Toward an Integrated Approach to Molecular Epidemiology

Christine B. Ambrosone and Fred F. Kadlubar

The emergence of "molecular epidemiology" as a scientific approach within the fields of epidemiology and toxicology has led to spirited discussion within the biomedical community, particularly in the area of cancer research. At scientific meetings and in peer-reviewed journals, numerous issues have been raised not only with regard to the practice of molecular epidemiology, but also with regard to its role in traditional epidemiology, toxicology, and risk assessment. Furthermore, the utility of information gleaned from such studies and the implications for public health have been the subject of considerable debate. Conceptual differences in how one views the function of epidemiologic and laboratory research may be reflected in discussions on the merits of molecular epidemiology. This commentary reviews some of the prevailing attitudes toward molecular epidemiology, with the goal of identifying areas of concern and suggesting means of achieving harmonization. The need for cross-training of epidemiologists and laboratory scientists is discussed, and suggestions are made for building successful collaborative relations across disciplines. Am J Epidemiol 1997;146:912-18.

Definitions and discussions of molecular epidemiology appear frequently in the literature, and the utilization of biomarkers in large, population-based studies is increasingly becoming more common. With the growing use of a molecular approach to epidemiologic research, conflicts regarding the shift of focus from the population to the individual often arise within the scientific community. Furthermore, the use of biomarkers in studies by epidemiologists and the application of molecular or biochemical assays to human populations by clinicians or laboratory scientists who are untrained in each other's disciplines may result in seriously flawed studies with improper interpretation of results. This may lead not only to presentations of incorrect data and erroneous conclusions, but also to skepticism and widespread disregard of results obtained from properly conducted molecular epidemiologic studies. We suggest that if molecular epidemiology is to receive recognition as a credible approach to research, it is necessary that it not be just "the employment of biomarkers in epidemiologic studies." Rather, there is a need for intellectual development of epidemiologists and laboratory scientists across disciplines and a well-defined collaborative relationship between these researchers.

THE ROLE OF EPIDEMIOLOGY

Epidemiology is a growing and evolving discipline that cannot escape being affected by changes and advances in scientific knowledge and technology. It is often mentioned that this scientific approach resulted from a realization among epidemiologists that substantive advances in our understanding of disease etiology were unlikely to be obtained from better questionnaires, and at the same time, laboratory scientists began to realize that humans are not always comparable with inbred strains of rodents and that there are interindividual differences in disease susceptibility as a consequence of genetic polymorphisms, environmental influences, or both. As illustrated in a number of papers in a recent issue of the American Journal of Public Health (1-4), it is clear that there is concern about the extent to which epidemiology should change and about what the goals and priorities of epidemiologic research should be. Key issues appear to be the apparent movement away from a public health orientation (1), marked by a shift of focus from the population to the individual, and the neglect of the role of socioeconomic factors in disease etiology (3). Concern that "risk factor" epidemiology leads to a biophysical reductionism (5) and "produces interventions that can be harmful" (3, p. 679) is further extended to...
molecular epidemiology, which in the view of individuals with this perspective epitomizes "dangerous trends" in epidemiology. For these researchers, molecular epidemiology is but a symptom of the growing use of epidemiology to unravel disease etiology, rather than an approach to address issues of disease occurrence within populations.

Is it not possible to allow epidemiology to grow on its own terms and to sprout into different directions? For some epidemiologists who often come from social science backgrounds, epidemiologic analyses at the population level may be useful to identify modifiable disease risk factors, particularly those related to specific demographic subpopulations, that may be addressed by society as a whole. There are also numerous researchers whose training and perspective stems from clinical medicine, the basic biomedical sciences, or both, and these epidemiologists may be most interested in elucidating the natural history of disease. Particularly for cancer research, progress in our understanding of carcinogenesis may require the harnessing of epidemiology and molecular toxicology as workhorses to 1) elucidate human variability in susceptibility to known risk factors, and 2) identify relations between genetic polymorphisms and the biologic effects of sustained interaction with exogenous and endogenous exposures. It is incongruous that these ends should be at odds with each other and that researchers from both disciplines cannot integrate approaches and endpoints in epidemiologic and laboratory studies. This ideologic conflict may be at the base of resistance or hostility toward a molecular epidemiologic approach to research.

