Molecular epidemiology

The science and art of molecular epidemiology

M L Slattery

This paper details some of the issues surrounding the growing field of molecular epidemiology.

Epidemiology is both a science and an art. The science of epidemiology entails applying classic epidemiological methods to understanding the distribution of diseases in populations. The art of epidemiology is interpreting the findings. Molecular epidemiology provides new opportunities for epidemiologists and other medical researchers to understand diseases and make public health recommendations for disease prevention and treatment. The value of molecular epidemiological studies, in terms of providing information that can be used to improve the health of populations, depends on how well both the science and the art are applied.

Molecular epidemiology, an area of epidemiology that is somewhat ambiguous, encompasses utilisation of biomarkers and genetics as tools to define both exposures (factors that are inherited) and outcomes (factors that are acquired). As noted by Porta and colleagues, there are an increasing number of published articles with molecular epidemiology as a key word. Molecular epidemiology has been applied to many diseases, although a large percentage of published studies have focused on cancer. Within the cancer arena, most molecular epidemiological studies involving genetics have examined inherited genetic variants or polymorphisms. These genetic variants are exposures, a host characteristic, that may independently or through combination with other diet, lifestyle, or environmental exposures change disease risk. While the hope was that these studies would explain some of the inconsistent diet and lifestyle associations reported in the literature, many have added their own element of confusion.

Evaluation of acquired tumour mutations as a disease end point with diet, lifestyle, and environmental exposure data can provide information about specific disease pathways. The central issue in the review by Porta and colleagues was classification of genetic mutations in tumours and appropriate inferences from this classification. Despite the growing number of published molecular epidemiology studies, studies looking at acquired alterations in tumours are limited. Molecular epidemiological studies of tumour mutations have provided information about the distribution of specific alterations in populations and how diet and lifestyle factors are associated with specific genetic alterations in tumours. These studies have the potential of providing support for previously identified risk factors and a better understanding of the carcinogenic process. However, as Porta and colleagues point out, lack of careful application of the science of epidemiology can limit the amount of useful information obtained from studies of molecular epidemiology.

What is the science of epidemiology? Epidemiology is based on observations in disease trends, incidence, and mortality rates in different populations that turn into testable hypotheses. Observations, such as the one by Porta, that K-ras mutations occur commonly in pancreatic cancer or that p53 mutations occur commonly in many solid tumours, can be the stepping stone for hypotheses that can be tested in analytical studies, using either a case-control or cohort study design. A critical part of the science of epidemiology is appropriate study design selection and an understanding of the strengths and limitations of the study design chosen (see references 19–22). All study designs have potential sources of bias; it is imperative that sources of bias are understood and if possible, evaluated within the context of the study being conducted.

The science of epidemiology entails carefully defining targeted populations and making appropriate inferences from these populations; this is central to molecular epidemiology. Inferences made to the population need to come from studies that are conducted in the population. Being aware of potential selection bias and the impact, if any, on inferences made from the study is critical. For instance, in a study of colon cancer and tumour alterations, it has been shown that when participants have to be re-consented in order to obtain tumour blocks, a greater percentage of people with a family history of cancer participate in the study. The implications of a less representative population are many, including different distribution of mutations in tumours, different diet and lifestyle associations resulting from family history status, or different associations with inherited factors; all have the potential for inappropriate inferences to the population. By starting at the population level, meaningful subsets, not just samples of convenience, can be identified based on their age, gender, family history, or other diet and lifestyle characteristics. Inferences about associations to the population at large, as well as to smaller defined subsets, can be made.

The science of epidemiology entails rigorous collection of data for all study aspects; erroneous associations can result if exposure data are collected haphazardly even if genetic or other molecular data are error free. To collect accurate exposure data, knowledge of the subject matter for all exposures of interest as well as understanding the population being studied is needed. Knowledge of potential confounders to the disease/exposure associations is needed so that information that could bias results is collected and considered in the analyses. Collection of additional sources of data in molecular epidemiology studies, including blood and tumour blocks, have their own set of challenges. Debate is ongoing about issues of informed consent, human subjects, and ability to use samples for future research as new information on disease processes becomes available. Finally, transitional studies that provide information on validity of markers, the interrelation of various markers, and the application of these markers to studies of causal associations in populations are needed. For instance, are results obtained from immunohistochemistry of p53 overexpression the same as those obtained from sequencing the p53 gene? What are the advantages and limitations of each method of p53 analyses? Some attempts have been made to resolve these issues.

Lack of rigorous application of the scientific principles of epidemiology can be the pitfalls of molecular epidemiology. Briefly some of these problems can be summarised as:

- Ill defined target population or samples of convenience. When these samples are used, the hazard of making inappropriate inferences to the population exists. Molecular epidemiological studies that use convenient tumour samples are especially prone to this pitfall.
- Subsets of study participants who actually participate in molecular aspects of study. From a practical perspective it is impossible to get all samples or tumour blocks targeted. However, attempts need to be made to
determine if the study population differs from the broader targeted population.

- Genes of convenience are often studied. It is often easier to study a gene that others have examined than to determine the importance of other genes along the hypothesised disease pathway. Efforts to identify and assess other genes that have functional variants and are thought to be involved in the pathway of interest are needed to better understand the disease process.

- Small sample sizes, leading to imprecision in associations. To determine precise associations, molecular epidemiological studies need large samples, especially if we hope to examine disease pathways.

- Statistical methods are inappropriate, leading to wrong conclusions. In addition to applying appropriate statistical methods, careful thought as to the interpretation of results in terms of potential bias and biological implications is often lacking.

- Lack of quality control over laboratory data as well as data from field components of study. Within the context of molecular epidemiology, sample tracking is a critical part of quality control so that samples are appropriately linked to other study data.

- Publication bias. There is tremendous difficulty in getting null or confirmatory studies published, resulting in a limited and often misleading body of information available.

- The assumption that anybody can do epidemiology. This may stem in part from the sense of non-epidemiologists that epidemiology is a “soft science” and is easier to do than “bench science”. Designing and conducting studies involves a scientific body of knowledge, which when ignored, can lead to flawed conclusions. Lack of application of the science of epidemiology can leave little hope for a meaningful application of the art of epidemiology.

- It is the art of epidemiology that pulls together the biological, clinical, and environmental information that will transcend epidemiology from defining associations to describing disease pathways. To do this epidemiologists must have an understanding not only of bias, but also of biology. They must develop a broad understanding of disease pathways being studied, so that data collection and analyses can be meaningful.

While working at the population level of exploration, molecular epidemiology must incorporate knowledge from many disciplines to obtain an understanding of the organism, the system, and the cell. Translating complex disease pathways into relevant public health messages should be the goal and the result of the art of epidemiology.

Wade Hampton Frost’s characterisation of epidemiology in 1936“ applies to many of our current attempts to understand disease. He described epidemiology as: “...something more than the total of its established facts. It includes their orderly arrangement into chains of inference which extend more or less beyond the bounds of direct observation. Such of those chains as are well and truly laid guide investigation to the facts of the future; those that are ill made fetter progress.” Molecular epidemiological studies, when based on the science and art of epidemiology, can leave little hope for a meaningful scientific return involving genetics. JAMA 2001;285:799–801.

