other epidemiologic studies of environmental exposures.

Coping with the Problems

Designs for Assessing Interactions and Dose–Response

In studies involving primary subject selection, power for detection of interactions can be increased by using special sampling plans. Unfortunately, a major obstacle in employing such designs is that they require a priori specification of a number of parameters that may be only vaguely known, if at all. For cohort studies, one must be able to specify likely values for the intercept and main-effect parameters (e.g., α, β, γ in model 18) in the model of interest, as well as a value for the interaction parameter (δ) for which one wishes to maximize power or precision. For case–control studies, the intercept need not be specified, but one must have some idea of the exposure distributions in the population serving as the source of cases and controls.

A considerable amount of literature exists for choosing optimal designs, at least in the cohort framework; Seber and Wild (*30*) provide references to the linear-model literature and also review design methods for nonlinear models. Although this literature is highly technical, a few general conclusions can be drawn, especially in the special case of studying departures from risk or rate additivity.

The optimal design for detecting departures from additivity will not correspond to the optimal design for detecting departures from linearity of the dose–response curve for each exposure. Nor will either of these designs correspond to the optimal design for detecting main effects; however, the presence of main effects will hopefully have been established before embarking on a specialized study of interactions.

Because one will have to simultaneously consider interaction and dose–response, as explained earlier, it may be best to select subjects to maximize precision of the estimated dose–response surface. In this approach, interaction represents but one of several potentially important departures from linearity of the joint dose–response surface relating smoke (x) and radon (z) to risk. For example, consider the quadratic-risk model given in model 19. A good design for studying such a model would select subjects to enhance the precision of estimates for β_2 and γ_2, as well as δ.

More generally, one would want to allow for response surfaces other than quadratic, including perhaps unanticipated shapes. One simple cohort design to help achieve this end would try and insure that subjects are distributed evenly across the joint range of smoke and radon levels (that is, across the combinations of x and z).

The case–control situation is not addressed as easily, for it is the case–control ratio rather than the joint exposure distribution that is controlled by the investigator. Nevertheless, if one is willing to sacrifice the ability to estimate the main effect of one of the exposures, one also may manipulate the marginal distribution of that exposure by, for example, case–control matching; see Smith and Day (*31*) and Thomas and Greenland (*32*) for some elementary studies of the impact of matching on interaction assessment in the context of log-linear interactions. For interaction assessment, one can expect that certain highly variable matching ratios will offer more precision than fixed ratios: Relatively few controls per case would be needed in strata with many cases, but relatively many controls per case would be needed in strata with few cases.

If one already knows the joint distribution of disease and one of the exposures in the source population, it may be most efficient to employ a two-stage design rather than a conventional matched design; see Cain and Breslow (*33*) for further discussion of this point.

Modeling Interactions and Dose–Response

The confounding of interactions and dose–response can be overcome if one has accurate information on the values of the variables (here, smoke and radon) over a reasonably broad range of combinations of the variables. Even with accurate and broad-ranging measurements, however, one must take care to employ a model form flexible enough to accurately approximate the true dose–response surface. Because the shape of the true surface usually is

unknown (and is in fact what is under study), a safe strategy would be to employ as flexible a model form as practical.

The most flexible approaches available are nonparametric regression methods, such as bivariate smoothers; for examples, see Hastie and Tibshirani (*34*). Unfortunately, these methods are not yet implemented widely in software, are impractical for handling more than a few regressors, and can require fairly large samples for reasonable performance. An easier approach, with somewhat less flexibilty, is generalized additive modeling (*34*). As an example, the generalized-additive analogue of model 1 would be

$$\log_e(R_{xz}) = \alpha_k + \beta(x) + \gamma(z), \quad [20]$$

where $\beta(x)$ and $\gamma(x)$ are now unspecified functions of x and z that will be estimated from the data. Unlike model 1, which constrains dose–response to be log linear, model 20 allows the dose–response for smoke and radon to be any shape at all. Both models 1 and 20 do, however, imply that the shape for the smoke dose–response does not change across levels of radon or covariates, and the shape for the radon dose–response does not change across levels of smoke or covariates; this set of constraints is called the no-additive-interaction or parallelism condition. Model 20 is easily fit using the GAIM software package (*35*). To generalize model 20 to allow for departures from additivity, one may add a product-term function to obtain

$$\log_e(R_{kxz}) = \alpha_k + \beta(x) + \gamma(z) + \delta(xz). \quad [21]$$

This is one of several possible generalized-additive analogues of model 1. Unlike model 20, it does not constrain the dose–response surface to contain parallel dose–response curves.

All the models given so far imply that the shape of the dose–response surface does not change across the covariate strata (i.e., there is no additive interaction with covariates). To get around this restriction, one could model the covariate effects in detail and add interaction terms between the covariates and exposures to the model. Among the drawbacks of this strategy is that the resulting model may have too many terms for the fitting procedure to work. Even if the model can be fit, the individual terms may be estimated with little accuracy. The individual terms also may be difficult to interpret, although this need not be a problem if one focuses on graphs of the response surfaces instead of on model terms.

Further extensions of the above models may be obtained by considering other transformations of the outcome measure, as in the additive logit model in which

$$\text{logit}R_{kxz} = \alpha_k + \beta(x) + \gamma(z), \quad [22]$$

where logit $R = \log_e[R/(1-R)]$. One also may employ incidence times or rates in place of risks as the outcome measure in the above models. The latter models often fit better and may even obviate the need for product terms in the model. They also allow for straightforward incorporation of time-dependent exposures in the model, an obvious advantage in longitudinal studies of exposures such as smoke and radon. Nevertheless, tests of the no-coaction hypothesis still correspond to testing the fit of an additive-risk model (such as model 17 or 18) (*36*).

Unfortunately, additive-risk models cannot be fit to case–control data unless one has sufficient external information to reconstruct the population risks from the data. For unmatched studies, all one needs is an estimate of the crude disease rate in the source population of cases and controls or knowledge of the case and control sample fractions. For matched studies, one must have the crude rates or sampling fractions within levels of the matching factors. Given this information, however, one may fit the same variety of model forms as used for cohort data (*37*).

For further discussions of modeling issues and techniques see Breslow and Day (*23*), McCullagh and Nelder (*2*), and Hastie and Tibshirani (*34*). Less technical overviews of modeling are given by Greenland (*38*) and Checkoway et al. (*39*).

Evaluating and Correcting for Measurement Error

The best means of coping with measurement error is, of course, not to have it. Because this ideal is not attainable in typical environmental and occupational studies, evaluation of measurement error and its effects is an essential component of any informative study. Most evaluations are limited to narrative review of factors influencing errors and the implications for bias; most commonly, these evaluations comprise arguments that exposure-measurement errors were independent and nondifferential and hence produced only bias towards the null. As shown earlier, however, such arguments are of little use in interaction assessment, because independent, nondifferential misclassification may bias interaction terms in any direction.

Much more can be done if data are available about the accuracy of the exposure and covariate measurements in the study. In the best situation a validation substudy is conducted in which exposure is remeasured in a subsample of subjects using criterion methods, that is, methods more accurate than the general methods applied to all subjects. The association of the criterion and general measurements, as estimated from the validation substudy, may then be used to correct coefficient estimates obtained from the full study cohort. Correction methods also may be applied if the criterion-general measurement association is estimated from data external to the study (although, in the latter case, one must assume that this association is the same in both the study and the external data). There is now an extraordinary variety of validation-based correction methods available (for example *40–42*).

If a criterion measurement is unavailable, it still may be possible to obtain a more limited correction of coefficient estimates using a reliability substudy in which replications of the general measurement are obtained on a subsample of subjects. Again, there is a variety of reliability-based correction methods (e.g., ref. *43*).

If neither validation data nor reliability data are available, but some educated guesses can be made about the distributions of exposure and covariate errors, one may conduct a sensitivity analysis of the study results. In such an analysis, various hypothesized error distributions are used to correct the study results; one thus sees how sensitive estimates are to assumptions about the error distribution. This analysis is conducted easily under various simplifying assumptions (*29*). If the study variables are discrete, matrix formulas for correcting contingency-table results can be applied (*40*, Appendix), and these are programmed easily in matrix languages such as GAUSS, SC, S-PLUS, and SAS IML.

Conclusions

Given the difficulties inherent in attempting to study interactions with epidemiologic data, design and analysis is best focused on accurate estimation of the entire dose–response surface relating incidence to covariates, rather than on isolated aspects of this surface, such as statistical interaction. One may, of course, test the departure of the data from surfaces predicted by various causal models, such as the no-coaction model (*7,9*) or the simple independent-action model (*44*), but the power and validity of these tests will be nearly optimal under the same conditions that insure accuracy of dose–response estimation, such as well-balanced exposure distributions and accurate exposure measurement.

Flexible modeling and, where possible, quantitative evaluation of measurement error

will help achieve the most accurate assessment of interaction possible with available data. Nevertheless, because of limitations of power and because of distortions produced by measurement error, one should be cautious about the potential of environmental epidemiology for interaction assessment.

Appendix

For simplicity, suppose we have just one stratum, and let $P(x,y \mid x_m,z_m)$ be the probability that a subject with measured smoke and radon exposures x_m and z_m has true levels x and z; note that $\Sigma_{xz}P(x,y \mid x_m,z_m) = 1$ (here, Σ_{xz} indicates the sum over all possible values of x and z). Let

$$\bar{x}(x_m,z_m) \equiv \sum_{xz} xP(x,y \mid x_m,z_m)$$

and

$$\bar{z}(x_m,z_m) \equiv \sum_{xz} zP(x,y \mid x_m,z_m) \quad [23]$$

be the means of the true smoke and radon levels among subjects with measured levels x_m and z_m; let R_{xz} be the average risk among subjects with true levels x and z; and suppose R_{xz} follows the no-interaction linear-risk model (model 18 with $\delta = 0$). Then the average risk among subjects with measured levels x_m and z_m will be

$$\begin{aligned} R(x_m,z_m) &= \sum_{xz} P(x,z \mid x_m,z_m)R_{xz} \\ &= \sum_{xz} P(x,z \mid x_m,z_m)\,(\alpha + \beta x + \gamma z) \\ &= \alpha \cdot 1 + \beta\sum_{xz} xP(x,z \mid x_m,z_m) \\ &\quad + \gamma\sum_{xz} zP(x,z \mid x_m,z_m) \\ &= \alpha + \beta\bar{x}(x_m,z_m) + \gamma\bar{z}(x_m,z_m). \end{aligned} \quad [24]$$

Now let $RD(x_m,z_m) = R(x_m,z_m) - R(0,0)$ be the risk difference between subjects with measured levels x_m,z_m and subjects measured as having no exposure. Then

$$\begin{aligned} &RD(x_m,z_m) \\ &\quad = \alpha + \beta\bar{x}(x_m,z_m) + \gamma\bar{z}(x_m,z_m) \\ &\quad - \left[\alpha + \beta\bar{x}(0,0) + \gamma\bar{z}(0,0)\right] \\ &\quad = \beta\left[\bar{x}(x_m,z_m) - \bar{x}(0,0)\right] \\ &\quad + \gamma\left[\bar{z}(x_m,z_m)\right] - \bar{z}(0,0)\Big]; \end{aligned} \quad [25]$$

in contrast, for subjects measured as exposed to only one of the two exposures, we have

$$\begin{aligned} &RD(x_m,0) + RD(0,z_m) \\ &\quad = \beta\left[\bar{x}(x_m,0) - \bar{x}(0,0)\right] \\ &\quad + \gamma\left[\bar{z}(x_m,0) - \bar{z}(0,0)\right] \\ &\quad + \beta\left[\bar{x}(0,z_m) - \bar{x}(0,0)\right] \\ &\quad + \gamma\left[\bar{z}(0,z_m) - \bar{z}(0,0)\right] \\ &\quad = \beta\left[\bar{x}(x_m,0) + \bar{x}(0,z_m) - 2\bar{x}(0,0)\right] \\ &\quad + \gamma\left[\bar{z}(x_m,0) + \bar{z}(0,z_m) - 2\bar{z}(0,0)\right]. \end{aligned} \quad [26]$$

Thus, except in certain special cases,

$$RD(x_m,z_m) \neq RD(x_m,0) + RD(0,z_m). \quad [27]$$

that is, the risks based on the measured exposures need not be additive, and this is so even if the measurement error is independent and nondifferential and the risks based on the true exposures are additive.

Additivity will be preserved (i.e., 25 will equal 26 under model 18 with $\delta = 0$) if the mean true levels $\bar{x}$ and $\bar{z}$ depend on the measured levels x_m and z_m in an additive fashion, for then

$$\begin{aligned} &\bar{x}(x_m,z_m) - \bar{x}(0,0) \\ &\quad = \bar{x}(x_m,0) + \bar{x}(0,z_m) - 2\bar{x}(0,0) \end{aligned} \quad [28]$$

and

$$\begin{aligned} &\bar{z}(x_m,z_m) - \bar{z}(0,0) \\ &\quad = \bar{z}(x_m,0) + \bar{z}(0,z_m) - 2\bar{z}(0,0) \end{aligned} \quad [29]$$

This would occur, for example, if the errors were independent nondifferential and x and z were unassociated, or if x and z were bivariate normal and their respective errors were independent normal with homogeneous variance. Additivity also will be preserved under "Berkson error" [see Armstrong (*29*) for discussion of Berkson error in the context of main-effect estimates].

REFERENCES

1. Thompson WD. Effect modification and the limits of biological inference from epidemiologic data. J Clin Epidemiol 44:221–232 (1991).
2. McCullagh P, Nelder JA. Generalized linear models, 2nd ed. New York: Chapman and Hall, 1989.
3. Rubin DB. Comment: Neyman and causal inference in experiments and observational studies. Stat Sci 5:472–480 (1990).
4. Lewis D. Causation. J Philos 70:556–567 (1973).
5. Rubin DB. Estimating causal effects of treatment in randomized and nonrandomized studies. J Educ Psych 66:688–701 (1974).
6. Rothman KJ. Causes. Am J Epidemiol 104:587–592 (1976).
7. Koopman J. Interaction between discrete causes. Am J Epidemiol 113:716–724 (1981).
8. Miettinen OS. Causal and preventive interdependence: elementary principles. Scand J Work Environ Health 18:159–168 (1982).
9. Greenland S, Poole C. Invariants and noninvariants in the concept of interdependent effects. Scand J Work Environ Health 14:125–129 (1988).
10. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol 15:412–418 (1986).
11. Robins JM. Control of confounding by intermediate variables. Stat Med 8:679–701 (1989).
12. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology 3:143–155 (1992).
13. Slud EB. Dependent competing risks and the latent-failure model (letter). Biometrics 44:1203–1204 (1988).
14. Prentice RL, Kalbfleisch JL. Letter to editor. Biometrics 44:1205 (1988).
15. Rothman KJ. Synergy and antagonism in cause–effect relationships. Am J Epidemiol 99:385–388 (1974).
16. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. Am J Epidemiol 112:467–470 (1980).
17. Rothman KJ. Modern epidemiology. Boston: Little, Brown, 1986.
18. Greenland S. Elementary models for biological interaction. J Haz Materials 10:449–454 (1985).
19. Greenland S. The effect of misclassification in the presence of covariates. Am J Epidemiol 112:564–569 (1980).
20. Kupper LL. Effects of the use of unreliable surrogate variables on the validity of epidemiologic research results. Am J Epidemiol 120: 643–648 (1984).
21. Savitz DA, Barón AE. Estimating and correcting for confounder misclassification. Am J Epidemiol 129:1062–1071 (1989).
22. Greenland S. Tests for interaction in epidemiologic studies: a review and a study of power. Stat Med 2:243–251 (1983).
23. Breslow NE, Day NE. Statistical methods in cancer research II: the design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, 1987.
24. Lubin JH, Samet JM, Weinberg CR. Design issues in epidemiologic studies of indoor exposure to radon and risk of lung cancer. Health Physics 59:807–817 (1990).
25. Thomas DC. Are dose-response, synergy, and latency confounded? Alexandria, VA: American Statistical Association, 1981.
26. Greenland S, Robins JM. Confounding and misclassification. Am J Epidemiol 122:495–506 (1985).

27. Greenland S. Interpreting time-related trends in effect estimates. J Chron Dis 40 (Suppl 2):17S–24S (1987).
28. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification always bias a true effect toward the null value? Am J Epidemiol 132:746–748 (1990).
29. Armstrong BG. The effects of measurement error on relative risk regressions. Am J Epidemiol 132:1176–1184 (1990).
30. Seber GAF, Wild CJ. Nonlinear regression. New York: Wiley, 1989.
31. Smith PG, Day NE. The design of case–control studies: the influence of confounding and interaction effects. Int J Epidemiol 13:356–365 (1984).
32. Thomas DC, Greenland S. The efficiency of matching in case–control studies of risk-factor interactions. J Chron Dis 38:569–574 (1985).
33. Cain KC, Breslow NE. Logistic regression analysis and efficient design for two-stage studies. Am J Epidemiol 128:1198–1206 (1988).
34. Hastie TJ, Tibshirani RJ. Generalized additive models. New York: Chapman and Hall, 1990.
35. Almudevar T, Tibshirani R. GAIM version 1.1. Toronto: S.N. Tibshirani Enterprises, 1991.
36. Greenland S. Additive risk versus additive relative risk models. Epidemiology 4:32–36 (1993).
37. Greenland S. Multivariate estimation of exposure-specific incidence. J Chron Dis 34:445–453 (1981).
38. Greenland S. Modeling and variable selection in epidemiologic studies. Am J Public Health 79:340–349 (1989).
39. Checkoway H, Pearce NE, Crawford-Brown D. Research methods in occupational epidemiology. New York: Oxford University Press, 1989.
40. Greenland S. Variance estimation for epidemiologic effect estimates under misclassification. Stat Med 7:745–757 (1988).
41. Rosner B, Spiegelman D, Willer WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. Am J Epidemiol 132:734–745 (1990).
42. Pepe MS, Fleming TR. A nonparametric method for dealing with mismeasured covariate data. J Am Stat Assoc 86:108–113(1991).
43. Thompson WD. Kappa and attenuation of the odds ratio. Epidemiology 1:357–369 (1990).
44. Weinberg CR. Applicability of the simple independent-action model to epidemiologic studies involving two factors and a dichotomous outcome. Am J Epidemiol 123:162–173 (1986).

The Place of Epidemiology in Environmental Decisions: Needed Support for the Development of Risk Assessment Policy

Raymond Richard Neutra[1*] and Dimitrios Trichopoulos[2]

[1]California Department of Health Services, 5900 Hollis St., Suite 8, Emeryville, CA 94704; [2]Department of Epidemiology, Harvard School of Public Health, 667 Huntington Avenue, Boston, MA 02115

Some of the most challenging problems in the use of epidemiology for regulatory policy concern summarizing epidemiological and other kinds of information to create a weight of evidence. Another frequent issue is whether to embark on epidemiological study. There are also concerns that negative results never see the light of day. These and other meta-issues are worthy of funded evaluation by expert work groups. — Environ Health Perspect 101(Suppl 4): 67–69 (1993).

Key Words: Epidemiology, risk assessment policy, research needs

Risk Assessment Policy Needs Go Beyond Research Needs

So far we have discussed areas of exposure analysis, study design, and data analysis in which methodological improvements are needed and worthy of research support. For the most part, this support will go to individual researchers to help finance the time spent developing and testing new methods.

In this paper we discuss scientific activities, such as the systematic and consistent summarization of a body of evidence, the decision to initiate more study, and the conclusion that enough is enough.

These are issues of risk assessment policy and research strategy, and while not as value laden as risk management (which we do not address in this document), they are by nature a matter of scientific consensus. For this reason, support for the formation of a risk assessment policy often involves development of draft documents by governmental and nongovernmental scientists and the systematic review and development of procedure by groups of researchers. We begin by raising some of the issues of a risk asssessment policy and discuss in the last section how this policy or related research needs could be supported. To place these considerations in context, we must remember that most risk assessment must proceed with only animal data. Therefore, the issues raised here, though important, apply to a small proportion of regulatory decisions.

This manuscript was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990 – September 1992.

*Author to whom correspondence should be addressed.

Putting It All Together: Methodological Issues in Weighing a Body of Evidence

Repeated strong epidemiological findings can implicate a remediable environmental exposure even without supporting animal toxicological evidence or identifying a responsible agent. An example would be the well-known strong association between carcinoma of the nasal sinus and cabinet making (*1*). Industrial hygienic precautions can be instituted even before a better understanding of the responsible mechanism is clarified. Here, epidemiology alone suffices to drive regulation. Strong consistent results require no ingenuity to summarize.

More frequently, the human epidemiological results are not distinguishable from the null, or the dose–response slope is so low that bias or confounding could plausibly account for the observed association. Alternatively, the results may implicate something in the general environment that cannot be avoided by some easy measure that could, like a cabinetmakers dust mask, be applied without an understanding of the responsible agent. In this case, usually other disciplines are needed to pinpoint the offending agent and its means of control. For all of these reasons, most environmental policy is set after considering the integrated information from clinical medicine, basic physical and biological science, animal toxicology, exposure analysis, and a body of epidemiological evidence. When all of this evidence points toward the same conclusion, policy decisions are simplified. Often, however, the evidence is conflicting. Some studies are called positive, and others are said to be negative, that is, indistinguishable from the null.

The terms positive and negative suggest a solidity that is misleading. One school of thought reserves these terms for statistically significant associations and tends to view any association that does not achieve the preset p value to be as good as no association at all. Another school suggests that information from all studies should be pooled and the decision to believe the results should be based on Bayesian approaches that consider prior plausibility and the cost of a false positive or a false negative result. This issue becomes particularly difficult when, as is the case in the current debate about low-frequency electromagnetic fields (*2*), the body of positive epidemiological evidence has very weak biological plausibility. Intuitively it is clear that a higher relative risk or a greater number of confirmatory studies are necessary for such a situation than would be for an agent that is similar to a previously studied agent whose mechanism of action is well understood. The acceptance of erionite as a carcinogenic mineral fiber comes to mind. The documentation of two villages with a rate ratio of 9000 and one animal study demonstrating carcinogenicity was sufficient for the International Agency for Research on Cancer (IARC) (*3*) to list erionite as a carcinogen. The biological plausibility weighed heavily here. Methodological research into how one uses epidemiological and other types of information to update prior probability assessments practically and intelligibly for environmental decision makers is, therefore, of high priority.

Agents that act by nonthreshold mechanisms, as is the case with many carcinogens, could conceivably produce unacceptable numbers of disease in the population when there is widespread exposure. This could happen even when the exposures are so low that the increase in relative risk is undetectable.

Although epidemiological studies cannot rule out effects of societal concern in this situation, such negative or null studies can sometimes help assess whether humans have higher or lower sensitivity to an agent than expected on the basis of animal bioassays. This is a special topic within the problem of summarizing evidence.

Another problem that often faces those who are trying to weigh a body of evidence is that negative studies (not distinguishable from the null) are thought to be less likely to be published than positive studies. Censoring may occur at the level of the researcher or the journal. For example, a researcher may suggest a positive association between a disease outcome and a variable that was not originally included in the main hypothesis. The researcher probably would not report a surprising lack of association under the same circumstances. An author whose main hypothesis was not supported in a study may decide not to submit an article or may become discouraged after receiving routine editorial criticisms. All of this could skew the available evidence. This problem has been discussed previously (*4*). It is time to see if the phenomenon is substantial and to evaluate which, if any, potential remedies should be applied.

Contexts in which Follow-up Environmental Epidemiologic Studies are Recommended

The usual motivation for academically based epidemiologic research is to pursue a credible hypothesis in a setting that promises a high likelihood of providing a persuasive answer because the amount of exposure, the size of the study population, and the ability to control bias and confounding are all favorable. Research priorities from funding agencies thus give as much attention to feasibility as to the potential importance of the project being funded. Hypotheses often derive from basic science considerations or from animal or other epidemiological evidence. Public agencies, on the other hand, are often directed to carry out studies whose answer would be of great policy interest even though the low biological credibility of the tested hypothesis or the conditions of study militate against the likelihood of a persuasive result.

Frequently, one or a few studies initiated in either way are not considered persuasive. When current epidemiological information is insufficient, one is faced with the problem of deciding if any additional epidemiological study is likely to be helpful (for example, where animal evidence suggests the possibility of an effect large enough to be of social concern but too small to be detected toxicologically or epidemiologically under usual exposure scenarios). If epidemiology offers hope for demonstrating an effect, how strong must a collection of positive studies be to implicate an agent or estimate its potency in humans? How strong must a collection of negative studies be to give a clean bill of health to an agent that is thought to act by a mechanism that should display a threshold of effect? In regulatory toxicology, no single study is ever considered definitive. Instead, a specified number of statistically significant studies in several species and laboratory settings is routinely required. This policy is pursued regardless of considerations such as strength of effect, prior plausibility, or the social costs of false positive or negative results. While this procedure would not be advocated for epidemiology, it does remind us that no single study, especially a screening study, is likely to be definitive and that initiating an investigation for scientific or public policy reasons usually commits one to a sequence of studies until a body of evidence has accumulated. The principles that guide the initiation and termination of this commitment need elucidation.

When concerned segments of the public demand a study in a particular setting, for instance the Love Canal or large areas of Los Angeles where there is aerial application of malathion for the Mediterranean fruit fly, they often have a legitimate desire to participate in the design, conduct, and interpretation of the results. They often have unique "shoe leather" experience as to routes of exposure, hypotheses about effects that they wish to be addressed, and susceptibility to and concerns about potential conflicts of interest in the analysis process. Often the results (or media interpretation of the results) from one such location can have profound impact on national policy.

Several methodological research questions arise from the involvement of members of the public. How well do various techniques of involvement work, practically, for citizen satisfaction, and with regard to avoiding bias in study results? How does the presentation of epidemiological results and the social setting in which that presentation takes place influence the understanding and acceptance of results? How can one safely assume that a local community correctly registers and remembers study results carried out on its behalf, given that its perception of the results influences local and national policy?

The availability of morbidity and mortality data bases offers both opportunities and dilemmas. Academic epidemiology has traditionally turned to available data bases to look for unsuspected variations with regard to person, place, and time as a first step to unearthing the activity of some causal agent. This approach was often useful in the early days of infectious disease epidemiology; and recently there have been some successes in China, where sharp regional variations in chronic disease rates (*5*) have led to the discovery of causal agents of indoor air pollution [female lung cancer in southern China (*6*)] or deficiencies or excesses in trace elements [selenium (*7*)].

Despite these isolated instances and the usefulness of these data bases in occupational and life-style epidemiology, it is striking how difficult it is to find examples in modern developed countries in which the routine surveillance of morbidity and mortality for temporal or spatial variations has led to the discovery of new causal factors in the physical environment. These sources of information have been more useful for following the course and control of disease of known origin and for testing specific hypotheses that arose from other considerations. Despite this unpromising record, there are politicians and scientists who have high hopes for screening studies based on the analysis of routine data. This can be fruitful even when there is no biologically based hypothesis, but the opportunity for chance associations from multiple comparisons is great. In a more hypothesis-testing mode, one can explore existing data to ask whether a single locality (or all localities) containing some environmental hazard has (or have) higher rates of a particular disease. The opportunities for chance associations after multiple comparisons also are great. The availability of such data also facilitates the generic clustering study in which one asks if a disease clusters more than chance or demographics would suggest. The answer allows the researcher to assess the hypothesis that a disease is spread from person to person or from a point source. The traditional academic hypothesis-generating strategies thus have potential for initiating a wild-goose chase in environmental epidemiology. There is a need for a clear rationale for initiating such endeavors while considering the subsequent commitment it entails.

Methodological Research Questions

We have raised some research and policy issues that arise when summarizing epidemiological evidence or deciding to initiate new studies in service to the regulatory process. We will briefly discuss them below.

How can funding agencies advance the state of knowledge and the quality of practices in these several areas? As mentioned before, science policy requires a consensus process and, therefore, requires the support of researchers working as a group. Efficiency dictates that some researchers be supported to prepare the

groundwork or oversee and summarize the consensus process while others are supported to participate in it. Care should be given to inviting participants with a range of disciplines and backgrounds so that persons with various kinds of practical experience are balanced by individuals who are not grounded in the more traditional ways of doing or thinking about things. A variety of mechanisms could be appropriate for the different areas dealt with below. These include supporting outside scientists to work with government scientists to develop draft guidelines or work out discussions of the rationale for approaching a particular problem. They include RFP support for individuals, the intensive work group sessions to resolve a problem procedure, or a workshop or conference to summarize a consensus. While support for such activities can be as costly as support for laboratory or clinical research, the yield can be useful for the regulatory and scientific processes.

Epidemiological Evidence

How should a body of epidemiological evidence be summarized for hazard identification and dose–response purposes? How should biological background information be incorporated?

These questions concern more than statistical meta-analysis. There are some very interesting issues relating to the ability of the human mind to integrate large bodies of information and summarize them in consistent ways. Prior opinions often influence the interpretation and weight of evidence. This area may benefit from an analysis of expert behavior and artificial intelligence. The issue of positive and negative studies arises. It is inappropriate to pass each study through a test of statistical significance and contrast the number of positive and negative studies. What alternatives are there to the semantics of this terminology and the practices that arise from them? To the extent that the process becomes more explicit, hidden biases will be less important and the process will be less arbitrary and capricious for regulatory purposes.

The Price of Sample Information

When should epidemiological study be supported rather than laboratory or clinical study, and how many studies of varying kind and size are needed to determine that a hazard exists at current doses or that a hazard is unlikely at existing doses?

We cannot hope for a cut-and-dried procedure to make these determinations, but the elements that should go into such decisions need broader and more fundamental discussion and understanding. We need to bring together decision analysts, epidemiologists, statisticians, regulatory lawyers, and toxicologists to study case histories and to propose theoretically sound and practical approaches to deciding when to start and how much is enough. A published discussion of the issues, similar to the National Research Council's (NRC) risk assessment/risk management report of 1983 (*8*), may reduce the frequency of false positive and negative epidemiological studies, inappropriate waiting for unneeded studies, and the misuse of an inadequate number of negative or positive studies to make decisions.

Documenting and Remedying Negative Publication Bias

Despite thoughtful discussions of the allegation that negative studies are less likely to be published, the extent of the problem and reasons (if any) for it have not been fully documented. Once this is better understood, along with the likely uses of negative results, alternative approaches should be proposed and a market survey conducted to determine the acceptability of the alternatives. A journal of abstracts is one option, and a peer-reviewed electronic data base with abstracts keyed into electronically retrievable detailed documentation is another. The benefit of understanding and remedying this problem is that negative studies, and reviewers' concerns about them, would be available as societal decisions are being made. However, an obvious danger is that negative studies may be negative because they are poorly conceived or poorly executed.

The Implications of Involving the Public

What evidence claims that public concern or involvement can improve or bias study results or influence response rates? Once a study is done, what is the prevalence of knowledge about its results, and what is the duration of that knowledge? What interventions influence this? Here lies the borderland between epidemiology and evaluation research, similar to research in antismoking or contraception campaigns, except that segments of the public may view effective efforts to disseminate epidemiological research information as propaganda for environmental inaction. The benefits of a better understanding would be improved public health practice and public decision making based on evidence rather than misconception. A caveat must be given here. Environmentally cautious actions are often warranted even in the face of negative epidemiological results that are well disseminated and understood by the public. It is nonetheless helpful to be clear, for future reference, as to whether epidemiological evidence influenced the decision.

Why Epidemiology Must Be Skeptical

Theory can predict that a certain proportion of comparisons will be statistically significant or that a shared methodological flaw can produce a false association in multiple studies. But both the scientific community and the general public are more convinced by empirical demonstrations of these theoretical predictions. For example, in how many occupational studies is a variable like month of birth associated with disease? How many examples can we find in which several studies seemed to implicate an agent but subsequent studies failed to confirm the initial findings? Case studies like these can help journalists, the public, and decision makers to better understand the importance of a solid body of evidence for making good policy.

REFERENCES

1. IARC. Monographs on the evaluation of carcinogenic risks to humans, Supplement 7. Lyon: International Agency for Research on Cancer, 1987; 380–381.
2. Office of Technology Assessment. Biological effects of power frequency electric and magnetic fields. Stock Number 052-003-01152-2. U.S. Government Printing Office, 1989.
3. IARC. Monographs on the Evaluation of Carcinogenic Risks to Humans, Supplement 7. Lyon: International Agency for Research on Cancer, 1987; 203–204.
4. Begg CB, Berlin JA. Publication bias: a problem interpreting medical data. J Royal S Soc Series A 151:419–463 (1988).
5. Li JY, Liu PQ, Li GY, Rong S, Cao D, eds. Atlas of cancer mortality in the People's Republic of China, ed 1. Shanghai, China: China Map Press, 1979.
6. He X, Wei C, Liu Z, Chapman R. An epidemiological study of lung cancer in Xuan Wei County. Environ Health Perspect 94:9–13 (1991).
7. Keshan Disease Research Group. Epidemiological studies on the etiological relationship of selenium and Keshan disease. Chin Med J 92:477–482.
8. National Research Council. Risk assessment in the federal government: managing the process. Washington, DC: National Academy Press; 196–203.

Health Effects Of Electric and Magnetic Fields: Overview Of Research Recommendations

David A. Savitz

Department of Epidemiology, Campus Box #7400, School of Public Health, University of North Carolina, Chapel Hill, NC 27599-7400

We developed a series of articles concerning epidemiologic research on potential health effects of electric and magnetic fields. Our goal was to identify methodological issues that have arisen through past studies of cancer, reproduction, and neurobehavioral outcomes in order to suggest strategies to extend knowledge. Following an overview of relevant physics and engineering principles, cancer epidemiology of electric and magnetic fields is discussed separately with a focus on epidemiologic methods and cancer biology, respectively. Reproductive health studies, many of which focus on exposure from video display terminals are then summarized, followed by an evaluation of the limited literature on neurobehavioral outcomes, including suicide and depression. Methodological issues in exposure assessment are discussed, focusing on the challenges in residential exposure assessment and interpretation of wire configuration codes. An overview offers recommendations for priorities across these topic areas, emphasizing the importance of resolving the question of wire codes and childhood cancer. Collectively, these articles provide an array of observations and suggestions regarding the epidemiologic literature, recognizing the potential benefits to science and public policy. — Environ Health Perspect 101(Suppl 4):71–72 (1993).

Key Words: Epidemiology, electromagnetic fields

Introduction

The scientific literature on potential health effects of electric and magnetic fields has evolved haphazardly, like many research pursuits. The origins of the epidemiologic evidence can be traced to studies of neuropsychological symptoms in Soviet electrical workers in the 1960s (*1,2*), with an important study of power lines and childhood cancer published by Wertheimer and Leeper in 1979 (*3*). Through the 1980s, the pace and scope of epidemiologic research accelerated to the point that there are now perhaps a dozen major ongoing epidemiologic studies focused on cancer and a smaller number addressing reproduction and neuropsychological function.

In an attempt to conceptualize and organize better the evolving evidence, we have developed a series of articles. The intent was not to review comprehensively the past research or to draw specific conclusions for decision-making purposes, but rather to focus on the frontiers of existing knowledge and make recommendations for how to extend those frontiers. Some degree of subjectivity is required to abstract the important observations from the literature and make recommendations about which of the many possible approaches is most likely to advance our understanding. Individual authors undoubtedly have different priorities about what would constitute an exciting discovery, but we all share an interest in the fundamental questions of whether exposures to power-frequency electric and magnetic fields affect human health in clinically important ways.

This manuscript was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990 – September 1992.

Overview of Articles

To orient readers unfamiliar with the physics and engineering aspects of electric and magnetic fields, the articles begin with Kaune's "Introduction to Power-Frequency Electric and Magnetic Fields" (*4*). This defines the key concepts, identifies principal sources of electric and magnetic fields and the levels of exposure typically encountered, analyzes how these fields affect humans, and describes the technology available for environmental measurements. The shielding of electric but not magnetic fields by biological tissues is noted, with some discussion of the processes by which the weak fields of concern might induce biological effects.