Resolution of these issues within the epidemiologic research community may be necessary before there is a general acceptance of molecular epidemiology as a useful and valid tool with which to study disease etiology. Surely the discipline is large enough to accommodate epidemiologists who come from a public health perspective and those whose primary interest is to use epidemiology as a vehicle for the study of disease etiology. If it can be agreed that epidemiology should play a role in studies of disease etiology, then not only risk factors but also biologic plausibility and host factors such as gene-environmental interactions could be incorporated into epidemiologic studies, along with the development and refinement of more appropriate biochemical and molecular biologic techniques.

THE ROLE OF TOXICOLOGY

Animal studies are most often employed to identify carcinogens and to assess the cancer risk of naturally occurring and synthetic chemicals to which humans may be exposed. With the more recent emphasis on identifying the mechanisms of action of these carcinogens in relation to the cancer process, it now becomes of paramount importance to determine whether or not such mechanisms are operative in humans at levels of human exposure to these substances. Comparative toxicology, namely the extrapolation of animal studies to humans, must also meet the challenge of providing relevant biomarkers of exposure, susceptibility, and effect that can be used across species to improve human risk assessment and support public health rules and guidelines issued by government regulatory agencies.

How molecular epidemiology can add to our understanding of disease

Numerous papers and books have been written in the last several years regarding the application of biomarkers to molecular epidemiologic studies of disease (6-16). Many of these texts have discussed molecular epidemiology in the context of cancer research and have demonstrated the utility of biomarkers in the elucidation of the cancer process in the continuum between exposure and disease outcome. This has generally consisted of 1) biomarkers of exposure, including those of internal, biologically effective, and target-tissue dose; 2) biomarkers of susceptibility, including polymorphisms in genes involved in carcinogen metabolism, in DNA repair, and in controlling cell growth; and 3) biomarkers of effect, such as early genetic alterations and modulation of immunologic and nutritional status, that lead to tumorigenesis. The above cited papers provide extensive, detailed information regarding the use of such biomarkers in epidemiologic studies, and it is not the purpose of this commentary to present basic information regarding each of these categories of biomarkers. However, a few points will be made regarding the general advances in our understanding of the cancer process that have been made by the use of molecular epidemiologic studies.

Identification of susceptible subgroups. In a recent review, Perera (12) presented convincing arguments that there is sufficient evidence from molecular epidemiologic studies to conclude that individuals are not uniformly susceptible to carcinogenic exposures. It has been known for some time that individuals vary from 10 to several hundredfold in their response to pharmaceutical agents (17), and data indicate that metabolic variability also affects carcinogenicity. However, this model of variability in susceptibility has not been consistently applied to epidemiologic or toxicologic studies. Except for factors such as age, gender, ethnicity, and menopausal status, which are often entered in regression models as possible confounding or
Identification of etiologic agents in carcinogenesis. Identification of subgroups of the population who may be more vulnerable to the effects of certain carcinogens may have important implications beyond risk assessment. By identification of an increased risk in certain subgroups, disease risk factors may be more clearly elucidated. This concept was demonstrated in the recent preliminary report of a clear association between breast cancer and cigarette smoking among women with a slower form of a detoxifying enzyme, N-acetyltransferase (22). As reviewed by Palmer and Rosenberg (23), there have been numerous studies of smoking and breast cancer with inconsistent results, and overall, tobacco smoke has not been considered a breast cancer risk factor. However, if results of the above-cited study are confirmed, this will illustrate that heterogeneity in response to carcinogenic exposures may dilute or mask the true effects among susceptible populations. This may be true for other risk factors and cancers at other organ sites as well. Assessment of exposure-disease associations among populations who may be susceptible, based on variability in metabolism, DNA repair, immunologic status, or other host factors, may more clearly elucidate risk factors that may be targeted for preventive public health initiatives.

