

# 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,<sup>1</sup> 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.<sup>2-8</sup>

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<sup>1</sup> 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<sup>9-12</sup> and how diet and lifestyle factors are associated with specific genetic alterations in tumours.<sup>13-18</sup> 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<sup>1</sup> 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<sup>1</sup> or that *p53* mutations occur commonly in many solid tumours,<sup>19</sup> 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<sup>20-22</sup>). 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 studying 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.<sup>23</sup> 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.<sup>24</sup> 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.<sup>21</sup> 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.<sup>25</sup>

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<sup>26</sup> 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 truly guide investigations into the future; if not, they may indeed fetter progress.

**ACKNOWLEDGEMENT**

The contents of this manuscript are solely the responsibility of the author and do not necessarily represent the official view of the National Cancer Institute.

*J Epidemiol Community Health* 2002;**56**:728-729

.....

**Author’s affiliations**

**M L Slattery**, Health Research Center, 375 Chipeta Way, Suite A, Salt Lake City, Utah 84108, USA

Funding: this study was funded by CA48998 and CA61757 to Dr Slattery.

Conflict of interest: none.

Correspondence to: Dr M L Slattery; Marty.Slattery@hrc.utah.edu

**REFERENCES**

- 1 **Porta M**, Malats N, Vioque J, *et al*. Incomplete overlapping of biological, clinical, and environmental information in molecular epidemiological studies: a variety of causes and a cascade of consequences. *J Epidemiol Community Health* 2002;**56**:734-8.
- 2 **Geisler SA**, Olshan AF. GSTM1, GSTT1, and risk to squamous cell carcinoma of the head and neck: mini-HuGE review. *Am J Epidemiol* 2001;**154**:95-105.
- 3 **Chen J**, Giovannucci E, Kelsey K, *et al*. A methylenetetrahydrofolate reductase polymorphism and the risk of colorectal cancer. *Cancer Res* 1996;**56**:4862-4.
- 4 **Slattery ML**, Potter JD, Samowitz WS, *et al*. Methylenetetrahydrofolate reductase, diet, and risk of colon cancer. *Cancer Epidemiol Biomarkers Prev* 1999;**8**:513-18.
- 5 **Kampman E**, Slattery ML, Bigler J, *et al*. Meat consumption, genetic susceptibility, and colon cancer risk: A US multi-center case-control study. *Cancer Epidemiol Biomarkers Prev* 1999;**8**:15-24.

- 6 **Chen J**, Stampfer MJ, Hough HL, *et al*. A prospective study of N-acetyltransferase genotype, red meat intake, and risk of colorectal cancer. *Cancer Res* 1998;**58**:3307-11.
- 7 **Furgerg AH**, Ambrosone CB. Molecular epidemiology, biomarkers and cancer prevention. *Trends Mol Med* 2001;**7**:517-21.
- 8 **Rothman N**, Wacholder S, Caporaso E, *et al*. The use of common genetic polymorphisms to enhance the epidemiologic study of environmental carcinogens. *Biochim Biophys Acta* 2001;**1471**:C1-10.
- 9 **Samowitz WS**, Curtin K, Ma KN, *et al*. Microsatellite instability in sporadic colon cancer is associated with an improved prognosis at the population level. *Cancer Epidemiol Biomarkers Prev* 2001;**10**:917-23.
- 10 **Samowitz WS**, Curtin K, Schaffer D, *et al*. Relationship of K-ras mutations in colon cancers to tumor location, stage and survival: a population-based study. *Cancer Epidemiol Biomarkers Prev* 2000;**9**:1193-8.
- 11 **Rashid A**, Zuhurak M, Goodman SN, *et al*. Genetic epidemiology of mutated K-ras proto-oncogene, altered suppressor genes, and microsatellite instability in colorectal adenomas. *Gut* 1999;**44**:826-33.
- 12 **Andreyev JHN**, Norman AR, Cunningham D, *et al*. Kirsten ras mutations in patients with colorectal cancer: the multi-center “RASCAL” study. *J Natl Cancer Inst* 1998;**90**:675-84.
- 13 **Slattery ML**, Curtin K, Anderson K, *et al*. Associations between Cigarette smoking, lifestyle factors, and microsatellite instability in colon tumors. *J Natl Cancer Inst* 2000;**92**:1831-5.
- 14 **Slattery ML**, Curtin K, Ma K, *et al*. Associations between dietary intake and Ki-ras mutations in colon tumors: a population-based study. *Cancer Res* 2000;**60**:6935-41.
- 15 **Martinez ME**, Malizman T, Marshall JR, *et al*. Risk factors for Ki-ras protooncogene mutation in sporadic colorectal adenomas. *Cancer Res* 1999;**59**:5181-5.
- 16 **Bautista D**, Obrador A, Moreno V, *et al*. Ki-ras mutation modifies the protective effect of dietary monounsaturated fat and calcium on sporadic colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 1997;**6**:57-61.
- 17 **Slattery ML**, Potter JD, Curtin K, *et al*. Estrogens reduce and withdrawal of estrogens increases risk of microsatellite instability-positive colon cancer. *Cancer Res* 2001;**61**:126-30.
- 18 **Slattery ML**, Anderson K, Curtin K, *et al*. Association between dietary intake and microsatellite instability in colon tumors. *Int J Cancer* 2001;**93**:601-7.
- 19 **Hollstein M**, Rice K, Greenblatt MS, *et al*. Database of p53 gene somatic mutations in human tumors and cell lines. *Nucleic Acids Res* 1994;**22**:3551-5.
- 20 **Rothman KJ**, Greenland S. *Modern epidemiology*. 2nd edn. Philadelphia, PA: Lippincott-Raven, 1998.
- 21 **Schulte PA**, Perera FP, eds. *Molecular epidemiology, principles and practices*. San Diego, CA: Academic Press, 1993.
- 22 **Slattery ML**. Does an apple a day keep breast cancer away? *JAMA* 201;**285**:799-801.
- 23 **Slattery ML**, Curtin K, Schaffer D, *et al*. Associations between family history of colorectal cancer and genetic alterations in tumors. *Int J Cancer* 2002;**97**:823-7.
- 24 **Beskow LM**, Burke W, Merz JF, *et al*. Informed consent for population-based research involving genetics. *JAMA* 2001;**286**:2315-21.
- 25 **Voskuil DW**, Kampman E, van Kraats AA, *et al*. p53 over-expression and p53 mutations in colon carcinomas: relation to dietary risk factors. *Int J Cancer* 1999;**81**:675-81.
- 26 **Frost WH**. *Introduction to Snow on cholera; being a reprint of two papers by John Snow MD*. New York: The Commonwealth Fund, 1936.