Prevention of Colorectal Cancer in Scotland: Strategies for those at Increased Genetic Risk

Rory J. Mitchell

Thesis Presented for the Degree of Doctor of Philosophy
University of Edinburgh
2004

Declaration

I hereby declare,

(i) That this thesis is composed by myself

(ii) That the work presented within this thesis is my own unless otherwise stated

(iii) That this work has not been submitted for any other degree or professional qualification

Rory J. Mitchell

February 2004

Acknowledgements

Many thanks to Harry Campbell, Mary Porteous and Malcolm Dunlop, for their supervision and support over the course of my studentship.

I would like to acknowledge the following people for their contribution to the research presented in this thesis:

Susan Farrington for her help in assessing pathogenicity of mismatch repair gene variants. Norma Brown and Alison Fordyce for their work in conducting interviews and constructing pedigree files during the initial construction of the family history data set. David Brewster and the Staff at ISD for organising the record linkage process, and for making me welcome at the ISD HQ. Jeff Lloyd for programming the computer model, and Gus Ferguson for his collaboration and advice throughout the computer modelling project.

Thanks also to Robin Mitchell and Richard Storton for help with proofreading my thesis, and extra thanks to Helen Rice for proofreading and providing support, encouragement and lasagne during my write-up.

A special thank-you to Rosa Bissett, who knows everything and was always willing to help me out. Thanks also to Rachel, Maureen and the rest of the SOCCS team.

Finally, I'd like to acknowledge Helen Eborall, Lisa Hanna, Lorna Sibbett, Igor Rudan, and the rest of my "PhD peers", for all their support.

Abbreviations

CI = Confidence Interval

COGS = Colorectal Cancer Genetic Susceptibility Study Programme

CSO = Chief Scientist Office

FAP = Familial Adenomatous Polyposis

FDR = First Degree relative

FOBT = Faecal Occult Blood Test

GUI = Graphical User Interface

HNPCC = Hereditary Non-Polyposis Colorectal Cancer

ISD = Information and Statistics Division

MMR = Mismatch Repair

MSI = Microsatellite Instability

NYSIIS = New York State Intelligence Information System

OR = Odds Ratio

OR_{MH} = Mantel Haenszel estimator for the Odds Ratio

RR = Relative Risk

SCR = Scottish Cancer Registry

SDR = Second Degree Relative

UML = Unified Modelling Language

Abstract

The identification of people at increased genetic risk of colorectal cancer and the provision of appropriate clinical screening represents one approach to the prevention of colorectal cancer in the Scottish population. This thesis aims to contribute to current knowledge regarding the available tools for identifying those at increased genetic risk in a population, namely genetic testing and family history assessment.

Key issues relating to the use of family history in this context were addressed through the analysis of a unique data set, comprising family history information reported by a colorectal cancer case or control subject at interview and the results of record linkage of this data to the Scottish Cancer Registry. Retrospective family history case-control analysis showed that individuals with an affected first-degree relative were at an increased risk of developing colorectal cancer (OR_{MH} 2.14, 95% CI = 1.11, 4.14). Prevalence of such a family history in control subjects was 9.4% (95% CI = 4.9, 13.9). Substantial under-reporting of family history was evident, with sensitivity of interview as a means of determining a history of colorectal cancer in a first-degree relative being approximately 0.55 for both cases and controls. These studies illustrate the potential advantages of targeting people with a family history, but also highlight some of the limitations of such an approach.

The genetic epidemiology of the mismatch repair genes hMLH1 and hMSH2 and their association with colorectal cancer was considered in a systematic literature review. Although conventional epidemiological studies are lacking, there is compelling evidence to implicate mutations in these genes in the aetiology of a sub-set of colorectal cancers, with penetrance of approximately 80% in males and 40% in females. A total of 550 different published gene variants were identified, and this high degree of heterogeneity was illustrated in a unique database. This review indicates that carriers of mismatch repair gene mutations merit particular consideration in the context of colorectal cancer prevention through targeting people at increased genetic risk.

Accordingly, the challenge of identifying asymptomatic mismatch repair gene mutation carriers in Scotland was addressed through the development of a computer model of cascade genetic testing, a strategy in which a mutation is identified in one family

member and systematically traced through a pedigree. The model predicts that application of cascade genetic testing to colorectal cancer cases < 55 years of age over a twenty- year period would involve testing 7142 patients and 849 relatives of known carriers, and would identify 321.2 (95%CI = 305.3, 337.1) asymptomatic mutation carriers, representing approximately 27% of the estimated 1209 carriers in Scotland. Model outcomes were highly sensitive to the prevalence and penetrance of mutations, and the participation rates of those offered testing. Overall, outcomes from this computer model suggest that cascade genetic testing is potentially a useful means of identifying asymptomatic mismatch repair gene mutation carriers in Scotland. Follow-up work should ensure that it is also of practical importance as a tool for planning research and health policy.

Identification and screening of mismatch repair gene mutation carriers is an important approach to colorectal cancer prevention, but is only relevant to a minority of people at increased genetic risk. Hence, despite inherent limitations, family history remains a crucial tool for genetic risk assessment in a population. An integrated approach to the prevention of colorectal cancer through targeting people at increased genetic risk can potentially provide substantial health benefits to a sub-group of the population, and thus contribute to the overall prevention of colorectal cancer in Scotland.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Genetics of Colorectal Cancer: A Public Health Perspective

Many common complex diseases have a hereditary genetic component to their aetiology, and hence a proportion of healthy individuals in a population are at increased genetic risk of developing a specific condition. At present, there is considerable interest in understanding the genetic basis of disease susceptibility, driven both by the desire to further scientific knowledge, and by the hope that this knowledge may lead to novel treatments and prevention strategies. One strategy for disease prevention is to identify people at increased genetic risk and thus provide an opportunity for intervention, either to reduce the absolute risk of disease to that individual or to identify and treat the disease at an early stage. Previously, people at increased genetic risk were identified solely on the basis of their family history, but recent advances in understanding of molecular genetics have provided the potential for more accurate risk assessment.

Colorectal cancer provides a paradigm for the use of genetic information to reduce the burden of a common complex disease. It is a significant public health problem, and a proportion of cases are directly attributable to genetic predisposition. The molecular basis of this predisposition is now understood to some extent, although clinically, most "high-risk" individuals are identified through their family history. Furthermore, clinical intervention strategies, considered in detail later, are available for prevention or early detection of colorectal tumours, and these can be targeted to the sub-group of the population at increased genetic risk.

Currently, devising and implementing strategies for preventing colorectal cancer through targeting people at increased genetic risk provides a major challenge and a significant opportunity for the scientific and medical community. Rapidly advancing knowledge of the molecular genetics involved offers huge potential in this regard, but genuine health benefits will only result if such knowledge is applied accurately, appropriately, and in a manner that is complementary to existing strategies based on familial aggregation. Conversely, the relevance and application of family history information must be further evaluated, and considered in the context of genetic testing.

1.2 Descriptive Epidemiology of Colorectal Cancer

Worldwide, colorectal cancer is a major public health problem, with a current annual incidence approaching 950,000 cases (118). The incidence rate of colorectal cancer increases with advancing age, and the disease is more common in males than in females. Globally, incidence rates are variable and are about four times higher in developed countries than in developing countries (119).

These trends are reflected in the cancer statistics for Scotland, where the annual number of colorectal cancer registrations currently exceeds 1800 cases in males, and 1600 cases in females, making this the third most common malignancy in both sexes (123). Over the ten-year period up to 1998, incidence rates have increased by 22.8% in males and 2.4% in females (123), an observation that may be partly due to the ageing population in Scotland. Incidence increases with age in both sexes, being exceedingly rare in people under thirty

years of age, and increasing rapidly beyond age 50 (102). Overall, lifetime risks of developing colorectal cancer are 5.2% and 4.6% in males and females respectively, and annual mortality over the last five years for which data are available (1997-2001) averaged 856 deaths in males and 780 deaths in females (123).

Colorectal cancer therefore presents a significant public health problem both globally and within Scotland, and consequently much research has been conducted with the aim of identifying and optimizing strategies for prevention of this disease.

1.3 Clinical Characteristics of Colorectal Cancer

The development of most colorectal cancers is considered to follow an adenoma-adenocarcinoma sequence (108). Broadly speaking, tumorigenesis follows the "two-hit" principle originally proposed by Knudson (144). The earliest detectable lesion is a disruption of the normal micro-architecture of the colon to form Aberrant Crypt Foci, which in turn may develop into adenomatous or non-adenomatous polyps. Adenomatous polyps occur quite frequently in the general population (311), and in many cases do not develop further. However, such lesions have the potential to undergo malignant transformation, thus completing the sequential development of colorectal carcinoma.

The process of colorectal cancer tumorigenesis is a slow one, with the timeframe for the adenoma-carcinoma sequence being in the region of 10-15 years. However, symptoms often do not develop until the process is well advanced. These factors provide a rationale

for prevention of colorectal cancer through screening for early lesions, which is considered in detail later.

In the absence of screening, patients first present to their general practitioners with symptoms suggestive of colorectal cancer, and are referred to secondary care for a full diagnostic work-up and treatment. The principal treatment method for colorectal cancer is surgical excision, and this may be supplemented with radiotherapy and/or chemotherapy. As with all cancerous lesions, the key to optimal treatment is early detection. To some extent, early diagnosis may be achieved by raising public awareness of the symptoms and risks of colorectal cancer. Providing rapid diagnosis and immediate treatment is, of course, also beneficial. Unfortunately, due to the late onset and non-specific nature of colorectal cancer symptoms, the majority of cases are not diagnosed until they reach an advanced stage, by which time metastasis may have taken place and removal of the primary tumour may be difficult or impossible. Overall, five-year survival following a diagnosis of colorectal cancer in Scotland is approximately 45% (123), with the major factor determining the success of treatment being stage at diagnosis.

1.4 Environmental Factors Influencing Colorectal Cancer Risk

While incidence rates are subject to geographical variation (119), and may also vary according to ethnicity (9), and there is compelling evidence that the observed differences are primarily due to the role of environmental factors. This hypothesis is supported by the rising incidence of colorectal cancer in populations undergoing rapid economic

development, with associated "westernization" of diet and lifestyle (118, 251). Further evidence for a strong environmental influence comes from migrant data; despite the relatively low incidence of colorectal cancer in Japan, incidence rates in Hawaiian Japanese are among the highest in the world (217). Accordingly, considerable effort and resources have been expended with the aim of elucidating the precise dietary and other variables responsible for the observed environmental influences on colorectal cancer incidence.

A report commissioned by the World Cancer Research Fund and the American Institute for Cancer Research concluded that evidence was sufficient to suggest that adherence to a diet high in vegetables and low in meat, together with regular physical activity and avoidance of alcohol, could substantially reduce colorectal cancer risk (314). Other reviews (223) and publications by the International Agency for Research on Cancer (124-126) have reached similar conclusions, although direct evidence supporting the hypothesis that dietary modification can prevent colorectal cancer is not available. Clinical intervention studies (8, 247) and observational cohort studies (79), as well as studies utilizing animal models (234, 281), have shown no evidence of polyp prevention related to diet. Nonetheless, polyp prevention may not be the best endpoint, so results of further clinical studies with cancer prevention as the endpoint are awaited.

In addition to dietary factors, an association between hormone replacement therapy and colorectal cancer has been reported by a number of case-control and cohort studies, with

the majority of these providing evidence in favor of a protective effect (223). The weight of evidence also suggests that smoking may be a significant risk factor (88).

Physical activity has consistently been associated with reduction of colorectal cancer risk (24, 61, 90, 101, 193, 208, 255, 273), although precise data relating to the type and duration of exercise required and the magnitude of the effect on risk remains comparatively scarce. Accumulating evidence also implicates obesity as a risk factor for colorectal cancer (85, 252), and a positive association may exist between colorectal cancer and diabetes (155, 156). It has been hypothesized that hyperinsulinaemia may be an important risk factor for colorectal cancer, and that this association may explain the observed effects of obesity, exercise and diabetes on colorectal cancer risk (87, 89, 143, 208).

1.5 Genetic Susceptibility to Colorectal Cancer

Colorectal cancer is a multifactorial condition, and while environmental factors are clearly important in the aetiology of the disease, there is a significant input from genetic factors. Twin studies have provided classic epidemiological evidence for genetic susceptibility to colorectal cancer (106, 168), with a recent large-scale twin study suggesting that about 35 percent of all colorectal cancer cases have a genetic component to their aetiology (168). The most compelling evidence of genetic susceptibility, however, comes from observations of familial aggregation and molecular genetic studies.

1.5.1 Familial Aggregation

The genetic factors involved in colorectal cancer are poorly understood and may include dominant genes, recessive genes, pathogenic mutations of low penetrance, and complex gene-gene and gene-environment interactions. Recently, research has begun to unravel some of these factors, particularly those that confer genetic susceptibility in an autosomal dominant manner. From an historical perspective, however, it is the tendency for familial aggregation of colorectal cancer that has formed the basis of our understanding of the genetic nature of the disease. The observed patterns of familial aggregation, and the associations between this aggregation and other clinical features, have traditionally provided a means of classification for hereditary cancer syndromes, and have informed genetic risk assessment and clinical practice. Although new understanding regarding the molecular genetic factors involved has prompted some revision of previous classification and clinical management, familial aggregation continues to be the mainstay of clinical genetics, providing a simple means of identifying people at increased genetic risk and evaluating their risk.

1.5.1.1 Rare Hereditary Colorectal Cancer Syndromes

Several "cancer family syndromes" can be clearly defined on the basis of their clinical characteristics. The most common of these syndromes is Familial Adenomatous Polyposis (FAP), which is characterized by a very early development of multiple adenomas, ranging up to several thousand in number. FAP is inherited in an autosomal dominant fashion. Penetrance approaches 100%, and the average age at onset of cancer is around 44 years.

Other rare syndromes which confer an increased risk of colorectal cancer include Peutz-Jeghers syndrome, characterized by mucocutaneous pigmentation and hamartomatous polyps of the large bowel, and Juvenile Polyposis, in which hamartomatous polyps occur during childhood.

Collectively, these syndromes account for approximately 0.1% of colorectal cancer cases (223). They are thus of limited importance from a wider public health perspective, and affected families have been extensively studied on account of the relative ease of clinical diagnosis. Consequently, this thesis is not concerned with the above syndromes, and focuses instead on the more common Hereditary Non-Polyposis Colorectal Cancer syndrome (HNPCC), and the wider aspects of familial aggregation and genetic risk of colorectal cancer.

1.5.1.2 Hereditary Non-Polyposis Colorectal Cancer (HNPCC)

As the name suggests, Hereditary Non-Polyposis Colorectal Cancer syndrome is a term applied to colorectal cancer cases that are apparently hereditary in origin but do not exhibit extensive polyposis and are thus not considered to have one of the above syndromes. In the HNPCC syndrome, affected kindreds have an unusually high occurrence of colorectal and certain extracolonic cancers, with a relatively early age of onset. Historically, the first documented HNPCC kindred was identified by the pathologist Alfred Warthin in 1913 (298), who was alerted to the presence of the affected family by the prediction of his seamstress, based on her family history, that she would die of cancer. The study of this

family was re-visited by Lynch and colleagues in the 1960's and 70's (183, 185). Henry T. Lynch is credited with characterizing the HNPCC syndrome; hence the term 'Lynch Syndrome' which is used to describe the condition in some publications.

The most common extracolonic malignancy that occurs as part of the HNPCC spectrum is endometrial cancer, and other tumours that occur more rarely include those of the ovary, stomach, hepatobiliary tract, brain, skin, renal pelvis and ureter. Other clinical and pathological features have been associated with colorectal tumours occurring in people with HNPCC, as opposed to ostensibly 'sporadic' colorectal cancer. These include poor differentiation, an excess of signet-cell and mucoid features, infiltrating leukocytes and a propensity for the right-sided colon. However, none of these features are pathognomic, and hence HNPCC has traditionally been diagnosed and defined exclusively on the basis of family history.

The limitations of family history as a means of diagnosing or defining a clinical condition are well recognized, and are considered in more detail later. Essentially, the number of cases within a family will depend upon chance and family size, as well as the extent of familial risk. Definitions of familial risk are thus open to interpretation, and consequently various criteria have been used in the past to define people at high risk, or to diagnose people with HNPCC. In 1991 a set of criteria were proposed by the International Collaborative Group on HNPCC, with a view to standardizing criteria used for epidemiological studies, and thus facilitating unbiased comparisons (283). The Amsterdam

Criteria, as these proposed criteria became known, have become widely accepted as the basis for a diagnosis of empirically defined HNPCC. Subsequently, revised criteria, referred to as the Modified Amsterdam Criteria, or Amsterdam II, have been published with a view to including families that exhibit extracolonic cancers as part of the HNPCC syndrome (15). Slightly different criteria have been used in several studies in Asia (Japanese Criteria)(82), and the less stringent "Bethesda" criteria have also been suggested to include a larger proportion of the spectrum of people at increased genetic risk (236). These criteria are presented in table 1.1.

Table 1.1 Clinical Criteria for Diagnosis of HNPCC

Name of criteria	Specific criteria	Published reference
Amsterdam	Three relatives with colorectal cancer, one of which is a first-degree relative of the other two; colorectal cancer affecting more than one generation; at least one colorectal cancer case diagnosed before age 50 years	Vasen et al. (283)
Modified Amsterdam*	Two colorectal cancer cases in first-degree relatives in very small families that cannot be expanded further; colorectal cancer affecting more than one generation; at least one colorectal cancer case diagnosed before age 55 years	Bellacosa et al. (15)
	Two first-degree relatives affected by colorectal cancer, plus a third relative with an unusually early-onset neoplasm or endometrial cancer	
Japanese†	Three or more colorectal cancer cases among first-degree relatives	Fujita et al. (82)
	Two or more colorectal cancers among first-degree relatives and any of the following: diagnosis before age 50 years; right colon involvement; synchronous or metachronous multiple colorectal cancers; association with extracolonic malignancy	
Bethesda*	Individuals from families that fulfill the Amsterdam criteria	Rodriguez-Bigas et al. (236)
	Individuals with two HNPCC-related cancers, including synchronous and metachronous colorectal cancers or associated extracolonic cancers	
	Individuals with colorectal cancer, plus colorectal cancer and/or HNPCC-related extracolonic cancer and/or colorectal adenoma in a first-degree relative; at least one of the cancers diagnosed before age 45 years and the adenoma diagnosed before age 40 years	
	Individuals with colorectal or endometrial cancer diagnosed before age 45 years	
	Individuals with right-sided colorectal cancer with an undifferentiated histopathologic pattern (solid/cribiform) diagnosed before age 45 years	
	Individuals with signet-ring cell type colorectal cancer diagnosed before age 45 years	
	Individuals with colorectal adenomas diagnosed before age 40 years	

* Fulfillment of any of the criteria listed is sufficient. † Cases can be classified as fulfilling either the first set of criteria or the second set and can be diagnosed with hereditary nonpolyposis colorectal cancer if they fulfill either set of criteria. Abbreviation: HNPCC, hereditary nonpolyposis colorectal cancer.

For research purposes, the Amsterdam criteria are the most widely used, and by this definition of HNPCC, the syndrome may account for 1–5 percent of all colorectal cancer cases (49, 138, 195).

The lifetime risk of colorectal cancer for an individual with HNPCC defined solely on the basis of meeting the empirical Amsterdam family history criteria is approximately 40% (291), and the average age of onset is approximately 44 years. In relative terms, the risk of colorectal cancer for an individual with HNPCC, as compared to the general population, will decline with age. Voskuil et al., found that the cumulative incidence ratio in HNPCC family members compared to the general population decreased from 148 at age 40 to 11 by age 75 (291). In reality, where the cause of HNPCC in a family is a gene defect inherited in a true autosomal dominant fashion, the family will comprise a mixture of people at an extremely high risk, and people without the defect, who will have an actual risk no greater than the rest of the general population. As discussed later, the ability to make this distinction is one of the major advantages of genetic testing.

1.5.1.3 Familial Risk In People Not Meeting Empirical Criteria For HNPCC

The complex molecular basis of genetic susceptibility is likely to involve numerous genes, which may be expressed in a recessive or dominant manner with varying penetrance. These genes may interact with each other, or with environmental risk factors. The extent of the genetic contribution to the aetiology of colorectal cancer, and the precise nature of that contribution, will therefore vary considerably between different cases. Accordingly, a spectrum of genetic risk is evident in a given population, and thus individuals with a family history of colorectal cancer that does not meet the specified criteria for HNPCC may still be at increased genetic risk of colorectal cancer. Quantifying the extent of genetic risk associated with a specific family history presents a major challenge both in the research and clinical context, and a body of published evidence addressing this issue is now available.

Numerous studies have suggested that a relatively high number of colorectal cancer patients have a family history of the disease, as opposed to controls, and such observations suggest a degree of heritability (153, 157). Better quality evidence is now available from case-control (22, 25, 70, 141, 149, 153, 157, 160, 205, 206, 222, 239, 256, 260, 272) and cohort studies (39, 78, 92, 136, 179), which have directly compared the colorectal cancer experience of relatives of cases with relatives of controls, or with the general population, thus quantifying the magnitude of the relative risk. A systematic review and meta-analysis of such studies, conducted by Johns et al., produced an estimate of relative risk for a first-degree relative of a colorectal cancer case of 2.25 (95% CI = 2.00, 2.53) (134). The same meta-analysis showed that individuals with more than one affected first-degree relative have a relative risk of 4.25 (95% CI = 2.40, 6.22), and that people with a first-degree relative diagnosed at less than 45 years of age have a relative risk of 3.87 (95% CI = 2.40, 6.22) (134).

Broadly speaking, familial risk increases with extent of family history, i.e. the number of affected relatives and the degree of relationship with the person in question, and the risk is also greater when the affected relative(s) were diagnosed at a young age. Although the above studies have provided information on the relative risk relating to various sub-groups of the population with a family history of colorectal cancer, assessment of the absolute risk of colorectal cancer in an individual is more complex and requires a skilled judgment from a clinical geneticist.

1.5.2 Molecular Genetics

Recent advances in understanding of genetics, and in the available technology for research in the field of genetic epidemiology, have made the definition of the molecular changes that underlie genetic susceptibility to complex disease a realistic goal.

Polymorphisms in genes coding for enzymes involved in carcinogen metabolism have been implicated in modification of colorectal cancer risk. Several carcinogens, including dietary heterocyclic amines, act as substrates for the N-acetyltransferase enzymes NAT1 and NAT2, providing a clear rationale for an association between polymorphisms in these genes and risk of colorectal cancer. However, evidence in support of this hypothesis is inconclusive, with recent literature reviews and meta-analyses concluding that no consistent association between N-acetyltransferase and colorectal cancer risk has been demonstrated (28, 48, 320) and that further studies in this area are required. Similarly, the role of the glutatione S-transferase family of enzymes in the metabolism of a range of carcinogens has prompted the investigation of the role of polymorphisms in susceptibility to colorectal cancer. Again, results have been inconsistent, and a systematic literature review (43) and a recent meta-analysis (321) provided no support to the hypothesis that polymorphisms in these genes influence colorectal cancer risk. Polymorphisms in various other genes, including Methyltetrahydrofolate reductase, Apolipoprotein E and Tumour Necrosis Factor-α have also been implicated in colorectal cancer susceptibility.

Whilst identification of polymorphisms that modify colorectal cancer risk is important from the broad perspective of understanding the genetic epidemiology of colorectal cancer, the effects are generally small or moderate, and are not well understood. Consequently, the genes considered above are not currently of clinical relevance, and the role of molecular genetics in current strategies for prevention of colorectal cancer focuses on autosomal dominant colorectal cancer syndromes.

Families exhibiting an autosomal dominant colorectal cancer syndrome are amenable to linkage analysis, which provides a useful method of determining the chromosomal location of the gene responsible. This technique was used to localize the gene for Familial Adenomatous Polyposis (18, 163), facilitating the subsequent identification of mutations in the adenomatous polyposis coli gene as the cause of this syndrome (95). Progress has also been made in understanding the genetic basis of other rare cancer family syndromes. Pathogenic mutations in the STK11 gene are evident in a large proportion of families with Peutz-Jehers syndrome (107, 303, 322), although this gene has been excluded as a cause in some families, suggesting a degree of genetic heterogeneity (17). Similarly, approximately 50% of families exhibiting the Juvenile Polyposis Syndrome have mutations in the SMAD4 gene, although the BMPR1A gene has also been implicated in some such families, and in others the genetic basis remains unknown (113, 238, 313).

From a public health perspective, the most important hereditary colorectal cancer syndrome is HNPCC, and consequently the genetic basis of this syndrome has been the subject of

much research. Recently, it has been established that a large proportion of families diagnosed with HNPCC harbor potentially pathogenic mutations in mismatch repair genes. This group of genes code for a complex of proteins that play an integral role in the recognition and repair of errors in DNA replication.

The originally identified, and best understood, mismatch repair system has been described in the bacterium, *Escherichia coli.* Human homologues of the mismatch repair genes present in this organism have since been identified. This process is reflected in the nomenclature, with human mismatch repair genes being classified according to their homology with the genes that comprise the bacterial system, as shown in table 1.2.

Table 1.2 Classification and Nomenclature of Human Mismatch Repair Genes

Bacterial Gene	Human Homologues
MutS	human MutS Homologue 2 (hMSH2)
	human MutS Homologue 3 (hMSH3)
	human MutS Homologue 6 (hMSH6)
MutL	human MutL Homologue 1 (hMLH1)
	Post Meiotic Segregation Increased 1 (PMS1)
	Post Meiotic Segregation Increased 2 (PMS2)
MutH	Proliferating Cell Nuclear Antigen (PCNA)
	Methyl-CpG-Binding Endonuclease
MutY	human MutY Homologue 1 (hMYH)

The precise mechanisms involved in the human mismatch repair system have yet to be ascertained. Other, as yet unidentified, mismatch repair genes may be involved, and

interactions with other genes may also have a role. However, a broad model for the function of the system has been established.

The primary role of the three MutS homologues is in the recognition and binding of any errors that may have occurred during DNA replication. hMSH2 can bind with one of two other mismatch repair proteins creating heterodimers comprising either hMSH2/hMSH6 or hMSH2/hMSH3. The former complex, sometimes referred to as hMutSα, preferentially recognizes mismatches involving only one base pair, whereas the hMSH2/hMSH3 complex, also known as hMutSβ, binds to larger DNA mismatches, typically involving 2-4 bases (147). In either case, a larger heterodimer formed by hMLH1 and hPMS2 is then recruited to excise the incorrect daughter sequence and replace it with the correct sequence, using the parental strand as a template. The hMLH1 protein has no known enzymatic activity itself and probably acts as a "molecular matchmaker," in that it recruits other DNA repair proteins to the mismatch repair complex. The protein product of hMYH is involved in the repair of oxidative DNA damage (259). At present the function of hPMS1 is largely unknown, and the role of other mismatch repair genes, including the putative MutH homologues remains to be characterized.

Of the pathogenic mutations identified so far in mismatch repair genes, the vast majority occur in hMLH1 and hMSH2, an observation that may reflect their central role in the mismatch repair process. Other mismatch repair genes that have been implicated as causes of colorectal cancer are hMSH6, hMYH, PMS1 and PMS2.

As discussed in detail in chapter 3, mutations in mismatch repair genes were originally implicated as the genetic defect underlying the HNPCC syndrome. However, the relationship between these genes and empirically defined HNPCC is not straightforward. Such mutations have been found only in a proportion of families that meet the Amsterdam family history criteria. Conversely, pathogenic mutations in hMLH1 or hMSH2 have also been identified in kindreds that do not meet the traditional criteria for diagnosis of HNPCC. This observation may be due to the inherent misclassification bias involved in diagnosing a condition on the basis of family history alone, particularly in small families. It is also possible that some mismatch repair gene mutations are not detected during analysis for technical reasons. Equally, the incomplete correlation between empirically defined HNPCC and known mismatch repair genes may be indicative of genetic heterogeneity. Kindreds exhibiting the HNPCC syndrome may harbour pathogenic mutations in other, as yet unidentified, mismatch repair genes, in other genes involved in the same biological pathway, or in genes unconnected with mismatch repair.

The association between mismatch repair genes and colorectal cancer is further complicated by the apparent role of microsatellite instability (MSI). MSI, characterized by instability of microsatellite repeats during cell replication, can occur as a consequence of mismatch repair deficiency (264) and is a feature of a large proportion of colorectal tumours occurring as part of the HNPCC syndrome (19). However, ostensibly pathogenic MMR gene mutations have been identified in the absence of MSI, and conversely MSI occurs in a minority of "sporadic" colorectal cancers (19). Thus, microsatellite instability and family history are both associated with the sub-set of colorectal cancers caused by

mismatch repair gene mutations, but are not consistent nor pathognomic features of mismatch repair deficiency.

This complexity and uncertainty is reflected in the current terminology. Strictly speaking, the term HNPCC refers to people meeting the empirical Amsterdam criteria, but since the identification of mismatch repair gene mutations as the causal factor in a proportion of such people, the term has also become synonymous with these mutations. An attempt to address this situation has been made, by suggesting that the term "Hereditary Mismatch Repair Deficiency Syndrome" be used to describe cases in which the cause is known to be a germline mutation in a mismatch repair gene (132). However, this term is not in widespread use, and the somewhat confusing terminology remains.

Since their identification, knowledge and understanding of mismatch repair genes and their role in colorectal cancer has advanced rapidly, but it remains limited in many respects. The mismatch repair genes hMLH1 and hMSH2 are highly heterogeneous, with more than 200 allelic variants identified in each. This factor introduces substantial obstacles to obtaining a clear understanding of the association between mismatch repair gene mutations and colorectal cancer. Complete mutation analysis of these genes is required to determine the known or suspected defects in the individual in question. This is both technically difficult and time-consuming. Consequently, for largely economic and practical reasons, mutation analysis has been carried out almost exclusively among colorectal disease patients, particularly those identified as being at high risk of harboring mutations. Lack of control

data, coupled with the heterogeneous nature of the gene variants concerned, makes the pathogenicity of each variant difficult to establish. Indeed, pathogenicity may vary according to the nature and location of the gene variant involved. Identification of the causal defect is therefore subject to interpretation based on clinical and molecular data that is often insufficient for such purposes.

Despite the above limitations, however, the identification of mismatch repair genes and their role in the aetiology of a subset of colorectal cancer cases provides the potential for improving genetic risk assessment, and as such offers new opportunities for prevention of colorectal cancer through targeting people at increased genetic risk.

1.6 Principles of Colorectal Cancer Prevention

There are several possible approaches to the challenge of reducing the burden of colorectal cancer in the Scottish population, and these can be broadly categorized as primary prevention, chemoprevention and secondary prevention:

1.6.1 Primary Prevention

Primary prevention centers on the reduction of colorectal cancer risk through the avoidance of known risk factors, and the adoption of a healthier lifestyle and diet. Although evidence from intervention studies is limited, it is reasonable, on the basis of the known relative risks of various factors, to suppose that primary prevention could theoretically play a major role in reducing colorectal cancer incidence and mortality. However, the difficulties inherent in

achieving significant lifestyle changes are well recognized, and whilst primary prevention is ultimately the most desirable method of reducing the burden of disease it may not be feasible to achieve a substantial effect in the near future.

1.6.2 Chemoprevention

Chemoprevention using Non-Steroidal Anti Inflammatory Drugs (NSAIDs) is another promising area of research, with the potential to reduce colorectal cancer incidence (13, 16, 63, 94, 103, 120, 176, 233, 244). Both epidemiologic evidence and experiments utilizing murine models have suggested that nonsteroidal antiinflammatory drugs have antitumour properties that may prevent colorectal cancer. Sulindac has been shown to inhibit tumour growth in experimental systems and to reduce adenoma counts in humans with familial adenomatous polyposis (86), as has a recent study of the specific cyclo-oxygenase-2 inhibitor, celecoxib (262). Two recent randomized controlled studies have demonstrated that aspirin reduces the risk of adenoma development or recurrence in people with a personal history of adenomas or colorectal cancer (13, 244). Although such studies provide compelling proof of the chemopreventive properties of NSAIDs, further clinical trials are required to determine if the benefits of this approach outweigh the inherent risks of long-term use. Thus, whilst chemoprevention may have a role as part of future strategies for colorectal cancer prevention it is at present largely confined to the research context.

1.6.3 Secondary Prevention

The adenoma-carcinoma sequence of colorectal cancer development provides a lengthy latent period of tumorigenesis before a symptomatic tumour becomes evident. During this

stage pre-neoplastic lesions and early-stage cancers may be detected. Colorectal cancer is thus an ideal candidate condition for screening, adequately meeting the criteria prepared by the UK National Screening Committee (279), which are in turn based on the principles of screening first outlined by the WHO (309). There is a clear rationale for 'secondary prevention' of colorectal cancer, through screening procedures designed to facilitate the identification and removal of preneoplastic lesions, and the early detection and treatment of existing tumours.

Screening has considerable potential as a means of reducing the disease burden of colorectal cancer, but the appropriate application of screening depends on the techniques involved, their effectiveness and their disadvantages. Several methods are available, that may be used alone, or in combination:

(i) Faecal Occult Blood Test (FOBT)

Intermittent bleeding is a common feature of colorectal adenomas and carcinomas, and may occur relatively early in cancer progression, before the tumour becomes symptomatic. The faecal occult blood test relies on the pseudoperoxidase activity of haemoglobin to test for occult blood in a stool sample. Positive FOBT results can be followed-up using more stringent diagnostic techniques.

The test is non-invasive and therefore has no risks directly associated with it. It is also fairly easy to use, relatively inexpensive, and can be sent through the post. Consequently, FOBT is relatively acceptable to the target population, and can potentially be used for

screening large numbers of people. However, the intermittent nature of bleeding means that FOBT may miss a significant proportion of cancers and precancerous lesions. The sensitivity of a one-time faecal occult blood test is in the region of 30-70%, and varies depending on the setting (at home or in medical practice) and the exact test and protocol used (169, 190, 226, 318). Perhaps more significantly, the FOBT also lacks specificity, since various other conditions can produce faecal blood, and the results can also be influenced by diet and medication. Individuals with a false positive result may be offered more invasive screening and diagnostic procedures unnecessarily, ultimately leading to increased costs and health risks.

(ii) Colonoscopy

Colonoscopy is usually considered to represent the 'gold-standard' for detection of early colorectal tumours and pre-neoplastic lesions. It is an invasive procedure requiring a skilled operator, and permits the visualisation of the majority of the colon. Colonoscopy has an estimated sensitivity of over 90% for detecting cancers and large adenomas (231). Biopsy and polypectomy can be performed at the time of examination. The technique is associated with a small but significant complication rate itself, and the necessity of sedation introduces additional risks. Nonetheless, colonoscopy constitutes the mainstay of secondary prevention of colorectal cancer, particularly in people deemed to be at a particularly high risk of the disease, as considered later.

(iii) Flexible Sigmoidoscopy

Flexible sigmoidoscopy is an invasive procedure used to visualise the descending (sigmoid) colon, and the main disadvantage of this technique is an inability to detect more proximal cancers. Compared to colonoscopy, flexible sigmoidoscopy is associated with fewer complications but is also less effective. The sensitivity of flexible sigmoidoscopy has been estimated at approximately 68-78% (121, 169). Again, biopsies can be taken at the time of examination.

(iv) Double Contrast Barium Enema

This technique provides an alternative, and non-invasive means of visualising the colon, but compares unfavourably with colonoscopy in terms of sensitivity and specificity, and obviously does not facilitate biopsy or polypectomy at the same time as visualization. Consequently, double contrast barium enema is primarily used as a complementary technique or when there are technical difficulties with performing colonoscopy.

(v) Virtual Colonoscopy

Virtual colonoscopy is a relatively new technique that uses helical computed tomography to generate images of the large bowel. The minimally invasive nature of the technique and the associated low complication rate offer considerable advantages over conventional colonoscopy. Virtual colonoscopy has been shown to have broadly comparable sensitivity to conventional colonoscopy in terms of identifying polyps over 6mm in size (64), but detection rates for smaller adenomas compare less favourably (64, 99, 232). At present virtual colonoscopy is expensive and largely limited to the research context, but it remains a promising technique that may form an aspect of future screening programmes (20, 310).

* * * * * *

For each individual who may be offered screening, the objective is to apply a tailored screening protocol that will maximize the benefits and minimize the associated costs, health risks and any adverse psychosocial effects. Ultimately, the major consideration in terms of benefit is the absolute risk of developing colorectal cancer. When absolute risk is low, there is only a very small chance that the individual concerned will develop cancer over the screening period, and consequently it is probable that no benefit will accrue from screening. In contrast, where absolute risk is very high, the likely benefits of screening may be sufficiently large to justify an intensive screening protocol, despite the financial costs and possible side effects of screening.

The principal determinant of absolute risk is age, with the incidence of colorectal cancer increasing markedly later in life. This trend remains consistent for all risk groups, including those deemed to be at increased genetic risk. Thus, young people with a high relative risk of developing colorectal cancer may still be less likely to develop colorectal cancer in the next ten years than older members of the general population. Consequently, the potential target groups for secondary prevention of colorectal cancer through screening include both individuals at increased risk and older members of the general population at average risk. It is necessary, therefore, to consider people at increased risk, whether due to genetic predisposition or environmental / dietary risk factors, within the context of any wider screening programme. The key issue is whether or not such people will benefit from more intensive screening, or intervention at a younger age.

1.7 Screening for Colorectal Cancer in People at Average Risk

An important consideration regarding screening in average-risk individuals is the large number of people who are eligible for such screening. In this context issues of cost are particularly relevant. Similarly, the balance between sensitivity and specificity of screening must be carefully weighted to maximize the health benefits to the population, whilst minimizing any adverse health effects. Low specificity of a test applied to a large population will result in numerous individuals with 'false positive' results, who may be unnecessarily subjected to further diagnostic procedures.

The exact age at which the population risk of colorectal cancer becomes sufficiently high to justify screening depends on the technique being employed, and the associated issues of cost, side-effects, and acceptability to the general public. For pragmatic reasons, 50 years of age is often considered as the point beyond which screening may be beneficial.

The non-invasive nature and relatively low cost of the faecal occult blood test makes this a particularly attractive method of screening at the population level. Evidence for the effectiveness of the FOBT in average-risk individuals comes from several large-scale randomised controlled trials (100, 142, 151, 190), with a UK study demonstrating a 15% reduction in cumulative CRC mortality in the screening group (odds ratio = 0.85, 95% CI = 0.74-0.98)(100). A pilot is currently being conducted in the UK, with a view to introducing FOBT screening nationally. Despite these encouraging results, concerns remain regarding the sensitivity of FOBT as a means of detecting colorectal cancer; in the randomised controlled trials considered above the overall sensitivity was in the region of 50%.

Invasive screening is unlikely to be acceptable to the general public or feasible to conduct on a population basis, although there is some evidence to suggest that flexible sigmoidoscopy may be effective in average risk individuals (166, 250).

1.8 Screening for Colorectal Cancer in People at Increased Genetic Risk

The concept of targeting colorectal cancer screening to people at increased genetic risk is well established. Considerable debate surrounds the appropriate screening protocol in this context, which centers on the costs and benefits of screening a particular risk group. As mentioned previously, the absolute risk to the individual being offered screening is the key factor in determining the extent of screening that is appropriate, and this risk is a synthesis of age and genetic predisposition. Genetic predisposition is highly complex and is thus difficult to quantify, meaning that a degree of uncertainty surrounds the calculation of absolute risk. In addition, the appropriate screening protocol for an individual with a known absolute risk is still not clear, as it depends not only on the effectiveness of the protocol, but also on the health risks, financial costs and psychosocial effects of screening.

Despite the uncertainties inherent in determining risk and offering screening to individuals who may be at increased genetic risks, the overall value of this approach is sufficiently high to have encouraged its use in research and in clinical practice. The appropriate protocols for specific rare syndromes such as FAP have been largely established, and due to the relative insignificance of these syndromes at the population level they are therefore not considered further in the context of this thesis. The evidence to support screening in various other risk groups is summarized below.

1.8.1 First Degree Relatives of Colorectal Cancer Patients

First-degree relatives of colorectal cancer patients have been shown to be at an increased risk of developing the disease themselves (25, 39, 78, 92, 136, 141, 149, 153, 157, 160, 179, 205, 206, 222, 239, 256, 260, 272), with a recent meta-analysis estimating the relative risk for people with one affected first-degree relative as 2.25 (95% CI = 2.00, 2.53) (134). In reality, the individuals within this group will have a broad spectrum of risk, ranging from average population risk for those people who have a relative affected by chance, to extremely high risk for those people who have a strong genetic predisposition that has, again by chance, not become manifest as a more extreme family history. This latter situation may arise in people from very small families. Consequently, targeting first-degree relatives of colorectal cancer cases is a somewhat crude approach, and does not involve offering a tailored screening protocol to individuals. Nonetheless, such people constitute a large sub-group of the population who may merit particular attention with respect to colorectal cancer screening.

It has been estimated that around 4-7% of people in the general population have an affected first-degree relative (112, 243, 293). As with people at average risk, the large size of this sub-group is an important consideration. Consequently, much attention has focused on faecal occult blood testing, since this approach meets the necessary criteria for screening a large group of people at a moderately increased risk, being inexpensive and non-invasive. Several studies have demonstrated that FOBT can be used to identify adenomas and carcinomas in individuals with a history of colorectal cancer in at least one first-degree relative (10, 45, 70, 111, 115, 239). Nakama et al., found that 8.5% of subjects with an

affected first-degree relative had a positive FOBT test, compared with 4.8% of subjects without such a history (204). Furthermore, the positive predictive value of the FOBT test for colorectal cancer was 6.8% in people with an affected first-degree relative; significantly greater than the value of 2.4% for people without such a family history (204). Although FOBT is a feasible and potentially effective means of screening in people with an affected first degree relative, this sub-group may contain individuals with greater degrees of family history. Consequently, many authors have queried whether or not the sensitivity and specificity of the test are sufficient to address the needs of this sub-group.

Identifying people with an affected first degree relative is a convenient means of providing a group of people at increased genetic risk, but ideally this group should be further stratified according to their family history. Thus, whilst FOBT may not be adequate as a means of screening all individuals with a history of colorectal cancer in at least one first-degree relative, it may have a role in screening people with only one affected first-degree relative, i.e. those at the lowest end of the increased familial risk spectrum.

Although primary evidence demonstrating the efficacy of colonoscopy as a screening tool for first-degree relatives of colorectal cancer cases is not available, various studies have found that the yield from colonoscopic screening, in terms of identifying adenomas, is greater in people with such a history than in control subjects (96, 111, 269, 316). For example, a controlled, prospective study calculated a relative risk of 3.49 ($P < 0.001$) for harbouring colonic adenomatous polyps among first degree relatives of colorectal cancer patients compared to controls without such a history (96). This implies that more invasive

and intensive screening protocols, involving the use of colonoscopy as an initial screening tool, are likely to be effective in people with one affected first degree relative. Flexible sigmoidoscopy provides another option for screening in this group, and is intermediate between FOBT and colonoscopy in terms of the balance between sensitivity and specificity. However, as with screening older people at average risk, the high cost and the potential for adverse complications means that invasive screening techniques may not be considered appropriate in people with a history of colorectal cancer in a first-degree relative.

Screening protocols for people at slightly increased genetic risk are likely to involve large numbers of people, and may overlap with strategies for people at average risk. Hence, it has been suggested that screening methods used for average risk individuals are also appropriate for those with an affected first-degree relative, but that screening should commence ten years earlier (312).

1.8.2 Strong Family History (More Than One Affected First-Degree Relative, or One First Degree Relative Affected at <45 Years of Age)

Although there is little primary evidence for a reduction in colorectal cancer incidence from screening people with a strong family history, proof of the concept that screening can be effective in this context comes from the observation that the yield from screening such groups is relatively high in comparison to controls groups and / or the general population (96, 312). It is reasonable to infer from such data that benefits in terms of reduced incidence and mortality of colorectal cancer may accrue from screening in this context.

The risk of colorectal cancer for people with more than one first-degree relative with colorectal cancer has consistently been shown to be greater than in people with only one affected first-degree relative (78, 157, 205, 206, 256, 260). In a meta-analysis conducted by Johns and Houlston, the pooled relative risk for this group compared to the general population was 4.25 (95% CI = 3.01, 6.02) (134). Similarly, several studies have shown that relatives of colorectal cancer cases who were diagnosed at less than 45 years of age are at a comparatively high risk (39, 78, 97, 205, 256), with the pooled estimate for this group being 3.87 (95% CI = 2.40, 6.22)(134). Both these figures are markedly greater than the pooled estimate of relative risk for people with one affected first-degree relative, calculated as 2.25 (95% CI = 2.00, 2.53) in the same meta-analysis (134). In view of the high relative risks in these sub-groups, screening methods that are appropriate for people with one affected first-degree relative may be inadequate. Conversely, the intensive screening programmes offered to people who meet the accepted criteria for HNPCC may be inappropriate. People with two affected first-degree relatives, or one affected at <45 years of age thus constitute a convenient sub-group, which merits particular consideration in the context of screening.

FOBT provides one option for screening in this risk group, and many studies have shown this approach to be capable of detecting adenomas and colorectal cancer in people with a family history (10, 38, 45, 70, 111, 115). As mentioned previously, though, FOBT may be inadequate to meet the needs of people with a strong family history. Houlston et al., found the negative predictive value of this test to be 78% in a group comprising people with two affected first-degree relatives or one first-degree relative affected at less than 45 years of

age, and concluded that the test was unsatisfactory (111). Flexible sigmoidoscopy provides another option, but the limitations of this technique in terms of sensitivity and extent of visualisation of the colon may preclude its use in this group. In contrast, colonoscopy is a relatively sensitive technique, and is known to be capable of detecting neoplasms in people with a family history of colorectal cancer (38, 54, 64, 83, 96, 116, 172, 181, 194, 263, 269). Consequently, this is generally considered the method of choice when screening people with a strong family history of colorectal cancer, and is extensively used in clinical practice.

The exact screening protocol that is appropriate for this risk group remains subject to the interpretation of various researchers and policy-makers, who must evaluate the costs and benefits. However, a growing consensus is focussed on the selective use of colonoscopic screening for this group; striking a balance between providing adequate surveillance and limiting the frequency of colonoscopy. A recent paper by Dunlop outlines a strong case for performing colonoscopy at around 35-40 years of age, with the procedure not being repeated until age 55 if findings are normal (56). This recommendation is comparable with current Scottish guidelines for colorectal cancer screening (249) (see appendix A1).

1.8.3 Hereditary Non-Polyposis Colorectal Cancer

The HNPCC syndrome is currently defined on the basis of the Amsterdam family history criteria, and this method of diagnosis may lead to misclassification. Nonetheless, the accepted criteria for HNPCC are a convenient means of identifying people likely to be at a

substantially increased genetic risk, and individuals meeting these criteria are generally considered to be eligible for relatively intensive screening.

Numerous studies have produced evidence to support the use of screening in individuals with a family history of cancer that meets the Amsterdam Criteria for HNPCC (93, 130, 131, 178, 203, 230, 282, 284, 287). Such individuals have a very high risk of developing colorectal cancer, and a large proportion of tumours develop in the proximal colon. Consequently, colonoscopy is the most appropriate screening method in this group (55, 130, 203, 284), and this approach has been specifically shown to facilitate early detection of colorectal cancer (287), and to reduce mortality (130, 230). A screening interval of one to three years is often recommended, although the observation that interval cancers can occur within eighteen months of colonoscopy (285) implies that screening should ideally be at least biennial in order to minimise the chances of such occurrences. Absolute risk of developing colorectal cancer increases rapidly after around 25-30 years of age in HNPCC subjects (3, 284), and screening programmes should begin at this time (34, 55, 312). This consensus is also reflected in the current guidelines for colorectal cancer screening in Scotland (249) (see also appendix A.1).

Prophylactic colectomy provides another potential option for clinical management of HNPCC subjects (96, 235). However, considering that effective screening protocols are available, prophylactic colectomy is not normally justified except for people who do not comply with colonoscopic screening or for whom such screening proves technically problematic (41, 186).

60-64	M	FDR	0.90	0.90	0.90	0.90
65-69	M	FDR	0.90	0.90	0.90	0.90
70-74	M	FDR	0.90	0.90	0.90	0.90
75-79	M	FDR	0.90	0.90	0.90	0.90
80-84	M	FDR	0.90	0.90	0.90	0.90
>85	M	FDR	0.90	0.90	0.90	0.90
0-4	F	FDR	0.90	0.90	0.90	0.90
5-9	F	FDR	0.90	0.90	0.90	0.90
10-14	F	FDR	0.90	0.90	0.90	0.90
15-19	F	FDR	0.90	0.90	0.90	0.90
20-24	F	FDR	0.90	0.90	0.90	0.90
25-29	F	FDR	0.90	0.90	0.90	0.90
30-34	F	FDR	0.90	0.90	0.90	0.90
35-39	F	FDR	0.90	0.90	0.90	0.90
40-44	F	FDR	0.90	0.90	0.90	0.90
45-49	F	FDR	0.90	0.90	0.90	0.90
50-54	F	FDR	0.90	0.90	0.90	0.90
55-59	F	FDR	0.90	0.90	0.90	0.90
60-64	F	FDR	0.90	0.90	0.90	0.90
65-69	F	FDR	0.90	0.90	0.90	0.90
70-74	F	FDR	0.90	0.90	0.90	0.90
75-79	F	FDR	0.90	0.90	0.90	0.90
80-84	F	FDR	0.90	0.90	0.90	0.90
>85	F	FDR	0.90	0.90	0.90	0.90
0-4	M	SDR	0.90	0.90	0.90	0.90
5-9	M	SDR	0.90	0.90	0.90	0.90
10-14	M	SDR	0.90	0.90	0.90	0.90
15-19	M	SDR	0.90	0.90	0.90	0.90
20-24	M	SDR	0.90	0.90	0.90	0.90
25-29	M	SDR	0.90	0.90	0.90	0.90
30-34	M	SDR	0.90	0.90	0.90	0.90
35-39	M	SDR	0.90	0.90	0.90	0.90
40-44	M	SDR	0.90	0.90	0.90	0.90
45-49	M	SDR	0.90	0.90	0.90	0.90
50-54	M	SDR	0.90	0.90	0.90	0.90
55-59	M	SDR	0.90	0.90	0.90	0.90
60-64	M	SDR	0.90	0.90	0.90	0.90
65-69	M	SDR	0.90	0.90	0.90	0.90
70-74	M	SDR	0.90	0.90	0.90	0.90
75-79	M	SDR	0.90	0.90	0.90	0.90
80-84	M	SDR	0.90	0.90	0.90	0.90
>85	M	SDR	0.90	0.90	0.90	0.90
0-4	F	SDR	0.90	0.90	0.90	0.90
5-9	F	SDR	0.90	0.90	0.90	0.90
10-14	F	SDR	0.90	0.90	0.90	0.90
15-19	F	SDR	0.90	0.90	0.90	0.90
20-24	F	SDR	0.90	0.90	0.90	0.90
25-29	F	SDR	0.90	0.90	0.90	0.90
30-34	F	SDR	0.90	0.90	0.90	0.90
35-39	F	SDR	0.90	0.90	0.90	0.90
40-44	F	SDR	0.90	0.90	0.90	0.90
45-49	F	SDR	0.90	0.90	0.90	0.90

50-54	F	SDR	0.90	0.90	0.90	0.90
55-59	F	SDR	0.90	0.90	0.90	0.90
60-64	F	SDR	0.90	0.90	0.90	0.90
65-69	F	SDR	0.90	0.90	0.90	0.90
70-74	F	SDR	0.90	0.90	0.90	0.90
75-79	F	SDR	0.90	0.90	0.90	0.90
80-84	F	SDR	0.90	0.90	0.90	0.90
>85	F	SDR	0.90	0.90	0.90	0.90

Abbreviations: FDR = First Degree relative, SDR = Second Degree Relative

Grant application form	**CSO reference number: CZH/4/145**

Project title (not more than 25 words):

The Development of a Computer Model of Cascade Genetic Screening in Complex Disease

Duration of project *(months)*: 20

Summary of costs:

Staff	Indirect costs *(if applicable)*	Consumables	Travel	Ex. items	Equipment	**Total**
£53,611	£21,444	£10,000	-	-	-	**£85,055**

Principal investigator:

Name and title	Position	Institution
Dr. Harry Campbell	Professor of Genetic Epidemiology	University of Edinburgh

Co-Applicants:

Name and title	Position	Institution
Professor Malcolm Dunlop	Professor of Coloproctology	University of Edinburgh
Dr. Mary Porteous	Consultant & Reader in Clinical Genetics	University of Edinburgh
Dr. R.K. Ferguson	Lecturer & Research Associate (Mathematics & Computer Sciences)	Heriot-Watt University
Professor Angus MacDonald	Professor of Actuarial Mathematics & Statistics	Heriot-Watt University
Rory J. Mitchell	PhD Student	University of Edinburgh

Project summary *(not more than 150 words)*:

A current challenge for the scientific and medical community is to utilise recent advances in our understanding of the genetic basis of complex disease by identifying carriers of pathogenic mutations and offering the appropriate management. Cascade genetic screening, in which mutations are detected in high-risk individuals and subsequently traced through families, is a useful strategy for identifying asymptomatic mutation carriers. The aim of this project is to address a gap in current knowledge by developing a comprehensive computer model to facilitate evaluation of this strategy. The model will initially be designed and validated with respect to mismatch repair gene mutations and colorectal cancer, but it will have sufficient flexibility to apply to any condition in which high penetrance mutations have been identified and effective interventions are available. The computer model will thus provide a valuable tool to inform decisions made by researchers and health policy makers regarding the use of cascade screening.

Introduction

Identification of individuals at an increased genetic risk of developing a particular complex disease can facilitate the targeting of appropriate intervention strategies to this sub-group of the population. Therefore, in conditions for which pathogenic mutations have been identified, and effective intervention strategies are available, genetic testing offers a feasible and effective method for reducing the burden of disease.

For both economic and ethical reasons, it is essential that genetic testing strategies are targeted to the people most likely to benefit. Population genetic screening is rarely justified, and stratified screening, in which individuals with a strong family history of disease are targeted, is subject to various practical and scientific difficulties. In contrast, cascade genetic screening is potentially both feasible and effective as a means of identifying asymptomatic carriers of pathogenic mutations. This process involves conducting mutation analysis in individuals, often patients or a sub-group of patients, who are at a relatively high risk of carrying a mutation. When a mutation carrier is found, genetic testing can be offered to their relatives, and the specific mutation present can thus be traced through an expanded pedigree in a 'cascade' fashion.

The concept of tracing a mutation through families is often used in a clinical genetic setting, and has been applied in the research context to various diseases, including cystic fibrosis (1; 2) and familial hypercholesterolaemia (3). Additionally, Krawczak et al., have produced equations designed to estimate the theoretical efficacy and efficiency of cascade screening (4). However, the outcomes of any cascade screening protocol will depend on numerous factors, including the genetic epidemiology of the condition in question, the available technology for genetic testing and the uptake of genetic testing by at-risk individuals. Furthermore, these factors will be superimposed on the background of population demographics, disease prevalence, other risk factors and existing screening strategies. A comprehensive evaluation of cascade screening in complex disease that considers all the above points has not yet been undertaken. The objective of the proposed project is to address this gap in the collective scientific knowledge. Computer modelling is the ideal approach to such an evaluation, permitting the complex nature of cascade genetic screening to be fully explored.

The computer model will be designed with the flexibility necessary for application to any complex condition, but it will initially be developed and validated with respect to colorectal cancer, building on ongoing research conducted by the applicants. Around 3400 colorectal cancer cases are diagnosed in Scotland each year, and a small but significant sub-set of these are caused by pathogenic mutations in mismatch repair (MMR) genes. The approximate lifetime risk of colorectal cancer in MMR gene mutation carriers if 80% for males and 40% for females (5; 6; 7). The prevalence of pathogenic MMR gene mutations in the Scottish population has been estimated as 1:3139 among the 15-74 years age group (95% CI = 1:1247, 1:7626) (8), implying that 1209 Scots in this age range harbour such mutations (95% CI = 498, 3044). Upon identification, mutation carriers can be offered colonoscopic screening, an intervention shown to have significant health benefits in terms of both reduced incidence rates and life-years saved (9). The UK National Screening Committee have acknowledged the potential of cascade genetic screening for MMR gene mutations (10), and our systematic literature review relating to MMR genes has supported this notion (11). It has been estimated that approximately 35% (95% CI = 10%, 38%) (12) of colorectal cancer cases have a primarily genetic basis, and it is widely accepted that other undiscovered genes contribute to this condition. Hence, our computer model will have potential relevance to new colorectal cancer genes as they are discovered, as well as to other multifactorial conditions.

Results of Pilot Studies

An initial prototype computer model of the cascade screening process as applied to MMR gene mutations has recently been developed as part of a CSO studentship conducted under the

supervision of the grant applicants. The methods outlined below were utilised and evaluated during prototype development, and a summary of the extent of the prototype can be found in appendix 1. The prototype has been designed to provide a core "engine" around which the rest of the model can be developed and implemented to form a complete and valid model of cascade genetic screening. It is restricted in scope to encompass only a few key elements of the process, and represents only a sub-section of a broader conceptual model. A graphical user interface has been created, enabling any user to alter various input parameters and view the model outputs in complete or summary form. The major limitations of the prototype model are that the identification of new mutation carriers is only considered through testing colorectal cancer cases, the family structures in the model are very basic, and the process of mutation tracing within a pedigree is over-simplified. These areas will constitute the focus of the next stage of development.

Aims:

(i) To create a detailed computer model of cascade genetic screening, as applicable to the challenge of identifying MMR gene mutation carriers in Scotland, using real data generated by our research group.
(ii) To utilise this model to evaluate the proportion of mutation carriers that could be identified over time under various cascade genetic screening protocols.
(iii) To evaluate the efficiency of cascade genetic screening.
(iv) To ensure that our model is broadly applicable in the context of identifying carriers of high penetrance mutations predisposing to complex disease.

Research Questions

(i) What proportion of mismatch repair gene carriers could be identified over time by applying a comprehensive cascade screening programme to the Scottish population?
(ii) How many non-carriers would undergo testing under such a programme?
(iii) What would be the cost of applying cascade screening to the Scottish population?
(iv) What effect does ascertainment criteria for cascade screening have on the above outcomes?
(v) What effect does penetrance, population prevalence and other variables have on the proportion of mutation carriers identified, and the efficiency with which this could be achieved?
(vi) What proportion of mutations in BRCA1 and BRCA2 could be detected using cascade screening, and how efficient would this process be?
(vii) What is the potential utility of cascade genetic screening in other complex diseases?

Plan of Investigation

The proposed project is an extension of ongoing pilot work. It is our intention to develop, test, and improve the model to the extent where it provides a thorough and accurate representation of the process of cascade screening.

The methods and software identified for development of the proposed model have been chosen on the basis of the expertise of the Heriot-Watt research team, who have extensive experience of computer modelling in the medical and biological domains. We have chosen to utilise object-orientated design, as this is a powerful and widely used technique for creating useful models of complex systems.

The first step in our overall strategy for developing a computer model using this technique is to develop a conceptual model. This is perhaps the most important stage as it precisely defines the requirements of the model, and it is normally completed by a 'domain expert' who has extensive knowledge of the system to be modelled. In this case, early versions of the conceptual model have already been produced as part of a CSO studentship, using input from the current applicants to determine the model's requirements. These inputs come from a variety of relevant

1.8.4 Mismatch Repair Gene Mutation Carriers

Much of the evidence in favour of screening individuals with HNPCC, and many of the issues considered above, are also applicable to mismatch repair gene mutation carriers. Indeed, the two definitions are often interchangeable in the current literature, with mismatch repair gene mutations being considered as the cause of the HNPCC syndrome. In reality, only a sub-set of empirically defined HNPCC patients will have MMR gene mutations, and only a proportion of MMR gene mutation carriers will meet the Amsterdam criteria for HNPCC. Carriers of mismatch repair genes are at an extremely high lifetime risk of developing colorectal cancer, estimated at 80% in males and 40% in females, and it follows that they should be offered intensive surveillance.

Jarvinen et al., have conducted a controlled screening trial on HNPCC families over a period of fifteen years, and have established the mutation status of the majority of study subjects (130, 131). This study has provided evidence in favour of screening MMR gene mutation carriers, finding a 55% (95% CI = 10%, 79%) reduction in risk of colorectal cancer in mutations carriers who underwent colonoscopic screening at three year intervals, compared to a control group of mutation carriers, and demonstrating that overall mortality was also significantly lower in the screening group (130). All participants in this study were offered screening, and any differences between those who accepted and those who did not comply in terms of lifestyle or illness behaviour will potentially lead to ascertainment bias. For obvious ethical reasons, this issue is also likely to affect future studies. Nonetheless, these findings constitute compelling evidence to support the intuitive view that MMR gene mutation carriers should be offered intensive screening for colorectal cancer due to their

perspectives, including clinical cancer genetics, genetic epidemiology and colorectal surgery. The conceptual model will be revised and overhauled as a first step towards developing a more comprehensive computer model of cascade genetic screening. This reflects an important aspect of the model development strategy, namely the testing and revision of the conceptual model in an iterative fashion. The Unified Modelling Language (UML), which is widely used and has previously been successfully utilised for modelling biological systems (13), will be employed at the conceptual stage. The conceptual model provides a framework for a functional model, which is usually developed by a computer programmer working under the guidance of the domain expert. In turn, the functional model forms the basis for software implementation using standard Java programming language. Again, model development will be undertaken in an iterative fashion.

Whilst the ultimate aim of this project is to create an adaptable model of cascade genetic screening that can be applied to any complex condition, model development will initially focus on cascade screening for mismatch repair gene mutations in the Scottish population. As discussed above, colorectal cancer is one area in which cascade screening is potentially of clinical benefit and the current applicants are in a unique position to evaluate cascade screening in this context.

A summary of key inputs to our computer model, and the available sources, is presented in appendix 2. Data on population demographics, and the incidence and mortality of colorectal cancer in Scotland are available in various publications by the Information and Statistics Division (ISD) of the Scottish Executive. Current guidelines the management of individuals at increased genetic risk due to family history will also be used, enabling a complete and accurate representation of the epidemiology of colorectal cancer in Scotland to be established as the background to our model. Information regarding the genetic epidemiology of mismatch repair gene mutations will come from the growing number of publications on this subject. An extensive literature review covering this issue has been conducted by the applicants (11), and has enabled the identification of working estimates. It is important to emphasize that the model will be deliberately constructed with the flexibility to alter such estimates, both to study the influence that such alterations could have on model outcomes, and to reflect future advances in our understanding of genetic epidemiology.

In addition to estimates obtained from the available literature, the proposed computer model will be informed by the existing expertise of the applicants and data from ongoing research into the genetic epidemiology of colorectal cancer and the application of genetic knowledge at the clinical level. We have access to valuable information from clinical genetics services, and have also amassed the largest collection of MMR gene mutation carriers in the UK, providing a unique data resource.

The Colorectal Cancer Genetic Susceptibility (COGS) study, a five-year (2003-2008) Cancer Research UK programme, is a large-scale study into the genetic basis of colorectal cancer. It is designed to conduct mutation analysis on all Scottish colorectal cancer patients who develop the disease under the age of 55 and to facilitate subsequent cascade genetic screening. This study will provide valuable and unique information regarding important aspects of the cascade screening process, including the acceptance rates of genetic testing among patients and various relative groups, the time-scale of genetic testing, and the laboratory and administrative costs involved. The COGS study has also generated detailed pedigree information relating to colorectal cancer cases and mutations carriers. This real data, and information on family structure relating to the general population will be utilised to simulate realistic pedigrees as part of the computer model. This will build on previous pedigree simulation work conducted by Professor MacDonald (14), and will ensure that the family structures used in the model mirror the real-life situation.

We intend to integrate the development of the computer model with the COGS study. The model will initially reflect cascade screening as it is currently being applied as part of this ongoing research. This will facilitate a high degree of external validation, since the predictions generated by the model can be compared with actual findings from the COGS study. Subsequently, input estimates will be manipulated within realistic boundaries to investigate the effects of altering various model parameters and determine the most favourable conditions for cascade screening in this context. The combination of domain expertise, 'real' data, and an iterative development strategy involving continual re-evaluation and revision of the model, will thus facilitate the expansion of the model to the extent where it constitute a full and realistic representation of cascade genetic screening for mismatch repair gene mutations.

Ethical approval for the COGS study has already been granted, and data used to inform inputs to the model will consist of summary statistics only. No subjects will be approached in connection with the current application.

The next step in model development will be to investigate the adaptability of the model by applying it to other complex conditions in which pathogenic mutations are known to cause a proportion of cases. The identification of mutations in the breast cancer genes BRCA1 and BRCA2 presents an ideal candidate for such an investigation, since the genetic epidemiology of this condition has been extensively studied. Application of the computer model to investigate the use of cascade screening in breast cancer will be conducted in a similar fashion to that outlined above. Inputs relating to breast cancer epidemiology, and the prevalence and penetrance of BRCA1/BRCA2 mutations etc. will be taken from the available literature. It is also intended to incorporate aspects of the "genISYS" project, funded by the CSO and undertaken by the Image Systems Engineering Laboratory (ISEL) at Heriot-Watt University with the aim of modelling breast cancer genetic risk analysis. Resulting model outcomes will be evaluated in the context of current knowledge about these genes.

Subsequently, it will be possible to apply the computer model to various other clinical situations, the only restriction being the availability of realistic input data. Once again, the model will be re-evaluated based on its response to different inputs, and will be adjusted accordingly if appropriate.

