Cancer Ecogenetics: Studying Genetic and Environment Interactions through Epidemiology

JOHN J MULVIHILL AND HRAFN TULINIUS

CANCER ECOCENETICS

Are cancer epidemiologists ready for molecular geneticists? This could have been the subtitle of a workshop, 'Genetic and Environmental Factors in Etiologic Studies of Cancer,' a European regional scientific meeting of the International Epidemiology Association, held in Reykjavik, Iceland, 13-15 June 1986. Under scrutiny was the intersection of two apparently contradictory trends in research on the origins of human cancer.

On the one hand, cancer, at least at the level of the cell, is a genetic disease. Most tumours have chromosomal abnormalities and, if the cancer cell did not pass on to its progeny new rules for bad cellular behaviour, cancer would not be a lethal disease. A score of oncogenes, discovered by laboratory scientists pursuing the hypothesis that cancer is an infectious disease, gave further insight that DNA sequences, long part of the human genome, could become activated and result in clinical cancer. Studies of cancer genetics proceeded mostly in laboratories, apart from any epidemiological assessment of potential environmental influences.

On the other hand, it is often said that 80 to 90% of human cancers are due to environmental factors, including behavioural ones such as personal diet. Geographical and temporal differences of cancer rates, occupational cancers, and, most compellingly, epidemiological studies of cancer in migrant populations point to its progeny new rules for bad cellular behaviour, cancer would not be a lethal disease. A score of oncogenes, discovered by laboratory scientists pursuing the hypothesis that cancer is an infectious disease, gave further insight that DNA sequences, long part of the human genome, could become activated and result in clinical cancer. Studies of cancer genetics proceeded mostly in laboratories, apart from any epidemiological assessment of potential environmental influences.

Any theory of carcinogenesis must allow for multiple steps, many of which can be modified by genetic or environmental factors. One possible sequence begins with activation of a protooncogene, which may affect or be affected by repair of genetic damage, for example, at fragile sites or at a mutation of an oncogene or antioncogene, a term coined by Knudson. Oncogenes influence carcinogenesis by becoming present and/or activated by (for example chromosomal translocation) and are non-hereditary but dominant mutations with a broad range of tissue specificities. In contrast, antioncogenes contribute to tumour formation by being inactivated or lost (for example, through chromosomal deletion), may be both hereditary and non-hereditary, and act as recessive genes with considerable tissue specificity. The existence of antioncogenes can be postulated because dominantly inherited cancers, for example retinoblastoma and Wilms' tumour, arise following the loss of genetic heterozygosity, that is a switch to hemi- and homozygosity in cells of specific tissues. Simply put, an antioncogene is a human gene that is normally active to block the development of cancer. Such candidate genes seem to cluster on chromosome 1p, 3p, 11p, and 13q.
Thorgeirsson from the National Cancer Institute summarized his progress in dissecting the complexity of just one step in another possible sequence of carcinogenesis: a procarcinogen undergoes metabolic activation to form an electrophilic intermediate that initiates a genetic change in a cell that, after promotion, results in a clinical cancer. By focusing on the model compound, 2-acetyl aminofluorene, Thorgeirsson saw differences in enzyme kinetics over a 10,000-fold range of dose that documented distinct activities for activation and detoxification. If homologous genes are found in human beings, a wide range of cancer susceptibilities to environmental agents can be anticipated, with good activators and poor detoxifiers representing the most susceptible class of individuals. In the end, Thorgeirsson predicts that population studies must be done, not with the present bioassay involving in vivo challenges, but with DNA polymorphisms.

Leppert from Salt Lake City gave details of the approach using DNA to delineate ecogenetic traits in human cancer. The simple strategy is to detect differences in the two copies (alleles) of each gene that a person inherits, one from the mother and one from the father. Modern DNA chemistry provides overwhelming possibilities for detecting such polymorphisms. In practice, the arduous task is to decide which restriction enzymes to apply to DNA and which DNA sequences should be used to probe the different fragment lengths. Some 2000 must actually be tried to achieve ideal coverage of the entire genome. The selection of specific probes can be guided by prior evidence, such as from formal family studies using red cell markers, case reports of individuals with specific chromosomal abnormalities, or by some pathogenetic hypothesis; or, it must be acknowledged, the choice may be by blind luck. Alternatively, the detection of abnormal chromosomes in a cancer or the loss of heterozygosity for specific genes in tumour cells compared to normal body cells could quickly suggest a chromosomal region, when the issue is cancer predisposition. There is a real need for developing resources to permit, for example, transforming lymphocytes, isolating DNA, and preparing gels with standard enzymatic digestions for shipment worldwide to laboratories with special panels of probes for a specific chromosome or a specific class of compounds.