Two chapters are devoted specifically to the study of cancer in relation to electric and magnetic field exposure. In "Epidemiologic Studies of Electric and Magnetic Fields and Cancer: Strategies for Extending Knowledge" (*5*), I focus on epidemiologic design and analysis issues that are in need of examination and improvement. The recommendation is made for two specific efforts concerning residential exposures and cancer: a comprehensive evaluation of sources and patterns of individual magnetic field exposures to identify exposure sources most worthy of study and to clarify the role of "wire codes" (based on power lines outside the residence) as an exposure source, and an examination of the sociology and geography of wire codes to evaluate confounding or selection bias or the possibility that wire codes influence cancer through mechanisms other than magnetic fields. Studies of occupational electric and magnetic field exposure would benefit from additional surveys of exposure patterns in diverse industrial settings and from additional empirical evidence on the patterns of cancer risk in relation to those exposures.

Stevens considers "Biologically Based Epidemiological Studies of Electric Power and Cancer" (*6*), in which he relates the indirect evidence from studies of DNA transcription and translation, calcium balance in cells, and pineal production of melatonin to modern concepts of cancer biology. The potential role of electric and magnetic fields in the carcinogenic process is examined in the context of a two-stage model for carcinogenesis, consisting of mutation of DNA and cell growth. Integration of laboratory evidence with this model of carcinogenesis leads to the following recommendations: *a*) given a number of points at which electric and magnetic fields might operate, exposures over a broad time period should be considered; *b*) effects on calcium balance encourage studies of acute nonlymphocytic leukemia; *c*) influences on pineal function suggest studies of hormone-dependent cancers (female breast, prostate) be conducted in conjunction with an evaluation of other influences on pineal function; and *d*) alteration of calcium homeostasis might lead to oxidative stress, which encourages study of

the interactive role of radiation and other agents that induce oxidative stress.

Shaw and Croen's article on "Human Adverse Reproductive Outcomes and Electromagnetic Field Exposure: Review of Epidemiologic Studies" (*7*) systematically reviews the epidemiologic evidence from exposures in residences, in workplaces, and to specific electrical devices such as electric blankets and video display terminals. Serious methodological deficiencies exist in many of the studies, especially those examining exposures other than video display terminals, so that research directions cannot be articulated with a high degree of specificity and confidence. Nonetheless, recommendations are made for: *a*) addressing the suggestion from laboratory studies of a possible adverse effect on growth by studying selected congenital anomalies, intrauterine growth retardation, and chromosomally normal spontaneous abortions; *b*) consideration of paternal residential exposure in relation to reproductive outcomes; and *c*) application to reproductive health outcomes of the exposure assessment methods for residential and occupational settings previously applied to cancer, ideally incorporating diverse sources of exposure.

"Neurobehavioral Effects of Power Frequency Electromagnetic Fields" by Paneth (*8*) summarizes the evidence for potential adverse effects on a number of indices of neurological and psychological parameters. The unique challenges of studying behavior and cognition include the problem of laboratory artifacts, the subtlety and transiency of many outcomes of interest, and the strong influence of social factors. The literature on neurobehavioral testing of experimentally exposed subjects, assessments of occupationally exposed workers, and the studies associating residential exposure with suicide are reviewed. Laboratory evidence suggesting effects on calcium efflux does not generate specific predictions, whereas the potential role of electric and magnetic fields in pineal function and circadian rhythms points directly toward depression as a plausible outcome. Thus, the recommendation is made that prospective studies of occupational exposure and depression be conducted, rather than pursuing additional studies of cognition in occupationally exposed groups, for which the results have been largely negative, or studying depression in relation to residential exposures, for which the social class influences would be difficult to remove.

Kaune summarizes the key issues regarding ascertainment of exposure in "Assessing Human Exposure to Power-Frequency Electric and Magnetic Fields" (*9*). Occupational exposures have been inferred largely from job titles. Residential exposure sources are reviewed, with a focus on the rationale for wire codes and spot measurements as indicators of long-term exposure. Recommendations are made for: *a*) development of job-exposure matrices for occupational exposure assessment based on direct measurements of workers in different occupational groups; *b*) evaluation of the ability of wiring codes and spot measurements to predict long-term historical exposure; *c*) an examination of exposures that are predicted by wire codes; *d*) an assessment of the contribution of residential and nonresidential exposures to total exposure; *e*) study of long-term temporal variation in residential exposure; and *f*) consideration of alternate exposure metrics associated more closely with wire code than is average magnetic field.

Siemiatycki offers his perspective on "Problems and Priorities in Epidemiologic Research on Human Health Effects Related to Wiring Code and Electric and Magnetic Fields" (*10*). He argues that the most pressing need is to verify the finding that wire codes are associated with childhood cancer because that possibility is the dominant basis for public concern. This could be achieved by reexamining data from past studies as well as by launching additional case–control studies that are responsive to concerns about control selection and incorporate measured fields and appliance exposures. Additional efforts are recommended for: *a*) reexamination of completed studies of wire codes and childhood cancer, *b*) new studies to examine the reported association between wire codes and childhood cancer, *c*) methodological research to evaluate the relation of wire codes to measured fields and indicators of historical exposure, *d*) occupational studies of cancer, *e*) documentation of exposure patterns in workers outside the electric utility industry, *f*) animal carcinogenicity studies, *g*) a broad survey of residential exposure and ecological studies of cancer; *h*) study of neurobehavioral effects, *i*) reproductive health studies focusing on residential wire codes primarily and other sources secondarily, and *j*) studies of adult cancer in relation to nonoccupational exposure, with items *a* to *f* of high priority and *g* to *j* of lower priority. Current impediments to the conduct of environmental epidemiology are noted, with the suggestion that large-scale monitoring systems are needed.

In all chapters, the authors were encouraged to express their own take on the literature and avoid the noncommittal tone of many previous committee recommendations. Neither the individual articles nor the summary represents a consensus but, rather, the product of individual work and critical responses to the ideas at several steps along the way. As a result, these chapters offer perspectives with which the reader may well disagree, but because the underlying assumptions that lead to the recommendations are provided, the debate itself should be a productive one. There was a consensus among Working Group members, however, about the basic premise that the research area is important (in part, because the public has decided that it is) and that well-designed and carefully conducted epidemiologic research will be beneficial to scientists and those concerned with the formulation of public policy on this issue.

REFERENCES

1. Asanova TP, Rakov AI. The state of health of persons working in the electric field of outdoor 400 and 500 kV switchyards. Gig Tr Prof Zabol 10:50–52 (1966).
2. Danilin VA, Voronin AK, Modorskii VS. The state of health of personnel working in high-voltage electric fields. Gig Tr Prof Zabol 13:51–52 (1969).
3. Wertheimer N, Leeper E. Electrical wiring configurations and childhood cancer. Am J Epidemiol 109:274–284 (1979).
4. Kaune WT. Introduction to power-frequency electric and magnetic fields. Environ Health Perspect 101(Suppl 4):73–81 (1993).
5. Savitz DA. Epidemiologic studies of electric and magnetic fields and cancer: strategies for extending knowledge. Environ Health Perspect 101(Suppl 4):83–91 (1993).
6. Stevens R. Biologically based epidemiological studies of electric power and cancer. Environ Health Perspect 101(Suppl 4):93–100 (1993).
7. Shaw GM, Croen L. Human adverse reproductive outcomes and electromagnetic field exposures: review of epidemiologic studies. Environ Health Perspect 101(Suppl 4):107–119 (1993).
8. Paneth N. Neurobehavorial effects of power-frequency electromagnetic fields. Environ Health Perspect 101(Suppl 4):101–106 (1993).
9. Kaune WT. Assessing human exposure to power-frequency electric and magnetic fields. Environ Health Perspect 101(Suppl 4):121–133 (1993).
10. Siemiatycki J. Problems and priorities in epidemiologic research on human health effects related to wiring code and electric and magnetic fields. Environ Health Perspect 101(Suppl.4):135–141 (1993).

Introduction To Power-Frequency Electric and Magnetic Fields

W.T. Kaune

EM Factors, 530 Lee Boulevard, Richland, WA 99352

This paper introduces the reader to electric and magnetic fields, particularly those fields produced by electric power systems and other sources using frequencies in the power-frequency range. Electric fields are produced by electric charges; a magnetic field also is produced if these charges are in motion. Electric fields exert forces on other charges; if in motion, these charges will experience magnetic forces. Power-frequency electric and magnetic fields induce electric currents in conducting bodies such as living organisms. The current density vector is used to describe the distribution of current within a body. The surface of the human body is an excellent shield for power-frequency electric fields, but power-frequency magnetic fields penetrate without significant attenuation; the electric fields induced inside the body by either exposure are comparable in magnitude. Electric fields induced inside a human by most environmental electric and magnetic fields appear to be small in magnitude compared to levels naturally occurring in living tissues. Detection of such fields thus would seem to require the existence of unknown biological mechanisms. Complete characterization of a power-frequency field requires measurement of the magnitudes and electrical phases of the fundamental and harmonic amplitudes of its three vector components. Most available instrumentation measures only a small subset, or some weighted average, of these quantities. Hand-held survey meters have been used widely to measure power-frequency electric and magnetic fields. Automated data-acquisition systems have come into use more recently to make electric- and magnetic-field recordings, covering periods of hours to days, in residences and other environments. Some of these systems are portable and can be worn by individuals for personal-exposure measurements. — Environ Health Perspect 101(Suppl 4):73–81 (1993).

Key Words: 60-Hz, health effects, epidemiology, tutorial

Introduction

Terms and concepts commonly used in the discussion of power-frequency electric and magnetic fields are introduced here. The interactions of these fields with matter, particularly living tissues, also are discussed. Finally, parameters that describe power-frequency fields are listed and instruments developed to measure one or more of these parameters are described.

Electric Fields

Definition

One of the fundamental properties of the particles that make up matter is their electric charges. Electrons and protons have negative and positive charge, respectively. Experiments have shown that the magnitudes of charges carried by these two particles are equal in magnitude. Furthermore, these particles seem to possess the smallest unit of electric charge that can be isolated: No smaller charge has ever been observed, and all larger charges apparently consist of integral multiples of the electronic charge. In the Standard International (SI) system of units, the electronic charge (i.e., charge of an electron) is -1.60×10^{-19} coulombs (C).

Electrically charged particles exert forces on each other. If two particles have charges of the opposite sign (e.g., a proton and an electron), the force between them is attractive. Otherwise, the force is repulsive. The electrical force between electrons and protons binds together the constituent particles of atoms and molecules. Electrical forces between charges are discussed using the concept of the electric field.

The electric field produced at a given point in space by a system of one or more electrically charged bodies is defined as the force that is exerted on a very small test body placed at this point and carrying a charge of exactly 1 C (Fig. 1). The electric field can be represented by an arrow that points in the direction of the electric force on the test body and whose length is in proportion to the strength of the electric force. These arrows are called vectors and will be denoted by letters printed in boldface; the magnitude of a vector (i.e., its length) will be denoted by the same letter in normal typeface. Thus, an electric field will be denoted by $\mathbf{E}$, while its magnitude will be written E.

The force, $\mathbf{F}$, in units of newtons (N), on any small particle placed in an electric field is given by the equation $\mathbf{F} = q\mathbf{E}$, where q is the particle's charge. Note that the force on a positive charge is in the same direction as the electric field, while the force on a negative charge is in the opposite direction. By definition, the fundamental units of the electric field are force divided by charge, that is, newtons per coulomb

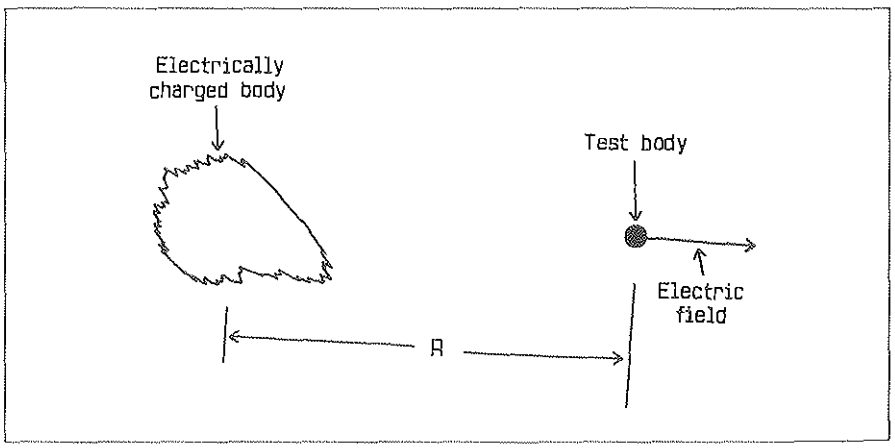


Figure 1. Electric field is force exerted on small test body carrying unit charge of 1 coulomb.

This manuscript was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990 – September 1992.

Table 1. Terms commonly used to describe frequency ranges below 300 kHz.

Term	Frequency range
Extremely low frequency	3 Hz – 3 kHz
Power frequency	50 Hz – 1000 Hz
Very low frequency	3 kHz – 30 kHz
Low frequency	30 kHz – 300 kHz

(N/C). (As will be explained shortly, the units of volts per meter are used more commonly for electric fields.)

A second concept that is intimately related to the electric field is the electric potential. The value of the potential at a single point has no significance, but the difference in potential between two points is related directly to the physical work (i.e., force acting through a distance) the electric field will do moving an electric charge between the two points. It is interesting that potential differences generally are easier to measure than electric fields, even though their definition is more abstract. It is customary to define the potential so that the earth (i.e., ground) is at zero.

Electric potential has been given its own unit, the volt (V). However, because potential difference is defined in terms of the work done by the electric field in moving a test body between two points, it also has the units of work per unit charge, that is, force per unit charge multiplied by distance. Because the units of an electric field are force per unit charge, we see that volts = electric field × meters. Thus, an alternative set of electric-field units is volts per meter (V/m).

Electrically conducting materials contain atoms and molecules with loosely bound electrons that can move from atom to atom under the influence of a force. Such movement, which will occur, for example, when an electric field is applied to the material, constitutes an electric current. The current passing through a specified cross-section of a body is defined as the total electric charge crossing this plane in one second. The fundamental unit of electric current is charge per unit time; in the SI system, this is given the name ampere, abbreviated A.

Often, the distribution of current within a body is of more interest than the total current through the body. This distribution is specified using the current-density vector, J, whose direction is that of current flow at a particular point and whose magnitude is equal to $\delta I/\delta A$, where δI is the current crossing a very small surface element of area δA oriented perpendicular to J. The units of current density are amperes per square meter, or A/m^2. J is directly proportional to E in a wide variety of materials. That is, $J = \delta E$ (Ohm's law), where the constant of proportionality, δ, is called the electrical conductivity of the medium. The units of δ are siemens per meter (S/m).

Living tissues are electrical conductors. Conductivities of living tissues, as measured by several groups, lie in the approximate range 0.01 to 1.5 S/m (*1–3*). By comparison, the conductivity of copper is about 60,000,000 S/m.

Electric fields whose magnitude and direction remain constant as time passes are called static. Mathematicians have shown that any quantity that changes over time may be represented as the sum of a (possibly infinite) number of sinusoidal functions of time, each characterized by a different frequency and magnitude. The frequency of a sinusoid is the number of complete cycles it goes through in one second. The SI unit of frequency is the hertz (Hz), where 1 Hz corresponds to exactly one complete cycle in one second. Frequency ranges are often categorized by terms such as extremely low frequency, very low frequency, etc. Table 1 defines the terms used to describe the frequency ranges of interest in this report (*4*).

Sinusoids approximately describe the time behavior of the voltages and currents produced by the electric generators used to energize electric power systems. These generators operate at the power frequencies of 50 or 60 Hz.

Electric-Field Sources

Experiments show that a vertical, almost static, electric field exists in the lower portion of the earth's atmosphere. The source of this field is electric charge carried from the ground to the upper atmosphere by thunderstorm activity. The mean strength of the ground-level atmospheric electric field is about 130 V/m (*5*). Ground-level field strengths in excess of 100 kV/m (i.e., 100,000 V/m) have been observed on flat, unobstructed surfaces during thunderstorms (*5,6*).

Electric fields with frequencies above about 30 Hz and extending above 100 GHz (i.e., 1×10^{11} Hz) predominantly have man-made sources. Considerable data have been published on electric fields produced by high-voltage transmission lines (*7–9*). Table 2 gives field intensities produced by typical power lines operating at several voltage levels (*10*). These data show that the largest electric fields produced at ground level by electric transmission lines now in service are about 10 kV/m. Electric fields under even higher voltage power lines that may be built in the future probably will not exceed this value significantly because of the need to limit shock hazards to personnel in the vicinity of the lines. However, as line voltages are increased, the widths of land on either side of transmission lines that are exposed to fields larger than, for example, 1 kV/m are increased. Ground-level electric fields found in substations or other electric power facilities usually do not exceed substantially the values listed in Table 2.

Work has been conducted in a number of countries to determine the actual exposures to electric fields of humans working and living in the vicinities of transmission lines and other electric-power facilities (*11–16*). This work has demonstrated that it is very difficult to estimate exposure using unperturbed field values (i.e., fields measured with no humans present) and simple estimates of a person's location as a function of time.

Another prominent source of electric fields in the extremely low frequency and very low frequency ranges is video display terminals (VDT). The display of information by a VDT is accomplished using a cathode ray tube (CRT). A beam of electrons, originating at the rear of a CRT, is directed onto the interior surface of its screen. The result is a spot of light at the point of impact whose intensity depends on the current in the electron beam. Magnetic fields are used to sweep the spot horizontally and vertically on the screen.

Static electric fields are produced by electrical charging of the screen of a VDT. Changes in the data being displayed by the VDT may result in modulation of these fields, usually at frequencies near 60 Hz. Finally, the circuitry used to generate the high voltages required by VDT produces electric fields at frequencies in the range of about 12 kHz to

Table 2. Typical electric and magnetic field strengths at ground level under electric-power transmission lines.

Voltage, kV	Current, A	Electric field, kV/m	Magnetic field, μT
115	200	1.5	5.0
230	300	2.5	6.1
345	400	3.4	6.8
500	550	6.7	8.4
765	750	10.0	10.0

Table 3. Electric and magnetic fields produced 30 cm in front of video display terminals (VDT). Fields produced by VDT differ greatly, so values listed below are not characteristic of all units.

Type of field	Frequency	Value
Electric[a]	0 Hz	≤1500 V/m
Electric	ELF[b]	≤70 V/m
Electric	VLF[c]	≤3 V/m
Magnetic	ELF[d]	≤0.7 μT
Magnetic	VLF[e]	≤0.25 μT

[a]One-half hour after turn on

[b]Extremely low frequency range, produced by changes in data being displayed by VDT

[c]Very low frequency range, produced by fly-back transformer and associated circuitry

[d]Extremely low frequency range, produced by vertical sweep circuitry

[e]Very low frequency range, produced by horizontal sweep circuitry

35 kHz. Table 3 summarizes electric field data from several publications (*17–19*).

There are, of course, many sources of electric fields in occupational and residential settings. Bowman et al. (*20*) have published the results of an electric (and magnetic) field survey of electrical occupations. They found that electric-field levels in most electrical occupations were similar to those in residential environments, except for occupations such as power line workers that involved work around very high voltages. General residential electric-field measurements have been published (*21*). One of the stronger electric-field sources in residences is electric blankets (*22*).

If necessary, shielding can be used to reduce the electric fields produced by common environmental sources. Throughout the frequency range of interest to this paper, a highly effective shield can be constructed by enclosing the source of interest in practically any conducting material.

Electric-Field Coupling to Living Organisms

Exposure of a living organism to an electric field is normally specified by the unperturbed field strength, that is, the field strength measured or calculated with the subject removed from the system. The use of this field to describe exposure is convenient, because it is relatively easy to measure or calculate. But, because of field perturbations, the unperturbed field is not equal to either the electric field that actually acts on the outer surface of the body or the electric field that is induced inside the body.

Electric fields with frequencies extending from 0 Hz to well above 300 kHz are altered strongly in the vicinity of almost any conducting body, including the bodies of humans or other living organisms. This perturbation occurs because the applied field, E, induces an electric charge density on the surface of the exposed body that generates a second electric field, E'. The total electric field is $E + E'$. Inside a conducting body, E and E' are nearly equal in magnitude but are directed oppositely. Consequently, their sum can be much smaller than either alone. In fact, for living tissues exposed to power-frequency electric fields, this cancellation is almost complete: The electric field induced inside the body is reduced relative to that outside it by at least a factor of 10,000 and, in most areas, by more than a factor of 1 million (*23*).

Outside the body, E and E' may add rather than cancel, so the applied field is enhanced. This enhancement tends to be greatest at the outer surface of the most sharply curved parts of the body. For example, the field at the top of the head of a person standing on the ground under a power line is enhanced by a factor of 15 to 20 (*24*).

Biophysical Analysis of Electric-Field Coupling

As discussed previously, the electric field acting on the surface of the body of a human or animal is enhanced over most of the body surface relative to the unperturbed electric field. Power-frequency electric fields can be perceived by humans (*25*) and by animals (*26–28*). One known mechanism of perception is hair stimulation (piloerection), that is, oscillatory hair movement by electric forces. The frequency of this vibration can be equal to or double the frequency of the applied electric field (*29,30*), depending on relative humidity and, possibly, other factors. Other modes of field perception have been investigated by Weigel et al. (*31*).

Another well-known mechanism of interaction between electric fields and biological tissues is the direct stimulation of excitable (e.g., neural) cells by the induction of voltages across their membranes sufficient to trigger their depolarizations. Such stimulation underlies the physiological responses of perception, shock, and electrocution that result from exposure to progressively larger electric currents. Most research on this mechanism covers electric-shock hazards (*32*). The basic dosimetric quantity is the current density in the affected part of the body. The threshold power-frequency current density required to stimulate most excitable cells is about 10 to 20 A/m^2. Very long nerve cells oriented parallel to the current-density vector may be sensitive to values as small as about 1 A/m^2 (*33–36*).

It is clear that the current densities directly induced in humans or other living organisms by externally applied power-frequency electric or magnetic fields with magnitudes similar to environmental levels are much smaller than levels required to excite neural tissues. For example, Kaune and Phillips (*37*) estimate that the current density induced in the ankle of a human standing on one foot directly under a higher voltage electric-power transmission line could be as high as 0.04 A/m^2, a value only about 4% of the level needed to excite very long nerve cells.

A person standing near a higher voltage transmission line may be exposed to a substantial body current (> 0.001 A) when touching a very large conducting object such as a truck or bus (*38,39*). To prevent electric shock hazards, the National Electrical Safety Code (*40*) requires that higher voltage electric-power transmission lines be designed so that their electric fields will not induce currents exceeding 0.005 A between the body of a grounded person and a bus or large truck.

Bernhardt (*35*) argued that extracellular electric fields induced by external fields could not be judged safe, *a priori*, unless they were substantially weaker than the fields generated by endogenous biological processes in living tissues. This author used electrocardiographic and electroencephalographic data to estimate endogenous fields in the brain and torso and arrived at a lower limit current density of about 0.001 A/m^2. Bernhardt's criterion, if valid, exempts most environmental human exposures from being of concern. However, current densities exceeding Bernhardt's limit are induced in the torso of a human standing under a higher voltage transmission line (*41*).

Several researchers (*42,43*) have pointed out recently that electric potentials arising across cell membranes from intrinsic thermal charge-density fluctuations are much larger than levels induced by most environmental electric and magnetic field sources. Some argue that this means that biological effects from such exposures are impossible. However, there is now substantial amount of literature indicating that

exposure of various living organisms to power-frequency electric or magnetic fields leads to various biological responses. Living tissues apparently possess some mechanism that enables them to detect signals below the ambient cellular noise. Cooperative interaction mechanisms involving the joint response of many cells are one possibility (*44–46*).

Magnetic Fields

Definition

Magnetic fields, like electric fields, are produced by electric charge but only electric charge in physical motion. Magnetic fields exert forces on other charges but, again, only charges in motion. Because the most common manifestation of electric charge in motion is an electric current, it is often said that magnetic fields are produced by electric currents and interact with other electric currents.

The magnitude, F, of the force acting on an electric charge moving perpendicular to the direction of a magnetic field is equal to the product of the magnitude, v, of the particle's velocity, the magnitude of its charge, $|q|$, and the strength, B, of the field's magnetic flux density (Fig. 2), that is, $F = |q|vB$. (More generally, $F = |q|vB\ sin\theta$, where θ is the angle between the directions of the velocity and the magnetic field.) Because the direction of this force is perpendicular to both the directions of the magnetic field and the particle's velocity, it can cause the particle neither to speed up nor slow down. That is, the magnetic field, by itself, can deliver no energy to a system with which it is interacting.

Time-varying magnetic fields generate electric fields through a process known as magnetic induction. The physical law that governs this phenomenon is Faraday's law. These electric fields can impart energy to a body with which they are interacting. The electric currents induced in the body of a human exposed to a time varying magnetic field are known as eddy currents.

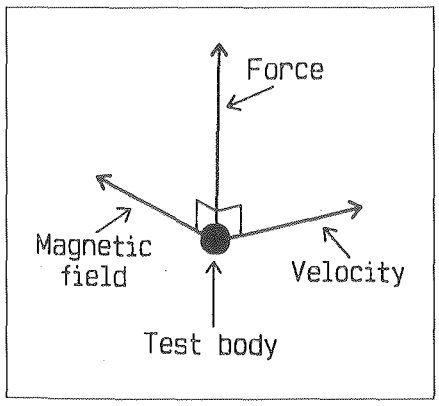


Figure 2. Force on small test body moving in a direction perpendicular to a magnetic field.

The complete specification of a magnetic field requires two vector quantities, the magnetic field intensity and magnetic flux density. Fortunately, these two are almost equivalent except in ferromagnetic materials such as iron. For purposes of describing human exposure, either may be used. In the SI system of units, the magnetic field intensity and flux density have units of amperes per meter (A/m) and tesla (T), respectively. The standard symbols for these two quantities are H (field intensity) and B (flux density). In vacuum, air, and to a lesser but still fully adequate approximation in nonmagnetic materials such as living tissues, $B/H = 4\pi \times 10^{-7}$.

At this time, most papers report the magnetic flux density in work related to low-frequency biological effects. A complication is that both the SI and CGS (i.e., centimeter-gram-second) systems of units have been and are still being used to report flux-density values. The CGS unit of flux density is the gauss (G), which equals exactly 0.0001 T. Magnetic flux densities found in typical residential environments have strengths of about 1 milligauss (1 mG = 0.001 G) or, equivalently, 0.1 microtesla (i.e., 0.1 μT).

Magnetic-Field Sources

The earth produces a static magnetic field known as the geomagnetic field. The strength of this field varies from about 30 μT (0.3 G) to 70 μT (0.7 G). Natural phenomena, such as thunderstorms and solar activity, produce time-varying magnetic fields with frequencies in the power-frequency range (*47*). Such fields are usually low strength, approximately 0.01 μT (0.1 mG). However, during intense magnetic storms (i.e., fluctuations in the earth's magnetic field resulting from solar activity), these fields can reach intensities of about 0.5 μT (5 mG) (*48*).

Of greater importance, in the context of possible biological effects, are the numerous static and alternating magnetic fields arising from man-made sources. In the lowest intensity range, generally less than 0.3 μT (3 mG), are alternating fields found in home and office environments (*49,50,21*). Electric blankets are one home source whose use can lead to sustained exposure to magnetic fields with somewhat elevated levels. Delpizzo (*51*) measured and calculated that a user would be exposed to a magnetic field produced by an Australian electric blanket of about 0.25 μT. Because U.S. blankets operate at lower voltage, and thus require more current, their fields are perhaps 2 to 3 times larger. Recently, electric blanket designs have been developed that produce only greatly reduced magnetic-field levels.

Higher flux densities can be produced by industrial processes using large magnets, induction motors, or heating devices. Particle accelerators use large magnets for several purposes (i.e., beam steering, momentum analysis of particles). Bowman et al. (*20*) recently published a survey of magnetic-field levels measured in the work areas of workers classified as electrical workers. The authors found these magnetic fields to be elevated significantly relative to those in typical residences. For example, arc welders were exposed to magnetic fields with a geometric mean of 4.1 μT (41 mG). Lovsund and co-workers (*52*) documented alternating magnetic fields from 8 to 70 mT (80 to 700 G) in the steel industry in Sweden.

Significant developments in specific areas of medical care have allowed the use of very strong static and pulsed magnetic fields for various diagnostic and treatment procedures (*53,54*). Static and pulsed flux densities from these new technologies range from about 0.5 to 2 T (5000–20,000 G) and 1 to 10 mT (10–100 G), respectively.

Electric power lines are a common source of power-frequency magnetic fields in developed societies (*55–57*). Table 2 lists typical magnetic field levels produced at ground level under various classes of transmission lines. The magnetic field produced by a single long, linear current-carrying conductor decreases in magnitude in proportion to the distance from the source. Partial cancellations between the magnetic fields produced by the multiple conductors of electric power lines result in the dependence between field strength, B, and distance, R, from the source being approximately $B \propto 1/R^2$. A method was developed recently for analyzing magnetic fields produced by such sources that enables power lines with decay characteristics of $1/R^3$ or $1/R^4$ to be designed easily (*58*). Such lines produce substantially reduced magnetic-field levels.

Another prominent source of magnetic fields with frequencies in the extremely low frequency and very low frequency ranges is VDTs. As noted earlier, the display of information by a VDT is accomplished using a CRT. A beam of electrons, originating at the rear of a CRT, is directed onto the interior surface of the display screen, producing a spot of light at the point of impact. The intensity of this spot is related to the current in the electron beam. Magnetic fields are used to sweep the spot quickly horizontally back and

forth across the screen. (The spot is moved from left to right at a constant rate and, at the end of this sweep, is returned to the left edge very quickly.) The number of complete horizontal traversals that occurs in one second is referred to as the horizontal sweep frequency. VDTs in use have horizontal sweep frequencies ranging from about 12 to 60 kHz. Users of VDTs and others in the vicinity may be exposed to the magnetic field used to control the horizontal movement of the unit's electron beam. The spatial orientation of this leakage magnetic field tends to be vertical.

As the electron beam is swept repeatedly and horizontally across the screen of a CRT, it also is swept slowly, vertically down the screen by a second magnetic field. The vertical sweep frequencies in VDTs now in use range from about 50 to 75 Hz. This frequency is the fundamental frequency of the leakage magnetic field from the vertical sweep circuitry. The spatial orientation of this field tends to be horizontal. Table 3 summarizes published values of magnetic fields produced 30 cm from the screens of an assortment of VDTs.

In contrast to the electric field case, it is quite difficult to construct magnetic-field shields of much effectiveness for frequencies in the power-frequency range. Ferromagnetic materials can be used to construct shields, but these shields tend to be physically large, very heavy, and of limited effectiveness. Substantial thicknesses of conducting materials also can provide magnetic-field shielding. A simple rule for such shields is that their thicknesses should be large with respect to the length $712\sqrt{f\sigma}$ in meters, where f and σ are frequency (Hz) and the shield conductivity (S/m), respectively.

Magnetic-Field Coupling to Living Organisms

In contrast to electric-field exposure, the bodies of humans, animals, and other living organisms cause almost no perturbation in a power-frequency magnetic field to which they are exposed. Faraday's law of induction states that time-varying magnetic fields generate electric fields through induction. Therefore, a living organism exposed to a magnetic field also will be exposed to an induced electric field that causes currents (called eddy currents) to flow in its body. These currents circulate in closed loops that tend to lie in planes perpendicular to the direction of the magnetic field.

Biophysical Analysis of Magnetic-Field Coupling

Alternating magnetic fields induce electric fields inside the bodies of exposed humans and animals. External alternating electric fields also induce electric fields inside bodies. The distributions of the fields induced by these two types of exposure are different, but at the level of the cell there would appear to be no fundamental difference. Thus, the biophysical analysis provided earlier in this paper for electric-field induction also can be applied to the electric fields induced by alternating magnetic fields.

How large must a magnetic field be to induce current densities sufficient to potentially stimulate excitable cells? Magnetic induction of currents can be modeled using a simple ellipsoidal approximation of a man. A typical man has a height of 1.7 m, a mass of 70 kg (*59*), and a body-width-to-body-thickness ratio of about two. An ellipsoid with semimajor axes of 0.85 cm, 0.20 cm, and 0.10 cm has the same body height, the same width-to-thickness ratio, and a body volume of about 7.1×10^4 cm^3. The maximum current density, J_{max}, induced in this model when exposed to a horizontal magnetic field, B, is given by the formula

$$J_{max} = 1.2 \times 10^{-5}\, \sigma B f, \qquad [1]$$

where the tissue conductivity, σ, has a value of 0.2 S/m and the frequency, f, is 60 Hz. As discussed earlier, a minimum current density of about 1 A/m^2 is required to excite long nerve cells. According to Equation 1, a whole-body magnetic field of about 0.07 T (700 G) would be required to achieve this level of induced current density. This flux density is much larger than magnetic flux densities produced by electric-power facilities. It is possible that exposures to fields of this size may occur in certain specialized industrial environments. However, these exposures normally would involve only small parts of the body and, thus, would not result in the induction of current densities as large as the values calculated with Equation 1 for whole-body exposures.

In addition to the induction, magnetic fields exert forces on charged particles that are in motion within a living organism. The most prevalent types of motion in matter are the motion of electrons in atoms, nucleons in nuclei, and the intrinsic spins of these particles. These motions lead to the existence of magnetic dipole moments that may be either permanent or induced by the applied magnetic field. A magnetic dipole in a magnetic field experiences a torque that attempts to align it parallel to the applied field. However, this alignment is resisted by random thermal motion and a statistical distribution of dipole directions is thereby established. The alignment of a dipole with a magnetic field can be calculated (*54,60*). At body temperature (37°C) and at a magnetic flux density of about 30 μT (0.3 G) that is characteristic of a heavily loaded transmission line, alignment is less than about 10^{-8} for electronic and nuclear magnetic moments that might occur in living tissues. Obviously, the effect on the magnetic dipoles that are part of the body of a subject exposed to such a magnetic field is very small. Of course, every single dipole is subject to this effect and, conceivably, some sort of process might exist that is sensitive to the average response of a large number of dipoles.

Charged particles also are carried by the bulk motion of various parts of the body. For example, charged ions are carried by blood flow. These ions are both positively and negatively charged and will experience magnetic forces in opposite directions, resulting in a separation of the two polarities of electric charge and, therefore, in the generation of electric potentials. These potentials can produce artifacts in the electrocardiograms of rats (*61*) exposed to static magnetic fields with flux densities above 0.3 T (3000 G).

One proposed mechanism of interaction between ac magnetic fields and living organisms is ion cyclotron resonance (*62*). There are several versions of this mechanism. The simplest proposes that there is an interaction between applied alternating and static magnetic fields that may cause a biological response if the frequency, f, of the ac field is close to the cyclotron resonance frequency, f_o, defined by the equation

$$f_o = \frac{1}{2\pi}\left(\frac{q}{m}\right) B_o, \qquad [2]$$

where q and m are the charge and mass of a biological ion of interest and B_o is the strength of the static magnetic field. In most situations, the static field is just the earth's magnetic field, which varies in flux density from about 30 to 70 μT over the surface of the earth. (Values outside this range can be found in the vicinity of ferromagnetic materials, such as those used in the construction of larger buildings.)

The name ion cyclotron resonance stems from the fact that a charged particle, such as an ion, traveling in a vacuum perpendicular to a static magnetic field will follow a circular path and will make f_o complete orbits in 1 sec. However, because ions in living tissues are not traveling in a vacuum but are, instead, moving through a highly viscous medium, attempts to apply this interaction picture have not been successful. However,

Table 4. Cyclotron resonance frequencies (Hz) for ions of biological interest and for static magnetic flux densities characteristic of various locations on the surface of earth.

