Pergamon

European Journal of Cancer Vol. 31A, Nos. 7/8, pp. 1047–1050, 1995

0959-8049/95 $9.50+0.00

0959-8049(95)00156-5

Genetic Epidemiology of Colorectal Cancer

G.M. Petersen

Genetic epidemiological methods have played an integral role in the characterisation of the genetic susceptibilities to colorectal cancer. Classic epidemiological approaches, such as case–control and prospective cohort studies, that utilise family history information have laid the foundation for the more specialised family-based genetic methods, segregation analysis and linkage analysis. The genetic epidemiology of colorectal cancer can be characterised by several themes: the consistently increased risk of colorectal cancer in first-degree relatives of patients with colorectal cancer; genetic predisposition to some, if not the majority of colorectal neoplasms; and genetic heterogeneity of the inherited colorectal cancer syndromes. With the rapid development of molecular genetic techniques, new opportunities for further research include studies to estimate the proportion of colorectal cancer that is accounted for by genetic susceptibility, the number of loci that may be involved, and most importantly, gene–environment interaction studies, not only of the inherited syndromes, but of common colorectal cancer.

Key words: colon, family history, genetic analysis, susceptibility
Eur J Cancer, Vol. 31A, Nos 7/8, pp. 1047–1050, 1995

INTRODUCTION

THE PAST DECADE has seen important new discoveries that have uncovered the molecular genetic basis of colorectal tumorigenesis. These include the delineation of a tumour progression model, and the genetic characterisation of inherited colorectal cancer (CRC) syndromes, as described elsewhere in this issue. The genetic epidemiology of CRC can be characterised by several themes: the consistently increased risk of CRC in first-degree relatives of patients with colorectal cancer; genetic predisposition to some, if not the majority of colorectal neoplasms; and genetic heterogeneity of the inherited colorectal cancer syndromes. A less consistent, but notable theme is the pattern of positive family history with younger age at diagnosis of CRC and tumour site. Genetic analyses of families have determined the mode of inheritance, and estimated the frequency of CRC susceptibility genes and their penetrance in gene carriers, again noting a relationship to younger age at onset. In addition to elucidating the genetic bases of the clinically well-defined inherited CRC syndromes, genetic epidemiological studies have generated new areas that merit investigation, such as the relationship between CRC gene mutations and clinical phenotypes in patient populations.

Genetic epidemiology is a relatively new discipline that has united the methods of epidemiology and human genetics in the study of complex diseases [1]. The goal of genetic epidemiological investigations is to characterise the aetiologies, both genetic and nongenetic, of diseases in humans. For CRC, these methods have included case–control studies, cohort studies, studies that use population-based registries, segregation analysis, and linkage analysis. This paper will review examples of each type of study, and discuss implications for future research directions.

Correspondence to G.M. Petersen at the Department of Epidemiology, Johns Hopkins School of Hygiene and Public Health and Oncology Center, Johns Hopkins School of Medicine, 615 N Wolfe St, Baltimore, Maryland 21205, U.S.A.

CASE–CONTROL STUDIES

One of the most consistent findings to emerge from epidemiological case–control studies of family history of CRC has been that first-degree relatives are themselves at a two to three-fold increased risk of CRC. The consistency of the magnitude of risk is remarkable, despite the various sampling frames, sample sizes, methods of data verification, analytical methods, and countries where the studies originated.

There have been numerous case–control studies of family history in CRC (reviewed in [2, 3]), but only illustrative examples will be presented here. Among the earliest studies was that of Woolf [4], who found, in pedigrees from Utah, that there was a higher number of deaths from CRC in the first-degree relatives of 242 patients who had died from CRC, compared with controls. Index cases were obtained from death certificates and family histories were determined from the records of the Mormon Genealogical Society. Causes of death for the relatives (145 fathers, 142 mothers, 209 brothers, 167 sisters) were also determined from death certificates. The control group consisted of sex-matched individuals who had died in the same country, the same year, and at approximately the same age as each of the deceased relatives of the propositi. Twenty-six first-degree relatives of the propositi (3.9%) had CRC compared to 8 (1.2%) controls ($P < 0.01$). The number of deaths due to CRC was increased over the controls in all four classes of relatives.