Appendix 3 lists some of the key outputs that will be generated by the model. The proportion of mutation carriers identified and the efficiency of this process will be crucial to the evaluation of cascade screening in any context, and this will form the focus of the model. Health benefits and associated risks resulting from surveillance of identified mutation carriers will vary according to particular condition in question, and the model will be designed to facilitate the inclusion of such information as part of future development. Our evaluation of cascade screening in the context of colorectal cancer will consider the benefits of colonoscopy in mismatch repair gene mutation carriers, using published estimates (9) supplemented by data from the COGS study. Similarly, any available information on the risks inherent in surveillance of mismatch repair gene carriers, including clinical complications and psychological distress will be incorporated into the evaluation.

The proposed model will include a component relating to administration and laboratory costs associated with the cascade genetic screening programme undergoing evaluation. A detailed economic evaluation relating to the costs and benefits of mutation carrier management in a specific context could be based on information generated by our model, or could be part of future development using the same methodology. However, a full economic evaluation is out-with the scope of the current project. The flexibility inherent in the methods we have chosen for model development also ensures that other features relevant to cascade genetic screening, such as gene-

gene and gene-environment interactions, could be built into the model if sufficient relevant data becomes available.

In accordance with recognised principles of good practice in model development, all assumptions and estimates built into the model will be carefully logged, and presented in as transparent a fashion as possible (15). Similarly, where the available data may not be accurate, or may not be sufficiently detailed, this uncertainty will be built into the model. Throughout the modelling process, particular attention will be paid to parameters that are potentially amenable to adjustment, such as acceptance rates and genetic test parameters.

The completed model will serve as a valuable tool for evaluating the feasibility and effectiveness of cascade screening in a variety of complex conditions. The model itself will be user-friendly, in the sense that no in-depth knowledge of computing will be required for operation. The model outcomes will, however, be sensitive to the input data provided. The availability and application of this data will thus be crucial to ensuring that the predictions made by the model are relevant to the corresponding real-life situation. Where the data parameters are unknown, or subject to wide confidence intervals, this uncertainty must be incorporated into the input parameters and considered carefully before drawing any conclusions. The model will thus be aimed at 'domain experts': individuals or groups with knowledge of the condition of interest and the ability to apply the model to that condition and interpret results accordingly.

Timetable of Work

Development of the prototype model has formed one aspect of the CSO studentship "Prevention of Colorectal Cancer in Scotland: Strategies for those at Increased Genetic Risk". The prototype will have been finalised and thoroughly tested by completion of this studentship on October 1st, 2003. The proposed project is intended to commence immediately, following on from work on the studentship, but not overlapping. The iterative process of model development will result in considerable overlap between the phases of the project outlined below.

Pre-Project Grant:	Funding from within existing resources in the PhD studentship has been secured for a four-month period to support the career development of Rory Mitchell. This time will be utilised to conduct a comprehensive review of the existing prototype conceptual model, and identify the requirements of the full version. Through consultation with all the applicants, and other individuals and groups with appropriate expertise, the required scope, limitations and outputs of the model will be carefully defined. During this period further training in computer modelling and Java programming will be undertaken.
0-3 Months:	An updated conceptual model will be represented in object-orientated form.
3-9 Months:	Using the conceptual model to provide the guidelines and framework, a functional computer model of cascade screening for MMR gene mutations will be developed.
9-18 Months:	During this latter phase, the computer model will be extensively tested and validated, with findings providing the feedback necessary to improve the model to the extent that it constitutes a complete and accurate representation of the process of cascade screening for mismatch repair gene mutations. Subsequently, the adaptability of the model, in terms of its applicability to other complex disease such as breast cancer, will be investigated. Findings at this stage will facilitate improvement of the model.

18-20 Months: In the concluding months of the project the user interface will be finalised, and findings will be disseminated in collaboration with the CSO.

Existing Facilities

The University of Edinburgh Colorectal Cancer Research Group is an expanding multidisciplinary team with considerable research experience in the field of colorectal cancer genetics. This group provides the ideal environment for the supervision of the proposed project. They have forged an ongoing functional collaboration with the co-applicants from the School of Mathematical and Computer Sciences at Heriot-Watt University. The latter researchers have the capacity to provide expert supervision of the computing aspects of the project, and to provide training in computer modelling where appropriate. Desk space, computing facilities and secretarial support are available at the MRC Human Genetics Unit in Edinburgh, where the research assistant to be employed on this grant is already based.

Justification of Requirements

Whilst all the current applicants will maintain an active interest in the project, it will be essential to employ a full time research assistant at grade AR1A to carry out the coordination of the project and to conduct the actual computer modelling work. Rory Mitchell is already established as a PhD student at the department of Community Health Sciences at the University of Edinburgh, and is currently completing a PhD thesis on "Prevention of Colorectal Cancer in Scotland: Strategies for those at Increased Genetic Risk" within the above research environment. He has been responsible for the pilot work for this project and is thus ideally placed to continue this line of research. Funding has also been requested for limited secretarial and computing support, totalling 10% of a full time salary for computing support (grade AD2), and 10% of a full time salary for secretarial support for the project (grade CN3). Such support will be necessary to ensure the efficient progress of this project. All staff will be employed on this grant for a period of 20 months.

Although the applicants themselves will undertake the majority of the modelling work, it will be necessary to engage the services of a computer modelling expert on an ad-hoc basis, to provide assistance and advice when required. The School of Mathematical and Computer Sciences at Heriot-Watt University has considerable experience in arranging such work, and the additional funding requested for this purpose is budgeted accordingly. Allowance has been made in the consumables section of the budget to purchase all software required for this project, and funding for miscellaneous office costs has also been included.

Research Outcomes and Implementation

Identification of people at high genetic risk of disease provides a target group who may benefit from clinical intervention, and cascade genetic screening provides a feasible option for accurately identifying such people. Detailed evaluation of the potential of this strategy, in terms of identifying mutation carriers and in terms of the associated costs and resources, is an essential prerequisite for introducing cascade genetic screening outside the research context. The proposed project will generate a detailed computer model of the cascade screening process and will include a thorough investigation of the process as applied to mismatch repair genes and colorectal cancer. In this context cascade genetic screening will be assessed in accordance with the National Screening Committee criteria for appraising the viability, effectiveness and appropriateness of screening programmes (16). The results of this investigation could directly inform NHS policy in Scotland. If the appropriate criteria are met cascade screening might be adopted to form an integral part of clinical genetics services for colorectal cancer, leading to health benefits for a small but significant proportion of the population.

In the wider context, our computer model will serve as a valuable tool for any researchers or policy-makers who are interested in the potential of cascade screening in various clinical

situations. The results and conclusions generated by the model could then be used to inform decisions regarding future research programmes and clinical treatment and screening policies. The completed model could be used to provide information on the likely impact of cascade screening for mutations predisposing to various complex diseases, and could be applied to any population. It is possible, therefore, that the model could be used for commercial exploitation.

Dissemination

The applicants will undertake to publish findings arising from this project in quality scientific journals, and to present at relevant conferences. We have recently co-hosted a meeting of the UK National Screening Committee on strategies for people at increased genetic risk of colorectal cancer, and we will report findings from this project directly to this committee.

Key References

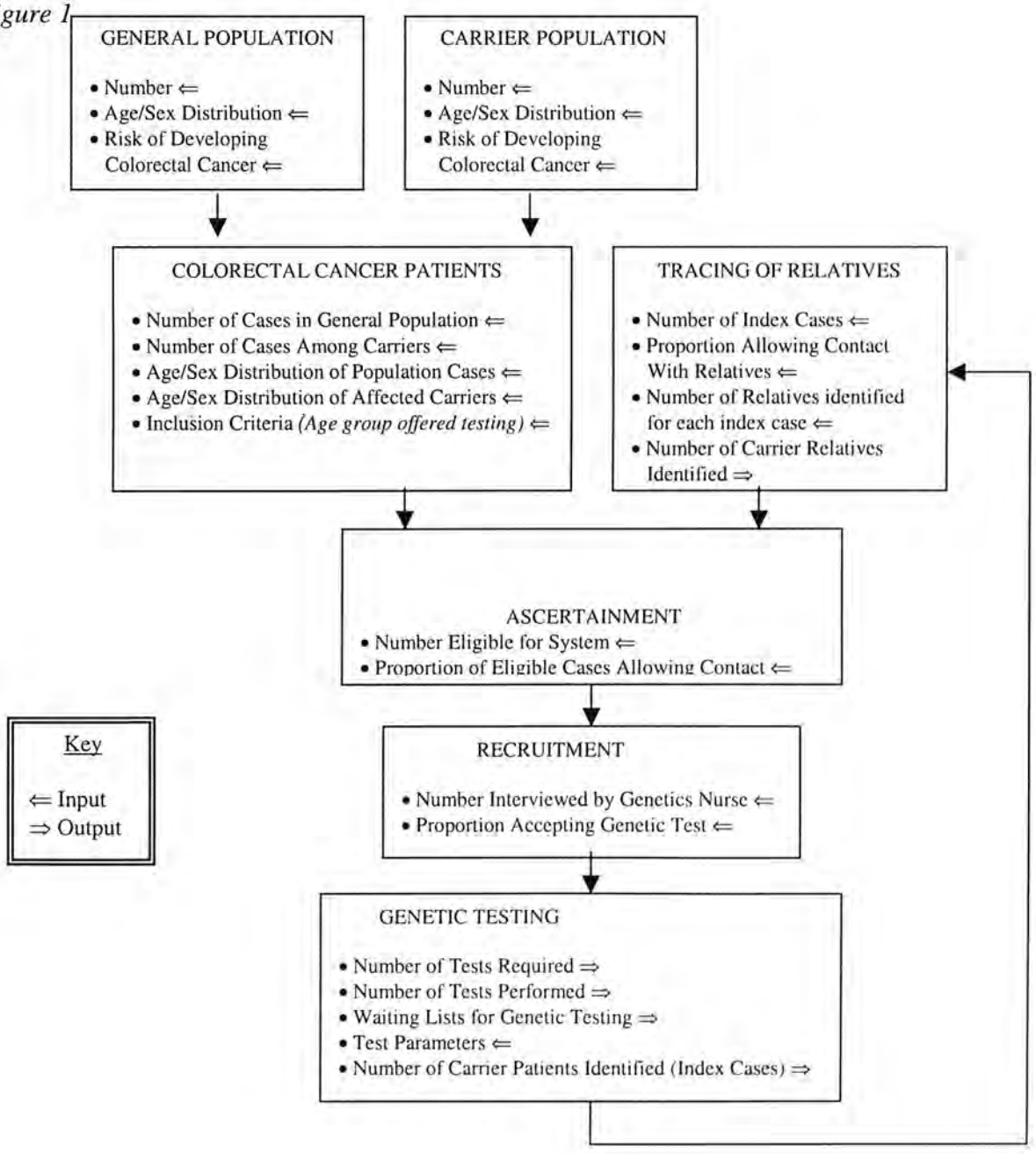
1. Holloway S, Brock DJ. Cascade testing for the identification of carriers of cystic fibrosis. J Med Screen 1994;1:159-164.
2. Super M, Schwarz MJ, Malone G, Roberts T, Haworth A, Dermody G. Active cascade testing for carriers of cystic fibrosis gene. BMJ 1994;308:1462-1467
3. Fouchier SW, Defesche JC, Umans-Eckenhausen MW, Kastelein JP. The molecular basis of familial hypercholesterolemia in the Netherlands. Hum Genet 2001;109:602-615.
4. Krawczak M, Cooper DN, Schmidtke J. Estimating the efficacy and efficiency of cascade genetic screening. Am J Hum Genet 2001:69:361-370
5. Dunlop MG, Farrington SM, Carothers AD, Wyllie AH, Sharp L, Burn J, Liu B, Kinzler KW, Vogelstein B. Cancer risk associated with germline DNA mismatch repair gene mutations. Hum Mol Genet 1997;6:105-110.
6. Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, Peltomaki P, Mecklin JP, Jarvinen HJ. Cancer risk in mutation carriers of DNA-mismatch-repair genes. Int J Cancer 1999;81:214-218.
7. Vasen HF, Wijnen JT, Menko FH, Kleibeuker JH, Taal BG, Griffioen G, Nagengast FM, Meijers-Heijboer EH, Bertario L, Varesco L, Bisgaard ML, Mohr J, Fodde R, Khan PM. Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis. Gastroenterology 1996;110:1020-1027.
8. Dunlop MG, Farrington SM, Nicholl I, Aaltonen L, Petersen G, Porteous M, Carothers A. Population carrier frequency of hMSH2 and hMLH1 mutations. Br J Cancer 2000;83:1643-1645.
9. Jarvinen HJ, Aarnio M, Mustonen H, Aktan-Collan K, Aaltonen LA, Peltomaki P, de la CA, Mecklin JP. Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer.[comment]. Gastroenterology 2000;118:829-834.
10. Rose P, Dunlop MG, Burton H, Haites N. Screening for late onset genetic disorders: Colorectal cancer. National Electronic Library for Health website (NHS) . 2002
11. Mitchell RJ, Farrington SM, Dunlop MG, Campbell H. Mismatch repair genes hMLH1 and hMSH2 and colorectal cancer: a HuGE review. Am J Epidemiol 2002;156:885-902.
12. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer-- analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78-85.
13. Lyddy P. Modelling 'The Central Dogma of Molecular Biology' using the Unified Modelling Language (UML). 2002. Heriot-Watt University.(Thesis)

14. Macdonald AS, Waters HR & Wekwete CTT. The genetics of breast and ovarian cancer I: A model of family history. Scandinavian Actuarial Journal 2003;1-27
15. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, Luce BR. Principles of good practice for decision analytic modeling in health- care evaluation: report of the ISPOR Task Force on Good Research Practices—Modeling Studies. Value Health 2003;6:9-17.
16. UK National Screening Committee. The criteria for appraising the viability, effectiveness and appropriateness of a screening programme. NSC website. 2002.
17. Lerman C, Hughes C, Trock BJ, Myers RE, Main D, Bonney A, Abbaszadegan MR, Harty AE, Franklin BA, Lynch JF, Lynch HT. Genetic testing in families with hereditary nonpolyposis colon cancer. JAMA 1999;281:1618-1622.
18. Farrington SM, Lin-Goerke J, Ling J, Wang Y, Burczak JD, Robbins DJ, Dunlop MG. Systematic analysis of hMSH2 and hMLH1 in young colon cancer patients and controls. American Journal of Human Genetics 1998;63:749-759.

Appendix1 **Scope of Prototype**

The figure below provides an overview of the scope of the prototype model of cascade genetic screening, listing some of the inputs required and the outputs generated by the system. The prototype model is designed to provide a narrow but detailed view of one aspect of cascade genetic screening as it could be applied to identifying mismatch repair gene carriers. The high-risk group concerned are colorectal cancer patients, and the model includes the capacity to limit genetic testing according to age at onset. Information provided by the model regarding the number of patients found to harbour pathogenic MMR gene mutation feeds back into the model, as indicated by the arrows, providing the starting point for tracing relatives and offering further genetic testing in a cascade fashion. The current prototype enables estimates to be generated regarding the likely outcomes of cascade screening in this context. However, it represents an oversimplification of the true situation. Initial expansion of the functionality of the model will include the use of real data will be used to inform estimates of test acceptance and to create a more realistic model of family structure and the process of tracing relatives. The scope of the model will also to be broadened in order to consider input from cancer genetics services.

Figure 1



Appendix 2 **Key Inputs to Computer Model**

Required Estimate	Source(s)
Size and age/sex distribution of Scottish Population	Government Statistics (ISD)
Population prevalence of MMR mutations	Literature (8), COGS and related analyses
Age/sex dependent penetrance of MMR gene mutations	Literature (5-7), COGS
Proportion of colorectal cancer cases meeting criteria for inclusion in cascade genetic testing programme for: (a) Non-carriers (b) Carriers	Estimates will vary according to criteria: (a) Government statistics (ISD), COGS, published data on MSI status (b) Literature, COGS
Number of relatives identified and traced per newly identified carrier*	Family structure simulation informed by government statistics, COGS
Proportion of cases meeting criteria, that are actually referred to cascade genetic screening programme	COGS
Acceptance rate of genetic testing among cancer cases*	Literature, COGS
Acceptance rate of genetic testing among relatives of known mutation carriers*	Literature (17), COGS
Proportion of individuals presenting to genetic services who would be offered genetic testing on the strength of their family history	Genetic Services Data
Acceptance rate of genetic testing among individuals presenting to genetic services*	Genetic Services Data
Sensitivity of non-specific mutation analyses	Literature (18), COGS
Specificity of non-specific mutation analyses	Literature (18), COGS
Sensitivity of specific mutation testing	Approximately 100%
Specificity of specific mutation testing	Approximately 100%
Traceability of relatives	COGS, previous research experience, Register General Office Scotland
Proportion of carriers accepting and complying with colonoscopic surveillance	COGS
Life-years saved as a result of surveillance programmes	Literature (9)
Cost of non-specific mutation analysis	COGS
Time required for non-specific mutation analysis	COGS
Cost of specific mutation test	COGS
Time required specific mutation test	COGS
Administration costs of genetic screening programme	COGS
Time required for recruitment	COGS

Abbreviations: COGS = Colorectal Cancer Genetic Susceptibility study, MSI = Microsatellite Instability, ISD = Information and Statistics Division

*Age /sex specific information will be incorporated where possible

Appendix 3 **Key Outputs of Computer Model**

STAGE	OUTPUTS
Ascertainment	Number of cases eligible for referral to cascade genetic screening programme
	Number of cases actually referred to cascade genetic screening programme
	Number of eligible cases accepting referral to cascade genetic screening programme
	Number of individuals presenting to cancer genetic services
	Number of individuals presenting to cancer genetic services who are eligible for referral to cascade genetic screening programme
	Number of individuals presenting to cancer genetic services who are actually referred to cascade genetic screening programme
Recruitment	Number accepting referral to cascade genetic screening programme
	Number accepting genetic counselling with genetic nurse
	Number accepting genetic testing and providing blood sample
Mutation Analysis	Non-specific Test results: Number of True Positives
	Non-specific Test results: Number of False Positives
	Non-specific Test results: Number of True Negatives
	Non-specific Test results: Number of False Negatives
Relative Recruitment	Total number of relatives of mutation carriers[#]
	Number of relatives of mutation carriers identified and traced[#]
	Acceptance rate of genetic testing among relatives[#]
Specific Testing	Specific Test results: Number of True Positives
	Specific Test results: Number of False Positives
	Specific Test results: Number of True Negatives
	Specific Test results: Number of False Negatives
Resource Allocation	Resources required for enrolment in cascade genetic screening programme
	Resources required for genetic counselling
	Resources required for non-specific genetic testing
	Resources required for specific genetic testing
Summary	Total number of mutation carriers identified*
	Total resources expended*
Follow-up	Number of mutation carriers accepting surveillance protocol*
	Predicted outcomes of surveillance*
	Predicted outcomes of non-surveillance*

*Information specific to age/sex distribution will be presented where appropriate

[#] Information specific to type of relationship with mutation carrier will be presented where appropriate

extremely high absolute risk of developing the disease. Further studies are, however, required to confirm the findings by Jarvinen et al., and to provide a more accurate assessment of the benefits of screening in mismatch repair gene carriers.

The accepted protocols for screening people who meet the Amsterdam criteria for HNPCC already constitute highly intensive surveillance, aimed at people with an extremely high genetic risk. It is unlikely, therefore, that MMR gene mutation carriers would benefit from yet more intensive screening. This reasoning is reflected in the current guidelines for colorectal cancer screening in Scotland (249), which endorse the same screening protocol in both circumstances. Hence, the current recommendation for screening in this group is biennial colonoscopy, commencing at around age 25 (34, 56, 249, 288, 310). As with empirically defined HNPCC, prophylactic colonectomy provides an alternative option for preventing colorectal cancer in MMR gene mutation carriers. Indeed, the more conclusive risk information provided by the detection of a pathogenic mismatch repair gene mutation may provide a more convincing rationale for prophylactic surgery in this group (41, 235).

1.8.5 Further Considerations

Acceptance rates and compliance with screening protocols are crucial considerations in the context of colorectal cancer screening targeted at people at increased genetic risk. Indeed, non-compliance may be the main limiting factor in terms of achieving successful outcomes (188). Numerous factors will influence whether or not an individual accepts screening. These are likely to include such general factors as age, sex and education (74, 188), but other key influences regarding this decision may be an individual's perception of their own

risk and the acceptability of the screening techniques being offered (302). Several studies have found that individuals with a family history of colorectal cancer are more likely to accept screening than those without such a history (188, 263, 302). This is consistent with the general observation that screening programmes aimed at people with a family history, particularly those diagnosed with HNPCC, tend to have a higher uptake than programmes aimed at the general population. Similarly, it has been noted that people who have declined screening on the basis that they have empirically defined HNPCC, may undergo screening after they are positively identified as MMR gene mutation carriers (130). Recommendation of screening by a health professional is also associated with a greater uptake (74). Many of the factors that influence acceptance and compliance are amenable to adjustment, primarily through patient education and raised public awareness.

In addition to offering screening, people at increased genetic risk may benefit from addressing their risk of developing colorectal cancer through primary prevention. Primary data to support this hypothesis are scarce, and the extent to which the dietary and environmental factors known to influence sporadic colorectal cancer also apply in the sub-group of the population at increased genetic risk remains poorly understood. Some evidence exists to suggest that the risk factors that affect the general population also apply in the subgroup of people with a family history (65), and that familial risk can be modified by dietary and lifestyle factors (66, 80, 159, 191, 258). Such findings are consistent with the observation that the spectrum of cancers in HNPCC families has altered during the last century, with the incidence of colorectal cancer, relative to other malignancies, increasing both in these families and in the general population. In both instances the observed trends

are presumably due to the influence of dietary and lifestyle changes. However, other studies have found no relationship between colorectal cancer risk and established risk factors among people with a family history of colorectal cancer (154, 206). Further investigations are required to determine the most appropriate methods of primary colorectal cancer prevention in those with a family history of the disease. Similarly, specific evidence regarding the influence of environmental factors on the risk of colorectal cancer in MMR gene mutation carriers is scarce. Environmental risk factors that may interact with mismatch repair genes, and their possible influence on penetrance of mismatch repair gene mutations, are considered further in chapter 3.

Despite this lack of clarity, though, it seems reasonable to recommend that individuals at increased genetic risk consider dietary and lifestyle changes, such as reducing fat and alcohol intake and taking exercise, with the aim of reducing their colorectal cancer risk. It is also probable that people who are aware that they have a relatively high genetic risk may be more highly motivated to effect such changes.

Although the specific effectiveness of chemoprevention in people with a family history of colorectal cancer has yet to be ascertained, preliminary evidence suggests that chemoprevention does lower risk in this sub-group (42, 240). Consequently, chemoprevention may prove to be a useful approach to colorectal cancer prevention in people at increased genetic risk (42, 103, 233). Again, evidence for a specific interaction between NSAIDs and mismatch repair genes is considered in chapter 3.

1.8.6 Summary

In summary, screening for colorectal cancer has the potential to offer significant benefits to individuals at increased genetic risk. Ideally, the extent of family history, and the associated risk, should be evaluated on an individual basis. For pragmatic reasons, however, it is necessary to provide broad guidelines for sub-groups meeting particular criteria, and several papers have attempted to formalise what appears to be a growing consensus regarding screening of people with a family history of colorectal cancer. The precise protocols offered will depend primarily on the absolute risk of the individual concerned, and will be influenced by factors such as compliance, expense and the potential risks associated with screening. Ideally, a specifically-tailored screening programme should be available to any individual who is at significant absolute risk of colorectal cancer. At present, however, considerably more research is required to inform the establishment of appropriate screening protocols.

1.9 Strategies For Identifying People At Increased Genetic Risk

Two principal strategies for identifying individuals at increased genetic risk of colorectal cancer are currently available, namely family history assessment and genetic testing for mismatch repair gene mutations.

1.9.1 Family History

The standard means of identifying people at increased genetic risk of colorectal cancer in Scotland, as elsewhere, has been through referrals from primary care to specialized genetic services, on the basis of family history. This relies largely on the individuals in question being aware of that they have a family history, and being sufficiently concerned in this regard to seek advice from their general practitioner. The general practitioner will then discuss this issue, and, if they deem it appropriate, may refer the patient to genetic services for further evaluation of their risk and genetic counseling.

Although the above situation does largely address the needs of people who are concerned about their family history of cancer, it is likely that a large proportion of people with a family history will not be brought to the attention of genetic services. This has been illustrated in several studies in which members of the general population, who have not previously sought or been offered advice about genetic risk, have reported having a family history in a research context (112, 135, 198, 243, 293). To some extent, this problem can be addressed by increasing public awareness of the genetic nature of colorectal cancer risk and the availability of strategies that may help people at increased susceptibility. Likewise,

people identified as having a family history of colorectal cancer, and people who have developed the disease themselves, can be advised to inform their close relatives so that they in turn can seek counselling and genetic risk assessment. Encouraging GP's to enquire about family history and refer patients where appropriate may also help increase the proportion of the high-risk sub-group who are offered the appropriate management. Overall, however, such interventions are likely to have a limited impact, and may produce the adverse effect of unnecessarily increasing anxiety amongst the public, which in turn could lead to over-stretching of resources.

The principal alternative for identifying people with a family history of colorectal cancer is though distribution of questionnaires. This approach has been explored in several studies (112, 162, 243, 293) and shown to be feasible in the research context. Response rates in these studies have been highly variable, ranging from 29% (162) to 84.7% (112). This disparity is likely to be due to differences in the design of questionnaires, the means of delivery (i.e. whether the questionnaires are mailed directly to potential participants or distributed through a general practice), and the populations under study. Nevertheless, these studies have underlined the main limiting factor of questionnaire studies, namely that only a proportion of the population will participate. A further option for identifying people with a family history is for a nurse to ask people attending health checks for details of their family history. Again, this approach has been shown to be feasible (135), but also has limitations, being labour-intensive and restricted to people attending health checks.

As a tool for identifying people at increased genetic risk, family history has the considerable advantages of being relatively easy to assess, and providing a broad measurement of genetic susceptibility that is not conditional on identification or understanding of the underlying molecular defects. However, family history is a rather crude instrument, and lacks sensitivity and specificity in terms of accurately predicting the absolute risk to an individual. Colorectal cancer is a common disease, and many people may have an affected relative, or even a strong family history, purely by chance. Conversely, a high degree of genetic susceptibility may not be reflected in family history, particularly when family size is small. Hence, offering screening on the basis of family history alone may lead to some individuals undergoing screening unnecessarily while others who are at a sufficiently high risk to justify such intervention are not included in screening programmes. A balance must therefore be struck between the sensitivity and specificity of family history, and must take into account the available resources and implications for the individuals concerned.

In addition to the limitations of family history as a means of identifying people at increased genetic risk, there are various practical issues involved with accurately establishing family history, which may impact on genetic risk assessment. In the clinical setting, family history is usually obtained directly from the individual concerned at interview with a geneticist, and in the research context family history is generally established by interview or questionnaire. Whilst interview is generally recognized as the preferable approach, it is still subject to underreporting and inaccuracies on the part of the interviewee (91, 140). This issue is crucial in the context of genetic risk assessment, and explored further in chapter 3.

1.9.2 Genetic Testing

Genetic testing has the potential to provide a more precise method of determining genetic risk, as it permits the analysis of molecular genetic risk factors at the individual level. Genetic testing may be used for further characterization of genetic susceptibility in people already known to be at increased familial risk of developing a particular condition, and also has the potential to identify people at increased genetic risk, irrespective of their family history. The recent identification of mismatch repair gene mutations as the causal defect in a small but significant sub-set of colorectal cancer cases therefore provides an opportunity for improving risk assessment and screening programme for people at increased risk.

From the clinical perspective, the ability to discriminate between people at different risks within families is a major advantage of genetic testing compared to empirical risk assessment based on family history. In autosomal dominant syndromes, a parent carrying the mutation in question will have only a 50% chance of passing the affected allele to each offspring. Thus, whilst an empirical diagnosis based on family history will classify all individuals within that family as at increased risk, the kindred will actually comprise a mixture of mutation carriers, who are at a greatly elevated risk, and non-carriers, with a risk no different from the general population. Once the causative mutation within a family is identified, genetic testing can be offered to all family members. Mutation carriers can then be offered the appropriate surveillance, whereas non-carriers can be reassured about their risk and removed from any family history-based screening they may have been undergoing. Genetic testing therefore theoretically facilitates a reduction in the total number of family members eligible for screening, as illustrated in studies demonstrating a reduction in the

overall screening effort applied to high-risk families following the introduction of genetic testing for mismatch repair gene mutations (225, 261). Although further research is required to inform economic evaluation of genetic testing for colorectal cancer susceptibility, it is possible that this approach may prove relatively cost-effective (31, 221).

Whilst genetic testing clearly has considerable potential as a tool for identifying people at increased genetic risk of colorectal cancer, it also has various disadvantages and limitations, associated with the technical difficulties inherent in both detecting and interpreting gene variants.

Mismatch repair gene mutations are detected in only a proportion of people who meet the Amsterdam criteria for HNPCC diagnosis, and are found in even fewer people who have a lesser degree of family history. Individuals in whom no mismatch repair gene mutation is detected may harbour pathogenic mutations in other genes, or indeed may have mismatch repair gene mutations that were 'missed' for technical reasons. Accordingly, apparently negative mutation status cannot be equated with lack of increased genetic risk, and individuals who meet empirical criteria for screening should maintain screening despite a negative genetic test result.

The pathogenicity of different mismatch repair gene variants remains largely unknown, and may vary considerably. Additionally, other genetic and environmental factors may influence penetrance, and at present data on these factors are extremely scarce.

Consequently, information regarding actual absolute risk associated with a particular mutation may be inaccurate, and may lead to an inappropriate level of anxiety and/or screening.

The psychosocial implications of genetic testing constitute crucial considerations in the context of applying genetic knowledge clinically. At the individual level, the identification of a pathogenic mismatch repair gene mutation may cause anxiety or depression regarding the risk of developing colorectal cancer, and may create additional concerns regarding insurance and the reaction of other family members (40, 165, 184, 227). The fact that a positive genetic test has implications for a whole family as well as the individual undergoing testing raises a number of complex issues. The perceptions of other family members, and the impact a positive test may have on them have been cited as a concern by people undergoing testing (227). Mutation carriers are likely to experience anxiety relating to the possibility that they may have passed the mutation on to their offspring, and may have associated feelings of guilt if this is found to be the case (165). Related issues of consent and confidentiality also arise from the familial nature of genetic testing. Family members may feel obliged to undergo testing for the benefit of their relatives, and it is possible that individuals who do not wish to undergo testing may become aware of their probable mutation status through results in other family members (40).

The uncertainty surrounding the results of genetic testing for mismatch repair gene mutations is another important issue that can potentially affect attitudes towards screening.

One concern is that people with a family history of colorectal cancer who have tested negative may be less inclined to comply with screening programmes. A qualitative study by Lerman et al., in which first-degree relatives of colorectal cancer patients were interviewed to ascertain their attitudes towards genetic testing, found that approximately half the participants anticipated that they would be less inclined to adhere to screening programmes or primary prevention strategies if they were to test negative (165). Conversely, many participants in this study felt that a negative test would not remove their anxiety surrounding their genetic risk (165).

Considerable uncertainty regarding the psychosocial impact of genetic testing for mismatch repair gene mutations remains, and several studies have indicated that understanding of genetics, genetic testing, and the concept of genetic risk amongst the general population is limited (46, 227, 274). Despite this, the potential benefits of genetic testing in this context are such that this approach is commonly used both clinically and for research purposes. Hence, there is a clear consensus that advocates the use of genetic counselling before and after genetic testing, in order to provide education and support to the individuals concerned (40, 165, 184, 220).

As discussed elsewhere, conducting mutation analysis of mismatch repair genes is also a lengthy and expensive process. Resulting practical and economic considerations, coupled with the scientific and psychosocial issues outlined above, have meant that mutation analysis is predominantly restricted to people deemed to be at high risk of having a

mismatch repair gene mutation. Hence, a major challenge in this context is that of devising strategies for efficiently targeting resources towards the identification of kindreds with such mutations. There are several options for targeting of resources in this manner. Firstly, the extent of family history may be used to identify people most likely to harbour a mutation. Similarly, the relatively early age of colorectal cancer onset evident in mismatch repair gene carriers means that young cases are relatively likely to carry a mutation. Additionally, the intermediate phenotype of microsatellite instability (MSI), which is considered further in the next chapter, is associated with mismatch repair gene deficiency, and this can provide another means of targeting resources. Each of the above approaches has been used in the context of epidemiological studies to identify mutation carriers, and to some extent have been applied in the clinical setting. Currently, most mutation carriers are identified by referral of patients with a family history of colorectal cancer to genetic services. Another option, under investigation in an ongoing program in Scotland, is to search for mismatch repair gene mutations among persons with early-onset colorectal cancer.

An overview of findings from one research group (174) provides a consideration of how best to utilize the above criteria to target mutation analysis to individuals at high risk of carrying a mutation, concluding that MSI is a useful predictor of mutation status in individuals meeting the Amsterdam criteria for HNPCC. Estimates of sensitivity and specificity of various family history criteria, as means of identifying mutation carriers, are also presented by Syngal et al., (267). At the present time, however, there is no consensus regarding the most efficient approach to identifying mutation carriers.

Applying mutation analysis to people deemed to be at high risk of carrying a mismatch repair gene mutation has so far been useful in two ways. Firstly, mutation analysis in colorectal cancer patients has enhanced our understanding of the genetic epidemiology of this disease. In addition, mutation analysis has permitted the characterization of the genetic defects responsible for the high degree of familial risk seen in a proportion of families with multiple colorectal cancer cases, and thus enabled screening procedures to be refined. However, from a public health perspective, the identification of all mutation carriers in a population at an asymptomatic stage is the ultimate goal of genetic testing.

There are several options for attempting to achieve such a goal, including population genetic testing, in which mutation analysis is conducted in every member of a population, and stratified population genetic testing, in which genetic testing is offered only to individuals meeting specified family history criteria. An alternative strategy to identifying asymptomatic mismatch repair gene mutation carriers in a population is cascade genetic testing, which involves performing mutation analysis in colorectal cancer patients or individuals otherwise deemed to be at high risk of harboring mutations, and subsequently tracing an identified mutation through a pedigree. The advantages and disadvantages of these strategies are considered in chapter 4, which is concerned with the evaluation of cascade genetic testing.

1.9.3 Summary

Family history is a useful tool for identifying people at increased genetic risk, but lacks sensitivity and specificity. Although it is frequently employed, various aspects of the association between family history and increased genetic risk remain to be elucidated. Gaps in the current knowledge include data relating to the precise absolute risks associated with various degrees of family history, the prevalence of family history in the general population, and the accuracy with which family history is reported and recorded. Genetic testing for mismatch repair gene mutations has recently provided another option for determining which individuals are at increased genetic risk, and an opportunity to improve the accuracy of genetic risk assessment. The general consensus is that both these strategies are relevant, and should be viewed as complementary. However, although understanding of the genetic epidemiology of colorectal cancer is advancing rapidly, the required data for informing strategies for people at increased genetic risk is incomplete. Consequently, considerable debate surrounds the appropriate strategies for people at increased genetic risk, and further evaluation of the various options is required.

1.10 Aims

The overarching aim of this thesis is to contribute to available data relevant to prevention of colorectal cancer in Scotland through targeting people at increased genetic risk. To this end, there are three core research components addressing gaps in current knowledge, organised around the central theme of evaluating available strategies for identifying people at increased genetic risk. Each component has specific research aims as follows:

(i) The role of mismatch repair gene mutations in colorectal cancer

To undertake a systematic literature review of the aetiological role of mutations in DNA mismatch repair genes in colorectal cancer. This comprises a collation of current knowledge of the role of the DNA mismatch repair genes, hMLH1 and hMSH2, and their genetic variation and a critical appraisal of published evidence of associations between mutations in these genes and colorectal cancer.

(ii) Family history as a tool for identifying people at increased genetic risk

To evaluate the use and limitations of family history assessment through analysis of a unique data set, comprising family history information reported by a colorectal cancer case or control subject at interview and information obtained through record linkage with the Scottish Cancer Registry. To estimate prevalence of family history of colorectal cancer in a population sample, to assess risk associated with a family history, and to investigate the accuracy with which family history of colorectal cancer is reported at interview.

(iii) Cascade genetic testing for DNA mismatch repair gene mutations

To develop a computer model of cascade genetic testing to assess the potential utility of this strategy as a means of identifying asymptomatic DNA mismatch repair gene mutation carriers in the Scottish Population, and to evaluate this strategy in the context of colorectal cancer prevention.

Chapter 2

Mismatch Repair Genes hMLH1 and hMSH2 and Colorectal Cancer: A Systematic Literature Review

A systematic literature review of the mismatch repair genes hMLH1 and hMSH2 and their role in colorectal cancer was conducted and presented in accordance with the guidelines presented by the Human Genome Epidemiology Network (HuGE Net). The completed review has been published in the American Journal of Epidemiology (199), and is also available on the HuGE Net Website (114). This chapter has been adapted from the published review, which is included as an appendix for reference (see appendix A2).

2.1 Abstract

Evidence to support a role for the mismatch repair genes human mutL homolog 1 (hMLH1) and human mutS homolog 2 (hMSH2) in the etiology of colorectal cancer has come from linkage analysis, segregation studies, and molecular biologic analysis. More recently, carriers of potentially pathogenic mutations in the hMLH1/hMSH2 genes have consistently been shown to be at a greatly increased risk of developing colorectal cancer compared with the general population. When considered together, the available evidence shows a strong, consistent, and biologically plausible association between mismatch repair gene mutations and colorectal cancer. The penetrance of mutations in hMLH1/hMSH2 is incomplete and is significantly higher in males (approximately 80%) than in females (approximately 40%). To date, evidence for gene-gene or gene-environment interactions is limited, although preliminary studies have revealed a number of avenues that merit exploration. Population screening for mutation carriers is not currently a feasible option, and mutation analysis remains restricted to either relatives of mutation carriers or colorectal cancer cases selected on the basis of phenotype.

2.2 Introduction

The mismatch repair genes human mutL homolog 1 (hMLH1) and human mutS homolog 2 (hMSH2) are integral components of the DNA mismatch repair pathway. So far, over 200 allelic variants have been identified for each gene, and the majority of these have been reported to be pathogenic in terms of colorectal cancer. The primary objectives of this review are to describe what is known about hMLH1 and hMSH2 and their variants in different populations and to examine the evidence implicating these genes as risk factors in the development of colorectal cancer.

2.3 Methods

2.3.1 Search strategy

The MEDLINE (National Library of Medicine), EMBASE (Excerpta Medica), and CANCERLIT (National Cancer Institute) databases were searched for papers published before December 31, 2001, using the keywords hMSH2 and hMLH1. Relevant papers were identified, critically appraised, and entered into a Reference Manager (ISI ResearchSoft, Berkeley, California) database. In addition, PubMed was searched via Reference Manager, by author name, for papers from research groups that had published several times on this subject. Finally, the database thus created was cross-referenced with papers cited in the International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer database of mutations (127).

For the "Gene Variants" section, a total of 109 papers were considered, which were identified by the above strategy and fulfilled the following selection criteria: 1) complete mutation analysis had been performed on more than five patients with colorectal cancer and 2) there was sufficient detail on the molecular nature of the genetic alteration. A database presenting details of all gene variants described in these published papers is shown in appendix A3, and posted on the website of the Human Genome Epidemiology Network (114).

For the "Associations" section, the above strategy led to the identification of eight studies that had conducted an analysis of the risk of developing colorectal cancer among carriers of mismatch repair gene mutations. These studies are summarized in table 2.2. A total of 77 papers that included results of complete mutation analysis performed on more than five colorectal cancer patients selected on the basis of family history, microsatellite instability (MSI), or age of onset, were also identified. A summary of these studies is presented in appendix A4, and on the HuGE Net website (114). Many papers included information relevant to both the gene variants and associations sections.

2.3.2 Classification of Gene Variants

For the purposes of this review, gene variants were classified into one of four categories. These categories are loosely based on the definitions given below, modified according to clinical observations.

1. Pathogenic mutation—generally frameshifts, nonsense mutations, and splice variants

2. Probable pathogenic mutation—generally nonconservative amino acid changes

3. Probable polymorphism—generally conservative changes, often observed in controls

4. Definite polymorphism—synonymous variants

2.4 Genes

2.4.1 hMSH2

The human mutS homolog 2 (hMSH2) gene is located at chromosome 2p21, an area initially identified as an important candidate region for genes involved in hereditary nonpolyposis colorectal cancer (HNPCC) syndrome by genetic linkage analysis within large affected families (161, 218).

2.4.2 hMLH1

The human mutL homolog 1 (hMLH1) gene is located at chromosome 3p21–23, an area also identified by genetic linkage analysis as an important candidate region within large HNPCC families that are not connected with the chromosome region 2p21–22 (171, 214).

2.5 Gene Variants

One conclusion generated by early attempts to identify precise genetic alterations in hMLH1 and hMSH2 was that variants in these genes are extremely heterogeneous. All 16 exons of the hMSH2 gene and 19 exons of the hMLH1 gene have been found to contain pathogenic mutations.

At present, there are no standard criteria for classifying variants as pathogenic mutations or polymorphisms, and consequently there is considerable variation in interpretation by different researchers. In general, categorization of alterations is based on the predicted effect on protein, with segregation of the mutation with colorectal cancer in the kindred in question and/or analysis of control subjects for that specific mutation also being considered when possible. However, the functional consequences of many mutations are difficult to predict accurately. It has been suggested that even alterations that do not affect the amino acid sequence could lead to aberrant splicing, and that the position of the mutation may be more significant than the type (211). In vitro functional assays have been developed and applied to the task of determining the pathogenicity of missense mutations (59, 213, 277) and may eventually facilitate accurate classification of such changes.

Appendix A3 lists all of the gene variants identified as part of this review, illustrating the extreme range of mutations identified and the fact that the observed spectrum of mutation is not entirely uniform. Some features of this table are summarized below. Figures 2.1 and 2.2 illustrate the distributions of unique gene variants that have been fully characterized at the molecular level in hMLH1 and hMSH2, respectively, according to their position on the gene. Figures 2.3 and 2.4 are designed to show the actual numbers of families in which pathogenic mutations have been identified.

Figure 2.1 Distribution of Unique Gene Variants in hMLH1

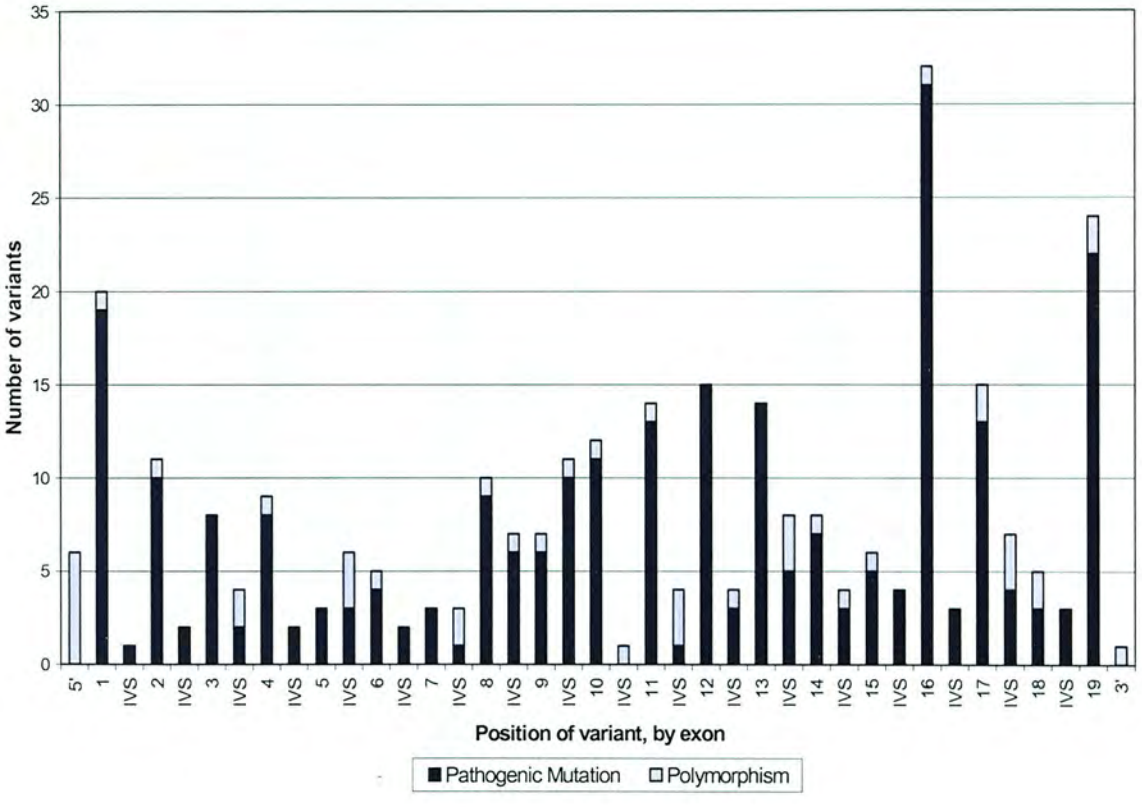


Figure 2.1 illustrates the distribution of all unique gene variants that have been identified and fully characterized in mutation analysis studies of colorectal cancer patients. Variants designated as categories 1, 1/2, 2, and 2/3 in appendix A3 are considered to be pathogenic for the purpose of this summary figure, and all other variants are described as polymorphisms. Exon deletions in which the underlying molecular variant was not known were excluded. Abbreviation: IVS, intervening sequence.

Figure 2.2 Distribution of Unique Gene Variants in hMSH2

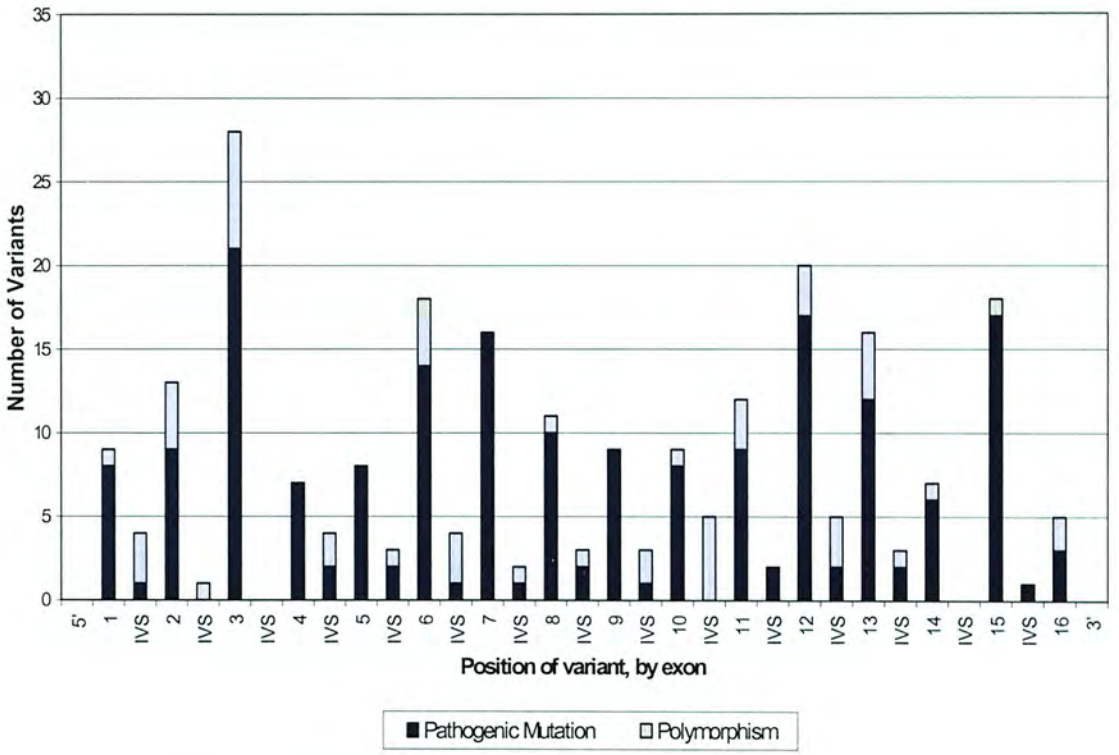


Figure 2.2 illustrates the distribution of all unique gene variants that have been identified and fully characterized in mutation analysis studies of colorectal cancer patients. Variants designated as categories 1, 1/2, 2, and 2/3 in Appendix A3 are considered to be pathogenic for the purpose of this summary figure, and all other variants are described as polymorphisms. Exon deletions in which the underlying molecular variant was not known were excluded. Abbreviation: IVS, intervening sequence.

Figure 2.3 Distribution of Pathogenic Mutations in hMLH1 by Number of Families Identified

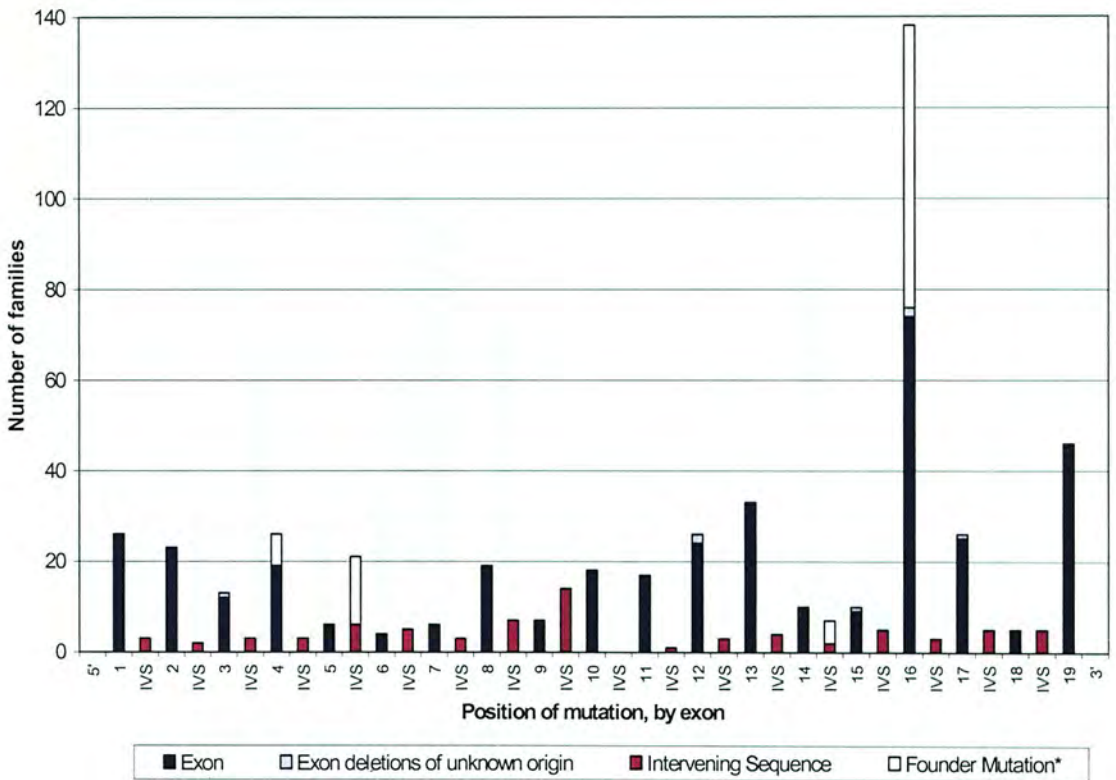


Figure 2.3 illustrates the distribution of pathogenic mutations according to the actual number of families in which a pathogenic mutation has been identified. These figures include all pathogenic mutations as defined in figure 2.1, plus exon deletions of unspecified origin. Deletions of more than one exon were excluded. *Families are deemed to have a "founder mutation" if they have a mutation which has been shown to have a founder effect in the same population. Abbreviation: IVS, intervening sequence.

Figure 2.4 Distribution of Pathogenic Mutations in *hMSH2* by Number of Families Identified

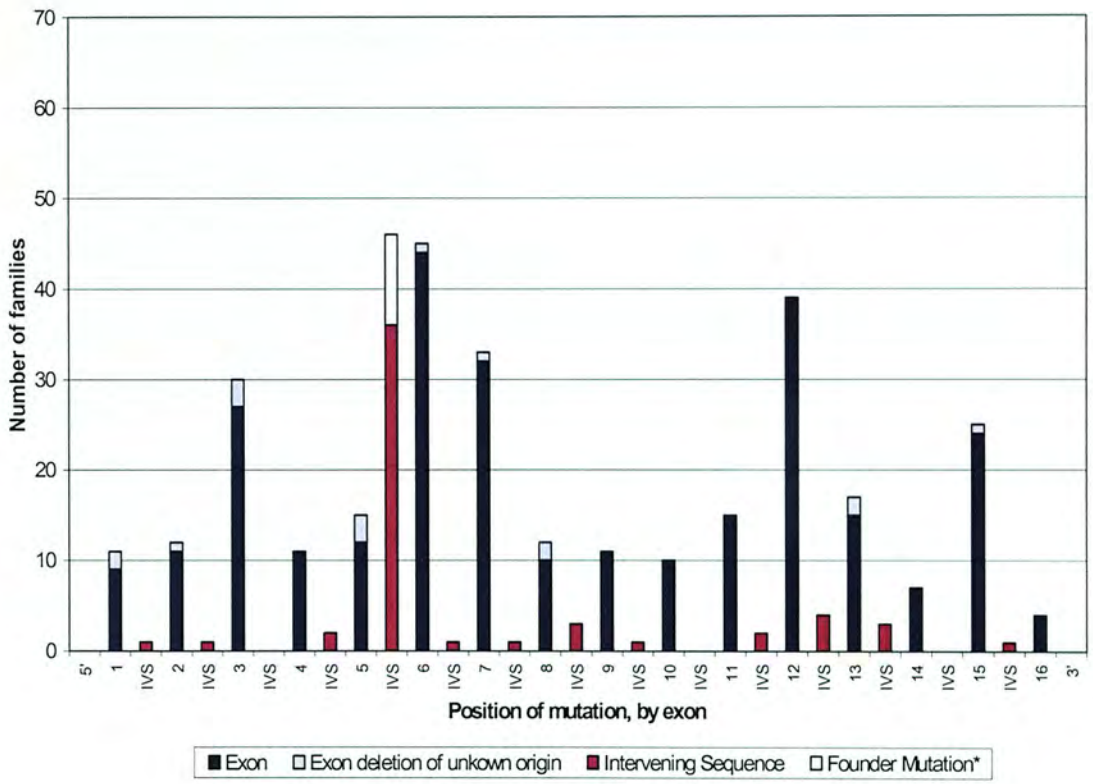


Figure 2.4 illustrates the distribution of pathogenic mutations according to the actual number of families in which a pathogenic mutation has been identified. These figures include all pathogenic mutations as defined in figure 2.2, plus exon deletions of unspecified origin. Deletions of more than one exon were excluded. *Families are deemed to have a "founder mutation" if they have a mutation which has been shown to have a founder effect in the same population. Abbreviation: IVS, Intervening Sequence

In total, 259 different pathogenic mutations, as defined above, have been identified in hMLH1, along with 45 polymorphisms. In hMSH2, 191 different pathogenic mutations and 55 polymorphisms have been characterized so far. This high degree of heterogeneity is similar to that found in the breast cancer genes BRCA1 and BRCA2, in each of which over 400 gene variants have been reported. When considering the range and type of gene variants listed in appendix A3, there are several important sources of bias that merit consideration. Firstly, a significant publication bias is likely to exist in favor of apparently pathogenic alterations. Highly penetrant mutations are also likely to be over-represented, since many studies involved conducting mutation analysis in patients selected on the basis of a strong family history of colorectal cancer. Secondly, genomic deletions in mismatch repair genes appear to occur relatively commonly, particularly in hMSH2, and such variants are not detected by many of the techniques commonly used for mutation analysis (304).

It is evident from the above figures that certain specific mutations have been identified in more than one kindred. Indeed, some mutations are found with a relatively high frequency. The most commonly observed mutations are summarized in table 2.1, which displays all mutations identified in more than four ostensibly independent kindreds.

Table 2.1 Commonly Observed Pathogenic Mutations

Exon	Nucleotide change	DNA change	No. of kindreds	Founder effect?	No. from founder population
		*HMLH1**			
Exon 16		3.5-kilobase deletion	63	Finland	62
Exon 16	Deletion A-A-G at nucleotide 1846	In-frame deletion in lysine codon 616	21		
IVS 5	g-a at nucleotide 4541	Out-of-frame deletion in exon 6 codon 152182	18	Finland	15
Exon 16	A-A-G-C at nucleotide 1852	Lys618Ala	15		
Exon 4	C-T at nucleotide 350	Thr117Met	12		
Exon 19	G-A at nucleotide 2146	Val716Met	12		
Exon 13	Insertion C at nucleotide 1490	Frameshift from codon 497	10		
Exon 4	T-G at nucleotide 320	Ile107Arg	7	Finland	7
Exon 13	C-T at nucleotide 1459	Arg487STOP	7		
Exon 17	C-T at nucleotide 1975	Arg659STOP	7		
Exon 19	G-A at nucleotide 2141	Trp714STOP	6		
Exon 8	C-T at nucleotide 676	Arg226STOP	6		
Exon 2	G-A at nucleotide 199	Gly67Arg	5		
Exon 2	C-T at nucleotide 184	Gln62STOP	5		
IVS 14	4-base-pair insertion/3-base-pair deletion at nucleotide 1667+2	Silencing of allele	5	Denmark	4
		*hMSH2**			
IVS 5	A-T at nucleotide 942+3	In-frame deletion in exon 5	46	Newfoundland	10
Exon 6	G-A at nucleotide 965	Gly322Asp	32		
Exon 12	Deletion A-A-T at nucleotide 1786	In-frame deletion in asparagine codon 596	11		
Exon 7	C-T at nucleotide 1216	Arg406STOP	6		

Abbreviations: hMLH1, human mutL homolog 1; IVS, intervening sequence; hMSH2, human mutS homolog 2; STOP, stop codon. Standard notation for nucleotides and amino acids is employed.

The observed spectrum of gene variants may be largely due to genuine differences in the mutability of specific nucleotides or sequences within the gene, but in some cases variants identified in apparently unrelated kindreds can be traced to a common ancestor. Such "founder effects" have been identified in the Finnish population, where two specific founder mutations in hMLH1 account for the vast majority of families in which mismatch repair gene mutations have been identified (215, 241). Another hMLH1 founder effect is evident in the Danish population (129). The extent to which founder effects are responsible for other frequently detected alterations is not entirely clear from the data currently available, and it is likely that some of the kindreds included in table 2.1 share a common ancestor. Interestingly, the intervening sequence 5 variant A-T at nucleotide 942+3 has been shown to occur as a founder mutation in Newfoundland (76), but another study found no evidence for a common haplotype in 10 carriers of this variant, of various origins, and concluded that the mutation also arises frequently de novo (51). This example underlines the notion that observations of mutation frequency are the result of both the probability of a mutation at a given nucleotide and the demographic history of the population in question.

Overall, little ethnic or population variation is apparent from the available gene variant data. However, the current biases towards highly penetrant mutations are such that the effect of the identified mutation is likely to transcend any population differences. Clearly, there is a need for accurate and extensive population-based data to elucidate any population differences in the spectrum and frequency of mismatch repair gene variants.

There is no clear evidence to suggest that any specific mismatch repair gene mutation produces a specific phenotype of colorectal cancer, although it has been suggested that some differences exist between the spectrum of extracolonic cancers associated with hMSH2 mutations in comparison with hMLH1 mutations (60, 248).

2.6 Associations

Evidence implying and supporting a causal role for hMLH1/hMSH2 in colorectal cancer comes from both epidemiologic studies and laboratory-based molecular studies. Initially, linkage studies revealed that disease expression in a proportion of HNPCC kindreds was linked to either chromosome 2p21 (161, 212, 218) or chromosome 3p21–23 (77, 171, 212, 214).

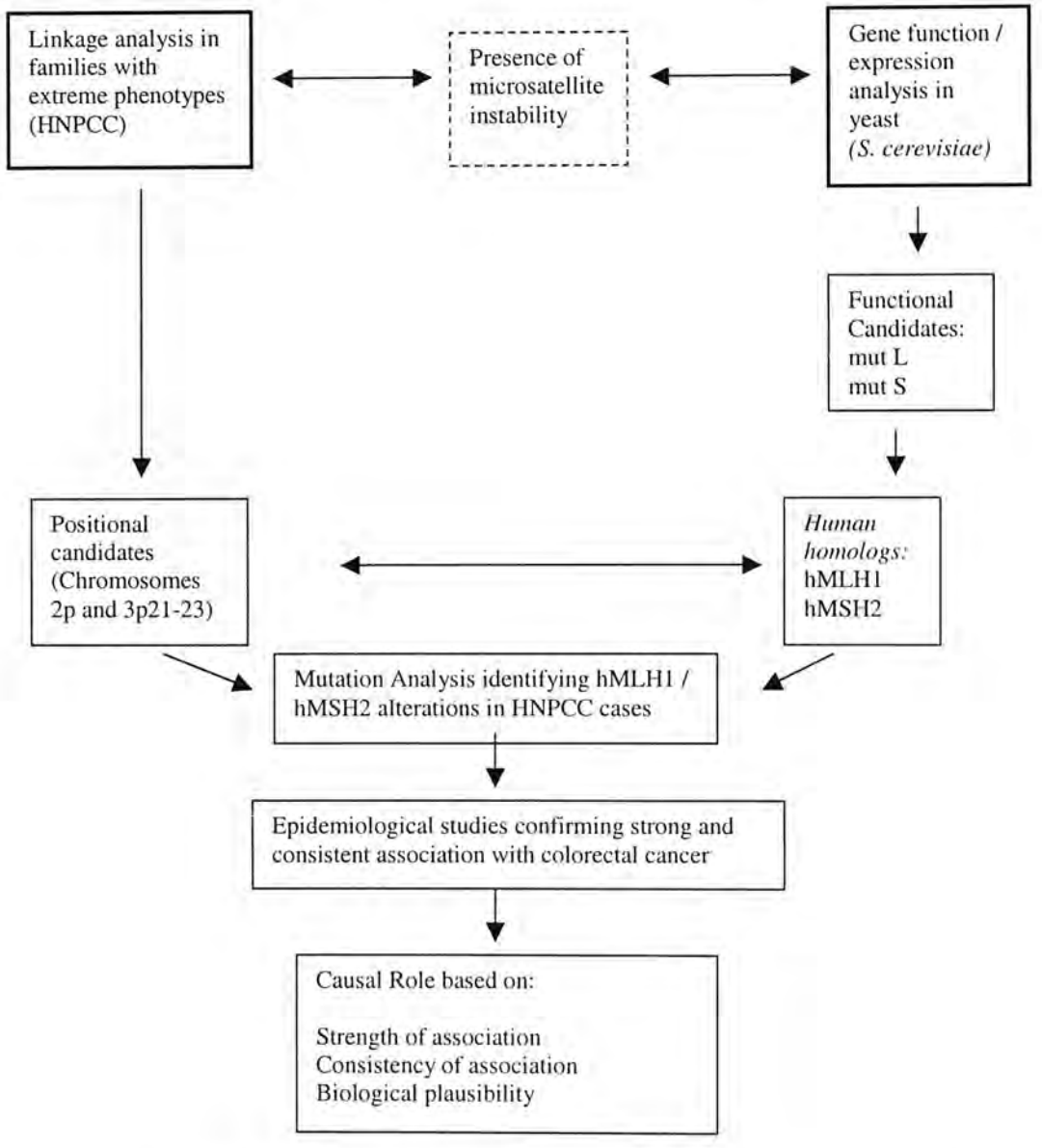
The connection between the HNPCC syndrome and mismatch repair arose from the observation that the majority of tumors from HNPCC families exhibited a replication error phenotype, a feature resulting from instability of microsatellite repeats during replication that is found only in a minority of "sporadic" colorectal cancer cases (1, 128). Previous molecular studies in the yeast *Saccharomyces cerevisiae* had led to the identification of a group of genes, known as mismatch repair genes, that were involved in maintaining the fidelity of DNA replication. Defects in yeast mismatch repair genes led to microsatellite instability (MSI), prompting formulation of the hypothesis that human homologs of these genes were involved in the HNPCC syndrome (264). Subsequently, several such homologs were identified, and two of them, hMLH1 and hMSH2, were shown to reside on chromosomes 3p21–23 and 2p21 respectively (29, 69, 161, 216). Further supportive

evidence came from the observation that pathogenic mutations in hMLH1 or hMSH2 could be identified and shown to segregate with disease in a high proportion of kindreds that had shown linkage to the corresponding chromosome (69, 161, 216).

The heterogeneity of mutations in mismatch repair genes means that screening for mutations in these genes is a lengthy and complicated process, and consequently mutation analysis has largely been restricted to colorectal cancer patients deemed to have a relatively high probability of carrying such a mutation. Only two studies identified in this review conducted mutation analysis among control subjects. Farrington et al. (62) found that none of 26 Scottish blood donors harbored previously identified mutations, although four variants of unknown significance were found. This was compared with the identification of potentially pathogenic mutations in 14 of 50 colorectal cancer patients diagnosed at less than 30 years of age. Similarly, no pathogenic mutations were reported in an analysis of 73 population controls from Utah (242).

Thus, the practical restrictions on mutation analysis, coupled with the low population prevalence of mismatch repair gene mutations and the fact that such mutations are found only in a minority of colorectal cancer patients, has meant that traditional cohort and case-control study designs have not been feasible. However, despite this lack of conventional epidemiologic evidence, subsequent studies have provided convincing evidence to support the hypothesis that mismatch repair gene mutations cause a subset of colorectal cancer cases. This evidence is summarized in figure 2.5.