Ion	Static magnetic flux density		
	30 μT	50 μT	70 μT
Li^+	66	111	155
Na^+	20	33	47
Mg^{2+}	38	63	88
Cl^-	13	22	30
K^+	12	20	27
Ca^{2+}	23	38	54

empirically, Equation 2 does describe certain behaviors of several biological systems (*62*).

Table 4 lists cyclotron resonance frequencies for a number of ions of biological interest and for static magnetic flux densities characteristic of those found on the surface of the earth. Note that the resonant frequencies of certain ions (Li^+, Mg^{2+}, Ca^{2+}) are near the power-line frequencies of 50 or 60 Hz at some locations on the earth.

Characterizing Power-Frequency Fields

This section describes methods for characterizing power-frequency electric and magnetic fields, and it describes several instruments that have been developed for this purpose.

Quantities Characterizing Power-Frequency Fields

Power frequency fields (either electric or magnetic) are vectors and therefore have lengths (magnitudes) and directions. Vectors can be decomposed into three orthogonal components that are usually labeled the x, y, and z components. Thus, three numbers (three measurements) are needed to characterize fully a field at any instant in time. However, fields of interest to this paper usually are not constant in time.

The next simplest case is when each component of a power-frequency field is a sinusoidal function of time: $B_k(t) = M_k \sin(2\pi ft + \phi_k)$, where the index k denotes the vector component (i.e., x, y, or z) under discussion, M_k is the peak magnitude of this component, f is the field's frequency, t is time, and ϕ is the phase angle of this component. Assuming the frequency is known, measurement of two parameters (M_k and ϕ) are needed to characterize each component of a power-frequency field.

Field magnitudes are almost always expressed in terms of root-mean-square (RMS) values rather than peak values. The relationship between these two values is $(M_{rms})_k = M_k/\sqrt{2}$. An RMS value can sometimes be interpreted as a measure of the time-averaged energy associated with the magnetic field.

Generally, the actual value of the phase angle of any component of a vector field is of no particular significance, but the relative phase angles between components are important. To see this, consider a field that has only nonzero x and y components. First, suppose that the x and y components are in phase (i.e., $\phi_x = \phi_y$). The two components' time behaviors will mimic each other: They will pass through zero at the same time, will reach their maximum values at the same time, and so on. This behavior is shown in the left side of Figure 3. It is not hard to see that, as time proceeds, the tip of the magnetic field vector will trace out a straight line that passes through the origin (Fig. 3). Accordingly, this state is referred to as linear polarization.

In the second case, assume the relative phase angle between the two components is 90°. Then, as shown in the right side of Figure 3, the time behaviors of the two components will be out of step. As one component reaches its maximum value, the other will pass through zero, and vice versa. Because the two components are never simultaneously zero, the total magnetic field always is different from zero. Figure 3 shows the elliptical path that the tip of the magnetic field vector follows as a function of time. This condition is given the name elliptical polarization. If the peak (or, equivalently, RMS) magnitudes of the x and y components are equal, the ellipse becomes a circle, and we have the limiting case of circular polarization.

The complete specification of a sinusoidal vector field thus requires the measurement of five quantities, three magnitudes, and two relative phase angles. Unfortunately, the situation often is more complicated because the power-frequency magnetic fields under study are distorted from pure sinusoids. Because this distortion tends to be the same, cycle after cycle, each component of a power-frequency field can be viewed as the sum of many different sinusoids, with the frequency of each successive term in this series being equal to the next larger integral multiple of the power frequency (50 or 60 Hz). That is,

$$B_k = M_{k1}\sin[2\pi(f)t + \phi_{k1}] + M_{k2}\sin[2\pi(2f)t + \phi_{k2}] + M_{k3}\sin[2\pi(3f)t + \phi_{k3}] + \ldots \quad [3]$$

The first term in this series, which involves just the power-frequency, is called the fundamental, and the terms involving integral multiples of this frequency are called harmonics. The second (third, fourth, ...) harmonic is the term whose frequency is 2 (3, 4, ...) times the power frequency.

Clearly, complete characterization of a power-frequency magnetic field is a formidable task requiring very sophisticated instrumentation. Instead, most instruments measure some average of the parameters that fully describe a power-frequency field.

The simplest type of meter responds only to the component of an electric or

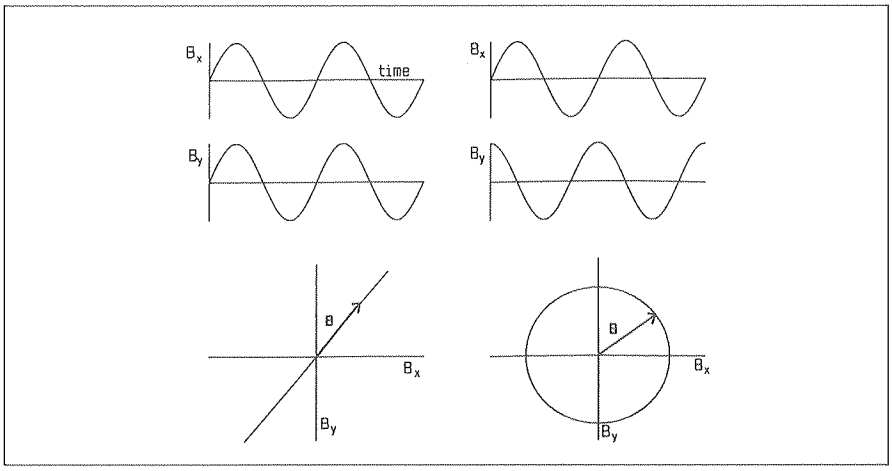


Figure 3. Left and right sides of figure show temporal and spatial patterns of components of linearly and circularly polarized vectors, respectively.

magnetic field that is parallel to the axis of the meter's probe. While some meters are designed to respond only to the 60-Hz fundamental, by using filters to reject all the terms in Equation 3 except the first, it is more common for a meter to accept a range of frequencies extending from about 30 to 40 Hz to a few hundred to a few thousand hertz. These latter meters have come to be referred to as broad-band instruments.

There are several ways to characterize the magnitude of a harmonically distorted signal. Some meters, known as average-responding meters, measure the average over a few cycles of the absolute value of the signal under study. Other meters measure the rms magnitude. Average responding meters are calibrated to display RMS values but do so without error only if there are no harmonics present.

There are several ways that a single-axis field meter, such as described in the preceding paragraph, can be used to characterize a three-dimensional electric or magnetic field. In one approach, the probe of the meter would be oriented in space to obtain the largest reading possible. This value is called the maximum field strength. Alternatively, the probe can be oriented in three perpendicular directions to measure the x, y, and z components of field strength. Then, a measure of the total field strength called the resultant field can be calculated using the formula

$$B = \sqrt{B_x^2 + B_y^2 + B_z^2} \qquad [4]$$

The maximum and resultant values are the same for linearly polarized fields. Otherwise, the resultant is always larger. In the extreme case of circular polarization, the resultant is 41% larger than the maximum value.

Survey Field Meters

A survey meter is a handheld instrument used to measure the electric or magnetic field at a particular point. The value so measured is displayed using either an analog or a digital display. Survey meters have no memory capability. Thus, any values that are to be retained must be written down by the user.

An interesting feature of some commercial meters is that their frequency responses can be set to be either flat or linear. With a flat response, all fields within the instrument's bandwidth are weighted equally. For example, a 0.1 μT (1 mG) magnetic field would be measured as 0.1 μT no matter what its frequency (as long as it was in the meter's bandwidth). With a linear response, this same field would be measured as 0.1 μT if its frequency were 60 Hz and 1.0 μT if its frequency were 600 Hz. By measuring with both bandwidths, qualitative information can be obtained about the harmonic distortion of the field under study.

Automated Data Acquisition Systems

Automated data acquisition systems are designed to be placed to acquire electric- or magnetic-field data for extended periods of time. The first system that was used in homes (*21*) was very large and consisted of several units that had to be wired together. Recent systems are much more sophisticated and can be used to characterize fully the magnitudes and phases of the fundamental and harmonics of all three vector components of the electric or magnetic field under study (*63*).

Personal Exposure Meters

Personal exposure meters can measure electric and/or magnetic fields while being worn by an individual. They therefore must be battery powered, small, and of low weight. The most powerful personal exposure meters available at this time essentially are battery-powered portable data acquisition systems. These meters incorporate onboard microcomputers and can be linked to other computers for the transfer of data.

Summary

An electrically charged particle exposed to an electric and a magnetic field will experience a force on it. If the particle is at rest, this force will be due to the electric field. Otherwise, both electric and magnetic forces will be present. All the electrically charged particles in living tissues will experience forces when exposed to electric and/or magnetic fields. Because living tissues are conductors, these forces will cause electric currents to flow.

Electric fields are specified by their field strengths in volts per meter (V/m). In the power-frequency region, magnetic fields are specified most often by their flux densities in units of tesla or gauss.

Static electric and magnetic fields are produced by the earth. In addition, very low levels of these fields are produced naturally with frequencies in the power-frequency range (i.e., 50–1000 Hz). Much stronger electric and magnetic fields are a byproduct of human use of electric power. The strengths of the electric and magnetic fields found in typical residences in developed countries lie in the approximate ranges of 0 to 10 V/m and 0 to 1 μT (0-10 mG), respectively. Electric and magnetic fields under high-voltage electric power transmission lines can reach levels as high as 10,000 V/m and 30 μT, respectively. Even stronger magnetic fields can be found in certain occupational environments.

Because the body of a person is a good conductor at power frequencies, its interior is shielded strongly from electric fields. The electric field induced inside most parts of the body by an external electric field is reduced by at least a factor of 1 million. This field is too small to excite nerve cells and is, apparently, considerably smaller than fields that naturally occur in tissues. Biological effects caused by such fields must be due to mechanisms of interaction not yet understood.

In contrast to electric-field exposure, power-frequency magnetic fields penetrate living tissues without significant perturbation and induce circulating electric fields and currents in the body of an exposed human. The sizes of these induced electric fields are similar to those induced by electric-field exposure. Moving particles within the body also will interact directly with the applied magnetic field, but the strength of this interaction is small relative to thermal interactions. ehp

REFERENCES

1. Schwan HP, Kay CF. Specific resistance of body tissue. Circ Res 4:664–670 (1956).
2. Geddes LA, Baker LE. The specific resistance of biological material—a compendium of data for the biomedical engineer and physiologist. Med Biol Eng Comput 5:271–293 (1967).
3. Schwan HP. Electrical properties of tissue and cell suspensions. Adv Biol Med Phys 5:147–209 (1977).
4. IEEE. IEEE standard dictionary of electrical and electronics terms. New York: The Institute of Electrical and Electronics Engineers, 1988.
5. Israel H. Atmospheric electricity. Springfield, VA: U.S. Department of Commerce, National Technical Information Service, 1973.
6. Toland RB, Vonnegut B. Measurement of maximum electric field intensities over water during thunderstorms. J Geophys Res 82:438–440 (1977).
7. EPRI. Transmission line reference book-345 kV and above. Palo Alto, CA: Electric Power Research Institute, 1975.
8. Bracken TD. Field measurements and calculations of electrostatic effects of overhead transmission lines. IEEE Power Appar Syst PAS 95:494–504 (1976).

9. Bridges JE, Preache M. Biological influences of power frequency electric fields—a tutorial review from a physical and experimental viewpoint. Proc IEEE 69:1092–1120 (1981).
10. Hauf R. Electric and magnetic fields at power frequencies with particular reference to 50 and 60 Hz. In: Nonionizing radiation protection, vol VIII (Seuss MJ, ed), ISBN 92-890-1101-7, Eur Ser 10. Copenhagen: World Health Organization, 1982: 175–198.
11. Lovstrand KG., Lundquist S, Bergstrom S, Birke E. Exposure of personnel to electric fields in Swedish extra high-voltage substations: field strength and dose measurements. In: Biological effects of extremely low-frequency electromagnetic fields (Phillips RD, Gillis MF, Kaune WT, Mahlum DD, eds), proceedings of the 18th Hanford Life Sciences Symposium, Richland, WA. CONF-781016. Springfield, VA: National Technical Information Service, 1979.
12. Looms JST. Power frequency electric fields: dosimetry. In: Biological effects and dosimetry of nonionizing radiation: radiofrequency and microwave energies. NATO Advanced Study Institute Ser. A, Life Sciences. New York: Plenum Press, 1983.
13. Deno DW, Silva M. Method for evaluating human exposure to 60-Hz electric fields. IEEE Trans Power Appar Syst PAS 103:1699–1706 (1984).
14. Bracken TD. Comparison of electric field exposure monitoring instrumentation. Final report, research project 799-19. Palo Alto, CA: Electric Power Research Institute, 1985.
15. Chartier VL, Bracken TD, Capon AS. BPA study of occupational exposure to 60-Hz electric fields. IEEE Trans Power Appar Syst PAS 104:733–744 (1985).
16. Silva M, Zaffanella LE, Hummon N. An activity systems model for estimating human exposure to 60-Hz electric fields. IEEE Trans Power Appar Syst 104:1923–1929 (1985).
17. Harvey SM, Starecky L. Survey of electric and magnetic fields at Ontario hydro VDU work stations. Report No. 84-74-K. Toronto, Ontario: Ontario Hydro Research Division, 1984.
18. Harvey SM. Characteristics of low frequency electrostatic and electromagnetic fields produced by video display terminals. Report No. 82-538-K. Toronto: Ontario Hydro Research Division, 1982.
19. Guy AW. Measurement and analysis of electromagnetic field emissions from 24 video display terminals in American Telephone and Telegraph Offices, Washington, DC. Report to The National Institute for Occupational Safety and Health, 1987.
20. Bowman JD, Garabrant DH, Sobel E, Peters JM. Exposures to extremely low frequency (ELF) electromagnetic fields in occupations with elevated leukemia rates. Appl Ind Hyg 3:189–194 (1988).
21. Kaune WT, Stevens RG, Callahan NJ, Severson RK, Thomas DB. Residential magnetic and electric fields. Bioelectromagnetics 8:315–335 (1987).
22. Florig HK, Hoburg JF, Morgan MG. Electric field exposure from electric blankets. IEEE Trans Power Del. PWRD-2:527–536 (1987).
23. Kaune WT, Gillis MF. General properties of the interaction between animals and ELF electric fields. Bioelectromagnetics 2:1–11 (1981).
24. Deno DW. Currents induced in the human body by high-voltage transmission-line electric field—measurement and calculation of distribution and dose. IEEE Trans Power Appar Syst. PAS-96:1517–1527 (1977).
25. Reilly JP. Electric and magnetic field coupling from high-voltage AC power transmission-lines—classification of short-term effects on people. IEEE Trans Power Appar Syst. PAS 97:2243–2252 (1978).
26. Kaune WT, Phillips RD, Hjeresen DL, Richardson RL, Beamer JL. A method for the exposure of miniature swine to vertical 60-Hz electric fields. IEEE Trans Biomed Eng BME 25:276–283 (1978).
27. Stern S, Laties VG. 60-Hz electric fields: detection by female rats. Bioelectromagnetics 6:99–103 (1985).
28. Sagan PM, Stell ME, Bryan GK, Adey WR. Detection of 60-Hz vertical electric fields by rats. Bioelectromagnetics 8:303–313 (1987).
29. Gillis MF, Kaune WT. Hair vibration in ELF electric fields. In: Abstracts, symposium on the biological effects of electromagnetic waves, 19th General Assembly of the International Union of Radio Sciences, August 1–8, 1978, Helsinki, Finland. Paris: URSI, 1978; 51.
30. Cabanes J, Gary C. La perception directe du champ électrique. CIGRE Report No. 233-08, International Conference on Large High Tension Electric Systems, Paris. Paris: CIGRE, 1981.
31. Weigel RJ, Jaffe RA, Lundstrom DL, Forsythe WC, Anderson LE. Stimulation of cutaneous mechanoreceptors by 60-Hz electric fields. Bioelectromagnetics 8:337–350 (1987).
32. Dalziel CF. Electric shock hazard. IEEE Spectrum: 41–50 (1972).
33. Schwan HP. Field interaction with biological matter. Ann NY Acad Sci 103:198–213 (1977).
34. Bernhardt JH. The direct influence of electromagnetic fields on nerve and muscle cells of man within the frequency range of 1 Hz to 30 MHz. Radiat Environ Biophys 16:309–323 (1979).
35. Bernhardt JH. On the rating of human exposition to electric and magnetic fields with frequencies below 100 kHz. Ispra, Italy: Commission of the European Communities, Joint Research Center, 1983.
36. Bernhardt JH. Assessment of experimentally observed bioeffects in view of their clinical relevance and the exposures at work places. In: Proceedings of the symposium on biological effects of static and ELF-magnetic fields (Bernhardt JH, ed). Munich: BGA-Schriftenreihe, MMV Medizin Verlag, 1985.
37. Kaune WT, Phillips RD. Comparison of the coupling of grounded humans, swine and rats to vertical, 60-Hz electric fields. Bioelectromagnetics 1:117–129 (1980).
38. Gandhi OP, Chatterjee I. Radio-frequency hazards in the VLF to MF band. Proc IEEE 70:1462–1464 (1982).
39. Guy AW, Chou CK. Hazard analysis: very low frequency through medium frequency range. Final report, USAF SAM contract F33615-78-D-0617, task 0065. Brooks Air Force Base, TX: Aerospace Medical Division, 1982.
40. IEEE. National electrical safety code. Piscataway, NJ: Institute of Electrical and Electronics Engineers, 1990.
41. Kaune WT, Forsythe WC. Current densities measured in human models exposed to 60-Hz electric fields. Bioelectromagnetics 6:13–32 (1985).
42. Weaver JC, Astumian RD. The response of living cells to very weak electric fields: the thermal noise limit. Science 247:459–462 (1990).
43. Adair RK. Constraints on biological effects of weak, extremely low-frequency electromagnetic fields. Physiol Rev A 43:1039–1048 (1991).
44. Postow E, Swicord ML. Modulated field and "window" effects. In: Handbook of biological effects of electromagnetic radiation (Polk C, Postow E, eds). Boca Raton, FL: CRC Press, 1985.
45. Swicord ML. Possible biophysical mechanisms of electromagnetic interactions with biological systems. In: Report from the committee on biological and human health effects of extremely low frequency electromagnetic fields. Arlington, VA: American Institute of Biological Sciences, 1985.
46. Swicord ML, Postow E. Hypothesis of weak electric and magnetic interactions with biological systems. IEEE Special Publications 86TH0139-6-PWR. Piscataway, NJ: IEEE Service Center, 1986.
47. Grandolfo M, Vecchia P. Natural and man-made environmental exposures to static and ELF electromagnetic fields. In: Biological effects and dosimetry of nonionizing radiation; static and ELF electromagnetic fields (Grandolfo M, Michaelson SM, Rindi A, eds). New York: Plenum Press, 1985.
48. Tenforde TS. Biological effects of ELF magnetic fields. In: Biological and human health effects of extremely low frequency electromagnetic fields. Arlington, VA: American Institute of Biological Sciences, 1985; 79–128.
49. Caola RJ Jr, Deno DW, Dymek VSW. Measurements of electric and magnetic fields in and around homes near a 500-kV transmission line. IEEE Trans Power Appar Syst PAS 102:3338–3347 (1983).
50. Male JC, Norris WT, Watts MW. Human exposure to power-frequency electric and magnetic fields. In: Interaction of biological systems with static and ELF electric and magnetic fields (Anderson LE, Weigel RJ, Kelman BJ, eds), Proceedings of the 23rd Hanford Life Sciences Symposium, Richland, WA. Springfield, VA: National Technical Information Service, 1987.
51. Delpizzo V. A model to assess personal exposure to ELF magnetic fields from common household sources. Bioelectromagnetics 11:139–147 (1990).
52. Lovsund P, Obert PA, Nilsson SEG. ELF magnetic fields in electrosteel and welding industries. Radio Sci 17:35S–38S (1982).
53. Margulis AR, Higgins CB, Kaufman L, Crooks LE, eds. Clinical magnetic resonance imaging. San Francisco: University of California Press, 1983.
54. Budinger TF, Lauterbur PC. Nuclear magnetic resonance technology for medical studies. Science 226:288–298 (1984).
55. Haubrich HJ. Das Magnetfeld im Nahbereich von Drehstromfreileitungen. Elektrizitaetswirtschaft 73:511–517 (1974).

56. Scott-Walton B, Clark KM, Holt BR, Jones DC, Kaplan SD, Krebs JS, Polson P, Shepher RA, Young JR. Potential environmental effects of 765-kV transmission lines: views before the New York State public service commission, cases 26529 and 26559, 1976–1978. DOE/EV-0065. Springfield, VA: National Technical Information Service, 1979.
57. Kaune WT. Coupling of living organisms to ELF electric and magnetic fields. In: Report from the committee on biological and human health effects of extremely low frequency electromagnetic fields. Arlington, VA: American Institute of Biological Sciences, 1985;25–60.
58. Kaune WT, Zaffanella LE. Analysis of magnetic fields produced far from electric power lines. IEEE Trans Power Delivery 7:2082–2091 (1992).
59. ICRP. Report of the task group on reference man, ICRP 23. Oxford: Pergamon Press, 1975.
60. Reitz JR, Milford FJ. Foundations of electromagnetic theory. PA: Addison-Wesley, 1960.
61. Gaffey CT, Tenforde TS. Alterations in the rat electrocardiogram induced by stationary magnetic fields. Bioelectromagnetics 2:357–370 (1981).
62. Liboff AR, McLeod BR, Smith SD. Ion cyclotron resonance effects of ELF fields. In: Extremely low frequency electromagnetic fields: the question of cancer. (Wilson BW, Stevens RG, Anderson LE, eds). Columbus, OH: Battelle Press, 1990.
63. Dietrich FM. EPRI magnetic field characterization waveform capture system. Report to Electric Power Research Institute by Electric Research & Management, Pittsburgh, PA, 1989.

Epidemiologic Studies of Electric and Magnetic Fields and Cancer: Strategies for Extending Knowledge

David A. Savitz

Department of Epidemiology, Campus Box 7400, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7400

Epidemiologic research concerning electric and magnetic fields in relation to cancer has focused on the potential etiologic roles of residential exposure on childhood cancer and occupational exposure on adult leukemia and brain cancer. Future residential studies must concentrate on exposure assessment that is enhanced by developing models of historical exposure, assessment of the relation between magnetic fields and wire codes, and consideration of alternate exposure indices. Study design issues deserving attention include possible biases in random digit dialing control selection, consideration of the temporal course of exposure and disease, and acquisition of the necessary information to assess the potential value of ecologic studies. Highest priorities are comprehensive evaluation of exposure patterns and sources and examination of the sociology and geography of residential wire codes. Future occupational studies should also concentrate on improved exposure assessment with increased attention to nonutility worker populations and development of historical exposure indicators that are superior to job titles alone. Potential carcinogens in the workplace that could act as confounders need to be more carefully examined. The temporal relation between exposure and disease and possible effect modification by other workplace agents should be incorporated into future studies. The most pressing need is for measurement of exposure patterns in a variety of worker populations and performance of traditional epidemiologic evaluations of cancer occurrence. The principal source of bias toward the null is nondifferential misclassification of exposure with improvements expected to enhance any true etiologic association that is present. Biases away from the null might include biased control selection in residential studies and chemical carcinogens acting as confounders in occupational studies. — Environ Health Perspect 101(Suppl 4):83–91 (1993).

Key Words: Brain cancer, leukemia, neoplasms, nonionizing radiation

Introduction

A number of reviews of the epidemiologic literature on electric and magnetic fields and cancer have been developed over the last several years. In contrast to substantive reviews that seek to summarize evidence and draw conclusions (*1,2*) or those that explore methodologic issues to assist in drawing conclusions about the evidence (*3,4*), this paper has the limited goals of defining current knowledge for the purpose of identifying gaps that future epidemiologic studies can fill.

Residential Exposure to Magnetic Fields and Cancer

Synopsis of Evidence

Wertheimer and Leeper (*5*) were the first to consider a possible relation between residential exposures to magnetic fields and cancer. They found that children who had died of cancer lived in homes imputed to have elevated magnetic fields based on wiring configuration codes more frequently than controls. Power lines in the vicinity of the home were examined to estimate current flow and distance to the homes as a marker of long-term average magnetic field levels in the home.

The approach to classifying wiring was presented in greater detail in a study of adult cancer (*6*) and has been used, with little modification, in several subsequent studies. Observable characteristics of the power lines serve as the basis for estimating the typical current flow along the lines in order to assign a wiring class based on such factors as the number of phases, the thickness of the wires, and the number of service drops between transformers. Categorizing the homes into levels they labeled as very high current configuration, ordinary high current configuration, ordinary low current configuration, and very low current configuration combines the wiring class with an estimate of the distance from the wires to the home. Homes in neighborhoods served by buried wires have been considered a separate, low-exposure, group. Diagrams and a more detailed description are provided in this volume by Kaune (*7*).

Subsequent studies of childhood cancer have provided mixed support. A case–control study of leukemia in Rhode Island (*8*) was reported as negative based on an exposure classification system taken from that developed by Wertheimer and Leeper (*5*) for Denver. In addition to concerns with the applicability of the Denver system to Rhode Island and reliance on analyses of residences rather than persons, the different occupancy dates for cases and controls appear to have biased their measures of association toward the null, i.e., toward the absence of any association (*9*).

Tomenius (*10*) conducted a study in Stockholm in which homes were classified based on proximity to electrical constructions and magnetic field measurements at their front doors. Electrical constructions (specifically, above-ground power lines) were more common near case than control homes, and measured fields above 3 mG were more common among cases not near electrical constructions than comparable controls. Average magnetic fields were virtually identical for case and control homes. The positive association based on measured fields was restricted to nervous system cancers with an inverse association found for leukemia.

A second study in Denver (*11*) supported the hypothesis that children living in homes with higher wiring configurations or higher measured in-home magnetic

This manuscript was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990 – September 1992.

The author would like to acknowledge the constructive reviews and discussion with the other members of the Working Group, including William Kaune, Nigel Paneth, Gary Shaw, Richard Stevens, and especially Jack Siemiatycki and Charles Poole.

Table 1. Comparison of results for measured magnetic fields and wiring configuration codes: residential characteristics and childhood cancer.

	Savitz et al. (*11*)			London et al. (*13*)		
Measure	Exposure level	OR	95% CI	Exposure level	OR	95% CI
Spot measurement of low power magnetic field (mG)	0–<0.65	1.0		0–0.32	1.0	
	0.65–<1.0	1.3	0.7–2.4	0.32–0.67	1.0	0.6–1.7
	1.0–<2.5	1.3	0.7–2.3	0.68–1.24	1.4	0.7–2.9
	2.5+	1.5	0.6–3.6	1.25+	1.2	0.5–2.8
24-hr measurement of magnetic field (mG)				<0.68	1.0	
				0.68–1.18	0.7	0.4–1.2
				1.19–2.67	0.9	0.5–1.7
				2.68+	1.5	0.7–3.3
Wire configuration code	Underground	1.0		Underground		
	Very low	1.6	0.8–3.1	+ Very low	1.0	
	Ordinary low	1.0	0.7–1.5	Ordinary low	0.9	0.5–1.7
	Ordinary high	1.5	0.9–2.4	Ordinary high	1.4	0.8–2.6
	Very high	2.2	0.9–5.2	Very high	2.2	1.1–4.3

fields under low power use were at increased risk of developing cancer, although the magnitude of association (odds ratios of 1.5–2.0) was lower than had been reported by Wertheimer and Leeper (odds ratios of 2.0–3.0) (Table 1). No association was found for electric fields or magnetic fields measured under high power use conditions. These results were not due to confounding by prenatal and childhood exposures reported by parents, but nonresponse and differential mobility of controls constitute important limitations in this study.

Myers et al. (*12*) recently provided results from a study conducted in the early 1980s in England. They interpreted their results as providing little evidence to support an association between childhood leukemia and residential electromagnetic field exposure, but their primary control group consisted of children with solid tissue tumors and diseases also potentially affected by this exposure. Elevated exposures were extremely rare, but there was modest evidence of increased risk with increased exposure.

The study completed most recently is methodologically the strongest and has a major bearing on the direction of future research. London et al. (*13*) recently reported the results of a case–control study of childhood leukemia and residential magnetic field exposure in Los Angeles County. The major improvements over Savitz et al. (*11*) consisted of the use of controls selected concurrently with case identification and a much more complete and extensive array of in-home magnetic field measurements. The results showed a clear association of wire codes with leukemia with more limited evidence of an association based on both spot measurements of magnetic fields and 24-hr measurements of magnetic fields (Table 1). In spite of their presumably greater accuracy as a reflection of long-term historical exposures, measurements taken over a 24-hr period failed to demonstrate a notably stronger association with disease than did spot measurements.

In addition to these studies of ambient background magnetic fields in homes, one study of childhood cancer reported on use of electrical appliances by the mother during pregnancy and by the child (*14*). Electric blanket use by the mother during pregnancy and by the child was associated with a modestly increased risk of developing childhood cancer, whereas heated water beds, bedside electric clocks, and other appliances used by the mother or child were not associated with increased risk.

There have been several studies of residential exposures and adult cancers. Wertheimer and Leeper (*6*) found modest positive associations between wiring codes and several types of cancer. Subsequent studies of residential exposures have been limited in quality of exposure assessment (*15*) or size (*16*) and generally are not supportive of such a link. Preston-Martin et al. (*17*) evaluated electric blanket use in relation to adult myelogenous leukemia and found no association.

Exposure Assessment Needs

If there is a causal relation between some aspect of electric or magnetic field exposure and cancer, there is no reason to believe that the exposure indicators used in past studies have captured it with precision. Modest associations in past studies may be masking a much more substantial effect that has been diluted by nondifferential misclassification, since our exposure indicators are only imperfect proxies for the potent exposure. This misclassification may operate on several levels, including the incorrect exposure metric (e.g., averages rather than peaks), failure to measure exposure comprehensively (e.g., ignoring sources other than the home), measuring exposure at the wrong period in the subject's life, as well as the familiar inability to measure precisely even in the desired places and times. Several strategies that would produce stronger measures of associations if such effects are actually present are available to improve exposure classification.

Models of Historical Exposures. All studies have relied on historical exposure reconstruction, including both the case–control studies and the cohort study (*15*). Instrumentation exists for relatively convenient acquisition of real-time individual exposure profiles over periods of several days (*7*). In spite of suggestions to collect individual-level data for prospective investigations (*18*), the rarity of the cancers that have been studied to date dictates that future studies are likely to continue to assess exposures retrospectively. Thus, the challenge faced by investigators is how to reconstruct an exposure history in the absence of direct measurements during the historical periods of interest.

One approach would be to develop predictive models of exposure in which the model inputs are amenable to historical ascertainment. The sources of electric and magnetic fields that we encounter in our daily lives are diverse and probably too complex to assess based solely on physics and

engineering principles. It might be more fruitful to develop statistical equations that relate patterns of location and activity to measured fields. For example, detailed diaries could be maintained in parallel with real-time measurements of field strengths for a sizable and diverse population of children or adults outside the context of a specific case–control study. Statistical models to estimate various field parameters of interest then could be developed based on locations and activities. The predictors in those models would have to be amenable to historical assessment in order to be useful, recognizing that in reality some would be (e.g., use of electric blankets) and some would not (e.g., how close the child sat to the television).

Such an approach would be enhanced by methods for reconstructing historical exposures within the most important (i.e., frequently occupied) environments. Wire configuration codes were intended as historically stable markers of in-home exposures, because power lines are rarely modified (*5*). Although the wire codes have been demonstrated to be associated with current magnetic fields in homes, the strength of prediction is quite limited. The structure of the original Wertheimer and Leeper (*5,6*) wire code was developed intuitively. Further examination of whether there are better ways to integrate information on observable characteristics of the wiring using physics and engineering principles should be undertaken, and empirical estimation [such as that developed by Kaune et al. (*19*)] offers promise for making more accurate inferences from the historically stable electrical constructions.

Exposure sources other than the ambient levels found in homes also should be considered. There is no information on whether some variant of a wiring configuration code could be developed for schools or commercial buildings where substantial periods of time are spent. Appliance use certainly has the potential to be incorporated into a comprehensive historical exposure assessment.

Wire Codes versus Measurements. The report by London et al. (*13*) of a clear association of childhood leukemia with wire codes and a weaker association with measured magnetic fields, even taken over a 24-hr period, highlights the need to better understand what aspects of past and present exposure are reflected by each (*7*). The logistical considerations are as follows: *a*) better response for wire codes (passive on the part of the respondent) than for in-home measurements (requiring the respondent's cooperation), which would increase study size and diminish the potential for bias; *b*) greater expense for in-home measurements due to scheduling inefficiencies and equipment; and *c*) less need for engineering expertise to use a meter than to develop a wire coding methodology.

The key unanswered question remains the validity of the different strategies as indicators of historical exposure. Prospective studies are needed and would be simple to conduct. A panel of homes, without concern for the health status of their occupants, could be recruited for long-term evaluation with a battery of measurements and wiring characteristics. Periodic repeat measurements would monitor changes over time. This should be coupled with individual monitoring of occupants of selected homes so that the residential information could be understood in the context of other exposure sources. The effect of differing patterns of room use over time, modification in the use of rooms and their physical arrangement, and changes due to shifts in the occupants could be assessed empirically for their bearing on historical exposure reconstruction. If predictable relations were found, they could be queried and incorporated into exposure assessment protocols. On the other hand, even if such sources of inaccuracy could not be remedied, at least they could be quantified and considered in interpretation of results. The necessary database would be rather tedious to initiate and maintain, but it would be of great value with the passage of time.

Alternative Exposure Parameters and Sources. For a number of practical and theoretical reasons, interest has focused on long-term average magnetic fields in homes. Because magnetic fields are largely unperturbed by trees, building materials, etc., the power lines have a systematic relation to in-home fields. Electric fields, equally ubiquitous, have no marker analogous to wire codes that would allow their effects to be examined. Empirical approaches to examining exposure sources might yield some historically applicable indicators of electric field exposure, although this is unlikely based on a knowledge of exposure sources and past epidemiologic study results (*11,19*).

The specific attributes of the magnetic fields predicted by wiring codes also remain mysterious (*7*). The relation of wire codes to peaks, transients, various percentiles, time above or below postulated thresholds, etc. could be examined to assist in interpretation. That ultimately may suggest modifications to wire codes to serve as surrogates for different types of indices. Guidance for laboratory investigators could be much more specific if the fields predicted by wire codes were better understood.

Finally, additional examination of appliance-based exposures is warranted. The effectiveness of asking about the use of such devices seems straightforward, but the validity of self-report on use of key appliances, including patterns of use (e.g., proximity to television set), warrants examination. The yield, in terms of refinements in exposure classification, from more sophisticated inquiries should be evaluated: Is it worthwhile to ask the brand of video display terminal or the setting on which electric blankets are used? It seems likely that appliances that do not contribute to average magnetic field still may contribute substantially to other indices, such as peaks or time above a given level (e.g., hair dryers). Such questions could be examined as part of an effort to develop statistical models of exposure starting with an effort to completely reconstruct sources of electric and magnetic fields.

Health End Points

Studies of both childhood and adult cancers have followed traditional approaches to disease classification. For childhood cancer, in particular, rarity of the disease has led to broader groupings than might be desirable. These groupings may dilute any effects of electric and magnetic field exposures on cancer subtypes. In some instances, all childhood cancers have been grouped together, although most investigators have also examined subtypes such as acute lymphocytic leukemia, lymphomas, brain tumors, etc. More attention should be paid to examining more refined disease subtypes. For example, among leukemias there is some suggestion that different cytogenetic subtypes have different etiologies (*20*). This implication encourages the evaluation of the role of electric and magnetic fields for those subtypes. Histologic categories of brain cancer recently have been shown to have markedly different associations with electrical occupations (*21*) and should be examined with respect to residential exposures as well. The practical challenge is assembling study populations that are large enough to have sufficient precision in examining subgroups. This also applies to the myriad forms of childhood cancer that are far too rare to consider in individual studies (e.g., osteosarcoma, Wilms' tumor). Meta-analysis using data from several completed studies would be one possible approach.