Lathrop from Salt Lake City emphasized some principles of genetic epidemiology that could influence the design of future studies. Two examples are the needs to identify individuals with double heterozygosity [two forms of the disease locus and two forms (alleles), of the gene marker] and to know the phase (which allele of the marker gene is linked to the disease gene). In addition, one must recognize the vast differences in resolution between the gene map as seen by physical methods (karyotyping) and genetic methods (recombination events).

**USING MONOGENIC PRENEOPLASTIC DISORDERS AS MODELS**

Some 200 single gene traits have neoplasia as a feature or complication. Individually rare, these traits may, in the aggregate, cause a significant fraction of cancers, even common ones like breast and colon cancer. Their study has great appeal because each host cell has the mutant gene that predisposes to cancer and whose action may be discoverable through genetic techniques. Moreover, the same initial mutation may occur in sporadic, environmentally-induced instances of the same cancer.

One such cancer is medullary carcinoma of the thyroid, according to Ponder (Sutton), who illustrated the complementary information that could be gained by studying sporadic versus familial forms of the same cancer. This cancer provides the opportunity for early detection and prevention of cancer by chemical screening (a rise in serum calcitonin following an injection of pentagastrin). Data from a UK registry of Sipple syndrome (multiple endocrine neoplasia, type 2, the familial occurrence of pheochromocytoma and medullary carcinoma of the thyroid) show that one third of Sipple syndrome gene carriers will not have presented clinically by the age of 60 years. At least 95% of gene carriers manifest the *biochemical* abnormality on screening by the age of 30 years, which allows confident exclusion of some branches of the family from further screening, and also the certain identification of gene carriers in genetic linkage studies.

In pursuing the gene for polyposis coli (the most powerful single known cause of colon cancer), Spurr et al (ICRF) are preparing specific DNA probes to clarify presently unmapped regions of chromosomes and, in particular, those genes present in a single copy (so that no doubt remains that the gene is on just one chromosome). Additional activities are the establishment of cell lines from informative families, preparation of human-rodent hybrid cell panels to speed the chromosomal localization of the observed genes, and the preparation of monoclonal antibodies to detect surface antigens that are specific to each chromosome. Of special note for epidemiologists is his call for equal attention to the maintenance of useful and efficient computerized records.
UNDERSTANDING FAMILIAL CANCER

The phenomenon of familial aggregation of cancer is more prevalent than single gene traits, yet so far more refractory to analytical insights. Since cancer is common (one in four individuals in a lifetime), most people will have some relatives with cancer, and by bad chance, some will have many.

Currently, syndromes of familial cancer are defined largely by intuition based on the numbers and types of cancers, the ages at diagnosis, and the occurrence of multiple primary cancers in some family members. The validity of a proposed syndrome can be best tested by long-time follow-up of initial families to see if cancers continue to occur to excess. Williams (Philadelphia) added statistical rigour to defining the Li-Fraumeni cancer family syndrome, by documenting and mathematically modelling the family histories in a series of 159 consecutive children with soft tissue sarcomas at the M. D. Anderson Hospital, Houston. Williams' segregation analysis showed a heritability of 0.13 (0.23 when the index was diagnosed under age 5 years). Nine second cancers were seen compared to 0.5 expected, and the autosomal gene model fit the observed patterns best.

In commenting, J. Peto from Oxford distinguished the genetic relative risk (rate in carriers/rate in non-carriers) from the epidemiological relative risk (rate in relatives of cancer patients/rate in general population). He illustrated how the low power of twin studies could easily result in overlooking major genetic determinants of cancer. He described a record linkage system in the UK that may test for familiality of common adult cancers with great power by virtue of its large numbers. All households had been enumerated in 1939 and subsequent deaths and cancer registrations have been linked to the register. Almost 400,000 cancers of all sites have been registered in the 1923-1939 birth cohort, including 5200 breast cancers in women under age 35 years.

Such nationwide surveys should give an accurate picture of familial cancer. In Iceland, with its small population but excellent genealogies, Tulinius said there was a total of 854 pedigrees of breast cancer patients with 62,300 identified people and 46,000 blood relatives. Early analysis of the nationwide data revealed the 2.6 relative risk for breast cancer in first degree relatives of breast cancer cases. Subsequent analyses could pursue the mixed model of segregation analysis illustrated by Williams. Breast cancer has now been documented in 1230 people (in the genetic resources), of which 45% are alive and potentially available for collection of biological specimens. Given the thrust of this workshop, the specimen thought to be of greatest use is DNA that could be isolated directly from fresh lymphocytes, or Epstein-Barr virus-transformed cultures.