Figure 2.5 Pathways of Epidemiologic and Biologic Research Identifying and Confirming the Causal Role of hMLH1/hMSH2 in Colorectal Cancer



Abbreviations: HNPCC, hereditary nonpolyposis colorectal cancer; *hMLH1*, human mutL homolog 1; *hMSH2*, human mutS homolog 2.

The most compelling supportive evidence comes from studies which demonstrate that mutation carriers are at greatly increased risk of developing colorectal cancer in comparison with the general population. Such studies are summarized in table 2.2.

Table 2.2 Risk Analysis Studies of hMLH1 / hMSH2 and Colorectal Cancer

Reference	Area of study	Ascertainment of index cases	No. of index cases	Ascertainment of mutation carriers	No. of mutation carriers	Penetrance* in mutation carriers	Source of data for comparison	Risk in comparison group	Standardized incidence ratio/relative risk
Aarnio et al. (4)	Finland	Members of HNPCC kindreds previously shown to have an *hMLH1* or *hMSH2* gene mutation.	50	Test-positive or obligate carriers.	360	♂ = 100%; ♀ = 54% (to age 70 years)	Finnish Cancer Registry data, 1991–1995	N/A	♀ + ♂ = 68 (95% CI: 56, 81)
Dunlop et al. (57)	Scotland	Colorectal cancer cases aged < 30 years identified through the Scottish National Cancer Registry between 1970 and 1993, excluding those with a family history fulfilling the Amsterdam criteria.	6	Relatives were traced, tested for mutation status where possible, and classified accordingly.	67	♂ = 74%; ♀ = 30% (to age 70 years)	United Kingdom cumulative incidence data published by EUCAN (118)	♂ = 2.53%; ♀ = 1.67% (to age 70 years)	♂ = 29#; ♀ = 18#
Aarnio et al. (3)	Finland	Families that fulfilled the Amsterdam criteria. In 24 of these, mutation analysis had demonstrated the segregation of *hMLH1* or *hMSH2*.	40	Cases of any cancer in relatives were identified and were included if adequate documentation was available "with the presumption that all tumor patients were HNPCC gene carriers."	293	♀ + ♂ = 78% (lifetime)	Finnish cumulative incidence data published by EUCAN (118)	♀ + ♂ = 2.6% (to age 75 years)	♀ + ♂ = 30#

Vasen et al. (286)	Netherlands	Families that fulfilled the Amsterdam criteria, identified through the Netherlands HNPCC registry and found to have a mutation in *hMLH1* or *hMSH2*.	19	Relatives were traced and tested for mutation carrier status where possible.	210	♂ = 92% ♀ = 83% (to age 75 years)	Netherlands cumulative incidence data published by EUCAN (118)	♂ = 4.41% ♀ = 3.28% (to age 75 years)	♂ = 21[#] ♀ = 25[#]
Vasen et al. (283)	Netherlands and Norway	Kindreds registered with the Netherlands HNPCC registry (n = 193) or suspected HNPCC families from the Clinical Genetic Centre, Radium Hospital, Norway (n = 58).	34 *hMLH1* carriers; 40 *hMSH2* carriers	Mutation carrier status was assigned to three groups: 1) tested carriers; 2) relatives with colorectal or endometrial cancer (excluding those tested negative); and 3) obligate carriers	362 *hMLH1* carriers; 301 *hMSH2* carriers	*hMLH1*: ♂ = 65%; ♀ = 55%; ♀ + ♂ = 60%; *hMSH2*: ♂ = 73%; ♀ = 54%; ♀ + ♂ = 65% (to age 70 years)	Netherlands cumulative incidence data published by EUCAN (118)	♂ = 2.81%; ♀ = 2.17% (to age 70 years)	*hMLH1*: ♂ = 23[#]; ♀ = 25[#]; *hMSH2*: ♂ = 26[#]; ♀ = 25[#]
Froggatt et al. (75)	England	Families that fulfilled the Amsterdam criteria, with mutations in *hMLH1* or *hMSH2*.	8	Subjects with mutations were included in the analysis. No further details were given.	50 (*hMLH1*: n = 23; *hMSH2*: n = 27)	*hMLH1*: ♀ + ♂ = 67%; *hMSH2*: ♀ + ♂ = 62%	United Kingdom cumulative incidence data published by EUCAN (118)	♀ + ♂ = 3.16%	*hMLH1*: 21[#]; *hMSH2*: 20[#]

Millar et al. (194)	Canada	Women with both colorectal cancer and endometrial cancer before age 70 years, identified through the Ontario Cancer Registry and/or the tumor registry at Princess Margaret Hospital, Toronto, and harboring *hMLH1* or *hMSH2* mutations.	7	First-degree relatives were identified. Carrier status was *not* determined.	N/A	N/A	Ontario provincial cancer rate	First-degree relatives of mutation carriers: 8.1 (95% CI: 3.5, 15.9); first-degree relatives of mutation-negative probands: 2.8 (95% CI: 1.7, 4.5)
Lin et al. (167)	United States	Kindreds were known to have mutations in *hMLH1* ($n = 2$) or *hMSH2* ($n = 2$). No further detail was given on how these kindreds were ascertained.	4	Mutation carriers were identified by testing ($n = 78$) or determined to be obligate carriers ($n = 27$).	105	*hMLH1*: ♂ = 94%; ♀ = 63%; ♀ + ♂ = 84%; *hMSH2*: ♂ = 96%; ♀ = 39%; ♀ + ♂ = 71%	N/A	N/A

Abbreviations: HNPCC, hereditary nonpolyposis colorectal cancer; *hMLH1*, human mutL homolog 1; *hMSH2*, human mutS homolog 2; N/A, not applicable; CI, confidence interval; EUCAN, European Network of Cancer Registries; ♂, penetrance in males only; ♀, penetrance in females only; ♀ + ♂, penetrance in group comprising both sexes.

[#] Where EUCAN data have been used for comparison, the estimate of the standardized incidence ratio is a crude one and does not take into account the age structure of the mutation carrier group. Because of the approximate nature of this comparison, we did not consider it appropriate to calculate confidence intervals for these estimates.

Aarnio et al. (4) calculated a standardized incidence ratio of 68 (95 percent confidence interval: 56, 81) for Finnish carriers of hMLH1 or hMSH2 mutations. In the other studies considered in table 2.2, researchers did not make a formal calculation of the standardized incidence ratio, but approximate estimates utilizing appropriate cancer registry data consistently show that the risk of colorectal cancer in mutation carriers is greatly in excess of the corresponding risk in the general population. The relative risk of 8.1 (95 percent confidence interval: 3.5, 15.9) for first-degree relatives of mutation carriers observed by Millar et al. (197) is consistent with a risk that is an order of magnitude greater in mutation carriers than in non-carriers.

The clinical presentation of colorectal cancer among mutation carriers appears to differ from that found among persons with sporadic cases in several respects, an observation that indirectly supports the hypothesis that mutations in mismatch repair genes account for a distinct subset of colorectal cancer cases. The most obvious clinical characteristic associated with colorectal cancer among mismatch repair gene mutation carriers is familial aggregation. Part *a* of Appendix A4 provides details on mutation analysis studies conducted among patients selected on the basis of family history. The results of these studies are summarized in tables 2.3 and 2.4.

Table 2.3 Association between the extent of family history of colorectal cancer and the prevalence of mismatch repair gene mutations

Family history criteria*	No. of studies	No. of index cases	*hMLH1* mutation carriers		*hMSH2* mutation carriers		Published References
			No.	%	No.	%	
Fulfillment of the Amsterdam criteria	27	534	145	27.2	87	16.3	(11, 12, 33, 35, 47, 49, 52, 75, 98, 105, 117, 146, 158, 175, 180, 189, 200, 202, 209, 219, 248, 268, 270, 292, 295, 297, 306)
Strong family history not fulfilling the Amsterdam criteria	25	494	46	9.3	43	8.7	(11, 12, 14, 35, 47, 49, 52, 98, 105, 117, 158, 175, 180, 189, 202, 209, 219, 248, 268, 270, 292, 294, 297, 306, 323)

*See table 1.1 for details

Table 2.4 Results of mutation analysis in patients fulfilling the Amsterdam criteria for colorectal cancer, by geographic origin

Country	No. of index cases	hMLH1 mutation carriers	hMSH2 mutation carriers	Published reference
		Asia		
Japan	15	1	8	Bai et al. (11)
Japan	11	5	0	Miyaki et al. (200)
Japan	4	0	1	Nomura et al. (209)
Korea	25	8	0	Han et al. (98)
Total	55	14 (25.5%)	9 (16.4%)	
		Europe		
Russia/Moldavia	7	1	3	Maliaka et al. (189)
Sweden	21	5	1	Tannergard et al. (270), Wahlberg et al. (292)
Sweden	7	1	0	Liu et al. (175)
Switzerland	10	3	3	Buerstedde et al. (33)
Switzerland	15	6	4	Heinimann et al. (105)
Switzerland	14	10	0	Hutter et al. (117)
Italy	14	4	3	Pensotti et al. (219)
Italy	18	1	2	de Leon et al. (49)
Italy	17	5	2	Viel et al. (290)

Italy	17	2	3	Curia et al. (47)
Italy	13	3	3	Calistri et al. (35)
France	10	3	2	Dieumegard et al. (52)
France	3	2	0	Wang et al., 1997 (294)
France	22	11	3	Wang et al., 1999 (295)
Holland / Norway	92	25	16	Wijnen et al. (306)
Germany	57	11	4	Lamberti et al. (158)
England	17	3	5	Froggatt et al. (75)
Total	344	96 (27.9%)	54 (15.7%)	
		Australia		
Australia	18	4	2	Kohonen-Corish et al. (146)
Australia	33	11	9	Scott et al. (248)
Total	51	15 (29.4%)	11 (21.6%)	
		North America		
USA	12	4	2	Luce et al. (180)
USA	28	10	1	Syngal et al. (268)
Canada	14	2	5	Bapat et al. (12)
Total	54	16 (29.6%)	8 (14.8%)	

The observed prevalence of potentially pathogenic mutations in individuals meeting the Amsterdam criteria is remarkably consistent across different populations (table 2.4). MSI is evident in 12–15 percent of sporadic colorectal cancer cases, compared with over 90 percent of cases defined, according to the Amsterdam criteria, as being from HNPCC kindreds (19). MSI is currently thought to result from defective mismatch repair, although evidence to support this hypothesis is limited by two factors. Firstly, the vast majority of studies that examine mutations in MSI-positive patients concentrate on HNPCC families, introducing considerable bias. Secondly, few investigators have looked systematically for mutations in patients with MSI-negative tumors. Interestingly, when this has been done, there have been a few instances in which tumors from patients with identified mutations in hMLH1 or

hMSH2 have not exhibited the MSI phenotype (12, 50, 62). Analysis of all published results from one research group showed that, among kindreds with suspected HNPCC, germline mutations could be detected in 16 out of 22 colorectal cancer patients with MSI-positive tumors, as compared with one out of 37 mutations in MSI-negative patients (174). The presence of mutations in MSI-negative cases may reflect mechanisms of tumorigenesis in people with mismatch repair gene mutations that do not require mutation instability. Mutation analysis studies involving patients selected on the basis of MSI are summarized in appendix A4, part *b*.

The association between early age of colorectal cancer onset and hMLH1/hMSH2 gene mutations is often confounded by the fact that the selection criteria have included family history, but a few studies have performed mutation analysis on patients selected solely on the basis of early age of onset. As is illustrated in table 2.5, these studies demonstrate a trend towards a higher pathogenic mutation detection rate in individuals diagnosed at a relatively young age, an observation that is consistent with the hypothesis that these genes are involved in colorectal cancer tumorigenesis. Details on the studies considered in table 2.5 can be found in appendix A4, part *c*.

Table 2.5 Association between age at onset of colorectal cancer and mismatch repair gene mutations

Age range (years)	No. of studies	No. of index cases	*hMLH1* mutation carriers		*hMSH2* mutation carriers		Published reference(s)
			No.	%	No.	%	
<30	1	50	7	14	7	14	Farrington et al. (62)
<40	1	12	1	8.3	1	8.3	Syngal et al. (268)
<45	1	38	1	2.6	2	5.3	Fornasarig et al. (72)
<50	6	135	6	4.4	6	4.4	Dieumegard et al. (52), Montera et al. (201), Tomlinson et al. (276), Wang et al., 1997 (294), Wang et al., 1999 (295), Weber et al. (299), and Yuan et al. (323)

2.6.1 Penetrance

While it has become widely accepted that mutations in the mismatch repair genes hMLH1 and hMSH2 play a causal role in a subset of colorectal cancer cases, the precise penetrance of these mutations remains unknown. A number of studies, summarized in Table 2.2, have addressed this issue. Results are presented differently for each study, so direct comparison is difficult. One consistent finding is that risk is higher among male mutation carriers (approximately 80 percent by age 70 years) than among females (approximately 40 percent by age 70 years), an observation with important implications for patient management and surveillance. Observed differences in penetrance between carriers of hMLH1or hMSH2 mutations (170, 286) await confirmation in future studies.

A study by Aarnio et al. (4) classified relatives of clinically defined HNPCC cases as being at a 25 percent, 50 percent, or 100 percent risk of being mutation carriers and

calculated the cumulative incidence of colorectal cancer amongst carriers up to age 70 years as being 100 percent and 54 percent for males and females, respectively. A potential source of bias in this particular study is the fact that the majority of the probands had one of the Finnish founder mutations. A similar study carried out in Amsterdam Dutch kindreds calculated risk of colorectal cancer among mutation carriers at age 75 years to be 92 percent in males and 83 percent in females (289).

These studies used family history as a selection criterion, an approach that introduces considerable ascertainment bias. Kindreds identified in this way will inherently have an unusually large number of colorectal cancer cases, and estimates of penetrance obtained in this way are likely to be falsely high. Dunlop et al. (57) used an alternative approach to identify mutation-carrying probands from the Scottish population, performing mutation analysis on colorectal cancer patients with a very early age of onset (< 30 years). The cumulative incidence of colorectal cancer among relatives proven to be mutation carriers was found to be 74 percent in males and 30 percent in females at age 70 (57).

The identification of families with mismatch repair gene mutations using any phenotypic selection criteria introduces ascertainment bias, and such kindreds may not be representative of all mutation-carrying families in the general population. Thus, there is a considerable need for estimates of penetrance based on systematically collected familial or population data.

2.6.2 Survival

Prior to the identification of mismatch repair genes, several studies suggested that the prognosis for patients with colorectal cancer due to HNPCC was more favorable than that for patients with sporadic colorectal cancer. Whether improved prognosis is specifically a feature of colorectal cancer in patients harboring mismatch repair gene mutations is not yet clear, although preliminary evidence suggests that this may be the case (105, 245).

A possible explanation for this phenomenon may be that the high frequency of mutations characteristic of mismatch repair-deficient tumors actually restricts tumor growth (245). However, kindreds included in survival analysis studies on the basis of a strong family history of colorectal cancer have, by definition, survived to produce a large family group for analysis. Therefore, these kindreds may not be representative of all mutation carriers, and there is a need for survival data from unselected, population-based cohort studies.

It has also been postulated that mismatch repair deficiency may have an effect on response to chemotherapy. Results are not entirely consistent, but several studies suggest an association between hMLH1/hMSH2 deficiency in cell lines and resistance to chemotherapeutic agents (5, 37, 68, 81).

2.7 Interactions

While the exact penetrance of specific mutations in hMLH1 and hMSH2 is unknown, it is not complete. Consequently, the age-related risk, pathologic features, and clinical outcomes associated with such mutations are subject to modification by other genetic and environmental factors.

The body of epidemiologic data regarding modification of disease resulting from mismatch repair gene mutations is somewhat limited. The effects of known environmental risk factors for colorectal cancer in mutation carriers are largely unstudied, and much of the suggestive evidence for interactions comes indirectly from studies using MSI-positive or clinically defined HNPCC cases as a surrogate for mutation carriers. Furthermore, the apparent presence of a statistical interaction between mismatch repair gene mutations and other genetic or environmental factors does not necessarily imply the existence of a biologic or causal interaction. Therefore, the studies considered below do not constitute evidence for true interactions involving hMLH1 and hMSH2, although they may prove useful in terms of identifying potential interactions that merit further investigation.

2.7.1 Gene-environment interactions

Reports by Ruschoff et al. (240) and Yamamoto et al. (317) have suggested that treatment of hMLH1- or hMSH2-deficient cell lines with nonsteroidal antiinflammatory drugs leads to a significant reduction in the proportion of cells

exhibiting MSI, indicating that this phenotypic manifestation of mismatch repair deficiency may be modified by these drugs.

Slattery et al. (254) have presented evidence suggesting that an interaction may exist between MSI and smoking. Compared with patients with MSI-negative tumors, patients with MSI-positive tumors were more likely to be heavy smokers: Odds ratios were 1.6 (95 percent confidence interval: 1.0, 2.5) in men and 2.2 (95 percent confidence interval: 1.4, 3.5) in women (254). These results are supported by those of another recent study (315), and the implication that smoking is specifically associated with a particular subset of colorectal cancer cases is consistent with the weak associations reported between smoking and sporadic colon cancer. It is possible that mismatch repair deficiency is involved in the observed association between smoking and MSI, but further studies involving known mutation carriers will be required to confirm this hypothesis.

Another recent paper by Slattery et al. (257) showed that the risk of MSI-positive colon cancer may be reduced by estrogens and increased by estrogen withdrawal. Dietary heterocyclic aromatic amines are another risk factor that requires further evaluation. Wu et al. (315) found that patients with MSI-positive tumors had received a relatively high dietary exposure to heterocyclic aromatic amines, an observation that remained significant after adjustment for smoking and red meat intake. This finding is consistent with laboratory studies, which have shown that rats exposed to particular heterocyclic amines showed the trait of MSI (36).

2.7.2 Gene-gene interactions

Risk of colorectal cancer among female hMLH1/hMSH2 mutation carriers is approximately half the risk in male mutation carriers (4, 57). In the absence of clear evidence of hormonal influence, the presence of a genetic modifier, X-linked or otherwise, remains a possibility.

The possibility of interaction between mismatch repair genes and other genes known to influence colorectal cancer susceptibility is an area that merits consideration. Initial studies have suggested that genes involved in carcinogen metabolism might modify the phenotypic expression of mismatch repair gene mutations. For example, Moisio et al. (23) demonstrated that a specific polymorphism in the gene encoding the xenobiotic enzyme *N*-acetyltransferase 1 was associated with a lower age of colorectal cancer onset in Finnish HNPCC kindreds with identified mutations in hMLH1. Similarly, an alteration in cyclin D1 has been associated with earlier age of onset in HNPCC cases; patients who harbour the mutant cyclin D1 allele have been shown to develop cancer an average of 11 years earlier than patients with two wild-type alleles (148).

Murine studies have demonstrated that MSH2 deficiency accelerates intestinal tumorigenesis in transgenic mice that are heterozygous for a germline mutation in the adenomatous polyposis coli gene (229). Similarly, Toft et al. (275) have used mice mutant for both MSH2 and p53 to demonstrate interaction between these genes. Additionally, in-vitro studies have suggested that interactions may exist

between mismatch repair genes and transforming growth factor-β receptor II (192). While these molecular studies demonstrate that gene-gene interactions may be worth further investigation, the above hypotheses have yet to be tested in human populations for relevance to cancer susceptibility.

2.8 Laboratory Testing

The heterogeneity of mutation types found in hMLH1 and hMSH2 has meant that many different techniques have been employed to test for mutations in these genes. A number of techniques are described below, along with their benefits and disadvantages.

2.8.1 In Vitro Synthesized Protein Assay

The in vitro synthesized protein assay technique uses an in vitro system to transcribe and translate a large polymerase chain reaction product containing several exons. The translated product is separated on a polyacrylamide gel electrophoresis system, and potential mutations are identified as truncated bands. These may represent a number of mutations that have the effect of altering splicing, therefore producing a translated fragment with certain exons deleted. Out-of-frame deletions or insertions, resulting in frameshifts or splice variants, will also be detected using this method.

In vitro synthesized protein assay does not detect missense mutations, in-frame deletions or insertions, large genomic deletions involving numerous exons, promoter

mutations, or mutations that silence the gene. The assay also requires the use of mRNA for the production of a cDNA polymerase chain reaction product.

2.8.2 Sequencing

cDNA sequencing also relies on mRNA being available. It will identify all mutation types except large genomic deletions, promoter mutations, and gene silencing mutations. Genomic sequencing detects even fewer changes than cDNA, but it does have the advantage of only requiring genomic DNA.

2.8.3 DNA Structure Techniques

A number of techniques rely on changes in DNA structure created by a mutation. These include denaturing gradient gel electrophoresis (305), including the adaptation of using two-dimensional gel electrophoresis (246), single-strand conformational polymorphism analysis (32), heteroduplex analysis, and denaturing high-performance liquid chromatography.

Table 2.6 summarizes the available information regarding the sensitivity of the above techniques. The use of various combinations of techniques may enhance sensitivity, but this is usually impractical. Recently, Yan et al. (319) demonstrated that the conversion of chromosomes from the diploid state to the haploid state, by fusion to a recipient rodent cell line, may facilitate improved sensitivity of current mutation detection techniques.

Table 2.6 Sensitivity of Mutation Detection Techniques

Technique	Sensitivity (%)	Published Reference
In vitro synthesized protein assay	69	Farrington et al. (62)
Genomic sequencing	80	Farrington et al. (62)
In vitro synthesized protein assay/genomic sequencing	93	Farrington et al. (62)
Denaturing gradient gel electrophoresis	>67	Fidalgo et al. (67)
Single-strand conformational polymorphism	>67	Fidalgo et al. (67)
Protein truncation test	50	Fidalgo et al. (67)
Heteroduplex analysis	19	Fidalgo et al. (67)
Two-dimensional DNA typing	—*	Sasaki et al. (246)

* Comparable to that of denaturing gradient gel electrophoresis.

2.9 Conclusions

The mismatch repair genes hMLH1 and hMSH2 are highly heterogeneous, and a proportion of gene variants constitute pathogenic mutations with a causal role in a sub-set of colorectal cancer cases. Carriers of pathogenic mismatch repair gene mutations are at substantially increased genetic risk of developing colorectal cancer, and thus constitute an important sub-group that merit particular consideration as part of strategies for preventing colorectal cancer through targeting those at increased genetic risk.

Chapter 3

Family History Studies

3.1 Introduction

Empirical risk assessment based on family history remains an integral part of current strategies to identify individuals who are at increased genetic risk of colorectal cancer. This approach is commonly employed both for clinical purposes, and as part of studies into the genetic epidemiology of this disease. A thorough understanding of this subject is essential for optimal utilisation of family history information in the broader context of preventing colorectal cancer at the population level.

Several gaps in the current knowledge and understanding of family history were addressed through the analysis of a unique data resource comprising family history information reported at interview and subsequently validated through record linkage to death records and the Scottish Cancer Registry. The following studies were undertaken as part of this analysis:

(i) Accuracy of Reporting of Family History of Colorectal Cancer

(ii) Prevalence of Family History of Colorectal Cancer

(iii) A Retrospective Family History Case-Control Study of the Risk Associated with Family History of Colorectal Cancer

(iv) A Prospective Study of Risk Associated with Family History of Colorectal Cancer

This chapter provides an overview of the construction of the data set (section 3.2). Subsequently, each of the above studies is presented in the general format of a scientific paper, with the rationale, specific methods, results and discussion

considered in turn. A discussion of the general methodology is presented in section 3.7, followed by a summary of conclusions from these studies and a consideration of further work (section 3.8).

3.2 General Methodology

The data set considered in this chapter was based on two parallel family history case-control studies, conceived and initiated in the 1990s. The author of this thesis organised data for record linkage and the conducted the analyses considered in this chapter, but was not involved in the design of the original study or the ascertainment of study subjects.

3.2.1 Ascertainment of Study Subjects

(i) Study "A" (Non-Age Selected Study Group)

A total of 199 consecutive colorectal cancer cases were ascertained between 1991 and 1993, from three hospitals in Southeast Scotland (St. Johns Hospital, The Royal Infirmary of Edinburgh and the Western General Hospital). Cases were referred to the study by their surgeons. The initial strategy employed to identify suitable control subjects was to use spouses of cases. However, this approach proved impractical, and consequently only 25 controls were identified by this means. A further 108 controls, matched by age (to within two years) and sex, were randomly selected from general practice lists in the Edinburgh area. Subjects were contacted and invited to attend an interview with a genetics nurse. Where a case or control subject did not respond, or declined to participate, another subject was selected. The initial participation rate was over 80%. At interview, subjects were asked to provide information on the name, date of birth and health status of each of their relatives. Specific enquires were then made regarding the cancer experience of each relative. Pedigrees were constructed and the associated health information provided by the interviewee was recorded.

(ii) Study "B" (Young Persons Study Group)

Study B utilised the Scottish Cancer Registry in order to specifically identify colorectal cancer cases from throughout Scotland who were diagnosed between 1971 and 1991, and had an extremely early age at onset (< 30 years). A total of 113 such cases were identified and included in the data set. A total of 100 control subjects were then identified through Scottish death records, by searching for sex matched individuals who had died of a traumatic cause and were born in the same year as the matched case. Subjects in this study were not interviewed.

In both studies, family pedigrees were constructed or extended by searching Scottish records of births, deaths and marriages for as many first and second-degree relatives of controls subjects as possible. Where appropriate, death certificates were examined, and cause of death was recorded.

Information regarding relatives of cases and controls was initially recorded in hard copy in the form of pedigrees and associated notes. Data relating to each relative was then entered into an Excel™ spreadsheet.

3.2.2 Record Linkage

The data files created in the above manner were subsequently matched with death records and the Scottish Cancer Registry, in order to establish whether each relative (or "data subject") was alive or dead, and to determine their personal cancer history. This was done through computerised record linkage performed by the Information and Statistics Division (ISD) of the Scottish Executive. Initial matching took place

in 1993, and the record linkage process was repeated in July 2001. All current analyses were based on this second linkage.

The Scottish Record Linkage System has been designed to bring together all records relating to hospital discharge, cancer registration and death for each patient. Using probability-matching techniques based on patient-specific identifying information, linkage of such records can be achieved with a rate of false-positive matches estimated at below 1% (12,13,14). The same methodology can be applied to linking research data with the health information held by ISD.

The record linkage protocol utilised at ISD is based on the principles of "probability matching" developed by Howard Newcombe (207), which state that the probability of a 'true' match increases each time two items of information agree, and decreases each time they disagree. The information generally used to match records includes surname, first initial (plus full forename and second initial if possible), sex, date of birth, and postcode. The databases created through the colorectal cancer case control family history studies did not contain postcode information, but with this qualification the linkage process used for these data followed the general procedure described below.

To begin with, the surnames were coded according to the New York State Intelligence Information System (NYSIIS) name compression algorithm. This program removes vowels and unifies commonly confused letter groups such as "ch"

and "sh". Subsequently the names are coded using the Soundex system (207), which assigns the same code to non-initial consonants with similar sounds.

The first step in the actual linkage involved 'blocking' of records that did not reach a minimal level of agreement with the record of interest. This was done in order to reduce the number of comparisons that need be made. Direct comparisons were then made between records that agree on all of Soundex/NYSIIS code, first initial and sex (block A) or all elements of date of birth (block B).

Odds ratios were calculated for each piece of information, providing a measure of the probability of matching, 'weighted' according to the probability of the result occurring by chance. For example, agreement of the "sex" item of information provides an odds ratio of 1.99, calculated by dividing 0.995 (the probability of this item matching if the records refer to the same person) divided by 0.5 (the probability of this item matching if the records refer to different people). Separate odds ratios were then combined, providing an overall estimate of the probability that the two records genuinely refer to the same person.

The above procedure can only provide relative estimates; the absolute probability of a true match will also depend on the size of the databases being searched, and the amount of information available for each entry. Sophisticated mathematical methods would be required to make a formal assessment of absolute probability, and such an approach is time consuming and not necessarily accurate. The alternative approach used by ISD was manual inspection of a representative selection of records and the

identification of a probability threshold at which the records are considered more likely than not to be a true match. Pairs of records with estimates above this threshold were accepted as matches. In practice, most records would be expected to clearly match or clearly belong to different people; only a few pairs of matches will be on the borderline.

Record linkage thus facilitated both the confirmation of cancers reported at interview, and the identification of previously unknown cancer cases in data subjects. There are, however, some limitations associated with this approach, which can be categorised as relating to the quality of the study data used to link records, the record linkage process itself, and the Scottish Cancer Registry. These limitations and their possible impact on the results presented in this chapter are considered further in section 3.7.

3.2.3 Definition of Colorectal Cancer

For the purposes of this study, the diagnosis of colorectal cancer was considered to include adenocarcinomas of the colon, rectum, recto-sigmoid junction and anus, as defined by the International Classification of Disease (ICD) system.

3.2.4 Ethical Approval

Ethical approval for the recruitment and interview of patients and controls was granted by the appropriate Local Research Ethics Committee (LREC) at the time the family history case-control studies were initiated.

Approval for record linkage was sought from the Privacy Advisory Committee at ISD. This committee was set up by the Chief Medical Officer to advise ISD and the Registrar General (for Scotland) on the release of patient-identifiable data and the linkage of datasets that have not previously been linked. The committee comprises a chairman and at least four members. The current membership includes a Professor in Public Health Medicine, a Clinician, a Consultant in Public Health Medicine, and three lay members. Approval for record linkage to be performed in connection with the current analyses was granted, with the condition that all linked data remained at ISD throughout the analyses to ensure the confidentiality of the matched information regarding relatives' cancer status.

3.3 Accuracy of Reporting of Family History of Colorectal Cancer

Elements of the research presented in this section have been published in the journal *Gut* (195). This paper is presented for reference in appendix A5.

3.3.1 Rationale

Family history is known to be an important significant risk factor for colorectal cancer, and is used in the clinical setting to inform decisions regarding the use of colonoscopic surveillance. The degree of personal risk relates to the extent of family history and the age of onset of affected relatives (134). Various guidelines based on degree of family history have been devised to determine when surveillance should be recommended (55, 56, 109, 249, 310). The accuracy of family history information is clearly critical whenever family history informs decisions about the necessity of colonoscopic surveillance. It is also an important consideration in the context of the epidemiological studies that inform the ongoing debate regarding the appropriate guidelines for offering surveillance.

In both situations, information on family history is often gathered by interview with a family member. This approach may be subject to error on the part of the interviewee, with under-reporting of family history being observed in previous studies (91, 140). Interview data is also potentially subject to systematic recall bias arising from the fact that people with raised awareness of colorectal cancer, such as patients, may be more likely to report a positive family history (71). Furthermore, the social stigma associated with bowel cancer may mean that this condition is

discussed less readily within families, and this factor could particularly affect reporting of family history.

Accuracy of reporting cancer in the family has been addressed in previous studies of people referred to genetics departments because of a cancer family history (53, 177, 253), people with a personal history of cancer (7, 71, 137, 138, 140, 145, 210), or close relatives of cancer cases (21, 91, 152). However, only a few studies have related specifically to colorectal cancer cases (7, 91, 137, 138, 140) or to community based consultands who have not been referred to a genetics clinic (140, 145). An additional limitation common to many published studies is that validation of the interviewee's report is only attempted for relatives reported to have had cancer. In such studies no information can be obtained regarding the sensitivity, specificity or negative predictive value of reports, and the question of under-reporting of cancer cannot be addressed.

In this analysis information on family history of colorectal cancer obtained from colorectal cancer cases and population controls at interview has been compared with death records and Scottish Cancer Registry, using record linkage. Information was obtained regarding the cancer experience of a total of 5637 first and second-degree relatives of 199 colorectal cancer cases and 133 controls. This data set facilitated an assessment of overall accuracy of reporting of family history of colorectal cancer at interview, including under-reporting. In addition, the effect of any inaccuracies was evaluated with respect to clinical genetic risk assessment. The findings have

considerable relevance to the methods used to validate family history and also have practical implications for surveillance guidelines based on degree of family history.

3.3.2 Methods

This study was based on subjects ascertained as part of study A. Summary statistics (sensitivity, specificity, positive predictive value and negative predictive value) relating to the accuracy of interviewee reports by various sub-groups of the data subjects were calculated and compared. Cancers occurring between the interview date and the record linkage were excluded from all analyses relating to accuracy of reporting. Confidence intervals were calculated using a normal approximation (308), which compares favourably with exact methods for this data (6).

3.3.3 Results

The mean age of the 199 cases at the time of interview was 64.0 years. There were 86 females and 113 males. In total, 3290 first and second-degree relatives of these cases were included in the database. 110 relatives were reported to be resident in England (n = 46) or overseas (n = 64), and the nurse constructing the pedigrees classified a further 251 as "untraceable".

The 133 controls had a mean age of 64.2 years at the time of interview, and consisted of 60 females and 73 males. In total, 2347 blood relatives (i.e. individuals related genetically, as opposed to by marriage) of controls were included in the database, of which 107 were reported to be resident in England (n = 61) or overseas (n = 46) and 91 were deemed to be untraceable. Individuals who have neither died

nor developed cancer are not matched through record linkage, and it is impossible to distinguish these individuals from those who cannot be traced. Hence, all relatives were included in the subsequent analysis, regardless of apparent 'traceability'.

3.3.3.1 Knowledge of Family Members' Health and Occurrence of Cancer

Cases and controls were asked at interview to briefly state their knowledge regarding the vital status and medical history of their relatives. The proportion of relatives for whom the interviewees were able to provide any health related information is shown in table 3.1, which also details the responses given by interviewees for all relatives found to have cancer by linking with central records.

Table 3.1 Knowledge of Health and Occurrence of Cancer in Relatives

Interviewee Group	Relative group	Number of relatives	Number (%) for whom any health information was supplied	Number of relatives with confirmed cancer*	Number (%) of affected relatives in which cancer was reported	Total number of cancers#	Number (%) of cancers accurately reported
Cases	FDR	1322	1250 (95%)	215	152 (71%)	240	106 (44%)
"	SDR	1968	713 (36%)	274	84 (31%)	293	42 (14%)
Controls	FDR	1037	991 (96%)	113	76 (67%)	124	51 (41%)
"	SDR	1310	671 (51%)	189	77 (41%)	202	36 (18%)

*This column refers to the total number of relatives in a particular group found by record linkage to have had cancer.
#This column describes the total number of primary cancers in relatives; including multiple primary cancers.
Abbreviations: FDR, First Degree Relatives; SDR, Second Degree Relatives

In total, 240 cancer cases in first-degree relatives (FDRs) were identified by ISD record linkage. 106 of these instances had been correctly reported at interview. In

contrast, 293 cancers were found in second-degree relatives (SDRs), of which only 42 were correctly reported. In the majority of instances where a cancer was not correctly reported, the interviewee either had no knowledge of the health of the relative in question, or was unaware that they had developed any type of cancer. However, in some cases a cancer was reported, but the site was incorrect or unknown. An indication of the extent to which this situation occurred is provided by the sixth column in table 3.1, which states the proportion of affected relatives who were reported as having had some form of cancer.

3.3.3.2 *Reporting of Colorectal Cancer Cases*

In total, there were 148 cases of colorectal cancer in first or second degree relatives, of which 62 (41.9%) were reported correctly by the interviewee. The mean age at onset of cases that were correctly reported was 63.3 years (95% CI = 60.5, 66.1), a value significantly different from the mean age of 70.2 years (95% CI = 67.8, 72.5) for cases that were not correctly reported. This observation is not unexpected, as cancer affecting more elderly relatives is less likely to be discussed within families. The suggestion that early-onset cases are more likely to be reported accurately at interview is of clinical interest, as such cases are more significant in terms of indicating increased genetic risk. A separate trend towards more accurate reporting in recent years was evident, although this effect was not statistically significant. Summary statistics associated with the accuracy of reporting of colorectal cancer in relatives were calculated from contingency tables, and are summarised in Table 3.2. The contingency tables themselves are presented for reference in appendix A6.

Table 3.2 Summary Statistics for Reporting of Colorectal Cancer

Group	Relative Group	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value (95% CI)	Negative Predictive Value (95% CI)
Cases	FDR (n = 1322)	0.566 (0.433, 0.690)	0.990 (0.983, 0.994)	0.698 (0.549, 0.814)	0.982 (0.973, 0.988)
Cases	SDR (n = 1968)	0.271 (0.166, 0.410)	0.996 (0.992, 0.998)	0.619 (0.409, 0.792)	0.982 (0.975, 0.987)
Controls	FDR (n = 1037)	0.529 (0.310, 0.738)	0.995 (0.989, 0.998)	0.643 (0.388, 0.837)	0.992 (0.985, 0.996)
Controls	SDR (n = 1310)	0.333 (0.192, 0.512)	0.995 (0.991, 0.995)	0.667 (0.417, 0.848)	0.985 (0.976, 0.990)

Abbreviation: CI, confidence interval

The data shown in Table 3.2 demonstrate substantial under-reporting of colorectal cancer in relatives. In both cases and controls, sensitivity of reporting in FDRs is around 50-60%, implying that a large proportion of cancers in FDRs go unreported. The poor sensitivity of reporting is even more striking in SDRs, with the majority of cases in SDRs of cases and controls not being reported at interview. The very high estimates of specificity and negative predictive value shown in Table 3.2 primarily reflect the fact that in absolute terms colorectal cancer affects only a small proportion of the population. However, even small effects on these parameters may have important implications for genetic risk assessment and resource allocation. For all relative groups, estimates of positive predictive value were in the range of 60-70%, indicating that approximately one third of reports of individual colorectal cancer cases are not confirmed using record linkage. There were no differences in accuracy of family history reporting between cases and controls.

The sensitivity of reporting of colorectal cancer compared to other common cancers is shown in table 3.3.

Table 3.3 Sensitivity of Interview as a Means of Identifying Familial Cancer Cases, by Site

(i) Cases

Site	Relative Group	Number of Cases*	Number of Cases Correctly Reported	Sensitivity of Interviewee Report (95% Confidence Interval)
Colorectal	FDR	53	30	**0.566** (0.433, 0.690)
"	SDR	48	13	**0.271** (0.166, 0.410)
Breast	FDR	23	17	**0.739** (0.535, 0.875)
"	SDR	22	8	**0.364** (0.197, 0.570)
Bronchus & Lung	FDR	33	19	**0.576** (0.408, 0.728)
"	SDR	37	6	**0.162** (0.077, 0.311)
Stomach	FDR	21	9	**0.429** (0.245, 0.635)
"	SDR	34	3	**0.088** (0.030, 0.230)

(ii) Controls

Site	Relative Group	Number of Cases*	Number of Cases Correctly Reported	Sensitivity of Interviewee Report (95% Confidence Interval)
Colorectal	FDR	17	9	**0.529** (0.310, 0.738)
"	SDR	30	10	**0.333** (0.192, 0.512)
Breast	FDR	5	4	**0.800** (0.376, 0.964)
"	SDR	15	3	**0.200** (0.070, 0.452)
Bronchus & Lung	FDR	33	18	**0.545** (0.380, 0.702)
"	SDR	30	4	**0.133** (0.053, 0.297)
Stomach	FDR	9	3	**0.333** (0.121, 0.646)
"	SDR	30	8	**0.267** (0.142, 0.444)

*Where more than one primary cancer occurred at the same site, it was not possible to determine whether the interviewee was aware of both these tumours. Therefore, where metachronous primary cancers occurred, only the first is considered in table 3.3.

Estimates of sensitivity for colorectal cancer were broadly comparable with the other common cancer types listed in Table 3.3, although numbers are small. However, it is noteworthy that breast cancer was more frequently reported than the other internal cancers in FDRs. This may reflect the more enigmatic presentation of visceral malignancy and possibly social stigma associated with bowel cancer in particular.

3.3.3.3 *Practical Implications of Inaccurate or Incomplete Reporting of Family History*

From a clinical perspective it is important to evaluate the validity of interviewee reporting as a means of identifying families that are eligible for genetic counselling, clinical screening and/or genetic testing. Various guidelines exist to help determine the extent of family history that warrants such interventions. For illustrative purposes simple family history criteria (two affected FDRs, or one affected FDR with age of onset <45 years) were applied to both cases and controls from this study. These criteria are used by the British Society of Gastroenterology, and the Association of Coloproctology of Great Britain & Ireland, to indicate a requirement for surveillance (55, 56). These criteria, which are concerned only with FDRs, were applied to the original data set. The risk categorisation of these individuals was then re-evaluated following record linkage.

Cases and controls were considered together since no differences were observed between these two groups in terms of the accuracy of reporting of colorectal cancer. In order to gauge the overall impact of inaccurate or incomplete reporting on surveillance recommendations, cases and controls were considered simply as consultands, rather than cases meriting post-surgical surveillance following their own personal history of colorectal cancer.

At interview, 5 of the interviewees reported a family history that met criteria indicating a need for surveillance. However, only 2 of these 5 interviewees were confirmed by record linkage to meet these criteria, giving an overall positive predictive value of 0.4 (95% CI = 0.12, 0.77). In addition, 4 further interviewees

who did not report a family history of colorectal cancer fulfilling the applied criteria actually did have such a family history based on record linkage data. Therefore, only 2 of 6 interviewees who met the applied criteria following record linkage were identified at interview, suggesting that the sensitivity of interview in terms identifying appropriate individuals was 0.33 (95% CI = 0.10, 0.70).

3.3.4 Discussion

This study has quantified the accuracy of reported family history of cancer in two important groups of people, namely those with colorectal cancer and those from the general population. Because cases reported to have colorectal cancer were confirmed, and cases that had not been reported by the interviewee were also identified, it was possible to systematically assess overall accuracy of reported family history of large bowel malignancy. Using this approach the accuracy of reporting of colorectal cancer has been determined in a large data set comprising 332 interviewees and 5637 first and second-degree relatives.

3.3.4.1 *Knowledge of Relative's Health Status*

Evaluation of the extent to which interviewees had any knowledge of their relatives' health forms a useful background to understanding the accuracy of cancer reporting, as it provides an indication of the extent to which people are aware of the health of their relatives.

Table 3.1 shows that interviewees could provide no useful information for approximately half of all SDRs, but did have some knowledge of the health status of

all but around 5% of FDRs. This consistent disparity suggests that many instances in which cancer in second degree relatives goes unreported are due to lack of contact with relatives, rather than ignorance of diagnosis in a known family member. The observation that positive predictive value is similar in FDRs and SDRs lends further support to this notion. Clearly, one would expect that interviewees would have greater knowledge about first degree relatives, and would be more likely to receive and maintain knowledge of a cancer diagnosis from such close family. It is also possible that the age structure of the relative groups in question will serve to exaggerate such effects, since FDRs will generally be relatively close in age to the interviewee. Disparity between FDRs and SDRs is evident throughout this study, and is consistent with findings from other published studies (53, 145, 177, 253).

The results presented in table 3.1 suggest that controls were actually slightly better than cases at providing knowledge of health status and reporting cancer events in relatives. However, this difference is not statistically significant, and is subtle when compared to the disparity between knowledge of FDRs and knowledge of SDRs.

3.3.4.2 *Reporting of Colorectal Cancer in Relatives*

In this study, estimates for sensitivity of reporting relating to colorectal cancer are low, being in the region of 50-60% in FDRs of both cases and controls, and around 30% for SDRs. These estimates were broadly comparable with the other common cancer types listed in table 3.3. These results imply, therefore, that a large proportion of cancers in FDRs, and the majority of cancers occurring in SDRs, go unreported.

A comparable approach to assessing accuracy of reporting of colorectal cancer, which includes identification of unreported cases as well as checking the accuracy of cases reported at interview, has been employed in one previous study (140). This study estimated the sensitivity of reporting a family history of colorectal cancer in FDRs as 0.65 (95% CI = 0.39, 0.85) for colon cancer cases, and 0.81 (95% CI = 0.54, 0.95) for controls, and the authors concluded that subjects were able to accurately report family history (140). However, this previous study did not consider SDRs, and no information is provided regarding the total number of relatives involved. Furthermore, the focus of this paper was on validation of an epidemiological study. The observed values for sensitivity of reporting may be less acceptable for genetic risk assessment where the objective is to determine the need for clinical intervention, particularly given the wide confidence intervals.

The finding that colorectal cancer cases that are correctly reported have a relatively young age at onset is of interest, as it suggests that early-onset cases are more likely to be reported, and such cases are more significant in terms of indicating increased genetic risk. However, the magnitude of differences observed in this study may have limited impact in a clinical setting.

Positive predictive value provides information about the validity of positive reports, and is thus a highly relevant measurement. The positive predictive values shown in table 3.3 indicate that a report of cancer at interview is equally likely to be correct for a first or second-degree relative. This observation supports the notion that differences in reporting of cancer in these groups arise largely from a lack of

knowledge about SDRs, as opposed to inaccurate reports. Estimates for positive predictive value were in the range of 60-70%, indicating that approximately one third of reports of individual colorectal cancer cases are not confirmed using cancer registry data. A proportion of these apparently false positive reports of colorectal cancer may be true positives not detected by the record linkage. However, such occurrences are unlikely to account for the number of false positive results observed. The validity of record linkage as a means of determining the cancer experience of a data subject is considered in section 3.7.

One of the important questions addressed by the current study is whether or not individuals who have had colorectal cancer are more likely than controls to provide false positive reports of the condition in their relatives. In the current study, 21 false positive reports were found among the 199 cases, compared with 11 false positive reports by controls among the 133 interviewed controls. There is thus no evidence to support the hypothesis that cases may over-report their family history.

In this study, family history documentation was optimal since a trained genetics nurse conducted interviews during a lengthy consultation at the interviewee's home. Reporting inaccuracies observed in this study may well be more extreme where family history is taken in a busy gastroenterology, surgical or general practice clinic.

3.3.4.3 *Genetic Risk Assessment*

From a clinical perspective, the information provided about the family as a whole is more important than the accuracy of individual reports. The results of this analysis

illustrate that incomplete or inaccurate interviewee reporting could have a substantial impact on genetic risk assessment. That only 2 of 6 families who actually met the applied criteria were identified at interview is a particular concern, implying that reliance on interview data in a clinical context could result in many families who actually meet criteria for surveillance being overlooked. Conversely, 2 of 5 families were mistakenly classified as meeting the chosen criteria due to false positive reports made at interview. In practice, such an effect could lead to surveillance being applied unnecessarily.

The statistical power of these analyses is extremely low, due to the small numbers concerned. Accordingly, the results should be regarded as indicative of a possible role for inaccuracy and under-reporting at interview in limiting the validity of genetic risk assessment, rather than conclusive evidence of such an effect.

3.3.4.4 *Conclusion*

The appropriate family history criteria for offering genetic counselling, colonoscopic surveillance or genetic testing is the subject of much current debate, and is likely to remain so. The findings of this study are highly relevant to this discussion, as they suggest that family history information obtained by interview may be misleading, and that verification of both positive and negative interviewee reports should be conducted whenever possible. In reality, any strategy used to establish family history of cancer will be a compromise between this ideal situation and what is actually feasible.

In principle, the use of central government records to establish family history of colorectal cancer is potentially useful from a scientific and clinical perspective, but it raises important issues of consent and privacy, and may not be considered justifiable at the population level. Under the Data Protection Act (1998) (278), the use of medical information without the specific consent of the individual to whom the data refers is severely restricted. This is problematic in the context of obtaining a complete family history, since it is unlikely to be possible to obtain consent from all family members of a study subject. In many situations, data is more readily available for deceased relatives.

Practically, such an approach would be difficult to implement in the UK, as it necessitates meticulous manual preparation of a data set for record linkage. Establishing family history through computerised central records is theoretically more realistic in some Scandinavian countries, in which births, deaths, marriages and health records are linked through a unique identifying number. Provided appropriate steps are taken to ensure privacy, a systematic, computerised means of establishing family history is conceivably a future option for public health policy in the UK.

Regardless of the techniques applied, the results of this analysis illustrate the importance of considering the methods used to establish a family history, particularly when the data is used to make clinical decisions.

3.4 Prevalence of Family History of Colorectal Cancer in the General Population

3.4.1 Rationale

Details of family history are regularly collected from colorectal cancer patients, but at present information regarding the prevalence of family history of colorectal cancer amongst members of the general population is less readily available. Obtaining such data is important as it provides an indication of the number of people in a population who might meet various family history criteria, thus informing evaluation of the potential of surveillance programmes in terms of costs and effectiveness at the population level. Addressing this issue is particularly relevant given the recent increase in public demand for information and advice regarding genetic risk of cancer. Increasing awareness of the genetic contribution to colorectal cancer potentially has major resource implications for genetic counselling and clinical screening.

Current knowledge regarding the prevalence of family history of colorectal cancer generally comes from interview or questionnaire-based data, often recorded as part of case-control studies. The extent of family history observed varies considerably, with most studies suggesting that between 3% and 10% of people in the general population have a first degree relative that has developed colorectal cancer (22, 78, 112, 135, 153, 205, 243, 256, 260, 293). This variation is likely to reflect differences in methods of identifying controls and obtaining their reports, as well as genuine differences in prevalence between the populations under study.

Interview, as compared to self-administered questionnaires, is generally considered to be a more valid method of obtaining information about family history. However, information gathered in this manner may also be both incomplete and unreliable, particularly with regard to second-degree relatives (91, 140, 198). This is potentially very relevant in the context of estimating prevalence of a family history of colorectal cancer in the general population, as it implies that relying on interview data will result in cases of colorectal cancer in relatives being overlooked. The true prevalence of family history will consequently be underestimated. Cancer registries and other databases can provide a more reliable source of family history information. Such an approach to estimating the prevalence of family history of colorectal cancer has previously been employed using data from the Utah Population Database (140, 256), although findings have not been repeated or validated in other populations.

There is, therefore, a paucity of unbiased information regarding the prevalence of a family history of colorectal cancer in healthy members of the general population. The purpose of the current study was to investigate the prevalence of family history of colorectal cancer in Scotland, using record linkage to death records and the Scottish Cancer Registry to determine the cancer experience of first and second-degree relatives of population-based control subjects. This approach removes much of the uncertainty and potential bias inherent in using information reported by the controls themselves to establish their extent of family history.

3.4.2 Methods

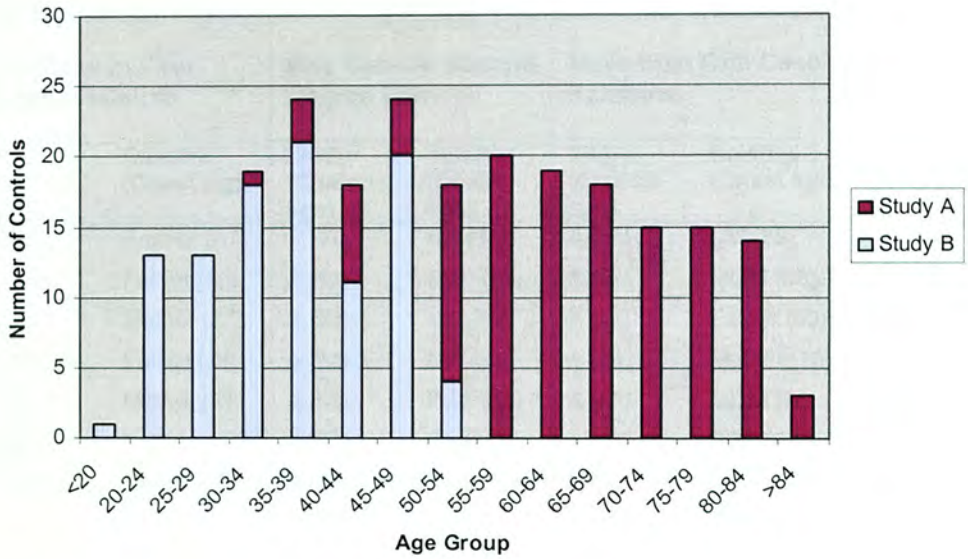
A descriptive study of the prevalence of family history of colorectal cancer in the study subjects was conducted. Confidence intervals for proportions were calculated using a normal approximation (308).

The observable extent of family history may vary according to the age of subjects. Furthermore, any population-based programme to identify individuals with a family history of colorectal cancer would probably be aimed at individuals in the 30-70 age range. For these reasons the primary study group for this prevalence study was limited to 160 control subjects within this age range from study A or B. The age of the control subjects is defined as the age at the time the subject was ascertained and the pedigree established. For Study A, this was the interview date in 1993, and for the deceased control subjects in Study B this was taken to be the age they would have been in 1994.

3.4.3 Results

The age distribution of all control subjects is illustrated in figure 4.1.

Figure 3.1 Age Distribution of Control Subjects



The average age of all the participating control subjects was 52.0 years, with a range of 19 to 85 years. The primary study group, restricted to the 160 control subjects within the 30-70 year age range, had an average age of 48.8 years.

A total of 61 colorectal cancer cases were identified in first or second-degree relatives of control subjects in the primary study group. Table 3.4 provides details of these cases, including the age of onset, age of control subject, and the relationship between the control subject and the affected relative.

Table 3.4 Summary of Colorectal Cancer Family Histories of Control Subjects

One Case in First Degree Relative		One Case in Second Degree Relative		More than One Case in First or Second-degree Relatives			
Study (Control age)	**Relative (Onset age)**	**Study (Control age)**	**Relative (Onset age)**	**Study (Control age)**	**Relative 1 (Onset age)**	**Relative 2 (Onset age)**	**Relative 3 (Onset age)**
A (61)	Brother (61)	A (47)	MA (61)	A (68)	MU (38)	PU (75)	
A (53)	Father (55)	A (51)	PGF (67)	A (67)	MGM (62)	PA (55)	
A (59)	Mother (77)	A (64)	MU (60)	A (41)	Father (60)	PU (73)	
A (60)	Father (48)	A (64)	MA (56)	A (64)	Mother (73)	MU (81)	
A (53)	Mother (77)	A (58)	PGF (83)	A (48)	MGM (75)	MGF (88)	
A (64)	Brother (47)	A (60)	PA (61)	A (58)	PGF (77)	MU (50)	
A (58)	Father (73)	A (68)	PGF (69)	**A (55)**	**MGM (77)**	**Mother (53)**	
		A (60)	MA (79)	A (55)	MGM (73)	Father (68)	PA (89)
		A (63)	PU (61)	A (68)	PGF (76)	Brother (64)	
		A (54)	PGF (69)	B (53)	Father (60)	PA (70)	
		A (52)	MA (62)	**B (33)**	**PGM (32)**	**Father (57)**	
		A (48)	MA (76)	B (47)	PGM (58)	PA (55)	
		A (54)	PU (67)	B (46)	Father (74)	PU (79)	
		B (48)	PGF (34)	B (40)	MGF (53)	PA (78)	
		B (37)	PGM (94)				
		B (32)	MA (82)				
		B (48)	PGF (34)				
		B (51)	PU (68)				
		B (46)	MU (54)				
		B (45)	PU (67)				
		B (48)	PGF (36)				
		B (42)	MGM (83)				
		B (39)	PGM (81)				
		B (38)	MA (68)				
		B (36)	PGM (84)				

BOLD TYPE = control subject classified as medium risk, according to current Cancer Genetics Guidelines for Scotland (249)
Abbreviations: MA, Maternal Aunt; MU, Maternal Uncle; MGM, Maternal Grandmother; MGF, Maternal Grandfather, PA = Paternal Aunt, PU = Paternal Uncle, PGM = Paternal Grandmother, PGF = Paternal Grandfather

The mean age of subjects with a positive family history was 52.3 years, a value not significantly different from the overall mean. Out of the 160 control subjects, 46 (28.8%; 95% CI = 21.7, 35.8) had a family history of colorectal cancer in any first or second-degree relative. 15 (9.4%; 95% CI = 4.9, 13.9) control subjects had an affected first degree relative, and 14 (8.8%; 95% CI = 4.4, 13.1) had more than one affected relative. There were no subjects with more than one affected first degree relative. Many of the multiple-affected families had affected members on different sides of the family, implying that the observed clustering is due to chance rather than a common genetic aetiology.

Applying the current Cancer Genetics Guidelines for Scotland (249) to this data set revealed that two (1.25%) of the above controls, marked in bold in table 3.4, would initially be classified as "Medium Risk", since they each have two affected relatives, of which one is a first-degree relative and one of which was diagnosed at less than 55 years of age. The recommended protocol for management of such individuals is referral to a genetics counsellor and subsequent colonoscopic surveillance comprising a single colonoscopy at 30-35 years of age, not repeated until 55 years of age if initial findings are normal (see appendix A1).

3.4.4 Discussion

This study attempts to provide an insight into the issue of population prevalence of colorectal cancer in the general population. The proportion of individuals with an affected first degree relative is often used as a convenient measurement of the prevalence of family history. In this study 9.4% (95% CI = 4.9, 13.9) of population

controls have been found to have a history of colorectal cancer in a first degree relative. This estimate is relatively high when compared to findings from two large scale questionnaire studies conducted in the UK by Sadhu et al., and House et al., in which the proportion of controls reporting a history of colorectal cancer in a first degree relative was 6.8% and 6.0% respectively (112, 243). The current estimate of the proportion of individuals with an affected first degree relative is also consistent with, but slightly higher than, the registry-based study by Kerber and Slattery, in which the corresponding value was 7.8% (140). However, as illustrated by the confidence intervals attached to the estimate, the observed difference between these results and those of these previous studies is not statistically significant and may therefore be due to chance.

A highly relevant study has recently been conducted with the specific aim of determining the prevalence of a family history of breast, colorectal and ovarian cancer among people aged 30-65 years in Scotland (293). Out of a total of 7620 participants, 608 (8.0%) reported a history of colorectal cancer in at least one first or second degree relative. Results were not reported for these relative groups separately. This estimate contrasts markedly with the current study, in which 28.8% (95% CI = 21.7, 35.8) of study participants were found to have an affected first or second degree relative. This difference is statistically significant.

The observed family history of colorectal cancer may be affected by the age distribution of the study group concerned, and cohort effects may also apply. Hence it is possible that the difference in the estimates produced by the current analyses

and the study by Wallace et al., (293) is partially the result of the fact that the study group considered in this analysis was ascertained earlier and included slightly older subjects. However, the actual differences between the age groups and the time of ascertainment are minor, and, furthermore, the hypothesis that older subjects would be more likely to report a family history of colorectal cancer is not supported by data from this study, since the mean age for those with a family history does not differ from the overall mean age.

Another potential cause of the observed differences between the results of these two studies is the potential for ascertainment bias in each, since the participation was not complete in this study group (approximately 80%), nor in the questionnaire study conducted by Wallace et al., (59%). However, both these rates are reasonably high, and the direction of any bias is unlikely to differ between the studies to an extent large enough to account for the observed differences.

The most compelling difference between these studies is in the methods used to ascertain details of family history. In this study, family history has been systematically established using record linkage to the Scottish Cancer Registry, whereas the study by Wallace et al., (293), and the majority of other previous studies designed to estimate the prevalence of family history, relied on information provided by population controls themselves. As demonstrated in section 3.3, colorectal cancer is often under-reported, particularly in second-degree relatives. The degree of under-reporting observed would account for the magnitude of the disparity observed between the current findings and those of the study by Wallace et al., (293).

Out of 160 population controls, 2 (1.25%) had a family history of colorectal cancer that would meet the required criteria for a classification of "medium risk", according to the Cancer Genetics Guidelines for Scotland (249). The scale of the current study is insufficient for this finding to be considered conclusive, but it is consistent with the findings of the study by Wallace et al., (293) who found that 45 of 7620 (0.6%) of participants met these same guidelines. Again, the higher prevalence of a particular family history observed in the present study is consistent with the hypothesis that family history of colorectal cancer is often under-reported.

Evaluation of family history provides a simple strategy for identifying people at increased genetic risk of developing colorectal cancer, and hence empirical risk assessment based on family history is an important aspect of current and future strategies for preventing colorectal cancer in this sub-group of the population. The appropriate inclusion criteria, and the methods used to establish family history in this context are crucial considerations in terms of the planning, funding and implementation of screening programmes. These issues have generated considerable debate, and continue to do so in the light of recent advances in our understanding of the genetics underlying colorectal cancer. Accurate, population specific information regarding the actual prevalence of a family history of colorectal cancer in the general population is valuable in terms of informing such debate.

In contrast to most previous studies, that are reliant on interview or questionnaire data, the current study is based on record linkage with information from a reliable cancer registry. This study has found that approximately 9.4% of people aged 30-70

have a family history of colorectal cancer in a first degree relative, and that approximately 28.8% of people have a first or second degree relative who has developed colorectal cancer. Furthermore, of 160 control subjects, two were found to have a family history of colorectal cancer that met current Scottish guidelines for referral to a genetic counsellor and possible inclusion in a colonoscopic surveillance programme. These findings should ideally be repeated by more extensive studies, but they currently provide unique and relatively unbiased information that could inform decisions made by researchers and policy makers regarding the provision of genetic services for colorectal cancer.

3.5 A Retrospective Family History Case-Control Study of the Risk Associated with Family History of Colorectal Cancer

3.5.1 Rationale

Knowledge of the extent of familial risk associated with various degrees of family history is crucial to facilitating accurate risk assessment in the clinical setting, and informing decisions made about screening protocols and genetic testing. Accurate information regarding the relationship between family history and colorectal cancer risk also furthers understanding of the genetic epidemiology of colorectal cancer. As discussed previously, various publications have addressed this issue (see meta-analysis by Johns et al., (134)). These have established the consensus that people with one or more affected relatives are at an increased risk of developing colorectal cancer, and provided some broad estimates for the magnitude of this risk.

Precise data relating to various family histories, and considering degree, relationship and age at onset of affected relatives, are comparatively scarce. Hence, there is a need for further studies in this area. Furthermore, it is possible that the degree of familial risk evident may vary with the population being considered, either due to genetic factors, or due to differences in exposure to environmental risk factors. Presently, data relating specifically to the Scottish population are not available.

Additionally, information regarding risk in people with a relative who has developed colorectal cancer at a very young age is scarce. Several studies have provided risk estimates for people with a relative affected at less than 45 years of age (39, 78, 97, 205, 256), and concluded that early onset of colorectal cancer in a relative is

associated with a higher degree of genetic risk than late or average onset of colorectal cancer in a relative. Study B offers the potential to provide unique information on risk for people with a relative affected at the extremely early age of 30 years or less. In Scotland, only 36 colorectal cancer cases in people under 30 years were diagnosed between 1986 and 1995, comprising just 0.24% of all colorectal cancer registrations in this period (102). Hence, the sub-group who exhibit such an extremely early onset are considered likely candidates for having an increased genetic risk.

Many estimates of relative risk associated with having a family history of colorectal cancer are based on information reported by cases and controls at interview or via questionnaire. In many studies, reported cases were confirmed using death certificates, medical records or cancer registry data, but only two previous case-control studies (256, 272) and two previous cohort studies (39, 92) have attempted to confirm negative reports of colorectal cancer in relatives. Reliance on reported data is a potentially important source of bias in risk analysis studies. As illustrated and discussed in the preceding sections, the use of record linkage provides a relatively reliable method of establishing family history. Consequently, the current family history case-control studies were designed to provide relatively unbiased and population-specific information regarding the risk associated with having a family history of colorectal cancer in Scotland.

3.5.2 Methods

The two studies considered in this chapter were analysed using the broad design and methods of a matched retrospective family history case-control analysis, as envisioned at the time the original study set was ascertained.

Cases and controls in study B were specifically matched at the time of ascertainment. Although controls in study A were originally matched to cases, details of these pairings were not available at the time of the current study. The latter were thus 're-matched', based on age and sex, and using a random generator to select pairs when more than one appropriate match existed.

The available family history data was analysed as a matched retrospective family history case-control study, using methods for calculating a mantel-haenszel estimator for the odds ratio (OR_{MN}) previously outlined by Liang et al., (167) and employed in similar study by St. John et al., (260). Importantly, this technique does not make any assumption of independence of cancer status for each relative, and thus allows for relatives who are part of the same family to be dependent in their cancer status. A matched analysis was conducted in order to maximise the power of the case-control study, since pooling relative data across case and control families is likely to underestimate the true odds ratio (167).

Each case subject, i, will have n_i relatives of which x_i will have developed colorectal cancer. The matched control subject, i, will have m_i relatives of which y_i will have developed colorectal cancer. This statistical notation is summarised in table 3.5:

Table 3.5 Statistical Notation for Retrospective Family History Case-Control Analysis

	Colorectal Cancer	No Colorectal Cancer	
Relative of Case	x_i	$(n_i - x_i)$	n_i
Relative of Control	y_i	$(m_i - y_i)$	m_i
	t_i	$(N_i - t_i)$	N_i

A series of 2 x 2 contingency tables were created for the matched pairs in each of study A and study B separately. The Mantel Haenszel estimator for the true common odds ratio is then given by:

$$OR_{MN} = \Sigma R_i / \Sigma S_i$$

Where: $R_i = x_i(m_i - y_i)/N_i$ and $S_i = y_i(n_i - x_i)/N_i$

An appropriate test statistic for the null hypothesis that $OR_{MN} = 1$, with an approximate chi-squared distribution and one degree of freedom, is calculated by:

$$T = \{\Sigma(x_i - n_i t_i/N_i)\}^2 / \{\Sigma(x_i - n_i t_i/N_i)^2\}$$

The approximate test-based standard error, SE, was calculated for $\log(OR_{MN})$, because the distribution of $\log(OR_{MN})$ is closer to the normal distribution than that of OR_{MN}. This value was calculated using the following equation:

$$T = (\log[OR_{MN}]/SE)^2$$

Thus the 95% confidence interval for the Mantel Haenszel estimator for the true common odds ratio is given by:

$$95\%\ CI\ for\ OR_{MN} = (e^{\log[OR_{MN}] - 1.96SE}, e^{\log[OR_{MN}] + 1.96SE})$$

Data were recorded in a Microsoft Excel™ spreadsheet, and all calculations were also carried out using this programme.