Although occupational literature suggests that electric and magnetic fields may be related to leukemia and brain cancer, these implications have not been pursued extensively in studies of residential exposures. There have been some studies focused

on residential magnetic field exposure and adult leukemia (*16,22*), and some studies of all forms of cancer (*6,15*). Although the results of residential studies have been largely negative, this avenue ought not be abandoned for several reasons. There is no biological reason to believe that childhood cancers are uniquely susceptible to magnetic fields. Given the rarity of childhood cancers and the consequent difficulty in conducting research, adult cancers are also worthy of concern. The continued support for a role of occupational exposures in the etiology of leukemia, especially acute myeloid leukemia; brain cancer; and, to a lesser degree, melanoma and lymphoma would encourage closer examination of residential exposures. Studies of residential exposure and adult brain cancer would be warranted by the evidence but have not yet been undertaken.

When biological understanding of the effects of electric and magnetic fields has progressed sufficiently, early, more highly prevalent disease indicators (biomarkers) of electric and magnetic field exposure may be identified. If such outcomes were sufficiently frequent, then prospective studies in which individual subjects would be monitored over days, or even weeks, for the development of the end point of interest could be developed, or banks of biological specimens could be exploited for nested case–control studies. Even if the marker were simply an integrator of exposure rather than a marker of a step in disease development, it potentially would be of value in identifying the most biologically potent aspect of exposure and in assisting in the design of studies of the more clinically significant end points. Some biological correlate of melatonin could serve as such a marker if more persuasive evidence linking the exposures and disease of interest to this pathway were to accrue, but a number of logistical issues would need to be overcome (*23*). Unfortunately, there are no other obvious candidates on the horizon, because the classic markers of genotoxicity (chromosomal aberrations, sister chromatid exchange, micronuclei) would not be expected to be useful in spite of some evidence for increased micronuclei formation among mice exposed to electric fields (*24*).

Considerations in Study Design

Several features of the past studies of residential magnetic field exposure and cancer have not been examined adequately as a potential basis for biased results. This includes several potential positive biases (which would produce spuriously elevated associations) and negative biases (which would produce spuriously reduced associations).

Control Selection and Wire Codes. The process of control selection in childhood cancer studies raises a number of concerns regarding the extent to which an unbiased sample from the study base has been attained. Several different methods have been used, including birth certificate controls (*5,8*), population register controls (*10*), and random digit dialing controls (*11,13*). Success in obtaining controls who are a random sample of the population from which the cases arose is difficult to demonstrate, given how little is known about the important determinants of exposure. In the study by Savitz et al. (*11*), for example, differential mobility of cases and controls may have introduced bias, but the presence or absence of such bias could not be demonstrated directly. In contrast, London et al. (*13*) obtained controls concurrently with case identification and obtained similar patterns of association, which suggests that differential mobility does not account for the pattern of results found by Savitz et al. (*11*).

Methodological evaluation of control selection for studies of residential magnetic field exposure should be undertaken. Because of our limited knowledge of the causes of childhood cancer as well as leukemia and brain cancer in adults, the exploration should extend beyond known risk factors to include: *a*) examination of a number of household characteristics in relation to magnetic fields, including socioeconomic characteristics, availability of a telephone, household composition (number and ages of children, etc.), patterns of residential movement and duration of occupancy, age of housing, and proclivity to participate in surveys; *b*) revaluation of the recently completed studies that used random digit dialing to assess the extent to which controls represented the general community and, particularly, households with children; and *c*) examination of alternative control selection strategies such as telephone directories, school records, and door-to-door canvassing. The costs of various alternatives and the yield of information should be examined.

Nonresponse and incomplete coverage are always possible sources of bias worthy of consideration. Although random digit dialing is a well-accepted technique, when the screening nonresponse and interview refusal rates are combined the losses can be rather severe (*25*). However, one of the advantages of using wiring codes as an exposure marker is that identifying an eligible address is sufficient to obtain the code; the home can be coded even if the respondent ultimately declines to be interviewed or allow in-home field measurements. The critical unknown relation is between nonresponse and wire code, and it is that uncertainty that makes evaluation of selection bias an important avenue to pursue.

A second product from a thorough examination of correlates of residential magnetic field exposures in the community would be improved guidance regarding possible confounders. Our knowledge of the etiology of childhood cancers in particular is quite limited. One approach to exploring possible confounders is to learn more about the characteristics of persons who live in homes classified as having elevated exposures. These characteristics would not, of course, necessarily be confounders, but they would satisfy at least one of the necessary criteria and could then be evaluated for their independent association with cancer risk.

Given the prominence of wire codes as a marker of cancer risk and the severely limited understanding of their implications, a broader evaluation of the sociology and geography of wire codes seems to be essential to identify the most valid approach to control selection and confounding. Understanding the patterns of wire codes within the community in broad and comprehensive terms of the spatial distribution of wires and wire codes, demographic and behavioral aspects of those who choose to live in homes of varying wire codes, and a comprehensive evaluation of empirical correlates of wire codes would serve several important research needs. Such knowledge would address a number of key methodological concerns simultaneously: the likelihood of selection bias in past case–control studies due to the constitution of the control group; suitable methods for selection of controls in future case–control studies; the likelihood of confounding by other exposures associated with wire codes; and the possibility that the impact of wire code on cancer risk is mediated through something other than the average magnetic field in the home.

Timing of Exposure. Timing of exposure, based either on when it occurs during the day or when it occurs during the individual's life, has received little attention. The possibility of an effect on melatonin synthesis suggests an emphasis on nighttime exposures, which is implicit in residential studies and would be especially applicable to studies of electric blankets and heated water beds [although it is questionable

whether the pineal gland is exposed from such sources (*26*)]. No studies of childhood cancer have examined adequately the temporal relation between exposure and disease by hypothesizing and testing expected induction and latent periods. Clearly, the temporal sequence of events leading to cancer is briefer than the corresponding multidecade process in adults.

There has been some cursory consideration of this issue in several studies (*5,11*), but none have obtained the desired lifetime residential exposure history to allow comprehensive evaluation of all potentially important time windows. Given our ignorance of the temporal course of disease induction, all exposure preceding disease is of potential interest, but presumably, some etiologically irrelevant exposures have been included in past studies and obscured any etiologic effects (*27*). The only well-established environmental cause of childhood cancer, exposure to ionizing radiation, has been shown to operate *in utero* (*28*), although the mechanisms of this type of exposure would be quite different than the possible effect of magnetic fields. Nonetheless, exposures *in utero* constitute one period of particular interest.

Logistically, assembling such lifetime histories is challenging and requires a residentially stable population. Urban areas with highly mobile populations such as Los Angeles or Denver are not good choices on this criterion, and many of the formerly occupied homes are outside the study region. On the other hand, without changes in residence, isolation of any effects of exposure in specified periods of life is impossible. Passive exposure assessment procedures (e.g., wire configuration codes) are more amenable to gathering a complete exposure history than procedures requiring respondent cooperation (e.g., in-home measurements). The presumption that any effects of magnetic fields only operate late in the etiologic process should be scrutinized since it seems to be based largely on the inability of such fields to cause mutations.

Ecologic Studies. Studies of disease patterns in populations over time or space, in which the group's magnetic field exposure is related to its cancer incidence, should be considered in spite of the well-known challenges posed by the ecologic fallacy (*29*) and from exposure misclassification (*30*). A principal motivation is to respond to critics who argue that secular changes in the use of electric power have been dramatic through this century and have produced a marked increase in electric and magnetic field exposure that has not yielded a corresponding increase in the cancers of concern (*31*). Obviously, this scenario would apply only if magnetic fields were a very strong contributor to cancer risk and if magnetic fields increased as electric power use increased, but such analyses still could place some upper bound on the magnitude of association.

This possibility could be examined empirically by isolating each of the assumptions and consequences. The argument that exposures have risen proportionately to the use of electric power has not been tested, and theoretical arguments against such a rise include the increasing suburbanization of America (with larger yards and greater distance from power lines), greater use of underground lines, higher voltages on lines, and more electricity-efficient appliances (such as microwave ovens and digital clocks). Wertheimer and Leeper's (*9*) examination of data from the Rhode Island study suggested that more recently occupied homes (which tended to be more suburban) had lower wire configuration codes on average.

Although historical measurements of individual exposures in the past are not available, the pattern of wire codes over time could be easily examined in several ways. Data on wiring configurations from the earliest studies conducted in the late 1970s could be compared to those conducted in the present, yielding a 10- to 15-year contrast. The housing stock at different historical periods could be estimated based on county tax assessment records to simulate the mix of wire codes in different historical periods. Even within completed studies, the dates of occupancy could be examined in relation to the wire codes of the homes.

Historical data on the cancers of interest are also necessary to conduct a study of time trends. A few cancer registries, such as the one in Connecticut, go back far enough in time. Mortality from childhood cancers (especially leukemias) has been so markedly changed by effective treatments that mortality data do not adequately reflect incidence. Additional challenges to making valid comparisons across long spans of time are posed by improving quality of diagnosis and the techniques used to classify cancers in different eras.

A study of geographic variation in residential exposure is more promising because potential confounders are likely to vary less markedly across the United States at the present time than over lengthy periods of interest. If there are systematic spatial differences in average exposure based on region, urban–rural differences, etc., then ecologic studies could examine efficiently the corresponding patterns of cancer incidence and mortality. For example, high current configurations appear to be more common in Los Angeles than Denver (*11,13*). Perhaps in large midwestern and northeastern cities, the prevalence of homes with such configurations may be greater still. Although limited to detecting gross differences, assessment of geographic patterns in mortality could yield some information if marked exposure gradients are present. The usual considerations in assessing the value of ecological studies apply: the opportunity to study wide exposure variation and a very different set of methodological concerns relative to those that apply to studies of individuals (*32*) is balanced against the effort required to conduct the research. The effort required to characterize spatial and temporal variation in exposure accurately is not known at present; but if not unduly demanding, it would at least produce a systematic evaluation of the argument that secular trends in exposure demonstrate the implausibility of an etiologic association with cancer.

An additional benefit of such an evaluation would be its use in pinpointing areas in which research might be most profitably conducted. Limited numbers of homes in the highest exposure groups have decreased the precision of all past studies, so areas with greater prevalence of higher magnetic field exposure would be most favorable for research.

Research Priorities on Residential Exposures

The preceding sections have sought to define comprehensively the issues deserving consideration and empirical research in order to advance our understanding of residential magnetic field exposures and cancer. Ideally, the reader should have sufficient information to reach independent conclusions regarding priorities. However, suggestions for the most pressing research needs are offered.

Two types of methodologic studies are needed to better interpret past studies and design future ones: assessment of individual magnetic field exposure sources and patterns and determination of correlates of wire code in the community. Either or both of these might be embedded into an ongoing study or conducted independently.

Evaluation of exposure sources through personal monitors and diaries combined with wire codes and home measurements would simultaneously define the exposure sources most worthy of study, address the relation between wire code and personal

exposure, and allow examination of different exposure metrics. Past studies relying on only one exposure source (e.g., wire codes, electric blankets) could be reinterpreted and future studies could focus on the most applicable exposures. Possible efforts to reduce exposures would also benefit from knowledge of how exposures are actually incurred.

Knowing the patterns of wire codes in the community is essential to evaluating theories of confounding, selection bias, and alternative causal pathways. Confounding would occur if wire codes were associated with an independent cause of cancer; selection bias would occur if study participants are unrepresentative of the target population; and an alternative causal pathway would operate if wire codes do not cause cancer through the resulting magnetic field exposure. Consideration of the sociological, geographical, and behavioral correlates of wire codes would provide empirical guidance to those who interpret studies that evaluate the possibility of bias as well as those who design future studies. It may seem more efficient to measure all possible confounders and adjust for them or to choose directly the correct control group, but without a clearer understanding of the likelihood of error, the inevitable design trade-offs cannot be made intelligently. Control groups, for example, may be selected in a limited number of ways, with random digit dialing the most popular for logistical reasons. The relationships among telephone access, social class, willingness to participate in surveys, and wire codes would be substantially valuable in determining whether some more onerus method of control selection is truly needed or whether studies need to be conducted in locales in which population registers are available. Without the pertinent background information on vulnerability to selection bias, selection of a single, appropriate, and credible control group is virtually impossible.

Occupational Exposure

Synopsis of Evidence

There has been a large number of studies in which job titles presumed to be indicative of above-background exposure to electric and magnetic fields have been examined in relation to cancer. Most commonly, such studies have focused on leukemia and brain cancer. The evidence has been reviewed in several publications (*1,2,33–35*). Most reviewers share the conclusion that these reports generally support an association between work in electrical occupations and the risk of leukemia, especially acute myeloid leukemia, and brain cancer. Magnitudes of association vary greatly from null to sizable increases, but elevations in risk on the order of 1.5 to 2.0 are commonly reported, especially in proportionate mortality, incidence, and case–control studies. Most cohort studies have not found the associations to exist to the same degree. Given the lack of sophistication in exposure assignment, the degree of consistency across diverse populations is notable. Other cancers, such as melanoma (*36,37*), lymphoma (*38*), and male breast cancer (*39–41*) have been implicated, but with less replication.

The structure of these studies has included registry-based examinations of proportionate mortality or incidence, registry- and community-based case–control studies, and historical cohort studies among electrical workers. Starting with Milham's (*42*) report, all have used job title as the exposure surrogate with some refinements in terms of more sophisticated classification systems (*43*) but largely relying on an intuitively developed listing of jobs thought to entail elevated electric and magnetic fields (e.g., electrician, lineman, television repairman).

A separate avenue of research on occupational exposures and cancer is that of paternal influences on childhood cancer risk. Spitz and Johnson (*44*) found that children who died from neuroblastoma more often had fathers who were employed in occupations thought to have electromagnetic field exposures than controls did. The increased risk was concentrated among electronics workers. These results were replicated to some extent by Wilkins and Hundley (*45*) in a similarly designed case–control study of neuroblastoma, but Bunin et al. (*46*) reported an absence of increased risk for neuroblastoma in relation to paternal exposure to electromagnetic fields. The mechanism for such an effect is tenuous, given that the agent is incapable of causing mutations in sperm, but perhaps some other mechanism of interfering with sperm production is applicable.

Exposure Assessment Needs

Evaluation of Nonutility Populations. Among electrical workers, electric utility workers have been most actively considered. While such studies are in progress, there is a need to identify additional groups of workers who are suitable for study. Studying other populations would provide an assessment of the replicability of the evidence from utility workers. More important, the actual exposure circumstances in terms of field frequencies and temporal patterns vary markedly among electrical workers, and if any such exposures are carcinogenic, there may be some more and some less potent forms of exposure among the different groups of workers. The variability in exposure circumstances, both quantitatively and qualitatively, is much greater in the workplace than in the home. This should produce more informative studies in the occupational setting if those exposures can be characterized adequately.

In the past, the interest has been in relatively rare cancers (leukemia and brain cancer), and the examined populations had to be sizable and there had to be a mechanism for identifying them (company or union records, for example). Starting with a roster of candidate worker groups, exposure measurement surveys would be essential to indicate that they truly have above-background exposures and to characterize the general patterns of that exposure. Candidate populations widely discussed include aluminum workers, electric railroad workers, arc welders, and other workers who work near electric motors.

Community-based studies may also address occupational exposures using some explicit or implicit job-exposure matrix that links job title to exposure. This is the basis of virtually all of the existing literature based on death certificates or cancer registries. It seems that the inability to characterize exposures generically across widely divergent job titles and industries gives these studies limited potential to advance the literature. Although it would be useful if widespread electric and magnetic field exposure surveys enabled us to characterize adequately exposures associated with specific job titles and industries, the variation within those groupings is likely to limit the value of broad job titles. In contrast, within an industry, the level of refinement can be much greater.

Improved Historical Markers. Because leukemia and brain cancer are so rare, it is not feasible to undertake prospective cohort studies for either, although future interest in more common cancers such as prostate cancer or female breast cancer would introduce that possibility. Nonetheless, most future research will continue to rely on historical markers of exposure, typically job titles and other documented information on work activities and locations. There is a need to validate such markers through present measurements and by other indirect means. There is also a need for imaginative approaches to reconstructing the historical exposures of interest, including simulation of past work environments and practices where possible. In a recent study of workers who used video display terminals, for example, exposures associated with outdated equipment were

estimated by retrieving some of the old equipment for measurements (*47*).

In addition to evaluating the adequacy of the surrogate marker in the future, it would be useful to assess just what exposures the surrogate marker predicts. Preliminary examination of job titles of electric utility workers (*48*) suggests that jobs thought to be exposed have a stronger gradient for magnetic than electric fields. Similarly, one might ask about the distinctions based on job title for different exposure indices (e.g., mean, median, peak) or for different frequencies. Future epidemiologic and laboratory research would benefit from obtaining more specific suggestions about the form of exposure reflected from job titles.

A more ambitious advancement would require incorporation of nonoccupational exposure sources into occupational studies. This is more feasible than the converse, extending residential studies to incorporate workplace exposures, since validated markers of nonoccupational exposure are available. Specifically, cohort mortality studies could be followed up with nested case–control studies that include interviews with living subjects or next of kin for decedents in which nonoccupational exposures would be estimated. Identification of the residences in which they had lived could be coupled with wire coding or measurements of those homes. Possible use of appliances such as electric blankets and heated water beds could be queried and probably could be accurately reported by a surviving spouse or child. Qualitative and probably even quantitative indices of exposure that combine diverse sources of exposure could then be developed and analyzed in relation to cancer risk. If some measure of total dose is the important one, such combined indices would be markedly better at predicting cancer risk than either component alone, since a sizable contribution to total dose comes from each source (*17,48*).

Health Outcomes

As seen in residential exposure, there is potential value in examining more specific subtypes of cancer. This is predicated on the possibility that more specific forms of cancer, defined by histology and cytogenetics, may show stronger relations to electric and magnetic field exposure. The generally stronger associations found for acute myeloid leukemia compared to other leukemias (*35*) and the recent evidence that astrocytomas show markedly stronger associations with electrical work than other forms of brain cancer (*21*) suggest that such efforts could yield important information.

On the other hand; there is no clear biological rationale for focusing on leukemia and brain cancer only, and as noted above, associations with melanoma, lymphoma, and male breast cancer have also been reported. We should retain the ability to discover that some other type of cancer is strongly related to exposure or to confirm or refute the reports of such associations in other studies. In general, studies that adequately can address many forms of cancer, such as historical cohort studies or case–control studies used with a case group of all cancers, are preferred to those that cannot.

In parallel with the suggestion for residential studies, markers of exposure or disease that are prevalent enough to be studied prospectively would be highly desirable. Current technology permits exposure assessment over periods as long as several weeks, but without some end point that can be observed in a short time frame and of adequate prevalence, such measurements can only serve to validate such exposure proxies as job title.

Study Design

Consideration of Other Exposures. Because of substantial contributions from the residence and appliance use (which may have substantial between-subject variability), it is understood that the workplace is not the only or even the dominant source of exposure (*49*). Logistical constraints make it difficult to integrate exposures across diverse sources, so further examination of the consequences of ignoring nonoccupational sources should be more carefully considered. Exposure surveys suggest the absence of any association between workplace and nonworkplace exposures among utility workers (*48,49*), but the consequences for studies of workplace exposures are not clear. There may also be reason to believe that some highly exposed occupational groups would tend to have exposures in their hobbies, such as operating radios or other electrical equipment. Considering a number of possible relations between work and nonwork exposures for their impact on observed dose–response gradients would assist in the interpretation of past studies and planning of future studies.

Confounding has been of particular interest in this literature, because a job title virtually always suggests exposures in addition to the electric and magnetic field exposure of interest. Many groups of electrical workers have the potential for occupational exposure to solvents, polychlorinated biphenyls (PCBs), soldering or welding fumes, etc. The impact of many of these agents on cancer risk is poorly understood, but future studies should examine them as effectively as possible. Nonetheless, hypothetical calculations for the magnitude of confounding by a carcinogen as potent as cigarette smoking (*50*) suggest that extreme and perhaps implausible scenarios are required to invoke such confounding as a critical threat to the validity of these studies.

An efficient approach to examining cancers that have not been thoroughly considered, such as lymphoma and melanoma, would be to pool results from the numerous surveys and cohort studies using meta-analysis. By examining a wide array of cancer types in that manner, patterns may emerge that were not previously appreciated. Even leukemia and brain cancer might be better understood through a more quantitative integration of the literature.

Effect-Modification by Timing and Other Agents. A critical deficiency in nearly all of the past studies that have relied on registry data is a failure to consider temporal aspects of exposure and disease. The extreme version is the death certificate-based study (*42,51*), in which there is no information whatsoever on the duration of employment in the listed job, when work in the job ceased, or what other jobs were held. Presumably, if there is an etiologic relation, like all others identified to date, there is some specificity for the induction and latent periods. Restricting the windows of exposure to the pertinent ones would enhance any causal associations. The assumption that electric and magnetic fields act at late stages should be addressed empirically by assessing cancer risk in relation to exposures in a recent time window. More generally, a flexible, trial and error approach to specifying the potential windows of importance should be adopted (*27*), especially when considering an agent for which the underlying biological processes are so poorly understood.

Finally, the possibility that the effects of electric and magnetic fields are enhanced by exposure to other agents should be evaluated. If such exposures are thought to act in concert with genotoxic agents, then one would expect some effect modification to be discernible. As noted above, there are a number of potentially carcinogenic agents such as solvents and PCBs thought to be prevalent among electrical workers, so the ability of such exposure to potentiate the effect of electric and magnetic fields could be examined. If future study designs permit consideration of nonoccupational exposures, then cigarette smoking would be of great interest as a potential effect modifier.

Research Priorities in Studies of Occupational Exposure

Among the suggestions offered for extending knowledge regarding occupational

exposure to electric and magnetic fields and cancer, two avenues are the highest priorities. First, exposure patterns of groups of workers potentially amenable to epidemiologic study need to be assessed. Identification of groups with above-background exposure is an absolute requirement, but the likely diversity of exposure forms and patterns (frequencies, temporal patterns of exposure, historical exposures) also should be understood in order for the epidemiologic studies to begin to clarify the circumstances in which adverse effects are and are not observed.

Second, following identification of suitable groups for epidemiologic study, more empirical information on occupational groups with a diversity of forms and patterns of electric and magnetic field exposure is needed. Traditional cohort or nested case–control studies within industries are most likely to be informative. By studying workers with elevated but distinct exposure patterns, the consistency of any increased risk of leukemia, brain cancer, or other cancers can be assessed and there is the possibility of identifying a group with a more potent exposure pattern.

Conclusions

The strategies outlined above can be divided into three groups by considering the effect the methodological deficiencies would have on the estimated measures of effect (i.e., the risk ratio or odds ratio). Improvements typically have one of the following goals: *a*) to reduce bias toward the null (false negative results), enhancing the magnitude of association if an underlying etiologic effect of electric and magnetic fields on cancer truly exists; *b*) to reduce bias away from the null (false positive results), diminishing positive associations reported in past literature if they are actually a result of errors in study design or execution; and *c*) to enhance precision of study results.

The principal source of potential bias toward the null is nondifferential exposure misclassification. Such nondifferential misclassification applies to all of the sources of discrepancy between operational measures of exposure in a given study (e.g., job title, wire code of residence) and the precise biological measure of dose that is etiologically effective (e.g., time-integrated total magnetic field, time above 2 mG). Any study design strategy that better approximates the biologically relevant dose (typically identified through trial and error) will enhance the magnitude of association if an etiologic effect is present. Better wire codes, more precise job titles, incorporation of other sources of exposure, consideration of different exposure parameters, and examination of varying time windows of exposure all have that intended effect. The search for effect modifiers can be viewed in this light, with the group in which the effect of electric and magnetic field is enhanced reflecting a stronger association with cancer. Also, the effort to define more specific subgroups of cancer more strongly associated with exposure fits into this category. If reducing misclassification increases the measures of association, then the likelihood that there is a true etiologic effect present is enhanced. Conversely, extensive unsuccessful efforts to identify a stronger relation could be interpreted as evidence against a causal effect, although it would not be clear when the search should be ended.

Potential sources of bias away from the null are less obvious in past studies of electric and magnetic fields and cancer. In the community-based studies, selection bias in the constitution of the control groups is an important consideration. The constitution of the control groups is challengeable (generally based on random digit dialing), as well as the potential bias due to nonresponse. For this to produce bias away from the null, a particular pattern (e.g., missing higher exposure controls) would have to be invoked. In both occupational and residential studies, the potential for unmeasured positive confounders should continue to be examined. Specific, testable candidates for sources of bias away from the null are needed to make progress in this area.

Finally, some of the above strategies are intended primarily to enhance precision. Identification of communities or workforces with a higher prevalence of elevated electric and magnetic field exposure should yield more precise estimates of effect. Meta-analyses of completed studies have the potential to yield increased precision in estimates of dose–response gradients. Finally, study of early disease markers could provide a much more common outcome than cancer, with consequently greater precision.

The strategies suggested in this paper are intended to open research avenues. Although some of the more obvious studies have been done or are in progress, there are some other pathways that would yield new insights regardless of the results obtained.

REFERENCES

1. Coleman M, Beral V. A review of epidemiological studies of the health effects of living near or working with electricity generation and transmission equipment. Int J Epidemiol 17:1–13 (1988).
2. U.S. EPA. Evaluation of the potential carcinogenicity of electromagnetic fields. Review draft EPA/600/6–90/005B. Washington, DC: U.S. Environmental Protection Agency, 1990.
3. Savitz DA, Pearce NE, Poole C. Methodological issues in the epidemiology of electromagnetic fields and cancer. Epidemiol Rev 11:59–78 (1989).
4. Kavet RI, Banks RS. Emerging issues in extremely low frequency electric and magnetic field health research. Environ Res 39:386–404 (1986).
5. Wertheimer N, Leeper E. Electrical wiring configurations and childhood cancer. Am J Epidemiol 109:273–284 (1979).
6. Wertheimer N, Leeper E. Adult cancer related to electrical wires near the home. Int J Epidemiol 11:345–355 (1982).
7. Kaune WT. Assessing human exposure to power-frequency electric and magnetic fields. Environ Health Perspect (Suppl 4):121–133 (1993).
8. Fulton JP, Cobb S, Preble L, Leone L, Forman E. Electrical wiring configurations and childhood leukemia in Rhode Island. Am J Epidemiol 11:292–296 (1980).
9. Wertheimer N, Leeper E. Electrical wiring configurations and childhood leukemia in Rhode Island (letter). Am J Epidemiol 111:461–462 (1980).
10. Tomenius L. 50-Hz electromagnetic environment and the incidence of childhood tumors in Stockholm County. Bioelectromagnetics 7:191–207 (1986).
11. Savitz DA, Wachtel H, Barnes FA, John EM, Tvrdik JG. Case–control study of childhood cancer and exposure to 60-Hz magnetic fields. Am J Epidemiol 128:21–38 (1988).
12. Myers A, Clayden AD, Cartwright RA, Cartwright SC. Childhood cancer and overhead powerlines: a case–control study. Br J Cancer 62:1008–1014 (1990).
13. London SJ, Thomas DC, Bowman JD, Sobel E, Cheng T-C, Peters J M. Exposure to residential electric and magnetic fields and risk of childhood leukemia. Am J Epidemiol 134:923–937 (1991).
14. Savitz DA, John EM, Kleckner RC. Magnetic field exposure from electric appliances and childhood cancer. Am J Epidemiol 131:763–773 (1990).
15. McDowall ME. Mortality of persons resident in the vicinity of electricity transmission facilities. Br J Cancer 53:271–279 (1986).
16. Severson RK, Stevens RG, Kaune WT, Thomas DB, Heuser L, Davis S, Sever LE. Acute nonlymphocytic leukemia and residential exposure to electromagnetic fields. Am J Epidemiol 128:10–20 (1988).
17. Preston-Martin S, Peters JM, Yu MC, Garabrant DH, Bowman JD. Myelogenous leukemia and electric blanket use. Bioelectromagnetics 9:207–213 (1988).

18. Monson RR. Editorial commentary: epidemiology and exposure to electromagnetic fields. Am J Epidemiol 131:774–775 (1990).
19. Kaune WT, Stevens RG, Callahan NJ, Severson RK, Thomas DB. Residential magnetic and electric fields. Bioelectromagnetics 8:315–335 (1987).
20. Sandler DP, Collman GW. Cytogenetic and environmental factors in the etiology of acute leukemias in adults. Am J Epidemiol 126:1017–1032 (1987).
21. Mack W, Preston-Martin S, Peters JM. Astrocytoma risk related to job exposure to electric and magnetic fields. Bioelectromagnetics 12:57–66 (1991).
22. Coleman MP, Bell CMJ, Taylor H-L, Primic-Zakelj M. Leukemia and residence near electricity transmission equipment: a case–control study. Br J Cancer 60:793–798 (1989).
23. Wilson BW, Anderson LE. ELF electromagnetic-field effects on the pineal gland. In: Extremely low frequency electromagnetic fields: the question of cancer (Wilson BW, Stevens RG, Anderson LE, eds). Columbus, OH: Battelle Press, 1990.
24. El Nahas SM, Oraby HA. Micronuclei formation in somatic cells of mice exposed to 50-Hz electric fields. Env Molecular Mutag 13:107–111 (1989).
25. Olson SH, Kelsey JL, Pearson TA, Levin B. Evaluation of random digit dailing as a method of control selection in case–control studies. Am J Epidemiol 135:210–222 (1992).
26. Florig HK, Hoburg JF. Power-frequency magnetic fields from electric blankets. Health Physics 58:493–502 (1990).
27. Rothman KJ. Induction and latent periods. Am J Epidemiol 114:253–259 (1981).
28. Committee on the Biological Effects of Ionizing Radiations. Health effects of exposure to low levels of ionizing radiation. Beir V. National Research Council. Washington, DC: National Academy Press, 1990.
29. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. Int J Epidemiol 18:269–274 (1989).
30. Brenner H, Savitz DA, Jöckel K-H, Greenland S. The effects of nondifferential exposure misclassification in ecological studies. Am J Epidemiol 135:85–96 (1992).
31. Poole C, Trichopoulos D. Extremely low frequency electric and magnetic fields and cancer. Cancer Causes Control 2:267–276 (1991).
32. Morgenstern H. Principles of study design in environmental epidemiology. Environ Health Perspect (Suppl 4):23–38 (1993).
33. Theriault GP. Health effects of electromagnetic radiation on workers: epidemiologic studies. In: Proceedings of the scientific workshop on the health effects of electric and magnetic fields on workers (Bierbaum PE Peters JM, eds.). U.S. DHHS, National Institute for Occupational Safety and Health, Cincinnati, OH: 1991; 191–24.
34. Savitz DA, Ahlbom A. Epidemiologic evidence on cancer in relation to residential and occupational exposures. In: Biologic effects of electric and magnetic fields. (Carpenter DO, ed.) New York: Academic Press (in press).
35. Savitz DA, Calle EE. Leukemia and occupational exposure to electromagnetic fields: review of epidemiological surveys. J Occup Med 29:47–51 (1987).
36. Olin R, Vagero D, Ahlbom A. Mortality experience of electrical engineers. Br J Ind Med 42:211–212 (1985).
37. Vagero D, Ahlbom A, Olin R, Sahlsten S. Cancer morbidity among workers in the telecommunications industry. Br J Ind Med 42:191–195 (1985).
38. Milham S Jr. Mortality in workers exposed to electromagnetic fields. Environ Health Perspect 62:297–300 (1985).
39. Demers PA, Thomas DB, Rosenblatt KA, Jiminez LM, McTiernan A, Stalsberg H, Stemhagen A, Thompson, WD, Curnen MGM, Satariano W, Austin DF, Isacson P, Greenberg RS, Key C, Kolonel LN, West DW. Occupational exposure to electromagnetic fields and breast cancer in men. Am J Epidemiol 134:340–348 (1991).
40. Tynes T, Andersen A. Electromagnetic fields and male breast cancer (letter). Lancet 2:1596 (1990).
41. Matanoski GM, Breysse PN, Elliott EA. Electromagnetic field exposure and male breast cancer (letter). Lancet 1:737 (1991).
42. Milham S. Mortality from leukemia in workers exposed to electrical and magnetic fields (letter). N Engl J Med 307:249 (1982).
43. Lin RS, Dischinger PC, Conde J, Farrell KP. Occupational exposure to electromagnetic fields and the occurrence of brain tumors. J Occup Med 27:413–419 (1985).
44. Spitz MR, Johnson CC. Neuroblastoma and paternal occupation: a case–control analysis. Am J Epidemiol. 121:924–929 (1985).
45. Wilkins JR, III, Hundley VD. Paternal occupational exposure to electromagnetic fields and neuroblastoma in offspring. Am J Epidemiol 131:995–1008 (1990).
46. Bunin GR, Ward E, Kramer S, Rhee CA, Meadows AT. Neuroblastoma and parental occupation. Am J Epidemiol 131:776–780 (1990).
47. Schnorr TM, Grajewski BA, Hornung RW, Thun MJ, Egeland GM, Murray WE, Conover DL, Halperin WE. Video display terminals and the risk of spontaneous abortion. N Engl J Med 324:727–33 (1991).
48. Deadman JE, Camus M, Armstrong BG, Heroux P, Cyr D, Plante M, Theriault G. Occuparional and residential 60-Hz electromagnetic fields and high-frequency transients: exposure assessment using a new dosimeter. Am Ind Hyg Assoc J 49:409–419 (1988).
49. Bracken TD. EMDEX Project Report, Volumes 1-3. Palo Alto, CA: Electric Power Research Institute, 1990.
50. Axelson O, Steenland K. Indirect methods of assessing the effects of tobacco use in occupational studies. Am J Ind Med 13:105–118 (1988).
51. Loomis DP, Savitz DA. Mortality from brain cancer and leukemia among electrical workers. Br J Ind Med 47:633–638 (1990).

Biologically Based Epidemiological Studies of Electric Power and Cancer

Richard G. Stevens

Pacific Northwest Laboratory, 906 Battelle Boulevard, Richland, WA 99352

As societies industrialize, the health profile of the population changes; in general, acute infectious disease declines and chronic disease increases. Use of electricity is a hallmark of the industrialization process, but there has been no suspicion that electricity could increase the risk of cancer. Recently, however, a number of epidemiologic studies have suggested that electromagnetic fields (EMF) may do just that. Although few cancer experiments have been done yet, there are a number of biological effects of EMF reported in the literature that might provide bases for designing cancer experiments and epidemiologic studies. These include effects of EMF on: *a)* DNA transcription and translation, *b)* calcium balance in cells, and *c)* pineal production of melatonin. Alterations in DNA transcription and translation could have pleiotropic effects. Disruption of calcium homeostasis has many implications including oncogene activation, promotional activity via protein kinases and ornithine decarboxylase (ODC), and increasing oxidative stress. Reduction of melatonin suggests a possible increased risk of cancers of hormone-dependent tissues such as breast and prostate. The idea that a cancer-causing agent must either be an initiator or a promoter should be discarded; indeed, the phenomenologic meaning of these two terms has become confused with imputed mechanistic necessity in recent years. Agents that affect division of normal cells or of fully transformed cells can play an important role in clinical cancer development quite apart from initiation or promotion. Epidemiologic studies of EMF and cancer should attempt to take account of other products of electric power (e.g., light at night) or factors associated with occupational EMF exposure (e.g., toxic chemicals) that may increase cancer risk and therefore act as cofactors or confounders. Epidemiology and laboratory studies should act synergistically in determining if there is a problem and identifying mitigation strategies if needed. — Environ Health Perspect 101(Suppl 4):93–100 (1993).