However, the use of mortality figures may not accurately reflect the prevalence of CRC among first-degree relatives, because affected relatives may still be alive at the time of the study, or may have died from another cause. One approach to this problem is to obtain information directly from patients or next of kin by questionnaire or interview. Duncan and Kyle [5]

studied 50 consecutive patients with adenocarcinoma of the colon or rectum in Scotland, excluding syndromic colon cancers. Family histories of 349 first-degree relatives (parents, siblings and children) were obtained by interview. Comparable controls, matched for sex and age, admitted for treatment of various non-malignant conditions were identified. Family histories of 386 first-degree relatives were similarly obtained. 8 of 50 patients in the cancer group (16%) and only one in the control group (2%) had a first-degree relative with colorectal cancer ($P < 0.05$). These investigators also examined clinical characteristics of their patients with positive family histories. They found that the mean age was similar to that of the whole group. In the patients with adenomatous polyps, 2 out of 12 (16%) had a positive family history of colorectal cancer, exactly the same percentage as the whole sample. There was also no right-sided predominance in the cancer group.

A number of other studies that vary this approach have examined the prevalence of adenomatous polyps or CRC in asymptomatic relatives of CRC cases compared with controls. For example, Rozen and associates [6] studied 471 asymptomatic adults who were first-degree relatives of patients having large bowel "neoplasia" (adenomatous polyps and/or cancer, excluding polyposis syndromes) in Israel. These first-degree relatives were screened by faecal occult blood examinations and flexible sigmoidoscopy, followed by colonoscopy when indicated. Adenomatous polyps or cancer were found in 8.1% of the study group as compared with 3.7% in a control group of 457 volunteers who did not have the same family history of neoplasia but were undergoing similar screening tests. Of the study group, the age-adjusted rate for colorectal adenomas and cancer increased 3-fold ($P < 0.001$) for subjects older than 40 years, and an even higher 5-fold relative risk was found for large bowel cancer alone ($P = 0.01$). Analysis by the number of affected first-degree relatives showed that those screened have a significant linear trend of increasing risk of colorectal neoplasia with increasing number of affected relatives. If only one relative was affected with large bowel neoplasia, the risk was 3-fold ($P < 0.01$), whereas with more than one affected relative, the risk increased to 5-fold ($P < 0.01$). When considering cancer only, the screenees with only one affected relative again had a 3-fold increased risk for CRC, but this did not reach statistical significance. However, the screenees with more than one relative were 9 times more at risk for cancer ($P < 0.01$).

These studies, which are examples of many in the literature, illustrate the variety of approaches, including case ascertainment and data collection, which characterise case–control studies. The magnitude of risk across studies is similar, and the patterns of risk associated with strength of family history, age at diagnosis, and adenomatous polyps present areas for further investigation.

COHORT STUDIES

The advantage of prospective cohort studies is that many of the potential study biases that can occur in case–control studies are obviated. In particular, the topic of interest, family history, can be collected in an unbiased, systematic way from all persons prior to disease outcome. One of the best examples of this approach was recently reported by Fuchs and associates [7], who analysed whether a family history of CRC (in first-degree relatives) was an independent risk factor for CRC. The subjects were derived from the Nurses' Health Study, a cohort of 121 700 U.S. women who were registered nurses aged 30–55, begun in 1976, and the Health Professionals Follow-up Study, a cohort of 51 269 U.S. men aged 40–75, begun in 1986. Members of both cohorts were asked to complete baseline and follow-up questionnaires on family history of cancer, and a number of other risk factors, including dietary intake, alcohol intake, aspirin use, physical activity and smoking. Both males and females reported a similar prevalence of CRC among first-degree relatives, 9.4% among men, and 10% among women. Over the study period, 148 males and 315 females developed CRC; 17% of this group had reported a family history of CRC, which was equivalent to a 1.7-fold increased risk for CRC, and was similar in males and females. This magnitude of risk is similar to that seen in the case–control studies discussed above. The relative risk of CRC increased with number of first-degree relatives, and was greatest for persons who were younger than 45 years of age. This excess risk was not affected by the other risk factors measured. This study provides further evidence that genetic susceptibility, as measured by family history, is an independent risk factor for CRC. This study is striking in its consistency with studies that utilised other genetic epidemiological methods.