3.5.3 Results

In total, 133 case-control pairs were identified as part of study A. Case-control pairs were matched by sex, and comprised 60 female and 73 male pairs. The subset of cases included in this study (133 from a total of 199 in the complete data set) had a mean age of 65.3 years at ascertainment in 1993. Controls were also matched with cases to within 2-3 years of age, and hence had the similar mean age of 64.9 years.

The data set for study B included 100 matched pairs, comprising 46 female pairs and 54 male pairs. For study B, age was taken as age at ascertainment for surviving cases and controls, or the age they would have been at this point for deceased individuals. The case subjects (a subset of 100 from 113 in the complete data set) had a mean age of 37.0 years, and control subjects had a mean age of 37.1 years.

A summary of the overall number of relatives included in this study, and their colorectal cancer experience, is shown in table 3.6:

Table 3.6 Summary of Colorectal Cancer Cases Among Relatives

Study	Relative Group	Relatives of Cases/Controls	Number of relatives	Number of Affected relatives
A	FDR	Case	850	29
A	FDR	Control	1037	17
A	SDR	Case	1339	40
A	SDR	Control	1310	30
A	All	Case	2189	69
A	All	Control	2347	47
A	Parents	Case	266	15
A	Parents	Control	268	13
A	Siblings	Case	355	12
A	Siblings	Control	451	3
B	FDR	Case	507	8
B	FDR	Control	496	3
B	SDR	Case	885	25
B	SDR	Control	1004	24
B	All	Case	1392	33
B	All	Control	1500	27
B	Parents	Case	198	8
B	Parents	Control	200	3
B	Siblings	Case	240	0
B	Siblings	Control	222	0

Abbreviations: FDR, first degree relatives; SDR, second degree relatives

Analysis of the data set as a matched retrospective family history case-control study yielded the following results:

Table 3.7 Results of Matched Retrospective Family History Case-Control Analysis

Study	Relative Group	OR_{MN} (95% CI)
A	First Degree	2.14 (1.11, 4.14)*
A	Second Degree	1.45 (0.76, 2.75)
A	Parents	1.16 (0.53, 2.50)
A	Siblings	12.02 (2.11, 68.44)*
A	All Relatives	1.59 (1.05, 2.41)*
B	First Degree	2.02 (0.51, 8.00)
B	Second Degree	1.04 (0.60, 1.77)
B	Parents	2.67 (0.67, 10.67)
B	Siblings	N/A (no cases in siblings)
B	All Relatives	1.20 (0.69, 2.07)

*Significant at the 5% level ($p < 0.05$)

Differences in observed risk of colorectal cancer between relatives of cases, compared with relatives of controls, were only significant at the 95% level for first-degree relatives, siblings, and all relatives together for Study A. Significant differences were not observed for any relative group from study B. This is largely due to a lack of statistical power, resulting from the small numbers of colorectal cancer cases developing in each relative group. Combination of the two studies produced an estimate of the Mantel Haenszel odds ratio (OR_{MN}) for all first-degree relatives of 2.19 (95% CI = 1.22, 3.93).

For comparison, the data set was also analysed as an unmatched retrospective family history case-control study, using conventional methods for a case control study. In both studies cases outnumber controls (study A: 66 additional cases, study B: 13 additional cases). Therefore, whilst matched analysis is a more powerful technique, unmatched analysis has the advantage of including additional information.

Table 3.8 Results of Unmatched Retrospective Family History Case-Control Analysis

Study	Relative Group	Odds Ratio (95% CI)
A	First Degree	2.17 (1.24, 3.79)*
A	Second Degree	1.96 (1.23, 3.12)*
A	Parents	1.26 (0.65, 2.46)
A	Siblings	6.13 (1.83, 20.56)*
A	All	2.10 (1.48, 2.99)*
B	First Degree	4.09 (1.17, 14.34)*
B	Second Degree	1.54 (0.90, 2.64)
B	Parents	3.21 (0.87, 11.87)
B	Siblings	N/A (no cases in cases or controls)
B	All	1.80 (1.11, 2.93)*

**Significant at the 5% level ($p<0.05$)*

Results from the unmatched analysis are similar to those from the matched analysis, although confidence intervals are slightly narrower. As considered further in the discussion section, the unmatched analysis is not entirely statistically valid, as it makes the false assumption that all observations are independent. Hence the matched analysis constitutes the focus of the retrospective family history case control study considered in this section, with unmatched results included for comparison only.

3.5.4 Discussion

For study A, the results presented in table 3.7 for first-degree relatives are consistent with the current literature. In a comparable family history case control study, in which the cases were not selected on the basis of age, the Mantel Haenszel estimator of the odds ratio (OR_{MN}) was reported as 2.1 (95% CI = 1.4, 3.1) for parents and

siblings together (260). The estimate for Odds Ratio for FDRs of cases compared to controls in study A, is also comparable to the relative risk of having an affected first-degree relative of 2.25 (95% CI = 2.00, 2.53) reported in a systematic meta-analysis (134). Hence, this study has provided support for the observation that first-degree relatives of non age-selected colorectal cancer cases are at an approximately two-fold increased risk of developing the disease themselves, and confirmed that this observation applies to the Scottish population.

Table 3.7 also shows a marked difference between the odds ratios for parents of colorectal cancer cases and those of siblings. Some previous studies have provided some evidence to suggest that sibling risks are greater than parent/offspring risks (39, 92, 149). In contrast, no obvious difference between risks in siblings and parents of controls was observed by St John et al., who calculated odds ratios of 2.4 (95% CI = 1.4, 4.2) for parents of cases and 1.9 (95% CI = 1.1, 3.5) for siblings of cases (134). True differences between sibling risks and parent/offspring risks may arise due to the influence of environmental effects that are shared more closely between siblings, or may reflect the mode of inheritance, with recessive inheritance resulting in a greater risk to siblings being observed. Age cohort effects may also account for such differences. However, in this study, the total number of colorectal cancer cases developing in individual relative groups is very low, as reflected in the wide confidence intervals. This implies that the observed degree of disparity may well have arisen by chance.

A slightly increased odds ratio among second-degree relatives of colorectal cancer cases compared to controls for study A is also evident in table 3.7, although this was not statistically significant. Second-degree relatives share 25% of their genes, compared to the 50% of genes shared by first-degree relatives. Accordingly, the magnitude of increased genetic risk in a group with a second-degree relative with colorectal cancer would be expected to be approximately half that of a group with an affected first-degree relative. The results obtained in this study are consistent with this expectation, although again the number of cancer cases observed in this group are not high, and the confidence interval for the odds ratio estimate is wide, precluding any clear conclusions. Data on risk in second-degree relatives of colorectal cancer cases is generally scarce, and second-degree relatives were not considered in the study by St John et al., (260).

Study B considers the risk associated with having a relative diagnosed with colorectal cancer at an extremely early age. In general, a relatively early age at onset of colorectal cancer is associated with a hereditary origin. Several studies have estimated relative risk of colorectal cancer in relatives of cases diagnosed at < 45 years of age (39, 78, 97, 205, 260), and a meta-analysis of data from these studies suggests that the relative risk is 3.87 (95% CI = 2.40, 6.22)(134). There are currently no published estimates of relative risk for relatives of colorectal cancer cases diagnosed at <30 years. The prior hypothesis to this study was that relative risk in these groups would be considerably greater than risk in relatives of non-age selected cases (study A). Unfortunately, this hypothesis cannot be accepted or rejected based on the estimate presented in table 3.7, due to the small number of colorectal cancer

cases in study subjects, and the resulting lack of statistical power. Indeed, this study has not shown a statistically significant increase in risk for relatives of cases compared to relatives of controls. Combination of the two studies in this analysis does produce an estimate of relative risk in FDRs with a narrower confidence interval. However, these two studies cannot be considered directly comparable, due to differences in age and ascertainment methods of cases and controls.

The fact that relatively few colorectal cancer cases were observed in study B, compared with study A, is likely to reflect the fact that the average age of cases and controls was younger at the time of ascertainment. Hence relatives are also comparatively young in age and have contributed fewer years at risk to the study. The same effect may account for the relative lack of cases in siblings compared with parents, and in first-degree relatives compared with second-degree relatives.

Overall, the results of this study are severely limited due to the small number of colorectal cancer events occurring in relatives and consequent lack of statistical power. In the matched analysis of study A, only 34 case-control pairs were informative for colorectal cancer in FDRs, and in study B only 10 case-control pairs were informative for colorectal cancer in FDRs. Neither study has the data required for meaningful analyses by specific relative group, i.e. siblings or parents.
An unmatched analysis was conducted to determine whether this technique would provide any more statistical power, and also to establish, through comparison of results, whether or not the sub-set of cases included in the matched study were representative of all cases ascertained. The pattern of results from this unmatched

analysis is very similar to those from the matched study. Confidence intervals are generally slightly narrower, indicating that the unmatched analysis had slightly more statistical power. The unmatched results indicate that the elevated risk observed in FDRs in study B is statistically significant at the 5% level ($p < 0.05$), although the confidence intervals are still wide. For siblings in study A, the relative risk was much lower in the unmatched analysis and closer to the values reported in previous studies. This suggests that the extremely high relative risk observed in the matched analysis is an artefact resulting from the small sample size.

The unmatched study does not allow for the dependence of cancer status between related individuals. This could in theory be addressed using sophisticated statistical techniques such as Generalized Estimating Equations (164). However, it is unlikely that such analysis would significantly improve the statistical power of the study. Another approach that could be employed is Kaplan-Meier survival analysis, as used by St John et al., (260). Again, this represents an alternative method of analysis, but the limitations on power imposed by the small numbers of colorectal cancer events would remain. Conclusions from the analysis of this data set by different methods would therefore be unlikely to differ from the analysis presented in this section, and further analyses were considered inappropriate.

In conclusion, the results of this study are consistent with published literature, and support the prior hypothesis that people with an affected relative are at an increased genetic risk of developing colorectal cancer. However, the lack of statistical power of this study means that only limited conclusions can be drawn from these results.

3.6 A Prospective Study of Risk Associated with Family History of Colorectal Cancer

3.6.1 Rationale

Information on relatives of all cases and controls in studies A and B was ascertained and initially matched at ISD in 1993. The same data underwent record linkage for the second time at ISD in 2001. Hence there is up to eight years of follow-up on relatives identified as part of these studies, who were alive at the time of ascertainment. In theory, this data provides the opportunity to conduct a prospective cohort study of the various relative groups.

There is a paucity of prospective data on familial risk of colorectal cancer in the literature. Fuchs et al., (78) have conducted a large-scale prospective study, and estimated the age-adjusted relative risk of colorectal cancer for people with an affected FDR, as compared to those without such a history, to be 1.72 (95% CI = 1.34, 2.19). However, this study obtained information on family history from interview, and reports of colorectal cancer in relatives were not confirmed.

The aim of this study was to provide an estimate for relative risk associated with family history of colorectal cancer from a prospective study conducted in Scotland.

3.6.2 Methods

All first-degree relatives of case or control subjects, who were alive and colorectal cancer free at time of ascertainment, were eligible for the prospective study. From Study A this group comprised 687 relatives of cases, and 617 relatives of controls. 411 relatives of cases and 366 relatives of controls from Study B were included.

The data set was analysed as a prospective cohort study with 'family history' as the risk factor of interest. For the purposes of this study, first-degree relatives of cases were considered to have a family history, and all first-degree relatives of controls were considered not to have a family history. Standard notation for a cohort study was adopted, as shown in table 3.9:

Table 3.9 Statistical Notation for Prospective Risk Analysis

	Disease	No Disease	
Exposed to risk factor (relative of case)	a	b	
Not exposed to risk factor (relative of control)	c	d	
			n

Relative risk (RR) was calculated by the formula:

$$RR = a(c+d) / [c(a+b)]$$

A Chi-squared statistic was calculated by:

$$X^2 = n(|ad-bc|-n/2)^2 / (a+b)(c+d)(a+c)(b+d)$$

Confidence intervals were calculated by standard methods for cohort studies.

3.6.3 Results

The results of the prospective cohort analysis are presented in table 3.10.

Table 3.10 Results of Prospective Risk Analysis

Study	Relative Group*	No. of Relatives of Cases (Total Person Years at Risk)	No. of Relatives of Cases with CRC	No. of Relatives of Controls (Total Person Years at Risk)	No. of Relatives of Controls with CRC	Relative Risk (95% CI)
A	FDR	687 (5612)	12	617 (4699)	5	2.16 (0.64, 7.23)
B	FDR	411 (3140)	4	366 (2655)	0	N/A

*Abbreviations: FDR, first degree relative; CRC, colorectal cancer

In study A, a total of 12 FDRs of cases and 5 FDRs of controls developed colorectal cancer during the follow-up period. This number was insufficient to yield a statistically significant estimate of relative risk. Since no cases occurred in relatives of control subjects from study B, calculation of a relative risk was not possible.

3.6.4 Discussion

The prospective nature of this study, and the use of record linkage to determine the true cancer experience of study subjects over time, makes this a unique and highly valid approach to the challenge of informing risk prediction on the basis of family history of colorectal cancer. Unfortunately, the scale of the study was insufficient to allow conclusions to be made from the results shown in table 3.10. The estimated relative risk of having an affected FDR from study A is similar to the results of a meta-analysis of other published studies, which quoted a figure of 2.25 (95% CI =

2.20-2.65) (134). However, the confidence intervals pertaining to this study are very wide, and the observed results could have occurred by chance.

Clearly, a longer period of follow-up, and / or a larger data set is required to generate statistically valid results from a prospective study of risk associated with family history of colorectal cancer. At present, however, prospective analysis of this data set lacks the statistical power to generate meaningful results.

3.7 Discussion of General Methodology

3.7.1 Ascertainment of Study Subjects

In study A, both cases and controls were contacted and invited to participate in the study, and individuals who declined, or did not respond to, this invitation were excluded, and a replacement was found. This situation may have introduced some degree of ascertainment bias into the study, since people who opted to take part may be more likely to have experience of colorectal cancer or have a greater knowledge about the subject. Similarly, based on experience in other studies, non-responders may also have been relatively young and more likely to be male. However, the total proportion excluded on this basis was less than 20%, and this high participation rate should limit any bias that may exist.

Since all controls were identified from the general population, and ascertainment was not based on hospital attendance or other health-related criteria, the combined analysis of the two groups of controls is considered to be valid for the purposes of the analysis of prevalence of family history of colorectal cancer.

3.7.2 Record Linkage

Within the Scottish Record Linkage System, the false positive rate of records "matched" via patient-specific identifying information has been estimated as less than 1% (104, 122, 139), when based on the probability matching algorithm alone. Data regarding the false negative rate is not available, but false negative "matches" are likely to occur at a similarly low level. Hence, records that refer to different

people may be matched in error, and records that genuinely refer to the same person may not be matched. At a practical level, such errors in record linkage often arise due to an inability to distinguish people with similar or identical names, who share similar or identical dates of birth.

The actual frequency of errors in record linkage of study subjects with death records and the Scottish Cancer Registry is impossible to assess, since such occurrences will depend on the size of the data set under study and the number of colorectal cancer registrations in the registry, as well as the accuracy of the record linkage process itself. The protocol employed by ISD, whereby a threshold for accepting a match is based on a subjective assessment of the absolute probability that two records match, is designed to limit the number of false positive matches that are accepted for the purposes of analysis.

Another limitation of record linkage concerns the inability to trace relatives who have left Scotland. Information gathered at interview suggests that around 10% of relatives in this data set may have been untraceable, either due to emigration or limitations in the quality of available information. The principal effect of an inability to trace a proportion of relatives will be to underestimate cancer incidence amongst data subjects. However, it is unlikely that either the fact that some relatives could not be traced, or the possible occurrence of mismatched records, would introduce any systematic bias to these studies.

The record linkage process is wholly reliant on the provision of detailed personal information for matching with health records. Where information this is provided solely by interview, information may be more complete for first-degree relatives, leading to systematic bias in the probability of matching success. In this data set, such an effect should be limited by the intermediate step of using central records to extend and verify pedigree information.

Record linkage currently provides the only realistic method of 'confirming' a negative report of cancer, and alternative methods of confirming reported cancers in relatives compare unfavourably, being labour intensive and potentially incomplete (confirmation may not always be found). Overall, record linkage has been found to be a useful and efficient means of determining the cancer experience of data subjects, which is not biased by relying on information from the subjects themselves. It does not, however, constitute a "gold standard" in this context, as potential for error remains.

3.7.3 Identification of Colorectal Cancer Cases Through Death Records and the Scottish Cancer Registry

The accuracy and completeness of cancer registry data itself is a crucial consideration for any study that uses such a resource to validate or confirm cancer diagnoses, particularly when such data is used to inform clinical policy or epidemiological studies. Information on data quality of cancer registries is generally scarce, but several studies have recently been undertaken in Scotland to evaluate the reliability and validity of the Scottish Cancer Registry.

The Scottish Cancer Registry was initiated in 1958, and ascertainment is considered to be sub-optimal prior to 1968. Although ascertainment of any registry is unlikely to reach 100%, methods of ascertainment have steadily improved since this time, and the Scottish Cancer Registry is now considered to be reasonably complete in recent years and to compare favourably with other registries. By comparing cancer registry data with information re-abstracted from medical records, Brewster et al., found that 5.5% (95% CI = 2.5%, 8.6%) of colorectal cancer cases recorded in the registry did not warrant this diagnosis upon re-evaluation (26). This study related to incidence in the year 1990. A separate study examined fourteen sources of potential cancer registrations within the Tayside health board area, and concluded that the ascertainment rate of colorectal cancer in this area in 1992 was approximately 98.5% (27). Other estimates of data quality can be obtained by examining the proportion of cases registered only from death certificates, with a low value being indicative of good data quality. For the period 1986-1995, 2.8% of colorectal cancer cases were registered only from death certificates, and a trend towards an increasing proportion with increasing age was evident (102). A high proportion of microscopically verified registrations is also considered to imply a high degree of data validity. Over the same time period, 84.5% of colorectal cancer cases in the Scottish Cancer Registry were verified in this manner (102).

Hence, although some limitations of the Scottish Cancer Registry are illustrated by the studies considered above, the data recorded are considered to have a comparatively high degree of accuracy. This point is supported by the inclusion of

this registry in the publication "Cancer Incidence in Five Continents", which maintains high standards for the quality of data included (119).

Cancer diagnoses made prior to the establishment of an effective cancer registry were only identified for the purposes of these analyses if they were recorded on a death certificate. Hence, the ascertainment rate of colorectal cancer cases before 1968 is likely to be relatively low, and data provided in this manner may be less accurate in terms of site of cancer. With respect to the assessment of the accuracy of reporting of family history of colorectal cancer, this effect is unlikely to introduce systematic bias in terms of sensitivity of reporting, but may have resulted in a slight underestimation of positive predictive value. Also, where confirmation of cancer was by death certificate as opposed to ISD matching, the age at onset was taken to be the date of death. This may have resulted in an overestimation of age at onset for such cases, and conceivably led to an overestimation of family history where age of onset forms part of the family history criteria. A final consideration is that cancer cases that occur outwith Scotland are invariably not recorded in the Scottish Cancer Registry. Again, this factor may have resulted in a slight underestimation of the true colorectal cancer incidence in the data set.

3.8 Conclusion and Further Work

The studies described in this chapter represent a comprehensive analysis of a unique data resource. The key conclusions of the analyses presented and discussed in this chapter can be summarised as follows:

(i) People in Scotland with a history of colorectal cancer in a first-degree relative are at a significantly increased risk of developing the disease

(ii) The population prevalence of having a history of colorectal cancer in a first-degree relative is approximately 9.4% (95% CI = 4.9, 13.9)

(iii) Family history is substantially under-reported at interview by both colorectal cancer patients and healthy controls, and is often reported inaccurately

These conclusions are of considerable relevance from a clinical perspective, as well as in the context of epidemiological studies, since family history is used as part of clinical genetic services and in research as a means of identifying a sub-group of the population at increased genetic risk.

The findings presented in this chapter also illustrate the need for further research into the prevalence of family history of colorectal cancer, the meaning of family history at the individual and population level, and the methods by which family history is established. Repetition of the prevalence and the accuracy studies in a larger data set is ideally required to confirm the findings and to quantify the results more precisely. Similarly, the analyses of the risk associated with having a family

history of colorectal cancer presented here lack statistical power and a larger study, employing similar methods, is necessary to provide precise and reliable results.

The methods utilised in the prospective study facilitate re-analysis of this data set after a longer period of follow-up, since record linkage can be repeated, providing updated information on the colorectal cancer status of study subjects. In theory, it is also possible to stratify study subjects by extent of family history and estimate relative risk in each group. Furthermore, the general methodology for constructing the data set described in this chapter can theoretically be employed to provide data on absolute risk associated with having a family history of colorectal cancer. This could be achieved by identifying cohorts of individuals defined by degree of family history and performing record linkage on this data set before and after a defined period of follow-up. Age would be a vital consideration in such a study, since absolute risk is determined by a synthesis of age, environmental risk and familial (genetic) risk. Information on absolute risk of colorectal cancer associated with family history could be used directly to inform genetic risk assessment. However, such analysis is not feasible for the current data set, given the prohibitively small numbers of colorectal cancer cases occurring in the study subjects.

In conclusion, the current data set has been thoroughly analysed and explored, yielding results that are highly relevant to the challenge of preventing colorectal cancer in Scotland through targeting people at increased genetic risk.

Chapter 4

The Development of a Computer Model of Cascade Genetic Testing

4.1 INTRODUCTION

4.1.1 Rationale for Identifying Mismatch Repair Gene Mutation Carriers

There is now compelling evidence to support a role for mismatch repair gene mutations in a small but important subset of colorectal cancer cases. Individuals who harbour such mutations are at a very high risk of developing colorectal cancer, with penetrance being approximately 80% in males and 40% in females (4, 57, 170, 286, 289). Furthermore, colorectal cancer cases that are associated with mismatch repair gene mutations are likely to occur at a relatively early age. Mismatch repair gene mutation carriers thus constitute an important sub-group of the population who may be targeted as part of strategies to prevent colorectal cancer.

Information regarding genetic susceptibility to a particular condition is of limited practical use in the absence of effective intervention strategies. As discussed previously, clinical screening can potentially reduce the incidence and mortality of colorectal cancer in people at high genetic risk of this disease. Genetic testing for mismatch repair gene mutations provides one strategy for identifying such people, and, in contrast to empirical risk assessment on the basis of family history, has the important ability to elucidate individual risks within families that harbour a mutation. These observations provide a compelling rationale for identifying mismatch repair gene mutation carriers and offering clinical screening.

In the Scottish Population, the prevalence of pathogenic MMR gene mutations in the 15-74 years age group has been estimated as 1:3139 (95% CI = 1:1247, 1:7626) (58). Extrapolation of this estimate to the Scottish population (2001 census) implies that the number of Scots in this age range who have such mutations is 1209 (95% CI = 498, 3044). Theoretically, identifying these people at an asymptomatic stage, and providing appropriate surveillance, would have significant benefits for this sub-group of the population in terms of colorectal cancer prevention. However, further evidence of effectiveness, and careful consideration of potential drawbacks, are required to inform decisions regarding the appropriateness of this approach.

Mismatch repair gene mutation carriers in Scotland are identified either through clinical cancer genetic services, or as part of research programmes. This situation is common to most developed countries, and often the approach to the ascertainment of carriers for research and clinical purposes is integrated. However, mutation analysis is often targeted to colorectal cancer patients, particularly in the research context, and, whilst this enhances understanding of the genetic basis of colorectal cancer, it is of limited clinical benefit to the individual concerned.

At the present time, asymptomatic individuals in Scotland who meet empirical criteria for "high" genetic risk of colorectal cancer, as defined by guidelines published by the Scottish Cancer Group (249), are eligible for genetic testing. However, only a proportion of asymptomatic MMR gene mutation carriers will meet these criteria, and the identification and testing of those that do is conditional on

them being aware of their family history and sufficiently concerned to seek medical advice.

Thus, ongoing research is designed to enhance understanding of the genetic epidemiology of colorectal cancer, and clinical genetic services offer genetic counselling and testing, where appropriate, to individuals who seek medical help regarding a high empirical risk of developing colorectal cancer. However, from a wider public health perspective, a more co-ordinated and systematic strategy is required to reduce the burden of colorectal cancer resulting from MMR gene mutations. The goal of such a strategy must be the identification of all MMR gene mutation carriers at an asymptomatic stage, and the provision of appropriate counselling and clinical screening for these people.

Numerous considerations surround the issue of who should be offered genetic testing. Simply offering a genetic test may lead to increased anxiety in the test subject, and this may be further compounded by the fact that a significant proportion of tests are inconclusive. Additionally, conducting complete mutation analysis to determine the MMR gene mutation status of an individual is both expensive and time-consuming. Hence, the capacity to minimise the number of people offered genetic testing in general, and complete mutation analysis in particular, is a critical feature of an appropriate strategy for identifying mismatch repair gene mutation carriers at the population level.

4.1.2 Strategies for Identifying MMR Gene Mutation Carriers

There are several theoretical options for attempting to identify asymptomatic MMR gene mutation carriers from within the Scottish population.

4.1.2.1 Population Genetic Testing

Population genetic testing, in which mutation analysis is conducted in every member of a population, is in theory the most comprehensive strategy for identifying asymptomatic mutation carriers at the population level. There are two practical strategies that could be employed to search for mutations in the context of population genetic testing; either conducting complete mutation analysis of the genes in question, or looking for specific mutations that are known or suspected to occur in the population. The latter option is far less expensive and labour-intensive and could be of particular benefit in countries where specific "founder" mutations are prevalent. However, the extreme heterogeneity of mismatch repair gene variants and the low allele frequency of individual mutations create considerable practical barriers to testing for specific MMR gene mutations at the population level. Published information and data from ongoing research in Scotland suggest that numerous different MMR gene mutations are present in the Scottish population, many of which will not previously have been identified. Conducting comprehensive population genetic testing based on identifying specific mutations would therefore necessitate performing numerous pre-symptomatic tests, and would not be capable of identifying novel mutations. At present, therefore, non-specific mutation analysis

represents the most appropriate option for population genetic testing in this context, despite the fact that mutation analysis is relatively time-consuming and expensive.

Regardless of the strategy used, conducting population-wide genetic screening for a very rare set of mutations, such as MMR mutations, is difficult to justify on practical and ethical grounds, since a large number of people would undergo genetic testing and only a small proportion would actually benefit. Genetic testing may also be unacceptable to a proportion of the public, resulting in incomplete participation. These factors, coupled with the practical difficulties outlined above and the current limitations in knowledge regarding the association between MMR gene mutations and colorectal cancer, mean that conventional population testing is unlikely to be recommended in the near future. Indeed, a recent UK National Screening Committee workshop concluded that there is currently no case to offer population screening in an attempt to identify MMR gene mutation carriers (237). Authors in the United States have reached similar conclusions, agreeing that more information regarding the prevalence and penetrance of mismatch repair gene mutations and more evidence of effective intervention strategies are essential prerequisites for implementing population genetic testing (30, 31, 44, 221, 271, 301).

4.1.2.2 Stratified Population Genetic Testing

Stratified population genetic testing, in which genetic testing is offered only to individuals with a certain extent of family history of colorectal cancer provides another option for identifying asymptomatic mismatch repair gene mutation carriers.

However, as discussed previously and highlighted in chapter 3, there are considerable practical difficulties in obtaining accurate family history information, and attempting to do so for the entire population may be unfeasible. Recording and assessment of family history in the primary care setting could potentially circumvent this problem in the future, but implementing such a policy is problematic in itself and at present family history information is not routinely gathered.

Additionally, family history has limited sensitivity and specificity as a tool for identifying MMR gene mutation carriers (199), and stratified population genetic testing may have the additional drawback of raising anxiety amongst people with a family history of colorectal cancer by targeting them specifically. The targeting of people with a particularly strong family history as part of a stratified genetic testing programme may be more acceptable in these respects, since relatively few people would be offered genetic testing and the probability of those people harbouring a mismatch repair gene mutation is relatively high. However, there are two major disadvantages to such an approach. Firstly, application of any family history criteria will result in a proportion of MMR gene mutation carriers, i.e. those who do not meet these criteria, remaining unidentified. In addition, if only the minority of people with a very strong family history of colorectal cancer are to be included in a stratified genetic testing system, it becomes difficult to justify the implementation of such a programme at the population level. Considerable overlap would also exist between highly stratified population genetic testing and clinical genetic services. Hence, whilst family history may be useful as a means of targeting potential MMR

gene mutation carriers in a clinical context, stratified genetic screening at the population level is unlikely to prove feasible or desirable at present.

4.1.2.3 Cascade Genetic Testing

Cascade genetic testing initially involves conducting mutation analysis in individuals who are at a relatively high risk of carrying a mutation. Often, people with a family history, people who have developed the disease themselves, or a sub-group of such patients, are targeted for this purpose. Once a mutation carrier has been identified, genetic testing can be offered to their relatives, and the specific mutation present can thus be traced through an expanded pedigree in a "cascade" fashion. The genetic proximity of the relatives offered testing is an important factor in determining the effectiveness and efficiency of cascade genetic testing, and is generally referred to in terms of 'depth'. For example, restricting testing to first-degree relatives of known carriers is highly efficient, since each individual tested will have a 50% probability of carrying the same mutation. However, this approach may miss carriers in more distant relatives. The ideal cascade genetic testing system would test all first-degree relatives of an index case, and subsequently test first-degree relatives of those found to have a mutation. In practice, such a system will be restricted, due to the fact that some first-degree relatives will be deceased, and some will decline to participate.

The concept of tracing a pathogenic mutation through families is a well-established aspect of clinical genetic services. Irrespective of how an individual mutation carrier

is identified, it is possible, and in many instances desirable, to test relatives of that carrier for the specific mutation they have been found to carry. Hence, one of the major advantages of cascade genetic testing as a means of identifying asymptomatic MMR gene mutation carriers is the potential of this approach to complement current strategies pursued by genetic services. Active cascade genetic testing to identify MMR gene mutation carriers is essentially a systematic and comprehensive continuation of the way in which current clinical practice has evolved.

Another major advantage of active cascade genetic testing over other strategies for identifying asymptomatic carriers of pathogenic mutations is its potential efficiency. By limiting genetic testing to either people deemed on clinical grounds to be at high risk of carrying a mutation, or close relatives of known carriers, it is theoretically possible to restrict the number of non-carriers who undergo testing while maximizing the number of carriers identified. This not only reduces the resources required to identify mutation carriers, but also reduces the number of asymptomatic non-carriers offered genetic testing.

Cascade genetic testing has a particular advantage in circumstances where pathogenic mutations in the gene(s) in question are heterogeneous and/or previously unknown pathogenic mutations may exist, as is the case with MMR genes. In this situation, identification of a pathogenic mutation usually requires complete mutation analysis of the entire gene(s), a procedure that is relatively time-consuming and expensive. At current NHS rates, mutation analysis of hMLH1 and hMSH2 costs

around £750 (Dr. M. Porteous, personal communication). Application of population or stratified genetic testing involves conducting such mutation analysis in each participant. In contrast, the design of cascade genetic testing is such that, once mutation analysis reveals a pathogenic mutation, subsequent genetic testing of relatives of the identified carrier involves only a simple test for that specific mutation concerned. Accordingly, such a test is relatively quick to perform, and compares favourably with mutation analysis in terms of cost, with each test for a known mutation costing £50 to £100 (Dr. M. Porteous, personal communication).

Throughout this thesis, the term "mutation analysis" or "non-specific test" refers to conducting complete analysis of the mismatch repair genes hMLH1 and hMSH2 in an attempt to detect any possible mutation in a colorectal cancer case. The term "pre-symptomatic test" or "specific test" refers to the process of testing a relative of a known carrier for the specific mutation known to occur in that carrier.

Mutation analysis of MMR genes may be inconclusive or incomplete in the sense that not all mismatch repair genes are considered, and results are subject to interpretation as considered in chapter 3. Hence a negative result at this stage cannot be equated with a lack of increased genetic risk, and, if the appropriate family history criteria are met, surveillance of the individual concerned must be maintained. In contrast, the low population prevalence of MMR gene mutations means that a relative offered a pre-symptomatic test for a pathogenic mutation known to occur in their family, and found to be negative, can be considered to be at no more than the

average population risk for colorectal cancer. This is true regardless of their family history of the condition, provided that the specific mutation in question segregates with disease in the family. In this manner, cascade genetic testing permits the identification of asymptomatic carriers of pathogenic mutations without the need for complete mutation analysis in each individual tested, and provides a conclusive test result for relatives of a known carrier.

A theoretical application of cascade genetic testing has been considered in a paper by Krawczak et al., (150). This publication provided a series of equations designed to estimate the theoretical efficacy and efficiency of cascade genetic testing for various hypothetical mutations, including autosomal dominant mutations predisposing to complex disease. MMR genes and colorectal cancer were not specifically considered. The authors calculated that a cascade genetic testing strategy involving testing of all children, siblings, grandchildren, nieces, nephews and first cousins of a mutation carrier would identify 63.4% of mutation carriers in the population concerned, assuming penetrance of this mutation to be 50% (150). It was further estimated that such a strategy would necessitate testing 3.5 non-carriers within a pedigree for each carrier identified (150). The proportion of mutations in the population that could be identified, and the efficiency of the process, varied with screening depth and penetrance. The results produced by the equations in this paper prompted the conclusion that cascade genetic testing may provide a "viable" approach to identifying asymptomatic carriers of mutations that predispose to complex disease (150).

Whilst Krawczak et al., provide a theoretical demonstration of the potential utility of cascade genetic testing, the equations used are insufficient to permit a comprehensive assessment of this approach in a specific clinical context such as MMR gene mutations and their association with colorectal cancer. The outcomes of any cascade genetic testing programme will depend on numerous factors, including the genetic epidemiology of the condition in question, the available technology for genetic testing and the uptake of genetic testing by at-risk individuals. These factors will be superimposed on the background of population demographics, diverse family structure, the age/sex distribution of cases and existing strategies for colorectal cancer prevention in people at increased genetic risk. Only a small proportion of these factors are considered by Krawczak et al. Furthermore, the equations presented in this paper make the key assumptions that participation will be complete, and that penetrance is the same in both males and female carriers.

Another crucial consideration that is not addressed by the equations provided by Krawczak et al., concerns the timescale of the cascade genetic testing programme. In reality, it will take time to recruit potential mutation carriers, perform mutation analysis, trace relatives and so on. As time progresses, the population will age and new cases of disease will arise. The timescale of a genetic testing programme is also an essential consideration in terms of planning and resource allocation. Hence, an evaluation based on the instantaneous completion of a cascade genetic testing programme is of limited practical use.

Several authors have taken a more clinically-focused approach to investigating the potential of cascade genetic testing as a means of identifying carriers of mutations predisposing to various diseases (110, 265, 266, 280, 307). Super et al., have outlined the case for cascade genetic testing to identify carriers of mutations that cause cystic fibrosis (265), and have applied this strategy in a clinical setting (266). Results of this practical application suggested that cascade genetic testing was both effective and efficient in this context, and prompted the authors to recommend widespread application of this approach (266). The utility of cascade genetic testing in cystic fibrosis has also been formally considered by Holloway and Brock (110), who concluded that cascade genetic testing was useful in individual families, but would be less appropriate at the population level in comparison with offering genetic testing to all couples (a form of population genetic testing).

Cystic fibrosis is a recessive condition that is present from birth, and the combined allele frequency of pathogenic mutations is in the order of 0.02. These factors contrast with the extreme rarity and autosomal dominant nature of mismatch repair gene mutations. Hence, whilst the potential utility of cascade genetic testing as a means of identifying mutation carriers can be illustrated using cystic fibrosis, it is not possible to extrapolate conclusions to MMR genes and colorectal cancer. This is particularly true of information regarding participation rates and the psychosocial effects of cascade genetic testing, since the issues faced by the individuals being offered testing are very different in these two situations.

Cascade genetic testing for Familial hypercholesterolaemia has been explored in the Netherlands as part of an ongoing project (280). This condition is characterized by mutations in the low-density lipoprotein receptor gene, which cause elevated blood cholesterol and thus predispose to atherosclerosis and vascular disease. Cascade genetic testing beginning with 237 mutation carriers identified 2039 additional carriers from a total of 5442 relatives (280). The protocol employed involved testing first-degree relatives of known carriers, with second-degree relatives being tested where testing of first degree relatives was not possible. The overall participation rate in this study was 90% (280).

In many respects, the low-density lipoprotein receptor gene and its role in familial hypercholesterolaemia is comparable with MMR gene mutations and their association with colorectal cancer. The condition is inherited in an autosomal dominant manner, and over 300 different mutations in the same gene have been found to be pathogenic. A small number of these mutations occur with a relatively high frequency, whereas the majority have only been identified in one family (73). Furthermore, the condition is rare, affecting 1:500 people in the population of the Netherlands, and it influences the probability of developing particular complex diseases in adult life. However, there are also important differences between these two situations. Familial hypercholesterolaemia can be effectively treated by a combination of lifestyle modification and the administration of drugs to lower cholesterol, whereas intervention for MMR gene mutation carriers involves regular invasive surveillance. This difference may affect participation rates, since the necessary interventions entailed by a diagnosis of familial hypercholesterolaemia

may be comparatively acceptable to potential participants. There are also practical differences associated with identifying carriers of mutations in MMR genes and the low-density lipoprotein receptor gene. While mutations in the latter gene are heterogeneous, it has been shown that in the Netherlands nine specific mutations account for 66.5% of familial hypercholesterolaemia, and that some mutations occur in specific geographical areas (73). The fact that specific mutations can be identified in the majority of low density lipoprotein receptor gene mutation carriers largely precludes the need for comprehensive mutation analysis.

Cascade genetic testing for mismatch repair gene mutations has not been conducted and formally evaluated on a population-wide scale, but an ongoing large-scale study into the genetic basis of colorectal cancer in Scotland, entitled the Colorectal Cancer Genetic Susceptibility (COGS) programme, can potentially provide some data relevant to such an evaluation. The COGS programme was initiated in 2000, and has been funded by two Cancer Research UK programme grants through to 2008. The Chief Scientist Office (CSO) has provided essential co-funding support to the COGS programme through project grants. The main aim of COGS is to conduct genetic testing for mismatch repair gene mutations in all Scottish colorectal cancer patients diagnosed under the age of 55 years and to characterise the impact of these mutations in terms of penetrance, survival and interactions. Another objective is the detection of novel genetic variants that cause colorectal cancer. Thus far, 26 definite mismatch repair gene mutation carriers have been identified.

The COGS programme is not specifically designed to conduct or evaluate cascade genetic testing. However, funding and ethical approval is in place for the recruitment and genetic testing of relatives of identified carriers, in order to better characterise the pathogenicity of mismatch repair gene mutations. Consequently, a considerable body of data that are relevant to the evaluation of cascade genetic testing are currently becoming available through the COGS programme.

Overall, there is some information from ongoing research programmes, published literature and expert recommendations to suggest that cascade genetic testing may be a feasible means of identifying asymptomatic carriers of mutations in MMR genes. However, a formal evaluation of cascade genetic testing in this context has not been undertaken, and there is no direct evidence of the effectiveness of this approach. Information regarding the outcomes of cascade genetic testing for mismatch repair gene mutations, particularly with respect to the number of mutation carriers that will be identified and the relative efficiency of this process, is not currently available. Such information is required to inform the planning of cascade genetic testing programmes, either for research purposes or as part of public health policy.

4.1.3 Aims and Objectives

The overall aim of the research presented in this chapter is to conduct an evaluation of cascade genetic testing for mismatch repair gene mutations through the development of a computer model of this process that runs on a chronological basis, starting at a particular point in time rather than with an initiating event. The

population demographics and genetic epidemiology of colorectal cancer at this point will be defined. A hypothetical cascade genetic testing programme, with adjustable parameters and scope defined by the user of the model, will then be applied and the outcomes of this system will be studied.

The vast majority of pathogenic mismatch repair gene mutations have been reported in hMLH1 and hMSH2, and data relating to other genes is comparatively scarce. Accordingly, the computer model considered in this chapter focuses on these two genes, and in the context of this model the term "mismatch repair genes" refers to hMLH1 and hMSH2.

The specific objectives of the computer model were to:

(i) Conduct a comprehensive "task analysis", outlining the required nature and scope of a computer model capable of evaluating a hypothetical cascade genetic testing programme for MMR gene mutations

(ii) Develop a computer model of cascade genetic testing as it applies to MMR gene mutations

(iii) Determine suitable quantitative 'default' estimates as inputs for this model, identified through the available literature and the use of 'real data' from the COGS study

(iv) Utilise this model to estimate the number of MMR gene carriers that would be identified over time under an active cascade genetic testing program in Scotland

(v) Estimate the efficiency, according to the computer model, with which MMR gene carriers could be identified by cascade genetic testing, in terms of the number of genetic tests required to achieve this

(vi) Investigate the sensitivity of key model outcomes to variations of input parameters on a realistic scale

(vii) Critically interpret the outcomes from the computer model and consider the wider implications of findings in the context of evaluating cascade genetic testing as a strategy for identifying asymptomatic MMR gene carriers in the Scottish population.

Defining the aims and scope of a proposed computer model in detail is an integral part of to model development. This process is referred to as the "task analysis". This is, essentially, a document designed to answer two crucial questions:

(i) What exactly is going to be modelled?

(ii) What are the requirements of the completed model?

A simplified version of the task analysis document is presented in the results section.

4.2.1 Strategy for Model Development

Computer modelling provides a valuable means to represent complex systems. Increasingly, this approach has been used in the field of biomedical research and health service provision. In some circumstances, biomedical researchers will have developed sufficient expertise to enable a computer model to be developed entirely by one individual, or within one research group. More commonly, computer modelling is used as one of many research tools by researchers whose area of expertise resides in the field to which the model relates, not in the actual process of computer modelling. Conversely, an expert in computer modelling is unlikely to have sufficient knowledge and understanding of the scientific research in question to enable them to create a meaningful model. In such circumstances, model development will be the result of collaboration between at least two people. A "domain expert", will provide the knowledge and expertise to understand and define the system to be modelled, and a computer programmer will undertake the technical task of writing the computer programme that will form a working model.

Throughout the development of this computer model, the author of this thesis assumed the role of "domain expert", and a computer programmer was employed on an ad-hoc basis to implement the model. Reconciling these different perspectives and facilitating productive communication between domain experts and computer

programmers is one of the major challenges in the task of modelling biological and medical systems.

The strategy chosen to overcome this challenge was to use a five-stage process of model development, broadly encompassing the following phases:

(i) Inception

(ii) Conception

(iii) Implementation

(iv) Evaluation

(v) Interpretation

In this approach, the design of the computer model comes from the domain expert. The computer programmer then implements the model according to pre-defined specifications and requirements, working in close collaboration with the domain expert. Finally, the domain expert evaluates the functions of the model and uses the model to generate relevant data for analysis and interpretation. Thus, the domain expert retains ultimate control of the process of model development.

The inception phase involved the production of a task analysis document. This was conducted in a systematic manner, with the aim of defining and recording the

precise requirements and scope of the model, and identifying sources of data to inform model parameters. In many ways, this is the most important aspect of model development, since it determines the capacity of the final computer model. The next stage in model development was the creation of a "conceptual model", which is a representation of the model from the perspective of the domain expert. To facilitate the creation of a conceptual model it was necessary for the domain expert to attain a thorough understanding of the system to be modelled, the factors that influence it, and the issues that may be encountered. This was the most creative stage, in which the broad concept of 'creating a model of cascade genetic testing' was given detail and definition. At this stage the quantitative estimates that provide inputs to the model were identified. The completed conceptual model facilitated communication of the model system to a computer programmer, who used the conceptual framework thus provided as the basis for creating a functional model. In turn, the functional model formed the basis for 'coding' the software that constitutes a working version of the model. Subsequently, the model was 'run' at various settings to generate the data that forms the basis of an evaluation of cascade genetic testing.

A crucial feature of this approach is the iterative and incremental fashion in which the computer model was developed. Both the conceptual and functional models were constantly revisited during development to enable unforeseen hurdles to be overcome, and to incorporate new ideas.

Details of the methods used at each stage of model development are considered below.

4.2.2 Inception (Task Analysis)

The task analysis was informed by published literature on the subject of the epidemiology of colorectal cancer, the genetic epidemiology of mismatch repair gene mutations, and the concept and practice of cascade genetic testing. Additional input at both the task analysis and conceptual model stage was sought from experts with a variety of relevant perspectives, including clinical cancer genetics, genetic epidemiology and colorectal surgery.

4.2.3 Development of Conceptual Model

The most effective and systematic manner in which to represent a conceptual model is to use a 'visual modelling' system that provides a structured way of representing systems organized around concepts that exist in the real world and can therefore be understood with comparative ease. The conceptual model of cascade genetic testing was developed and represented using an object-orientated visual modelling design. Object-orientated design is a powerful technique for creating useful models of complex systems (224). Specifically, the conceptual model was created using Rational Rose software (Rational™), which employs the Unified Modelling Language (UML) to create detailed visual models. The Unified Modelling Language is widely used, and has previously been successfully utilised for modelling biological systems (182).

A major advantage of the above approach is that the UML can be used both to create a conceptual model, and to represent a functional model. This dual use ensures optimal communication between domain expert and computer programmer, and facilitates a smooth transition from conceptual model to functional model.

Development of a conceptual computer model took a bottom-up approach, starting with the various parts of the cascade genetic testing system, and trying to understand and represent how the properties of the system arise from the interaction of its components. In object-orientated design, objects are used to represent entities, including their state and behaviour, in a real-world system. Objects that share common properties, behaviour, semantics and relationships can be grouped together into classes. This reflects an important concept of object-orientated modelling, whereby such characteristics are detailed only once in the model, and are shared by each object that uses them. In the context of modelling cascade genetic testing, an example of an object may be a member of the population, referred to in the conceptual model as a "Scot". Each Scot will have similar properties, or "attributes", and similar interactions with other objects in the cascade genetic testing system. Hence a "class" that includes the shared characteristics of Scots can be created.

The methods employed in this project were ideally suited to the production of a conceptual model by a domain expert, since they employ the language of the domain, rather than the terminology of computer programming. They provided a rigid structure to the conceptual model, which enabled it to be expressed and

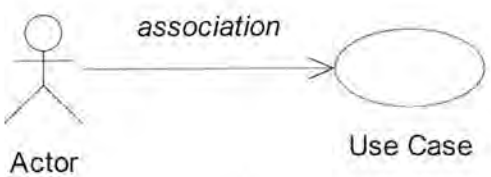
understood in the systematic way required by a computer programmer, and indeed by a computer programme, in the implementation phase.

Using the UML, the conceptual model can be represented in various different ways, using diagrams that view the model from distinct but related perspectives. The types of diagrams used, and their basic UML notation, are outlined below:

(i) Use Case Diagrams

Use case diagrams are used to document the behaviour of the system, in terms of its intended functions (use cases), people who interact with the system (actors) and the relationships between these. These diagrams form the starting point of the conceptual model, and communicate the functional objectives of the model.

Figure 4.2.1 Use Case Diagram Notation

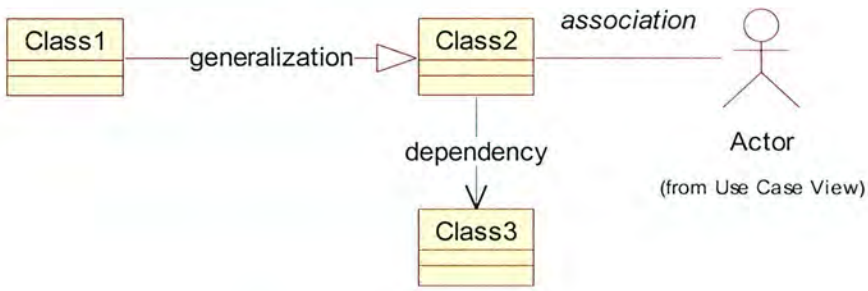


An example of a use case diagram is given by figure 4.3.3 (page 182).

(ii) Class Diagrams

Class diagrams illustrate a sub-set of classes that are included in the conceptual model. They often relate to use cases, and include each class involved in a particular use case. Classes and class diagrams contain much of the information that provides the framework for creating a functional model.

Figure 4.2.2 Class Diagram Notation

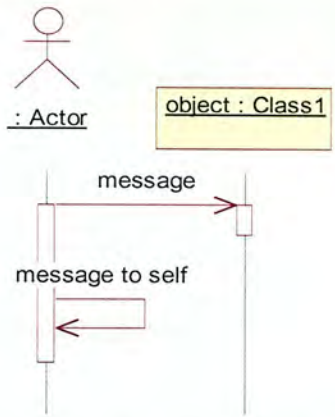


An example of a class diagram is given by figure 4.3.7 (page 187).

(iii) Sequence Diagrams

A sequence diagram displays the chronology of interactions between objects during a particular scenario. These diagrams invariably correspond to a particular use case, and document the sequence of interactions and communications between objects that are required to complete the use case in question.

Figure 4.2.3 Sequence Diagram Notation

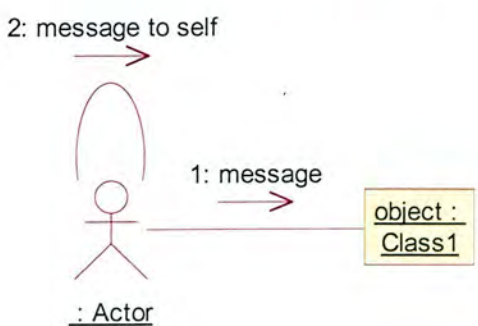


An example of a sequence diagram is given by figure 4.3.4 (page 183).

(iv) Collaboration Diagrams

Collaboration diagrams essentially provide an alternative view of a scenario described by a sequence diagram. Both sequence and collaboration diagrams may be referred to as scenario diagrams. The key difference is that a collaboration diagram illustrates object interactions arranged around the objects in question, and their links to one another. Hence they are able to represent the scenario in broader terms.

Figure 4.2.4 Collaboration Diagram Notation



An example of a collaboration diagram is given by figure 4.3.5 (page 184).

The conceptual model was represented using a combination of the diagrams outlined above, with associated documentation. Further details of the model were communicated verbally during discussions between the domain expert and the computer programmer.

4.2.4 Input Estimates

The validity of any computer model is crucially dependent on the accuracy and quality of the data used to provide inputs and set parameters for the model system. Resources identified at the task analysis stage; principally published literature, data from the COGS study, and expert opinion, were used to define the estimates required as inputs to the model. In accordance with recognised principles of good practice in model development (300), all assumptions and estimates built into the model were logged, and presented in as transparent a fashion as possible. The estimates identified and used in this computer model are presented in the results section. The number of expected colorectal cancer cases in the population, and the distribution of these cases by age, sex and mutation status was calculated using formulae entered into a Microsoft Excel™ spreadsheet.

4.2.5 Development of Functional Model

The methods employed were designed to ensure a logical transition from conceptual model, which represents the domain expert's view of the cascade genetic testing system, to functional model, which represents the same system from the perspective of the functions of the model. In essence, the conceptual model was designed to

describe the real-life cascade genetic testing system, whereas the functional model provided a blueprint for writing the computer programme ("coding") that constitutes the working model software.

At this stage, communication between domain expert and computer programmer was essential. This was primarily achieved through the conceptual model itself and the associated documentation that describes in detail what the objects, classes and diagrams represent. The use of the UML for both the conceptual and functional models ensured rapid and efficient development, since many of the objects and classes already identified as part of the conceptual model were directly utilized or adapted for use in the functional model.

4.2.6 Coding of Functional Model

The completed functional model provided the framework for a working version of the model system. It is, however, still represented as a visual model, which needs to be translated into a functional computer language that is capable of conducting calculations, reading spreadsheets, and presenting outputs. This process is referred to as "coding". Java programming language was used in this phase of model implementation. The software used was Sun ONE Studio 4 Update 1 (Sun™) (133).

4.2.7 Evaluation of Model Function

The software with which the model was developed has in-built debugging facilities to ensure that the consistency of the code is maintained. Thus, technical and typing errors in the code are brought to the attention of the computer programmer and subsequently resolved. Scope for human error in model development was minimized by review of the model throughout development, and the completed model was tested to ensure that outputs were realistic and responded logically to alterations in model parameters and in the code itself.

4.2.8 Analysis of Model

Default settings for model inputs were defined by analysing and interpreting available information (see results section), and the influence of various input parameters was studied by incrementally altering the value of interest and examining the results. In order to make the model as realistic as possible, a random element was incorporated into the calculation of probabilities. This is discussed in more detail in the results section, but the general process was as follows. A probability is defined, and the model generates a random value between 0 and 1. If this random value is less than the defined probability, the event (e.g. test acceptance) will take place; if it is greater than the defined probability it will not (e.g. test refusal). This procedure was employed at numerous points in the model, ensuring that a range of possible outcomes exist for each specified set of input values.

The operator of the working model must initiate each 'run' of the working computer model, and manually process the output data to facilitate analysis. Hence, an acceptable balance between the time-consuming process of repeatedly running the model and the statistical power required of the observed results was required. Accordingly, the model was run ten times for each set of input values considered in this project in order to standardize the recording and interpretation of results. This strategy for generating model outcomes is considered further in the discussion section.

Data from each run of the model was saved in an excel spreadsheet (Microsoft Excel™). Mean values, and the associated standard deviation and confidence intervals, were calculated by manually constructing mathematical functions within this software, using standard statistical methods. Where the estimate of interest was calculated from other estimates (e.g. ratio between two values), this was done separately for each run, and an overall mean was calculated.

Cascade genetic testing is a long-term strategy for identifying asymptomatic mutation carriers, and therefore results over a long period of time are of interest. However, it is not feasible to assume that conditions defined by the model will remain constant over very long periods of time. To reflect a balance between these two considerations, results were generated and analysed over a hypothetical time scale of twenty years. This issue is considered further in the discussion section.

The computer model generates a wide range of outcomes values. Two key output values were used to facilitate a broad comparison between results under different input conditions. These were "yield", measured as the total number of asymptomatic mismatch repair gene mutation carriers identified over time, and "relative efficiency", measured in a simplified manner by dividing the number of asymptomatic mismatch repair gene mutation carriers identified by the number of mutation analyses conducted. Mutation analyses are the most expensive and time consuming aspects of the cascade genetic testing system, and the ratio of mutation analyses performed to carriers identified thus provides a useful indicator of overall efficiency.

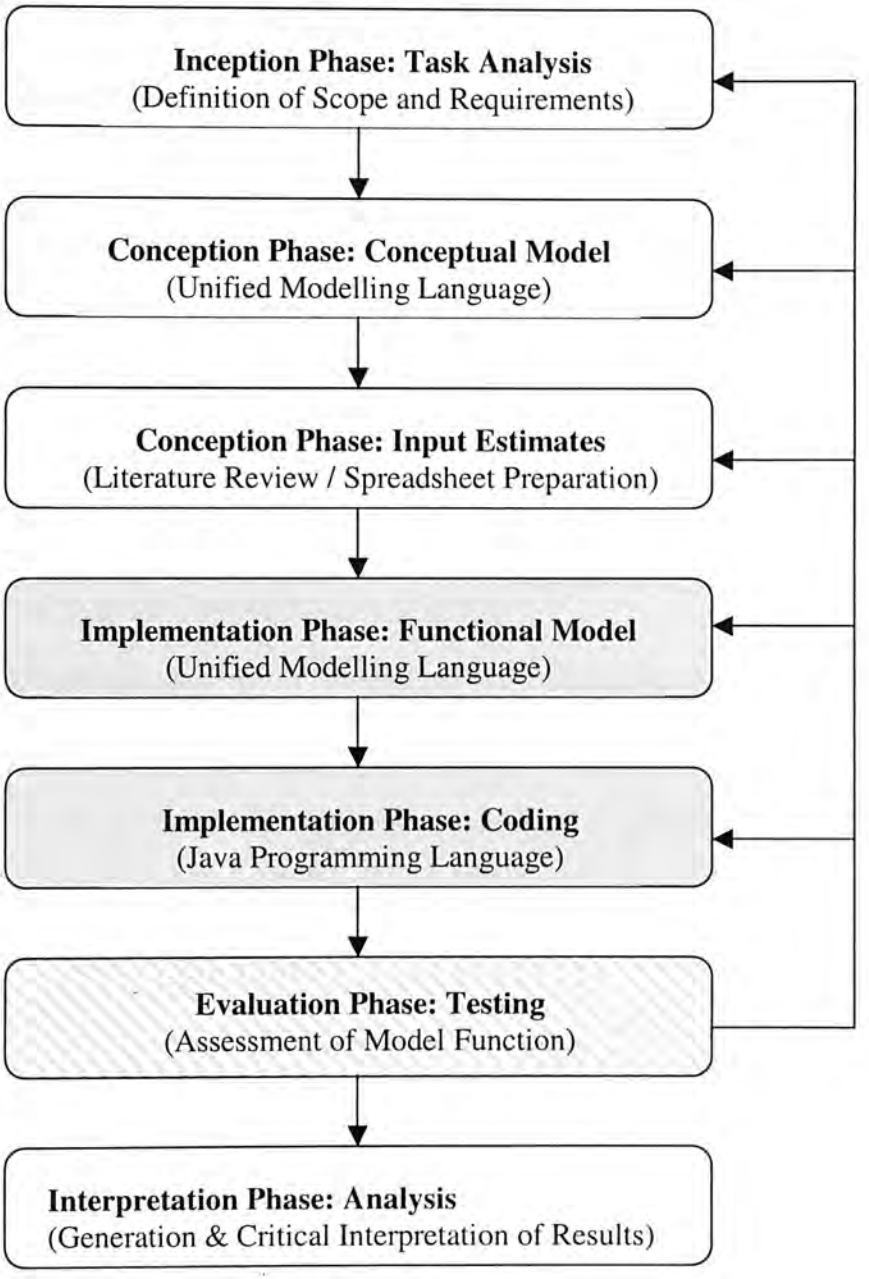
4.2.9 Interpretation of Model Results

Model results were interpreted in the context of the identification of asymptomatic mismatch repair gene mutation carriers at the population level, as considered in the discussion section.

4.2.10 Summary

The stepwise process of computer model development is summarized in figure 4.2.5.

Figure 4.2.5 Strategy For Computer Model Development



- ☐ Phase conducted by author ("domain expert")
- ☐ Phase conducted by computer programmer in collaboration with author
- ☐ Phase conducted by both author and computer programmer

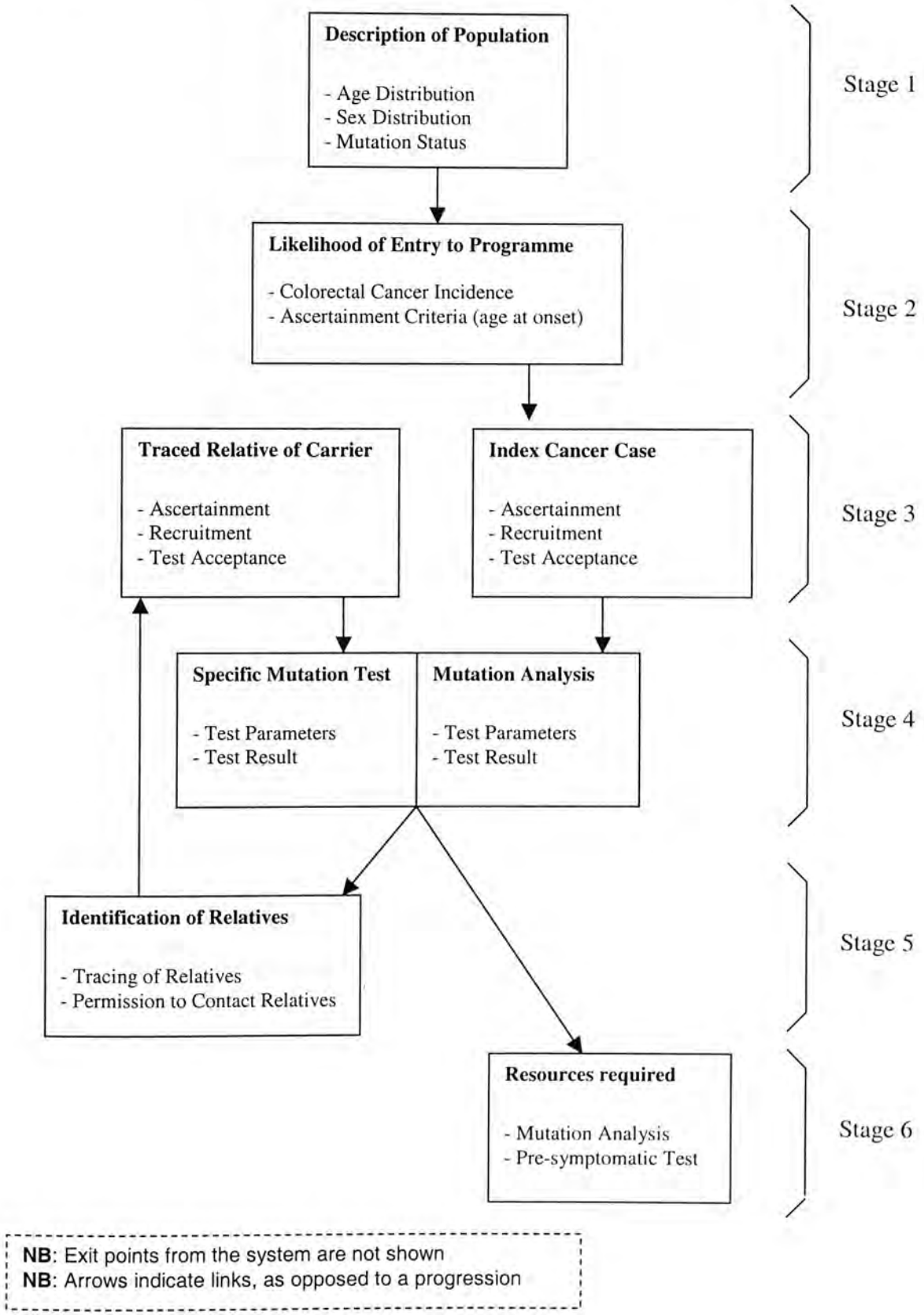
4.3.1 Task Analysis

The task analysis was used to set the terms of reference for model development, outlining the requirements and defining the scope of the model. The requirements and scope were subject to alteration during model development to reflect new ideas, revised estimates, or difficulties encountered in the later stages.

4.3.1.1 Scope and Requirements of Model

Figure 4.3.1 provides an overview of the task analysis constructed. This schematic diagram illustrates the key stages of cascade genetic testing to be included in the model, and indicates the main factors that must considered at each stage. These stages are considered in more detail in the subsequent text.

Figure 4.3.1 Task Analysis: Overview



NB: Exit points from the system are not shown
NB: Arrows indicate links, as opposed to a progression

Stage 1: Description of Population

The size of the Scottish population, and its age and sex distribution, will be presented in the form of a spreadsheet, or "look-up table", so that the computer model can access this data. The sub-group within the Scottish population who carry a pathogenic MMR gene mutation will also be defined in these terms (i.e. total number, and age/sex distribution).

Stage 2: Probability of Entering the Cascade Genetic Testing Programme

In cascade genetic testing, mutation carriers are identified by targeting individuals deemed to be at high-risk of having a mutation. In the context of MMR gene mutations, the principal options are to target individuals with a family history of colorectal cancer, or to target colorectal cancer cases. For practical reasons (considered further in the discussion section) this computer model will consider only the latter situation, an approach that is comparable to the ongoing COGS study. Conducting mutation analysis is therefore limited to colorectal cancer cases. For the purposes of the computer model, cases in which mutation analysis reveals a mismatch repair gene mutation are referred to as "index cases", since these people provide a route into a mutation-carrying family.

Since a relatively high proportion of colorectal cancer cases occurring at a young age are of hereditary origin, potential index cases can be targeted more efficiently by considering age at onset of colorectal cancer as the criteria for inclusion in the model

cascade genetic testing system. Variable age limits will thus provide one of the crucial inputs to the model.

The probability of developing colorectal cancer for individuals without pathogenic MMR gene mutations (non-carriers) can be assumed to be equal to the population incidence of colorectal cancer. People with pathogenic MMR gene mutations (carriers) will have an increased likelihood of developing colorectal cancer, determined by the penetrance of the mutations in question. The number and age/sex distribution of carriers and non-carriers that are predicted to occur over a period of time will be calculated and recorded in a spreadsheet for use by the model.

Stage 3: Ascertainment & Recruitment

Figure 4.3.1 illustrates the two potential routes into the cascade genetic testing programme; either as a potential index case, or as a relative of a known carrier. If the appropriate criteria are met, the individual concerned should be referred for genetic testing, but ascertainment will depend on several factors. Firstly, the potential participants must be willing and able to allow contact with the system. A small proportion of patients may reject such contact, and patients who are very ill, or who die before referral for genetic testing may not be included in the cascade genetic testing programme. Secondly, potential participants must attend an interview and genetic counselling session with a genetic nurse, and must consent to genetic testing and provide a blood sample for this purpose. Each of these stages constitutes a possible exit point from the programme, and the probability of completing each

stage will be defined by acceptance rates set for each individual. Acceptance rates may vary by age, sex, whether mutation analysis or a pre-symptomatic test is being offered and, in the latter situation, by distance of relationship to a known carrier. Accordingly, a spreadsheet that records acceptance rates and has the capacity to vary rates by these factors will be created.