Key Words: Epidemiology, pineal, calcium, cancer, DNA transcription

Introduction

Industrialization is associated with changes in the health status of a population. There are large differences among countries in the rates of cancer of specific sites (*1*). The ratio of colon cancer incidence in Connecticut (high rates) to that in Nigeria (low rates) is 10; the ratio of breast cancer incidence in Canada (high rates) to that in the non-Jews in Israel (low rates) is 7. Total cancer incidence varies by about 3-fold among countries. Migrants from one area take on the cancer rates of their new homes, and there are known, or strongly suspected, causes of some of the most common cancers in each society such as hepatitis B virus and liver cancer in Taiwan, and cigarette smoking and lung cancer in the United States. These observations suggest racial factors do not account for the large variations in cancer risk around the world. Rather, lifestyle is believed to play a major role in cancer occurrence. However, it is not clear how much of the variation in risks can be accounted for by specific lifestyles. At one extreme is the confidence that cigarette smoking accounts for 90% of lung cancer in high risk societies; at the other extreme is the lack of understanding of societal differences in stomach cancer risk and why it has declined in the industrialized world. Understanding of breast cancer lies somewhere in the middle. Although major risk factors have been identified for breast cancer, they do not appear to account for the observed international variation in breast cancer rates (*2*). The reason for the large differences in risk among countries remains a mystery.

Until recently there was no suspicion that any aspect of the use of electric power might play a role in explaining cancer differences among societies. Motivated by the seminal, although limited, observations of Wertheimer and Leeper (*3,4*) and by recently emerging evidence for biological effects of electromagnetic fields (EMF) from the laboratory (*5*), interest in the possible role of electric power in cancer risk has increased dramatically.

This paper is deliberately speculative and suggests new avenues for epidemiological inquiry. Biological effects of EMF are described then reexamined after a presentation of the two-stage model for cancer (*6*) in order to determine how the biology might relate to carcinogenesis. Implications for timing of exposure, cancer types, cofactors (effect modifiers), and confounders in epidemiologic studies are suggested.

This manuscript was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990 – September 1992.

This work was supported by the U.S. Department of Energy under contract no. DE-AC06-76RLO 1830 and the Health Effects Institute, Cambridge, MA.

EMF Biological Effects

The biophysical mechanisms of interaction of EMF with biological systems is the subject of intense interest, but is beyond the scope of this paper (*7–10*). Many biological effects of EMF have been reported in the literature over the years and it is not possible here to evaluate each fairly as it might affect cancer risk. Many of these are single reports that have not yet been assessed in other laboratories. However, several biological effects that do have such implications are discussed as they might relate to cancer risk and as they might be accommodated in the design of epidemiological studies. The following discussion is selective and brief.

DNA Mutation

There is evidence that EMF does not damage DNA directly (*11,12*), although there is one report of increased micronuclei production by 50-Hz electric field exposure (*13*). Effects of EMF on chromosome segregation might lead indirectly to mutation, although this possibility is in the very early stages of evaluation. In addition, there has been very little study of possible effects of EMF on DNA repair (*12*).

DNA Transcription and Translation

Although direct DNA damage from EMF has not been demonstrated, there have been reports of altered mRNA and protein synthesis (*14,15*). Goodman et al. (*16*) reported increased RNA transcription in dipteran salivary gland cells after exposure to a magnetic field. Alterations in protein synthesis have also been observed, and there is a shift to proteins of lower molecular weight with a higher net charge. These results are consistent with

a model in which translation of protein from mRNA is interrupted (*17*). Such an interruption of protein synthesis would be interesting in light of the emerging theory of cell cycle regulation by a complex interplay of Rb, the protein product of the retinoblastoma antioncogene, with a series of other nuclear proteins (*18*). If the interaction of Rb with other proteins can be disrupted by EMF due to shortening of the polypeptides and consequent changes in conformation, then effects on cell cycle regulation would be expected.

Calcium

Since the report by Bawin and Adey (*19*), there has been growing interest in the role of EMF in calcium balance in cells and in ion flow in general (*20*). A window of effect that depends on frequency and amplitude seems to determine whether EMF will affect calcium balance (*21*). This idea was tested by Smith et al. (*22*) who reported that diatom mobility, which is dependent on calcium concentration in the growth medium, can be affected by EMF at specified frequency-intensity windows. Reese et al. (*23*) replicated this result with considerably more variability in response.

Electromagnetic stimulation has been used clinically to improve bone healing. Luben et al. (*24*) examined the biological mechanisms for this and suggested that the effects are mediated at the plasma membrane either by interference with hormone binding to osteoblasts or by blocking receptor-cyclase coupling in the membrane. They further speculate that Ca^{2+} movement may be involved. Lyle et al. (*25*) reported that a 13.6 Hz magnetic field with a peak intensity of 20 μT doubled calcium uptake of normal lymphocytes and of a lymphoma cell line. These researchers point out that calcium concentration and intracellular distribution affects many cellular functions, in particular the protein kinase C, which is important to lymphocyte activation and proliferation.

Ornithine Decarboxylase

Ornithine decarboxylase (ODC) is required for polyamine biosynthesis, and levels of it are high in rapidly dividing cells. Tumor promoters such as 12-*O*-tetradecanoylphorbol-13-acetate (TPA) rapidly increase ODC activity in cells. Byus et al. (*26*) reported that a 1-hr exposure of several different cell lines to a low-intensity (10 mV/cm) 60-Hz electric field increased ODC activity (although still far below the levels induced by TPA). They also reported that in Reuber H35 hepatoma cells, 3 hr of exposure first led to an increase, then to a decrease in ODC activity. The monotonically increasing dose–response that might be expected from a toxic chemical model is not evident. Instead, a 1-hr exposure to 10 and to 0.1 mV/cm increased activity, and 5 and 1 mV/cm had no effect.

Immune Function

Lyle et al. (*27*) reported that exposure to a 450-MHz field amplitude modulated at 60 Hz inhibited the toxicity of T-lymphocytes by 20%. The carrier wave had no effect, and amplitude modulation higher (up to 100 Hz) and lower (down to 3 Hz) showed less inhibition than at 60 Hz. Exposure to a 60-Hz electric field yielded similar results (*28*). Effects of EMF on intracellular calcium concentration might account for this inhibition (*25*). Experiments in animals have not shown effects of EMF on immune function (*29*). In humans, however, there is a report of impaired immune function in aluminum reduction plant workers who have high magnetic field and volatilized aromatic hydrocarbon exposures (*30*). The study was undertaken because five cases of B-cell lymphoma occurred over a 7-year period in an aluminum plant in Washington state when only 0.2 were expected. Among 23 apparently healthy volunteers from this plant, there was a significant increase in mean T8 and T4 levels, and there was a significant alteration in T4/T8 ratios in ten of the subjects due to disproportionate elevations of the T8 subpopulation. Another investigation, from Yugoslavia, reported that workers occupationally exposed to microwaves had greater numbers of micronuclei and other genomic abnormalities in a sample of their peripheral lymphocytes than controls (*31*).

Pineal Function

The pineal gland is a neuroendocrine transducer that provides a hormonal signal synchronized to the daily light and dark cycle (*32*). Melatonin, the principal pineal hormone, exerts a generally suppressive action on other endocrine glands. Reduced circulating concentrations of melatonin can result in increased prolactin release by the pituitary and increased estrogen and testosterone release by the gonads (*33,34*). Production of melatonin is suppressed by light perceived by the retina. Hence, circulating melatonin concentrations are low in daylight and higher at night (*35*).

Cohen et al. (*36*) suggested that reduced pineal melatonin production might increase human breast cancer risk because lower melatonin output would lead to an increase in circulating estrogen levels and would stimulate the proliferation of breast tissue. Indeed, early menarche, late menopause, and nulliparity are each associated with an increased risk of breast cancer (*37*), and all result in a longer period for proliferation of the breast epithelial stem cells at risk (*38*).

A number of investigators have reported that EMF, under some circumstances, can reduce or suppress melatonin production by the pineal gland (*39–45*). These observations provided a natural framework for postulating that EMF may influence risk of certain cancers, in particular breast and prostate cancer (*46,47*). Melatonin inhibits the growth of Dunning prostatic cancer cells transplanted into rats (*48*), is oncostatic to human breast cancer cells in vitro (*49*), and inhibits chemically induced breast cancer in rats (*50*). Light at night (LAN) suppresses melatonin production (*34,51*). Thus, two products of electric power, EMF and LAN, may reduce melatonin in humans and influence risks of breast and prostate cancers.

Two groups have reported experiments in rats in which a 50-Hz magnetic field increased chemically induced mammary cancer yield (*52,53*). Beniashvili et al. (*52*) used a 20-μT magnetic field, whereas Mevissen et al. (unpublished data) used a 100 μT magnetic field. Should these findings be replicable in other laboratories they will be important.

Two-Stage Model for Carcinogenesis

The term *carcinogenesis* is meant to convey the entire process from a beginning with normal cells in a healthy tissue to an ending with a diagnosed malignant tumor. The terms *initiation* and *promotion* were originally defined on strictly phenomenological grounds: an initiator is an agent that alone does not produce tumors, yet when followed by a promoter yields many tumors. A promoter is an agent that alone yields no tumors, yet when preceded by an initiator yields many tumors. A promoter followed by an initiator also yields no tumors. The tumors on mouse skin originally used to define initiator and promoter were not in fact cancer; they were benign lesions, many, if not most, of which regressed upon cessation of the promoter. In the last two decades, the terms initiator and promoter have come by some to be used as if they offer deep insight into the process of carcinogenesis, as if a cancer-causing agent is either one or the other, and as if both must be necessary for cancer to occur. An alternative view is that the original definitions of these agents offer evidence on the process of carcinogenesis, but do not define it. The following paragraphs present a model for cancer and a specific interpretation of what initiators and promoters are and how they fit into the larger scheme of carcinogenesis according to this model.

Offering a definition of cancer in quantitative or even qualitative terms runs the risk of ignoring some or many of the myriad characteristics of cancer that have been reported.

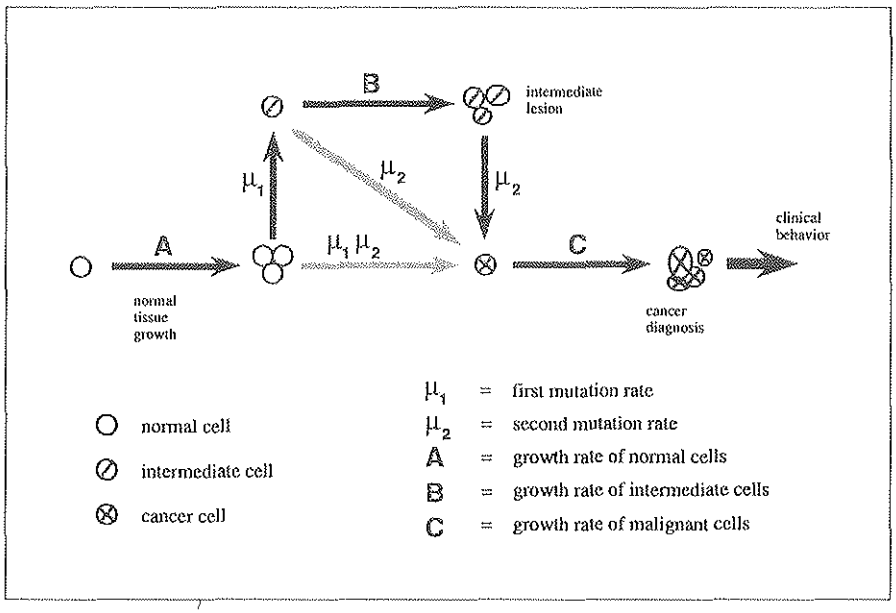


Figure 1. Two-stage model for cancer, adapted from Moolgavkar and Knudson (*6*). A normal cell divides as part of a normal tissue. With small probability, a normal cell may divide to give one normal and one intermediate cell, which in turn can divide to become an intermediate lesion. With small probability, an intermediate cell may divide to give one intermediate and one cancer cell. A promoter greatly increases the chance that a malignant cell will arise by increasing the pool of intermediate cells.

However, without a model it is difficult to formulate meaningful studies that might contribute new understanding. Figure 1 is adapted from Stevens et al. (*54*) and depicts the two-stage model for cancer developed by Moolgavkar and Knudson (*6*). This model is biologically simple without, so far, being shown to be simplistic; it is currently the most parsimonious model consistent with the body of knowledge on cancer. It provides an appealing framework within which to evaluate the potential carcinogenicity of a putative cancer-causing agent, and to design studies based on specific mechanisms.

According to the model, the drivers of cancer appearance are mutations to DNA and the growth kinetics of normal, intermediate, and cancer cells. A normal cell divides with growth rate A to maintain healthy turnover of a normal tissue. With low probability μ_1, a normal cell may divide and produce a normal cell and an intermediate cell that has suffered one of the two DNA mutations required for malignant transformation. This intermediate cell divides with growth rate B to form an intermediate lesion. An intermediate cell may also divide with low probability μ_2 to yield an intermediate cell and a malignant cell that has suffered the second necessary mutation to DNA. A normal cell may suffer both DNA mutations at a single division with very low probability, μ_1 times μ_2. Malignant cells can divide with growth rate C to form a malignant tumor (i.e., progression). After clinical diagnosis, behavior of the tumor is affected by the clinical treatment as well as all the endogenous factors previously affecting its growth.

Within the context of the two-stage model, an initiator is a mutagen delivered at low dose, thus increasing mutation rates μ_1 and μ_2, and the probability of mutating both of the two genes necessary for malignant transformation is low. A promoter is an agent that increases the proliferation of intermediate cells, B (and perhaps normal cells, A). This increases the chance that the second mutation will occur by, for example, mitotic recombination if the two events must occur in both homologues of the same gene (e.g., an antioncogene, also known as a tumor-suppressor gene), or by mutation of a second necessary gene (e.g., a second protooncogene). A complete carcinogen is either a mutagen delivered at high dose or an agent that is both mutagenic and mitogenic. An intermediate cell may not be subjected to a promoter but may still suffer the second mutation to become malignant (depicted by the diagonal arrow leading from the single intermediate cell to the single malignant cell in Figure 1). A premalignant lesion is a proliferation of intermediate cells that are heterozygous for an antioncogene, or that have only one of two different and necessary protooncogenes activated. A malignant tumor is a proliferation of malignant cells which have both necessary mutations. At present, behavior of a malignant tumor is not addressed in the two-stage model. In particular, there may be further genetic alterations necessary for the ability to metastasize. It must be noted that cancer can arise in the absence of application of a promoter since normal cell turnover will still allow for mutation of DNA by a mutagen. Cancer can also arise in the absence of a mutagen since spontaneous DNA mutations do occur.

A prediction based on the two-stage model is that application of a low dose of a mutagen to a benign tumor will greatly increase malignant conversion of the tumor [so-called initiation–promotion–initiation (*6*)]. Experiments have confirmed this prediction (*55*). It must be stressed that agents that increase proliferation of normal or intermediate tissue will increase cancer risk apart from any direct effect on DNA. Such proliferation stimulating agents (promoters are one class) may account for a greater proportion of cancer cases than strictly genotoxic agents in the environment (*56*).

The darkness of the arrows in Figure 1 provides a very rough sense of the relative probabilities of the respective pathways leading to cancer. There is growing evidence that cancer arises from the malignant conversion of a single cell (*57*). If this is true, then although the probability of transformation of a particular cell is extremely low, the probability that at least one cell of a tissue becomes transformed is much higher. The darkest arrow leads from a normal cell to a normal tissue since this is the normal process. The probability that at least one normal cell will become malignant at a single cell division cycle is the product of μ_1 and μ_2. Typical mutation rates are approximately 10^{-7} per cell per division (*58*), so that even with approximately 10^{10} cells in a given tissue, the chance of a cancer cell arising is very low. However, the chance that an intermediate cell will arise is not so low. Intermediate cells divide as do their normal counterparts, and they may have a growth advantage, as when a promoter is applied and an intermediate lesion appears. Within an intermediate lesion, the chance that a malignant cell will arise depends on the second mutation rate and the rate of division of the cells of the lesion.

Some hereditary cancer syndromes can be equated to a germ-line mutation with inheritance of all cells in the intermediate stage. Retinoblastoma is the model for this growing list of cancers (*59*). The Rb gene confers a 100,000-fold increased risk for retinoblastoma in those who inherit one absent or defective homologue. Whereas the probability that a particular cell will suffer the second mutation and become malignant is very low (and therefore the syndrome is recessive at the cellular level), the probability that at least one cell of the tissue will suffer the second

Table 1. Speculative table of how bioeffects of electromagnetic fields might be related to cancer site, time of relevant exposure, cofactors, and confounders.

	Site	Time[a]	Cofactors (effect modifiers)	Confounders
Transcription/translation		Promotion		Toxic chemicals
Calcium	Acute nonlymphocytic leukemia, oxygenated tissue (e.g., lung)	Promotion	Ionizing radiation; body iron stores	Free radical producing toxic chemicals
Ornithine decarboxylase		Promotion	Chemical promoters	
Pineal				Shift work, alcohol, light-at-night
Cancer cells	Estrogen receptor breast	Progression		
Stem cells	All breast/prostate	Prior to initiation		
Immune	Non-Hodgkin's lymphoma	Progression		

[a]When time is prior to initiation, then exposures in the distant past are likely to be important; when time is progression, then very recent exposures, perhaps even less than 1 year, are likely to be important.

mutation over the life of the individual is extremely high, approaching one (and therefore dominant at the level of the tissue).

Epidemiological Study Considerations

Consideration of biological effects of EMF suggests particular design features for epidemiologic studies (Table 1). Within the context of the two-stage model described above, temporal sequence of exposure, interaction with other agents, and confounding factors depend on how EMF is hypothesized to affect cancer risk.

The lack of direct effects on mutation suggests that if EMF increases cancer risk, it is not by increasing μ_1. However, if the accurate functioning of mitosis is affected, EMF might indirectly lead to mitotic recombination, and the fixation of an antioncogene if one is involved; an intermediate cell may yield a normal and a malignant progeny. In this way, μ_2 may be affected and not μ_1.

Time of Exposure

Tamarkin et al. (*50*) investigated the effect of melatonin on chemically-induced mammary cancer in rats. Sixty female Sprague- Dawley rats received 15 mg of dimethylbenzanthracene (DMBA) in peanut oil by intragastric intubation at age 50 days. Following the DMBA administration, 30 rats received daily injections of melatonin and 30 received vehicle injection. Ninety days after DMBA administration, 50% of the vehicle-treated rats had developed tumors, whereas none of the melatonin-treated group had tumors. Melatonin was discontinued in the latter group at this time, and tumors began to appear later. Their next experiment examined the effect of reducing melatonin. Thirty-six pinealectomized and 36 sham-operated rats received 7 mg of DMBA administered as before. Two months after DMBA was administered, 48% of the pinealectomized rats had mammary tumors, whereas none of the sham group had tumors. By day 240 (termination of the experiment), 88% of the pinealectomized rats had tumors, whereas only 22% of the sham group had tumors. These experimental observations showed that melatonin suppresses mammary tumorigenesis in rats, and lack of melatonin increases tumor formation.

Despite the animal and *in vitro* evidence, it is not clear that alterations in melatonin affect risk of breast cancer in humans. However, if suppression of melatonin does raise human breast cancer risk, then the mechanism by which it acts has important implications for the conduct of epidemiological studies of electric power. As shown in Figure 2, if a mechanism of action is by virtue of a general stimulation of estrogen production, then the increased turnover of the normal breast epithelial stem cells at risk could increase cancer risk. Past exposures to agents such as EMF or light-at-night (LAN) that might suppress melatonin would be important. This can be viewed in Figure 1 as EMF increasing A, the growth of normal cells. Exposures as early as puberty might be crucial. EMF exposure many years in the past should be assessed if stem cell turnover is thought to be affected.

However, if suppressed melatonin production increased breast cancer risk by virtue of releasing estrogen-receptor positive (ER+) breast cancer cells from a quiescent state [i.e., progression (*60*)], then very recent exposures could be crucial. This would be an increase in C in the two-stage model. Similarly, direct effects of EMF on immune cell function (*27*) also might increase C. Very recent EMF exposures should be assessed if cancer-cell growth is thought to be affected (*61*). Perhaps even exposure within one year of diagnosis should be assessed.

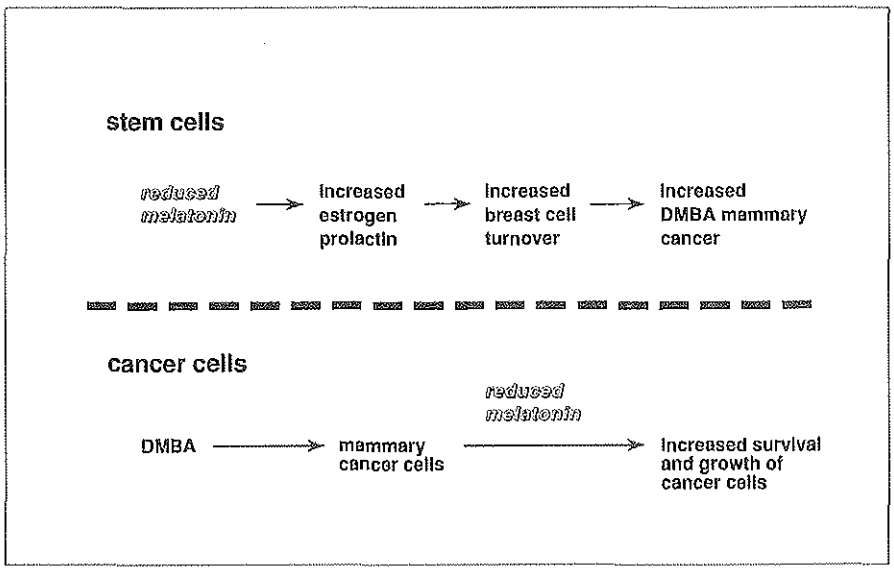


Figure 2. Two possible mechanistic explanations for the observation of Tamarkin et al. (*52*). In the first, reduced melatonin leads to an increase in estrogen and/or prolactin that leads to an increase in the breast epithelial stem cells at risk. This results in increased DMBA-induced cancer cell production. The second mechanism postulates reduced melatonin leading to the release of existing DMBA-induced cancer cells from their quiescent state. Melatonin also may affect the immune function.

Effects of EMF on calcium balance that might influence promotion via increased oxidative stress or disrupted signal transduction might affect the growth of intermediate cells, B. In studies of these possible mechanisms, time of exposure, cofactors, and cancer types are all influenced (see below). Similarly, effects on ornithine decarboxylase (ODC) might affect B. Effects on promotion predict a time frame for relevant exposure that lies between the distant past and the very recent past, perhaps 2 to 10 years prior to cancer diagnosis.

Cancer Types

Certain leukocytes, such as neutrophils and macrophages, generate oxygen radicals in order to kill foreign cells (*62*). In contrast, cell-mediated killing by lymphocytes (e.g., cytotoxic T-lymphocytes) appears to depend not on oxidative bursts but rather on the calcium-dependent release of a pore-forming protein that perforates the membrane of the target cell (*63*). On this basis, nonlymphocytes that use oxidative bursts for their function may be more susceptible to the disruption of their own oxidative defence mechanisms than other cell types. Intracellular calcium concentration and distribution modulate cellular oxidative activity, degranulation, phagocytosis, and mobility (*64*). Therefore, EMF-induced disruption of calcium balance and increased oxidative stress may affect these nonlymphocytes more than other hematopoietic tissues (*65*). The observations of increased nonlymphocytic leukemia in occupational studies of EMF and cancer (*66*) are consistent with this possibility. A study of residential EMF exposure and acute nonlymphocytic leukemia (ANLL) did not find evidence for an association (*67*); however, occupational exposures were not assessed. Although carefully done, this study was small and does not, by itself, offer strong evidence against a role for EMF in adult leukemia.

Calcium is an important second messenger for gene expression; Morgan and Curran (*68*) reported that c-*fos* protooncogene expression in PC12 cells is induced by either receptor–ligand interaction or by alterations of voltage-dependent calcium channels. There is a report that exposure of human polymorphonuclear leukocytes to a static magnetic field of 0.1 T dramatically increased degranulation in a time-dependent manner; the effect was inhibited by calcium-channel antagonists (*69*). A case–control study of acute nonlymphocytic leukemia in adults that takes account of both residential and occupational exposures should be performed.

There have been three biological effects of melatonin investigated that might influence cancer risk (*2*) and account for the experimental observations of Tamarkin et al. (*50*), which showed melatonin inhibited DMBA-induced mammary carcinogenesis in rats. There is evidence that melatonin can *a*) stop the growth of hormone-dependent cancer cells, *b*) suppress production of sex hormones (e.g., estrogen, testosterone), and *c*) augment immune function. These three mechanisms have different implications for cancer types that might be affected (Fig. 3). If EMF or LAN suppresses melatonin and the first mechanism applies, then risk of hormone-dependent tumors could be increased (e.g., ER+ breast cancer). If the second mechanism applies, then risk of cancers of tissues that are dependent on hormones for growth could be increased (e.g., all tumors of breast and prostate). Compromise of immune function was long assumed to increase risk for cancer in general and for all sites. However, recent evidence shows a clear and large increased risk only of non-Hodgkin's lymphoma, particularly of the brain. There also may be an elevation of some mesenchymal tumors and perhaps melanoma. It is significant to this discussion that the evidence is against a role for immunologic suppression in the etiology of most common cancers such as breast and colon cancer (*70*).

Among 370 patients diagnosed with malignant melanoma who were followed prospectively, a second primary cancer of breast later developed at a rate six times that expected (*71*), which suggests a common etiology for melanoma and breast cancer that may involve reduced melatonin. Melatonin has been reported to slow the growth of transplanted melanoma cells in athymic mice (*72*). Thus, studies of breast cancer and of malignant melanoma in females and in males and of prostate cancer in males should be performed.

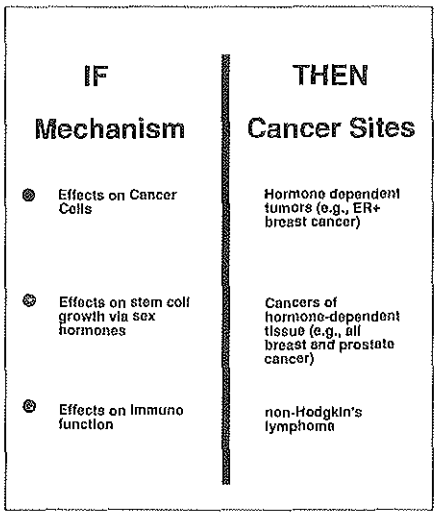


Figure 3. The cancer types that might be influenced by agents that suppress melatonin depend on melatonin's action. Thus, both the timing of exposure and the cancer site chosen for epidemiological study depend on the postulated mechanism of action.

Cofactors (Effect Modifiers)

Calcium is important in many aspects of cellular physiology, and it plays an important role as an intracellular messenger for the activation of genes and in intercellular communication. Calcium homeostasis also is important for protection from oxidative stress (*73,74*). EMF-induced disruptions of calcium balance that might lead to increases in free radicals may inhibit a cell's ability to protect itself from some other oxidative attack such as a toxic chemical or ionizing radiation. Many tumor promoters are agents that increase radical production in cells (*75*); this may then provide a mechanism for EMF promotion or copromotion. A proposed test of whether EMF can increase oxidative stress is based on a hepatocyte toxicity assay. It was long speculated that the final event in the death of hepatocytes after exposure to toxins was the influx of extracellular calcium. However, Smith et al. (*76*) reported that incubation of freshly isolated hepatocytes in a calcium-free medium greatly increased their susceptibility to damage from carbon tetrachloride and other hepatotoxins. The calcium-free medium was not itself toxic in the absence of the chemical toxins. Reed and Fariss (*77*) speculated about these results and suggested that the calcium-free medium led to a disruption of intracellular calcium, which led to increased oxidative stress and susceptibility to the toxins. If EMF can disrupt intracellular calcium balance and lead to increased oxidative stress, then EMF may increase the toxicity of carbon tetrachloride to freshly isolated hepatocytes (*65*).

Subtle effects of EMF on calcium homeostasis that might lead to an increase in oxidative stress could go entirely undetected unless they occurred in conjunction with another agent. By this mechanism, EMF may stress a cell not to the point of damage but to the point of increasing susceptibility to other agents. This reasoning leads to the speculation that EMF, under some circumstances, might be a radiosensitizer by increasing oxidative stress and reducing the cellular compliment of reducing equivalents (*65*).

Increases in oxidative stress could promote or copromote by increasing B, the growth rate of intermediate cells. Balcer-Kubiczek and Harrison (*78*) used a transformation assay of C3H/10T1/2 cells to determine the interactions of an X-ray, TPA, and a 2.45-GHz microwave pulse modulated at 120 Hz. They found that the microwave alone had no effect and did not increase the effect of the X-ray. However, the microwave

significantly increased transformation when used in combination with the X-ray and TPA as compared to the X-ray plus TPA without microwave. In addition, microwave and TPA together increased transformation in the absence of the X-ray; neither increased transformation alone. Stuchly et al. (*79*) have reported copromotion by magnetic fields in the mouse skin.

Adey (*80,81*) suggested a synergism of EMF with chemical tumor promoters that leads to autonomous cell growth via disruptions of intercellular communication. Recent data in support of this speculation (*82*) showed that intermittent exposure to a 1-G, 60-Hz magnetic field significantly increased the expression of transformation by TPA of C3H/10T1/2 fibroblasts.

Another possible cofactor for oxidative stress resulting from effects on calcium balance is body iron stores. There is evidence that high body iron stores increase cancer risk (*83*), and the role of iron in catalyzing oxygen radicals is one possible mechanism (*65*). Phillips et al. (*84*) reported that exposure to a 60-Hz magnetic field or to a combined electric and magnetic field produced constitutive expression of transferrin receptors on human colon cancer cells *in vitro*.

Finally, EMF-induced loss of iron from its intracellular storage protein, ferritin, might increase oxidative stress. Therefore, higher iron might increase the effect of any EMF increases in oxidative stress due to disruption of calcium, and, in contrast, EMF itself may increase reactive iron availability within the cell and cause further oxidative stress. There is a need for a study of EMF effects on susceptibility to radiation-induced cancer and cancer induced by chemicals that increase oxidative stress. Body iron stores also should be assessed in these studies.

Confounding

The group of confounders that are based on increased oxidative stress from disruption of calcium balance would include chemicals associated with occupational exposure to EMF. These would be chemicals that are known or suspected to derive their toxicity in whole or in part from generation of free radicals. These chemicals may also be cofactors (effect modifiers) as opposed to confounders.

In studies of leukemia, detailed information on exposure to agents that might increase oxidative stress should be gathered. If possible, biomarkers of exposure to these agents should be used in these studies.

If EMF is thought to increase cancer risk via an effect on pineal function, then possible confounders include LAN, shift work, alcohol consumption, and any other factor that has been shown to affect pineal function. These factors should be taken into account in epidemiological studies of EMF.

Also, in studies of breast and prostate cancers, other agents that affect pineal function should be assessed.

Future Directions

Epidemiology can be conducted fruitfully in the absence of a biological rationale. Without much understanding of what was bad about cigarette smoke, early studies of smoking and lung cancer made great contributions to understanding people's health and eventually improving it. However, a biological rationale can aid in designing epidemiological studies, particularly in an area such as electric power and cancer, because the design is so challenging. In studies of smoking, it is simple and relatively inexpensive to gather data, because the answer to a question can yield good exposure information. There are also biomarkers of recent exposure to cigarette smoke. However, it is very difficult to assess exposure to EMF (even if there were consensus on what constitutes a relevant exposure); questionnaires are of limited use, and there is no identifiable biomarker of exposure. Therefore, biological considerations enable focused epidemiological studies of specific hypotheses.

There is a place for both epidemiological and laboratory studies in most research programs into the causes of cancer; there should be synergy between the two. Epidemiology can address directly the question of whether increased risk of cancer is associated with some aspect of the human environment such as exposure to EMF. These studies are necessarily crude and can rarely determine precisely what component of the exposure is the culprit because pure, single-agent exposures in human populations are virtually nonexistent. EMF is too broad a definition of exposure in epidemiological studies to be very helpful in mitigating exposure, and there is no doubt that electricity will continue to be used whether or not a consensus eventually emerges from epidemiology that EMF increases cancer risk. However, laboratory studies can address what can be done about a problem of increased cancer risk by isolating what component of exposure causes the problem. If laboratory studies could isolate successfully a particular feature of exposure to EMF that accounted for an increased risk of cancer, then mitigation of exposure to that feature might be feasible. The reverse also is true: Laboratory studies can identify previously unsuspected cancer-causing agents that should be investigated epidemiologically.

Biological considerations for epidemiological studies are illustrated in Table 1. This table is speculative and selective. It is not intended to exclude any ideas on possible EMF studies. Rather, it is intended to provide an example of the kind of reasoning that might go from biological rationale to epidemiological design.

Suggested new avenues for study include hormone-dependent cancers and cancers of hormone-dependent tissues, increased susceptibility to radiation-induced cancer, and increased susceptibility to free-radical producing toxic chemicals.

REFERENCES

1. Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. J Natl Cancer Inst 66:1191–1308 (1981).
2. Stevens RG, Davis S, Thomas DB, Anderson LE, Wilson BW. Electric power, pineal function, and the risk of breast cancer. FASEB J 6:853–860 (1992).
3. Wertheimer N, Leeper ED. Electrical wiring configurations and childhood cancer. Am J Epidemiol 109:273–284 (1979).
4. Wertheimer N, Leeper ED. Adult cancer related to electrical wires near the home. Int J Epidemiol 11:345–355 (1982).
5. Nair I, Morgan MG, Florig HK. Biological effects of power frequency electric and magnetic fields—background paper. Washington, DC: U.S. Government Printing Office, 1989.
6. Moolgavkar SH, Knudson AG. Mutation and cancer: a model for human carcinogenesis. J Natl Cancer Inst 66:1037–1052 (1981).
7. Tenforde TS, Kaune WT. Interaction of extremely low frequency electric and magnetic fields with humans. Health Phys 53:585–606 (1987).
8. Blank M. Theory of frequency-dependent ion concentration changes in oscillating electric fields. J Electrochem Soc 134:1112–1117 (1987).
9. Weaver JC, Astumian RD. The response of living cells to very weak electric fields: the thermal noise limit. Science 247:459–462 (1990).
10. Liboff AR, McLeod BR, Smith SD. Ion cyclotron resonance effects of ELF fields in biological systems. In: Extremely low frequency electromagnetic fields: the question of cancer (Wilson BW, Stevens RG, Anderson LE, eds). Columbus, OH: Battelle Press, 1990; 251–289.