Identification of etiologic agents by characterization of DNA adducts. DNA adducts, the products of reactive chemicals and DNA that contain one or more covalent bonds between the moieties (24), are often evaluated in molecular epidemiologic studies as markers of biologically effective dose. It has been demonstrated that the presence of adducts may predict a carcinogenic response (25), but the association between adduct formation and carcinogenesis may depend upon the location of the adducts and the genes affected (24), as well as on cell proliferation and immunologic, hormonal, and nutritional status. Furthermore, there are many intervening variables in the cancer process, and clearly, substantial work is required to elucidate the complex relation between adduct levels and cancer risk. For these reasons, as well for issues of temporality and expense, DNA or surrogate tissue adduct measurements may not always be optimal biomarkers of exposure for use in epidemiologic studies, as suggested by some critics (13). However, the characterization of DNA adducts in target tissue may be extremely useful for identification of etiologic factors in specific cancers. As exposure assessment from interview data can be used to identify disease risk factors in epidemiologic studies, so too can tissue-specific adducts lend clues to etiologic factors in disease. For example, identification of tobacco smoke-related adducts in the breast suggest that tobacco smoke could be a breast cancer risk factor (26, 27). Furthermore, the characterization of endogenous malondialdehyde-DNA adducts derived from oxidative stress and lipid peroxidation, in addition to smoking-related adducts in the same tissue, yield further insights into breast cancer etiology (28). Similarly, recent studies have identified adducts in prostate tissue, a disease for which little is known regarding risk factors (29); thus, the characterization of these adducts may greatly enhance our understanding of prostate carcinogenesis. An illustration of the utility of DNA adducts in identification of etiologic agents in carcinogenesis is the recent work in which benzo-(a)pyrene adducts were identified at specific codons in the p53 gene that are the major mutational hot spots in human lung cancer, clearly providing a direct link between a tobacco smoke carcinogen and lung cancer etiology (30). In the future, transitional studies may be used to evaluate levels of adducts in human tissue in relation to exposure data and to polymorphisms in genes involved in carcinogen and hormone metabolism. Studies such as these should provide powerful insights into tissue-specific carcinogenesis.

TOWARD AN INTEGRATED APPROACH TO MOLECULAR EPIDEMIOLOGY

Molecular epidemiology is novel in that it means many things to many people. The term "molecular epidemiology" has been applied to case-control or cohort studies performed by traditional epidemiologists that include biomarkers of exposure, susceptibility, or effect. It is also used by laboratory scientists to...
describe studies of metabolism or other biochemical processes in human subjects. Finally, clinicians may term a study molecular epidemiology when it applies molecular findings to patient health outcomes. Researchers from each of these perspectives may make important contributions to the literature, yet each working independently of the other may result in sometimes serious flaws in study design, implementation, and interpretation. Interdisciplinary, collaborative studies may also produce more coherent and global insights into disease etiology rather than fragmented, narrowly limited pieces of information from multiple disciplines. However, advancement of the field and the performance of valid studies is clearly dependent on cross-training of scientists in each other's disciplines.

Biochemistry and molecular biology for the epidemiologist. There is ongoing debate within departments of public health and preventive medicine regarding the necessity of basic training in biomedical sciences for chronic disease epidemiologists. For years, it had not been considered necessary to understand the pathophysiology of disease to evaluate exposure-disease associations (1). For example, John Snow, often hailed as the "Father of Epidemiology," traced the source of cholera to use of water from a particular pump (31). There was little understanding of contamination from poor sanitation and the effects that bacteria in the water would have on public health; removal of the pump handle alleviated the problem. However, as the complexity of disease becomes apparent, it would require blinders not to acknowledge that chronic diseases, such as cancer, diabetes, and cardiovascular disease, are multifactorial. Unless we opt to disregard these findings and continue to evaluate single exposures in relation to disease outcome, it is necessary to be cognizant of the plethora of factors that influence the initiation and progression of disease. Particularly for epidemiologists who use biomarkers in their studies, training and basic skills in biochemistry and molecular biology are becoming requisites. For the epidemiologist who may be interested in evaluating gene-environment interactions, it is further suggested that training be acquired in molecular genetics and toxicology. While collaboration with a skilled molecular biologist or toxicologist is essential to a successful molecular epidemiologic study, it is also important that the epidemiologist understands the underpinnings of the laboratory assays and is able to evaluate the genetic results and interpret them in a meaningful and valid way. For the cancer epidemiologist, familiarity with principles of oncology, toxicology, and molecular genetics will also enhance one's creative abilities to construct etiologic hypotheses regarding gene-environment interactions in cancer and other diseases.