Stage 4: Genetic Testing

The results of genetic testing will clearly depend on the mutation status of the individual in question, but will also be influenced by the parameters of the test being applied. False negative and false positive results may occur if sensitivity and specificity of the genetic testing are less than 100%. Consequently the capacity to adjust test parameters will be incorporated into the model.

Stage 5: Identification and Tracing of Relatives

The 'depth' of any cascade genetic testing program is an important consideration. For the purposes of this model, cascade genetic testing will be restricted to first and second-degree relatives of a known carrier. This option represents a pragmatic approach to cascade genetic testing. Testing first-degree relatives only is likely to limit the effectiveness of cascade genetic testing, due to the fact that some first-degree relatives will be unavailable or will decline testing. Similarly, extending testing beyond second-degree relatives is problematic since more distant relatives will be harder to trace and contact, and may be less willing to participate. Testing first and second-degree relatives is also the approach used as part of the COGS

programme, and hence modelling this strategy will ultimately facilitate comparison between the model and this programme.

Current ethical guidelines regarding consent to genetic testing mean that relatives of known carriers cannot usually be contacted and offered testing without the permission of the index case or known carrier in question (see discussion). Thus, the model will include a facility for blocking the cascade genetic testing process within a family if such consent is not forthcoming. The probability that consent to contact relatives will be given will depend on age, sex, and possibly the route of testing (i.e. whether the known carrier themselves underwent mutation analysis or a pre-symptomatic test). These estimates will be recorded in spreadsheet form. The traceabilty of relatives is another issue that must be acknowledged in the model. Relatives who are eligible for testing may be untraceable, or may not respond to contact from the cascade genetic testing program. These factors will be dealt with at the ascertainment stage, by providing an estimate of probability of contact with the cascade genetic testing programme that is specific to relatives.

Stage 6: Resources

The computer model of cascade genetic testing is designed to facilitate the evaluation of this strategy in terms of the number of carriers that will be identified and the relative efficiency with which this is achieved. Decisions on whether or not such a program should be implemented in real life will depend on a wider assessment of the health benefits and economic issues involved. These issues are

considered in the discussion section of this chapter, but a detailed economic evaluation is beyond the scope of the proposed modelling project. However, a component relating to the resources required for direct implementation of cascade genetic testing will be included in the model. This will be designed to facilitate investigation of the effect of resource limitations on the outcomes of cascade genetic testing, and to provide an indication, in relative terms, of the appropriate allocation of available resources.

4.3.1.2 Required Estimates and Sources of Data

The data sources available to inform the inputs and parameters of the model were also considered at the task analysis stage, and are presented in table 4.3.1. In many instances, data are scarce or the available estimates have wide confidence intervals. The need to build the resulting uncertainty and flexibility into the model was acknowledged and emphasized at the task analysis stage.

Table 4.3.1 Required Estimates and Data Sources For Model Inputs

Stage (See Figure 4.3.1)	Required Estimate	Source(s)
1	Size and age/sex distribution of Scottish Population	Government Statistics (ISD)
1	Population prevalence of MMR mutations	Literature (58), COGS and related analyses
2	Age/sex dependent penetrance of MMR gene mutations	Literature (3, 4, 57, 289), COGS
2	Proportion of colorectal cancer cases meeting criteria for inclusion in cascade genetic testing programme for: (a) Non-carriers (b) Carriers	Estimates will vary according to criteria: (a) Government statistics (ISD), COGS (b) Literature, COGS
3	Delay before presenting to system	COGS
3	Proportion of cases meeting criteria that are actually referred to cascade genetic testing programme	COGS
3	Acceptance rate of genetic testing amongst colorectal cancer cases*	COGS
3	Acceptance rate of genetic testing among relatives of known mutation carriers*	Literature (164), COGS
3	Time required for recruitment	COGS
4	Test parameters for mutation analyses	Literature (62), COGS
4	Time required for mutation analysis	COGS
4	Test Parameters for pre-symptomatic test	Literature, COGS
4	Time required for pre-symptomatic test	COGS
5	Number of relatives identified and traced per newly identified carrier*	Family structure simulation
5	Acceptance rate for allowing contact with relatives*	COGS
5	Traceability of relatives	COGS, previous research experience, Register General Office Scotland

*Age /sex specific information will be incorporated where possible

Abbreviations: COGS, Colorectal Cancer Genetic Susceptibility study; ISD, Information and Statistics Division.

4.3.2 Conceptual Model

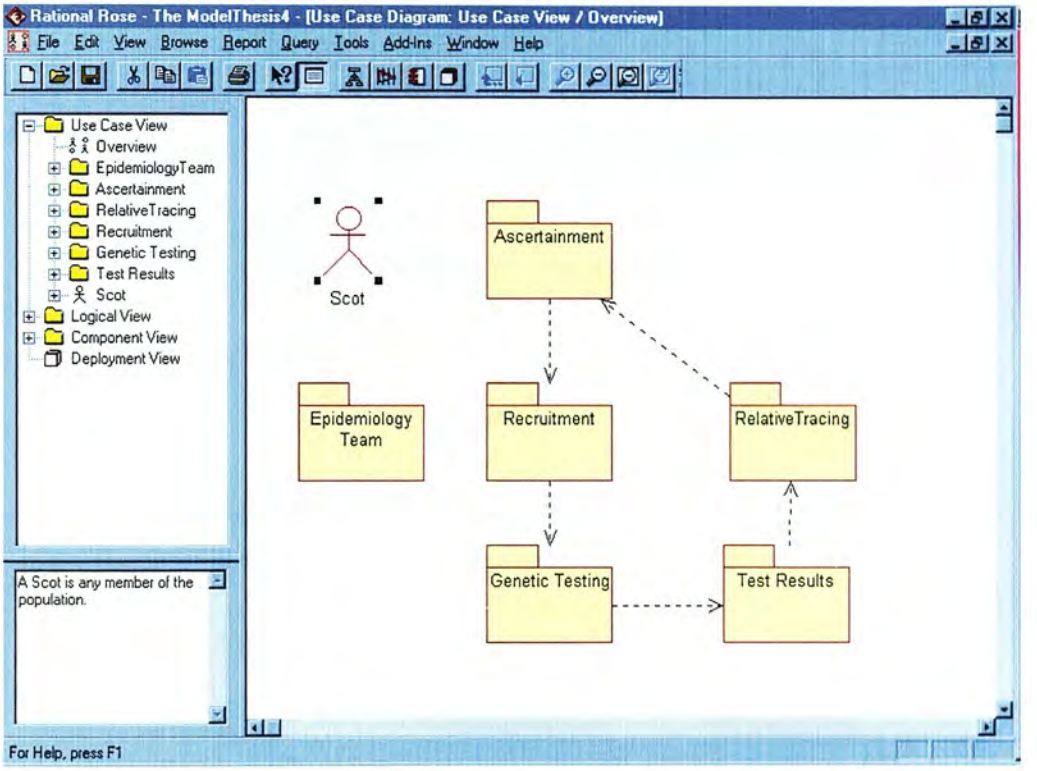
The iterative nature of model development dictates that the conceptual model evolves through numerous versions. Various aspects of the model that were initially included were discarded in subsequent iterations, and conversely other aspects were incorporated in the later stages. The final conceptual model, presented in this thesis, was designed to provide an efficient model that includes all the information required to inform development of the functional model, with minimal surplus material.

A visual model is best communicated and understood through direct exploration of the model using the appropriate software. It is not feasible to represent the entire conceptual model in the main text of this thesis, and for this reason details of the conceptual model are confined to appendix A7. The objective of this section is to give an overview of the conceptual model, illustrating the scope of the model, and focusing on some key areas as examples of how the model is constructed.

4.3.2.1 The Use Case View

The use case view represents the conceptual model form the domain perspective, and describes the actors and use cases involved and their relationships. To organise the conceptual model, and facilitate its logical presentation, the aforementioned actors and use cases are grouped into various packages. The packages created in the use case view are illustrated in the screenshot below, both in the browser view and in diagrammatic form.

Figure 4.3.2 Overview of Packages Created in Use Case View



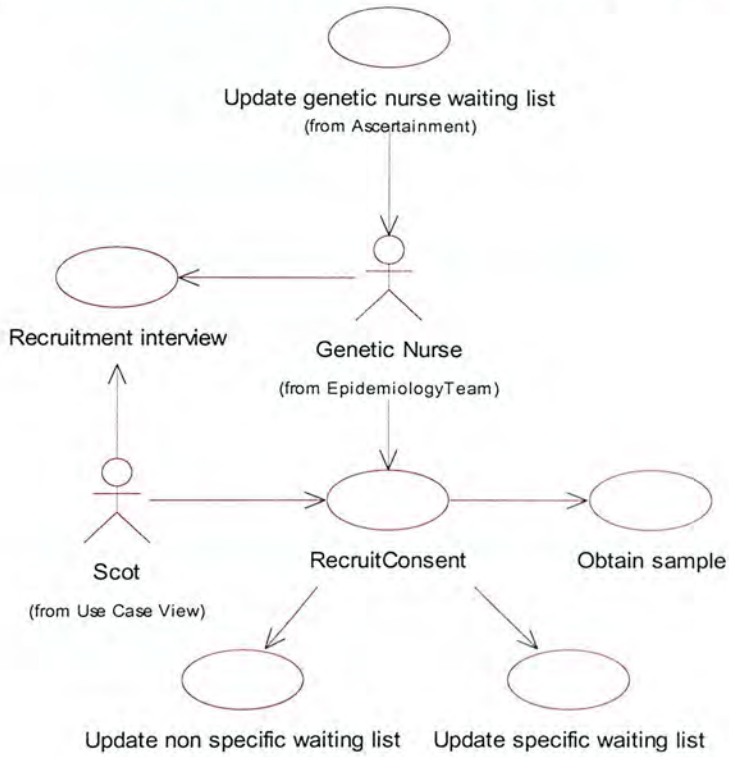
NB: The "Scot" actor represents any member of the Scottish population that interacts with the cascade genetic testing system, and is represented in the overview because it is common to several packages.

Figure 4.3.2 illustrates the inter-dependency of the five main packages. The "Epidemiology Team" package contains details of all 'actors' involved in implementing cascade genetic testing, including health professionals such as a genetics nurse. It should be emphasised that actors are created to carry out the functionality required in the model system, and do not directly correspond to actual people who may implement a real-life cascade genetic testing programme.

To illustrate how each package is represented in the use case view, the actors and use cases created in the recruitment package, which describes the process of

recruiting an eligible Scot to the cascade genetic testing programme, are displayed in the following use-case diagram.

Figure 4.3.3 Use Case Diagram for Recruitment Package



The general message of the above diagram is that the genetic nurse will receive information from the genetic nurse waiting list regarding a Scot who is eligible for genetic testing and has agreed to an interview. The genetic nurse will arrange and conduct this interview, and will ask the Scot if they consent to genetic testing. If consent is forthcoming, the genetic nurse will initiate genetic testing by taking a blood sample and updating the appropriate waiting list.

The detail of the recruitment process is contained within each specific use case, and is represented as the documentation, sequence diagram, collaboration diagram and

flow of events pro-forma for each use case. The vital "Recruit Consent" use case is represented in these terms below, by way of an example.

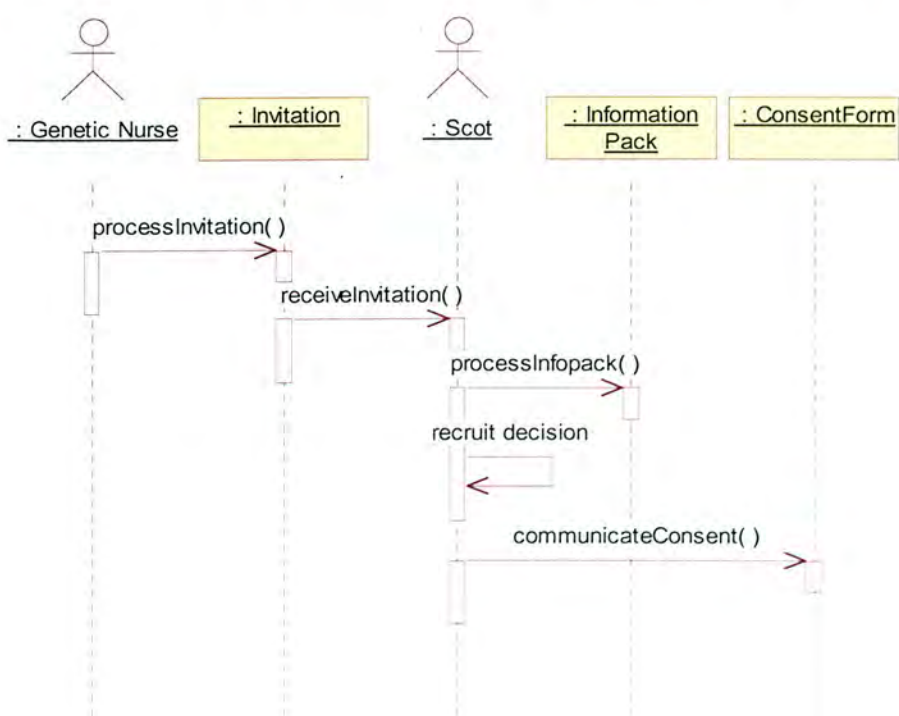
Documentation for "Recruit Consent" Use Case

The "Consent" use case is a vital one, and can be adapted to several similar situations in the model. At various stages in the model system, the "Scot" in question will be asked to give their consent to proceed to the next stage. Consent at the recruitment stage is concerned with obtaining consent to undergo genetic testing. The probability that this consent will be forthcoming is defined in an "Acceptance" look-up table.

Sequence Diagram for "Recruit Consent" Use Case

Scenarios relating to the Recruit Consent use case were illustrated using sequence diagrams, as shown in figure 4.3.4.

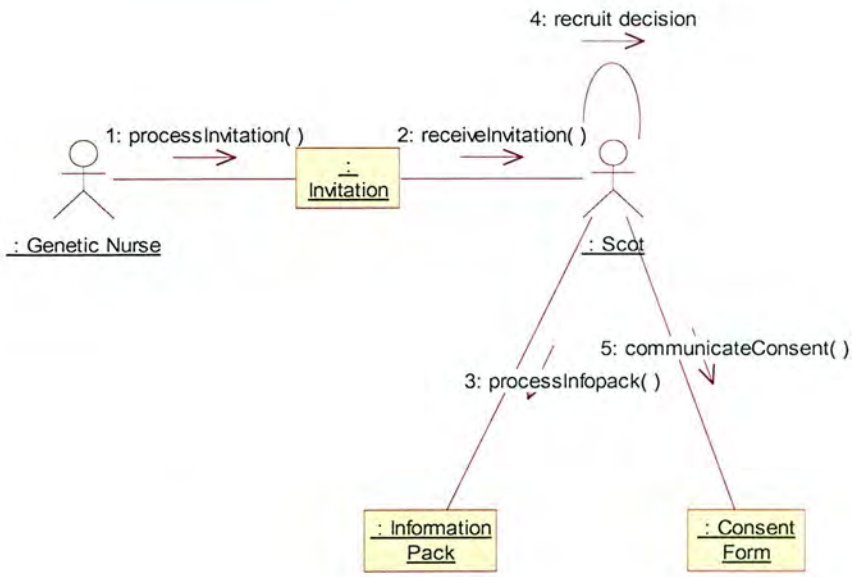
Figure 4.3.4 Sequence Diagram for Recruit Consent Use Case



Collaboration Diagram for "Recruit Consent" Use Case

An alternative means of representing a scenario of the "Recruit Consent" use case is provided by a collaboration diagram, as illustrated in figure 4.3.5.

Figure 4.3.5 Collaboration Diagram for Recruitment Consent Use Case



Flow of Events for "Recruit Consent" Use Case

A textual description of the various possible paths through a use case is provided by the "Flow of Events", which has a strict presentation format. The flow of events for the Recruit Consent use case is as follows:

1.1 Preconditions

The "Recruit Invitation" use case must have been completed successfully.

1.2 Main Flow

This use case is begun by a "Scot" who has developed colorectal cancer, upon being invited to undergo genetic testing by a genetic nurse. The Scot in question will study the information pack (i.e. obtain and process the relevant information from the genetic nurse) and then consider whether to take part. They will then complete a consent form specifying their decision. If they agree to take part the S-1: Accept sub-flow will be applied, if they do not the S-2: Decline sub-flow will be used.

1.3 Sub-flows

S-1: Accept: The Scot is prepared to participate in the next stage of the cascade genetic testing programme (i.e. undergo testing), and completes a consent form to that effect.

S-2: Decline: The Scot does not wish to participate in the next stage of the cascade genetic testing programme, and completes a consent form to that effect.

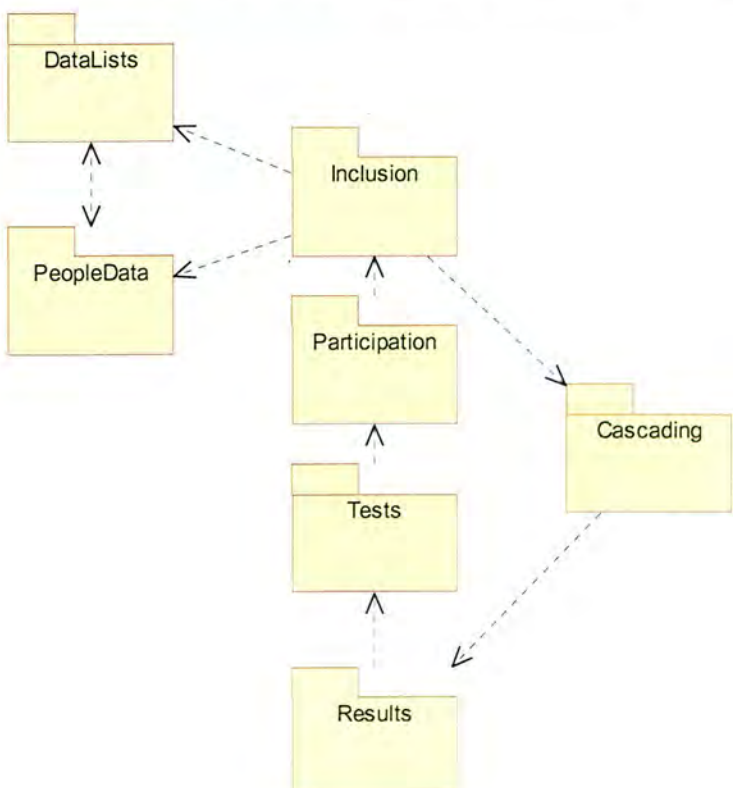
1.4 Alternative flows

E-1: The Scot does not complete a consent form. This can occur whether the invitation is made by post or in person, as there is no guarantee that a decision will be reached at interview. In this scenario the individual in question is assumed to have declined the invitation and accordingly takes no further part in the system.

4.3.2.2 The Logical (Class) View

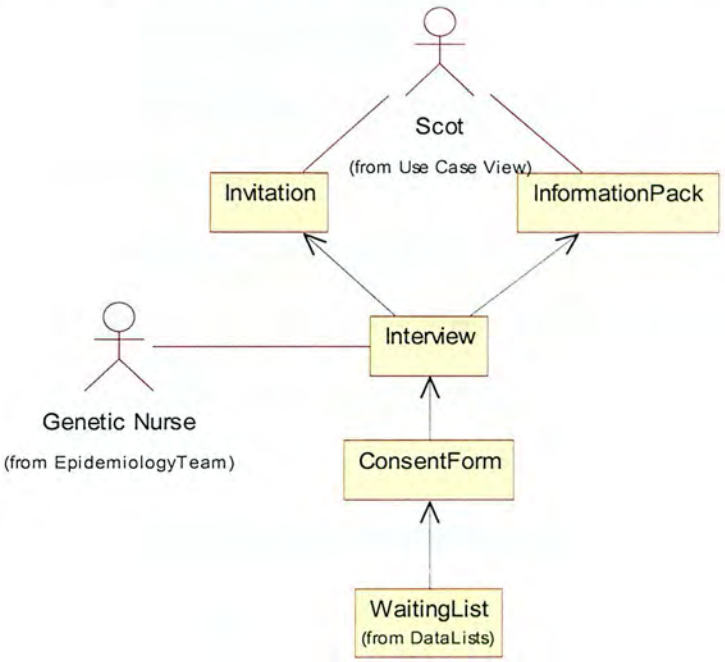
This view is primarily designed as an intermediate step between the domain perspective and the computer programming perspective. Essentially, constructing a "logical view" of the model involves representing the largely abstract functionality described in the use case view in a systematic object-orientated manner. As considered previously, objects can be organised into classes, based on their attributes, behaviour and semantics. Thus classes, their functions and their relationships, constitute the key feature of object-orientated model development. The structure of a class is defined by its 'attributes', and the functionality of the class (the role of the class in the model) is defined by its 'operations'. Generally, operations are mapped to messages in the scenario diagrams of the use case view. Classes can also be grouped into packages to facilitate clear presentation of the logical view. The packages created for the model are presented in figure 4.3.6.

Figure 4.3.6 Main Class Diagram of Computer Model



The classes contained within the 'Participation' package, and their interactions, are presented in the following class diagram for illustrative purposes.

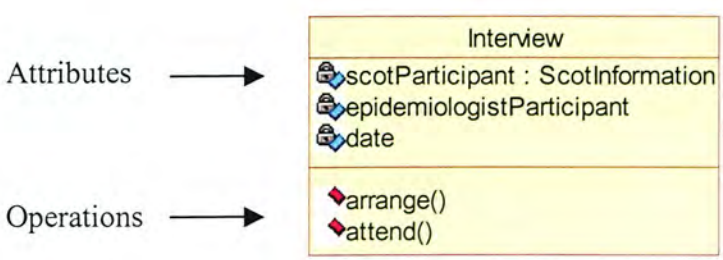
Figure 4.3.7 Class Diagram for Participation Package



Further detail of the "Interview" class was communicated through the attributes, operations and documentation of this class, as illustrated below.

Attributes and Operations for "Interview" Class

Figure 4.3.8 Attributes and Operations of "Interview" Class



Documentation for "Interview" Class

This class represents any interview that takes place during ascertainment or recruitment, between a 'Scot' and a member of the 'Epidemiology Team'.

4.3.2.3 Further Considerations

The above examples provide an overview of how the conceptual model was developed. Further details of the conceptual model were provided verbally at discussions between the author and the computer programmer. In addition, the capacity provided by the UML to document specific attributes and operations, and to specify associated data types and other detail, was used in some circumstances.

4.3.3 Estimates and Default Settings

The estimates and default settings used for the computer model are presented in this section. The validity, accuracy and limitations of these estimates are considered further in the discussion section.

4.3.3.1 Population Demographics & Genetic Epidemiology of Colorectal Cancer

Data on the size and age/sex distribution of the Scottish population came from the 2001 census (228), and information regarding the incidence and age/sex distribution of colorectal cancer in Scotland was obtained from statistics provided by ISD (123). For the purposes of the model, these figures were assumed to remain constant over time, and the implications of this are considered in the discussion.

Dunlop et al., (2001) (58) have provided the only available estimate of the population prevalence of MMR gene mutations. Hence this figure of 1:3139 (95%CI

= 1:1247, 1:7626) was directly used as a model input. Estimates of penetrance were obtained from the consensus of published studies (3, 4, 57, 289) which suggest that penetrance is approximately 80% in males and 40% in females. The cumulative incidence of colorectal cancer in MMR gene mutation carriers was presented in another paper by Dunlop et al., (57). The shape of the chart of cumulative incidence from this paper was used to estimate the lifetime cumulative incidence in various age groups (considered equal to cumulative incidence at age 70 for the purposes of the model).

Using the above estimates it was possible to estimate the number of mutation carriers and the number of non-carriers in a particular age group that would be expected to develop colorectal cancer per year. These estimates were set out in an excel (Microsoft™) spreadsheet, and functions were set up to allow the estimates pertaining to mutation carriers to be re-calculated according to defined changes in prevalence and penetrance. Estimates were calculated separately for males and females. This spreadsheet constitutes a vital input to the computer model, as it essentially defines the demographics of the population in question and the genetic epidemiology of colorectal cancer with respect to MMR gene mutations. Results from this spreadsheet, calculated for default settings of penetrance and prevalence, are presented in table 4.3.2.

Table 4.3.2 Expected Colorectal Cancer Cases in Scotland

(i) Males

Age group	Males, 2001 Census	Annual Expected Cases in Male Non-Carriers	No. of Male Carriers in Population	Cumulative Incidence in Mutation Carriers - Male	Average annual incidence in mutation carriers - males	Annual Expected Cases in Male Carriers
All ages	2,432,494	1,622				
0 - 4	142,360	0		0.000	0.000	
5 - 9	157,030	0		0.000	0.000	
10 - 14	165,583	0		0.011	0.002	
15 - 19	160,935	1	51	0.022	0.002	0.111
20 - 24	157,116	1	50	0.032	0.002	0.108
25 - 29	154,112	2	49	0.065	0.006	0.318
30 - 34	184,674	5	59	0.141	0.015	0.890
35 - 39	194,618	13	62	0.249	0.022	1.341
40 - 44	184,176	23	59	0.422	0.035	2.030
45 - 49	166,925	45	53	0.616	0.039	2.070
50 - 54	174,118	77	55	0.703	0.017	0.959
55 - 59	140,835	137	45	0.757	0.011	0.485
60 - 64	124,651	197	40	0.789	0.006	0.258
65 - 69	110,009	274	35	0.800	0.002	0.076
70 - 74	90,053	298	29			
75 - 79	66,057	261				
80 - 84	36,355	182				
85 & over	22,887	107				
15 - 74	1,842,222	1,072	587			8.65

(ii) Females

Age group	Females, 2001 Census	Annual Expected Cases in Female Non-Carriers	No. of Female Carriers in Population	Cumulative Incidence in Mutation Carriers - Female	Average annual incidence in mutation carriers - Females	Annual Expected Cases in Female Carriers
All ages	2,629,517	1,606				
0 - 4	134,514	0		0.000	0.000	
5 - 9	150,108	0		0.000	0.000	
10 - 14	157,287	0		0.000	0.000	
15 - 19	156,338	1	50	0.000	0.000	0.000
20 - 24	157,271	1	50	0.007	0.001	0.067
25 - 29	163,191	3	52	0.020	0.003	0.139
30 - 34	197,420	4	63	0.060	0.008	0.503
35 - 39	208,336	12	66	0.120	0.012	0.796
40 - 44	193,734	22	62	0.187	0.013	0.823
45 - 49	170,544	37	54	0.280	0.019	1.014
50 - 54	176,989	61	56	0.347	0.013	0.752
55 - 59	147,164	100	47	0.373	0.005	0.250
60 - 64	137,082	153	44	0.393	0.004	0.175
65 - 69	129,107	207	41	0.400	0.001	0.055
70 - 74	116,864	254	37			
75 - 79	99,466	260				
80 - 84	68,634	253				
85 & over	65,468	240				
15 - 74	1,954,040	853	623			4.57

NB: The sources and references for the data contained in table 4.3.2 are stated in the preceding text, and the limitations of these estimates are considered in the discussion section. Blank cells indicate that the relevant data was unavailable.

4.3.3.2 Age Limits

The age criteria for inclusion of colorectal cancer cases in a cascade genetic testing programme are arbitrary, in the sense that they are chosen rather than dictated by available data. At default settings, the age limit for eligibility to the hypothetical cascade genetic testing programme was set to 55 years, to reflect the pragmatic age limit chosen for the COGS study. The age limit for recruiting relatives of mutation carriers was set to 70 years, reflecting the fact that it may be considered

inappropriate to contact elderly relatives and/or offer clinical screening in this group. The minimum age of recruitment for both colorectal cancer cases and relatives of known mutation carriers was set to 18 years at default.

4.3.3.3 Acceptance Rates

Information regarding the uptake of cascade genetic testing for mismatch repair gene mutations is very limited. For the purposes of the computer model, the probability of acceptance at each of four stages was specified. These stages were:

(i) Probability of having contact with cascade genetic testing programme

(ii) Probability of attending interview with genetic nurse

(iii) Probability of accepting genetic test

(iv) Probability of allowing contact with other family members

Several factors have the potential to influence acceptance rates at each of these stages, including age, sex, whether the individual in question is a colorectal cancer patient or relative of a known carrier, and, in the latter situation, the degree of relationship between the known carrier and the relative being offered the test. Reliable information about acceptance rates in this context is not available, and estimates for this model were largely based on limited data from the ongoing COGS study. These data suggest that around 74% of colorectal cancer patients diagnosed under the age of 55 will undergo genetic testing. Accordingly, simplified and conservative default values that were consistent with this estimate of overall acceptance rate were applied to the model. The default probability of acceptance at

each stage was thus set to 0.9, regardless of the age, sex and circumstance of the individual in question. This gives an overall probability of accepting genetic testing of $(0.9)^3 = 0.73$. Acceptance rates were recorded in table form in an excel file, which was converted to a text file to be 'read' by the computer model. This file, which illustrates both the default settings and the capacity to include detailed acceptance information at a later stage, is presented in Appendix A8.

4.3.3.4 Waiting Times (Delays)

In practice, any cascade genetic testing programme is likely to feature delays at various stages, notably during the time consuming process of conducting mutation analysis. Consequently, delays, often manifest as time on a particular waiting list, were incorporated into the model. This feature contributes to the realism of the model, and manipulating the delays and studying the effect on model outcomes should provide data on the effect that delays will have on the overall utility of cascade genetic testing. Default values for delays were again based on experience with the ongoing COGS study, although in practice delays will depend on various financial and administrative factors, and may vary considerably. These default values are presented in table 4.3.3.

Table 4.3.3 Default Values for Delays

Stage	Explanation	Delay (weeks)
EC Registration	Referral of potential index case or relative of known carrier to cascade genetic testing system, and subsequent registration by the epidemiology coordinator (EC)	4
Genetics Nurse	Administrative processing of individual eligible for genetic testing, and arrangement of recruitment interview with genetics nurse	4
Mutation analysis	Time taken to perform mutation analysis	52
Pre-symptomatic test	Time taken to perform pre-symptomatic test	4
EC Lab Results	Time for interpretation and processing of lab results	2
Relative Presentation	Delay between identification of an index case and presentation of relatives of that mutation carrier to the cascade genetic testing system	4, 8 or 12

The estimate of one year for performing mutation analysis on a sample from a potential index case is a conservative one, reflecting the maximum waiting time at this stage in the ongoing COGS programme. The other delays are also conservative. In theory the administrative tasks and pre-symptomatic test could all be performed within a week, but in reality some delays are likely to occur, and this is reflected in the default estimates. Each delay outlined in table 4.3.3 can be manipulated by changing the default settings in the graphical user interface of the completed computer model.

One further source of delay occurs with the presentation of relatives of known carriers to the system. This process is staggered, with relatives presenting at either 4, 8 or 12 weeks after a mutation was found in their relative. This reflects the likely pattern of presentation that would occur in practice, since newly identified mutation

carriers may not inform their relatives of their potential risk immediately after diagnosis, and relatives in turn may take time to consider whether or not to become involved in the cascade genetic testing programme. An additional advantage of this staggered presentation is that relatives will enter the model system over a period of weeks, and will not overload the available resources by presenting at one time.

4.3.3.5 Resources

The component of the computer model relating to 'resources' is designed to facilitate an assessment of the impact that limitations in available resources may have on the outcomes of the model system. Resource limitations have the potential to severely affect model outcomes. At default settings, however, the model was intended to run as if resources were effectively unlimited. This was achieved by setting all resources to 1000 resource units per week, where one resource unit is sufficient to carry out a specific task (for example, genetic nurse interview) once in the given week. Resource inputs and default settings are presented in table 4.3.4.

Table 4.3.4 Resource Inputs and Default Settings

Resource	Detail	Default Setting
Referrals	The number of colorectal cancer cases that can be ascertained and processed per week	1000
Lab Results	The number of lab results that can be received and processed per week	1000
Genetic Nurse Interview	The number of genetic nurse (recruitment) interviews that can be conducted per week	1000
Mutation analysis	The number of mutation analyses that can be carried out per week	1000
Pre-symptomatic test	The number of pre-symptomatic tests that can be carried out per week	1000

In a real cascade genetic testing programme, the mutation analysis resource is likely to be crucial, since this is likely to be the most time consuming and expensive aspect of such an undertaking. Resources and the impact of their limitations are considered further in the discussion section.

4.3.3.6 Test Parameters

Pre-symptomatic testing for a specific mutation is a straightforward laboratory process, and for the purposes of the model the sensitivity and specificity of this test is assumed to be 100%. The model retains the capacity to set different values for these test parameters.

Mutation analysis is more complex. The test parameters will depend on the laboratory protocol and techniques employed, and the results may be open to different interpretations. However, there is no "gold standard" method for determining whether or not an individual has a pathogenic MMR gene mutation, and hence no systematic means of determining the test parameters. For the purposes of the model, therefore, it was decided to set the default values for sensitivity and specificity of mutation analysis to 100%. This is not considered to be an accurate, or even realistic, estimate, but rather is designed to standardize this aspect of the model and thus facilitate study of the effect of other factors. The implications of this approach are considered further in the discussion section.

4.3.3.7 Summary

A summary of default model settings is presented in table 4.3.5.

Table 4.3.5 Default Settings for Model Inputs

Input	Default Value
Population Demographics	
Population size	5,064,200
Population age/sex distribution	See table 4.3.2
Genetic Epidemiology	
Prevalence of mismatch repair gene mutations	1:3139
Penetrance (male)	0.8
Penetrance (female)	0.4
*Acceptance Rates**	
Probability of Allowing Contact	0.9
Probability of Attending Genetics Nurse Interview	0.9
Probability of Accepting Genetic Testing	0.9
Probability of Allowing Relative Contact	0.9
Waiting Times (delays)	
EC Registration delay	4 weeks
Genetics Nurse	4 weeks
Mutation analysis	52 weeks
Pre-symptomatic test	4 weeks
EC Lab Results	2 weeks
Age Criteria (eligibility for recruitment)	
Minimum age of colorectal cancer cases	18
Maximum age of colorectal cancer cases	55
Minimum age of relatives	18
Maximum age of relatives	70
Resources	
Referrals	1000
Lab Results	1000
Genetic Nurse	1000
Mutation analysis	1000
Pre-symptomatic test	1000

*Acceptance rates can vary according to the age and sex of the individual in question, and also according to whether this individual is a potential index case, a first-degree relative of a known carrier, or a second-degree of a known carrier. However, at default settings acceptance rates are constant, irrespective of these factors. See appendix A8 for more detail.

4.3.4 Functional Model

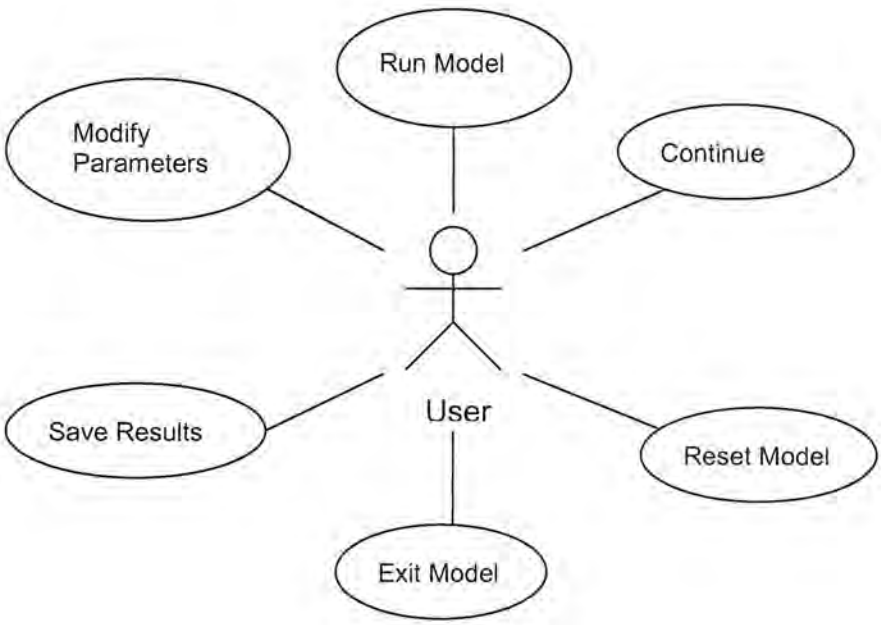
4.3.4.1 Introduction

As detailed in the methods section, the functional computer model was developed and implemented by a computer programmer, working in conjunction with the author of this thesis. Throughout this process, the functionality of the model was controlled and specified by the author, whereas the technical aspects of model development were the responsibility of the computer programmer. Reflecting this division of labour, this section consists of a description of the model's functionality from the domain perspective, focussing on *what* the model does rather than *how* the computer programme works.

The model itself is designed using the Java programming language, utilising Sun ONE Studio 4 Update 1 software. The model can be accessed through this software, facilitating the examination and alteration of the functional model and programming code using the "Editing" tab for the appropriate project. A separate "Edit Graphical User Interface (GUI)" tab allows access to the code of this aspect of the working model. These commands can be used to alter the appearance or functionality of the model at the "coding" level, and are mainly included to facilitate future model development by the author, the computer programmer, or other researchers. For the purposes of this thesis, no such alterations were made and the coding remains the original work of the computer programmer. Consequently these aspects of the model are not considered in this section, which is primarily concerned with the operation or "running" of the model from the domain perspective.

In a functional model, the requirements of the model user are represented as use cases. These high-level use cases are shown in figure 4.3.9.

Figure 4.3.9* User-Orientated Use Cases in Functional Model



*This figure has been modified from a design by Jeff Lloyd, who acted as the computer programmer for the development of the functional model

The functional model is centred on an "engine" class that actually drives the model.

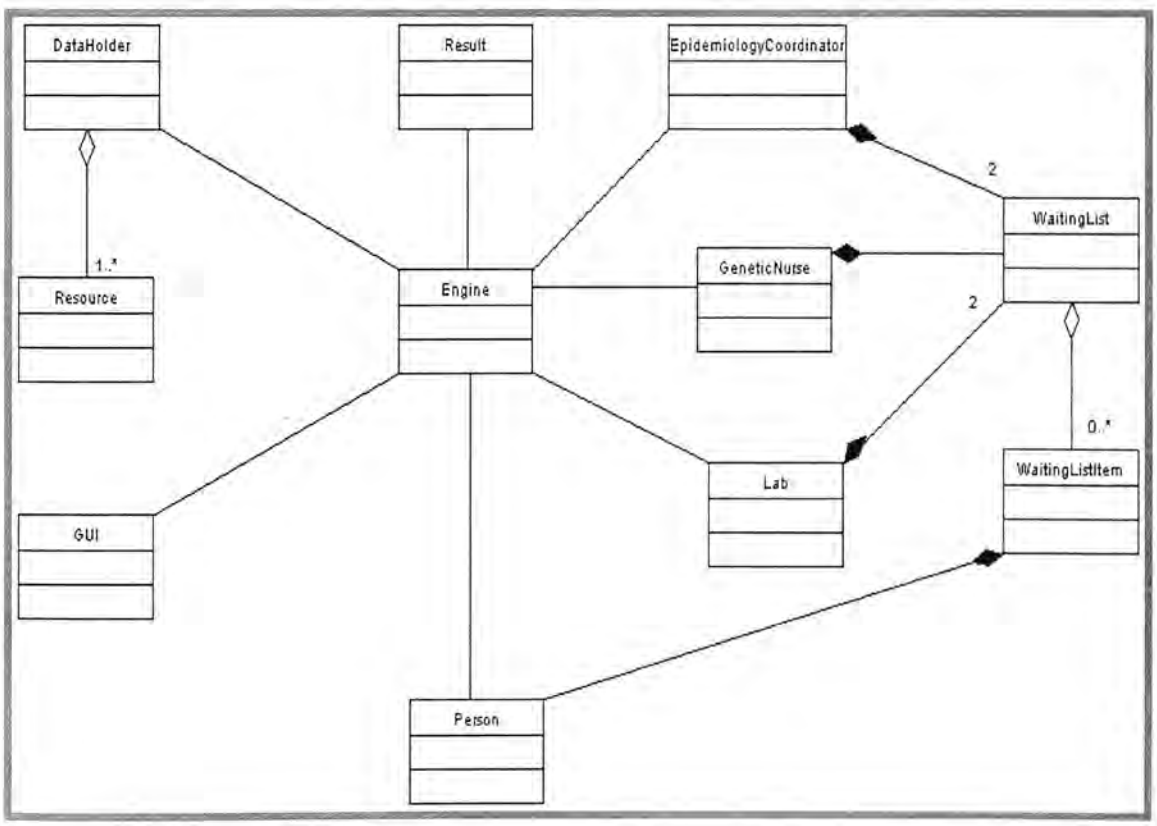
The following classes were created to implement the functional model.

- Graphical User Interface (GUI)
- Engine (responsible for running the model)
- Epidemiology Co-ordinator (responsible for holding the coordinator waiting lists and performing the tasks of the co-ordinator)
- Genetic Nurse (responsible for holding the genetic nurse waiting lists and performing the tasks of the genetic nurse)
- Lab (responsible for holding the laboratory waiting lists and performing the tasks of the lab worker)
- Waiting List (a simulation of a "first-in-first-out" waiting list)

- Waiting List Item
- Person (responsible for holding the methods and operations of a "Scot").
- DataHolder (accesses data files and holds input values)
- Resources (allocates resources to waiting lists on a weekly basis)
- Result (responsible for recording and presenting model outcomes)

A class diagram illustrating these classes, and the relationships between them is presented in figure 4.3.10.

Figure 4.3.10* Class Diagram: Overview of Functional Model

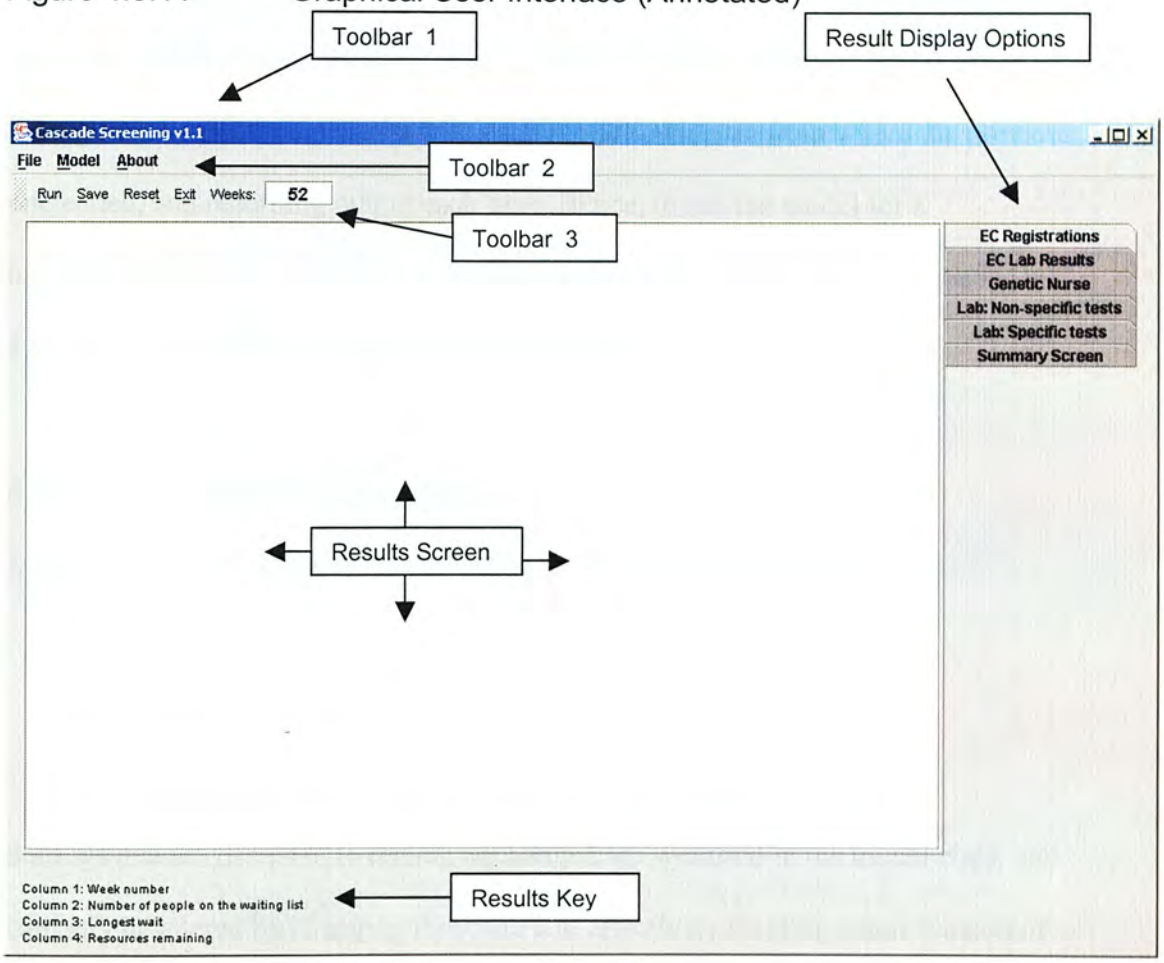


*This figure has been reproduced from a design by Jeff Lloyd, who acted as the computer programmer for the development of a functional model
Abbreviation: GUI, Graphical User Interface.

4.3.4.2 Graphical User Interface (GUI)

The model is 'run' using the "execute project" command, which opens up the Graphical User Interface (GUI). This window is shown in figure 4.3.11, below.

Figure 4.3.11 Graphical User Interface (Annotated)



Toolbar 1 forms the address bar at the top of the GUI, and states the model and version that is running. Cascade Screening v1.1 is the name of the final version of the model used for the subsequent analyses. Toolbar 2 contains the "About" button, which states the technical information about the model; the "File" button, which gives the standard options of "Save", Save As" or "Exit", and finally the Model

button, which opens a new window entitled "Settings and Resources". This latter window is considered in more detail later. The third tool bar provides the user with the option to "Run", "Save", "Reset", or "Exit" the model, and also contains a window in which the user can specify the number of weeks the model is to be run. This is set to 52 weeks as a default. Clicking on the "Run" button will run the model for the specified length of time. If it is clicked more than once the model will run through additional time periods, with each run following on from where the previous one ended, and recording output each time. Hence, to run the model for a hypothetical ten-year period, and record results annually, the run time can be set to 52 weeks and the "Run" button clicked ten times.

4.3.4.3 Settings & Resources

Model settings or inputs can be determined at three levels, namely:

(i) Code

(ii) Look-up tables

(iii) GUI

Settings that are designed to remain unchanged are specified in the model code, and can only be altered by changing this code and effectively creating a new version of the model. An example such a setting is the probability of a first-degree relative of a known MMR gene mutation carrier being a carrier themselves: this is always maintained at 0.5.

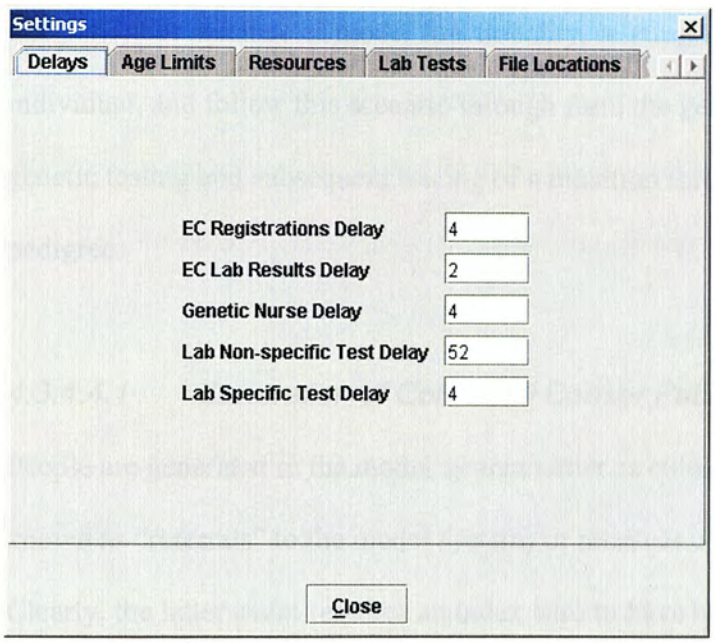
Complex inputs that the user may wish to vary include population demographics, estimates related to the genetic epidemiology of MMR gene mutations (prevalence,

penetrance), and acceptances rates according to age, sex and circumstance. Such inputs are defined in text files or "look-up tables", and can be changed by either altering the look-up tables themselves, or using the "Settings and Resources" and "Files" commands to specify a different look-up table for the model to access. For instance, in the analysis stage, a number of look-up files were created, each representing the "expected colorectal cancer cases" file (see table 4.3.2) with different assumptions of the prevalence of mismatch repair gene mutations in the Scottish population.

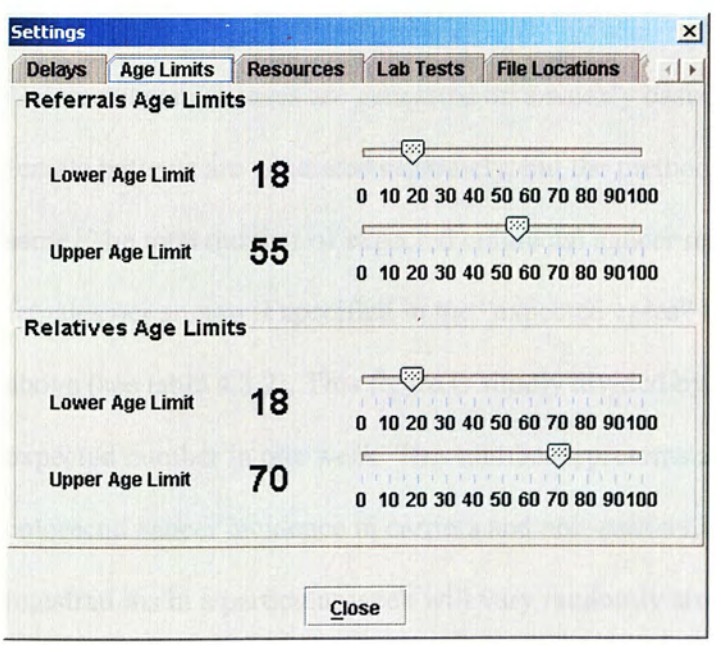
Finally, settings that the user may wish to change regularly, such as delays, resources and age limits can be adjusted directly using the "Settings and Resources" option of the GUI. Settings may be specified using simple entry boxes, as used for delays, or a sliding scale, as used for age limits. Windows for these two parameters are illustrated in figure 4.3.12.

Figure 4.3.12 Settings Adjustment Windows

(i) Delays



(ii) Age Limits



4.3.4.4 Running the Model

The starting point for the model system is defined chronologically. However, to understand and communicate what the model actually does it is useful to consider one particular instance of model functionality, relating to one hypothetical individual, and follow this scenario through from the generation of this person to genetic testing and subsequent tracing of a mutation through their extended pedigree.

4.3.4.4.1 Generation of Colorectal Cancer Patients

People are generated in the model system either as colorectal cancer patients (also known as "referrals" to the model system) or relatives of known mutation carriers. Clearly, the latter event requires an index case to have been generated, and so the initiating event in terms of the function of the model is the development of colorectal cancer in a member of the Scottish population.

Colorectal cancer cases are generated on a weekly basis. In the model, male and female patients are generated separately, but the methodology used is exactly the same. The total number of expected colorectal cancer registrations for males or females per annum is specified in the "expected cases" look-up table discussed above (see table 4.3.2). This figure is simply divided by 52 in order to determine the expected number in one week. This number represents a prediction, based on colorectal cancer incidence in carriers and non-carriers, and the actual number of registrations in a particular week will vary randomly around this prediction. This

random element is incorporated into the model through the use of an algorithm 'loop', which is presented below, written in "pseudocode".

Step 1: Calculate X = 1/(expected number + 1)

Step 2: Generate Y = Random number between zero and one.

Step 3: If X < Y a new case will be created and the algorithm is repeated

If X > Y no further cases will be created and the algorithm will terminate.

Hence the number of cases generated varies randomly from week to week, but, on average, the model will produce the expected number of registrations specified in the look-up table.

For each referral, age is specified on the basis of the population data presented in the expected cases look-up table. Age is allocated by firstly determining the cumulative proportion of expected cases in each age group. For example, a negligible number of registrations would be expected to be under 20 years old, but 0.00-0.01 referrals may be in the 20-24 age bracket, 0.01-0.03 may be 25-29, 0.03-0.06 may be in the 30-34 age group and so on. A random number between zero and one is then generated, and the age group that includes this number becomes the age group of the referral. In this manner, age group is assigned on a random basis, but following the age distribution of colorectal cancer occurrence specified in the look-up table. Hence, the majority of referrals will be elderly, and the average age will be 60-64, reflecting the average age at onset of colorectal cancer in Scotland. Once the age group has been established, the age of the referral is assigned randomly within the group parameters

(i.e. a referral in the 50-54 age group will have a 20% chance of being assigned an age of 50, 51, 52, 53 and 54 years).

Age criteria, as specified by the model user, are applied to new referrals. If the referral is older than the maximum age limit, they take no further part in the cascade genetic testing process, although their existence is recorded. Referrals with an age below the minimum criteria are withheld from the cascade genetic testing system until they reach the minimum age. In practice, referrals below the minimum age limits are extremely rare.

The probability that a given referral will be a mutation carrier varies according to the age at onset (i.e. the age group of the referral). This probability can be determined by simply dividing the number of expected colorectal cancer cases in mutation carriers by the total number of expected cases for the specified age/sex group. A random number between zero and one can then be generated, and if this number is less than the calculated probability of being a mutation carrier the referral in question is assigned mutation carrier status. Once again, this algorithm ensures that the attributes of the referral (in this case mutation status) are assigned randomly but follow the demographic and epidemiological data specified in the expected cases look-up table.

4.3.4.4.2 *Progression of Referrals Through Cascade Genetic Testing Programme*

Essentially, the progression of a newly created referral through the model system can be described in terms of waiting lists and acceptance probabilities. The

acceptance algorithm for allowing contact with the cascade genetic testing system is applied to each new referral of an age within the specified inclusion criteria. If the individual in question is deemed to 'accept' at this stage, they are placed on the 'Epidemiology Co-ordinator Waiting List'. At such time as the referral has been on the waiting list for the minimum period (delay) and there are sufficient resources available, they will be removed from this waiting list and the acceptance algorithm for "probability of accepting genetic nurse interview" will be applied. If the individual in question is deemed to be willing to attend a genetic nurse interview they will be placed on the 'Genetic Nurse Waiting List". Once again, they will remain on this waiting list until the minimum waiting period has elapsed and there are adequate resources to allow the genetic nurse to perform the interview. Whether or not the individual in question, having attended the genetic nurse interview, will accept genetic testing is determined by the acceptance algorithm for genetic testing. If this process determines that they will undergo genetic testing, they will progress to the 'Lab Waiting List', and will finally undergo testing as and when the minimum period has elapsed and there are sufficient laboratory resources. At each stage involving an acceptance algorithm, a result of 'non-acceptance' will result in removal from the waiting list, and the individual concerned will take no further part in the system.

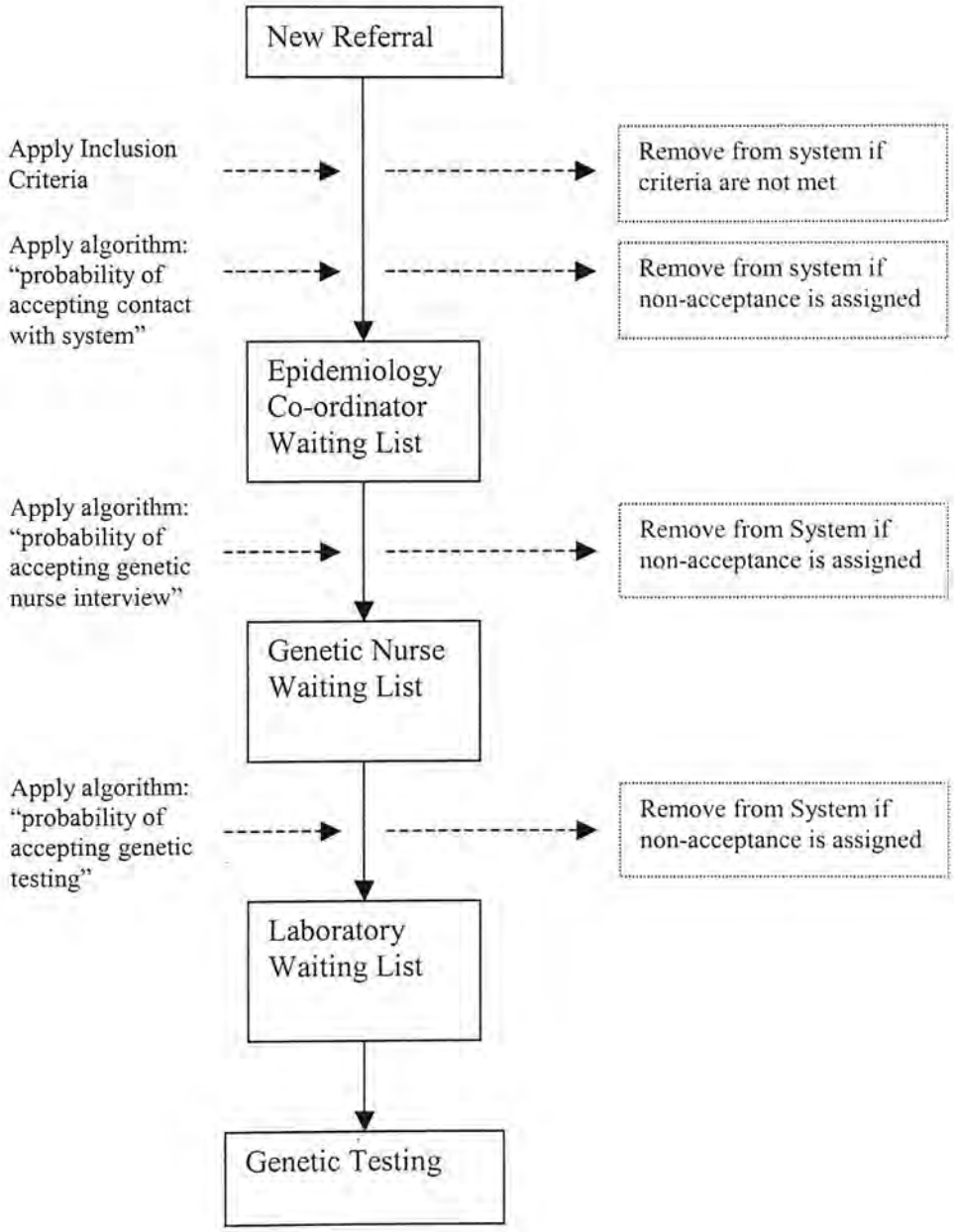
Acceptance algorithms are applied as follows. Once the age, sex and mutation status of a new referral has been established, the "Acceptance" look-up table will be used to determine the probability that the particular referral will participate in various stages of the cascade genetic testing system. At each stage the relevant probability is

applied by generating a random number. A decision of acceptance will be assigned if the random number is less than the specified probability, and a decision of non-acceptance will be assigned if it is greater.

All waiting lists in the model operate on a first-in-first-out basis. Each week the model is running, a check is made to establish if there are any entries on each waiting list. If there are, the length of time that the person at the top of the list has been waiting is obtained. If this time is less than the minimum waiting time specified by the 'delay' settings of the model the list is left as it is. If the individual at the top of the list has been waiting for longer than the minimum delay period they are removed from the list and progress to the next stage of the cascade genetic testing process, provided that there are sufficient resources available for this to be done. If resources are inadequate they will be left on the list and considered again the following week. The above process is repeated until either there are no further people on the list who have waited the minimum period, or until there are no resources left.

The progression of a new referral through the cascade genetic testing system can be presented as a decision tree, as illustrated in figure 4.3.13.

Figure 4.3.13 Progression of Referrals Through Cascade Genetic Testing Programme



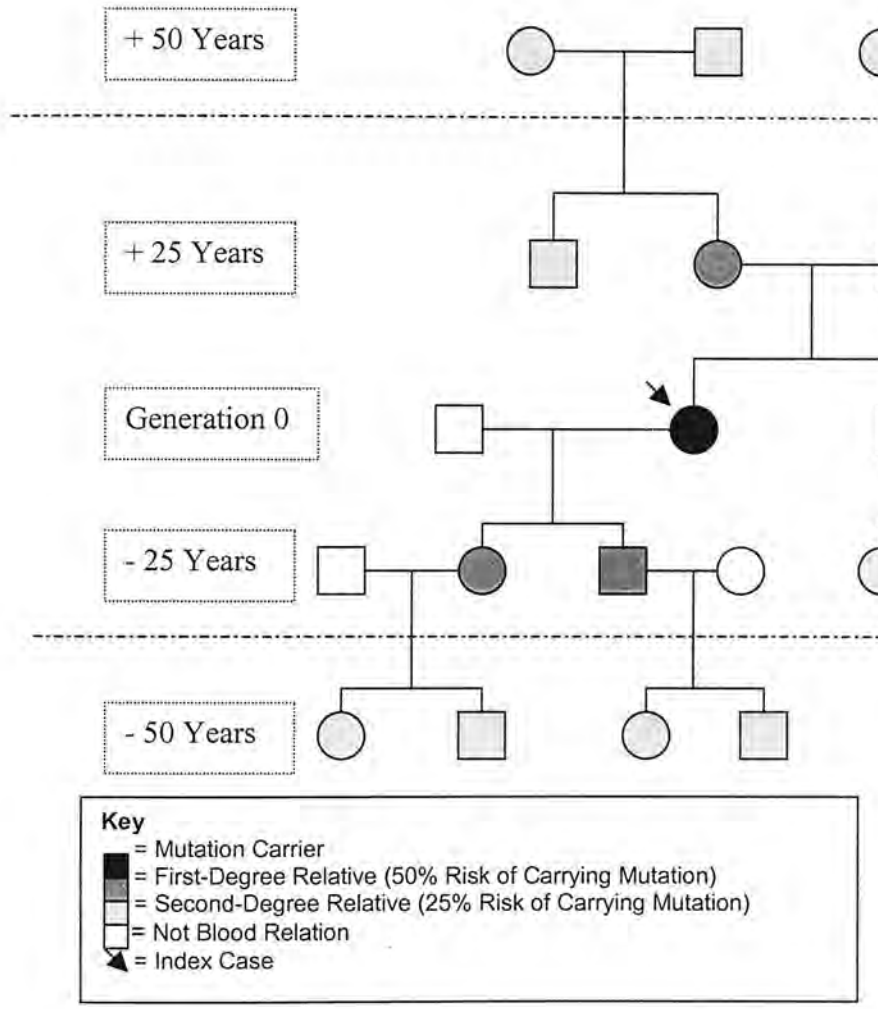
NB: Progression from one waiting list to another is subject to the minimum delay period and to the availability of adequate resources.

The user defines the length of time required to perform mutation analysis. After this delay, a test result is produced, based on the mutation status of the individual concerned, and the test parameters (for the purposes of the analyses considered in this thesis the sensitivity and specificity of all tests are considered to be 100%). If the test result is negative, this fact is recorded, and the individual concerned takes no further part in the system. However, if the results reveal that the referral is a MMR gene mutation carrier they become an 'index case', eligible for the next stage of the cascade genetic testing system, namely tracing of relatives. At this stage the algorithm "probability of allowing contact with relatives" is applied to the new index case. If they are deemed not to accept such contact, the index case will take no further part in the model system. However, if the index case does 'accept' contact with relatives a family tree is created as described below.

4.3.4.4.3 *Generation of Pedigrees*

A highly simplified representation of family structure was applied by the model, which makes the key assumptions that each sibship consists of two siblings (i.e. the population is stable) and the generation gap is 25 +/- 5 years. The pedigree initially created for a new index case consists of all first and second-degree relatives, as shown in figure 4.3.14. Standard pedigree notation, whereby square symbols are used for males and circular symbols for females are used in all pedigree diagrams.

Figure 4.3.14 Pedigree for Index Case



For each new index case, the model generates five first-degree relatives (two parents, one sibling and two children), and 12 second degree relatives (four grandparents, two uncles/aunts, two nieces/nephews and four grandchildren). Relatives are assigned an age based on the age of the index case and the generation difference, with a random variation of +/- 5 years. Relatives are created even if they are unfeasibly old, or of a negative age (not yet born). They will, however, only interact with the model system if and when they meet the specified age criteria.

The sex of relatives is random, but is conditional on the sex of their spouse where applicable. For example, the first parent created will have a 50% chance of being designated as female. If they are deemed to be female, the other parent will automatically become male.

Mutation status is assigned to relatives as followed:

- Siblings and children of known carriers have a 50% probability of being a mutation carrier. A random number between 0 and 1 is thrown, and if this number is < 0.5 they are deemed to be a mutation carrier.
- The first parent has a 50% probability of being a mutation carrier, decided as above. The mutation status of the second parent will be the opposite of the first, ensuring that each carrier has one, and only one, carrier parent.
- Once carrier status in first-degree relatives has been assigned as above, all second degree relatives will have a first-degree relative with a carrier status defined by the model system, and their own mutation status becomes conditional on this factor. For example, the mutation status of uncles/aunts is conditional on the status of their sibling (parent of index case). If their sibling is a carrier, the probability of being a carrier themselves is 50%, again randomly assigned. If their sibling is not a carrier, they are not either.
- The principle that one and only one parent of each known carrier is also a carrier is applied to all couples, including grandparents.
- Non-blood relatives (i.e. spouses/partners) are ignored by the model system.