11. Reese JA, Jostes RF, Frazier ME. Exposure of mammalian cells to 60-Hz magnetic or electric fields: analysis for DNA single-strand breaks. Bioelectromagnetics 9:237–247 (1988).
12. Frazier ME, Reese JA, Morris JE, Jostes RF, Miller DL. Exposure of mammalian cells to 60-Hz magnetic or electric fields: analysis for DNA repair of induced, single-strand breaks. Bioelectromagnetics 11:229–234 (1990).
13. Nahas SM, Oraby HA. Micronuclei formation in somatic cells of mice exposed to 50-Hz electric fields. Environ Mol Mutagen 13:107–111 (1989).
14. Goodman R, Bassett CAL, Henderson AS. Pulsing electromagnetic fields induce cellular transcription. Science 220:1283–1285 (1983).
15. Goodman R, Shirley–Henderson A. Exposure of cells to extremely low frequency electromagnetic fields: relationship to malignancy? Cancer Cells 2:355–359 (1990).
16. Goodman R, Abbott J, Henderson AS. Transcriptional patterns in the X chromosome of *Sciara coprophila* following exposure to magnetic fields. Bioelectromagnetics 8:1–7 (1987).
17. Blank M, Goodman R. New and missing proteins in the electromagnetic and thermal stimulation of biosynthesis. Bioelec Bioenerg 21:307–317 (1989).
18. Wagner S, Green MR. A transcriptional tryst. Nature 352:189–190 (1991).
19. Bawin SM, Adey WR. Sensitivity of calcium binding in cerebral tissue to weak electric fields oscillating at low frequency. Proc Natl Acad Sci USA 73:1999–2003 (1976).
20. Blank M, Soo L. Ion activation of the Na,K-ATPase in alternating currents. Bioelec Bioenerg 24:51–61 (1990).
21. Blackman CF. ELF effects on calcium homeostasis. In: Extremely low frequency electromagnetic fields: the question of cancer (Wilson BW, Stevens RG, Anderson LE, eds). Columbus, OH: Battelle Press, 1990; 187–208.
22. Smith SD, McLeod BR, Liboff AR, Cooksey K. Calcium cyclotron resonance and diatom mobility. Bioelectromagnetics 8:215–227 (1987).
23. Reese JA, Frazier ME, Morris JE, Buschbom RL, Miller DL. Evaluation of changes in diatom mobility after exposure to 16-Hz electromagnetic fields. Bioelectromagnetics 12:21–25 (1991).
24. Luben RA, Cain CD, Chen M, Rosen DM, Adey WR. Effects of electromagnetic stimuli on bone and bone cells in vitro: inhibition of responses to parathyroid hormone by low-energy, low-frequency fields. Proc Natl Acad Sci (USA) 79:4180–4184 (1982).
25. Lyle DB, Wang X, Ayotte RD, Sheppard AR, Adey WR. Calcium uptake by leukemic and normal T-lymphocytes exposed to low frequency magnetic fields. Bioelectromagnetics 12:145–156 (1991).
26. Byus CV, Pieper SE, Adey WR. The effects of low-energy 60-Hz environmental electromagnetic fields upon the growth-related enzyme ornithine decarboxylase. Carcinogenesis 8:1385–1389 (1987).
27. Lyle DB, Schechter P, Adey WR, Lundak RL. Suppression of T-lymphocyte cytotoxicity following exposure to sinusoidally amplitude-modulated fields. Bioelectromagnetics 4:281–292 (1983).
28. Lyle DB, Ayotte RD, Sheppard AR, Adey WR. Suppression of T-lymphocyte cytotoxicity following exposure to 60-Hz sinusoidal electric fields. Bioelectromagnetics 9:303–313 (1988).
29. Morris JE, Phillips RD. Effects of 60-Hz electric fields on specific humoral and cellular components of the immune system. Bioelectromagnetics 3:341–347 (1982).
30. Davis RL, Milham S. Altered immune status in aluminum reduction plant workers. Am J Ind Med 18:79–85 (1990).
31. Garaj-Vrhovac V, Fucic A, Horvat D. Comparison of chromosome aberration and micronucleus induction in human lymphocytes after occupational exposure to vinyl chloride monomer and microwave radiation. Periodicum Biologorum 92:411–416 (1990).
32. Wurtman RJ, Axelrod J. The pineal gland. Sci Amer 213:50–60 (1965).
33. Reiter RJ, ed. The pineal gland, vols I, I, III. Boca Raton, FL: CRC Press, 1981.
34. Reiter RJ. Effects of light and stress on pineal function. In: Extremely low frequency electromagnetic fields: the question of cancer (Wilson BW, Stevens RG, Anderson LE, eds). Columbus, OH: Battelle Press, 1990; 87–107
35. Tamarkin L, Baird CJ, Almeida OF. Melatonin: a coordinating signal for mammalian reproduction? Science 227:714–720 (1985).
36. Cohen M, Lippmann M, Chabner B. Role of the pineal gland in the aetiology and treatment of breast cancer. Lancet 2:814–816 (1978).
37. MacMahon B, Cole P, Brown J. Etiology of human breast cancer: a review. J Natl Cancer Inst 50:21–42 (1973).
38. Moolgavkar SH, Day NE, Stevens RG. Two-stage model for carcinogenesis: epidemiology of breast cancer in females. J Natl Cancer Inst 65:559–569 (1980).
39. Wilson BW, Anderson LE, Hilton DI, Phillips RD. Chronic exposure to 60-Hz electric fields: effects on pineal function in the rat. Bioelectromagnetics 2:371–380 (1981); 4:293 (1983).
40. Welker HA, Semm P, Willig RP, Commentz JC, Wiltschko W, Vollrath L. Effects of an artificial magnetic field on seratonin N-acetyltransferase activity and melatonin content of the rat pineal gland. Exp Brain Res 50:426–432 (1983).
41. Reuss S, Olcese J. Magnetic field effects on the rat pineal gland: role of retinal activation by light. Neuroscience Lett 64:97–101 (1986).
42. Reiter RJ, Anderson LE, Buschbom RL, Wilson BW. Reduction of the nocturnal rise in pineal melatonin levels in rats exposed to 60-Hz electric fields in utero and for 23 days after birth. Life Sciences 42:2203–2206 (1988).
43. Lerchl A, Nonaka KO, Stokkan KA, Reiter RJ. Marked rapid alterations in nocturnal pineal serotonin metabolism in mice and rats exposed to weak intermittent magnetic fields. Biochem Biophys Res Comm 169:102–108 (1990).
44. Lerchl A, Nonaka KO, Reiter RJ. Pineal gland: its apparent magnetosensitivity to static magnetic fields is a consequence of induced electric currents (eddy currents). J Pineal Res 10:109–116 (1991).
45. Kato M, Honma K, Shigemitsu T, Shiga, Y. Effects of exposure to a circularly polarized 40-Hz magnetic field on plasma and pineal melatonin levels in rats. Bioelectromagnetics 14:97–106 (1993).
46. Stevens RG. Review and Commentary: electric power use and breast cancer: a hypothesis. Am J Epidemiol 125:556–561 (1987).
47. Stevens RG. Electric power, melatonin, and breast cancer. In: The pineal gland and cancer (Gupta D, Attanasio A, Reiter RJ, eds). London: Brain Research Promotion, 1988; 233–244.
48. Philo R, Berkowitz AS. Inhibition of Dunning tumor growth by melatonin. J Urol 139:1099–1102 (1988).
49. Hill SM, Blask DE. Effects of the pineal hormone melatonin on the proliferation and morphological characteristics of human breast cancer cells (MCF-7) in culture. Cancer Res 48:6121–6 (1988).
50. Tamarkin L, Cohen M, Roselle D, Reichert C, Lippman M, Chabner B. Melatonin inhibition and pinealectomy enhancement of 7,12 dimethylbenz(a)anthracene-induced mammary tumors in the rat. Cancer Res 41:4432–4436 (1981).
51. Reiter RJ. Melatonin: the chemical expression of darkness. Mole Cell Endocrinol 79:C153–C158 (1991).
52. Beniashvili DS, Bilanishvili VG, Henabde MZ. Low-frequency electromagnetic radiation enhances the induction of rat mammary tumors by nitrosomethyl urea. Cancer Letters 61:75–79 (1991).
53. Mevissen M, Stamm A, Buntenkötter S, Zwingelberg R, Wahnschaffe U, Löscher W. Effects of magnetic fields on mammary tumor development induced by 7,12–dimethylbenz(a)anthracene in rats. Bioelectromagnetics 14:131–143 (1993).
54. Stevens RG, Wilson BW, Anderson LE. The question of cancer. In: Extremely low frequency electromagnetic fields: the question of cancer (Wilson BW, Stevens RG, Anderson LE, eds). Columbus, OH: Battelle Press, 1990; 361–370.
55. Hennings H, Shores R, Wenk ML, Spangler EF, Tarone R, Yuspa SH. Malignant conversion of mouse skin tumors is increased by tumor initiators and unaffected by tumor promoters. Nature 304:67–69 (1983).
56. Pitot HC, Dragan YP. Facts and theories concerning the mechanisms of carcinogenesis. FASEB J 4:2280–2286 (1991).
57. Wainscoat JS, Fey MF. Assessment of clonality in human tumors: a review. Cancer Res 50:1355–1360 (1990).
58. Oller AR, Rastogi P, Morgenthaler S, Thilly WG. A statistical model to estimate variance in long term–low dose mutation assays: testing of the model in a human lymphoblastoid mutation assay. Mutation Res 216:149–161 (1989).
59. Knudson AG. Hereditary cancer, oncogenes, and antioncogenes. Cancer Res 45:1437–1443 (1985).
60. Blask DE. The emerging role of the pineal gland and melatonin in oncogenesis. In: Extremely low frequency electromagnetic fields: the question of cancer (Wilson BW, Stevens RG, Anderson LE, eds). Columbus, OH: Battelle Press, 1990; 319–335.
61. Liburdy RP, Sloma TR, Sokolic R, Yaswen P. ELF magnetic fields, breast cancer, and melatonin: 60 Hz fields block melatonin's oncostatic action on ER+ breast cancer cell proliferation. J Pineal Res 14:89–97 (1993).

62. Johnston RB, Keele BB, Misra HP, Lehmeyer JE, Webb LS, Baehner RL, Rajagopalan KV. The role of superoxide anion generation in phagocytic bactericidal activity. J Clin Invest 55:1357–1372 (1975).
63. Young JD, Cohn ZA. Cell–mediated killing: a common mechanism? Cell 46:64–642 (1986).
64. Sawyer DW, Sullivan JA, Mandell GL. Intracellular free calcium localization in neutrophils during phagocytosis. Science 230:663–666 (1985).
65. Stevens RG, Kalkwarf DR. Iron, radiation, and cancer. Environ Health Perspect 87:291–300 (1990).
66. Savitz DA, Calle EE. Leukemia and occupational exposure to electromagnetic fields: review of epidemiologic surveys. J Occup Med 29:47–51 (1987).
67. Severson RK, Stevens RG, Kaune WT, Thomas DB, Heuser L, Davis S, Sever LE. Acute nonlymphocytic leukemia and residential exposure to power frequency magnetic fields. Am J Epidemiol 128:10–20 (1988).
68. Morgan JI, Curran T. Role of ion flux in the control of *c-fos* expression. Nature 322:552–555 (1986).
69. Papatheofanis FJ. Use of calcium channel antagonists as magnetoprotective agents. Radiat Res 122:24–28 (1990).
70. Kinlen LJ. Immunologic factors. In: Cancer epidemiology and prevention (Schottenfeld, D, Fraumeni JF, eds). Phildelphia, PA: W. B. Saunders Co., 1982; 494–505.
71. Gutman M, Cnaan A, Inbar M, Shafir R, Chaitchik S, Rozin RR, Klausner JM. Are malignant melanoma patients at higher risk for a second cancer? Cancer 68:660–665 (1991).
72. Narita T, Kudo H. Effect of melatonin on B16 melanoma growth in athymic mice. Cancer Res 45:4175–4177 (1985).
73. Pascoe GA. Calcium homeostasis and oxidative stress. In: Extremely low frequency electromagnetic fields: the question of cancer (Wilson BW, Stevens RG, Anderson LE, eds). Columbus, OH: Battelle Press, 1990; 337–360.
74. Nicotera P, Bellomo G, Orrenius S. The role of Ca^{2+} in cell killing. Chem Res Toxicol 3:484–494 (1990).
75. Cerutti P. Prooxidant states and tumor promotion. Science 227:375–381 (1985).
76. Smith MT, Thor H, Orrenius S. Toxic injury to isolated hepatocytes is not dependent on extracellular calcium. Science 213:1257–1259 (1981).
77. Reed DJ, Fariss MW. Glutathione depletion and susceptibility. Pharm Rev 36:25S–33S (1984).
78. Balcer–Kubiczek EK, Harrison GH. Neoplastic transformation of C3H/10T1/2 cells following exposure to 120-Hz modulated 2.45-GHz microwaves and phorbol ester tumor promoter. Radiat Res 126:65–72 (1991).
79. Stuchly MA, McLean JRN, Burnett R, Goodard M, Mitchel REJ, Lecuyer DW. Modification of tumor promotion in the mouse skin by exposure to an alternating magnetic field. Cancer Letters 65:1–7 (1992).
80. Adey WR. Joint actions of environmental nonionizing electromagnetic fields and chemical pollution in cancer promotion. Environ Health Perspect 86:297–305 (1990).
81. Adey WR. Electromagnetic fields, cell membrane amplification, and cancer promotion. In: Extremely low frequency electromagnetic fields: the question of cancer (Wilson BW, Stevens RG, Anderson LE, eds). Columbus, OH: Battelle Press, 1990; 211–249.
82. Cain CD, Adey WR. 60-Hz magnetic fields enhance tumor promoter-induced focus formation in C3H10T1/2 fibroblast coculture system. Department of Energy's Annual Research Review, Milwaukee, WI, November 3–7, 1991.
83. Stevens RG, Jones DY, Micozzi MS, Taylor PR. Body iron stores and the risk of cancer. New Engl J Med 319:1047–1052 (1988).
84. Phillips JL, Rutledge L, Winters WD. Transferrin binding to two human colon carcinoma cell lines: characterization and effect of 60-Hz electromagnetic fields. Cancer Res 46:239–244 (1986).

Neurobehavioral Effects of Power-Frequency Electromagnetic Fields

by Nigel Paneth

Program in Epidemiology, A206 E. Fee Hall, Michigan State University, East Lansing, MI 48824

Some laboratory experiments have suggested that power-frequency electric and magnetic fields (EMF) may be capable of influencing calcium efflux from cell membranes, pineal function, and circadian rhythms. As yet, however, no consistent, replicable laboratory model has been developed for any of these effects. Most assessments of human volunteers exposed to EMF have been negative, but occasional effects on vigilance or alertness and some modest effects on circadian rhythmicity have been reported. Several carefully performed studies of workers occupationally exposed to high electric-field strengths have failed to find effects on behavior or cognitive functioning. Although the bulk of human research on the effects of EMF on cognitive performance is negative, there has been less assessment of behavior and psychiatric symptomatology. Because some studies, in both humans and animals, have described effects of EMF on circadian rhythms, future research might concentrate profitably on the assessment of EMF in relation to depression and other cyclically mediated psychiatric disorders. — Environ Health Perspect 101(Suppl 4):101–106 (1993).

Key Words: Electromagnetic fields, 60-Hz fields, cognitive testing, behavior, suicide

Introduction

Assessment of the relationship between exposure to electromagnetic fields (EMF) and behavior and cognition in humans and animals is especially difficult. Three methodologic problems stand out. The first, which applies for the most part to animal studies, is that laboratory EMF exposure is not easily separated from concomitant exposures, such as noise, vibration, hair stimulation, and even mild electric shocks. When the outcome of interest is cancer or birth defects, these secondary exposures are unlikely to be confounding; but when changes in motor activity or circadian rhythm are found, or animals appear averse to an exposed location, the role of these other potentially noxious stimuli must be considered.

The second difficulty, a feature of both animal and human studies, is the subjective and transient nature of many behavioral and cognitive measures. Although mismeasurement can occur with any human health outcome, the condition most studied in relation to EMF, cancer, has the virtue of being based, in most cases, on tissue diagnosis. Characteristics of behavioral and cognitive measures, however, are their dependence on the specific conditions under which they are obtained, the cooperativeness of the subject, and the skills of the examiner.

The third difficulty, specific to human studies, is that many measures of cognition and behavior are correlated powerfully with socioeconomic status and educational background. Teasing out the impact of these factors can prove a challenge to investigators. These concerns should be kept in mind as one reviews the specific studies below.

A remarkable amount of the research in this field is not found in the peer-reviewed literature but in the form of reports to government agencies. Not only does this make retrieval of the findings difficult, but it indicates that many of the results described may not have been subjected to adequate peer review.

Suggestions from the Laboratory

A wide variety of experiments has been performed to investigate the effects of EMF on several cellular, subcellular, and whole animal preparations. Major differences in experimental set-up, in frequency and intensity of EMF exposure, and in choice of end points characterize the literature as a whole. Although many findings have been reported, the vast majority have not been replicated consistently, partly because of the heterogeneity of the experimental models currently in use. So far, no single experimental model can be pointed to as a consistent biological marker of EMF exposure. In searching for clues for the more effective framing of epidemiological research, this paper concentrates on those laboratory findings that have been at least partly replicated in more than one laboratory.

Calcium Efflux

Among the few replicated laboratory findings in this field is the effect of 60-Hz electric-field exposure on the efflux of calcium across brain tissue cell membranes (*1–4*). Because calcium efflux serves as a messenger of electrochemical signals in the brain, it could be the basis for effects of EMF on behavior in the whole animal, although it is uncertain precisely what dysfunction would be predicted by this biochemical disturbance. Lovely has suggested a possible effect on memory skills (*5*).

Although the calcium efflux effect has been replicated, its direction has not. Bawin et al. found, in chick brain tissue, that calcium efflux decreased with EMF exposure (*1*), but Blackman et al. found, in the same preparation, an increase (*2,3*). Another surprising finding was an absence of gradient of risk with gradient of exposure; rather, only specific combinations of frequency and intensity produced the effects described. Moreover, these effect windows were not identical in the two sets of experiments. For Bawin et al., effects were seen at 10 V/m with frequencies of 6 and 16 Hz; for Blackman et al., an intensity window at 60 Hz was found for 35 to 43 V/m. Inasmuch as these precise effect windows were not hypothesized prior to the experiment, one cannot rule out the role of chance in their specification.

Blackman, in replicated experiments, also found that the calcium efflux effect induced by EMF exposure was sensitive to prenatal (egg) exposure to 50- and 60-Hz fields at a variety of intensities (*6*). Gunderson, working with chick spinal cord (as contrasted to chick brain), failed to find effects of EMF exposure on calcium efflux (*4*).

Pineal Function and Circadian Rhythms

Two studies have noted depression in nighttime melatonin production in rats (*7,8*). Related changes seen in single investigations include period shifts in norepinephrine, serotonin, and dopamine secretion (*9*), and reductions in cerebrospinal

This manuscript was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990 – September 1992.

fluid and 5-hydroxyindole acetic acid without period shifts (*10*). These studies, which point to EMF effects on the pineal gland or the suprachiasmatic nucleus, are paralleled by laboratory findings of the effect of EMF exposure on circadian rhythms.

Dowse and Palmer (*11*) and Ehret (*12*) were able to show either entrainment of mice, phase delays, or other changes in metabolic rhythms with electric-field exposure. The Ehret studies used very high field strength (130 kV/m), which may limit the importance of their findings. The Dowse and Palmer studies are among those that have been criticized for lack of singularity of the experimental exposure; Roberts (*13*) has argued that noise and corona discharges might account for the findings. Dowse (*14*) later showed phase changes in the locomotor activity of drosophila. Ehret and co-workers' success in entraining mice rhythms was not paralleled by their experiments in rats (*15,16*), whose body temperature, activity, and food intake cycles were not affected by exposures of 4.5 to 55 kV/m at 60 Hz.

Sulzman and Murrish (*17*) showed some effects on the circadian rhythmicity of food intake and oxygen consumption in squirrel monkeys with fairly high exposures (26–39 kV/m), but not all exposed animals showed the effect. The circadian cycle of fungal spore formation, however, seems unperturbed by magnetic fields as high as 32 G (*18*).

Although not a study of rhythmicity, the work of Thomas et al. (*19*) has implications for human cyclical disorders, such as bipolar depression. Four of five rats exposed to a combined 60-Hz magnetic field and magnetostatic field of 26 μT showed consistent, large increases in their responses to a differential reinforcement of low rate schedule in which a food pellet resulted only if the rat pressed a lever twice 18 to 24 sec apart. Because the exposure corresponded to the cyclotron resonance condition for lithium ions, the authors speculated that the effect might be related to efflux of lithium ions in brain cells, paralleling the changes that might be occurring with lithium carbonate treatment of bipolar disorders.

Human Studies

Three kinds of investigations in humans have been reported: *a*) neurobehavioral testing or assessment of experimentally exposed volunteers, *b*) assessments of occupationally exposed workers, *c*) correlational studies of EMF residential exposure with suicide.

Neurobehavioral Effects in Volunteers

Human volunteers exposed to EMF have been studied for changes in circadian rhythmicity, verbal reasoning, attention skills, mood, and perception of electric fields. The literature is modest, and several reports are not in the peer-reviewed press [e.g., Johansson et al. (*29*) and Rupilius (*30*)—cited by Stollery (*20*), also Sander et al. (*21*), Fotopoulos et al. (*22*), and Graham et al. (*23–25*)—cited in Gamberale et al. (*26*), and Roberge (*35*), and Stopps and Janischewskyj (*36*)—cited in Broadbent et al. (*33*)].

Several of the human studies do not choose as outcomes the best-known and most reliably administered behavioral and cognitive measures. Idiosyncratic choice of test procedure seems the norm, and replication of the same cognitive or behavioral test by more than one investigator is rare.

A remarkable and much-cited study of the effect of EMF (*27*) exposure on circadian rhythm is an example of this experimental idiosyncrasy. The investigators constructed two underground chambers, one of which was lined with material that prevented entry of the earth's natural electro-magnetic field into the chamber. Both chambers avoided any indication of solar time (no windows, clocks, etc.). With humans living in these chambers for 3 to 8 weeks at a time, it was found that the naturally occurring circadian rhythms were lengthened about 15 minutes by shielding but shortened about 70 minutes when an electric field of 2.5 V/m at 10 Hz was applied to the shielded room. These changes represent less than a 5% change from the natural 24- to 25-hour cycle and are not known to carry any clinical significance.

This work is considered "probably the most significant work on the effects of electromagnetic fields on circadian rhythms" (*28*); yet it is unlikely to be replicated by other investigators because of the expense of the experimental arrangement, the need for human volunteers to spend a great deal of time in what is surely an unpleasant setting, and the modest effects observed even under these strenuous conditions.

Stollery (*20*) examined 76 volunteers in a cross-over trial. The subjects (all male) were exposed to a 500-μa current (50 Hz) via skin electrodes. In the control situation, no current was passed, but the blindness of the experiment was partly compromised by the ability of some subjects to perceive the electric field. No effects were found on self-reported stress, semantic reasoning, vigilance, or concentration. Some effects were found on a subset of the vigilance test (time taken to identify nontarget numbers), on arousal, and on some parts of the syntactic reasoning test. However, these effects were restricted to the second day of this 2 day experiment and thus were found only in one half of the experiments.

Studies by Johansson (*29*) in Sweden and by Rupilius (*30*) in Germany apparently failed to show neurobehavioral effects of exposures of 20 kV/m. With exposure to 100 kV/M, Kanz (*31*), cited in Knave et al. (*32*), found subjective reports of changes in hearing and taste and "pains in the nerves." How the subjects knew that the origin of their pain was neural is not clear.

Graham and colleagues have performed several experiments on young, male, human volunteers exposed to 60-Hz electric and magnetic fields at a variety of field strengths (*23–25*). Among behavioral and cognitive parameters assessed, mood, simple reaction time, memory span, fatigue, and decision-making ability were not consistently affected by the exposure, though some effects were seen in some exposure conditions, particularly with a fast, intermittent pattern of electric-field exposure. More consistent effects were noted in slowing of the heart rate.

Studies of Occupationally Exposed Workers

A wider variety of neurobehavioral outcomes has been studied in workers exposed occupationally to EMF. These studies have assessed reaction time, vigilance, short-term memory, perception, psychiatric symptoms, self-reported memory loss, and manual dexterity. Several of these tests have been used in more than one study. EEG findings have been assessed in two studies. Moreover, a group of four occupational studies, to be described in detail, verified the occupational exposure through measurement of electrical and/or magnetic fields in the workplace.

Early studies of electrical workers in the Soviet Union have not been considered useful contributions. These studies described nonspecific symptoms, such as dizziness and headache, in workers with unmeasured and uncertain exposure, and the studies failed to include controls.

Knave et al. (*32*) assessed 53 Swedish men who had worked for more than five years in high-voltage (400 kV) substations. An equal-sized control group consisting of low-voltage (220/380 V) distribution workers was matched individually to the exposed cohort by location, age, and duration of employment. No adverse effects of high-voltage exposure were found in eight psychological performance tests (reaction time, two memory tests, manual dexterity, addition, tapping, perceptual speed, and matrices), nor on EEG examinations, nor on self-reported (on a standardized questionnaire) anxious or

depressive symptoms. In fact, for several tests, scores were higher for the exposed group, who had a higher level of educational achievement than the controls. Control for educational background in the analyses was not attempted.

Broadbent et al. (*33*) interviewed 390 electrical power transmission and distribution workers and obtained exposure measurements in 287 of them. No correlation was found between either measured or estimated EMF exposures and self-reported headaches, anxiety, obsessional or somatic symptoms, depression, or episodes of forgetfulness.

Baroncelli et al. (*34*) examined four groups of male employees who worked in and around the 258 electric power substations (220 kV) of the Italian State Railways. The four groups were defined by the number of hours per week they were estimated to have been exposed to maximum electric-field strength. One group, employed in the same departments as the others, had no exposure, while the other three groups were exposed for 1, 10, and 20 hr per week. No differences were found among the four groups in acoustic reaction time, visual reaction time, IPAT-anxiety state, and state-trait anxiety.

Gamberale et al. (*26*) studied 26 linesmen during 2 working days immediately before and immediately after performing a simulated routine inspection on a 400 kV power line. On one of the days the power line was in operation; on the other, it was not. The workers worked on the line from 10:00 a.m. to 2:30 p.m. except for a 30-min lunch break, which was taken in a trailer placed under the line to insure continuity of exposure. No differences were found in the exposure and control conditions in a variety of self-reported symptoms, such as wakefulness, stress, concentration, energy level, headache, and anorexia. No changes were noted in a variety of tests of color word vigilance, time to pair symbols and digits previously presented as pairs, and number of presented digits that could be remembered. Simple reaction time did not differ overall between the two conditions, but the improvement from morning to afternoon testing seen in both conditions (possibly a practice effect) was slightly, but significantly, less in the exposed condition, even though both reaction times were better in the exposure condition. This latter observation was the only significant difference among the dozens of behavioral variables assessed.

Each of these four studies included a measure of exposure over and above the worker's job classification. In three studies, Knave et al. (*32*), Broadbent et al. (*33*), and Gamberale et al. (*26*), study subjects wore an exposure meter during or shortly before the period of investigation.

Knave et al. (*32*) measured electric-field strengths at a height of 1.8 m in the high-voltage substations using a field intensity meter but did not provide the findings in their paper. The study subjects also wore dosimeters for an unspecified period of time, during which the percent of time spent exposed to four ranges of electric-field strength (<5, 5–10, 10–15, 15–20 kV/m) was measured for four different types of work. It was found that less than 5% of work time was spent in a field of 10 kV/m or greater during inspection work, "everyday" work, and resting, but that revision of breakers (which involves ascent to the same level as the breaker, 6–8 m) involved exposure above 10 kV/m for 34% of the time and above 15 kV/m 16% of the time.

Broadbent et al. (*33*) had their workers wear a single-channel electrochemical exposure meter strapped to their left arms for two weeks closely preceding the questionnaire administration. About 10% of the sample received exposures above 6.6 kV/m. They concluded that their subjects received exposures an order of magnitude lower than those of Knave et al. (*32*). Gamberale et al. (*26*) used a BE-log dosimeter, which detects both electric and magnetic fields, and which was worn during both the exposed and control conditions. Average exposure was 2.8 + 0.35 kV/m, and 23.3 + 4.2 μT in the exposed condition.

Baroncelli et al. (*34*) did not obtain personal dosimetry but measured electric-field and magnetic-flux density at two substations (all Italian railway substations apparently were built to identical specifications). Electric fields ranged from 1 to 5 kV/m and magnetic fields ranged from 4 to 15 μT at 1.5 m. Two unpublished Canadian studies cited by Broadbent (*33*) failed to find neurobehavioral effects in workers.

Epidemiologic Studies of Suicide and Depression

Two English studies have assessed, each in a different way, the distance between the last known address of 598 suicides and of 598 control subjects (selected randomly from the electoral register) and overhead high-voltage transmission lines.

In the first publication (*35*), exposure was assessed based on calculations of the maximum electric- and magnetic-field strength at any address using the known configurations of voltage, current, orientation, and other factors specific to the nearest high-voltage transmission line. Three relationships are presented: the proportion of addresses in both series with estimated electric fields exceeding 0.1, 0.5, and 1.0 V/m; the number of rank-ordered pairs (based on ranking the case and control series separately on electric-field exposure) in which the value for the suicide address exceeded the control address; and, within each of the three defined electric-field thresholds (0.1, 0.5, 1.0 V/m), the number of suicides and controls fitting into each decile of exposure.

The proportion of addresses above the 0.5 and 1.0 V/m threshold is slightly higher among controls, but no statistical test or measure of association is presented. The number of pairs in which the controls exceeded the suicide in electric-field exposure was significantly more than the converse condition. Within each exposure level, a significant difference between the two groups for exposure is reported, but inspection of the data does not indicate this is necessarily based on higher exposure deciles among the suicide victims.

This fairly strong result, indicating suicides were less exposed to electric and magnetic fields, is interpreted by the authors as evidence that a correlation between the two variables had been established, but that it is a relationship whose direction (i.e., whether suicides were less or more exposed to EMF than controls) is uncertain. They state their conclusion as follows: "It is not possible to determine whether more or less than the expected number of suicides occurred at the higher field-strength addresses" (*35*).

In a second paper (*36*), the group measured magnetic-field strength 0.5 m from the front door of all but 12 of the 1196 subjects described in the previous study. The mean magnetic-field strength was 867 μG in the suicides, 709 μG in the controls, a significant ($p < 0.05$) difference, but one that represents just one-seventh of the pooled standard deviation of more than 1000 μG. Of the suicides, 47% had magnetic-field exposures above the median as compared to 39% of controls ($p < 0.01$). The authors provided a calculation [challenged by Bonnell et al. (*37*)] indicating the median magnetic field measured in their study (400 μG) could induce an electric field inside the human body similar to that found to affect behavior in monkeys (i.e., 3.5–4.0 V/m). Noteworthy are the authors' conclusions that most of this magnetic field must have come from indoor appliances and wiring and not the high-voltage transmission lines, which were assumed to produce no more than 50 μG in the residences.

Perry et al. (*38*) continued the theme of this work in the same region of England

by measuring magnetic fields at the addresses of patients discharged from the hospital with depression and myocardial infarction. Again, the addresses were compared with addresses of controls obtained from electoral registers. The only statistical result presented is the one-sided probability (0.033) associated with the regression coefficient for suicide case status prediction of measured magnetic field, controlled for electoral ward and distance from the nearest roadway (thought to contribute to noise and pollution and thus possibly confounded with depression). No relationship was found for myocardial infarction. The average field strength was considerably higher than in the previous study, 2.26 μG in the depressive patients, 2.07 μG in controls.

In an earlier paper, Perry and Pearl (*39*) found that residents of an apartment block nearer to the main electrical supply cable, and to the corresponding higher magnetic field, were more likely to be admitted to the hospital both for depressive illness and myocardial infarction than residents of the same block living at a greater distance from the supply cable.

This series of epidemiologic papers leaves much to be desired. The influence of confounding factors that might link depression or suicide to place of residence or to use of electrical appliances is not addressed. Indeed, there is virtually no exploration of the sociodemographic characteristics of the compared populations. Suicide victims and patients with depression are unlikely to be comparable to a random sample of individuals who are well and stable enough to be entered onto the electoral rolls. Although little work has been done on the socioeconomic aspects of power-line and electric-cable siting, it would not be surprising if there were associations between, for example, urbanization and crowding and location of EMF exposure sources. These variables might, in turn, be linked to depression or suicide.

More recently, a British study of suicide by occupational classification reported essentially no relation with job titles associated with EMF exposure. A slight excess of suicide was seen in radio and TV mechanics in two vital data sets obtained a decade apart, but this group did not have particularly high occupational EMF exposure (*40*).

Implications for Future Epidemiologic Research

Hypotheses Worth Pursuing

At times, epidemiologic research is in a position to pursue in the population clues about disease causation that are reflective of well-established pathophysiological mechanisms. More often, however, having such variables available is an unattainable luxury. Indeed, in many of the triumphs of epidemiology, the biological mechanisms were revealed after, and not before, the epidemiologic association had been established. This was true for smoking and lung cancer, prenatal diethylstilbestrol (DES) and vaginal cancer, aspirin and Reyes syndrome, prenatal rubella and cataract, cholera and water supply, and a host of other epidemiologic discoveries.

However, in each of the examples listed above, and perhaps in virtually all epidemiologic discoveries, there was a body of knowledge that made the association biologically plausible. Both cigarette smoke and DES were known laboratory carcinogens (though not known to produce the specific disease studied by epidemiologists); aspirin is a liver mitochondrial toxin, and prenatal viral infections have long been known to produce fetal damage. A reading of John Snow's 1854 treatise on cholera and water supply will show how he adhered closely to known biological principles and evidence even though microbes were not yet known causes of disease. True epidemiologic associations do not emerge out of the blue.

Thus, even if the mechanisms of the disease in question have not been worked out fully, an important epidemiologic precept is, or ought to be, that epidemiologic studies must have a serious biological basis. Studies that simultaneously examine suicide and ischemic heart disease, such as those of Perry et al. (*38,39*), appear to have no basis in plausible biology. No laboratory experiment points to an effect of EMF on a biological mechanism common to suicide and to ischemic heart disease, if such a mechanism exists.

A particular risk in exposure-based epidemiology, such as environmental epidemiology, is that the exposure of interest will be assessed in relationship to any disease, symptom, or complaint available for study, regardless of biological plausibility. Under such circumstances, whether because of chance or bias, associations surely will be demonstrated. Biological plausibility is an important constraint not just on the interpretation of results but on the design of studies.

In the field of neurobehavioral relationships to EMF, there has been considerable expenditure on laboratory experiments but a failure to develop a consistent experimental model that would lend itself to extrapolation to epidemiologic research. In the absence of this mechanism, what directions ought to be pursued?

There appear to exist two sets of biological mechanisms in this area where findings have been replicated. The first is calcium efflux across the cell membrane; the second is in the function of that part of the nervous system involved in circadian rhythmicity.

A limitation of the calcium efflux model is that it seems to lend itself to a large variety of disease states. Pending the prediction of a specific neurobehavioral finding based on this model, even whole animal researchers, let alone epidemiologists, have nowhere to turn if they wish to verify this model on their study populations.

On the other hand, rhythmicity, both circadian and seasonal, is a potentially powerful mediator of psychiatric state in humans. Psychiatric disorders of rhythmicity, such as seasonal-affective disorder and premenstrual syndrome, are well established. There exists, therefore, a plausible biological basis for linking these disorders, or similar ones, to EMF exposure. In the present state of laboratory-based knowledge, the hypothesis that EMF exposure might be a contributing cause to depressive illnesses seems worthy of epidemiologic assessment.

Studies Worth Undertaking

Although most scientists have a conscious or unconscious bias in favor of studies that produce positive findings, the most impressive human studies in the EMF-neurobehavior field are the careful investigations of cognitive performance in occupationally exposed men, studies whose results are negative.

Inasmuch as the subjects of these studies had been exposed occupationally for a considerable period of time, usually years, and their exposure in typical work situations had been measured and found to exceed the general population exposure by orders of magnitude, the absence of any real effects on a variety of cognitive measures must be viewed as a strong and reassuring negative result. These studies suggest there is little or no value to larger scale epidemiologic research that attempts to link cognitive outcomes to EMF exposure.

The strength of these studies indicates that populations occupationally exposed to high intensities of electric- and magnetic-field exposure are likely to be excellent candidates for studies of the incidence or prevalence of depression in relation to EMF exposure. A few questionnaire items in several of the occupational studies [e.g., Broadbent et al. (*33*), Knave et al. (*32*), Gamberale et al. (*26*)] assessed mood and similar parameters. However, standardized instruments for the

detection of depression were not used, and none of the studies had a large enough sample to detect elevated levels of clinical depression. Inasmuch as women are more liable to depression, at least at younger ages, it will be important to look carefully for exposed female populations. To clarify the linkages to the abnormalities of rhythmicity observed in the laboratory setting, it will be necessary to screen especially for seasonal disorders, and this may require studies that cover several seasons or that take account of the menstrual cycle in the assessment of depressive symptoms.