POPULATION-BASED REGISTRY STUDIES

The use of population-based cancer registries has provided an alternative means to examine questions about shared family environment and cancer risk. One example of this is the study by Mellemgaard and colleagues [8], which used the Danish Cancer Registry and marital information from the population census register to characterise CRC incidence among 8345 spouses of 8529 CRC patients, identified from the cancer registry. Relative risks could be estimated from the observed number of CRC in the spouses compared to expected number of cases from age, sex, and time-specific incidence rates for Denmark. Interestingly, this study found no increased risk of colon cancer among spouses (relative risk of 0.96 for husbands, and 0.94 for wives), even when duration of marriage was considered.

In the U.S.A., a unique opportunity is offered by the Utah Population Database, which contains genealogical information on over 1.2 million persons. This database can be linked to the statewide tumour registry to explore hypotheses on the relationship of CRC and family history [9, 10]. Cannon-Albright and associates estimated a "genealogical index" from these data in order to compare the relatedness of 1800 colon cancer patients grouped by colon site (proximal versus distal) and early age of onset. While their analyses revealed a clustering of common colon cancer in families, as evidenced by a larger genealogical index, there did not appear to be preferential clustering by site or age of onset. They concluded that familial history is also a risk factor in the non-syndromic forms of colon cancer.

Slattery and Kerber [10] used the same, but updated Utah databases in a case–control study of colon cancer, and estimated a "familial standardised incidence ratio" for each group. Cases were 1299 women and 1244 men with primary colon cancer, and controls were matched for age, place of birth, marital status and sex. Their analyses showed that most of the increased risk of colon cancer derives from first-degree relationship (odds ratios of 2.5 to 2.9), and that the risk increases with number of affected relatives. They also found that increased risk is associated with younger age of onset, and for both men and women, a family history of breast, uterine, ovarian or prostate cancer increased colon cancer risk.

SEGREGATION ANALYSES

The use of pedigree data is an important strategy in genetic epidemiology. Segregation analysis is a method which tests

various models of transmission of a phenotype within a family for consistency with genetic and non-genetic aetiologies [1]. Results of these analyses vary, depending upon the phenotype of interest and the models tested. One of the most well-known studies which used this method is that of Cannon-Albright and associates [11]. These workers tested the hypothesis that there is a genetic basis for non-polyposis colorectal adenomas. They ascertained kindreds through patients with colon cancer and adenomas, and screened by flexible sigmoidoscopy all available first-degree relatives over 25 years of age, and spouses of family members as controls. One or more adenomatous polyps were found in 21% of family members (41 of 191), but in only 9% of controls (12 of 132) ($P < 0.005$). Segregation analysis of these kindreds suggested that the observed excess of discrete adenomatous polyps and CRC was the result of an inherited dominant gene for susceptibility, rather than an inherited recessive gene for susceptibility or a chance occurrence.

More recently, segregation analysis has been applied to the hereditary non-polyposis colorectal cancer (HNPCC) phenotype. Although one of the goals is to establish mode of inheritance, the value of these studies, particularly those derived from population-based registries, is to estimate the frequency of the gene, and the penetrance of the phenotype. Among the earliest studies is that of Bailey-Wilson and associates [12] who used a highly selected sample of eleven extended families comprising 2762 members. This study found that the best model assumed autosomal dominant inheritance, with penetrance of 71–79%, and a gene frequency of 0.008.

Houlston and colleagues [13] examined 203 pedigrees, and found that the best model also included autosomal dominant inheritance, with a penetrance of 63%, and a gene frequency of 0.006. They also found that these genes accounted for a large proportion (81%) of young onset (under age 35) colon cancer. Another study of population-based registry data from Modena, Italy [14, 15] found slightly different results, but consistent with the overall pattern of susceptibility. Of 605 CRC families in the registry, 28 pedigrees were suspected of being HNPCC families. Using both sources of data, their segregation analysis concluded that two loci contribute to HNPCC, with a major codominant locus having a gene frequency of 0.0044, and lifetime penetrance of 72.8%.