Principles of study design for the laboratory scientist and the clinician. Enthusiasm for and acceptance of molecular epidemiology, particularly for cancer research, may be greatest among scientists whose training comes from the laboratory. In animal and in vitro models, these scientists have studied chemical carcinogenesis and clearly understand the role of carcinogens and molecular alterations involved in the cancer process. To these researchers, the application of such findings to human population studies may be the next logical step in the validation and extrapolation from laboratory models to humans. These studies may be hampered, however, by the lack of training in methods of epidemiologic and statistical design. Unless one has had the experience of observing bias in data resulting from improper choice of a control group, or at least has done extensive reading on the subject, such issues may be disregarded in transitional studies. Similarly, other factors such as the use of validated questionnaires, timing of interviews in relation to diagnosis of disease, and proper control for possible confounding factors or those that may cause effect modification could have profound effects on the validity of study results. In studies of genetic polymorphisms, it may also be essential to gather exposure data to evaluate gene-environment interactions rather than simply to evaluate the effects of the genotype on disease risk unless the study is of a high-risk (high-exposure) population. Thus, it is important for laboratory scientists and clinicians to be aware of important study design issues and, ideally, to consult with an epidemiologist before embarking on data collection.

Statistical consideration of sample size and power. Perhaps the most important issue in molecular epidemiologic studies, and the most difficult to address, is the appropriate statistical consideration of sample size and power. This is an issue of which researchers without an epidemiologic background are often unaware or tend to dismiss as unimportant. Epidemiologists, on the other hand, may be only too well aware of the importance of an adequate sample size, but are limited by funds and subject accrual. Inadequate power is likely to result not just because the molecular assays are expensive and thus, restrict the number of subjects recruited, but because the means of analysis automatically cuts the population in half or, in many instances, well below that, in stratified analyses. In many cases, genetic polymorphisms alone do not increase risk, but act in concert with environmental exposures, as in the case of NAT2, breast cancer, and smoking cited previously (22). In such instances, cases and controls are stratified by genotype, and associa-
tions between the risk factor and disease status are evaluated separately within each strata. Clearly, even in large studies, the numbers of subjects in each cell will be drastically reduced, affecting the odds of a type I or type II error. Although Fisher’s exact test may be used to examine significant associations in 2 × 2 tables with expected small numbers, there are few analytic approaches with which to address these problems. Some statistical approaches have been proffered to be used in molecular epidemiologic studies, such as case-series analyses, and these provide important new contributions to the literature (32–34).

There is a rapidly growing understanding of the genetic basis of biologic functioning, spurred on by the Human Genome Project as well as by explosive advances in technology. There is little doubt that in the coming years, numerous genes and polymorphisms in those genes will be identified, including those involved in biotransformation of endogenous and exogenous substrates, cell cycle control, DNA repair, and immune surveillance. Furthermore, biochemical studies will continue to elucidate complex gene-gene interactions. It is likely that, with identification of the entire genome, DNA chips will be used to genotype for polymorphisms in a multitude of these genes. This will undoubtedly lead to a trend to incorporate markers of susceptibility into epidemiologic studies, with tools available to study gene-gene interactions as well, in addition to gene-environment interactions. The implications of these applications are staggering. It is becoming clear that chronic disease is multifactorial, and multiple exposures, as well as variability in susceptibility based on polymorphisms in numerous genes, interact in disease etiology. Thus, molecular epidemiologists will have the ability to evaluate complex interactions of numerous genetic polymorphisms and environmental exposures, which could lead to groundbreaking understandings of disease etiology as well as to a plethora of poorly designed studies with misinterpreted results. This challenge can be met only by interdisciplinary collaborations and cross-training of epidemiologists, clinicians, and laboratory scientists in each others’ fields.