- The computer model assigns carrier status to relatives prior to any consideration of genetic testing. Therefore, although obligate carriers only interact with the model system if they undergo testing themselves, they are indirectly identified at this stage.

Hence, a set of relatives, with known age, sex and mutation status, is created for each index case identified. The probability that each relative will 'accept' any contact with the cascade genetic testing system is specified by the user in the acceptance look-up table, and may vary according to age, sex, and/or degree of relationship to known carrier. This probability is applied in the manner described previously. Similarly, the age limits set by the user for inclusion of relatives in the system are applied. As with referrals, relatives of known mutation carriers who meet these criteria are placed on the Epidemiology Co-ordinator waiting list, and will progress through the cascade genetic testing system in the same way, subject to delays and resources. In order to stagger the presentation of relatives to the system, and to make the process more realistic, a randomly assigned delay of 4, 8 or 12 weeks will be applied between the identification of the index case and the presentation of each relative.

When a relative undergoes genetic testing and is found to be a mutation carrier, they become the focal point for the creation of new first and second-degree relatives. At this stage, the "allow contact with relatives" algorithm will be applied to the relative in question, and cascade genetic testing will only proceed if acceptance is forthcoming. By definition, many of the relatives of this new mutation carrier will

have already been offered testing due to their relationship with the index case. However, the identification of a new mutation carrier will facilitate the expansion of the pedigree to include new family members who are at 25% or 50% risk of having a mutation. This tracing of mutation carriers through an extended family constitutes the core principle of cascade genetic testing.

The following figures illustrate the new at-risk family members who are generated by the model after various relatives of index cases are demonstrated to have a mutation themselves. For the purposes of the model, no relatives who are more than two generations distant from the index case are included, since such relatives are almost certain to be too old or too young to participate in cascade genetic testing within a meaningful timeframe.

Figure 4.3.15 Expansion of Pedigree After a Mutation is Identified in Sibling of Index Case

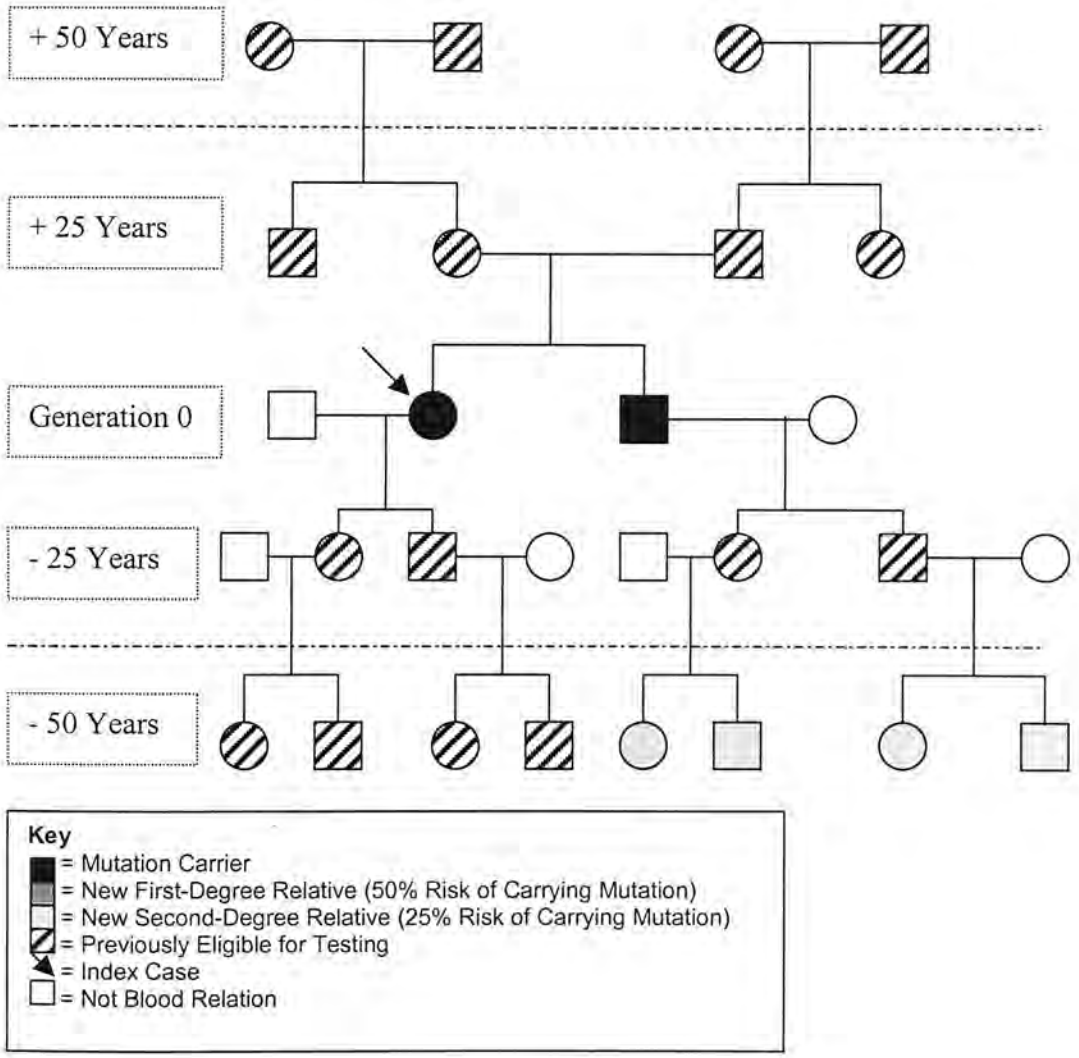
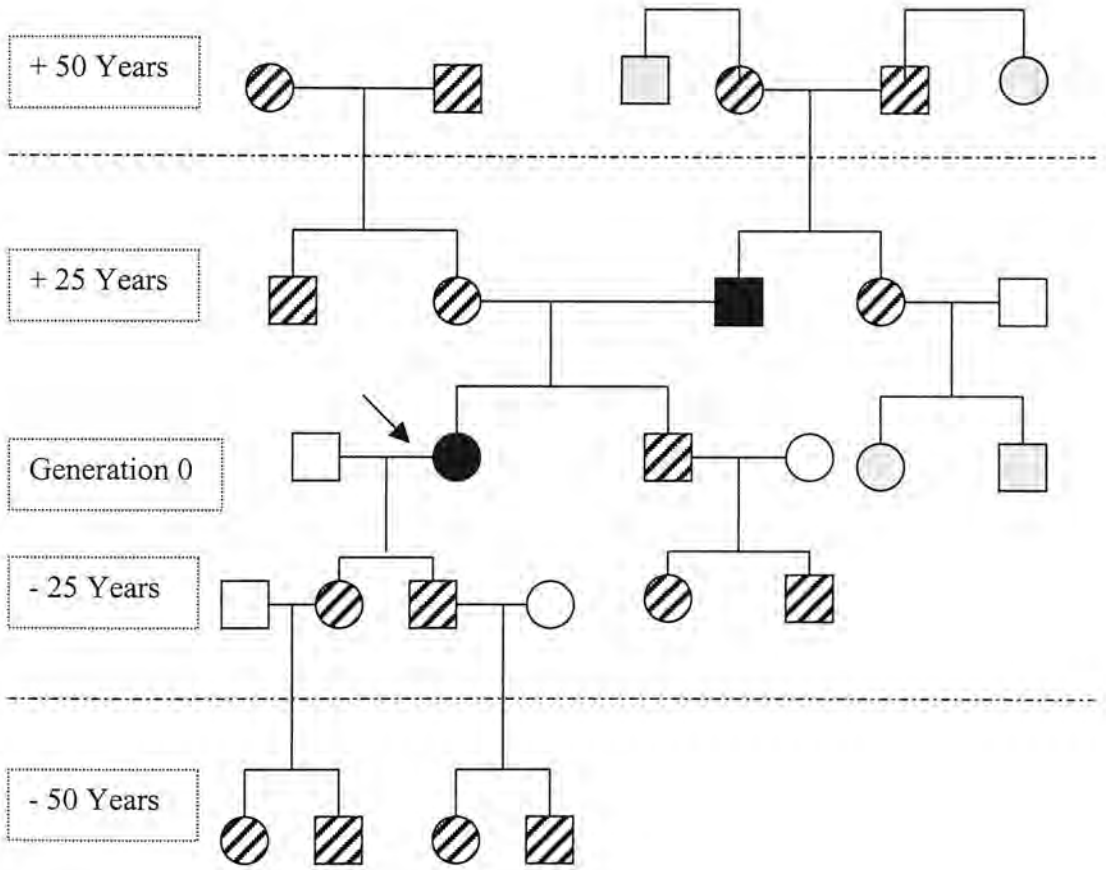


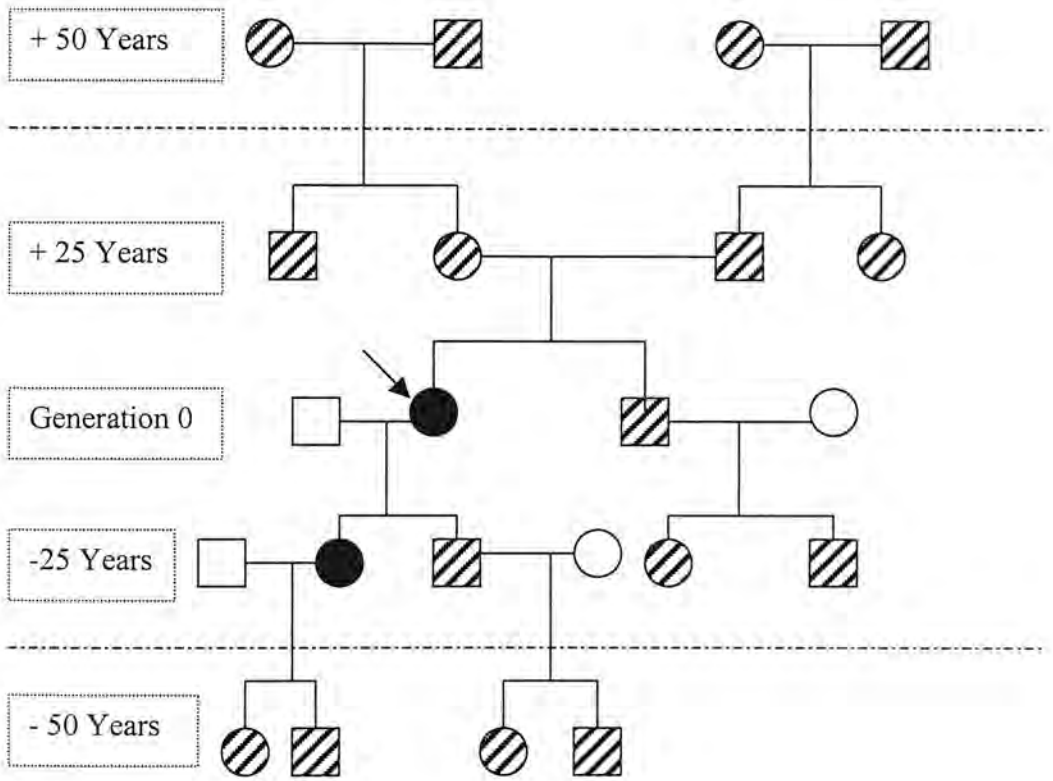
Figure 4.3.16 Expansion of Pedigree After a Mutation is Identified in Parent of Index Case



Key

- = Mutation Carrier
- = New First-Degree Relative (50% Risk of Carrying Mutation)
- = New Second-Degree Relative (25% Risk of Carrying Mutation)
- = Previously Eligible for Testing
- = Index Case
- = Not Blood Relation

Figure 4.3.17 Expansion of Pedigree After a Mutation is Identified in Child of Index Case



Key

- = Mutation Carrier
- = New First-Degree Relative (50% Risk of Carrying Mutation)
- = New Second-Degree Relative (25% Risk of Carrying Mutation)
- = Previously Eligible for Testing
- = Index Case
- = Not Blood Relation

Figure 4.3.18 Expansion of Pedigree After a Mutation is Identified in Aunt/Uncle of Index Case

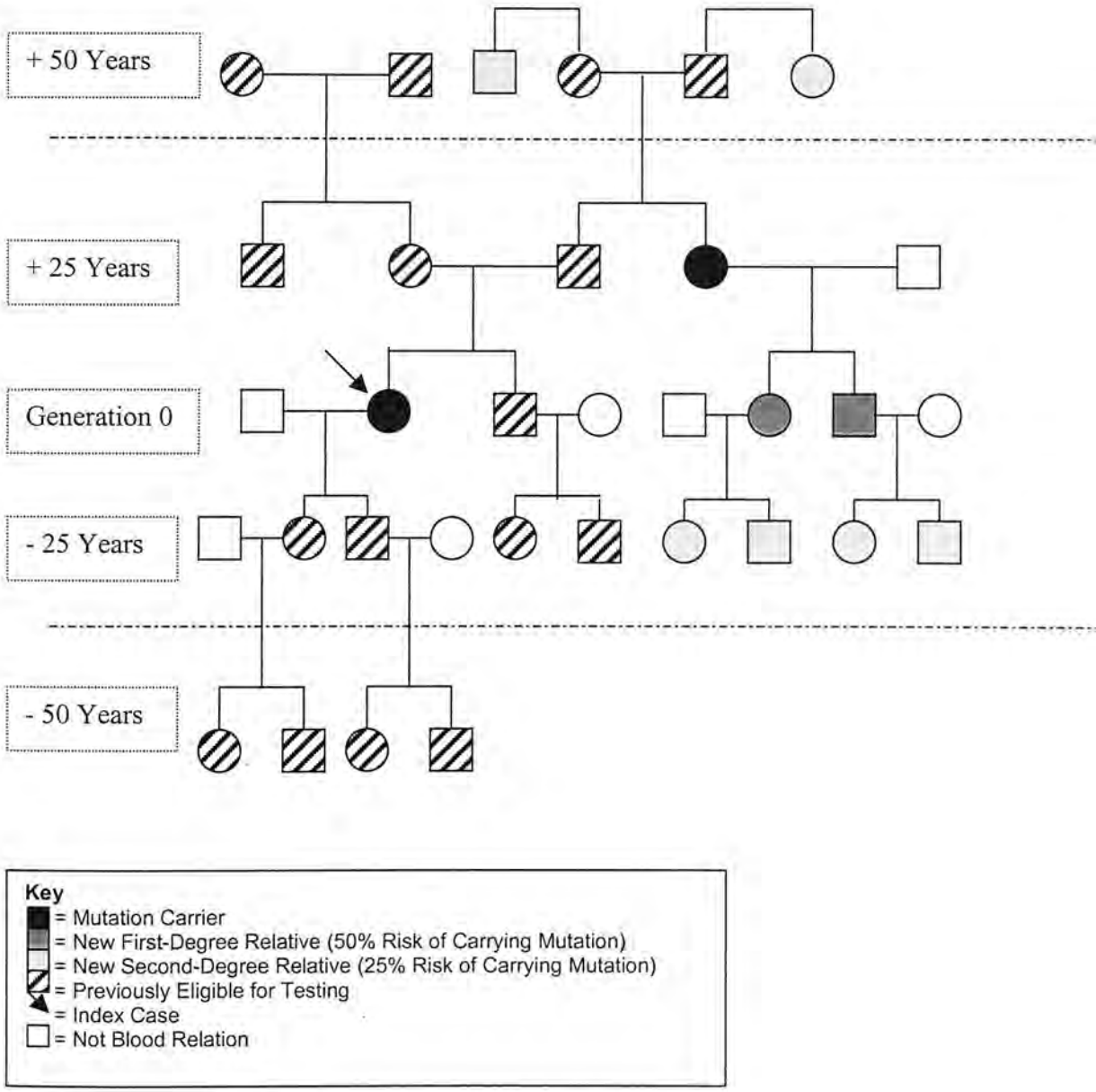
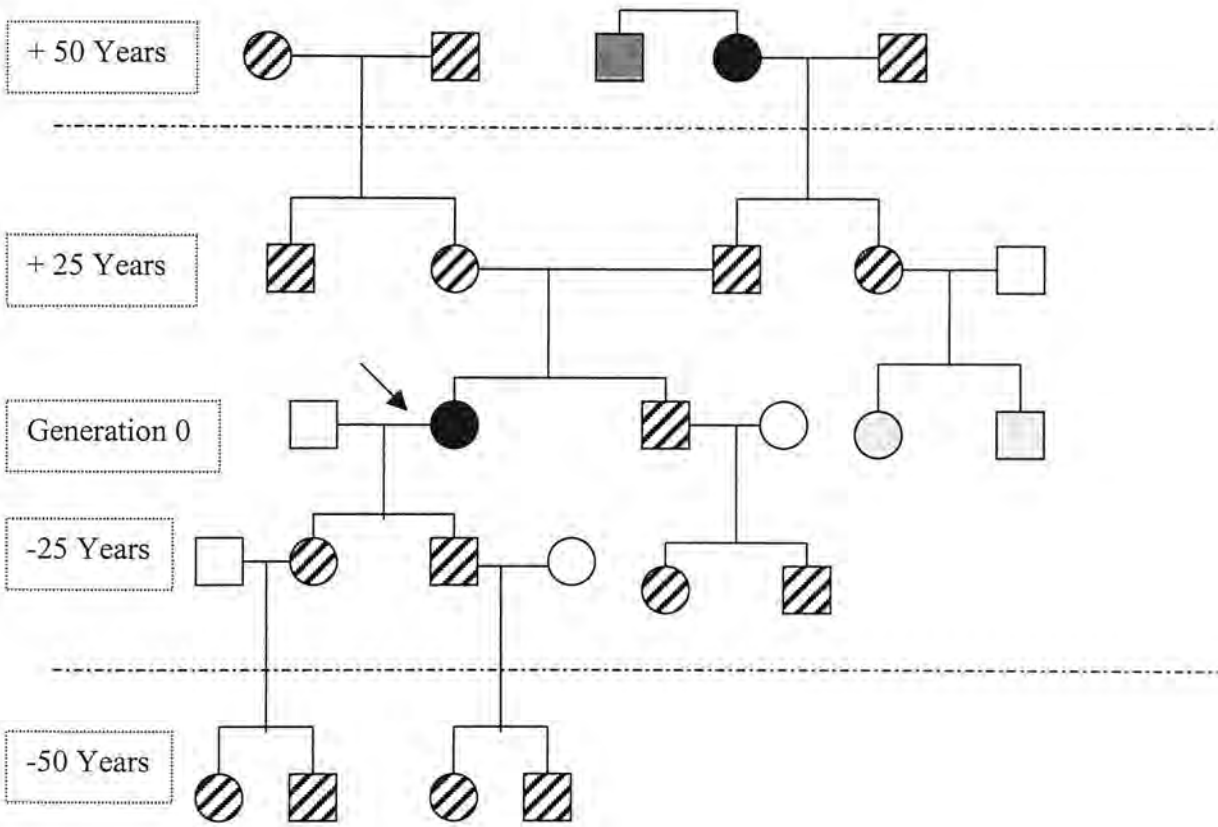
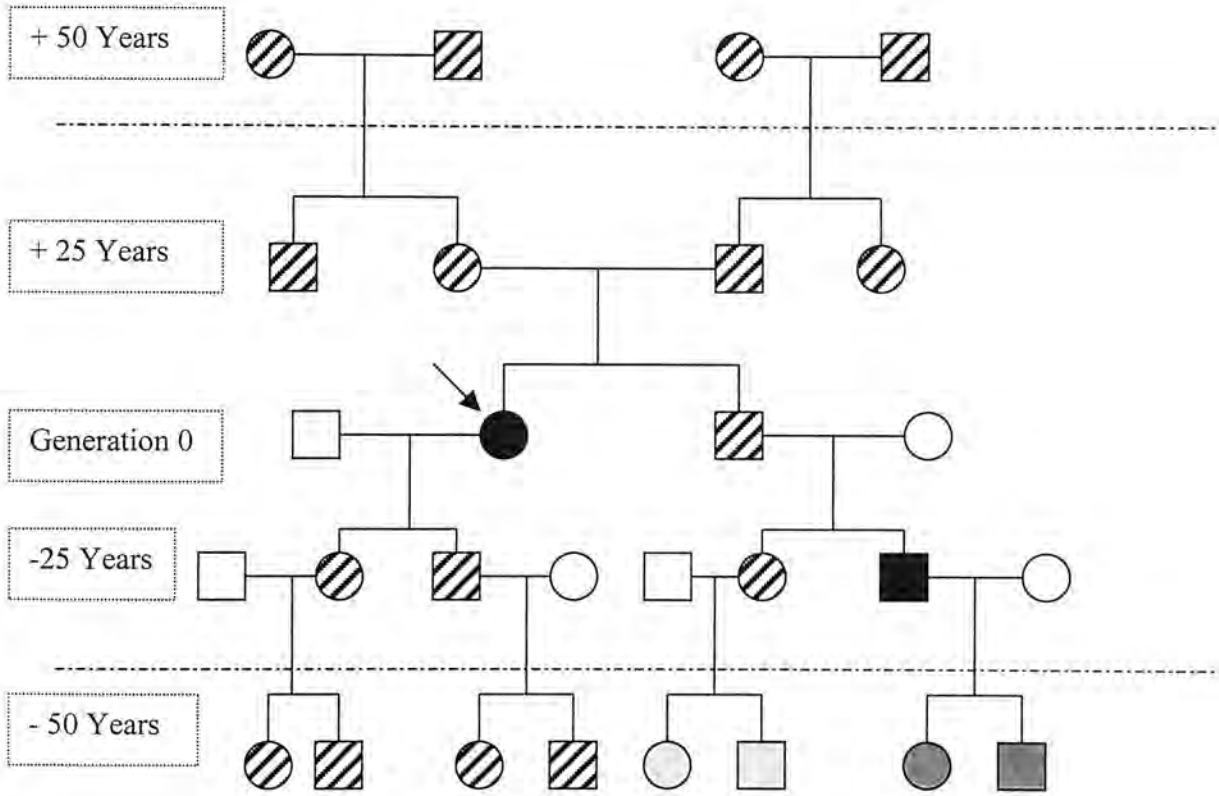


Figure 4.3.19 Expansion of Pedigree After a Mutation is Identified in Grandparent of Index Case



Key
= Mutation Carrier
= New First-Degree Relative (50% Risk of Carrying Mutation)
= New Second-Degree Relative (25% Risk of Carrying Mutation)
= Previously Eligible for Testing
= Index Case
= Not Blood Relation

Figure 4.3.20 Expansion of Pedigree After a Mutation is Identified in Niece/Nephew of Index Case



Key
- = Mutation Carrier
- = New First-Degree Relative (50% Risk of Carrying Mutation)
- = New Second-Degree Relative (25% Risk of Carrying Mutation)
- = Previously Eligible for Testing
- = Index Case
- = Not Blood Relation

The process of identifying index cases and tracing their specific mutation through an extended pedigree where appropriate continues throughout the running period of the computer model, restricted only by delays, acceptance rates and age criteria.

4.3.4.5 Presentation of Model Outcomes

Model outputs are recorded weekly, and summary data is recorded after each run period. For clarity, the results display is broken up into various stages of the model system, as specified using the "Results Display Options" on the right of the GUI (see figure 4.3.11). The key to the results displayed is shown at the foot of the screen. The possible displays are as follows:

(i) EC Registrations

(ii) EC Lab Results

(iii) Genetic Nurse

(iv) Lab: Non-Specific Tests [Mutation analyses]

(v) Lab: Specific Tests [Pre-symptomatic tests]

For each set of results, the GUI will display the week number, the number of people on that particular waiting list, the longest waiting period and the resources that remain at the end of that week.

The "Summary Screen" displays all results in a user-friendly manner, providing details of all the relevant model outcomes as they stand at the end of each run period. Usually the run period is set to 52 weeks, and so this feature acts as an

annual report of the status and progress of the cascade genetic testing system. A typical "Summary Screen" display, after running the model at default settings for one 52-week period, is shown in figure 4.3.21.

Figure 4.3.21 Summary Screen Display

```
*******************************************
Week 52
*******************************************
Number of people generated, 2517 ( number of individuals entering the system)
Total males generated, 1390
  Referrals, 1390          (colorectal cancer patients)
  Relatives, 0     (relatives of known carriers)

Total females generated, 1127
  Referrals, 1127
  Relatives, 0

Male carriers, 6
  Referrals, 6
  Relatives, 0

Female carriers, 3
  Referrals, 3
  Relatives, 0

Below minimum age limits:
  Male Referrals , 1
  Female Referrals, 0
  Male Relatives , 0
  Female Relatives, 0

Within age limits:
  Male Referrals, 244
  Female Referrals, 171
  Male Relatives, 0
  Female Relatives, 0

Above age limits:
  Male Referrals, 1145
  Female Referrals, 956
  Male Relatives, 0
  Female Relatives, 0

Numbers accepting invitation to see Epidemiology Co-ordinator, 368
  Referrals, 368
  Relatives, 0
Numbers not accepting invitation to see Epidemiology Co-ordinator, 36
  Referrals, 47
  Relatives, 0
```

Numbers accepting invitation to see Genetic Nurse, 306
Referrals, 306
Relatives, 0

Numbers not accepting invitation to see Genetic Nurse, 41
Referrals, 41
Relatives, 0

Numbers still waiting to be invited to see Genetic Nurse, 21
Referrals, 21
Relatives, 0

Number of referrals accepting non-specific blood test, 245
Number of referrals not accepting non-specific blood test, 33
Number of referrals still waiting to be seen by Genetic Nurse, 28

Number of relatives accepting specific blood test, 0
Number of relatives not accepting specific blood test, 0
Number of relatives still waiting to be seen by Genetic Nurse, 0

Mutation Analysis True Positives, 0
Mutation Analysis False Positives, 0
Mutation Analysis True Negatives, 0
Mutation Analysis False Negatives, 0
Number of people still waiting for mutation analysis, 245

Pre-symptomatic test True Positives, 0
Pre-symptomatic test False Positives, 0
Pre-symptomatic test True Negatives, 0
Pre-symptomatic test False Negatives, 0
Number of people still waiting for pre-symptomatic test, 0

//

The summary display can be saved as a text file or an excel file, or can be copied and pasted into an existing excel spreadsheet.

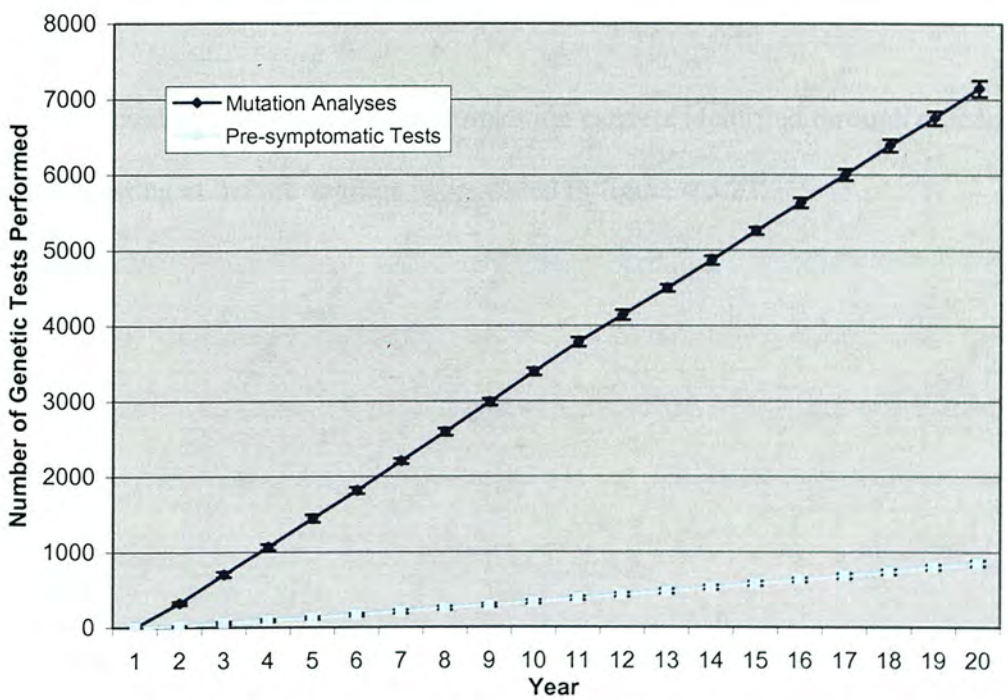
4.3.5 Analysis of Model

The broad strategy for analysis of the model was to firstly examine the model outcomes at default settings in detail, and then systematically alter various inputs in turn and observe the effects of such alterations on the main model outcomes. Results are presented below. On all graphs, Y-error bars represent 95% confidence intervals.

4.3.5.1 Default Settings

Figure 4.3.22 illustrates the total number of genetic tests that the computer model estimated would be conducted during a cascade genetic testing programme.

Figure 4.3.22 Predicted Number of Genetic Tests Performed at Default Settings

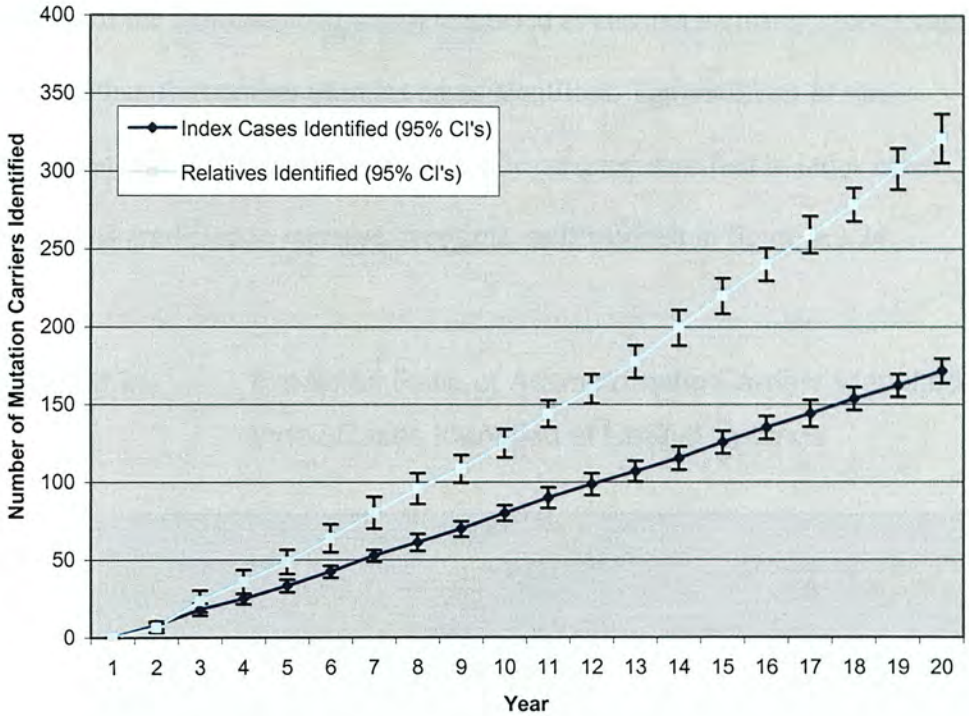


As expected under the assumption that the incidence of colorectal cancer in Scotland remained constant over the period of the hypothetical cascade genetic testing programme, the number of mutation analyses performed increases over time in a linear fashion. Since a delay of 52 weeks for mutation analysis is specified in the default settings of the model, no such tests are recorded in year 1. Subsequently, the number of mutation analyses performed was predicted to increase at the rate of approximately 380 per year, reaching 1443 (95% CI = 1394, 1490) after 5 years, 3387 (95% CI = 3339, 3436) after ten years, and 7142 (95% CI = 7037, 7250) by 20 years.

The predicted number of pre-symptomatic tests performed averaged 126 (95% CI = 109, 143), 334 (95% CI = 312, 356) and 849 (95% CI = 812, 886) after 5 years, 10 years and 20 years respectively. There was a slight trend towards an exponential increase in the number of pre-symptomatic tests performed over time.

The predicted number of MMR gene mutation carriers identified through cascade genetic testing at default settings is presented in figure 4.3.23.

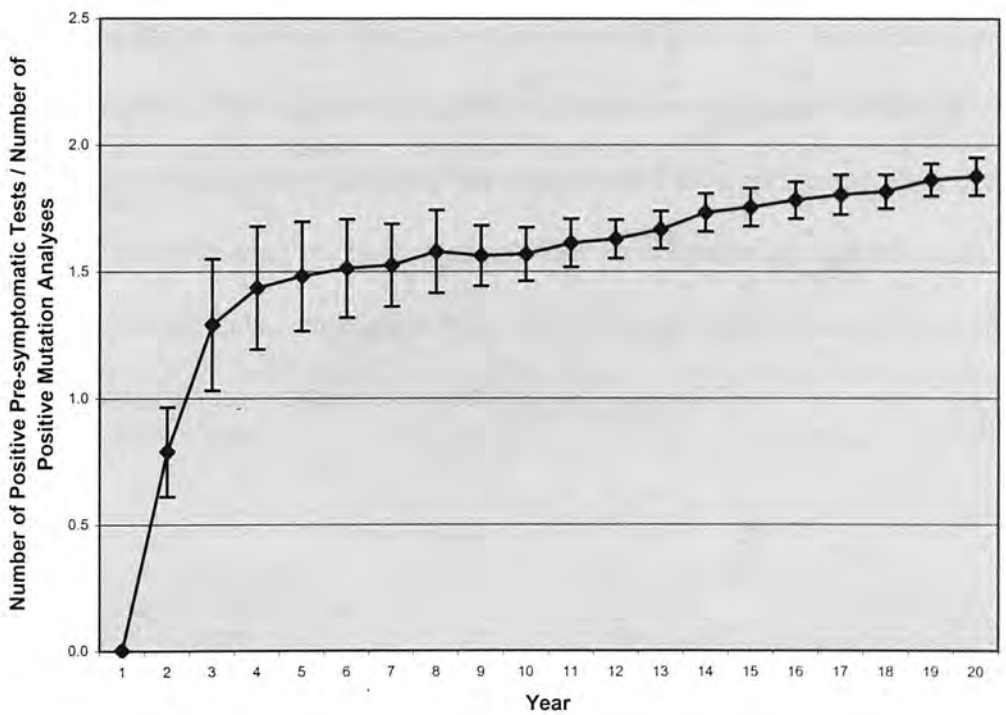
Figure 4.3.23 Predicted Number of Mutation Carriers Identified at Default Settings



The computer model indicated that an average of 79.9 (95% CI = 74.8, 85.0) index cases (i.e. colorectal cancer patients who harbour a pathogenic mismatch repair gene mutation) would be identified after 10 years of cascade genetic testing, with this number increasing to 171.5 (95% CI = 163.6, 179.4) after twenty years. The number of relatives identified as carriers was predicted to average 124.9 (95% CI = 115.8, 134.0) after ten years and 321.2 (95% CI = 305.3, 337.1) after twenty years. This is the most crucial outcome, as the identification of asymptomatic mutation carriers is the primary goal of cascade genetic testing. In total, the model indicates that the number of carriers identified would average 204.8 (95% CI = 192.4, 217.2) after ten years and 492.7 (95% CI = 468.9, 516.5) carriers would be identified after 20 years.

The divergent nature of the two measurements presented in figure 4.3.23 is an important observation, as it indicates that the number of relatives (i.e. asymptomatic members of the Scottish population) identified as carriers increases more steeply over time than the number of index cases identified. The net effect of this observation is that the ratio of asymptomatic carriers identified to index cases identified is predicted to increase over time, as illustrated in figure 4.3.24.

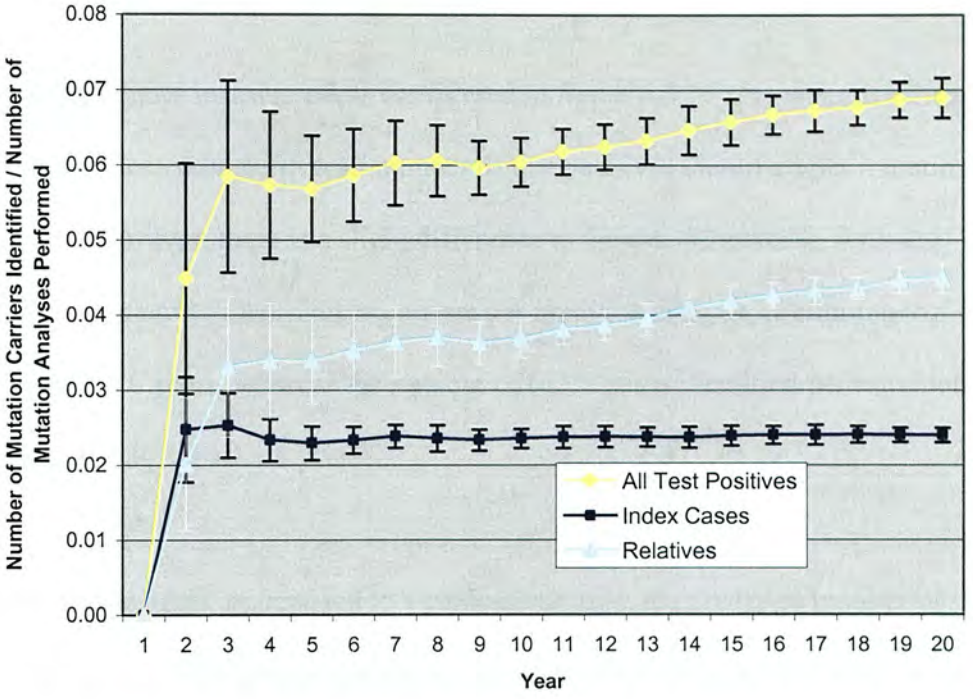
Figure 4.3.24 Predicted Ratio of Asymptomatic Carriers Identified to Index Cases Identified at Default Settings



The model estimated that an average of 1.88 (95% CI = 1.80, 1.95) asymptomatic mutation carriers would be identified per index case after 20 years.

One of the primary objectives of the computer model is to evaluate the efficiency of cascade genetic testing. A calculation of absolute efficiency would necessitate a detailed assessment of the economic aspects of the process that is beyond the scope of the current model. However, an indication of relative efficiency is readily obtained from the model outcomes. A simple and useful index of efficiency is given by comparing the number of asymptomatic MMR gene mutation carriers identified to the number of mutation analyses performed to achieve this. This approach disregards the ascertainment and testing of relatives of index cases, which in reality will form a substantial part of the work undertaken as part of the cascade genetic testing programme. However, the majority of resource is likely to be expended on mutation analysis. Furthermore, the number of mutation analyses performed is directly proportional to the number of pre-symptomatic tests performed. Hence, the number of mutation analyses performed provides an indication of total resources expended. The predicted cumulative ratio of MMR gene mutation carriers identified per mutation analyses performed is illustrated in figure 4.3.25.

Figure 4.3.25 Predicted Number of Gene Mutation Carriers Identified Per Mutation Analyses Performed at Default Settings



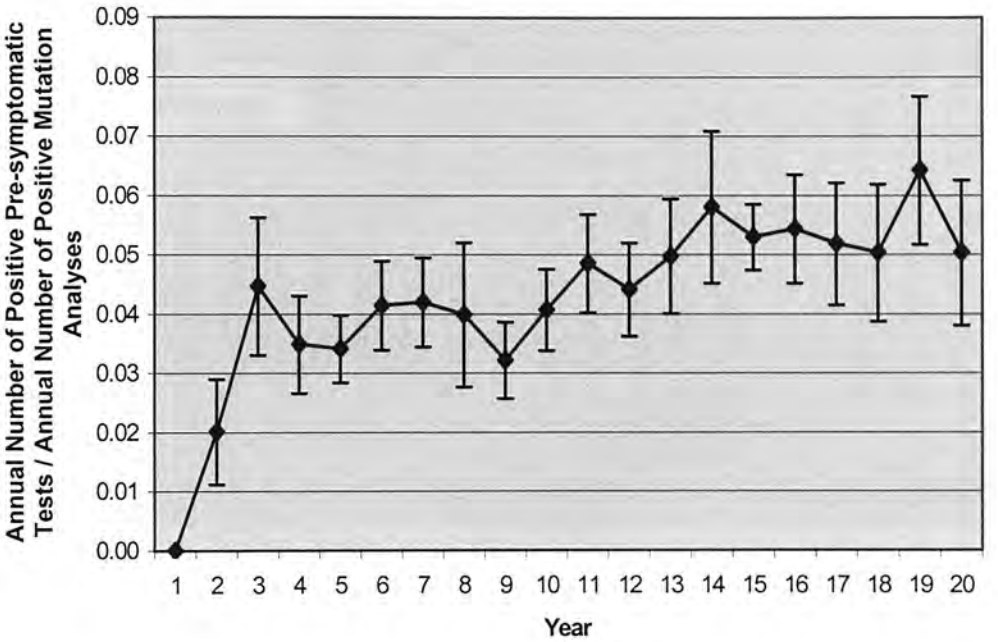
The number of index cases identified per mutation analysis provides an indication of the capacity of the chosen ascertainment criteria, in this model based on the age at colorectal cancer diagnosis, to target carriers of MMR gene mutations. As shown in figure 4.3.25, results provided by the model at default settings imply that the number of index cases identified per mutation analysis after 20 years would be 0.0240 (95% CI = 0.0230, 0.0250). This equates to approximately one positive result for every 44 mutation analyses performed. Over the entire twenty-year period the model predicted that 171.5 (95% CI = 163.6, 179.4) index cases would identified, and 7144 (95% CI = 7037, 7250) mutation analyses would be performed. As illustrated in figure 4.3.25, the cumulative ratio of asymptomatic MMR gene mutation carriers identified per mutation analyses performed was predicted to be 0.0339 (95% CI =

0.0282, 0.0395), 0.0369 (95% CI = 0.0343, 0.0394) and 0.0449 (95% CI = 0.0429, 0.0469) after 5, 10 and 20 years respectively.

The plots of both measurements considered in figure 4.3.25 take a similar shape, which is reflected in the overall number of test positives identified per mutation analysis. However, there is a slight difference in the rate of increase, with the number of relatives identified as carriers per mutation analysis continuing to increase to 20 years, whereas the number of index cases identified per mutation analysis levels out after 5 years.

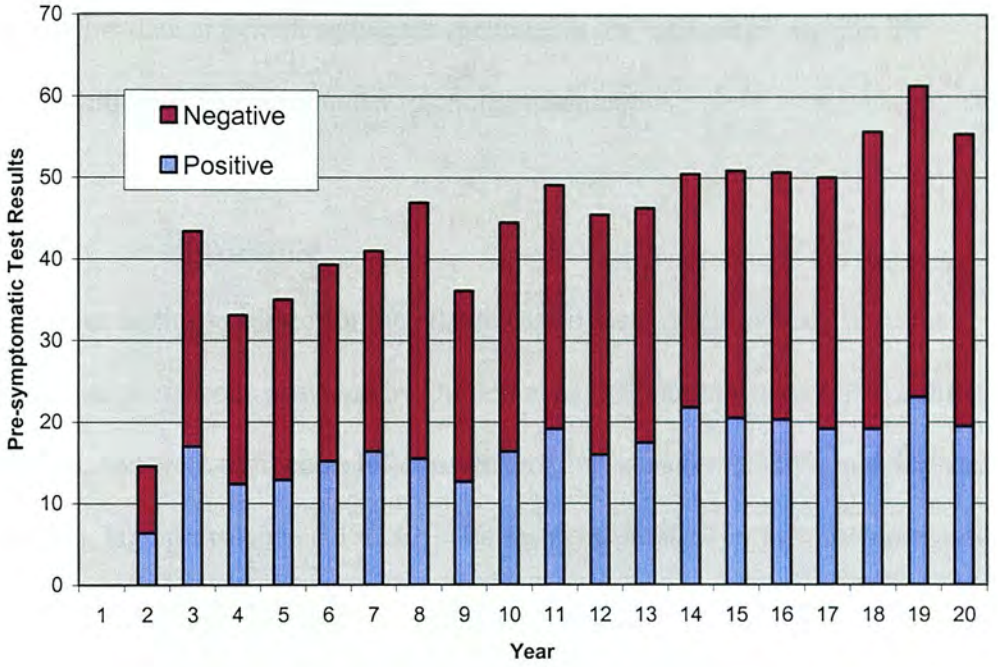
On an annual basis, as opposed to a cumulative total, the predicted number of asymptomatic MMR gene mutation carriers identified each year per number of mutation analyses performed in that year was as illustrated in figure 4.3.26.

Figure 4.3.26 Predicted Annual Number of Asymptomatic MMR Gene Mutation Carriers Identified Per Mutation Analysis Performed at Default Settings



One of the potential advantages of cascade genetic testing is the capacity to limit the number of asymptomatic non-carriers who undergo genetic testing, whilst maximising the number of asymptomatic carriers identified. Accordingly, the proportion of people undergoing a pre-symptomatic test who actually have a MMR gene mutation is of considerable relevance. This proportion, according to the predictions of the computer model, is illustrated below.

Figure 4.3.27 Predicted Annual Pre-symptomatic Test Results at Default Settings



After 20 years, the overall proportion of pre-symptomatic tests with a positive result was 0.38 (95% CI = 0.37, 0.39).

The average number of interviews conducted by a genetic nurse as part of a cascade genetic testing programme at default settings was predicted to be 2197 (95% CI = 2141, 2253) after five years, 4608 (95%CI = 4514, 4702) after ten years and 9349 (95% CI = 9174, 9523) after twenty years. Hence the average number of genetic nurse interviews conducted each year would be around 468.

Default settings for resources are deliberately set at very high levels (1000 events per week) so that there are effectively no restrictions on the model system due to limited resources. Consequently, the number of people on a waiting list at default

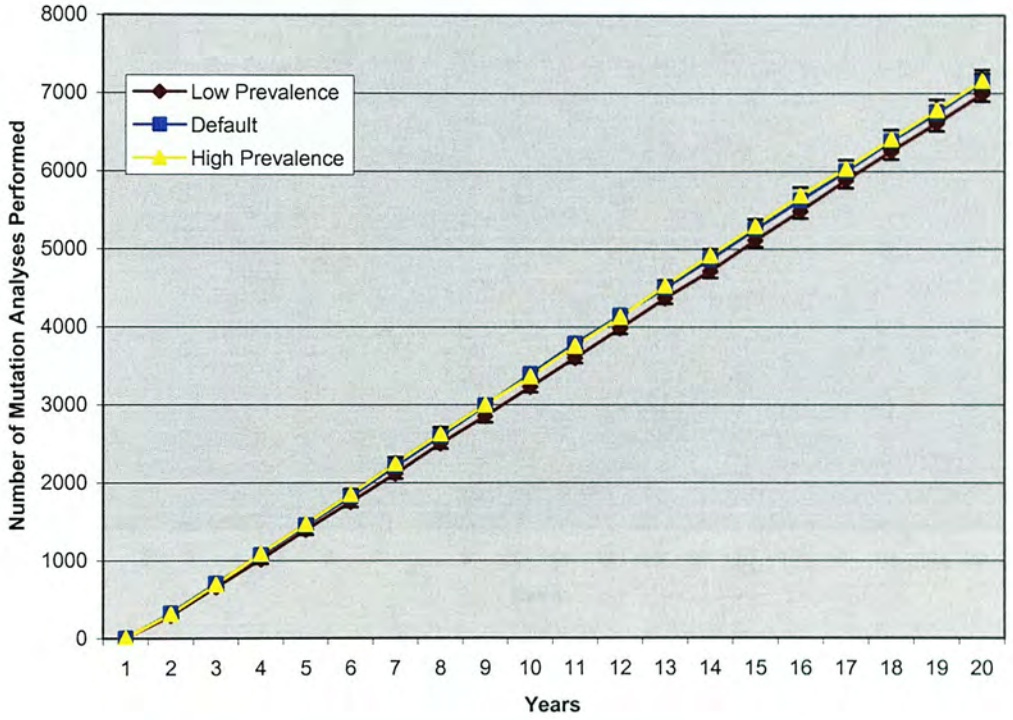
settings will simply be a reflection of the number of people who are eligible for a particular stage of the model, and the delay that is built into the model at that stage. Waiting list data at default settings is included in the 'resources' section for comparison with outcomes under restricted resources.

4.3.5.2 Prevalence

Prevalence settings utilised for model evaluation were taken directly from the estimate of prevalence provided by Dunlop et al., (58) and the associated confidence intervals, and were defined as follows: default prevalence = 1:3139, low prevalence = 1:7626, high prevalence = 1:1247. The results generated by the computer model at these low, default and high prevalence settings are presented below, and form the basis of an evaluation of the effect of prevalence on model outcomes.

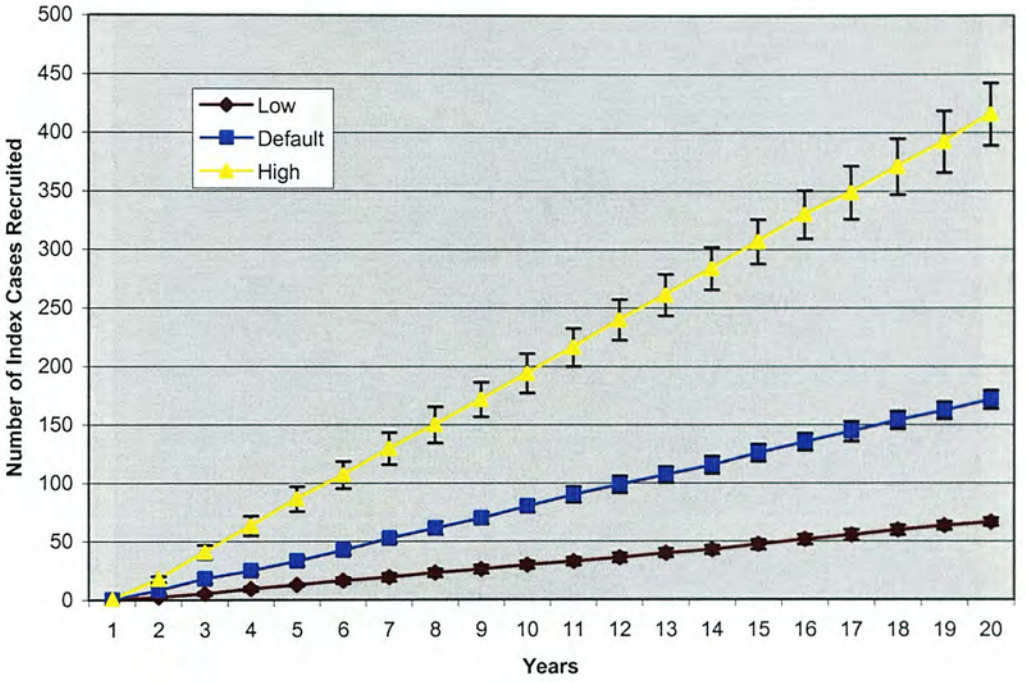
Even when the prevalence of MMR gene mutations is estimated to be relatively high, the vast majority of colorectal cancer cases that are likely to occur in the Scottish population will be sporadic in origin. Hence, prevalence would be expected to have only a minimal impact on the total number of mutation analyses performed during the modelled cascade genetic testing process. The model, as illustrated in figure 4.3.28, upholds this expectation.

Figure 4.3.28 Predicted Number of Mutation Analyses Performed at Low, Default and High Prevalence



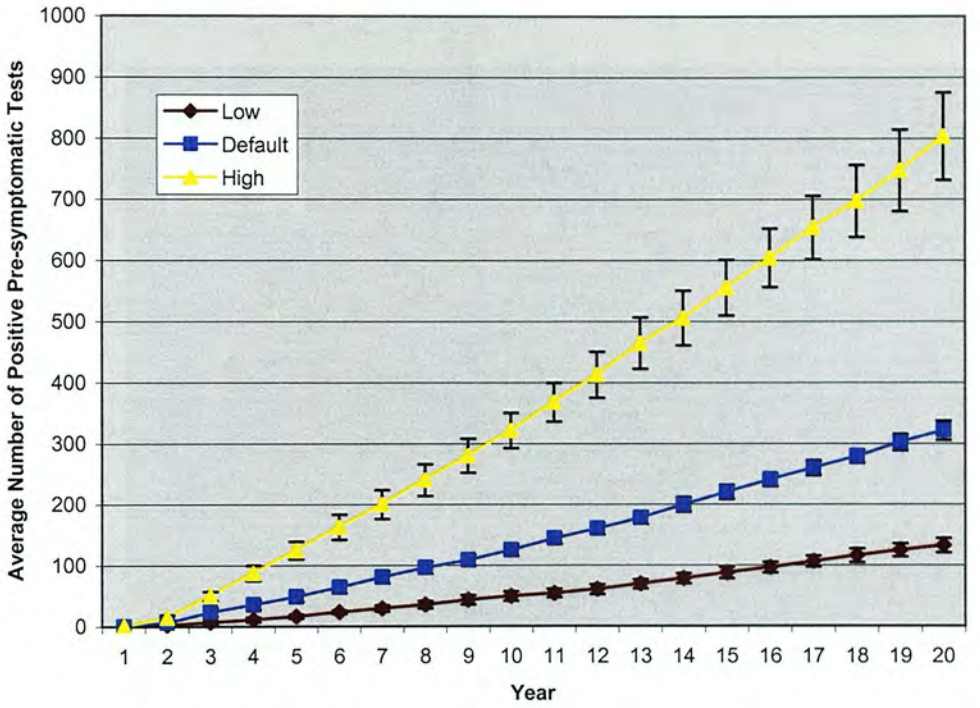
Conversely, the proportion of MMR gene mutation carriers amongst the colorectal cancer cases that present to the cascade genetic testing system will increase with the prevalence of such mutations in the population. Figure 4.3.29 illustrates the effect, according to the model, of various prevalence estimates on the number of index cases identified. Y-error bars displaying 95% confidence intervals are shown for all estimates in this figure.

Figure 4.3.29 Predicted Number of Index Cases Recruited at Low, Default and High Prevalence



The number of relatives identified as MMR gene mutation carriers is directly related to the number of pre-symptomatic tests performed, which in turn is related to the number of index cases identified. Therefore, the pattern observed in figure above is reflected in figure 4.3.30, which shows the actual number of relatives identified as carriers.

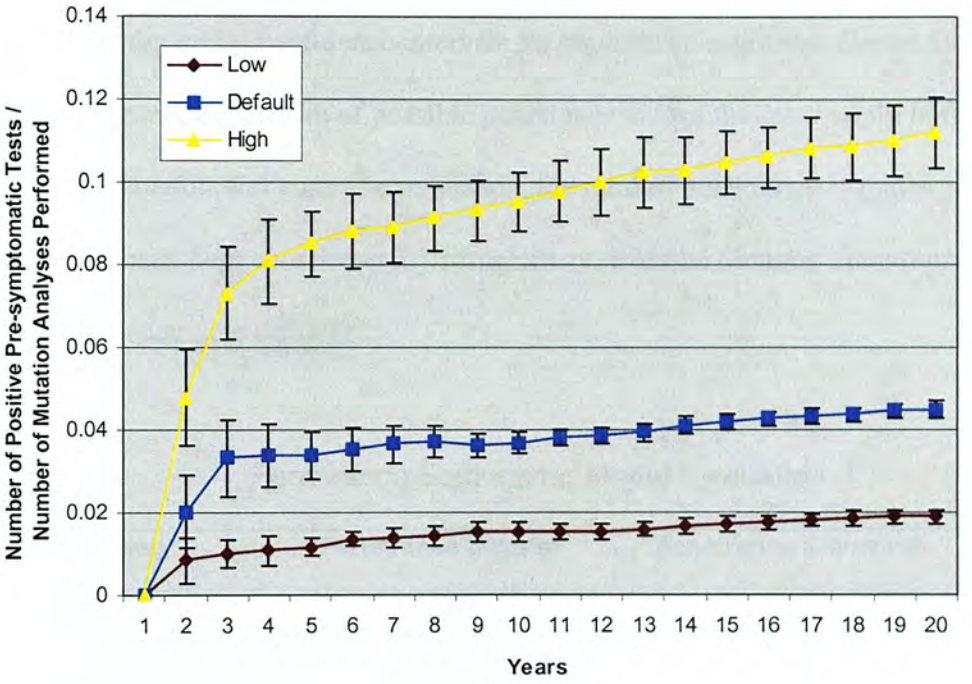
Figure 4.3.30 Predicted Number of Asymptomatic Mutation Carriers Identified at Low, Default and High Prevalence



The model predicts that the total number of asymptomatic mutation carriers identified through cascade genetic testing will vary considerably according to the prevalence of such mutations. After twenty years, the number of asymptomatic mutation carriers identified at low, default and high prevalence respectively were predicted to be 132.5 (95%CI = 120.7, 144.3), 321.2 (95% CI = 305.3, 337.1) and 803.4 (95% CI = 731.8, 875.0).

As considered previously, an insight into the relative efficiency of cascade genetic testing can be obtained by considering the ratio of asymptomatic MMR gene mutation carriers identified to mutation analyses performed. This data, relating to each prevalence estimate, is shown in figure 4.3.31.

Figure 4.3.31 Predicted Number of Asymptomatic Mutation Carriers Identified Per Mutation Analysis Performed at Low, Default and High Prevalence



A similar pattern is evident at all prevalence levels considered in figure 4.3.31, with the number of positive pre-symptomatic tests per mutation analysis performed increasing rapidly from zero over the first two or three years, and then increasing at a diminishing rate (i.e. levelling off) over the remainder of the modelled cascade genetic testing programme. The predicted number of asymptomatic mutation carriers identified per mutation analysis performed after 20 years was 0.019 (95% CI = 0.017, 0.020) at low prevalence, 0.045 (95% CI = 0.043, 0.047) at default settings, and 0.112 (95% CI = 0.104, 0.120) at high prevalence.

4.3.5.3 Penetrance

To investigate the effect of penetrance on the model system, the model was run ten times at each of the various penetrance settings, listed in table 4.3.6. There are no reliable data regarding confidence intervals for penetrance estimates. Hence these settings represent a spectrum of possible penetrance values that may apply in the real-world situation. For ease of comparison, the ratio of penetrance in males and females has been kept constant at 2:1 respectively, with the obvious exception of the 'complete' penetrance estimate.

Table 4.3.6 Penetrance Settings for Model Evaluation

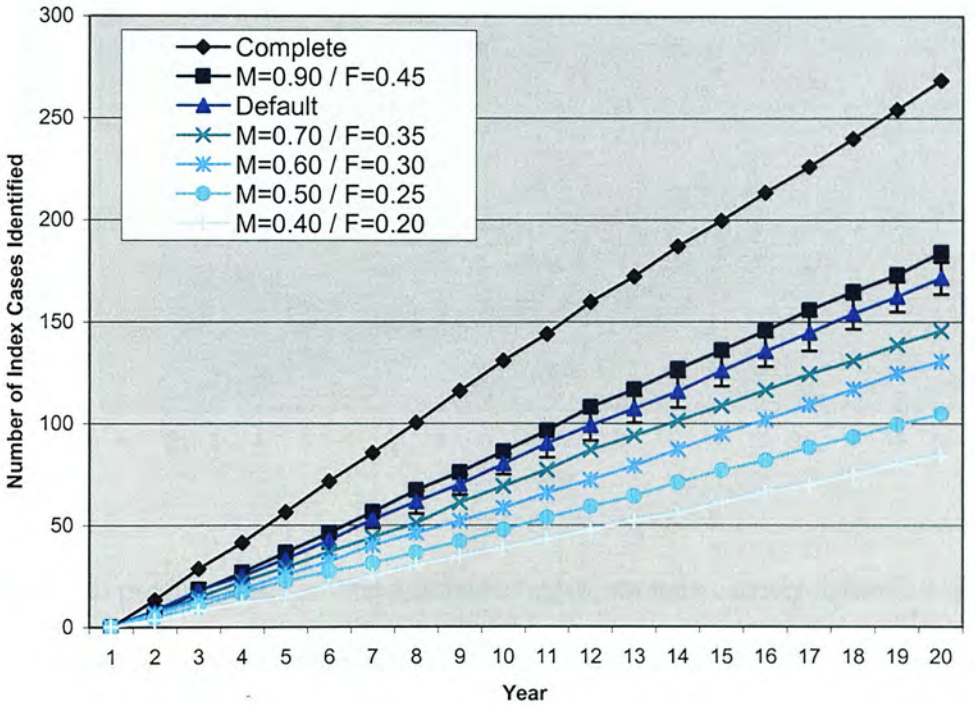
Setting	Penetrance (Males)	Penetrance (Females)
M=0.40 / F=0.20	0.40	0.20
M=0.50 / F=0.25	0.50	0.25
M=0.60 / F=0.30	0.60	0.30
M=0.70 / F=0.35	0.70	0.35
Default	0.80	0.40
M=0.90 / F=0.45	0.90	0.45
Complete	1.0	1.0

The penetrance of MMR gene mutations directly influences the number of carriers expected to develop colorectal cancer each year. However, the fact that prevalence of such mutations is estimated to be low (58) means that, even at high penetrance, mutation carriers will constitute a small proportion of all colorectal cancer cases. Consequently, different penetrance settings had no significant effect on model predictions of the total number of mutation analyses performed (data not shown).

The proportion of colorectal cancer cases which have a MMR gene mutation would be expected to increase in line with increasing penetrance, as shown in figure 4.3.32.

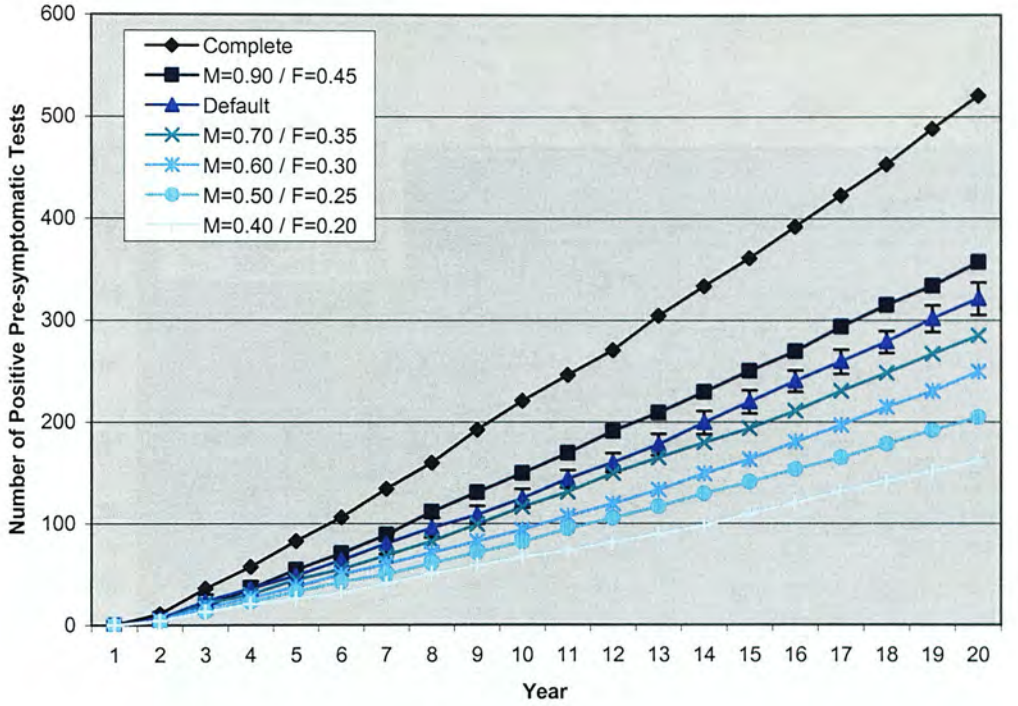
For clarity, in this figure, and subsequent figures in which numerous results are presented, confidence intervals are shown only for default settings.

Figure 4.3.32 Predicted Number of Index Cases Identified at Various Penetrance



As before, since the average number of relatives identified as carriers per index case remains constant (although individual measurements can vary at random), the effect of varying penetrance on the number of relatives identified as mutation carriers will follow a similar pattern to that for index cases, as shown below.

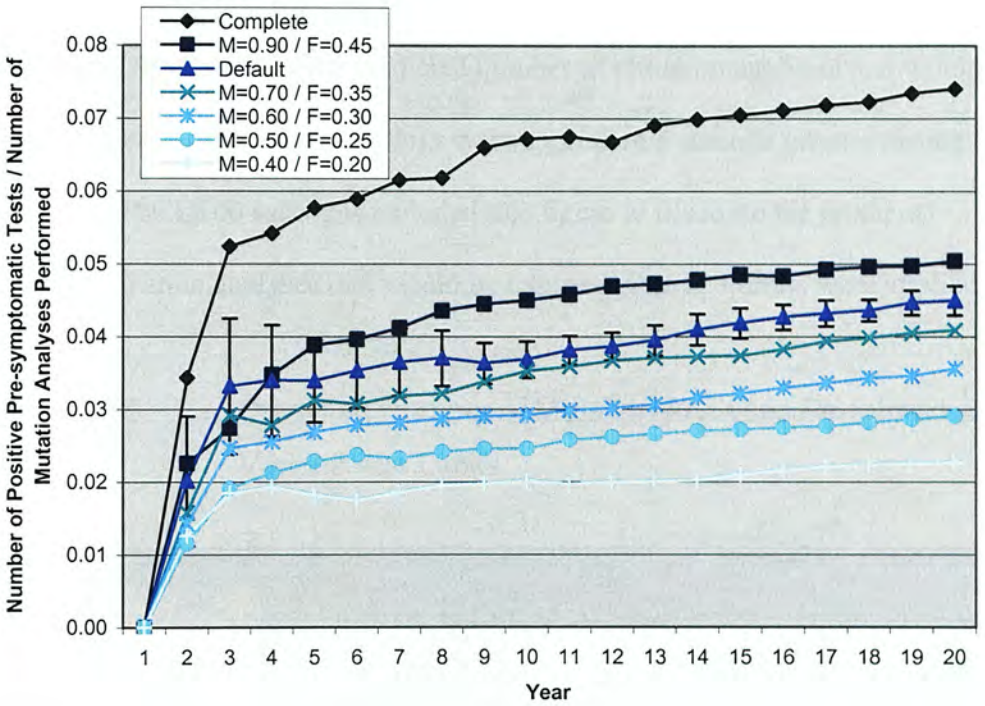
Figure 4.3.33 Predicted Number of Asymptomatic Mutation Carriers Identified at Various Penetrance



The model predicts that the total number of asymptomatic carriers identified if the default estimate of penetrance applied would be 321.2 (95% CI 305.3, 337.1). If penetrance was complete, 521.1 (95% CI = 502.7, 539.5) asymptomatic carriers were predicted to be identified in this time. At the opposite end of the spectrum, with penetrance being only 0.4 in males and 0.2 in females, this figure is predicted to be 161.6 (95% CI = 138.6, 184.7).