LINKAGE ANALYSES AND BIOMARKERS OF CRC

Linkage analysis of CRC families has been a particularly fruitful strategy for identifying predisposing genes. In particular, this has led to the identification of the gene for familial adenomatous polyposis [16–18] and HNPCC [19–22], described elsewhere in this issue (pp. 1039–1046). Because of the increased availability of mapped, highly polymorphic genetic markers, improved technologies for rapid molecular analysis, and computerised linkage analysis, the yield of this approach promises to be high, as CRC families that demonstrate non-Mendelian familial aggregation are studied.

Other biomarkers of risk that may prove useful include enzyme polymorphisms and tumour markers. Ilett and associates [23] conducted a study of 49 CRC patients and two control groups matched for sex, age, and race for metabolic clearance of sulphamethazine. They found a significantly larger proportion of rapid acetylators among the colon cancer group compared with the controls. The authors speculated that this phenotype may reflect differences in cancer patients' ability to metabolise xenobiotics, and thus provide a marker of susceptibility. More recently, we and others have found that microsatellite instability, or replication error (RER) can be assayed in colorectal tumours, using molecular analysis of microsatellite (dinucleotide repeat) polymorphisms in tumour DNA compared with constitutional DNA [20]. The RER phenotype occurs in virtually all colorectal tumours obtained from patients with HNPCC, while a much smaller proportion of sporadic tumours are RER positive. Whether this biomarker is a reflection of susceptibility remains to be elucidated. We have recently completed a study in which the majority of apparently sporadic CRC patients under the age of 35 have evidence of replication errors in their colorectal tumours [23a].

FUTURE DIRECTIONS

The results of all these studies confirm the usefulness of the genetic epidemiological approach to colon cancer. From a practical standpoint, these studies justify the clinical practice of obtaining a family history from CRC patients in order to identify at-risk individuals who could benefit from regular surveillance for CRC and adenomatous polyps [24]. Identification of genetic markers of risk, such as RER phenotype of colorectal tumours [20], could facilitate selection of high risk families. Improved methods for risk assessment will be a consequence of genetic epidemiological studies.

A number of areas for future research have become apparent. For example, how large is the proportion of CRC accounted for by genetic susceptibility? Epidemiological studies, such as that conducted by Mecklin [25], may be further enhanced now that genetic epidemiological studies have characterised features such as age at onset, and molecular studies have identified predisposing CRC genes. In his study, Mecklin [25] found a frequency of 3.8–5.5% for cancer family syndrome among all CRC patients diagnosed in one Finnish province during a 10-year period. Among 584 family members, 150 (25.7%) had some form of malignancy, 86 (14.7%) being CRC. The frequencies of familial adenomatosis and ulcerative colitis were 0.2 and 0.6%, respectively. Mecklin concluded that approximately 4–6% of all CRC seemed to be due to specific genetic syndromes. More recently, Aaltonen and colleagues [26] used age at onset of CRC as a means of identifying families with HNPCC from the Finnish Cancer Registry. Six families with HNPCC were identified from a cohort of 227 CRC patients, with age of onset less than 45 years. Their analysis suggested that 0.5–0.9% of all CRCs are due to HNPCC.

A second area that genetic epidemiology can help address is that of characterising genetic and clinical aspects of inherited syndromes. This has been particularly notable in advancing our understanding of familial adenomatous polyposis since the *APC* gene was identified. A number of studies have been published that have examined correlations between *APC* mutations and phenotypic characteristics of FAP patients, such as congenital hypertrophy of the retinal pigment epithelium, desmoid tumours, and other extracolonic lesions [27–30]. Now that several genes have been identified for HNPCC, there is great promise for further understanding of this phenotype, which has been a more complex problem because of the frequency of phenocopies. A recent study by Mecklin [31] attempted to correlate numbers of CRC and other cancers in patients with different mutations of *MSH2* and *MLH1* genes.