These studies will be more valuable if conducted using prospective cohort or matched-exposure approaches, as contrasted to the case–control method. Important information is conveyed by using a sampling of the actual work experience to demonstrate that the exposures are indeed high rather than relying on estimates based on work classification. Case–control studies of depression in relation to occupational exposure might be of some value but are unlikely to have access to measured exposures. Moreover, employees in whom exposure causes major health effects may leave the work force and thus be excluded from case–control studies.

Occupational studies obviate some of the unresolved issues surrounding residential EMF exposure. It is surprising that wiring codes and electrical distribution patterns in neighborhoods have not been studied sociologically. Is it likely that neighborhoods in the United States, so carefully segregated by income and class, have identical distributions of overhead transmission lines, transformers, and electrical substations? One needs only to think of the expression "the other side of the tracks" to note the possible association of social class with exposure to the electricity associated with railroads. Until we have a better understanding of these patterns of association, all epidemiologic studies based on residential exposure will remain suspect for confounding, particularly for neurobehavioral health issues, all of which are powerfully linked to social class and economic opportunity.

For this reason, case–control studies of depression in relation to residential exposure to EMF do not seem a promising avenue of research, at least until the above-noted issues have been addressed. ehp

REFERENCES

1. Bawin SM, Adey WR. Sensitivity of calcium binding cerebral tissue to weak environmental electrical fields oscillating at low frequency. Proc Natl Acad Sci USA 73:1999–2003 (1976).
2. Blackman CF, Benane SG, Kinney LS, Joines WT, House DE. Effects of ELF fields on calcium-ion efflux. Radiat Res 92:510–520 (1982).
3. Blackman CF. The role for the magnetic component in the field-induced efflux of calcium ions from brain tissue. In: Electromagnetic fields and neurobehavioral function (Lovely RH, O'Connor ME, eds). New York: Alan R Liss, 1985.
4. Gunderson R, Greenebaum B, Goodman E. Effects of 60-Hz electromagnetic fields on calcium efflux and neurotransmitter release. Technical report prepared for the New York State Power Lines Project, Albany, 1986.
5. Lovely RH. Behavioral toxicology of ELF electric and magnetic fields. In: Electromagnetic fields and neurobehavioral function (Lovely RH, O'Connor ME, eds). New York: Alan R Liss, 1985.
6. Blackman CF, Benane SG, House DE. Effect of ambient levels of power-frequency electric fields on a developing vertebrate. Bioelectromagnetics 9(2):129–140 (1988).
7. Wilson BW, Anderson LE, Hilton DI, Phillips, RD. Chronic exposure to 60-Hz electric fields: effect on pineal function in the rat. Bioelectromagnetics 2(4):371–380 (1981). See also erratum: Bioelectromagnetics 4(3):293 (1983).
8. Reiter RJ, Anderson LE, Buschbom RL, Wilson BW. Reduction in the nocturnal rise in pineal melatonin levels in rates exposed to 60-Hz electric fields in utero and for 23 days after birth. Life Sci 42:2203–2206 (1988).
9. Vasquez BJ, Anderson LE, Lowery C, Adey AR. Effects of 60-Hertz fields on the daily rhythms of opiate receptors. Soc Neurosci Abstr 11:443 (1986).
10. Wolpaw JR, Seegal RF, Dowman R, Satya-Murti S. Chronic effects of 60-Hz electric and magnetic fields on primate central nervous system function. Technical report prepared for the New York State Power Lines Project, Albany, 1968.
11. Dowse HB, Palmer JD. Entrainment of locomotor activity rhythms in mice electrostatic fields. Nature 222:564–566 (1969).
12. Ehret CF. High strength electric fields are circadian zeitgebers. In: Biological effects from electric fields associated with high voltage transmission lines: DOE contractors review. Washington, DC: Division of Electric Energy Systems.
13. Roberts AM. Effect of electric fields on mice. Nature 223:639 (1969).
14. Dowse HB. The effects of phase shifts in a 10-Hz electric field cycle on locomotor activity rhythm in Drosophila melanogaster. J Interdiscip Cycle Res 13:257–264 (1982).
15. Ehret CF, Rosenberg RS, Sacher GA, Duffy PH, Groh KR, Russell JJ. Biomedical effects associated with energy transmission systems: effects of 60-Hz electric fields on circadian and ultradian physiological and behavioral function in small rodents. DOE/TIC/1027653 (DOE81027653). Springfield, VA: National Technical Information Service, 1981.
16. Ehret CF, Rosenberg RS, Sacher GA, Duffy PH, Groh KR, Russell JJ, Malarkey D, Rosado JO, Tiefentahl CA. Effects of extremely low frequency electric fields on ultradian rhythms of physiological and behavioral functions in rodents. Arkh Patol 43(11):33–36 (1981).
17. Sulzman FM, Murrish DE. Circadian rhythms. Technical report prepared for the New York State Power Lines Project, Albany, NY, 1986.
18. Bitz, DM, Sargent ML. A failure to detect an influence of magnetic fields on the growth rate and circadian rhythm of Neurospora crassa. Plant Physiol 53:15–157 (1974).
19. Thomas JR, Schrot J, Liboff AR. Low-intensity magnetic fields alter operant behavior in rats. Bioelectromagnetics 7:349–357 (1986).
20. Stollery BT. Effects of 50-Hz electric currents on mood and verbal reasoning skills. Br J Ind Med 43:339–349 (1986).
21. Sander R, Brinkman J, Kuhne B. Laboratory studies and animals and human beings exposed to 50-Hz electric and magnetic fields. In: Proceeding of the International Congress on Large High Voltage Electric Systems, CIGRE paper 36-01. Paris: CIGRE, 1982.
22. Fotopoulos SS, Cook MR, Gerkovich MM, Koontz E, Graham C, Cohen HD. Effect of 60-Hz electric and magnetic fields on human biochemical activity. In: Biological effects from electric and magnetic fields associated with high-voltage transmission lines. Contractors review. St. Louis, MO, 1984. Washington, DC: U.S. Department of Energy, 1984.
23. Graham C, Cohen HD, Cook MR, Gerkovich MM, Phelps JW, Fotopoulos SS. Human exposure to a 60-Hz fields: effects on performance, physiology and subjective state. In: Biological effects from electric and magnetic fields associated with high-voltage transmission lines. Contractors review. St. Louis, MO, 1984. Washington, DC: U.S. Department of Energy, 1984.
24. Graham C, Cohen HD, Cook MR, Gerkovich MM, Phelps JW. Effects of 60-Hz electric and magnetic fields at two intensity levels. In: Biological effects from electrical and magnetic fields, air ions, and ion currents associated with high transmission lines. Contractors review. Kansas City, MO, 1987. Washington, DC: U.S. Department of Energy, 1987.
25. Graham C, Cook MR, Cohen HD. Final report: immunological and biochemical effects of 60-Hz electric and magnetic fields in humans. Report No. DOE/CE/76246-T1. U.S. Department of Energy Office of Scientific and Technical Information, 1990.
26. Gamberale F, Olson BA, Eneroth P, Lindh T, Wennberg A. Acute effects of ELF electromagnetic fields: a field study of linesmen working with 400-kV power lines. Br J Ind Med 46:729–737 (1989).

27. Wever R. The Circadian system in man. New York: Springer-Verlag, 1979.
28. Sulzman FM. Effects of electric fields on circadian rhythms. In: Electromagnetic fields and neurobehavioral function (Lovely RH, O'Connor ME, eds). New York: Alan R Liss, 1985.
29. Johansson R, Lundquist AG, Lundquist S, Scuka V. Is there a connection between the electricity in the atmosphere and the function of man? Part 3. 50-Hz field variations. FOA Report C267-H5. Stockholm: Foreign Operation Administration, 1973.
30. Rupilius JP. Research on the action of electric and magnetic 50-hz alternating fields on man. Master's thesis. Freiberg, Germany: Albert-Ludwig University, 1976.
31. Kanz E. Neues uber die Witterwirkung auf die psychologische Reaktionslage des Menschen. Arch Phys Ther (Leipzig) 211:3 (1951).
32. Knave B, Gamberale F, Bergstrom S, Birke E, Iregren A, Kolmodin-Hedman B, Wennberg A. Long-tern exposure to electric fields: a cross-sectional epidemiologic investigation of occupationally exposed workers in high-voltage substations. Scand J Environ Health 5:115–125 (1979).
33. Broadbent DE, Broadbent MHP, Male JC, Jones MRL. Health of workers exposed to electric fields. Br J Ind Med 42:75–84 (1985).
34. Baroncelli P, Battisti S, Checcucci A, Comba P, Grandolfo M, Serio A, Vecchia P. A health examination of railway high-voltage substation workers exposed to ELF electromagnetic fields. Am J Ind Med 10:45–55 (1986).
35. Reichmanis M, Perry FS, Marino AA, Becker RO. Relation between suicide and the electromagnetic field of overhead power lines. Biol Chem Phys 11:395–403 (1979).
36. Perry FS, Reichmanis M, Marino AA, Becker RO. Environmental power-frequency magnetic fields and suicide. Health Phys 41:267–277 (1981).
37. Bonnell JA, Norris WT, Pickles JH, Male JC. Comments on environmental power-frequency magnetic fields and suicide. Health Phys 44:698–699 (1983).
38. Perry S, Pearl L, Binns R. Power frequency magnetic field: depressive illness and myocardial infarction. Public Health 103:177–180 (1989).
39. Perry S, Pearl L. Power frequency magnetic field and illness in multistory blocks. Public Health 102:11–18 (1988).
40. Baris D, Armstrong B. Suicide among electric utility workers in England and Wales. Br J Ind Med 47:788–789 (1990).

Human Adverse Reproductive Outcomes and Electromagnetic Field Exposures: Review of Epidemiologic Studies

Gary M. Shaw* and Lisa A. Croen

March of Dimes Birth Defects Foundation, California Birth Defects Monitoring Program, Emeryville, CA 94608

Concerns have been raised regarding a relation between residential and occupational electromagnetic (EM) field exposures and adverse reproductive effects. This paper reviews the epidemiologic evidence for this possible relation, including some pertinent methodologic issues, notes relevant findings from the experimental literature, and discusses areas for future research. Evidence is lacking for a strong association between a woman's use of a video display terminal (VDT) during pregnancy and spontaneous abortion. The evidence for a strong association between a women's use of a VDT and other adverse reproductive endpoints is also lacking, with some suggestive findings for congenital malformations and too few data to reach a conclusion about other endpoints. With respect to low-level EM field exposures other than VDTs, the paucity of data prevents one from determining whether there are reproductive health risks associated with such exposures. Therefore, this is an area that needs further investigation. Given that altered growth may be an underlying biologic effect of EM field exposures, endpoints that might be pursued in future studies include congenital malformations not associated with chromosomal anomalies, intrauterine growth retardation, and chromosomally normal spontaneous abortions. — Environ Health Perspect 101(Suppl 4):107–119 (1993).

Key Words: Pregnancy, spontaneous abortions, congenital malformations, electromagnetic fields, adverse reproductive outcomes

Introduction

In the last 10 years, public concern regarding possible human health effects from exposures to nonionizing electromagnetic (EM) fields has been mounting. Although the primary focus has been on potential carcinogenic effects, there has also been concern that exposures from electric and magnetic fields will result in adverse reproductive effects. This concern is at least partially attributable to numerous reports of clusters of female video display terminal (VDT) users who have experienced a spontaneous abortion. Interest has also heightened because of the ubiquity of EM fields and the consequent prevalence of exposure. Clearly, even if risks from exposures are low, the extent of populations exposed in modern society could result in a large disease burden.

Since Wertheimer and Leeper (*1*) reported the initial observation of an association between electric power lines and childhood cancer, several other investigators have looked for associations between various EM field exposures and childhood cancers (*1–6*), adult cancers (*7–11*), and adverse reproductive outcomes (*12–36*), including congenital malformations, spontaneous abortion, reduced birth weight, and prematurity.

A number of reviews (*37–41*) offer an overview of the epidemiologic literature for EM fields and human health effects. This paper will: *a*) summarize some of the epidemiologic literature on reproductive endpoints and low-level EM fields, *b*) note some findings from experimental work possibly relevant to humans, *c*) discuss some methodologic issues related to epidemiologic research of reproductive outcomes and EM fields, and *d*) identify areas for future research regarding adverse reproductive outcomes and EM field exposures.

Previous Epidemiologic Research

Studies on reproductive health have investigated a variety of residential and occupational EM field exposure sources including electric blankets, type of home heating, occupations in electrical industries and occupational use of VDTs. For the purposes of summarizing previous epidemiologic findings, studies are grouped into those examining residential and occupational EM field exposure sources (excluding VDTs) and those examining exposures from VDTs. Potential EM field exposures in these two groups are not mutually exclusive (i.e., VDT use may result in exposures to both extremely low frequency and very low frequency EM fields).

Residential and Occupational EM Field Exposure Studies

Studies that have investigated a possible relation between EM fields and reproductive end points are summarized in Table 1. Wertheimer and Leeper conducted two investigations of residential exposures to EM fields (*14,15*). The first (*14*) involved an examination of possible fetal effects from parental use of electric blankets or heated waterbeds. Parents whose baby's birth was announced in the local newspaper were contacted and interviewed about electric blanket and waterbed use. Data were collected from 1256 index birth families, representing 29% of the total number of births in the study catchment area and time period. Length of gestation was observed to be somewhat longer for infants whose parents used electrically heated beds when those infants were conceived during a season when the need for electric bed heating was greatest. The biologic basis for that observation was not given. This study found no difference between the user and nonuser groups in the proportions of infants weighing less than 2500 g at birth. It did find, however, that among low-weight infants in the user group 46% had gestations of 37 weeks or more, whereas 21% of the low-weight infants in the nonuser group had term gestations. This suggests that parental use of an electrically heated bed may be associated with having a child that is growth retarded. The prevalence of congenital malformations was too low to evaluate among the 528 siblings. Abortions (induced or spontaneous) occurring in the one year preceding conception of a live birth (index birth or sibling) were more common among

This manuscript was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990 – September 1992.

*Address reprint requests to G. Shaw, California Birth Defects Monitoring Program, 5900 Hollis Street, Suite A, Emeryville, CA 94608.

This work was supported in part by the Health Effects Institute, Boston, Massachusetts.

Table 1. Studies of electric and magnetic field exposures

Reference	Study design	Subjects included in study	Study outcome definition	Source of outcome information	Study period
Residential exposures					
Wertheimer and Leeper (*14*)	Cohort	1256 Denver births 692 earlier siblings	Birthweight, previous abortions, gestational length	Birth announcements, earlier birth certificates, questionnaire	1982 for index births 1975–1981 for previous pregnancies
Wertheimer and Leeper (*15*)	Cohort	1879 Oregon liveborns with traceable addresses of and information on type of home heating	Fetal losses < 20 weeks gestation 1 year prior to conception of live birth	Birth certificates	1983, 1985
Dlugosz et al. (*16*)	Matched case–control	542 cases, 542 controls, New York	Neural tube and oral cleft defects	New York Malformation Registry	1983–1986
Eckert (*34*)	Case study	294 infants who died from Sudden Infant Death Syndrome, Germany	Sudden Infant Death Syndrome	?	1961–1967
Occuptional exposures					
Knave et al. (*12*)	Cross-sectional	53 workers in high voltage jobs in Sweden, 53 workers in low voltage job	Fertility measured by number of children and sex ratio	Questionnaire	1952–1975
Nordstrom et al. (*13*)	Historical cohort	880 pregnancies among 372 male workers' spouses in Sweden power occupations	Birthweight, perinatal death, spontaneous abortions, and congenital malformations	Mail survey and medical records verification	1953–1979
Buiatti et al. (*17*)	Case–control	112 infertile males and 127 male controls with normal sperm counts, Italy	Male infertility	Single hospital/clinic population	1979–1981
Hemminki et al. (*18*)	Historical cohort	195 spontaneous abortions among 35,000 female members of Union of Metal Workers, Finland	Spontaneous abortion	Hospital discharge registry	1973–1976

(residential and occupational) and adverse reproductive outcomes.

Exposure definition	Source of exposure information	Timing of exposure	Maternal or paternal exposure	Results
Use of an electric blanket or heated previous 8 years	Telephone interview	Selected periods around pregnancy during	Maternal and paternal	No difference noted for low birthweight, some evidence for intrauterine growth retardation among exposed births, abortions higher in previous pregnancies among exposed, and abortions followed a seasonal pattern of need for blanket use
Home with ceiling cable heat	Assessor's office files	Time of fetal loss	Assumed maternal	Abortion frequency not higher in homes with ceiling cable heat. A positive correlation was found between the fetal loss ratio for ceiling cable heated homes and the increase in heating degree days
Use of electric blankets and waterbeds by setting and season	Mail survey	First trimester	Maternal	Risks for neural tube defects (odds ratio = 0.9), cleft lip +/– palate (odds ratio = 0.7), and cleft palate (odds ratio = 0.8) not increased for electric blanket users vs. nonusers; results similar for waterbed use
Floor level of residence	Address books	Time of infant death	?	Larger number of cases lived on basement or first floors than on higher floors
Employment in high-voltage job for more than 5 years	Employment records	Before child's birth	Paternal	Fewer children born to exposed fathers but difference present before "exposed" workers had started high voltage jobs; fewer boys among exposed fathers
Occupation groups: high-voltage switchyard workers (400 kV), construction workers in switchyard (130–200 kV), other electric field exposure (<70 kV)	Employment records	Conception	Paternal	Smaller proportion of male births among swtichyard workers, 3-fold higher prevalence of congenital malformations in switchyard offspring, switchyard workers reported more difficulty fathering children no other endpoint elevated among exposed
Occupation in the radioelectric industry	Face-to-face interview	Longest held occupation	Paternal	Increased risk observed for men employed in radioelectric industry; odds ratio = 5.9, 95% confidence interval (0.86, 40.2)
Occupations in electronic branch of metal industry	Union records	While member of union	Maternal	Age-standardized rate of spontaneous abortions for electronics workers was 8.0% and for radio and television workers 12.8% compared with 5.5% and 4.8% before becoming a member

electric blanket users (7.8%) and waterbed users (6.3%) than among nonusers (4.2%). An uneven seasonal distribution of abortions was observed among electric blanket users, with all 24 occurring between September and June. This uneven distribution was not observed among waterbed users or among nonusers.

While this study suggests an association between electric bed-heating devices and fetal growth or fetal loss, the following methodologic considerations may have influenced results in a direction that is difficult to predict: the method for identifying subjects (i.e., published birth announcements), the source of reproductive end-point and covariate information (i.e., birth certificates), the nonindependence between index births and their siblings, and the exclusion of fetal losses among primagravida women. In addition, as the authors note, results do not differentiate whether the associations might be related to the thermal effect of the bed heating devices or to the EM field produced by the bed heaters.

In an effort to disentangle thermal versus EM field effects, Wertheimer and Leeper (*15*) investigated whether homes with ceiling cable heat (EM fields are higher from this heating source than from baseboard heating) were associated with increased fetal losses compared to homes with other types of heating sources. The investigators selected a sample of Oregon births from 1983 and 1985 and identified prior fetal losses (less than 20 weeks gestation) among their mothers, as reported on the birth certificate of the index birth. They compared the ratio of fetal losses in cable heat homes to the ratio in homes with other types of heating and did not find a difference (0.076 versus 0.075). (The number of previous fetal losses was the numerator, but the denominator did not include the number of births from the same cohort of conceptions as the fetal losses; rather, it was the number of index births.) A positive correlation was observed between a monthly ratio of fetal loss at ceiling cable heat homes (relative to homes without cable heat) and the increase in heating degree days reported for that month.

These data, along with their previous observations from Colorado, were interpreted by the investigators as evidence that the number of fetal losses (not computed as a rate) was elevated in months when cold weather was increasing, a time period they argue was also associated with increasing EM field exposures resulting from increased use of ceiling cable heat. As in their previous study, history of pregnancy loss was obtained from a cohort of women who had a more recent pregnancy resulting in a live birth. Thus, fetal losses among primiparous women were not eligible for this study. Data on possible confounders, such as infections or possible risk factors that might follow similar seasonal distributions (but also would have to be related to homes with ceiling cable heat), were not available. Nevertheless, these observations are suggestive and need follow-up by more rigorous study methods.

The only study done to date that has investigated a possible relation between congenital malformations and residential exposures to extremely low frequency EM fields was performed by Dlugosz et al. (*16*). However, details of this investigation have not yet been published so that a critical assessment of the study's null results is premature.

Another study of residential exposures to EM fields concerned sudden infant death syndrome (*34*). Although the sudden death of an infant is a postnatal event, the possibility of death resulting from a prenatal event may warrant consideration of this condition as an adverse reproductive outcome (*42*). The investigator observed that among 294 infants who died of sudden infant death syndrome in Hamburg, Germany, more had lived in basement or first floors, and he argued for an association between EM fields and sudden infant death based on the assumption that lower floors are more likely to have "uncommon magnetic fields or stray electric currents in the ground." The crude exposure definition and the comparisons of number of infant deaths by floors without an estimate of the population at risk make the interpretation of this study difficult.

The earliest epidemiologic study of reproductive health effects from occupational exposures to EM fields was conducted by Knave et al. (*12*). They sought to identify whether fertility (and other chronic health effects) was compromised among men working in high-voltage (50 Hz and 400 kV) substations in Sweden for more than 5 years as compared to men having occupations with low-voltage EM fields. Results showed that men exposed to high-voltage EM fields had fewer children than men working in low-voltage occupations. This difference, however, was seen prior to the date the exposed men began work in a job that involved high-voltage EM fields. The difference in number of children might have indicated that the men working in high-voltage conditions, as a group, were less frequently attempting to have children. The investigators also observed an altered sex ratio, with a paucity of male children, and concluded that work in high-voltage substations could not be ruled out as a possible explanation for this result. This nonspecific result may be suggestive of reproductive damage associated with occupations in areas having high-voltage EM fields.

Nordstrom et al. (*13*) also studied men working in high-voltage areas for increased frequencies of spontaneous abortions, perinatal death, congenital malformations, altered sex ratio, and reported fertility problems. A cohort of 542 male employees of Swedish power facilities completed a questionnaire about employment characteristics, pregnancy and fertility problems (spontaneous abortions verified in medical records), and possible confounders. Pregnancies among spouses of male switchyard workers in 400-kV substations were considered exposed. Pregnancies among spouses of males not employed as switchyard workers, but employed with the power facilities, were considered the reference group. Some pregnancies among the spouses of switchyard workers were also considered in this group for the periods the male workers were not employed in the switchyard. The prevalence of congenital malformations among children of switchyard workers was three times that of the reference group. The malformations among these infants reflected a heterogeneous group of diagnoses. Switchyard workers were about twice as likely to report fertility problems and had somewhat fewer male offspring compared to reference workers. The latter finding is consistent with the altered sex ratio noted by Knave et al. (*12*). Spontaneous abortions were not more common among switchyard workers. Results were not influenced by adjustment for parental cigarette smoking, alcohol use, medication use, maternal age, or a variety of other possible confounders. These data, although too sparse for analyses of specific end points such as detailed malformation groupings, are suggestive of an association with males' preconceptional exposures and subsequent malformed offspring. Because the separation of groups with differing exposure levels may have been incomplete, a true association could have been underestimated in this study.

In a case–control study of 112 infertile males (azospermic or oligospermic) and 127 males with normal sperm counts, Buiatti et al. (*17*) found an odds ratio of 5.9 associated with employment in the radioelectric industry. The elevated risk was imprecise (95% confidence interval = 0.86, 40.2), and no details were provided about the possible electric and magnetic field exposures that might be associated with occupations in the radioelectric industry.

Hemminki and colleagues (*18*) observed an increased rate of spontaneous abortion (spontaneous abortions/births + induced abortions + spontaneous abortions) for electronics workers among a cohort of 35,000 female Finnish members of the Union of Metal Workers. The increase was primarily among women involved with the production of radios and televisions. The age-standardized rate among radio and television workers was 12%, compared with 4.8% in the same women before joining the Union.

Within the subgroup of radio and television production workers, the investigators noted an excess of spontaneous abortions among women who were exposed to solder. Finding an increase of spontaneous abortions among electronics workers may be interesting in light of increased risks for women employed in the semiconductor industry found by others (*43*), but it does not specifically implicate EM field exposures.

Video Display Terminal Use Studies

Many studies have researched reproductive effects related to VDT use during pregnancy and have focused primarlily on potential risks for spontaneous abortions (*19,20,22–28,35,36*) and congenital malformations (*23,24,26,27,29–33*) and given less attention to other specific adverse reproductive outcomes (*20,21,26–28,32,33*). The technical aspects associated with the use of VDTs along with a discussion of possible health implications of working with a VDT have been reviewed by Bergqvist (*44*). Further, others (*45,46*) have reviewed much of the epidemiologic research regarding VDT exposures, and a detailed summary of the studies done to date is provided in Table 2. Because the recent study by Schnorr et al. (*25*) has not been reviewed elsewhere, details of this study are summarized below.

Schnorr et al. (*25*) compared the prevalence of spontaneous abortions in a cohort of female telephone operators who used VDTs at work to a cohort of female telephone operators who did not use VDTs at work. These two cohorts had similar work situations with the exception of VDT use; therefore, some control for physical and psychologic stress was obtained. A total of 730 married women between the ages of 18 and 33 who had pregnancies during the study time period participated in the study. Among the 323 VDT users, 14.8% of the pregnancies ended in spontaneous abortion, compared to 15.9% among the 407 who did not use a VDT. The percentage of spontaneous abortions was also similar between nonusers and users of 1 to 25 hours per week (15.6% versus 17.2%) and >25 hours per week (15.6 versus 15.4%). Analyses involving potential confounders, early versus late abortions, and only physician-verified abortions did not alter the findings. A sample of the work environments in this study were subjected to measurements of electric and magnetic fields. The VDT users were exposed to levels of extremely low-frequency emissions in the range of exposures similar to exposures in the home and for comparison operators. Unlike others, this study provided some control for psychosocial stress and ergonomic factors and some measurement data on EM field emissions from VDTs. However, it did not have the power to rule out the weak association (approximately a 20% increased risk) observed in earlier investigations.

Findings from Experimental Work

Studies using various animal models have been conducted to investigate potential biologic effects resulting from exposures to electric and magnetic fields, and conflicting evidence regarding the ability of these fields to influence prenatal development has emerged. Juutilanen (*47*) has recently summarized the effects of low-level fields on embryonic development. While some studies, mainly of chick embryos, have reported developmental effects (*48–62*), several other studies found no effects (*63–69*). In studies where effects have been reported, the observed developmental deficits include various congenital anomalies, developmental delay, altered sex ratio, fetal loss, reduced fertility, and demasculinization. The mechanisms underlying these developmental abnormalities are unknown (*70*). These studies have included a variety of exposure paradigms, such as different waveforms, field intensities, exposure durations, and field types (electric, magnetic, or both). The effect of the earth's magnetic field has not always been considered, yet has been shown to influence results in some studies (*71*). Dose–response relations have generally not been observed. Rather, effect windows have been noted in some studies where animal embryonic models subjected to particular field strengths exhibited aberrant development, while animals exposed to field strengths above or below did not (*48,57*).

Studies with cellular test systems have shown that electric and magnetic fields at specific frequencies and intensities are capable of resulting in biologic effects (*72,73*). Although the exact mechanisms for these interactions are unknown, one theory suggests that cells have their own weak electrical signals that enable them to "whisper together," which allows for cell-to-cell communication for normal health (*73*). Disrupted communication by exposure to EM fields may result in unregulated growth (*73*). Evidence from other studies suggest that electric and magnetic fields may alter growth (*74,75*), enhance DNA synthesis (*76*), and influence the modulation of calcium binding to cell surface molecules (*77*). Growth, DNA synthesis, and calcium binding are relevant mechanistically to normal embryogenesis, and alterations to these processes by EM fields suggest that exposure to these fields could result in adverse reproductive effects.

Conclusions that may be drawn from the experimental studies are limited by the inconsistent findings resulting from differences in study design, animal model, and exposure paradigm considered. Many of the positive associations observed among animal models have not been replicated across laboratories, and effects observed from *in vitro* studies have been less evident or not present in *in vivo* systems (*78*), which suggests that the developing fetus may be protected by physical or physiologic maternal attributes. It is clear that exposure to various aspects of EM fields can produce biologic effects in experimental systems; however, the interpretation of these effects in terms of risks to human reproduction needs substantial clarification.

Methodologic Issues

Previous epidemiologic research in this area can be interpreted in the methodologic contexts of: *a*) issues with study endpoints, *b*) issues with exposure assessment, and *c*) other design and analytic considerations. The emphasis of discussion here is on the more common end points studied (i.e., spontaneous abortions and congenital malformations).

Study End Point Issues

The decision to investigate a particular reproductive health effect in relation to a putative exposure is generally based upon the biologic plausibility of such an association as suggested by previous epidemiologic or teratologic data, anecdotal reports made by astute clinical observers, or reports of disease clusters. Based on the concern derived from reports of seemingly unusual aggregations of spontaneous abortions and congenital malformations occurring among VDT users (*44*), and on the observed association between childhood cancer and residential exposure to EM fields (*1,3,4,6*), a variety of reproductive health effects from both residential and occupational EM field exposures have been investigated (Tables 1, 2).

An issue relevant to some of the end points studied is the specificity with which they have been defined for analysis. For example, among the studies that investigated congenital malformations in relation to various EM field exposures (Tables 1, 2), many considered all malformations as a single analytic group. Based on observations with known teratogens, exposures are unlikely to result in a general increase in all types of malformations, although, depending on the timing of exposure, they may increase the risk of more than one type. However, based on the nonspecific biologic effects observed in experimental work of EM exposures, it is not clear how one might better define malformation groups for analysis. Studies of spontaneous abortions might also be criticized for heterogeneous endpoint definitions. Although a few studies (*20,25*) have analyzed early and late spontaneous abortions separately, no study has included karyotype information on aborted fetuses. Spontaneous abortions are etiologically heterogeneous, with approximately 30% to 50% being chromosomally

Table 2. Studies of exposures to video display terminals (VDTs)

Reference	Study design	Subjects included in study	Study outcome definition	Source of outcome information	Study period
Lewis et al. (*35*)	Case–control	30 South Australian spontaneous abortion cases and 60 controls matched on maternal age and on delivery date	Spontaneous abortions (≤ 20 weeks)	Questionnaire (self-administered)	1960–1978
McDonald et al. (*24*)	Historical cohort	9471 pregnancies from among 56,012 women in Montreal; also 8161 previous pregnancies from same women	Clinically recognized spontaneous abortions, congenital malformations	Medical records maternal interview	1982–1984
Kurppa et al. (*29,31*)	Case–control	1475 Finnish infants with malformations; 1475 paired referents	Malformations of central nervous system, cardio-vascular system, skeleton or oral clefts	Finnish Register of Congenital Malformations	1976–1982
Ericson and Kallen (*26*)	Historical cohort	10,025 Swedish women working in selected white collar occupations	Stillbirth, neonatal death, birthweight <2500 g, malformations, clinically recognized spontaneous abortions	Swedish Registries	1976–1977 1980–1981
Ericson and Kallen (*27*)	Case–control	429 cases and 926 controls of similar maternal age who worked outside home in Sweden	Stillbirth, neonatal death birthweight <1500 g, malformations, clinically recognized spontaneous abortions	Swedish Registries	1980–1981
Westerholm and Ericson (*36*)	Historical cohort	4117 pregnancies among female clerks at social security bureaus in Sweden	Hospitalized spontaneous abortion rate, malforma-tions, birthweight, perinatal mortality, stillbirth	Swedish Registries	1980–1983
Butler and Brix (*28*)	Historical cohort	817 pregnancies among 728 female clerical workers in Michigan	Stillbirth, spontaneoos abortions	Questionnarie, medical records	1980–1985
Goldhaber et al. (*23*)	Case–control	452 cases and 723 liveborn controls	Spontaneous abortions, malformations	Medical records, birth certificates	1981–1982
Nurminen and Korppa (*21*)	Historical cohort	Finnish women, referents from earlier malformation study, 60 VDT working mothers, 179 non-VDT working mothers	Threatened abortion, length of gestation, birthweight, placental weight, and maternal blood pressure	Medical records, questionnaire	1976–1982

and adverse reproductive outcomes.

Exposure definition	Source of exposure information	Timing of exposure	Maternal or paternal exposure	Results
Average number of hr/wk using VDT at work	Questionnaire	During pregnacy	Maternal	Odds ratio for any VDT use was 1.7 for all spontaneous abortions
Number of hr/wk using VDT at work	Questionnaire	When woman first suspected she was pregnant	Maternal	Spontaneous abortions among current pregnancies: observed/expected for nonusers was 0.89, for users 1 to 6 hr/wk 1.24,for users 7 to 29 hr/wk and 1.25, for 30+ hr/wk 1.12. Malformations: among current pregnancies prevalence was 3.3% in VDT users vs. 3.7% in nonusers. Among previous pregnancies the prevalences were 3.8% vs. 3.6%.
VDT use determined by industrial hygienist, based on occupational title grouped as 1 to 4 hr use/day or 4+ hr use/day	Questionnaire	First trimester	Maternal	Odds ratios for any VDT use were 0.4 for central nervous system, 0.9 for oral clefts, 0.8 for skeletal defects, and 1.6 for cardiovascular defects
Occupational groups assumed to have high, medium, low potential for VDT use	Census data	One or two years prior to delivery	Maternal	Higher frequency of infants born weighing <1500 g or <2500 g among women in medium exposure group; no excesses noted for malformations, spontaneous abortions, or neonatal deaths
Number of hr/wk using VDT at work	Mail survey	Selected periods during pregancy	Maternal	Odds ratios for VDT use 20+ hr/wk = 2.3 for birth defects, 1.2 for spontaneous abortions and for any VDT use, 1.6 for birth defects and 1.1 for spon taneous abortions
VDT use at work classified into 5 exposure groups based on frequency of use	Trade union and employment records	During 1980–1983	Maternal	No deviance from expected for stillbirths, perinatal mortality, spontaneous abortions or birthweight; a 2-fold excess was observed for significant malformations among exposed <10 hr/wk
Number of hr/wk using VDT at work	Questionnaire, employment records	During pregnancy	Maternal	Stillbirths and spontaneous abortions 1.2 times higher than expected among users of VDTs >20 hr/wk
Number of hr/wk using VDT at work	Mail/telephone questionnaire	First trimester	Maternal	Odds ratios for VDT use 20+ hr/wk = 1.8 for spontaneous abortion, 1.4 for birth defects, and for any VDT use, 1.2 for spontaneous abortions, and 1.2 for birth defects
VDT use determined by industrial hygienist based on occupational title grouped as 1 to 4 hr use/day or 4+ hr use/day	Questionnaire	First trimester	Maternal	No difference noted between VDT workers and non-VDT workers

(Continued)

Table 2. Studies of exposures to video display terminals (VDT's)

Reference	Study design	Subjects included in study	Study outcome definition	Source of outcome information	Study period
Bryant and Love (*22*)	Case–control	334 spontaneous abortion cases, 334 prenatal controls, and 334 postnatal controls in Canada	Spontaneous abortions where women required hospitalization	Medical records	1984–1985
Nielson and Brandt (*19*)	Case–base	666 spontaneous abortion cases and 764 pregnancies in referent group from clerical and commercial workers in Denmark	Clinically recognized spontaneous abortions	Medical registers	1983–1985
Brandt and Nielson (*30*)	Case–base	421 congenital malformations and 1365 pregnancies in referent group from Denmark	Malformations	Medical registers	1983–1985
Tikkanen et al. (*32*)	Case–control	500 infants with cardiovascular malformations; 1055 nonmalformed infants from Finland	Cardiovascular malformations, length of gestation birthweight, placental weight	Finland malformation registry and children's cardiac registry	1982–1984
Bjerkedal and Egenaes (*33*)	Historical cohort	1820 pregnancies among female postal employees in Norway	Malformations, stillbirth, early neonatal death, perinatal death, birthweight, prematurity, multiple birth	Norway malformation registry, medical birth registry	1967–1984
Windham et al. (*20*)	Case–control	439 spontaneous abortion cases; 909 live-born controls in a California county	Spontaneous abortion (≤20 weeks), low birthweight, intrauterine growth retardation	Pathology labs	1986–1987
Schnorr et al. (*25*)	Cohort	323 VDT using operators and 407 non-VDT using operators in U.S.	Spontaneous abortion (≤28 weeks)	Questionnaire, vital records	1983–1986

abnormal (*79*). One would expect that if exposures to EM fields increase the risk for spontaneous abortions then the timing of exposures would be different for early and late spontaneous abortions. The inability to form etiologically homogeneous endpoint groups for analysis may result in an attenuated measure of a true effect.