Figure 4.3.34 provides an illustration of the predicted effect of penetrance on the relative efficiency of cascade genetic testing. For clarity, 95% confidence intervals are only displayed for default settings in this figure.

Figure 4.3.34 Predicted Number of Asymptomatic Mutation Carriers Identified Per Mutation Analysis Performed, at Various Penetrance



As illustrated in figure 4.3.34, the computer model supports the hypothesis that the relative efficiency of cascade genetic testing is directly proportional to penetrance.

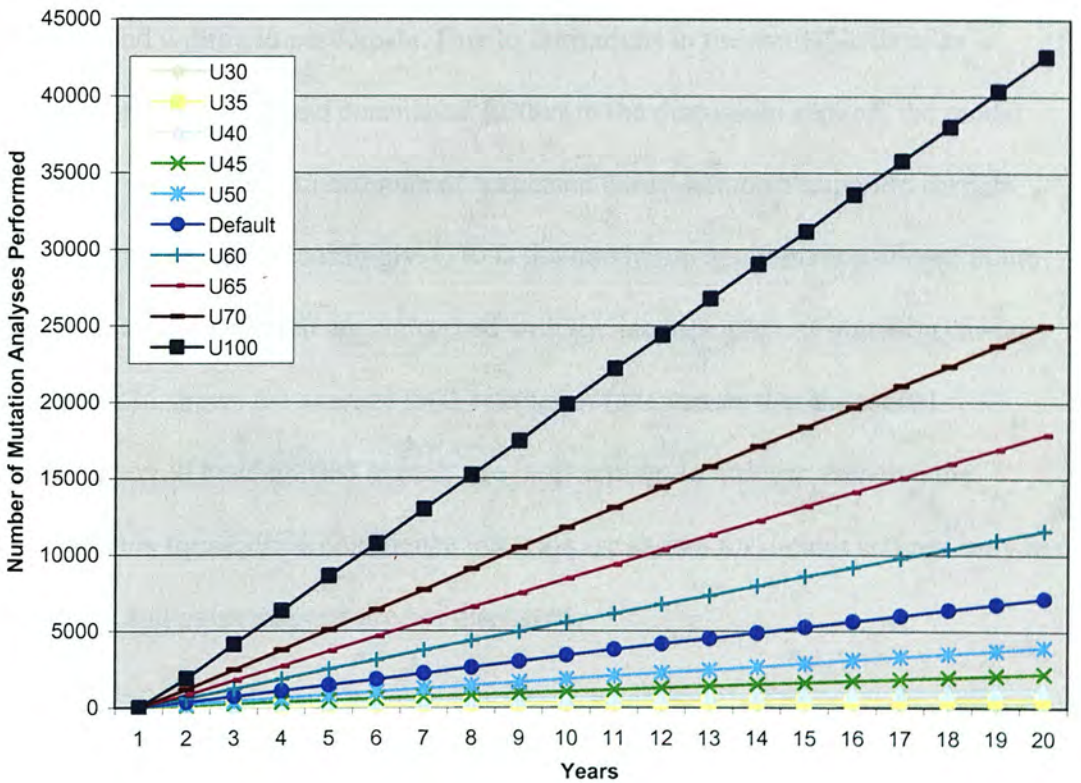
4.3.5.4 Age Limits

Sporadic colorectal cancer incidence increases rapidly with advancing age, whereas cases that are primarily hereditary in origin tend to manifest at a relatively young age. Hence, the proportion of all colorectal cancer cases that are caused by MMR gene mutations would be expected to be greater in relatively young cases, providing a useful and established means of targeting potential carriers. To illustrate this point, and to investigate the potential effect of various age limit criteria on the outcomes of

cascade genetic testing, the model was run ten times each with the maximum age limit set to 30, 35, 40, 45, 50, 55 (default), 60, 70 and 100 years. Throughout this chapter, these settings are referred to as "under 'x' years" or "Ux".

Figure 4.3.35 demonstrates the predicted number of mutation analyses that would be performed if each of the above settings were applied in a cascade genetic testing programme. The U100 setting is included this figure to illustrate the predicted number of mutation analyses that would be required if no age limits were applied.

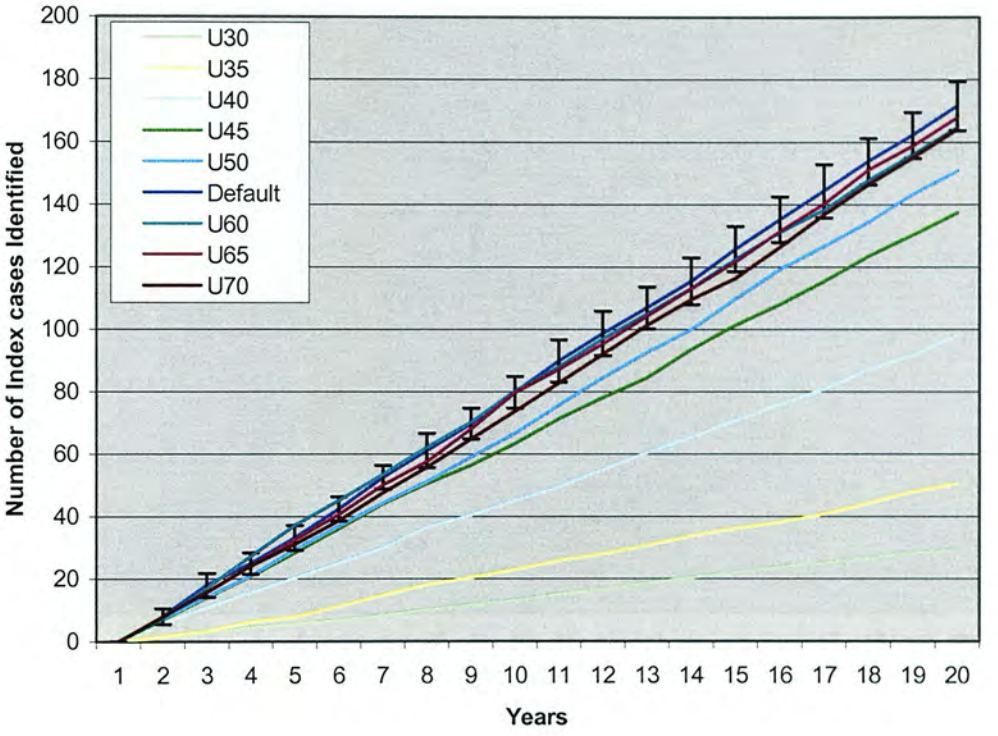
Figure 4.3.35 Predicted Number of Mutation Analyses Performed at Various Age Limits



At the U100 setting the model predicted that an average of 42574 (95% CI = 42007, 43140) mutation analyses would be performed over the twenty-year period. In contrast, when the maximum age limit was set to 30 years, only 199 (95% CI = 190, 208) mutation analyses were predicted to be performed over the same time period. The underlying hypothesis that lowering the age limit will improve the efficiency with which index cases are identified, at the expense of reducing the overall yield of the programme is explored below.

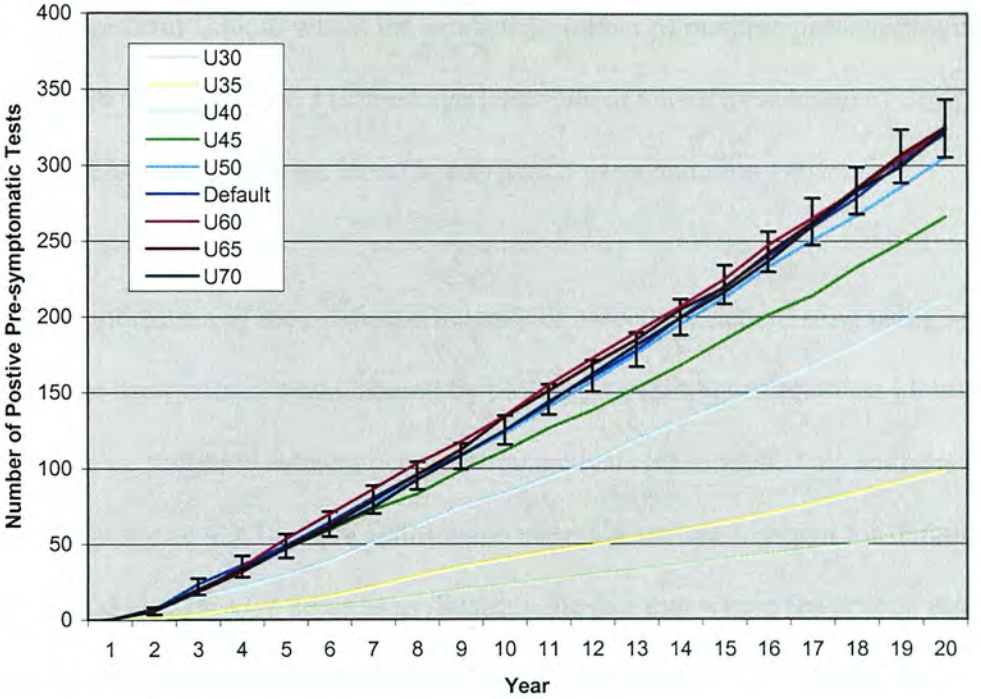
In practical terms, cases developing after age 70 are unlikely to be included in a genetic testing programme, due to the fact that only a very small proportion of such cases are likely to harbour MMR gene mutations, and that elderly patients may be less able and willing to participate. Due to limitations in the available data, as indicated in table 4.3.2 and considered further in the discussion section, the model itself does not generate an estimate of 'expected cases' amongst mutation carriers over 70 years of age. Accordingly, U70 is the maximum age limit considered in the subsequent figures, which are concerned with the identification of mutation carriers. Figure 4.3.36 shows the average total number of index cases that the model predicted would be identified at each age limit setting. In order to preserve the clarity of this figure, 95% confidence intervals are shown for default settings only in this figure and point markers are not displayed.

Figure 4.3.36 Predicted Number of Index Cases Identified at Various Age Limits



The most striking feature of the above figure is the clustering of estimated index cases identified for all age limits above 55 years. Indeed, values for the U60, U65 and U70 settings generally fall within the 95% confidence intervals for the results at default (U55) settings. Predictions of the number of asymptomatic MMR gene mutation carriers identified at various age limits follow a similar pattern, as illustrated in figure 4.3.37. Again, 95% confidence intervals are shown for default settings only in this figure and point markers are not displayed.

Figure 4.3.37 Predicted Number of Asymptomatic Mutation Carriers Identified at Different Age Limits

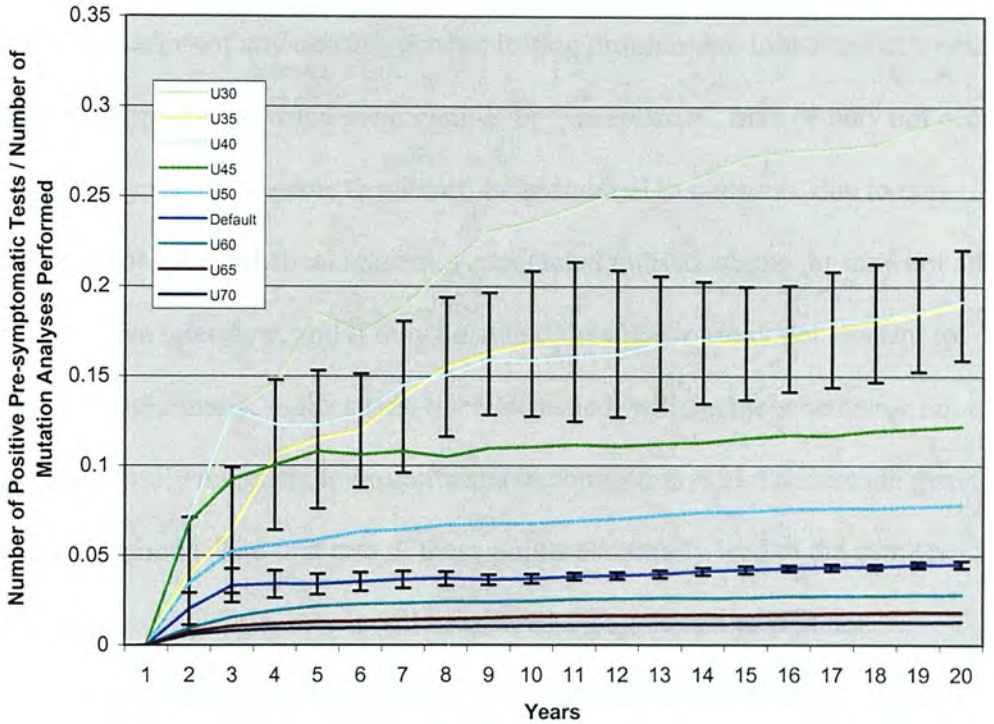


The above figure shows that cascade genetic testing with age limits of U60, U65 and U70 are predicted to yield a total number of positive pre-symptomatic tests that is not significantly different from the yield at default (U55) settings. Although a statistically significant difference may become evident based on a larger number of simulations (runs), this finding implies that applying maximum age limits for recruitment of index cases of greater than 55 years would have only a minor effect in terms of increasing yield. For almost all of the twenty year 'run' period of the model, the predicted number of positive pre-symptomatic tests at the U50 setting are also within the confidence intervals for default settings, suggesting that reducing the default age limits by five years may have limited impact on yield. In contrast, setting the model age limits to just 30 or 35 results in a substantial reduction in predicted

yield, with less than a third of the number of asymptomatic mutation carriers being identified compared to default settings. A large increase in overall yield is evident between age limit U35, at which the predicted number of positive pre-symptomatic tests was 98 (95% CI = 84, 112) and age limit U40 at which an average of 213 (95% CI = 194, 232) relatives were identified as MMR gene mutation carriers.

Again, an indication of the relative efficiency of cascade genetic testing using various age limit criteria was obtained by plotting the number of relatives identified as MMR gene mutation carriers per mutation analysis performed. This analysis is presented in figure 4.3.38. 95% Confidence intervals are again shown for default settings, and also for U35 settings to illustrate the fact that where fewer total tests have been performed confidence intervals are relatively wide. For clarity, point markers and other confidence intervals are not displayed.

Figure 4.3.38 Predicted Number of Asymptomatic Mutation Carriers Identified Per Mutation Analysis Performed, at Various Age Limits



At the U30 setting, the average number of asymptomatic carriers identified per mutation analysis performed reached 0.2940 (95% CI = 0.2433, 0.3447) after twenty years, implying that only around four mutation analyses would be required to identify one asymptomatic mutation carrier if such criteria were implemented. This high relative efficiency is inevitably offset by a low overall yield, as illustrated in figure 4.3.37. At the other extreme, the model suggests that if no maximum age limit were applied to colorectal cancer cases in a cascade genetic testing program, as represented by the U100 setting, the average number of asymptomatic carriers identified per mutation analysis performed would be just 0.0076 (95% CI = 0.0071, 0.0080).

4.3.5.5 Acceptance Rates

The extent to which eligible individuals participate is a crucial issue consideration in terms of the utility of any cascade genetic testing programme. In the model system, there are four points at which participation, or "acceptance", may or may not occur. Firstly, it may not be possible to contact the individual in question, due to non-cooperation or other practical reasons. A contacted individual may or may not attend a genetic nurse interview, and if they do attend they may or may not consent to genetic testing. Finally, index cases, once identified, will decide whether or not to allow their family members to be contacted in connection with the cascade genetic testing programme. The first two of these points essentially lead to the same result; either the individual in question will be offered a genetic test or will not. A conservative estimate of 0.9 is applied to the probability of acceptance at each of these points at default settings. With the exception of the "acceptance 100" setting, at which 100% participation is assumed at all stages, these two settings have been kept constant throughout the following analysis to facilitate an investigation of the effect on model outcomes of participation at the 'genetic testing' and 'allowing contact with relatives' stages. The acceptance settings used as model input in this investigation are shown in table 4.3.7.

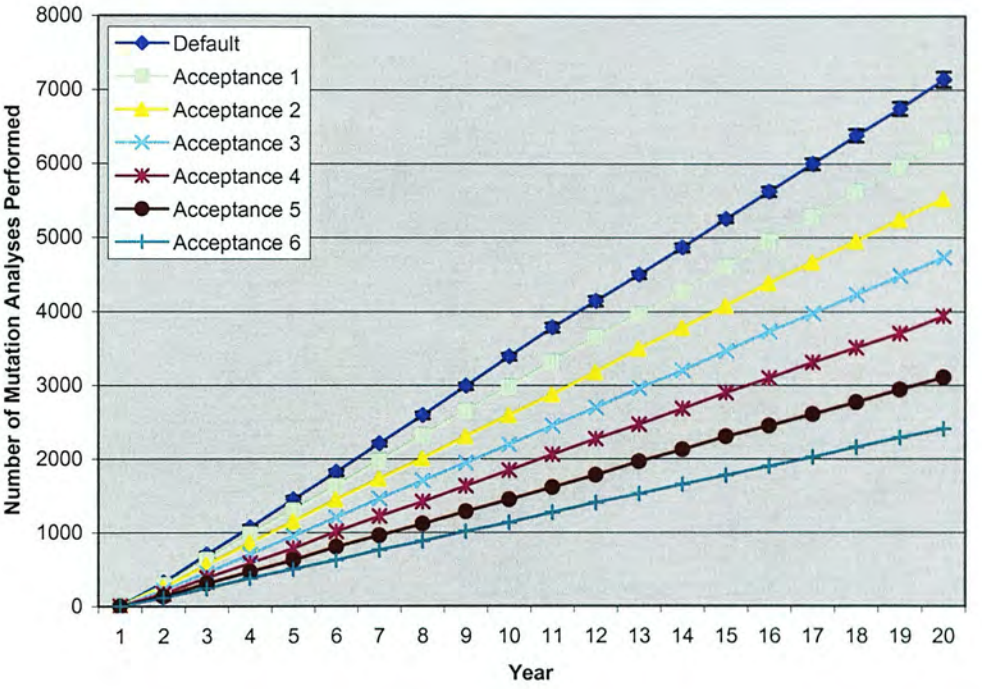
Table 4.3.7 Acceptance Settings for Model Evaluation

Setting	Probability of Contact	Probability of Attending Genetic Nurse Interview	Probability of Accepting Test	Probability of Allowing Contact With Relatives
Acceptance100	1.00	1.00	1.00	1.00
Default	0.90	0.90	0.90	0.90
Acceptance 1	0.90	0.90	0.80	0.90
Acceptance 2	0.90	0.90	0.70	0.90
Acceptance 3	0.90	0.90	0.60	0.90
Acceptance 4	0.90	0.90	0.50	0.90
Acceptance 5	0.90	0.90	0.40	0.90
Acceptance 6	0.90	0.90	0.30	0.90
Acceptance 7	0.90	0.90	0.90	0.80
Acceptance 8	0.90	0.90	0.90	0.70
Acceptance 9	0.90	0.90	0.90	0.60
Acceptance 10	0.90	0.90	0.90	0.50
Acceptance 11	0.90	0.90	0.50	0.50
Acceptance 12	0.90	0.90	0.9 Patient, 0.5 FDR/SDR	0.90
Acceptance 13	0.90	0.90	0.9 Patient/FDR, 0.5 SDR	0.90
Acceptance 14	0.90	0.90	0.5 Patient, 0.9 FDR/SDR	1.90

Abbreviations: FDR, first degree relative; SDR, second degree relative

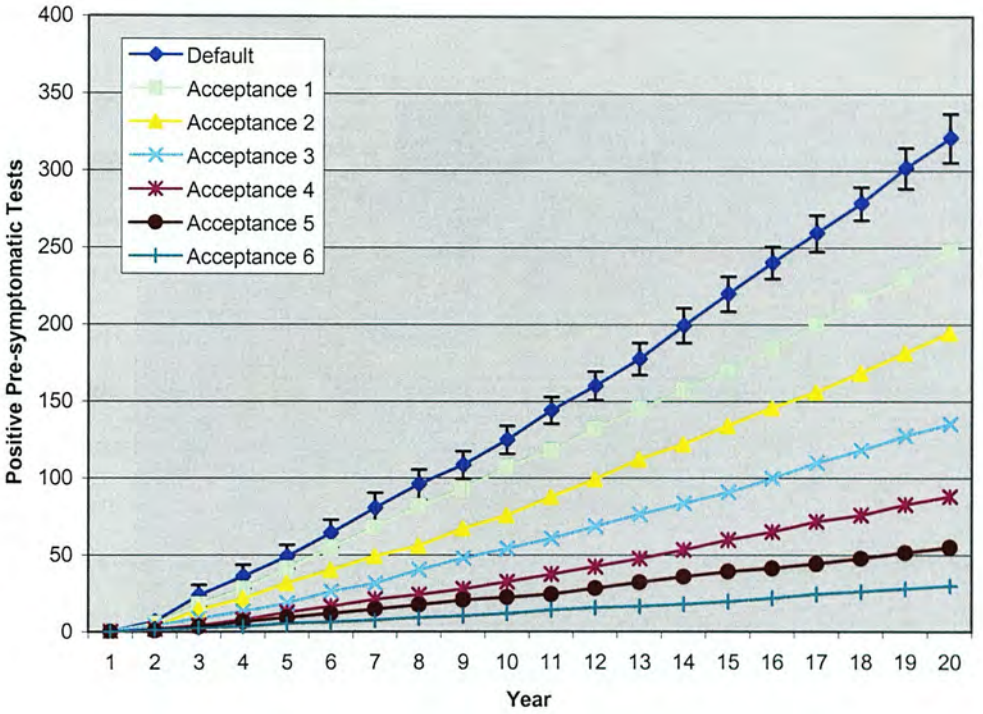
Model outcomes relating to the predicted number of mutation analyses that would be performed at various acceptance settings for genetic testing, with other acceptance probabilities remaining constant at 0.9, are presented in figure 4.3.39.

Figure 4.3.39 Predicted Number of Mutation Analyses Performed at Various Acceptance Rates for Genetic Testing



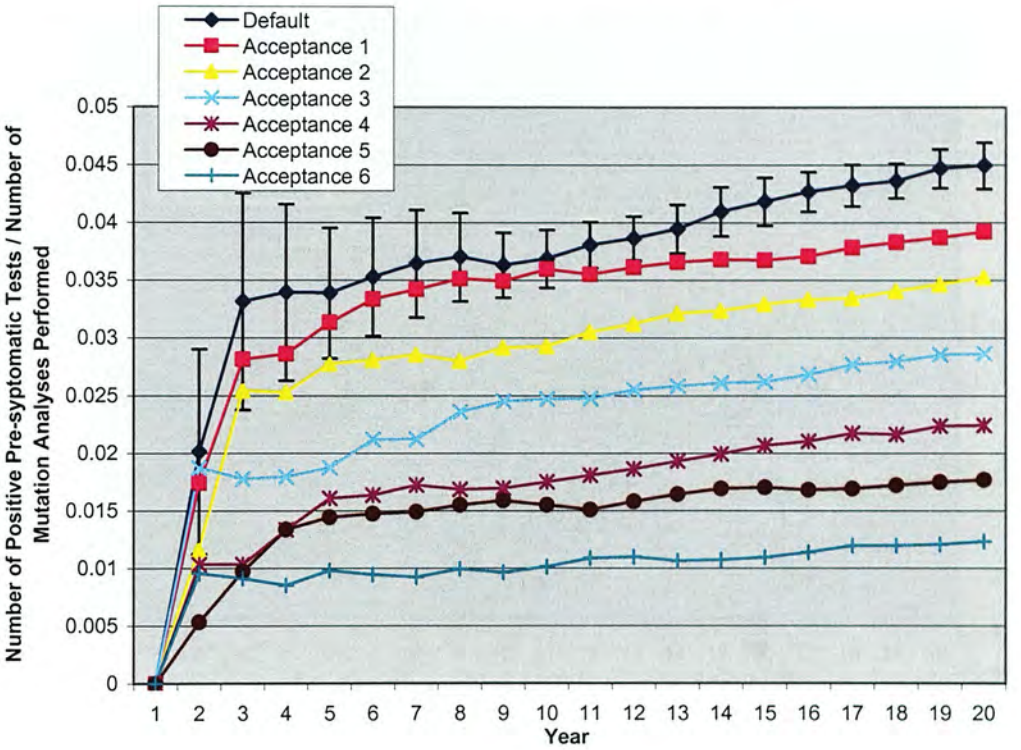
The predicted number of mutation analyses performed decreases in direct proportion to the probability that eligible patients will accept the genetic test. The predicted impact of this effect on yield from cascade genetic testing is illustrated in figure 4.3.40.

Figure 4.3.40 Predicted Number of Asymptomatic Mutation Carriers Identified at Various Acceptance Rates for Genetic Testing



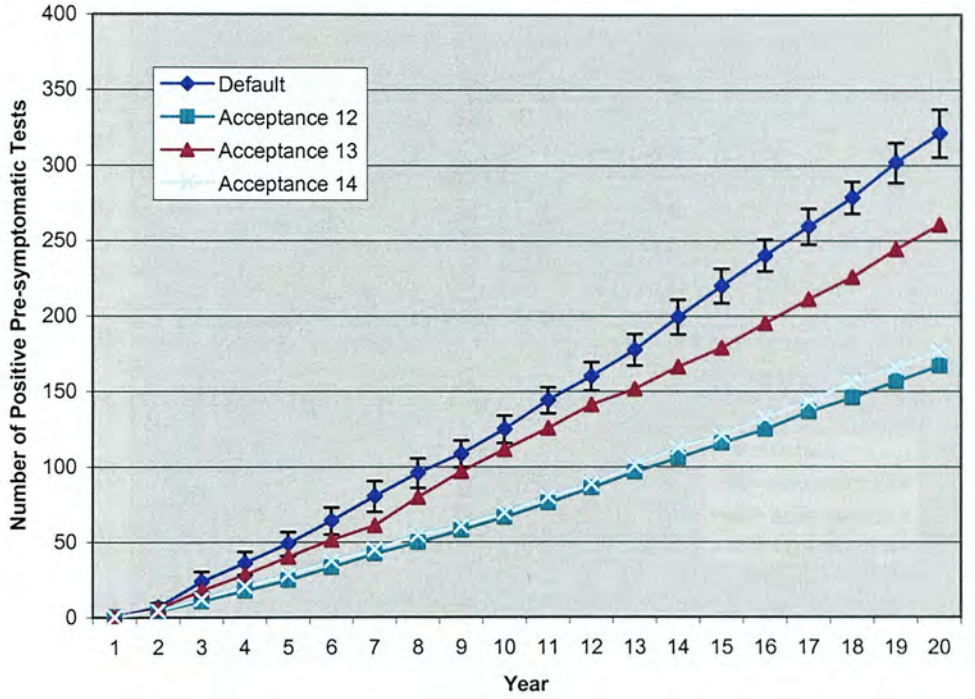
In general, the pattern observed for number of pre-symptomatic test positives is similar to that for number of mutation analyses performed. However, a linear increase in probability of test acceptance results in a subtle exponential increase in positive pre-symptomatic tests, due to the fact that this probability applies to both potential index cases and relatives of known MMR gene mutation carriers. A consequence of this feature is that cascade genetic testing is predicted to be more efficient, in relative terms, when the probability of test acceptance is high, as illustrated in figure 4.3.41.

Figure 4.3.41 Predicted Number of Asymptomatic Mutation Carriers Identified Per Mutation Analysis Performed, at Various Acceptance Rates for Genetic Testing



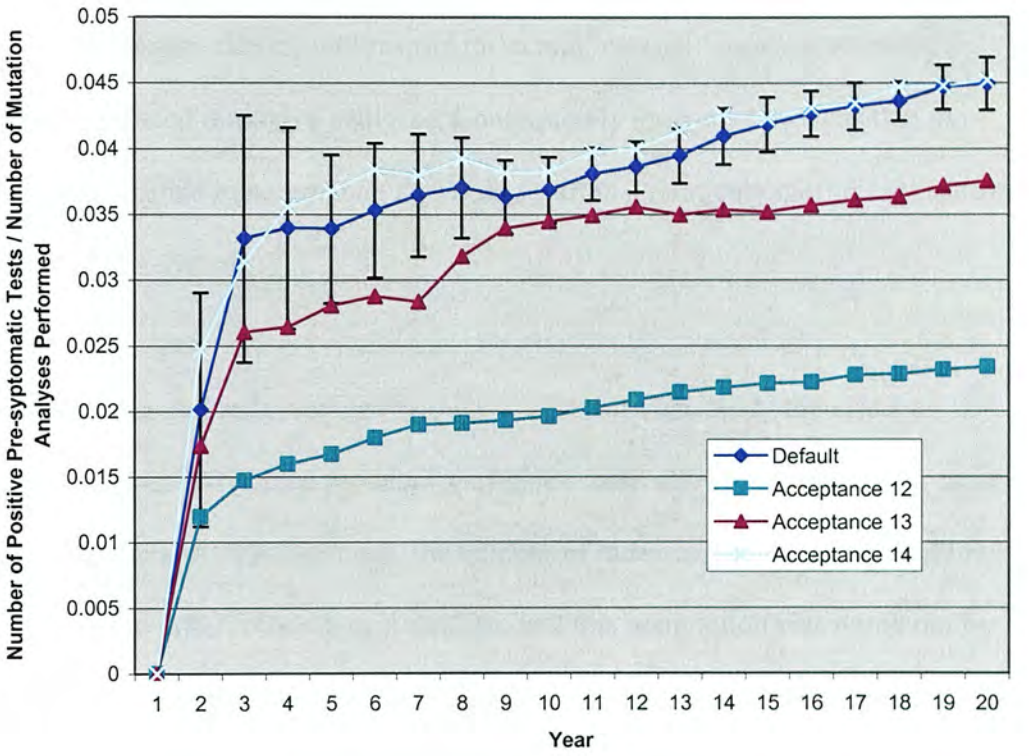
Model settings 12, 13 and 14 were designed to further explore the issue of genetic test acceptance by specifying different probabilities for undergoing the test depending on whether the individual being offered testing is a colorectal cancer patient (potential index case), a first degree relative of a known carrier, or a second degree relative of a known carrier. The predicted numbers of asymptomatic mutation carriers identified at these settings are presented in figure 4.3.42.

Figure 4.3.42 Predicted Number of Asymptomatic Mutation Carriers Identified at Additional Acceptance Rates for Genetic Testing



As shown by the close proximity of the plots of acceptance 12 and acceptance 14, the model predicts that the effect of the probability of test acceptance being limited to 0.5 for potential index cases would be very similar to the effect of the probability of test acceptance by relatives of known carriers being limited to this level. The plot of acceptance 13 indicates that if this probability was 0.9 (default) for potential index cases and first-degree relatives of known carriers, but only 0.5 for second-degree relatives, the yield would still be significantly less than at default settings. Figure 4.3.43 illustrates the relative efficiency of cascade genetic testing at these settings.

Figure 4.3.43 Predicted Number of Asymptomatic Mutation Carriers Identified Per Mutation Analysis Performed at Additional Acceptance Rates for Genetic Testing

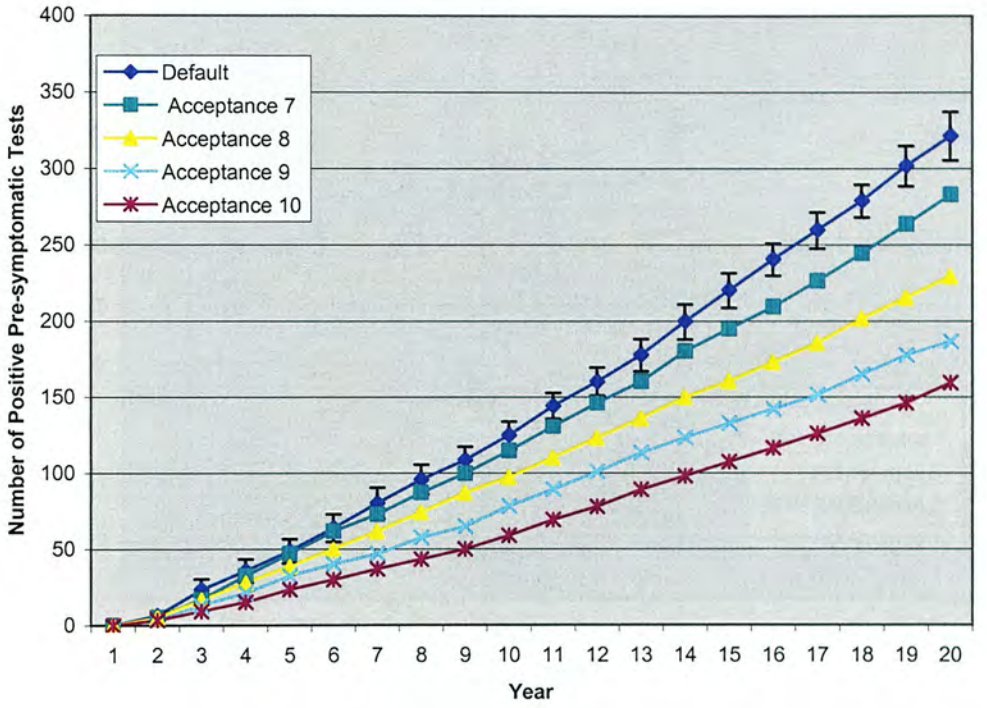


At the acceptance 13 setting, the reduced yield associated with fewer second-degree relatives accepting a genetic test is reflected as a proportional reduction in relative efficiency compared with default settings. No significant difference in relative efficiency is apparent between default settings and acceptance 14. This result would be expected, since the probability of test acceptance for relatives is the same at each of these settings. Hence, these results clearly demonstrate the intuitive point that the yield of cascade genetic testing will vary with the probability of test acceptance by potential index cases, but that efficiency in terms of laboratory resources will not be affected. There are, of course, likely to be some administration time and resources expended on colorectal cancer patients who do not accept genetic testing, so in this

respect the cascade genetic testing may be slightly more efficient when acceptance at this stage is higher. In contrast, to these findings, the acceptance 12 setting illustrates the fact that a decrease in probability of accepting genetic testing among relatives of known carriers will restrict the actual 'cascade' process, whereby a mutation is traced through a pedigree. Consequently the model predicts that the estimates specified in acceptance 12 would result in a relatively inefficient cascade genetic testing system.

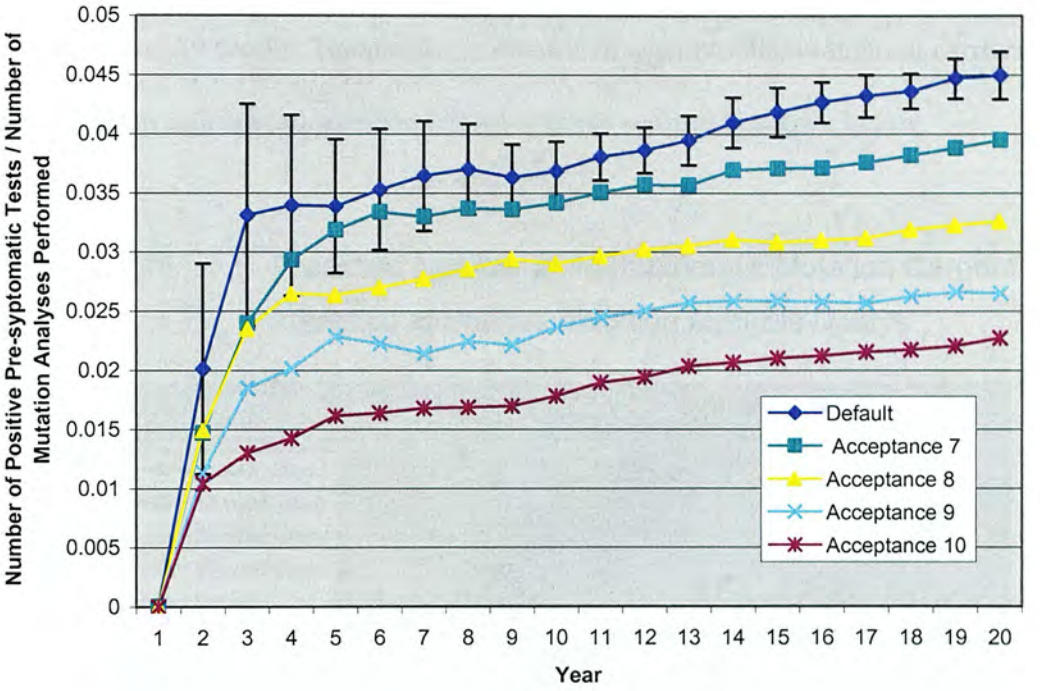
The model was run at acceptance settings 7, 8, 9 and 10 to study the effect on key model outcomes of reduced probability of index cases allowing contact with other family members. At these settings, the number of index cases identified would not be expected to differ, other than at random, and this assumption was borne out by the model results (data not shown). The average number of asymptomatic mutation carriers predicted to be identified at these settings is shown in figure 4.3.44.

Figure 4.3.44 Predicted Number of Asymptomatic Mutation Carriers Identified, at Various Acceptance Rates for Allowing Contact With Relatives



The computer model predicts that the number of asymptomatic mutation carriers identified will decrease with the probability that index cases will allow contact with their relatives. For example, at the acceptance 10 setting, when this probability is just 0.5, the yield was approximately half that at default settings. The predicted impact of probability of index cases allowing contact with their relatives on relative efficiency is shown in figure 4.3.45.

Figure 4.3.45 Predicted Number of Asymptomatic Mutation Carriers Identified Per Mutation Analysis Performed, at Various Acceptance Rates for Allowing Contact With Relatives



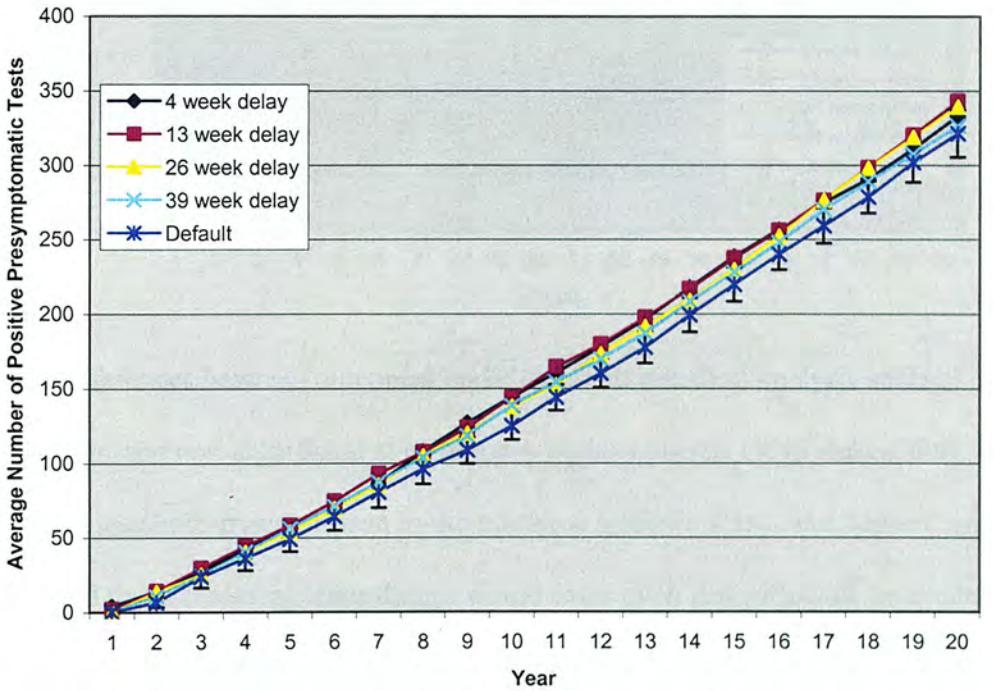
The above results indicate that cascade genetic testing becomes markedly less efficient in relative terms as the probability of mutation carriers allowing contact with their relatives decreases.

4.3.5.6 Delays

At default settings, the model has several inbuilt delays, designed to reflect the approximate time that various tasks will take in a real-life situation. The most significant delay is associated with conducting mutation analysis, and the default time for this process is 52 weeks. Intuitively, it would be expected that delays in the model system would influence the time taken to reach certain targets, but would not otherwise affect the model outcomes, and would hence have no impact on the

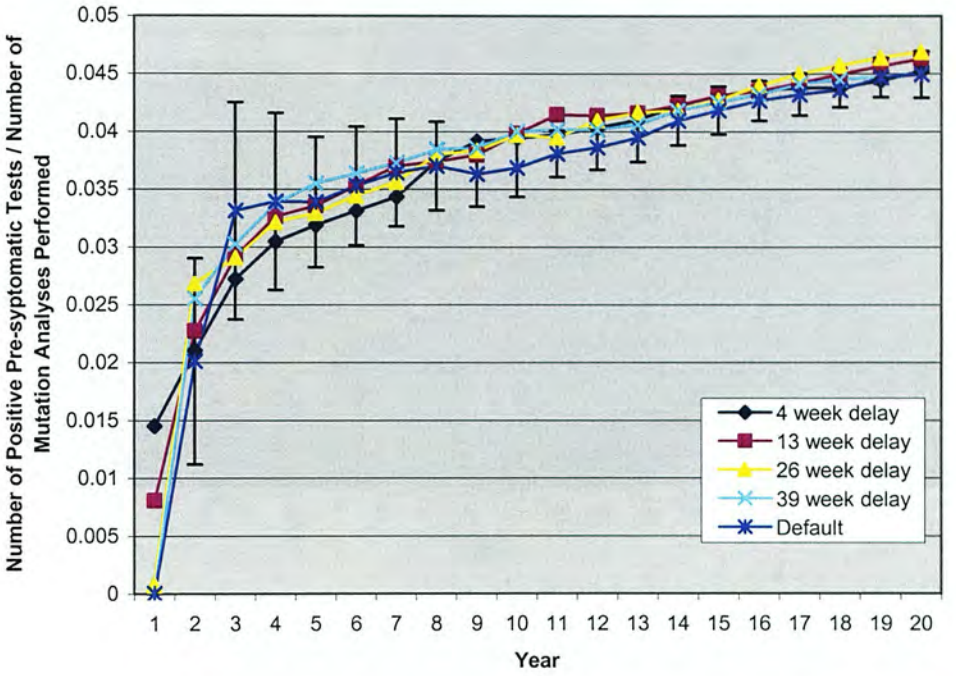
overall utility of cascade genetic testing. To test this hypothesis, the model was run ten times at each of the following mutation analysis delay settings: 4 weeks (considered the minimum realistic time for conducting mutation analysis), 13 weeks, 26 weeks, and 39 weeks. The predicted number of asymptomatic mutation carriers identified through cascade genetic testing at these settings is shown below.

Figure 4.3.46 Predicted Number of Asymptomatic Mutation Carriers Identified at Various Mutation Analysis Delays



This chart shows a slight tendency towards increased yield when mutation analysis delay settings are relatively short. However, the differences observed are minor, and are only significant at some of the annual time points at which results are recorded. The influence of such differences on the relative efficiency of the model system are presented in figure 4.3.47.

Figure 4.3.47 Predicted Number of Asymptomatic Mutation Carriers Identified Per Mutation Analysis Performed, at Various Mutation Analysis Delays



Again, differences between outcomes under different mutation analysis settings were minor, and non-significant at most measurement points. Other delays built into the model are short in comparison to the mutation analysis delay, and hence it was anticipated that decreasing these delays would exert even less influence on model outcomes. To demonstrate this, the model was run ten times with all delays except mutation analysis delay set to just one week. Finally, to determine the maximum effect of reducing delays in the model system, it was run ten times at "minimum delay" settings, with a mutation analysis delay of 4 weeks and all other delays being 1 week. Results, in terms of yield and relative efficiency, are presented in the following two figures.

Figure 4.3.48 Predicted Number of Asymptomatic Mutation Carriers Identified at Various Delays

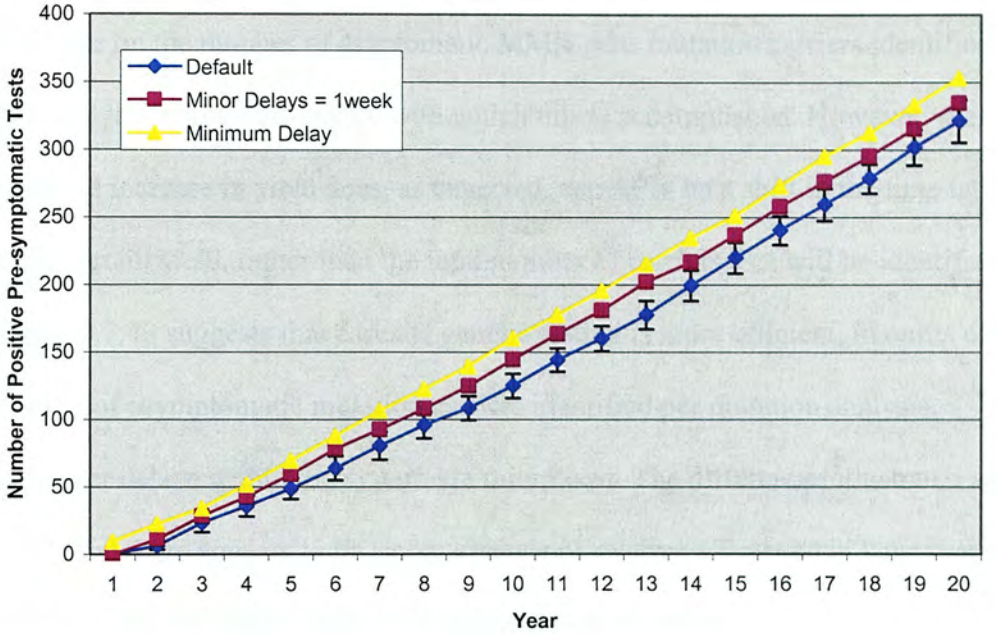
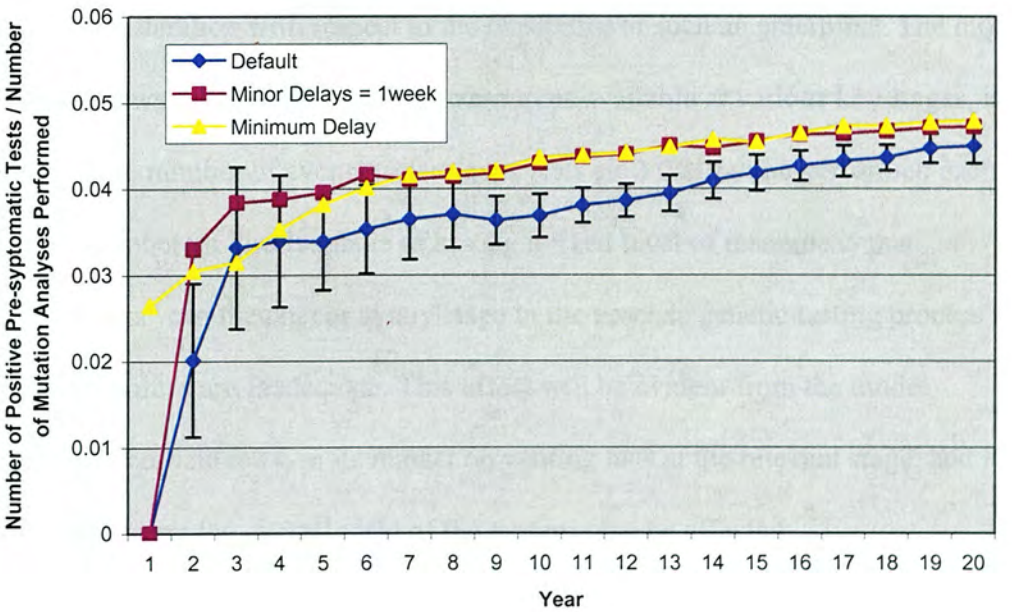


Figure 4.3.49 Predicted Number of Asymptomatic Mutation Carriers Identified Per Mutation Analysis Performed at Various Delays



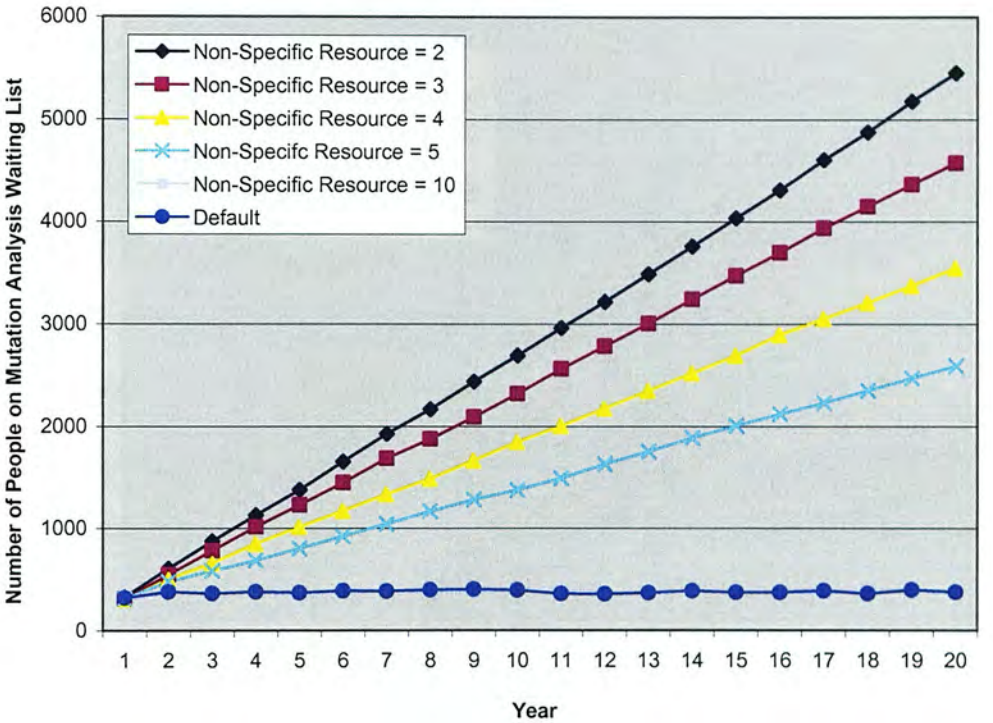
As illustrated in figure 4.3.48, the computer model predicts that minimising the delays within the cascade genetic testing system exerts a small but significant influence on the number of asyptomatic MMR gene mutation carriers identified over time and the relative efficiency with which this is accomplished. However, the observed increase in yield does, as expected, appear to be a shift in the time taken to reach certain yield, rather than the total number of carriers that will be identified. Figure 4.3.49 suggests that cascade genetic testing is more efficient, in terms of the number of asymptomatic mutation carriers identified per mutation analysis, whenever delays within the system are minimised. The differences illustrated are small, and there appears to be a convergence of relative efficiency in later years, implying that the impact in the long run would not be large.

4.3.5.7 Resources

The resources required to implement a cascade genetic testing programme are a major consideration with respect to the feasibility of such an enterprise. The model system allows the user to specify the resources available at various key stages, in terms of the number of events (interviews, tests etc.) that can be performed each week. An inherent disadvantage of having a fixed level of resource is that "bottlenecks" can then occur at any stage in the cascade genetic testing process at which resources are inadequate. This effect will be evident from the model outcomes, both directly, in its impact on waiting lists at the relevant stage, and also indirectly, since the overall yield of the system may be affected.

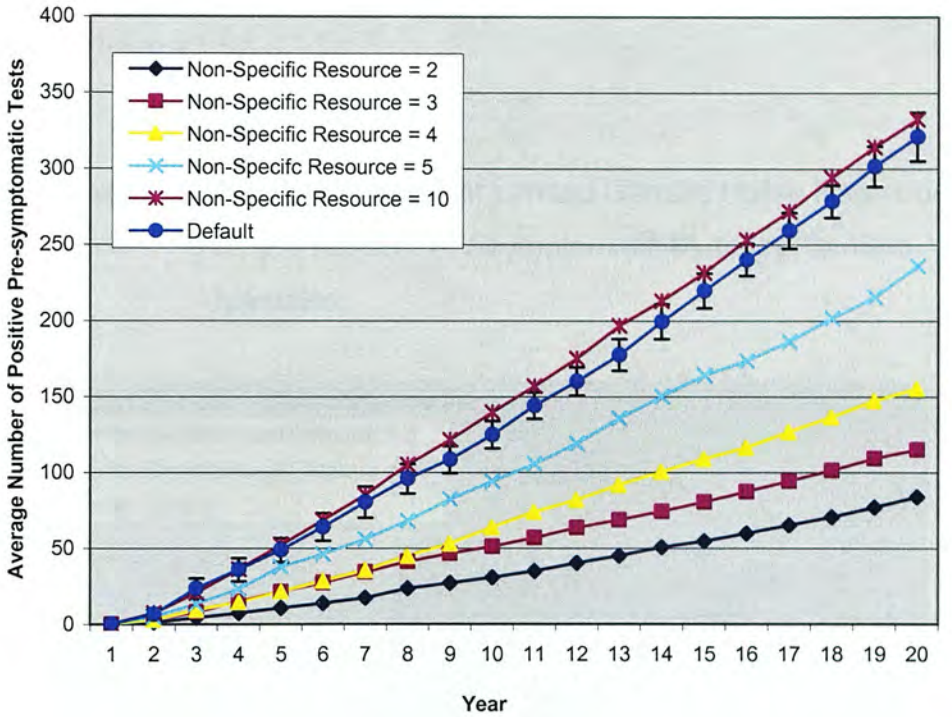
At default settings, resources are set at an extremely high level, to ensure that the performance of the system is not restricted by limitations in the resources allocated. It was anticipated that, for each point at which resources are utilised, there will be a threshold beyond which the resources will be unable to cope with the demand, resulting in increasing waiting lists and potentially reduced yield. In a real life situation, the bulk of the available resources, both in terms of time and money, are likely to be expended on conducting mutation analysis. Hence, the model was run ten times each at various resources settings for this stage, in order to determine the aforementioned threshold and investigate the effect that restricted resources would have on the overall number of asymptomatic MMR gene mutation carriers identified through the cascade genetic testing system. The effect on the number of people on the waiting list for mutation analysis of having tests resources limited to 2, 3, 4, 5, 10 and 1000 (default) tests per week is illustrated in figure 4.3.50.

Figure 4.3.50 Predicted Impact of Limited Mutation Analysis Resources on the Waiting List Length of the Mutation Analysis Waiting List



The above figure illustrates the principle that limiting resources will have no effect on the functionality of a cascade genetic testing system, provided they remain above the demand threshold. The number of people on the mutation analysis waiting list when resources are limited to 10 mutation analyses per week does not differ from default settings, when up to 1000 mutation analyses can be performed. In contrast, if the resource level is further reduced to 5 mutation analyses per week, demand exceeds the capacity for supply, and the number of people on the waiting list increases rapidly. Predictably, when resources are reduced further the rate of increase in the size of the waiting list becomes proportionally greater. The predicted effect of the increased waiting list size on the potential yield from cascade genetic testing is shown in figure 4.3.51.

Figure 4.3.51 Predicted Impact of Limited Mutation Analysis Resources on the Number of Asymptomatic Mutation Carriers Identified

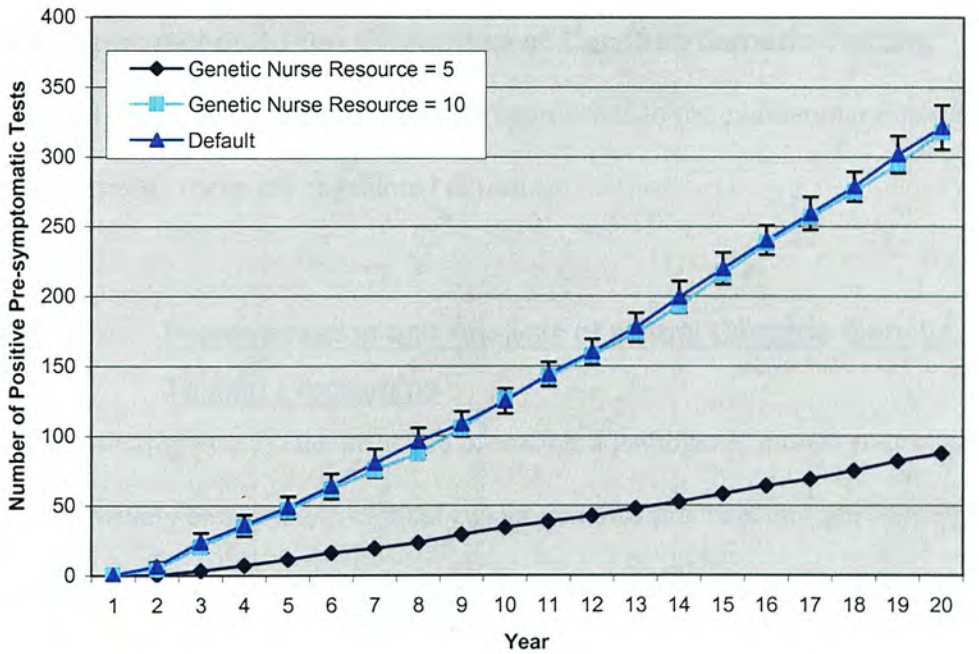


Again, there is no consistently significant difference between yield at default settings and yield when mutation analysis resources are limited to ten per week. Reducing resources to five per week, and below, results in a substantially reduced number of asymptomatic mutation carriers being identified. Hence, the model predicts that a cascade genetic testing programme would require the capacity to perform up to ten mutation analyses per week, in order to operate to its full potential.

Since a genetic nurse interview is a pre-requisite for mutation analysis, and only 10% of individuals attending a genetic nurse interview decline genetic testing, it can be anticipated that the hypothetical cascade genetic testing programme also requires the capacity to conduct approximately ten genetic nurse interviews a week.

Conversely, reduction of genetic nurse interview resources to five per week would be expected to introduce a 'bottleneck' effect. These inferences were borne out by running the model ten times each with resources at this stage set to 5 and 10 weeks, as shown in figure 4.3.52.

Figure 4.3.52 Predicted Impact of Limited Genetic Nurse Resources on the Number of Asymptomatic Mutation Carriers Identified



An indication of the minimum resources required for maintaining model performance can be obtained using the general principle of analysing the resource expended at default settings. For example, at default settings, an average of 850 pre-symptomatic tests were performed over the twenty-year period: a rate of 0.82 per week. Hence, it can be assumed that a resource capacity of just one pre-symptomatic test per week would be adequate to maintain yield attained at default settings.

4.4 DISCUSSION

Cascade genetic testing provides a potential strategy for identifying asymptomatic carriers of MMR gene mutations in the Scottish population. A thorough evaluation of the potential utility of cascade genetic testing in this context has not yet been undertaken, and the development and analysis of the computer model described as part of this thesis was designed to address this issue.

4.4.1 Approaches to the Evaluation of Cascade Genetic Testing

In general terms, there are three possible approaches to the evaluation of cascade genetic testing. These are considered in turn.

4.4.1.1 Implementation and Analysis of a Real Cascade Genetic Testing Programme

As discussed previously, the principle of tracing a pathogenic mutation through a family is widely employed in clinical cancer genetics practice, and introducing a systematic population-wide programme of cascade genetic testing may be a logical next step.

In many ways, simply implementing and analysing a real cascade genetic testing programme provides an ideal means of evaluating this strategy. Certainly, this real-life approach has numerous advantages over computer modelling, as no computer model can be considered a substitute for practical experience. However, there are two major drawbacks to such a strategy. Firstly, whilst implementing cascade genetic testing will provide a wealth of information about the application of a

specific programme in a particular context, it will not be capable, other than by inference, of providing information on cascade genetic testing under alternative conditions. The other major drawback to evaluating cascade genetic testing through direct implementation arises from the time-scale. At the population level, where the goal is the identification of asymptomatic mutation carriers, the medium and long-term outcomes of cascade genetic testing are of primary importance, and these may be measured over many years. Hence, it would be necessary to wait for many years for data from a current cascade genetic testing programme to mature sufficiently to inform predictions of medium and long-term outcomes for another programme.

Thus, the implementation of cascade genetic testing will provide valuable information on the application of a particular programme in a specific context, and this information may help inform the planning of future programmes. However, inference and assumptions will remain a necessary part of planning a future programme, since the population, genetic epidemiology, human behaviour and inclusion criteria involved may all be different. Furthermore, there is currently a paucity of data on the long-term results of cascade genetic testing, and this will inherently take many years to resolve, even if systematic cascade genetic testing is introduced immediately or is presently ongoing. A thorough and useful evaluation of the potential utility of cascade genetic testing must include the capacity to project results forward to provide an indication of the long-term outcomes, and the ability to predict the effects of applying various cascade genetic testing programmes under a variety of conditions.

4.4.1.2 Mathematical Evaluation

A mathematical evaluation, based on a series of equations to calculate theoretical outcomes, offers an alternative approach to the evaluation of cascade genetic testing. This approach has previously been utilised by Krawczak et al., (150). Straightforward mathematical evaluation, as opposed to computer modelling, has the advantage of presenting a relatively simple and transparent evaluation of cascade genetic testing. The main disadvantage of this approach is that even complex equations are inherently limited in the extent to which they can represent a real-life cascade genetic testing programme. Cascade genetic testing is inherently a highly complex system, as it incorporates such elements as family structure, population demographics, genetic epidemiology and participant behaviour. Consequently, a series of equations that do not consider all such factors, and do not incorporate data specific to a particular clinical context, represent a gross oversimplification of cascade genetic testing as it may be applied to the challenge of identifying asymptomatic mismatch repair gene mutation carriers in the Scottish population.

Furthermore, the time-scale of a cascade genetic testing programme is an essential consideration from a practical perspective, as the yield from any cascade genetic testing programme will be conditional on the length of time the programme is in operation. This factor is not addressed by a theoretical mathematical approach. A meaningful evaluation must address not just the potential achievements of a cascade genetic testing programme, but also the length of time required to reach such achievements. Hence, a mathematical evaluation is of limited practical use in terms of predicting outcomes and informing practical planning for cascade genetic testing.

4.4.1.3 Computer Modelling

Computer modelling provides a very powerful tool for representing, understanding and studying complex biological and medical systems. This ability to cope with complexity enables numerous factors and their interactions to be considered, and provides the capacity to conduct an evaluation that is highly specific to a particular situation; in this case the evaluation of cascade genetic testing for mismatch repair gene mutations in Scotland. The capacity to incorporate practical experience and real data into a computer model can be used to ensure that the model is representative of the real system under consideration. Another major advantage of developing a computer model as a tool for evaluating a complex system is flexibility. A computer model can readily be updated to incorporate new ideas and information, and the model inputs can be manipulated to facilitate the evaluation of various cascade genetic testing programmes under different conditions.

There are, however, drawbacks and limitations to computer modelling. Firstly, a computer model is, in essence, a simplification of a real-life system, and as such model outcomes may not be entirely realistic. Computer modelling is also a somewhat subjective process, as the model created will represent the system as it is understood by the modeller. A further crucial limitation of computer modelling is that the accuracy and validity of the information provided by the model is dependent on the quality of the data used as inputs to the model system. Each of these limitations is considered in subsequent sections of this discussion.

Overall, the development of a computer model of cascade genetic testing for mismatch repair gene mutations in the Scottish population currently represents the most appropriate approach to the evaluation of this strategy. The computer model described in this thesis represents an attempt to integrate this approach with real data relevant to cascade genetic testing, by developing a computer model that is partly based on and informed by the ongoing COGS programme.

4.4.2 Justification of Methods

The development of a computer model of cascade genetic testing was based on an object-orientated approach to visual modelling. The object-orientated approach in general provides a useful and proven technique for modelling complex systems. The principal merit of object orientated modelling in the current context is that it is based around "real-world" entities, with the model system being described by these entities and their interactions. This approach to modelling is designed to reflect the way in which the human brain itself understands systems: through objects and their interactions. Consequently, object-orientated modelling enabled a researcher with a background in the 'domain' to develop a computer model and communicate this both to other researchers and computer programmers.

The strategy outlined in the section 4.2 facilitated collaboration between the domain expert and the computer programmer, whilst the stepwise process of model development enabled the domain expert to retain overall control of the model's design and functionality. The unified modelling language (UML) is a widely used and user-friendly language for object orientated modelling (196, 224, 296). Hence,

the utilisation of UML to develop an object-orientated visual model of cascade genetic testing constituted the most appropriate strategy for model development. Based on the experience of developing this computer model, the decision to use object-orientated modelling and the UML has been fully justified. It is intended to utilise these tools for further work on the model (see 'further work' section), reflecting the success of the overall approach.

4.4.3 Limitations and Sources of Error

In essence there are three potential sources of error that may be present in the completed computer model of cascade genetic testing:

(i) Model Conception – Does the model accurately represent reality?