Finally, an extremely important question that may be addressed by multidisciplinary studies in the future, includes the number of polygenic loci that are involved in susceptibility to common CRC, and how these genes interact with each other and the environment. Molecular epidemiology [32] offers a framework for investigating both susceptibility markers, inter-

mediate markers of disease, and biomarkers of environmental effect. A recent example of this type of approach is that of Spitz and associates [33] who conducted a study of 108 patients with upper aerodigestive tract cancers, and an equal number of age and sex-matched controls. They examined the effect of environmental factors on DNA repair capacity as measured by cytogenetic analysis of chromosomal breakage in the presence of mutagen (bleomycin). They found that mutagen sensitivity was increased in the cancer patients (69%) compared with controls (44%). Interestingly, they found evidence for an interaction between mutagen sensitivity (odds ratio 3.2) and cigarette smoking (odds ratio 8.1), with a combined odds ratio of 23. It will not be surprising to find similar types of interactions in the future between biomarkers, such as DNA repair, genetic markers, and measures of environmental exposure in CRC. These types of studies have great potential for improving our understanding of CRC, with consequent implications for prevention and treatment.

In summary, genetic epidemiology has provided important methodological approaches to understanding genetic factors that predispose to CRC. CRC encompasses a number of disorders with different aetiological bases. While there are several inherited syndromes that involve colon cancer, there appears to be a genetic predisposition even to common colon cancer. We have reviewed exemplary studies that have demonstrated the importance of family history of CRC and genetic markers in understanding aetiology and in developing criteria for improved risk assessment for relatives.