Another issue of general concern regards the source of endpoint information. Studies have used a variety of sources, including vital records, parental interviews, medical records, pathology records, and malformation registries (Tables 1, 2), to ascertain reproductive health effects. The use of different sources probably results in different prevalence estimates for a given adverse reproductive outcome. A woman's report of a spontaneous abortion, her child's birthweight, or congenital malformation has been shown to be susceptible to error when compared to hospital records (*80,81*). This finding would also suggest that birth certificate data on prior pregnancy loss, as reported by mothers, would also be suspect. At least two studies relied on such information (*14,15*).

Further, the studies done to date have ascertained clinically recognized spontaneous abortions. However, many pregnancy losses occur prior to the recognition of pregnancy (*82*). Interestingly, at least one study (*20*) of spontaneous abortions from VDT use found a higher risk for early (<13 weeks gestation) spontaneous abortions than for late abortions. Even among recognized spontaneous abortions, the timing of diagnosis during pregnancy has varied in these studies [e.g., ≤ 20 weeks (*20*) or ≤ 28 weeks (*25*)]. The opportunity for bias is present if a subject's exposure is related to when her pregnancy is recognized. The earlier a pregnancy is recognized the greater the likelihood is that an early spontaneous abortion will be recognized. Goldhaber et al. (*23*) noted that pregnancies among VDT users were diagnosed an average 4 days earlier than nonusers, and they therefore controlled for this difference analytically. Cultural, ethnic, or educational variations might result in

and adverse reproductive outcomes (**continued**).

Exposure definition	Source of exposure information	Timing of exposure	Maternal or paternal exposure	Results
Number of hr/wk using VDT at home or work	Questionnaire	Period between 3 months prior and 4 months post-LMP	Maternal	Odds ratio for any VDT use was 0.81 using prenatal controls and 1.1 using postnatal controls; odds ratio for >21 hr/wk use was 1.1 using either prenatal or postnatal controls
Number of hr/wk using VDT at work	Questionnaire	During pregnancy	Maternal	Odds ratio for any VDT use was 0.92, and 0.76 for >30 hr/wk compared to no VDT use
Number of hr/wk using VDT at work	Questionnaire	During pregnancy	Maternal	Odds ratio for any VDT use was 0.96, and 1.0 for >30 hr/wk compared to no VDT
VDT use determined by industrial hygienist, based on occupational title grouped as 1 to 4 hr use/day or 4+ hr use/day	Questionnaire	First trimester	Maternal	Odds ratio for any VDT use was 1.2 and for 20+ hr/wk was 1.4 compared to no VDT use for cardiovascular malformations; no relation between VDT use and fetal growth indicators
Time period when VDTs present in work place 1979–1984	Employment records	Employed at end of of pregnancy	Maternal	Perinatal mortality 1.7% vs. 0.8%, low birth weight 1.1% vs. 0.8%, prematurity 9.0% vs. 6.5%, congenital malforma tions 6.2% vs. 5.2% among exposed vs. before period or after employment
VDT use at work defined as number of hr/wk and number of weeks during pregnancy	Questionnaire	During pregnancy	Maternal	Odds ratios for all abortions: <20 hr/wk = 1.2, ≥20 hr/wk = odds ratio for early (<13wk) abortions: <20 hr/wk = 1.5, ≥ 20 hr/wk = 1.4
Hr/wk and weeks during first trimester employed as a directory-assistance operator	Employment records	First 28 weeks of pregnancy	Maternal	Percent of spontaneous abortions were 14.8% vs. 15.9% for VDT users vs. nonusers; 17.2 vs. 15.6% for 1-25 hr/wk use and 15.4 vs. 15.6% for 25+ hr/wk use among users versus nonusers

differences of when a pregnancy is diagnosed, and therefore result in differences in the observed number of clinically recognized spontaneous abortions (*45,83*). If such variations are related to a woman's exposure, the opportunity for biased effect estimates may exist.

Exposure Assessment Issues

Differences in biologic effects are associated with various frequencies in the EM spectrum (*84*). Power frequency fields (50–60 Hz), in contrast to ionizing radiations like X rays, are incapable of breaking chemical bonds, and in contrast to microwaves, do not cause tissue heating (*84*). Also noteworthy is that some physical and biologic properties of electric and magnetic fields are distinct. Methodologic issues relevant to studies concerning exposure pertain to the definition of exposure: what it is; to whom (mother, father, or fetus) and when in the periconceptional period it occurs; the amount (dose) experienced by a study subject; how it is measured; from where information about it was obtained.

Definitions of exposure (Tables 1,2) have involved seemingly heterogeneous exposures, such as working in a 400-kvV switchyard (*13*) and home use of an electric blanket (*14,16*). Studies have differed in both the frequency range studied and the type of field (electric or magnetic) investigated. Those examining risks from VDT use were primarily interested in very low frequency EM fields, and investigations of electrically heated bed-warmers focused on exposure to extremely low frequency fields. In the occupational studies of electrical workers (*2,12,13,17,18*), it is not clear whether risks from electrical fields or magnetic fields were being assessed. In addition, higher frequency field exposures have been examined by other investigators (*85–90*). Even though there is a general lack of information regarding the biologic effects on reproduction from certain areas of the EM spectrum, a discussion of these exposures and studies is beyond the scope of this review.

Aspects surrounding when and to whom an exposure occurred are also important to

consider. A general principle of teratogenesis is that the ability of an exposure to result in an adverse effect is conditional on the timing of that exposure relative to the developing fetus (*90*). Failure to specifically obtain exposure information for a critical time period may produce a risk estimate biased toward the null value. EM field exposures to a man or woman prior to conception may be relevant to reproductive health effects, such as subfecundity or germ cell damage, but may not be relevant to chromosomal damage in the conceptus, given that EM field exposures have not been shown to be mutagenic (*91*). In general, various types of exposures (e.g., medications) a mother encounters during specific periods of pregnancy are plausible risk factors for congenital malformations (*92*), spontaneous abortions (*79*), reduced birth weight (*93*), and certain childhood cancers (*94*). However, there is less support for the notion that various exposures the father or mother receives prior to conception or the father receives during pregnancy may act as risk factors for these outcomes. Studies on risks of VDTs (Table 2) were exclusively interested in maternal use and predominantly for exposure during early pregnancy [but not exclusively (*19,30,35*)], whereas studies involving other EM field exposures (Table 1) involved both maternal and paternal exposures for several periconceptional time periods.

Another relevant issue concerns the extent or dose of exposure. There are multiple sources of low-level EM fields in our environment (*84*), resulting in some level of exposure to most individuals. With the exception of the study by Bryant and Love (*22*), who sought information on both home and work exposures to VDTs, no study considered other sources of possible exposures to EM fields beyond the single exposure of interest. The implication of neglecting these other sources is that variations in EM field exposure among individuals in the study populations might be overlooked. Such misclassification errors probably would reduce the magnitude of estimated effects (*95*). It is unclear, however, what the etiologically relevant measure of exposure is (e.g., time-weighted average, peak frequency, or some other parameter of exposure). Some experimental work (*48*) implies that certain biologic effects are not monotonically related to dose but depend on windows in the exposure range. In terms of measuring exposures, only one study (*25*) incorporated direct field exposure measures. Most studies have relied on self-reported surrogate exposures (e.g., number of hours per week using a VDT) (*19,20,22–24,27,30*), occupation in electronics industry (*17*), or residence in cable heat homes (*15*).

Potential errors in assessing exposure in these studies may be associated with the source of exposure information. Exposure information (Tables 1, 2) was derived primarily from self-reported questionnaire data, employment records, and other existing data sources. Querying study subjects allows for collection of detailed data on characteristics of possible field exposures, such as source and timing. However, this approach may be susceptible to information bias (*96*), as suggested by some investigators (*24,27*). Although studies concerning reproductive outcomes are often criticized on the basis that the group with the endpoint under study will remember past events better than those in the referent group, studies that have tested specifically for recall bias have been unable to demonstrate its presence (*97–99*). However, potential for recall bias may increase as public concern about EM field exposures heightens. Studies that used employment records to obtain exposure information are not subject to problems with recall, but information on an individual's work habits that may be relevant to EM field exposures may not be available. Similarly, the assignment of exposure status by an industrial hygienist (*21,26,29,31,32*) based on occupational title or group information may not account adequately for individual variations in exposure within an occupational classification. Either limitation might result in misclassification of exposure status and a resultant dilution of measured effect estimates (*95*).

Other Design and Analytic Issues

Because all humans are exposed to electric and magnetic fields to some extent, the selection of a completely unexposed reference population is impossible. If a sizable proportion of the reference population is also exposed, ability to uncover risks associated with those exposures may be diminished. However, if information on multiple sources of exposure is available for both the case and referent populations, estimation of risks from selected exposures will be enhanced.

An examination of possible confounders in studies of risks from EM fields, although important, may be problematic for two reasons. First, because there is a lack of information on what the exposure is, it is difficult to identify other factors that are related to the exposure and increase the risk for the reproductive outcome of interest. A second problem is that, in general, there are not many established risk factors for adverse reproductive outcomes other than low birth weight.

Sample size is often a problem in studies of specific reproductive endpoints. Because some reproductive endpoints are rare, the practice of broadly grouping endpoints is often undertaken. However, this is of dubious biologic validity. This issue is of particular concern for some of the studies that investigated congenital malformations (*13,24*), but it also may be relevant to studies of spontaneous abortions that do not include karyotype information. Teratogenic agents appear to increase the risk for specific malformations. One would not expect that all types of malformations would be increased from low-level EM field exposures during pregnancy, nor would one expect to see all types of spontaneous abortions (i.e., early versus late or chromosomally normal versus abnormal) increased. The determination of sample size requires consideration of these biologic issues.

Future Research

Concerns regarding the potential for residential or occupational EM field exposures to result in adverse reproductive effects have been raised. Overall, epidemiologic evidence is lacking for a strong association between a woman's VDT use during pregnancy and for that pregnancy to end in spontaneous abortion. The relations noted thus far have, with the exception of Goldhaber et al. (*23*), shown about a 20% increased risk for VDT users. If this small increase reflects causality and is not due to uncontrolled confounding or to artifact, it would be important from a public health perspective, given the large number of women who use VDTs during pregnancy. Evidence for increased risks for other adverse reproductive end points from VDT use is equivocal. Some suggestive findings for malformations will require follow-up. Even increased risks for early spontaneous abortions (*20*), which may be indicative of a teratogenic effect if the frequency of malformed fetuses is greater among these earlier abortuses than among later abortuses (*100*), will need follow-up. Too few data exist to reach a conclusion about other reproductive health effects from VDT use.

To elucidate the potential relation between VDT use and spontaneous abortion risk, further investigations will require large numbers of study subjects, improved measures of exposure (e.g., direct measurement of field emissions from VDTs, distance from VDT, proximity to other sources of EM fields, orientation of worker to VDT), consideration of the heterogeneous nature of the end point studied (i.e., chromosomal versus nonchromosomal), inclusion of early as well as late spontaneous abortions, and consideration of the competing hypotheses related to physical and psychosocial stress (*101,102*).

Monson (*103*) has suggested that future investigations of reproductive risks in relation to VDT use will need to be prospective in their design. Although common adverse reproductive effects such as spontaneous abortion or reduced birth weight may be amenable to a prospective design, the rarity of specific malformations (e.g., prevalences at birth of 1 per 1000 for common malformations), argues against the prospective approach based on pragmatic and economic considerations. Unless large occupational cohorts are available, prospective studies of malformation risk resulting from maternal VDT use would be hard to justify.

Clearly, the heightened public awareness about possible reproductive risks from VDT use may introduce difficulty in obtaining accurate exposure information from retrospective studies. Therefore, retrospective studies involving reproductive end points, such as congenital malformations, will need to be creative in validating exposure information. For example, in addition to collecting detailed information from a parent about VDT use, studies might attempt to acquire information from employers about a parent's work schedule and conditions surrounding VDT use. Other reproductive end points like low birthweight and intrauterine growth retardation may be studied economically by using reference groups from retrospective investigations of malformations and spontaneous abortions, as some investigators have done (*20,21,32*).

The epidemiologic evidence regarding the potential reproductive risks from exposure to low-level EM fields other than VDTs is even more lacking (Table 1). These studies reflect a variety of endpoints and field exposure sources, yet taken as a whole, they do not suggest strong associations between adverse reproductive effects and EM field exposure. However, it seems prudent to resist drawing conclusions about the potential reproductive risks from low-level field exposures until additional investigations are performed.

To date, only three studies have investigated residential exposure to EM fields and adverse reproductive effects: two explored exposures from electric bedwarmers and specific congenital malformations (*16*) and fetal loss (*14*), and one examined pregnancy loss and power-frequency field exposures from home heating (*15*). All three studies focused only on maternal exposures. In contrast, most occupational studies have investigated paternal exposure to EM fields.

There is a need for investigations of other sources of residential EM field exposures (e.g., transmission and distribution lines, electric appliances and power tools, or electric train lines). In addition, future research needs to incorporate additional methods, such as personal monitors or spot measurements or wiring codes, for assessing the exposure (or its surrogate) that appears to be related to childhood cancers (*1,3*). Retrospective studies could provide much needed descriptive data (primarily from the referent group) on the prevalence of exposures from these and other sources. Without this information, the choice of exposed groups to follow in prospective studies of adverse reproductive effects would be difficult to make.

Although there is little support for paternally mediated adverse reproductive effects (*104*), some studies have suggested altered sex ratios among the offspring of exposed fathers (*12,13*), and at least one study has suggested an association between paternal EM field exposures and childhood cancers in offspring (*2*). Thus, paternal residential exposure to EM fields, from a variety of sources and preconceptional time periods, may be an avenue worthy of additional investigation.

Future occupational studies might focus on maternal exposures to EM fields. Job exposure matrices, as used for studies of other occupational exposures and disease endpoints, would be useful to develop. However, to avoid errors associated with exposure misclassification, a sizable amount of exposure-based research would first have to be completed to ensure that occupations were accurately classified with respect to EM field exposures.

No study has considered any adverse reproductive health effect from combined home and work exposures to low-level fields. Schnorr et al. (*25*) reported that VDT users had abdominal measurements in the range of exposures that would be experienced at home, but no study has attempted to estimate a person's total exposure to low-level fields. Certainly, studies that have not found associations between EM field exposures and adverse reproductive effects may have failed to do so because total exposures were not considered.

Given what little is known about the attributes of the exposures being estimated, it seems unlikely that more descriptive types of epidemiologic study designs employing existing data sources will be very revealing. For example, studying seasonal variations for reproductive endpoints based on the assumptions about risk used by others (*14,15*) is simple if existing data sources can be utilized, but the interpretation is quite difficult given the nonspecificity of the analysis and the underlying assumptions about seasonal variation in risk.

The experimental work thus far offers little direction to epidemiologists interested in reproductive health effects, except it shows that numerous aspects about exposures appear to be relevant (e.g., waveform, frequency) and that growth may be altered by EM fields. Altered growth may be mechanistically important to many adverse reproductive endpoints, but it is nonspecific. Given that EM field exposures do not seem to result in mutation, candidate study end points would not include chromosomal abnormalities. Endpoints in which abnormal embryonic growth—such as congenital malformations not associated with chromosomal anomalies, intrauterine growth retardation, and perhaps chromosomally normal spontaneous abortions—is thought to be an underlying mechanism that would be suitable for study. To study the latter group, all abortions would have to be karyotyped, which might be prohibitively expensive.

Exposure assessment is a crucial issue, although it is not clear what the ideal, or perhaps even the most relevant, exposure measure should be. It seems important that both occupational and residential exposures be considered. Community concern has emphasized the dangers of high tension power lines. Although these exposure sources are less prevalent than many others, they should not be ignored. For research to be revealing, emphasis needs to be placed on developing exposure assessment techniques that are adaptable for epidemiologic investigations of reproductive outcomes and that will not suffer from large measurement errors and consequently reduce our ability to identify positive relations that may exist (e.g., items that can be included in questionnaires and can discriminate levels of field exposures). After more information how a person's exposure changes over time is available, direct measures obtained after the critical period of fetal development may be useful. Consideration of the many methodologic issues discussed above (e.g., sufficient sample size) may allow for greater specificity of findings regarding these common low-level field exposures.

In general, there are many concerns about reproductive health. The possibility that the normal reproductive process may be perturbed by EM field exposures has heightened these concerns further. Given the lack of epidemiologic data to address these concerns and the experimental evidence that certainly does not argue against a possible effect from these exposures, there seems to be sufficient justification for additional study in this area.

REFERENCES

1. Wertheimer N, Leeper E. Electrical wiring configuration and childhood cancer. Am J Epidemiol 109:273–284 (1979).
2. Spitz MR, Johnson CC. Neuroblastoma and paternal occupation: a case-control analysis. Am J Epidemiol 121:924–929 (1985).
3. Savitz DA, Wachtel H, Barnes FA, John EM, Tvrdik JG. Case–control study of childhood cancer and exposure to 60-Hz magnetic fields. Am J Epidemiol 128:21–38 (1988).
4. Savitz DA, John EM, Kleckner RC. Magnetic field exposure from electric appliances and childhood cancer. Am J Epidemiol 131:763–773 (1990).
5. Fulton JP, Cobb S, Preble L, Leone L, Forman E. Electrical wiring configurations and childhood leukemia in Rhode Island. Am J Epidemiol 111:292–296 (1980).
6. Tomenius L. 50-Hz electromagnetic environment and the incidence of childhood tumors in Stockholm County. Bioelectromagnetics 7:191–207 (1986).
7. Savitz DA, Calle EE. Leukemia and occupational exposure to electromagnetic fields: review of epidemiologic surveys. J Occup Med 29:47–51 (1987).
8. Wertheimer N, Leeper E. Adult cancer related to electrical wires near the home. Int J Epidemiol 11:345–355 (1982).
9. Stern RM. Cancer incidence among welders: possible effects of exposure to extremely low frequency electromagnetic radiation (ELF) and to welding fumes. Environ Health Perspect 76:221–229 (1987).
10. Thomas TL, Stolley PD, Stemhagen A, Fontham ETH, Bleecker ML, Stewart PA, Hoover RN. Brain tumor mortality risk among men with electrical and electronics jobs: a case–control study. J. Natl Cancer Inst 79:233–238 (1987).
11. Severson RK, Stevens RG, Kaune WT, Thomas DB, Heuser L, Davis S, Sever LE. Acute nonlymphocytic leukemia and residential exposure to power-frequency magnetic fields. Am J Epidemiol 128:10–20 (1988).
12. Knave B, Gamberale F, Bergstrom S, Birke E, Iregren A, Kolmodin-Hedman B, Wennberg A. Long-term exposure to electric fields: a cross-sectional epidemiologic investigation of occupationally exposed workers in high voltage substations. Scand J Work Environ Health 5:115–125 (1979).
13. Nordstrom S, Birke E, Gustavsson L. Reproductive hazards among workers at high voltage substations. Bioelectromagnetics 4:91–101 (1983).
14. Wertheimer N, Leeper E. Possible effects of electric blankets and heated waterbeds on fetal development. Bioelectromagnetics 7:13–22 (1986).
15. Wertheimer N, Leeper E. Fetal loss associated with two seasonal sources of electromagnetic field exposure. Am J Epidemiol 129:220–224 (1989).
16. Dlugosz L, Vena J, Zielezny M, Byers T, Roth G, Marshall E. Congenital malformations and periconceptional use of electric blankets and heated waterbeds: a register based case–control study. Am J Epidemiol 132:763 (1990).
17. Buiatti E, Barchielli A, Geddes M, Nastasi L, Kriebel D, Franchini M, Scarselli G. Risk factors in male fertility: a case–control study. Arch Environ Health 39:266–270 (1984).
18. Hemminki K, Niemi ML, Koskinen K, Vainio H. Spontaneous abortions among women employed in the metal industry in Finland. Int Arch Occup Environ Health 47:53–60 (1980).
19. Nielsen CV, Brandt LPA. Spontaneous abortion among women using video display terminals. Scand J Work Environ Health 16:323–328 (1990).
20. Windham GC, Fenster L, Swan SH, Neutra RR. Use of video display terminals during pregnancy and the risk of spontaneous abortion, low birthweight, or intrauterine growth retardation. Am J Ind Med 18:675–688 (1990).
21. Nurminen T, Kurppa K. Office employment, work with video display terminals, and course of pregnancy. Reference mothers' experience from a Finnish case-referent study of birth defects. Scand J Work Environ Health 14:293–298 (1988).
22. Bryant HE, Love EJ. Video display terminal use and spontaneous abortion risk. Int J Epidemiol 18:132–138 (1988).
23. Goldhaber MK, Polen MR, Hiatt RA. The risk of miscarriage and birth defects among women who use visual display terminals during pregnancy. Am J Ind Med 13:695–706 (1988).
24. McDonald AD, Cherry NM, Delorme C, McDonald JC. Visual display units and pregnancy: evidence from the Montreal survey. J Occup Med 28:1226–1231 (1986).
25. Schnorr TM, Grajewski BA, Hornung RW, Thun MJ, Egeland GM, Murray WE, Conover DL, Halperin WE. Video display terminals and the risk of spontaneous abortion. New Engl. J Med 324:727–733 (1991).
26. Ericson A, Kallen B. An epidemiological study of work with video screens and pregnancy outcome: I. A registry study. Am J Ind Med 9:447–457 (1986).
27. Ericson A, Kallen B. An epidemiological study of work with video screens and pregnancy outcome: II. A case-control study. Am J Ind Med 9:459–475 (1986).
28. Butler WJ, Brix KA. Reproductive outcomes of video display workers. Invited paper. Las Vegas, NV: American Public Health Association, 1986.
29. Kurppa K, Holmberg PC, Rantala K, Nurminen T, Saxen L. Birth defects and exposure to video display terminals during pregnancy. A Finnish case-referent study. Scand J Work Environ Health 11:353–356 (1985).
30. Brandt LPA, Nielson CV. Congenital malformations among children of women working with video display terminals. Scand J Work Environ Health 16:329–333 (1990).
31. Kurppa K, Holmberg PC, Rantala K, Nurminen T. Birth defects and video display terminals. Lancet 2:1339 (1984).
32. Tikkanen J, Heinonen OP, Kurppa K, Rantala K. Cardiovascular malformations and maternal exposure to video display terminals during pregnancy. Eur J Epidemiol 6:61–66 (1990).
33. Bjerkedal T, Egenaes J. Video display terminals and birth defects: a study of pregnancy outcomes of employees of the Postal-Giro-Center, Oslo, Norway. In: Work with display units 86 (Knave B, Wideback PG, eds). Amsterdam: Elsevier Science Publishers, 1987; 111–114.
34. Eckert EE. Sudden unexpected death of infants and electromagnetic fields. German Med Klin 71:1500–1505 (1976).
35. Lewis MJ, Esterman AJ, Dorsch MM. A survey of the health consequences to females of operating visual display units. Community Health Stud 6:130–134 (1982).
36. Westerholm P, Ericson A. Pregnancy outcomes and VDU-work in a cohort of insurance clerks. In: Work with display units 86 (Knave B, Wideback PG, eds). Amsterdam: Elsevier Science Publishers, 1987; 104–110.
37. Savitz DA, Pearce NE, Poole C. Methodological issues in the epidemiology of electromagnetic fields and cancer. Epidemiol Rev 11:59–78 (1989).
38. Coleman M, Beral V. A review of epidemiological studies of the health effects of living near or working with electricity generation and transmission equipment. Int J Epidemiol 17:1–13 (1988).
39. Aldrich TE, Easterly CE. Electromagnetic fields and public health. Environ Health Perspect 75:159–171 (1987).
40. Ahlbom A. A review of the epidemiologic literature on magnetic fields and cancer. Scand J Work Environ Health 14:337–343 (1988).
41. Meyer RE, Aldrich TE, Easterly CE. Effects of noise and electromagnetic fields on reproductive outcomes. EHP 81:193–200 (1989).
42. Peterson DR. Sudden infant death syndrome. In: Perinatal epidemiology (Bracken MB, ed). New York: Oxford, 1984; 339–356.
43. Pastides H, Calabrese EJ, Hosmer DW Jr, Harris DR. Spontaneous abortion and general illness symptoms among semiconductor manufacturers. J Occup Med 30:543–551 (1988).
44. Bergqvist UOV. Video display terminals and health. A technical and medical appraisal of the state of the art. Scand J Work Environ Health (Suppl) 10:1–87 (1984).
45. Marcus M. Epidemiologic studies of VDT use and pregnancy outcome. Repro Toxicol 4:51–56 (1990).
46. Windham GC. Does using VDTs during pregnancy affect the fetus? Health Environ Digest 4:1–3 (1990).
47. Juutilanen J. Effects of low-frequency magnetic fields on embryonic development and pregnancy. Scand J Work Environ Health 17:149–158 (1991).
48. Delgado JMR, Leal J, Monteagudo JL, Gracia MG. Embryological changes induced by weak, extremely low frequency electromagnetic fields. J Anat 134:533–551 (1982).
49. Ubeda A, Leal J, Trillo MA, Jimenez MA, Delgado JMR. Pulse shape of magnetic fields influences chick embryogenesis. J Anat 137:513–515 (1983).
50. Martin AH. Magnetic fields and time dependent effects on development. Bioelectromagnetics 9:393–396 (1988).

51. Juutilainen JP, Laara E, Saali K. Relationship between field strength and abnormal development in chick embryos exposed to 50 Hz magnetic fields. Int J Radiat Biol 52:787–793 (1987).
52. Tribukait B, Cekan E, Paulsson LE. Effects of pulsed magnetic fields on embryonic development in mice. In: Work with display units 86 (Knave B, Wideback PG, eds). Amsterdam: Elsevier Science Publishers, 1987; 129–134.
53. Juutilanen JP, Haari M, Saali K, Lahtinen T. Effects of 100-Hz magnetic fields with various waveforms on the development of chick embryos. Radiat Environ Biophys 25:65–74 (1986).
54. Blackman CF, House DE, Benane SG, Joines WT, Spiegel RJ. Effect of ambient levels of power-line-frequency electric fields on a developing vertebrate. Bioelectromagnetics 9:129–140 (1988).
55. Cameron IL, Hunter KE, Winters WD. Retardation of embryogenesis by extremely low frequency 60-Hz electromagnetic fields. Physiol Chem Phys Med NMR 17:135–138 (1985).
56. Rommereim DN, Kaune WT, Buschbom RL, Phillips RD, Sikov MR. Reproduction and development in rats chronically exposed to 60-Hz electric fields. Bioelectromagnetics 8:243–258 (1987).
57. Juutilanen J, Saali K. Development of chick embryos in 1-Hz to 100-Hz magnetic fields. Radiat Environ Biophys 25:135–140 (1986).
58. Sikov MR, Rommereim DN, Beamer JL, Buschbom RL, Kaune WT, Phillips RD. Developmental studies of Hanford miniature swine exposed to 60-Hz electric fields. Bioelectromagnetics 8:229–242 (1987).
59. Berman E, Chacon L, House D, Koch BA, Koch WE, Leal J, Lovtrup S, Mantiply E, Martin AH, Martucci GI, Mild KH, Monahan JC, Sandstrom M, Shamsaifar K, Tell R, Trillo MA, Ubeda A, Wagner P. Development of chicken embryos in a pulsed magnetic field. Bioelectromagnetics 11:169–187 (1990).
60. McGivern RF, Sokol RZ, Adey WR. Prenatal exposure to a low-frequency electromagnetic field demasculinizes adult scent marking behavior and increases accessory sex organ weights in rats. Teratol 41:1–8 (1990).
61. Soeradi O, Tadjudin MK. Congenital anomalies in the offspring of rats after exposure of the testis to an electrostatic field. Int J Andrology 9:152–160 (1986).
62. Martin AH. Variable effects of electromagnetic fields on child development. Teratology 43:433 (1991).
63. Stuchly MA, Ruddick J, Villaneuve D, Robinson K, Reed B, Lecuyer DW, Tan K, Wong J. Teratological assessment of exposure to time-varying magnetic field. Teratology 38:461–466 (1988).
64. Maffeo S, Brayman AA, Miller MW, Carstensen EL, Ciaravino V, Cox C. Weak, low-frequency electromagnetic fields and chick embryogenesis: failure to reproduce findings. J Anat 157:101–104 (1988).
65. Sisken BF, Fowler J, Mayaud C, Ryaby JP, Ryaby J, Pilla AA. Pulsed electromagnetic fields and normal chick development. J Bioelectricity 5:25–34 (1986).
66. Maffeo S, Miller MW, Carstensen EL. Lack of effect of weak, low-frequency electromagnetic fields on chick embryogenesis. J Anat 139:613–618 (1984).
67. Sandstrom M, Mild KH, Lovtrup S. Effects of weak pulsed magnetic fields on chick embryogenesis. In: Work with display units 86. (Knave B, and Wideback PG, eds), Amsterdam: Elsevier Science Publishers, 1986; 135–140.
68. Seto YJ, Majeau-Chargois D, Lymangrover JR, Dunlap WP, Walker CF, Hsieh ST. Investigation of fertility and in utero effects in rats chronically exposed to a high-intensity 60-Hz electric field. IEEE Trans Biomed Eng BME 31:693–702 (1984).
69. Rommereim DN, Rommereim RL, Sikov MR, Buschbom RL, Anderson LE. Reproduction, growth, and development of rats during chronic exposure to multiple field strengths of 60-Hz electric fields. Fund Appl Toxicol 14:608–621 (1990).
70. Berman E. The developmental effect of pulsed magnetic fields on animal embryos. Reprod Toxicol 4:45–49 (1990).
71. Leal J, Trillo MA, Ubeda A, Abraira V, Shamsaifar K, Chacon L. Magnetic environment and embryonic development: a role of the Earth's field. IRCS Med Sci 14:1145–1146 (1986).
72. Adey WR. Tissue interaction with nonionizing electromagnetic fields. Physiol Rev 61:435–514 (1981).
73. Adey WR. Joint actions of environmental nonionizing electromagnetic fields on chemical pollution in cancer promotion. Environ Health Perspect 86:297–305 (1990).
74. Greenbaum B, Goodman EM, Maron MT. Magnetic field effects on mitotic cycle length physarum. Eur J Cell Biol 27:156–160 (1982).
75. Byus CV, Pieper SE, Adey WR. The effects of low-energy 60-Hz enviromental electromagnetic fields upon the growth-related enzyme ornitine decarboxylase. Carcinogenesis 8:1385–1389 (1987).
76. Liboff AR, Williams T, Strong DM, Wistar R. Time-varying magnetic fields: effect on DNA synthesis. Science 223:818–820 (1984).
77. Blackman CF, Benane SG, House DE, Joines WD. Effects of ELF (1-120Hz) and modulated (50 Hz) RF fields on the efflux of calcium ions from brain tissue in vitro. Bioelectromagnetics 6:1–11 (1985).
78. Zusman I, Yaffe P, Pinus H, Ornoy A. Effects of pulsing electromagnetic fields on the prenatal and postnatal development in mice and rats: in vivo and in vitro studies. Teratol 42:157–170 (1990).
79. Kline JK. Maternal occupation: effects on spontaneous abortions and malformations. Occup Med Rev 1:381–403 (1986).
80. Axelson G, Rylander R. Validation of questionnaire reported miscarriage, malformation, and birth weight. Int J Epidemiol 13:94–98 (1984).
81. Rasmussen SA, Mulinare J, Khoury MJ, Maloney EK. Evaluation of birth defect histories obtained through maternal interviews. Am J Hum Genet 46:478–485 (1990).
82. Wilcox AJ, Weinberg CR, O'Connor JF, Baird DD, Schlatterer JP, Canfield RE, Armstrong EG, Nisula BC. Incidence of early loss of pregnancy. New Engl. J Med 319:189–194 (1988).
83. Hatch M, Marcus M. Occupational exposures and reproduction. In: Reproductive and perinatal epidemiology (Kiely M, ed). Boca Raton, FL: CRC Press, 1991; 131–142.
84. U.S. Congress, Office of Technology Assessment. Biological effects of power frequency electric and magnetic fields. Background paper. OTA-BP-E-53. Washington, DC: U.S. Government Printing Office, 1989.
85. Logue JN, Hamburger S, Silverman PM, Chiacchierini RP. Congenital anomalies and paternal occupational exposure to shortwave, microwave, infrared, and acoustic radiation. J Occup Med 27:451–452 (1985).
86. Kurppa K, Holmberg PC, Hernberg S, Rantala K, Riala R, Nurminen T. Screening for occupational exposures and congenital malformations: preliminary results from a nationwide case-referent study. Scand J Work Environ Health 9:89–93 (1983).
87. Lancranjan I, Maicanescu M, Rafaila E, Klepsch I, Popescu HI. Gonadic function in workmen with long-term exposure to microwaves. Health Physics 29:381–383 (1975).
88. Barron CI, Baraff AA. Medical considerations of exposure to microwaves (radar). JAMA 168:1194–1199 (1958).
89. Kallen B, Malmquist G, Moritz U. Delivery outcome among physiotherapists in Sweden: is non-ionizing radiation a fetal hazard? Arch Environ Health 37:81–85 (1982).
90. Wilson J, Fraser FC. Handbook of Teratology, vol 1: general principles and etiology. New York: Plenum Press, 1977.
91. Easterly CE. Cancer link to magnetic field exposure: a hyporhesis. Am J Epidemiol 114:169–174 (1981).
92. Janerich DT, Polednak AP. Epidemiology of birth defects. Epidemiol Rev 5:16–37 (1983).
93. Kramer MS. Determinants of low birthweight: methodological assessment and meta-analysis. Bull WHO 65:663–737 (1987).
94. Greenberg RS, Shuster JL Jr. Epidemiology of cancer in children. Epidemiol Rev 7:22–48 (1985).
95. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. Am J Epidemiol 105:488–495 (1977).
96. Schlesselman JJ. Case–control studies. Oxford: Oxford University Press, 1982.
97. Zierler S, Rothman KJ. Congenital heart disease in relation to maternal use of Bendectin and other drugs in early pregnancy. New Engl. J Med 313:347–352 (1985).
98. Tilley BC, Barnes AB, Bergstrath E, Labarthe D, Noller KL, Colton T, Adam E. A comparison of pregnancy history recall and medical records. Am J Epidemiol 121:269–281 (1985).
99. Klemetti A, Saxen L. Prospective versus retrospective approach in the search for environmental causes of malformations. Am J Public Health 57:2071–2075 (1967).
100. Creasy MR, Alberman ED. Congenital malformations of the central nervous system in spontaneous abortions. J Med Genet 13:9–16 (1976).
101. Johansson G, Aronsson G. Stress reactions in computerized administrative work. J Occup Behavior 5:159–181 (1984).
102. Mackay C. The alleged reproductive hazards of VDUs. Work & Stress 1:49–57 (1987).
103. Monson RR. Editorial commentary: epidemiology and exposure to electromagnetic fields. Am J Epidemiol 131:774–775 (1990).
104. Pearn JH. Teratogens and the male: an analysis with special reference to herbicide exposure. Med J Aust 2:16–20 (1983).