LINKING CANCER RECORDS

The occurrence of additional primary tumours in cancer survivors is an unfortunate clinical challenge, as well as a chance for cancer epidemiologists to distinguish genetic from environmental determinants. Such studies, especially in childhood, require international collaboration, as reported by Meadows (Philadelphia), for the Late Effects Study Group. Among survivors of childhood cancer, 299 individuals developed 322 additional cancers. Some appeared to be related to radiation therapy and, perhaps, to chemotherapy for the first tumour; others clearly arose from the same host factors that contributed to the initial cancer. For example, 48 subsequent bone cancers were observed compared to 0.4 expected.

Jensen from Copenhagen commented on prospects of record linkage in the Danish Cancer Registry which now contains over 700,000 cancers diagnosed as long ago as 1943. In contrast to Meadows' study of second cancers in children, his survey of multiple cancers in adults could not properly address the issue of genetic predisposition because, for example, a family history was not collected. Nonetheless, patterns of multiple primary tumours associated with tobacco use and radiotherapy could be identified as well as clusters of colon, endometrium, and breast cancers, resembling patterns seen in the cancer family syndrome of Lynch.

CONCLUSION

Mulvihill from Bethesda launched the final discussion of the workshop by contrasting the theory and practice of classic epidemiology and human genetics. There are large, fundamental similarities in these fields: each studies variations in disease in populations, relies on and stimulates progress in statistics, and makes special use of the twinning phenomenon. In current practice, cancer epidemiologists tend to neglect genetic hypotheses and geneticists fail to study environmental determinants (for example, in the age of onset of certain features of a genetic disease). Molecular genetics has enormous potential and has developed a vocabulary and strategy that may sound foreign to analytical epidemiologists. An ideal for future studies would be to have a senior geneticist and epidemiologist contribute to design of major efforts. Investigators of cardiovascular disease and mental illness have done so with advantage. As a general practice, epidemiologists should strongly consider, as these studies develop, collecting and documenting medical family history and
storing DNA in some form. Environmental epidemiologists might likewise advise their enthusiastic geneticist colleagues in matters of proper sampling procedures, data management, sources of population information of special interest, and proper development of controls and denominators.
Use of Quantified and Frequency Indices of Vitamin A Intake in a Case-Control Study of Lung Cancer

CHARLES G HUMBLE*, JONATHAN M SAMET† AND BETTY E SKIPPER


Recent publications have examined the extent of food frequency data needed to estimate nutrient intakes in epidemiological studies. The need for amount (usual portion size) data to supplement information on average frequency of intake has been questioned. In a case-control study of risk factors for lung cancer, we have collected data on frequency, amount and past pattern of intake for common or rich sources of vitamin A. These data have been combined with standard content information to calculate three different types of indices: one based on frequency alone, a quantitative index that included both frequency and amount, and a past-weighted index that combined the frequency, amount and data on relative consumption in the past. Odds ratios by tercile of carotene, retinol and total vitamin A intake varied little between the frequency index and the quantified and past-weighted indices. Higher intakes of carotene and total vitamin A were associated with a lower risk for lung cancer regardless of index type. This consistency of results is explained by a stability of relative intakes regardless of index type. Spearman rank order correlation coefficients between the frequency and quantitative indices exceeded 0.90 for total vitamin A, carotene, and retinol. This stability is attributable to parallel trends of increased frequency of consumption and portion size at higher levels of nutrient intake. Because similar trends were observed in five main food groups included in these analyses, these findings may be generalizable to other nutrients.

Food frequency questionnaires have been widely used in large scale epidemiological investigations of diet and disease. Several recent publications have addressed the extent of information needed with this type of questionnaire to estimate intake of specific nutrients in the context of an aetiological investigation. Some investigators have argued that quantitative measures incorporating both frequency of consumption and portion size are required; others have suggested that frequency alone may be satisfactory for some nutrients, particularly when analysis will be based on relative measures of intake.

We have recently completed a case-control study of lung cancer in New Mexico that examined the relationship between vitamin A intake and risk for this malignancy. The food frequency questionnaire obtained usual frequency of consumption, usual portion size, and stability of consumption for 55 different foods. We have used the responses to calculate indices of vitamin A intake based on frequency alone, on frequency and amount, and on frequency, amount, and stability of consumption. Using quantified (i.e., frequency and amount based) indices we found that higher intakes of carotene were protective against lung cancer in non-Hispanic white males and females who had stopped smoking two or more years before the interview. Here we have repeated the original case-control analyses to empirically assess the consequences of using dietary measures based on frequency alone rather than quantified measures that combine frequency and portion size.

METHODS
Detailed descriptions of subject selection, data collection, and methods used to calculate nutrient indices are provided in earlier reports. Briefly, in a population-based case-control study of lung cancer in New Mexico, 467 cases and 762 controls provided complete dietary data in response to a food frequency questionnaire (Table 1). Subjects described the average frequency of intake, usual amount, and stability of intake for each of 55 foods considered to be important sources of vitamin A, both carotene and retinol, in the local diet. Usual frequency of intake was ascertained on a