Components of successful collaboration. Cross-training of epidemiologists and laboratory scientists in the other’s respective discipline will only enhance collaborative research. While both contribute the strength of their skills to the research, an understanding of the entire breadth of the project by all parties allows for successful communication, hypothesis generation, and a clearer grasp of the perspective of one’s collaborator. A lack of such an understanding is often at the root of conflicts that may arise among collaborators. For an epidemiologist who has spent some time in the laboratory or who has at least worked and communicated extensively with those who have, the question, Why is this taking so long? will seldom be raised. Similarly, a laboratory scientist who turns over results to an epidemiologist to be incorporated into a database collected by interview may be frustrated by the length of time it takes to “run a logistic regression” unless he or she has worked with large databases and has an understanding of biostatistics and data analysis. An understanding of the complexities of both fields will encourage mutual respect and understanding, as well as enable fruitful communication.

Successful partnerships are also characterized by a degree of flexibility and a willingness to learn from one another. For example, when discussing research possibilities associated with an understanding of multiple enzymes involved in metabolic pathways for particular substrates, a biochemist may wax enthusiastically about the multiple gene-gene/gene-environment interactions involved in determining disease susceptibility. This scientist may be strongly inclined to forge ahead and evaluate polymorphisms in several genes that together may infer high risk with relevant exposure. On the other hand, the epidemiologist may only envision the nightmare of trying to relate exposures, combined “at risk” genotypes for several enzymes, and disease risk in a meaningful way. Such analysis, while biologically relevant, is analytically impossible unless the study consists of thousands of individuals. Flexibility on the part of scientists from both disciplines may result in identification of the means to begin to address such issues. Ultimately, while perspectives of collaborators may differ, the common ground for working together should be biologic plausibility.

One area that often presents difficulties in young collaborative relations and yet is the least likely to be addressed at the beginning of the project is the question of authorship of papers resulting from joint work. If first authorship belongs to “the one who did the work,” the epidemiologist may consider that his or her role, realizing the work involved in study design and administration as well as in data analysis. To some, the molecular biologist may be viewed as merely the genotyping laboratory. On the other hand, with assays that require development and troubleshooting as well as skill in interpretation of gels, the laboratory scientist may consider the epidemiologist as only a “numbers cruncher.” Again, training across disciplines may alleviate some of these naive views, but authorship should be discussed and agreed upon at the beginning of a collaboration. One approach taken by an interdisciplinary research group is an agreement that when an individual from one group is first author, a researcher

Am J Epidemiol Vol. 146, No. 11, 1997
from the other group will be last author, with a footnote that the project was a collaborative effort with equal input from both groups.

All of these issues can best be met and addressed by meaningful communication between the epidemiologic and laboratory science communities. It is likely that progress will be made in this field by the recent establishment of the international, multidisciplinary “Molecular Epidemiology Group,” whose goal is to foster an interdisciplinary approach to the study of chronic disease etiology by providing a forum for discussion and development of sound approaches to molecular epidemiology, sponsoring educational activities, and fostering of partnerships between scientists from various disciplines.

CONCLUSIONS

It is expected that the use of molecular markers in epidemiologic studies will expand as advances are made in our understanding of the molecular basis of disease, as well as in the technology to study it. The ability to keep pace with these advances will be served best by the establishment or use of existing cohorts or study groups and by establishment of biologic specimen banks. A cohort with biologic specimens will allow for nested case-control studies incorporating molecular markers to investigate new areas of research as they arise. A biologic specimen bank as a source of DNA will also allow for future investigations of questions that are yet to be posed, but may become of paramount importance as scientific knowledge advances.

With its rapid growth and implementation by numerous researchers from a variety of backgrounds, the field of molecular epidemiology often appears undis- ciplined and untrained. This area of research could be likened to a teenager, eager to explore the vast unknown, yet unschooled in the method of approach and lacking in the wisdom that only comes with years of experience. It is hoped that through future collaborations and discussions between epidemiologists, statisticians, laboratory scientists, and clinicians, and with a willingness and openness to explore ways to advance technology without stifling creativity, we may make progress in understanding and preventing chronic diseases that greatly affect the lives of millions of individuals in various populations.

REFERENCES