1. Khoury MJ, Beaty TH, Cohen BH. *Fundamentals of Genetic Epidemiology*. New York, Oxford University Press, 1993.
2. Burt RW, Bishop DT, Cannon-Albright L, *et al.* Population genetics of colonic cancer. *Cancer* 1992, **70**, 1719–1722.
3. Peipins LA, Sandler RS. Epidemiology of colorectal adenomas. *Epidemiol Rev* 1994, **16**, 273–297.
4. Woolf CM. A genetic study of carcinoma of the large intestine. *Am J Hum Genet* 1958, **10**, 42–47.
5. Duncan JL, Kyle J. Family incidence of carcinoma of the colon and rectum in north-east Scotland. *Gut* 1982, **23**, 169–171.
6. Rozen P, Fireman Z, Figer A, Legum C, Ron E, Lynch H. Family history of colorectal cancer as a marker of potential malignancy within a screening program. *Cancer* 1987, **60**, 248–254.
7. Fuchs CS, Giovannucci EL, Colditz GA, Hunter DJ, Speizer FE, Willett WC. A prospective study of family history and the risk of colorectal cancer. *N Engl J Med* 1994, **331**, 1669–1674.
8. Mellemgaard A, Jensen OM, Lynge E. Cancer incidence among spouses of patients with colorectal cancer. *Int J Cancer* 1989, **44**, 225–228.
9. Cannon-Albright LA, Thomas A, Goldgar DE, *et al.* Familiality of cancer in Utah. *Cancer Res* 1994, **54**, 2378–2385.
10. Slattery ML, Kerber RA. Family history of cancer and colon cancer risk: the Utah Population Database. *J Natl Cancer Inst* 1994, **86**, 1618–1626.
11. Cannon-Albright LA, Skolink MH, Bishop DT, Lee RG, Burt RW. Common inheritance of susceptibility to colonic adenomatous polyps and associated colorectal cancers. *N Engl J Med* 1988, **319**, 533–537.
12. Bailey-Wilson JE, Elston RC, Schuelke GS, *et al.* Segregation analysis of hereditary nonpolyposis colorectal cancer. *Genet Epidemiol* 1986, **3**, 27–38.
13. Houlston RS, Collins A, Slack J, Morton NE. Dominant genes for colorectal cancer are not rare. *Ann Hum Genet* 1992, **56**, 99–103.
14. Ponz de Leon M, Scapoli C, Zanghieri G, Sassatelli R, Sacchetti C, Barrai I. Genetic transmission of colorectal cancer: exploratory data analysis from a population based registry. *J Med Genet* 1992, **29**, 531–538.
15. Scapoli C, Ponz de Leon M, Sassatelli R, *et al.* Genetic epidemiology of hereditary non-polyposis colorectal cancer syndromes in Modena, Italy: results of a complex segregation analysis. *Ann Hum Genet* 1994, **58**, 275–295.
16. Bodmer WF, Bailey CJ, Bodmer J, *et al.* Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature* 1987, **328**, 614–616.
17. Kinzler KW, Nilbert MC, Su L-K, *et al.* Identification of FAP locus genes from chromosome 5q21. *Science* 1991, **253**, 661–665.
18. Groden J, Thliveris A, Samowitz W, *et al.* Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* 1991, **66**, 589–600.
19. Peltomaki P, Aaltonen LA, Sistonen P, *et al.* Genetic mapping of a locus predisposing to human colorectal cancer. *Science* 1993, **260**, 810–812.
20. Aaltonen LA, Peltomaki P, Leach FS, *et al.* Clues to the pathogenesis of familial colorectal cancer. *Science* 1993, **260**, 812–816.
21. Lindblom A, Tannergard P, Werelius B, Nordenskjold M. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nature Genet* 1993, **5**, 279–282.
22. Nicolaides NC, Papadopoulos N, Liu B, *et al.* Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* 1994, **371**, 75–80.
23. Ilett KF, David BM, Detchon P, Castleden WM, Kwa R. Acetylation phenotype in colorectal carcinoma. *Cancer Res* 1987, **47**, 1466–1469.
23a. Liu B, Farrington SM, Petersen GM, *et al.* Genetic instability occurs in the majority of young patients with colorectal cancer. *Nature Med* 1995, **1**, 348–352.
24. Houlston RS, Murday V, Harocopos C, Williams CB, Slack J. Screening and genetic counselling for relatives of patients with colorectal cancer in a family cancer clinic. *Br Med J* 1990, **301**, 366–368.
25. Mecklin J-P. Frequency of hereditary colorectal carcinoma. *Gastroenterology* 1987, **93**, 1021–1025.
26. Aaltonen LA, Sankila R, Mecklin J-P, *et al.* A novel approach to estimate the proportion of hereditary nonpolyposis colorectal cancer of total colorectal cancer burden. *Cancer Detect Prevent* 1994, **18**, 57–63.
27. Gurbuz AK, Giardiello FM, Petersen GM, *et al.* Desmoid tumors in familial adenomatous polyposis. *Gut* 1994, **35**, 377–381.
28. Giardiello FM, Krush AJ, Petersen GM, *et al.* Phenotypic variability of familial adenomatous polyposis in 11 unrelated families with identical APC gene mutation. *Gastroenterology* 1994, **106**, 1542–1547.
29. Olschwang S, Tiret A, Laurent-Puig P, Muleris M, Parc R, Thomas G. Restriction of ocular fundus lesions to a specific subgroup of APC mutations in adenomatous polyposis coli patients. *Cell* 1993, **75**, 959–968.
30. Nugent KP, Phillips RKS, Hodgson SV, *et al.* Phenotypic expression in familial adenomatous polyposis: partial prediction by mutation analysis. *Gut* 1994, **35**, 1622–1623.
31. Mecklin J-P. Molecular genotype/clinical phenotype relationship in HNPCC. *Genetics of FAP and HNPCC*, EuroFAP Meeting Abstracts, 1994.
32. Schulte PA, Perera FP. *Molecular Epidemiology. Principles and Practices*. San Diego, Academic Press, 1993.
33. Spitz MR, Fueger JJ, Halabi S, Schantz SP, Sample D, Hsu TC. Mutagen sensitivity in upper aerodigestive tract cancer: a case–control analysis. *Cancer Epidemiol Biomarker Prev* 1993, **2**, 329–333.

Acknowledgements—The author's research is supported by a Faculty Development Fund award, the Clayton Fund, Cancer Research Foundation of America, and NIH grants R01 CA63721 and R01 HG01197.