(ii) Model Application – Does the working model function properly and does it accurately reflect the conceptual model?

(iii) Model Inputs – Is the data that informs the model complete and accurate?

The limitations of the model with respect to these aspects are considered below.

4.4.3.1 Conception

A computer model is essentially a representation of a complex system as it is understood by the modeller. Consequently limitations, omissions and inaccuracies in this understanding will become manifest as such in the computer model. It is imperative, therefore, that the modeller first attains a thorough and accurate understanding of the system to be modelled. The methods employed, specifically the

task analysis and the development of the conceptual model, are designed to ensure that this level of understanding is attained. Discussion with various collaborators with expertise in clinical cancer genetics, genetic epidemiology and conducting genetic testing also helped to achieve the goal of a conceptual model that provides a comprehensive, detailed and accurate representation of the system in question. However, it must be acknowledged that the model created does constitute one interpretation of cascade genetic testing, and is thus subject to human error with respect to the accuracy with which the model reflects reality.

Once a thorough and accurate understanding of the system to be modelled has been attained, the next requirement is that this understanding is properly represented using the chosen modelling tools. Again, this aspect of model development is difficult to monitor in an objective manner. Effectively, the success, or otherwise, of the visual representation of the abstract model is determined subjectively by the computer modeller. Therefore, work on the conceptual model continued in an iterative fashion until the modeller was satisfied that the visual model was indeed an accurate representation of the system.

As before, a degree of external validity was provided by ensuring, through discussion with collaborators, that the conceptual model conveyed the desired information. The research environment in which the computer model was developed was ideal in this respect, since the research group and collaborators involved included experts from a variety of disciplines, including coloproctology, genetic epidemiology, clinical genetics and computer modelling. Hence, through discussion

and collaboration with other researchers it was possible to develop a model that was valid from a variety of relevant perspectives. The existence of alternative conceptual models of cascade genetic testing is inevitable, but the conceptual model considered herein is considered to be both valid and complete for the purpose of evaluating cascade genetic testing within the terms of reference set out in the 'aims' section.

4.4.3.1.1 Scope of Conceptual Model

For the purpose of creating a computer model, cascade genetic testing can be considered at a number of levels. At its simplest, this would include only the tracing of a known mutation through a pedigree. However, to provide a meaningful evaluation of cascade genetic testing in a particular context the population concerned, the genetic epidemiology of the genes involved, and the means by which index cases are identified must also be considered. Other aspects relating to cascade genetic testing in a wider context, such as clinical disease, alternative strategies for identifying mutation carriers, interventions in known mutation carriers and concurrent strategies for disease prevention, may also be relevant. However, it is not feasible to represent the complexity of cascade genetic testing in a particular context in its entirety, and pragmatic decisions regarding the aspects of cascade genetic testing that warrant inclusion in the model must be made. Therefore, whilst the conceptual computer model is considered to be a reasonably accurate representation of cascade genetic testing from the domain perspective, the scope of the conceptual model is inherently limited, and these limitations form a crucial consideration in the model's evaluation.

Early iterations of the conceptual model included a consideration of the process of colorectal cancer diagnosis. However, the end point of this section, or "package", of the conceptual model was cancer registration. Since cancer registry data can be accessed directly, both in real life and in the context of a model, the putative model of the diagnostic process was considered redundant. This provides an example of how a conceptual model can be refined and restricted without detriment to its overall utility.

In contrast, the effective absence of cancer genetic services from the model presents a major limitation in terms of how closely the model system represents reality. Individuals who are concerned about their family history of colorectal cancer and consult a general practitioner in this regard, or individuals who are otherwise found to have such a history at a primary care consultation, may be referred to cancer genetic services. Subsequently, individuals deemed to meet particular criteria may be offered genetic testing. Essentially, therefore, cancer genetic services provide an alternative strategy for identifying mismatch repair gene mutation carriers that is based on family history. A cascade genetic testing programme can potentially ascertain mutation carriers in this way as well as, or instead of, testing a sub-group of colorectal cancer cases.

Initially, the task analysis and early iterations of the conceptual model included the capacity to identify mutation carriers by targeting both colorectal cancer cases and individuals with a family history of the disease. For practical reasons, it was later decided to disregard family history in order to focus on the evaluation of cascade

genetic testing based on identifying index cases through testing colorectal cancer cases. This was done for several reasons. Firstly, the computer model was largely informed by the ongoing COGS study, which is concerned only with genetic testing in colorectal cancer cases and their families. Also, the inclusion of cancer genetic services would add another layer of complexity to the model, making it relatively difficult to evaluate specific hypothetical cascade genetic testing programmes against this background. Similarly, clinical information on the identification of MMR gene mutation carriers through genetic services is limited and its inclusion would add yet more uncertainty to the model. The decision to limit the scope of the computer model to exclude cancer genetic services was also a practical one, made on the basis that there was insufficient time available to model this aspect of cascade genetic testing in the appropriate level of detail. In this respect the completed model represents a balance between comprehensiveness and coherence.

The inability of the current model to consider cancer genetic services as a means of identifying mutation carriers constitutes a significant limitation. This will be addressed in future model development. Accordingly, results currently generated by the model must be interpreted on the understanding that they represent one particular approach to the identification of mutation carriers in cascade genetic testing.

Another aspect of cascade genetic testing that is not considered as part of the computer model is the follow-up surveillance of mutation carriers. As discussed elsewhere, the appropriate intervention strategy for MMR gene mutation carriers remains the subject of debate, but it is likely to involve regular colonoscopic

surveillance. The surveillance implications of cascade genetic testing are of enormous importance in the wider context of using this strategy to reduce the burden of colorectal cancer in the population, and health risks as well as benefits may accrue from colonoscopic surveillance. Modelling of clinical services is a major undertaking, and accordingly the relevant aspects of colonoscopic surveillance could not feasibly be included in the current model. The predicted outcomes from the model do, however, form a potential starting point for assessing the surveillance implications of cascade genetic testing at the population level.

Realistically, the decision to implement cascade genetic testing at a national level would be based on the costs involved, as well as the predicted health benefits. Again, early versions of the task analysis and conceptual model contained packages relating to the economics of cascade genetic testing. However, a thorough economic evaluation of cascade genetic testing would have to consider not just the direct laboratory, staff and administrative costs involved in implementing the programme itself, but also the costs of clinical interventions for mutation carriers and conversely the economic implications of preventing symptomatic colorectal cancer in this group. This would involve developing a model that includes the financial aspects of cancer genetic services and health care provision for colorectal cancer in Scotland. Such a model would constitute a major project on its own, and thus an economic evaluation of the cascade genetic testing was considered well outwith the scope of the current model.

Whilst the current model does not provide detailed economic information, it is designed so that the outcomes can be used to inform economic considerations. The yield of a particular cascade genetic testing programme states the number of mutation carriers identified over time, and the model also predicts when and how each carrier will be identified. Although it is not a direct outcome of the current version, the computer model also retains the capacity to predict the age and sex of newly identified mutation carriers. This information, coupled with surveillance guidelines for mutation carriers, could feasibly be used to determine the number of surveillance colonoscopies that would be performed over time as a direct result of cascade genetic testing. In turn, such estimates could be used to calculate the associated costs and staff time required.

Direct costs associated with implementing a particular cascade genetic testing programme can be estimated in a similar way. The model will provide data regarding the number of events (interviews, tests etc.) that are required, and this information, coupled with estimates of the cost per event can give an overall indication of the costs involved.

The inclusion of a 'resources' output was designed to facilitate economic considerations by providing a concise weekly report of the resource requirements of the programme in question. Thus, as well as providing estimates that can be used to inform the economic assessment of cascade genetic testing, the model can also provide information, in relative terms, regarding the resources required for various cascade genetic testing programmes under different conditions.

The estimated number of mutation analyses performed provides a useful indicator of the total resources expended as part of a cascade genetic testing programme. This is partly because conducting mutation analysis is likely to be the most expensive and time-consuming single event in the programme, and also because the number of mutation analyses performed is directly related to the number of other events (interviews, pre-symptomatic tests etc.). The relationship between the number of mutation analyses performed and the number of asymptomatic mutation carriers identified can thus be used as an indicator of relative efficiency, as illustrated throughout the analysis of model outcomes in this thesis.

Ultimately, the computer model is not capable of providing a detailed evaluation of the financial requirements and economic implications of a cascade genetic testing programme. This can be considered a significant limitation from the point of view of planning such a programme. However, whilst the economic aspects of cascade genetic testing do not form part of the computer model per se, it is possible to use the model outcomes provide to data to inform such considerations. Furthermore, the methods used to develop the computer model are designed to facilitate the expansion of the scope of the model, and consequently an economic component could be added as part of future work.

Cascade genetic testing may be associated with various psychosocial issues (40, 165, 184, 227). The risk of anxiety inherent in genetic testing, and the psychological impact of being identified as a mutation carrier and undergoing surveillance colonoscopy, are relevant considerations in the planning of any cascade genetic

testing programme. Additional considerations with respect to cascade genetic testing arise from the decision to involve family members in the programme. Much of the available information regarding the psychosocial aspects of genetic testing comes from qualitative studies, and thus appropriate data for inclusion in a computer model is scarce. Furthermore, specific information on cascade genetic testing for mismatch repair gene mutations is not currently available. Accordingly, the model does not include a component relating to psychosocial issues, although as with economic considerations model outputs can potentially be used to provide a quantitative assessment of the psychosocial impact of cascade genetic testing at the population level.

Additional aspects related to cascade genetic testing that could not feasibly be included in the current model include the possible presence of concurrent colorectal cancer prevention strategies, such as FOBT screening, and the inclusion of extracolonic cancers that are associated with MMR gene mutations in ascertainment criteria for genetic testing. Once again, these considerations could not be included in the computer model for pragmatic reasons, but could feasibly be incorporated at a later stage.

4.4.3.1.2 Simplification of Conceptual Model

Oversimplification of aspects of cascade genetic testing that are included constitutes a further source of potential limitation of the computer model. This is particularly evident with respect to construction of pedigrees. A key assumption that is built into the computer model relates to family size and structure. The process of tracing a

mutation through a pedigree is central to cascade genetic testing as it effectively determines the number of asymptomatic mutation carriers that will be identified for a particular index case. For practical reasons, the model generates a simple, stylised pedigree for each mutation carrier identified. This pedigree assumes that the population remains constant, and that the generation gap is 25 +/- 5 years. These assumptions are not unreasonable in themselves, since the Scottish population is relatively stable and the majority of children are born to parents between the ages of 20 and 30 years (84). However, the model does not reflect the true diversity of the Scottish population in terms of family size and structure.

A real-life cascade genetic testing programme would feature considerable variation in the number of relatives identified for each index case, reflecting differences in pedigree size, structure and the age of family members. Large numbers of asymptomatic carriers may be identified from one index case, whereas other index cases may yield no further carriers. In contrast, although a random element is introduced with respect to the generation of age and acceptance rates of relatives, the model applies the same uniform pedigree structure to all mutation carriers, and hence each index case has the same potential for the subsequent identification of carriers amongst relatives. The impact of this limitation on the outcomes form the computer model is difficult to predict, but it is apparent that, in terms of pedigree structure, the current model represents an oversimplification of real life. A more accurate and detailed means of generating pedigrees for mutation carriers, which is informed by real pedigrees from the COGS programme, will form a crucial aspect of future model development.

The interrelatedness of individuals within the population provides a further concern. Currently, the model generates entirely separate pedigrees for each index case. These pedigrees are extended outwards through cascade genetic testing and do not overlap, regardless of how large they become. In reality, there is a distinct possibility that different index cases may share a common ancestor, and may thus have inherited their MMR gene mutation from the same source. In such circumstances, the index cases will be related to some extent, although this may not be apparent at the time they are identified. Such an occurrence has already been observed in the Scottish population, with mutation carriers identified through genetic services subsequently being found to be distantly related (Dr. M Porteous, personal communication).

The extent to which index cases identified as part of a cascade genetic testing programme are likely to be related depends on the genetic variation of the Scottish population in general and the genetic epidemiology of colorectal cancer in particular. Considerable heterogeneity of mismatch repair gene mutations has been observed in previous mutation analysis studies involving Scottish subjects (57, 62, 173), and no founder effect has been established in the Scottish population. This suggests that a large proportion of MMR gene mutations in the Scottish population have arisen de novo, resulting in discrete 'mutation families'. However, in many cases these mutation events will have taken place many generations ago, and the possibility that mutation carriers in Scotland may be distantly related persists.

The probability that interrelatedness will become apparent through cascade genetic testing will be determined by the 'depth' to which cascade genetic testing is capable of tracing mutations through families. Outcomes from the computer model suggest that an average of 1.88 (95% CI = 1.80, 1.95) carrier relatives are identified for each index case, implying that pedigrees in the model do not generally become extensive and that overlapping of pedigrees is unlikely.

From a more practical perspective, it is clear that the high risk of colorectal cancer conferred by a pathogenic mismatch repair gene mutation means that there is a high probability of more than one family member developing colorectal cancer within the time-frame of a cascade genetic testing programme, and thus potentially presenting as separate index cases.

In disregarding the potential relatedness of index cases, the computer model may overestimate the yield of cascade genetic testing, since the assumption that each referral or relative identified through cascade genetic testing is 'new' to the system may be violated. Again, the impact of this effect on the validity of the model's outputs is difficult to assess, although it is not likely to exert a major influence in comparison with other potential limitations. Ideally, the computer model should have the capacity to generate information on the interrelatedness of the entire Scottish population, but this is unlikely to be feasible. Future model development will take a more realistic approach, utilising findings from the COGS programme to predict the convergence of pedigrees.

4.4.3.1.3 Summary of Limitations at Conception Stage

In conclusion, since the conceptual model was developed in an iterative fashion using the powerful object-orientated approach, and expert consultation as well as extensive research was used to define and understand cascade genetic testing prior to model development, it is considered to be as accurate a reflection of the real-life process of cascade genetic testing as can be reasonably achieved. The model is, however, inherently limited in scope and thus does not constitute a comprehensive representation of all the aspects of cascade genetic testing that may be of interest. Inevitably, these limitations are reflected in the model outcomes and the conclusions that can be directly drawn from these results. Nonetheless, the key model outcomes are often highly relevant to aspects of cascade genetic testing that are not included in the model, and to some extent may be used to inform estimates and calculations outwith the model's frame of reference. In addition, the model's design retains the capacity for additional functionality to be incorporated at a later stage, and some of the aspects of cascade genetic testing discussed above will form part of future model development.

4.4.3.2 Application

There are two major issues in terms of assessing the potential for error in application of the conceptual model. Firstly, the extent to which the functional model reflects the conceptual design must be considered. In addition, the potential for undetected programming errors to occur must also be evaluated.

Although the conceptual model was designed to be comprehensive and thoroughly documented, a degree of scope for interpretation remains. Similarly, for each piece of functionality specified by the conceptual model (e.g. construction of pedigrees), there may be more than one possible approach to programming of that functionality. Consequently, the question of whether or not the functional model accurately reflects the conceptual model is a somewhat subjective one. Frequent meetings between the domain expert and the computer programmer ensured that the functional model is a complete and accurate representation of the conceptual model from each of these perspectives. Each package and diagram created at the conceptual design stage is included in the functional model, and the objectives of the project outlined in the conceptual model are met by the working version. To some extent, an external assessment of the transition from concept to function can be made through comparison of the functional model outlined in the results section and the conceptual model presented in appendix A7.

The functional model was directly translated into Java code by the computer programmer to create the working version of the computer model. The Sun ONE Studio 4 Update 1 (Sun™) (133) integrated development environment contains a robust "debugging" function that identifies coding errors. This was used throughout the development and implementation of the functional model, ensuring that at a technical level the model does function correctly. Inevitably, there is some potential for human error at the coding stage, but this technical task was performed by an experienced programmer and the potential for error was further minimised by the use of frequent review and extensive documentation of the code. Hence, no

functional alterations to the model are anticipated to have occurred at this stage. Additional checks on the functionality of the computer programme were made at the evaluation stage, through consistency checking and adjusting the model parameters and/or code to ensure the model reacts in a logical manner.

A final consideration with respect to the application of the computer model concerns the manner in which results for analysis were obtained. A random element is incorporated into the model, such that the actual results from any 'run' of the model at the same settings will vary. This feature makes the model system more realistic, but also means that a particular set of results may not be representative of all model outcomes at those settings. To account for this random element, the model was 'run' ten times at each setting, and the mean values and associated confidence intervals were used for analyses. The confidence intervals associated with the key estimates provided by the model were reasonably narrow, and therefore ten runs at each setting were generally adequate to observe statistically significant differences between the outcomes of interest at different settings. Hence, whilst further repetition of the cascade model would undoubtedly result in more accurate estimates and narrower confidence intervals, this was considered unnecessary for the purpose of this analysis. The capacity to automatically run numerous simulations will be incorporated as part of further development of the computer model.

4.4.3.3 Inputs

As stated previously, outcomes from a computer model are only as valid as the information used to inform it. Consequently, a critical evaluation of the inputs used

in the development of the computer model forms a crucial aspect of the overall interpretation of the model results. In this context 'inputs' refer to all assumptions and all numerical estimates used to determine the model's behaviour and output, including estimates that are modified by the model user.

A general consideration concerning the quality of data relating to the genetic epidemiology of mismatch repair genes stems from the potential for bias in the published studies that provide the source for much of this data. Inherently, all the data relating to MMR genes and their epidemiology that are used to inform the computer model come from mutations that have been identified and are generally considered to be pathogenic. There is an inherent ascertainment bias in many such studies, since potential mutation carriers were often targeted on the basis of family history or microsatellite instability. Further bias may be associated with the specific laboratory techniques employed. This constitutes a limitation of the entire model, and emphasizes the need for unbiased information relating to mismatch repair gene mutations and their role in colorectal cancer.

The quality of data available to provide input estimates is highly variable, and in some cases such data has a high degree of uncertainty attached, or is simply not available. In general terms, the computer model can accommodate such limitations in data quality by including a range of possible values designed to cover all realistic inputs. Each run of the model thus provides information on the behaviour and outcomes of cascade genetic testing under a particular set of assumptions. The assumptions identified as default values, and the alterations made to default values

during analysis of the effects of various inputs are presented in the results section, and considered in the subsequent discussion and interpretation. This transparency is an essential aspect of the evaluation of cascade genetic testing using a computer model, and ensures that model analysis adheres to recognised principles of good practice in computer modelling (300).

4.4.3.3.1 *Population Demographics*

Data relating to the demographics of the Scottish population are considered to be robust, since they were obtained directly from the 2001 census (228). The only caveat to the reliability of this data is that in the model assumes that the 2001 figures remain constant throughout the 'run' period, i.e. a static, rather than a dynamic, population is modelled. The Scottish population is following the general pattern of an ageing population in a developed country, and is predicted to decrease by around 3.3% over the next twenty-year period (84). Although changing population demographics may influence outcomes from cascade genetic testing this effect is likely to be minor.

4.4.3.3.2 *Colorectal Cancer Incidence*

Cancer incidence data comes directly from the Scottish Cancer Registry, which is considered to be approximately 96.5% complete (27), and therefore constitutes a reliable and detailed source of information. Through the 'expected cases' look-up table, the computer model utilises average figures from 1989-1998 to predict the annual incidence and age/sex distribution of colorectal cancer in the Scottish population. These figures were the most up-to-date available at the time the model

was developed, but are obviously slightly dated. Again, the model does not attempt to project forward cancer incidence data, but rather assumes that the incidence of colorectal cancer remains constant throughout the run period of the model. It is possible that incidence will change over the next twenty years, with the current pattern of increasing incidence and decreasing mortality possibly continuing. Future interventions, such as national screening using the faecal occult blood test, may also have an impact on the epidemiology of colorectal cancer. Hence, the assumption that colorectal cancer incidence will remain constant throughout the period of cascade genetic testing is not entirely valid.

4.4.3.3.3 Penetrance

The number of MMR gene mutation carriers expected to develop colorectal cancer each year is calculated from available data on the prevalence and penetrance of MMR gene mutations, and is thus subject to the limitations and uncertainty of these data. Several studies have indicated that penetrance in terms of colorectal cancer is greater in males than in females (4, 57, 170, 286, 289) and a conservative estimate of penetrance based on the consensus of these studies is 80% in males and 40% in females (199). This estimate also reflects the results of the only one of the above studies that is not biased by family history status, in which penetrance to age 70 was estimated to be 74% in males and 30% in females (57). Consequently the estimates of 80% penetrance in males and 40% penetrance in females were used as default settings for the model, and, with the exception of the complete penetrance setting, the assumption that penetrance in males was double that in females was made for all input settings.

However, the applied estimates of penetrance reflect a consensus based on results from only a few studies conducted in different populations, and a range of estimates have been reported (see table 2.2 for details). Statistical evaluations of the accuracy of these estimates are not reported, but each study is relatively small in terms of the actual number of mutation carriers considered, and the number of these carriers who have developed colorectal cancer. The low number of index cases included in these studies also means that the carriers involved may not be representative of all mutation carriers. Similarly, there is potential for geographical variation in mismatch repair gene mutations, and therefore the extent to which the specific mutations considered in previous studies are representative of the Scottish population is unclear. Furthermore, the ascertainment of these cases, which, with one exception (57) usually involved family history criteria, creates a further source of bias. Consequently, the estimates of penetrance used as model inputs cannot be considered accurate, and model outcomes must be interpreted accordingly.

A related issue is that of competing mortality. The model makes the assumption that all individuals predicted to develop colorectal cancer on the basis of carrier status and penetrance estimates will do so, but this is not necessarily the case since they may die of unrelated causes beforehand. Hence there is a possibility that the estimates of penetrance used to inform the model may slightly overestimate the true number of carriers developing colorectal cancer. Another consideration in terms of 'expected cases' arises from the potential for clinical interventions to alter the number of MMR gene mutation carriers who actually develop cancer. If a significant proportion of mutation carriers were undergoing surveillance due to having a strong

family history of colorectal cancer, the effective penetrance of MMR gene mutation in this group may be reduced. However, the inconsistent relationship between family history and carrier status, and the fact that not all those with a strong family history are aware of it and actively engaged in clinical surveillance, means that this effect may not exert a major influence on the number of cases occurring amongst carriers.

Inaccuracy of penetrance estimates used to inform the model may result in systematic bias in model results in comparison to the outcomes of a real-life cascade genetic testing programme. If the true penetrance of MMR gene carriers is greater or less than expected, the number of carriers developing colorectal cancer each year will differ systematically from the observed model inputs. To some extent, this possibility is dealt with by the range of penetrance estimates used in the analysis of the model. Similarly, the assumption that penetrance in males is approximately double that in females may be incorrect, in which case the number of expected cases occurring each year will be correspondingly inaccurate compared to real life. Again, whilst this assumption is made for almost all input settings used in the analysis of the model presented herein, the model at least retains the capacity to alter penetrance estimates to determine outcomes over a realistic range, and also to incorporate new and improved estimates of penetrance obtained from future research.

A further and more complex consideration is that penetrance is likely to vary according to the mutation in question. There is some evidence to suggest that penetrance of hMLH1 mutations differs from that of hMSH2 mutations (75, 170, 286), and it is possible that penetrance of different mutations within the same gene

will vary according to their precise site and molecular nature. Hence, the computer model's assumption that penetrance is constant in all mutation carriers may not be valid.

The impact that variation in penetrance according to the specific mutation involved may have on model outcomes is difficult to predict. If penetrance varies according to the mutation involved, then the estimate of penetrance used as a model input can be considered as the mean of all individual penetrance estimates. However, the clinical interpretation of model results would be significantly affected if penetrance varied by mutation. Although the overall number of MMR gene mutation carriers developing colorectal cancer each year may remain constant, the majority of these cases would be in people with mutations of relatively high penetrance. Conversely, carriers of low-penetrance mutations are less likely to develop cancer and present to a cascade genetic testing programme. This may lead to a clinical benefit, since the people most likely to develop colorectal cancer would be identified early, but it would result in a decrease in the number of expected cases in MMR gene mutation carriers, due to the fact that the remaining carriers in the population have a low penetrance. Such an effect would compound the lack of feedback in the computer model, which is considered later.

4.4.3.3.4 *Prevalence*

Data regarding the prevalence of MMR gene mutations were only available from one paper by Dunlop et al., 2000 (58). This was calculated using the following equation, derived from Bayes' theorem, for population carrier frequency:

Population Carrier Frequency = [(carrier frequency in colorectal cancer patients) x (population prevalence of colorectal cancer)] / (prevalence of colorectal cancer in carriers)

The values used for carrier frequency in colorectal cancer patients were obtained from two studies by Aaltonen et al., 1998 (2) and Liu et al., 1995 (173) that conducted mutation analysis on systematically collected cohorts from Finland and the USA respectively. These studies in themselves are subject to bias. The observed carrier frequency in colorectal cancer patients is specific to the population concerned and thus may not be representative of the Scottish population. In particular, the Finnish population is noted for a high prevalence of founder mutations. Also, microsatellite instability was used as a prerequisite to genetic testing in both studies, potentially resulting in a proportion of carriers being overlooked. The population prevalence of colorectal cancer used in the above expression was obtained from the Scottish Cancer Registry, and can therefore be considered robust. However, the prevalence of colorectal cancer in carriers was estimated from data on only 48 Scottish MMR gene mutation carriers, seven of whom had developed the condition, and this estimate is therefore subject to considerable uncertainty.

A further limitation of the available data on the population carrier frequency stems from the fact that the estimate provided by Dunlop et al., (58) refers to the 15-74 age group. Intuitively, it seems possible that the prevalence of MMR gene mutations in the population may decrease with age, due to the relatively high mortality associated with carrying such a mutation. However, there are no data available concerning the

age distribution of MMR gene mutation carriers in the population that might confirm or disprove this hypothesis, and consequently the assumption that carrier prevalence remains constant in the 15-74 age group was made for the purpose of the model.

For the reasons outlined above, the population carrier frequency calculated by Dunlop et al., (58), and subsequently utilised as the default input for the computer model of cascade genetic testing, can be considered an indication, but not a reliable estimate, of the population prevalence of MMR gene mutations . To a large extent, this is reflected in the wide confidence intervals around the estimate. Accordingly, this level of uncertainty is incorporated into the model, with the upper and lower confidence intervals corresponding to the high and low prevalence inputs. However, due to the limitations and potential sources of bias associated with the prevalence estimate, it is conceivable that the true prevalence may be outside these inputs. Again, the estimate of prevalence used in the model settings must be considered as an assumption, rather than a conclusive input, and the results must be interpreted with this in mind.

4.4.3.3.5 *Cumulative Incidence*

As well as the penetrance and prevalence settings themselves, the expected number of colorectal cancer cases in MMR gene mutation carriers is subject to the distribution of cumulative incidence obtained from the charts of cumulative incidence constructed by Dunlop et al., (1997) (57). These charts are calculated from 67 gene carriers from six families and thus may not be representative of the cumulative incidence in all MMR gene mutation carriers. However, the shape of the

charts is consistent with the cumulative incidence of a disease with a genetic origin, and hence the age distribution of expected colorectal cancer cases used as a model input is considered to be a reasonable, if approximate, estimate of the true distribution.

4.4.3.3.6 *Calculation of 'Expected Cases'*

Limitations at this stage are the product of inaccuracies and uncertainty in the above data, since these data are used to calculate the expected number of cases, as outlined in the methods section. An additional restriction stems from the lack of detail in some of the available estimates, specifically from the fact that data relating to population demographics and colorectal cancer incidence are entered in five-year age groups. Cumulative incidence of colorectal cancer cases in mutation carriers was also estimated for each age group, and the average annual incidence over a five-year period was calculated, as opposed to the actual expected incidence each year. Consequently, the model lacks precision with respect to the number of colorectal cancer cases expected, since this estimate depends on the five-year age group of the individual concerned rather than their exact age. However, this effect is unlikely to have a significant influence on the model outcomes, particularly in comparison to the potential influence of other limitations relating to input data quality.

No cumulative incidence data are available for MMR gene carriers over 70 years of age, and information on prevalence of MMR gene mutations is unavailable after age 74. Consequently, the model does not consider the possible occurrence of colorectal cancer in mutation carriers over 70 years of age. This is unlikely to have a major

effect on the validity of the model outcomes, since the number of MMR gene carriers developing cancer after the age of 70 is likely to be very small, both in absolute terms and in relation to the number of sporadic colorectal cancer cases occurring in this age group. Most age criteria for inclusion in a cascade genetic testing programme, including default settings for the model, are unlikely to include patients over the age of seventy due to the large number of sporadic colorectal cancer cases in elderly people and competing morbidity in this age group. Hence the lack of data in elderly mutation carriers does not constitute a major limitation of the computer model in most circumstances.

A lack of feedback in the computer model in relation to the estimation of expected colorectal cancer cases in mismatch repair gene carriers constitutes a further important limitation of the calculation of expected cases. In reality, as asymptomatic mutation carriers are identified in the course of a cascade genetic testing programme, the number of carriers remaining in the population will decrease accordingly. For practical reasons, such feedback is not included in the computer model, although it constitutes an important aspect of future work. Removing carriers from the population will exert a self-limiting influence on the yield from cascade genetic testing, since as the number of unidentified carriers decreases, the expected number of colorectal cancer cases in this group will also decrease. The magnitude of this effect depends on the ratio of identified carriers to unidentified carriers. Extrapolation of data from Dunlop et al., (58), suggests that there are an estimated 1209 (95% CI = 498, 3044) mutation carriers aged 15-74 in the Scottish population. At default settings, the computer model predicts that 204.8 (95% CI = 192.4, 217.2)

mutation carriers would be identified through cascade genetic testing after ten years, and 492.7 (95% CI = 468.9, 516.5) would be identified after twenty years. This represents a substantial proportion of all mutation carriers in the population. In a real cascade genetic testing programme, the identification of this number of carriers may have a significant impact on the number of colorectal cancer cases expected to occur in the 'unidentified' carrier population. In this respect, the computer model tends to overestimate the number of index cases that would be identified by cascade genetic testing, particularly in later years. A feedback 'loop' to account for the removal of mutation carriers from the population as they are identified by the model system will constitute a vital part of future model development.

4.4.3.3.7 *Acceptance Rates*

The level of participation on the part of the target population is a crucial factor in determining outcomes of a cascade genetic testing programme. The computer model applies a 'probability of participation' algorithm at each of four stages, namely the probability of initial contact with the system, the probability of attendance at an interview with a genetic nurse, the probability of accepting genetic testing and the probability of allowing contact with relatives. This straightforward pattern is broadly representative of a real-life cascade genetic testing programme. However, the actual figures utilised by the model for determining the probability of participation at each stage are subject to error. As discussed previously, although some general information regarding the acceptance rates of genetic testing is available, data that relate specifically to MMR gene mutations were available only from the COGS study. The strategy devised to accommodate the uncertainty surrounding acceptance

rates in the computer model was to apply deliberately conservative estimates that are consistent with the COGS data at default settings, and then systematically investigate the effect of various acceptance rates on model outcomes. At default settings, the overall acceptance rate of genetic testing in the computer model is 72%, a rate close to the preliminary estimate of 74% from the COGS study. Data on the probability that an index case will allow contact with relatives are even more scarce, although initial experience with the COGS study suggests that refusal at this stage is very rare. Hence, the acceptance rate of 0.9 applied at default settings is a realistic yet conservative estimate.

Conservative and simple estimates as inputs in the absence of other reliable data were adopted in several situations to simplify the model and facilitate straightforward comparison between various cascade genetic testing programmes and various input settings. The rationale for this approach is that clarity with respect to inputs is an important consideration in the evaluation of the model, and that the use of simple, conservative inputs is preferable to the inclusion of complex estimates that are not necessarily accurate. Hence, the model is considered to comprise a realistic representation of the complex process of recruitment in cascade genetic testing, but the acceptance rates applied may not accurately represent acceptance rates in a real programme. The model has, however, been designed with the capacity to incorporate detailed information on acceptance rates, and such data can be utilised as they become available in the future.

4.4.3.3.8 *Test Parameters*

A pre-symptomatic genetic test for a specific mutation is technically very straightforward to perform using PCR-based methods, and, although there is some potential for laboratory error, the model's assumption that the sensitivity and specificity of such a test is 100% is justified. Mutation analysis is far more complicated, and worthy of further consideration. Essentially, for the purposes of the computer model, a MMR gene mutation is either present or absent, and will either be detected by mutation analysis or will not. This, again, is an oversimplification of reality. Numerous genes are involved in the mismatch repair system, not all of which have been identified, and various laboratory techniques, each with their own test parameters, may be applied as part of an overall mutation analysis protocol. Complete analysis of all known mismatch repair genes may not be possible and the results may be inconclusive. The potential also exists for misclassification of gene variants, such that an essentially harmless variant is assigned pathogenic status or vice versa. Furthermore, as considered previously, a spectrum of pathogenicity may exist with some gene variants having relatively modest effects on risk of colorectal cancer.

In many ways it is not appropriate to consider mutation analysis in terms of sensitivity and specificity, since the actual situation is far more complicated than this implies. However, once again due to time constraints it was not possible to incorporate such additional complexity into the model. Accordingly, these test parameters were assumed to be to 100% throughout the analysis of the computer model, in order to simplify this part of the model and thus facilitate evaluation of

other aspects of cascade genetic testing. This assumption is not considered to be entirely valid, and the possibility that false positive or false negative results could occur in a real cascade genetic testing programme provides a potential source of inaccuracy in outcomes from the model system. The capacity to model the process of mutation analysis in detail, and consider the implications of inaccuracy at this stage, will form part of future model development.

4.4.3.3.9 *Delays*

The inclusion of delays at various stages of the model system is essential as these delays partly determine the chronological nature of the computer model. Where possible, delays were based on experience with the COGS programme. However, these delays are a simplified representation of cascade genetic testing in real life, and as such the timescale of cascade genetic testing as predicted by the model may not be accurate. The model retains the capacity to adjust these delays to reflect improved estimates or explore the effects of different delays on model outcomes.

4.4.3.3.10 *Summary*

In view of the limitations considered above, the current version of the computer model as considered in this thesis may thus be regarded as a "prototype", and future work is required to develop a model that provides detailed and accurate predictions of outcomes from a real-life cascade genetic testing programme. Nonetheless, whilst the results of the current model must be interpreted with caution, it does constitute a unique evaluation of the potential application of cascade genetic testing, and has the capability to provide a considerable body of relevant data for analysis.

4.4.4 Evaluation of Model Outcomes

4.4.4.1 Default Settings

At default settings, the computer model is designed to provide estimates for the outcomes of a cascade genetic testing programme in which index cases are identified through offering mutation analysis to all colorectal cancer cases diagnosed under the age of 55.

The number of mutation analyses performed over time is directly determined by colorectal cancer incidence, inclusion criteria and the specified participation rates. The model predicts that approximately 380 such tests would be completed per year (not including the first year in which no tests would be completed due to the delay involved), and that the total number of mutation analyses performed over a 20-year cascade genetic testing programme would be 7142 (95% CI = 7037, 7250). Using current laboratory protocols, this represents a substantial amount of work and involves exposing a large number of colorectal cancer patients to genetic testing.

An average of 171.5 (95% CI = 163.6, 179.4) patients undergoing mutation analysis were predicted to be MMR gene mutation carriers, implying that approximately 2.4% (95% CI = 2.3%, 2.5%) of mutation analyses performed in colorectal cancer cases diagnosed at less than 55 years of age yielded a positive result. This estimate is considerably less than that found in the initial stages of the COGS programme, in which 11% of colorectal cancer cases undergoing mutation analyses were found to harbour a mutation in hMLH1 or hMSH2. There are several possible explanations

for this difference. Firstly, the COGS data is only preliminary and includes missense mutations of unclear pathogenicity. The resulting figure of 11% is thus not necessarily representative of long-term outcomes, and may overestimate the number of truly pathogenic mutations identified. Conversely, the estimate generated by the model may be inaccurate, reflecting the limitations of the computer model itself and the quality of data used to determine inputs.

Published data regarding the proportion of colorectal cancer cases that have a MMR gene mutation is scarce and so it is difficult to evaluate the external validity of the computer model in this respect. There are, however, several published studies that provide such data for colorectal cancer cases occurring under the age of 50 (52, 201, 276, 294, 295, 299, 323), and pooling of these data suggests that 8.9% of these harbour a mutation in hMLH1 or hMSH2 (199). The mean corresponding estimate produced by the computer model for this age group is approximately 3.9%. This reinforces the previous observation, based on comparison between the model at default settings and the COGS study, and implies that the computer model may underestimate the proportion of colorectal cancer cases that carry a mutation. Whilst the precise reason for this disparity is not clear, both the estimates provided by the computer model and those obtained from real-life studies are potentially subject to bias and errors of an order of magnitude that could account for the observed differences. Irrespective of the observed disparity, it is apparent that the vast majority of colorectal cancer patients diagnosed at less than 55 years of age do not harbour a mismatch repair gene mutation.

From a clinical perspective, conducting mutation analysis on colorectal cancer patients is of limited use in the absence of cascade genetic testing, since, although identification of a mutation may prompt intensive screening and help prevent metachronous malignancies, it is inherently too late to prevent colorectal cancer. In the context of cascade genetic testing, where the primary aim is the identification of asymptomatic mutation carriers, mutation analysis can be considered as a sacrifice in terms of time and resources, which is necessary to identify index cases.

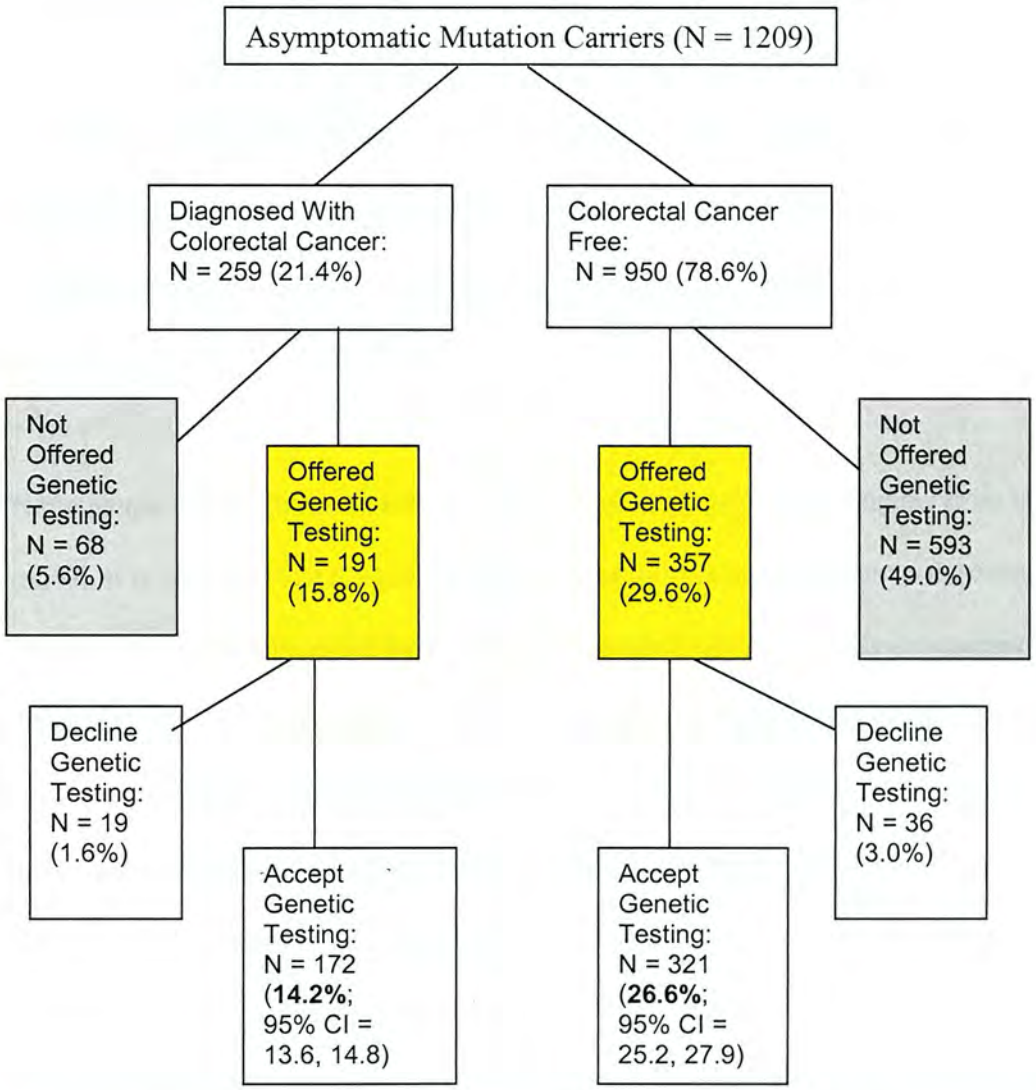
In comparison to the number of mutation analyses performed over time, the number of pre-symptomatic tests predicted to be performed during a twenty-year cascade genetic testing programme is small, averaging 849 (95% CI = 812, 886). Hence, at the population level, the impact of cascade genetic testing on asymptomatic members of the population is minimal, since only a tiny proportion of the population will be offered genetic testing. This observation illustrates one of the major advantages of cascade genetic testing over the alternative options for identifying asymptomatic carriers in a population. Population-wide genetic testing would inherently involve offering genetic testing to the entire population, with the number undergoing testing limited only by participation rates. Similarly, population genetic testing stratified on the basis of family history would necessitate testing a large proportion of the population. Extrapolation of the estimate of 9.4% for people with a history of colorectal cancer in a first degree relative, presented in chapter 4, to the Scottish population implies that nearly 250,000 people aged 30-70 would be offered testing if this criterion was used to determine eligibility for stratified genetic testing. Hence, both these approaches would involve conducting genetic testing in a much

larger proportion of the ostensibly healthy population, as compared to cascade genetic testing. The number of people offered genetic testing, and thus having to decide whether or not to participate, would also be much larger.

Theoretically, an exponential effect with respect to the cumulative number of pre-symptomatic tests performed may be anticipated, since cascade genetic testing involves offering genetic testing to an increasing number of family members as the mutation is traced through various relatives causing an expansion of pedigrees. Such an effect is evident in figure 4.3.22, but is very subtle. The probable reason for this is that the small size of sibships and the incomplete acceptance of genetic testing assumed by the computer model combine to limit the expansion of pedigrees through cascade genetic testing. The consequence of this would be that cascade genetic testing in a particular pedigree will be halted after one or two rounds of genetic testing, and hence all the relatives that will be tested for a particular index case will be tested within a short space of time.

Figure 4.3.22 shows that the average number of asymptomatic carriers predicted by the model to be identified during a 20-year cascade genetic testing programme at default settings was 321.2 (95% CI = 305.3, 337.1). Figure 4.4.1 illustrates the overall impact of a 20-year cascade genetic testing programme on the carrier population, as predicted by the computer model at default settings.

Figure 4.4.1 Involvement of Mutation Carrier Population In Cascade Genetic Testing at Default Settings



NB: All percentages refer to the percentage of all asymptomatic mutation carriers (i.e. 100*n / 1209). The number of carriers diagnosed with colorectal cancer and the number undergoing genetic testing are taken directly from summary model outputs; other values were subsequently calculated only for the purpose of figure 4.4.1, using the assumption of 90% acceptance amongst people offered genetic testing. Hence it was not possible to calculate confidence intervals for the majority of values displayed in this figure and these are provided for numbers accepting genetic testing only. Abbreviation: CI, confidence interval.

The above figure indicates that 45.4% of all mutation carriers will be offered genetic testing over the course of a 20-year cascade genetic testing programme, and implies

that 26.6% of all carriers will be offered testing whilst asymptomatic. However, as discussed previously, the model makes the assumption that index cases are unrelated, and accordingly that relatives identified as carriers through cascade genetic testing have not developed colorectal cancer themselves. In reality, some relatives of index cases will have a personal history of colorectal cancer, and consequently the estimated yield of cascade genetic testing provided by the computer model, in terms of the number of genuinely asymptomatic carriers identified, may be an overestimate.

Nonetheless, although limitations in terms of accuracy apply to estimates from the model, it is apparent that a significant proportion of mutation carriers in the Scottish population may be identified through cascade genetic testing at an asymptomatic stage. Offering and applying clinical surveillance to this group may significantly reduce their risk of colorectal cancer, and thus contribute towards the prevention of colorectal cancer at the population level. The actual health benefits that would result from the identification of a sub-group of MMR gene mutation carriers, in terms of colorectal cancer prevention and overall mortality, are difficult to assess accurately. A recent controlled trial found that the incidence of colorectal cancer over a 15-year period in MMR gene carriers undergoing colonoscopic screening was 18%, significantly less than the 41% incidence observed in carriers that did not undergo screening (130). This provides support for the hypothesis that the identification and screening of MMR gene mutation carriers provides a means of preventing colorectal cancer in this sub-group of the population, yet also illustrates another issue that may limit this approach; namely, that health benefits are conditional on compliance with

screening. Thus, at the population level, any reduction in morbidity and/or mortality that is associated with identification of MMR gene carriers will depend not only on the number of carriers concerned, but also on their age/sex distribution, the effectiveness of screening and the extent of compliance. The computer model does not, at present, have the capacity to consider these additional issues.

Ultimately, whether or not a cascade genetic testing programme is justifiable or desirable will depend on the costs involved in the identification of mutation carriers and their follow-up, as well as the health benefits that are expected to accrue. As mentioned previously, an economic evaluation of cascade genetic testing is outwith the scope of the computer model. The model is, however, designed to inform the planning of cascade genetic testing through the estimation of the yield of a particular cascade genetic testing programme and the resources required to achieve this. The relationship between these two key outcomes constitutes a crucial consideration in the context of planning the implementation of cascade genetic testing at the population level.

Throughout the results section, the number of asymptomatic mutation carriers identified per mutation analysis performed is used as an estimate of the relative efficiency of the cascade genetic testing programme in question. This is an inexact measurement of efficiency, but does provide a useful indication of the resources expended in the implementation of a cascade genetic testing programme in comparison to the yield. This estimate rises steeply over the first two years of cascade genetic testing, reflecting the year-long delay applied to conducting

mutation analysis. After this point, the ratio of number of asymptomatic carriers identified to mutation analyses performed continues to increase at a slower rate, peaking at 0.0449 (95% CI = 0.0429, 0.0469) after twenty years. This observation suggests that cascade genetic testing becomes relatively more efficient the longer the programme lasts for, lending support to the notion that cascade genetic testing should be viewed as a long-term strategy for identifying MMR gene mutation carriers. It is evident, however, that cascade genetic testing at default settings is not very efficient in terms of the ratio of yield to mutation analyses performed, with over 20 such tests being required to identify one asymptomatic mutation carrier even in a mature cascade genetic testing programme.

In the model system, pre-symptomatic tests are offered to first and second-degree relatives of known carriers. Hence, people undergoing pre-symptomatic testing will be at a 50% or 25% risk of harbouring a mutation themselves. The observation that the proportion of pre-symptomatic tests with a positive result was 0.38 (95% CI = 0.37, 0.39) thus implies that the ratio of first to second-degree relatives undergoing genetic testing is approximately 1:1. This high proportion of positive tests is one of the main advantages of cascade genetic testing, which is that relatives of known carriers form a sub-group of the population at an extremely high risk of harbouring a MMR gene mutation and thus present an ideal target for genetic testing.

4.4.4.2 Prevalence

Results provided by the computer model and presented in figures 4.3.29, 4.3.30 and 4.3.31 illustrate the major impact that the true prevalence of MMR gene mutations in

the Scottish population is likely to have on the outcomes of cascade genetic testing. Outcomes from the model at high and low prevalence settings differ considerably from outcomes at default settings with respect to the number of index cases identified. This in turn leads to significant differences in the number of pre-symptomatic tests performed and the overall yield of cascade genetic testing. The model predicts that the increase in yield associated with higher prevalence would be achieved with only a minor effect on the total number of mutation analyses required, as shown in figure 4.3.28, and that consequently the relative efficiency of cascade genetic testing is positively correlated with prevalence. This latter point is illustrated in figure 4.3.31. The true prevalence of MMR gene mutations in the population is thus a vital consideration in any assessment of the utility of cascade genetic testing. It is noteworthy that differences in prevalence do not, in theory, affect the proportion of MMR gene mutation carriers that are detected through cascade genetic testing.

There is clearly a need for precise information on the population prevalence of MMR gene mutations. For the purposes of planning cascade genetic testing, this should be specific to the population in question and should ideally be obtained directly from mutation analysis in randomly selected control subjects. At present such information is unavailable, largely due to the cost involved in mutation analysis. Improvements in laboratory technology, and the resources provided by large-scale programmes designed to obtain and analyse DNA from numerous healthy population controls (e.g. UK Biobank), may facilitate the gathering of such data in the future. An updated prevalence estimate based on COGS data will also be calculated as part of future work and incorporated into the computer model.

4.4.4.3 Penetrance

Alterations in penetrance of MMR gene mutations can be predicted to have a similar effect to prevalence on the yield and relative efficiency of cascade genetic testing, since higher penetrance will also result in a greater number of carriers developing colorectal cancer each year with a relatively minor increase in the overall number of mutation analyses required. Figure 4.3.33 illustrates the predicted magnitude of this effect on yield. In comparison with default settings, complete penetrance is predicted to result in an additional 200 asymptomatic mutation carriers being identified over a twenty-year period. The model also predicts that if true penetrance was only half the default estimates (i.e. 0.4 in males and 0.2 in females), 160 fewer asymptomatic carriers would be identified, a reduction of approximately half. However, whilst uncertainty remains over the true penetrance of MMR gene mutations, the vast majority of studies in this area have concluded that penetrance is incomplete (3, 4, 57, 289), and hence the high yield predicted by the model under the assumption of complete penetrance is unrealistic. Similarly, the penetrance is unlikely to be as low as 0.8 in males and 0.4 in females. A further consideration is that penetrance is a major factor in determining the appropriateness and extent of surveillance. Therefore, the identification of carriers of low penetrance mutations may not be considered appropriate.

Overall, the model predicts that penetrance will have a significant impact on the yield and efficiency of cascade genetic testing. When considered within probable boundaries, disparity between current estimates and true penetrance is likely to result in a difference in predicted yield of up to approximately 100 in either

direction. The magnitude of such differences are not as great as those caused by realistic variations in the prevalence estimate, but are highly relevant in terms of planning and implementing a real cascade genetic testing programme.

4.4.4.4 Age Limits

The age criteria for inclusion in any cascade genetic testing programme directly determine the number of mutation analyses that are required. Because early onset is a feature of colorectal cancer of hereditary origin, the proportion of MMR gene mutation carriers in a sub-group of colorectal cancer cases selected on the basis of age will increase as the age cut-off for inclusion is decreased. The resultant increase in the efficiency with which index cases are identified as part of a cascade genetic testing programme is counterbalanced by a decrease in the overall yield.

The necessity for applying some degree of age criteria to the ascertainment of colorectal cancer cases is amply illustrated by figure 4.3.35, which shows that approximately 42,500 mutation analyses would be required over twenty years if no such criteria were applied, at a rate of over 2100 per year. The resources and time required to conduct this number of tests are likely to be prohibitive. No difference is evident between the mean number of index cases predicted to be identified at default setting and with no age cut-off. Although repeated simulations may reveal a statistically significant difference in this regard, the model clearly shows that undertaking mutation analyses in colorectal cancer cases over the age of 55 is not likely to substantially increase the number of index cases identified as part of a cascade genetic testing programme. The model also predicts that limiting the age

cut-off value to less than 55 would result in considerably fewer mutation analyses being required. However, even a five-year decrease in age cut-off would result in a small but significant reduction in the number of index cases identified. The inherent trade-off between yield and efficiency that occurs when age limits are applied in this context is clearly evident in figure 4.3.38, which shows that the relative efficiency of cascade genetic testing is inversely proportional to the age limits applied.

Ultimately, deciding on which age criteria to apply to a cascade genetic testing programme is a pragmatic decision that must be made according to the resources available, and must also consider the appropriateness of offering testing to older colorectal cancer cases who are comparatively unlikely to harbour a MMR gene mutation. The default setting for age cut-offs of 55 years initially came from the COGS study, and the computer model supports the use of this value in the sense that it appears to facilitate the ascertainment of the vast majority of colorectal cancer cases with MMR gene mutations whilst keeping the number of mutation analyses performed to manageable levels.

4.4.4.5 Acceptance Rates

The level of participation on behalf of the target population is crucial to any genetic testing strategy, yet it is one of the most difficult aspects of cascade genetic testing to predict. The probability of an individual undergoing genetic testing may depend on numerous factors. Some of these, like age, sex and the context in which testing is offered, can be accommodated by the current model. Other factors that may influence acceptance include family history, ethnicity, social class, awareness of

genetic testing programmes and knowledge of colorectal cancer. The latter two factors in this list are of particular interest as they can potentially be manipulated through education and publicity generated in connection with a cascade genetic testing programme. Whilst the probability of participation at various stages of cascade genetic testing should ideally consider the above factors and be calculated specifically for each individual, inclusion of this level of detail is restricted by the paucity of relevant data and is currently outwith the scope of the model.

With all other variables, including other acceptance rates, being constant, the predicted number of mutation analyses performed is directly related to the acceptance rate of genetic testing by eligible colorectal cancer cases, and the predicted number of index cases identified will vary accordingly. However, as illustrated by figures 4.3.39 and 4.3.40, the family-based nature of cascade genetic testing will serve to exaggerate the effect of genetic test acceptance at this stage, if acceptance rates also apply to relatives of index cases. In a real-life setting, acceptance rates for index cases and relatives will probably differ, since the risk of having a mutation, the manner of approach and the situation (asymptomatic individual or cancer patient) will be entirely different. However, it is possible that external factors such as knowledge of colorectal cancer and awareness of the cascade genetic testing programme may influence acceptance rates of both potential index cases and relatives of mutation carriers.

In the cascade genetic testing system represented by the computer model, the major resource investment is required for the identification of index cases, whereas the

yield is crucially dependent on the actual process of tracing a mutation through asymptomatic individuals in the pedigree of a known carrier. Therefore, whilst yield is affected by the acceptance rate of genetic testing amongst both potential index cases and relatives of known carriers, the predominant effect on efficiency will stem from the acceptance rates amongst relatives. This point is illustrated in figures 4.3.42 and 4.3.43. In practical terms, this observation suggests that ensuring high participation amongst relatives should be of the highest priority.

A problem specific to cascade genetic testing arises from the possibility that a known mutation carrier will not consent to their relatives being contacted. The computer model predicts that this effect will lead to a reduction in both the yield and the efficiency of cascade genetic testing, with the magnitude of the effect being determined by the probability of allowing contact (see figures 4.3.44 & 4.3.45). Accordingly, the procedure for obtaining consent to contact relatives, and the rate at which such consent is forthcoming, may have implications for the utility of cascade genetic testing.

4.4.4.6 Delays

As demonstrated in the results section, delays in the model system affect the time taken to achieve certain model outcomes. However, such delays will have limited influence on the overall utility of the model system, and are mostly significant from a clinical perspective, rather than being a major consideration in terms of the outcomes of a cascade genetic testing programme. Additional resources may be employed to reduce delays in comparison with the default estimates obtained from

the COGS study. Additionally, future advancements in the technology used to conduct genetic testing may reduce the time required for this process. However, the incidence rate of colorectal cancer in the population will remain the primary limiting factor in terms of the timescale of any cascade genetic testing programme based on identifying index cases by targeting a sub-set of colorectal cancer patients.

4.4.4.7 Resources

The 'resources' component of the computer model provides the capacity to determine the requirements of a particular cascade genetic testing programme. The computer model demonstrates that inadequate resources at any stage can potentially cause a 'bottleneck' effect, whereby the entire programme is delayed. Estimates from the computer model are thus theoretically useful as a guide to the actual resources required by a particular cascade genetic testing programme. For example, it is apparent from figure 4.3.50 that a programme at default settings must have the capacity to conduct up to ten mutation analyses per week. Of course, as with other model outcomes, such estimates are subject to the limitations of the computer model outlined above, and must be interpreted with caution. In reality, it is unlikely that resource allocations would be entirely static. It is conceivable, for example, that a base resource designed to conduct five mutation analyses but with the capacity to stretch to ten when necessary would be adequate in the above circumstance. Overall, the model confirms the intuitive notion that the provision of a minimum resource level is crucial to the effective application of cascade genetic testing, but that increasing resources beyond the demand threshold will have no effect on outcomes.

4.4.5 Conclusions

The methods used for developing a computer model of cascade genetic testing are appropriate and the overall approach has been found to be highly effective. Limitations with respect to the input estimates and scope of the model mean that the model outcomes are subject to considerable uncertainty and potential bias, precluding the direct interpretation of these outcomes as accurate and reliable predictions of cascade genetic testing in a real-life situation. However, whilst the quantitative estimates provided by the computer model must be interpreted with caution, the computer model provides a detailed and realistic representation of cascade genetic testing, and the broad conclusions and observations obtained from the model are thus of relevance to understanding and planning the implementation of this strategy.

The computer model broadly supports the prior hypothesis that cascade genetic testing is potentially both feasible and effective as a means of identifying asymptomatic MMR gene mutation carriers in the Scottish population. At default settings, the model outcomes suggest that approximately 27% of all mutation carriers in the population will be identified in the course of a 20-year cascade genetic testing programme. The actual number of asymptomatic mutation carriers identified is predicted to total approximately 321.2 (95% CI = 305.3, 337.1). Applying colonoscopic surveillance to a group of mutation carriers of this size may confer significant health benefits at the individual and group level, and would potentially make a minor contribution towards the prevention of colorectal cancer in the population. To achieve this yield, the model predicts that around 7142 mutation

analyses would be required, which constitutes a major investment in terms of the time and resources involved.

The application of age limits to the ascertainment of colorectal cancer cases is a logical approach, since a relatively low proportion of older patients are likely to harbour MMR gene mutations. In this context, adjustment of the age limits facilitates the manipulation of the relationship between the yield from a cascade genetic testing programme and the number of mutation analyses required as part of that programme. Lowering the age limits decreases the yield but improves the efficiency, and vice versa. Outcomes from the model suggest that an age limit of 50 or 55 may represent an appropriate balance between yield and efficiency. Decisions regarding the age limits applied in a real cascade genetic testing programme could be informed by the model, but would be made on a pragmatic and subjective basis.

The computer model demonstrates that various other factors, including prevalence, penetrance and the acceptability of cascade genetic testing to the population concerned, will have a significant impact on yield and relative efficiency of cascade genetic testing. This emphasises the need for accurate data regarding these factors. The current unavailability of such data presents a considerable challenge to the evaluation of any strategy for identifying asymptomatic MMR gene mutation carriers, as well as limiting the utility of the computer model. In the absence of accurate and reliable input estimates, a broad spectrum of possible outcomes are predicted by the computer model, depending on the conditions under consideration.

The computer model is designed to provide quantitative estimates to inform decisions regarding the potential implementation of cascade genetic testing, and is not intended to make such a decision itself. Ultimately, the utility of cascade genetic testing must be judged in terms of the acceptability, effectiveness and cost of the interventions available for mutation carriers, as well as on the feasibility and cost of the cascade genetic testing process itself. There is some direct evidence to suggest that interventions in mutation carriers can lead to prevention of colorectal cancer in this group (130). It can thus be inferred that cascade genetic testing may be justifiable in the above terms, but the limitations of the current model preclude a conclusion in this regard.

Cascade genetic testing approaches the identification of MMR gene mutation carriers from the family perspective, as well as at the individual level. In this respect, identification of a mutation-carrying family may have beneficial effects beyond the timescale of the active cascade genetic testing programme, since specific genetic testing could theoretically be offered to family members in the future. Genetic testing of relatives of known carriers can potentially benefit both mutation carriers, who can be offered the appropriate level of clinical screening, and non-carriers, who can be reassured regarding their genetic risk of colorectal cancer, and removed from existing screening protocols.

The computer model only considers the identification of index cases through genetic testing of colorectal cancer patients. In reality, however, a cascade genetic testing programme applied to the Scottish population would be run against the background

of cancer genetic services. Mutation carriers identified through presentation to cancer genetic services with a strong family history of colorectal cancer could potentially be incorporated into a cascade genetic testing programme. The integration of these two strategies for identifying mutation carriers may constitute an ideal approach to cascade genetic testing.

4.4.6 Future Work

Outcomes from the current computer model provide a degree of insight into the potential outcomes of cascade genetic testing, with a chronological component permitting the assessment of the likely time-scale for this strategy. However, various limitations exist, both in the structure of the model and in the input estimates used to inform it. Hence the current model provides relevant data for analysis as part of this thesis, but in the wider context must be viewed as a prototype of an accurate computer model with practical applications.

Improved input estimates, a realistic method of generating pedigrees and defining the interrelatedness of mutation carriers, and a consideration of mutation analyses performed as part of clinical genetic services are required to provide truly accurate estimates relating to the outcomes of cascade genetic testing. Further work is thus necessary to address these limitations, as indicated in preceding sections of the text. Additionally, the flexibility inherent in the methods used for model development provide the capacity to adapt the current model to evaluate cascade genetic testing in any complex disease in which a sub-set of cases have a known genetic aetiology and for which effective interventions are available. On this premise, a two-year project

grant has been sought and obtained from the Chief Scientist Office (CSO) to pursue the further development of the current computer model of cascade genetic testing. The proposal for this grant is presented for reference in appendix A9.

Although the COGS programme is not designed to evaluate cascade genetic testing directly, it does involve the recruitment and genetic testing of families of known mutation carriers. This process is funded by an additional grant, and ethics approval was obtained in August 2003. Over the next few years, the COGS programme will provide valuable and unique information regarding important aspects of cascade genetic testing, including the acceptance rates of genetic testing among patients and various relative groups, the time-scale of genetic testing, and the laboratory and administrative costs involved. In addition, data from the COGS programme will provide the opportunity to calculate estimates of the prevalence and penetrance of mismatch repair gene mutation that are robust and specific to the Scottish population.

The COGS programme has also generated detailed pedigree information relating to colorectal cancer cases and mutations carriers. This real data will be utilised to simulate realistic pedigrees as part of the computer model. This will build on previous pedigree simulation work conducted by co-investigators on the CSO project grant (187), and will ensure that the family structures used in the model mirror the real-life situation.

Thus, the future development of the computer model will be integrated with the COGS programme. The model will initially reflect cascade genetic testing as it is currently being applied as part of this ongoing research. This will facilitate a high degree of external validation, since the predictions generated by the model can be compared with actual findings from the COGS programme. Subsequently, input estimates will be manipulated within realistic boundaries to investigate the effects of altering various model parameters. The combination of domain expertise, 'real' data, and an iterative development strategy involving continual re-evaluation and revision of the model, will thus facilitate the expansion of the model to the extent where it constitutes a full and realistic representation of cascade genetic testing for mismatch repair gene mutations. Conversely, the computer model will have the capacity to generate long-term estimates of the likely outcomes of the COGS programme, and will thus inform future planning of this and related programmes.

The next step in model development will be to investigate the adaptability of the model by applying it to other complex conditions in which pathogenic mutations are known to cause a proportion of cases. The identification of mutations in the breast cancer genes BRCA1 and BRCA2 presents an ideal candidate for such an investigation, since the genetic epidemiology of this condition has been extensively studied. Application of the computer model to investigate the use of cascade genetic testing in breast cancer will be conducted in a similar fashion to that outlined above. Inputs relating to breast cancer epidemiology, and the prevalence and penetrance of BRCA1/BRCA2 mutations etc. will be taken from the available literature. It is also intended to incorporate aspects of the "genISYS" project, funded by the CSO and

undertaken by the Image Systems Engineering Laboratory at Heriot-Watt University, with the aim of modelling breast cancer genetic risk analysis. Resulting model outcomes will be evaluated in the context of current knowledge about these genes.

Subsequently, it will be possible to apply the computer model to various other clinical situations, the only restriction being the availability of realistic input data. Once again, the model will be re-evaluated based on its response to different inputs, and will be adjusted accordingly if appropriate.

Ultimately, the completed computer model is intended to serve as a valuable tool for evaluating the feasibility and effectiveness of cascade genetic testing in a variety of complex conditions.

Chapter 5

Conclusion

5.1 Rationale

The central tenet of this thesis is that the identification of people at increased genetic risk of colorectal cancer, and the subsequent provision of appropriate clinical screening, provides an opportunity to reduce the incidence of colorectal cancer in this sub-group of the population, and consequently to contribute to the overall prevention of colorectal cancer in Scotland. At present, family history assessment and genetic testing for mismatch repair gene mutations constitute the two principal strategies for identifying individuals at increased genetic risk of colorectal cancer. Both approaches have been applied in the research context and in the clinical setting. However, current understanding of the associations between family history, mismatch repair gene mutations and risk of colorectal cancer is limited, and thus optimisation of strategies for people at increased genetic risk remains an ongoing challenge. The research presented within this thesis endeavours to contribute to current knowledge and understanding of the strategies available for identifying people at increased genetic risk of colorectal cancer in the Scottish population.

5.2 Family History Assessment

Family history has traditionally been the mainstay of genetic risk assessment, and it remains an integral part of current practice in clinical cancer genetics. The research presented in chapter 3 demonstrates the potential utility of family history as a means of identifying people at increased genetic risk, yet also illustrates some of the associated limitations and practical difficulties of this approach. This research confirms that family history is a significant risk factor for colorectal cancer in the Scottish population. The prevalence of a family history of colorectal cancer in the

general population of Scotland is shown to be high, with 9.4% of population controls aged 30-70 years having an affected first degree relative. These findings imply that people with a family history of colorectal cancer constitute a large sub-group of the Scottish population who are at significantly increased genetic risk, yet also emphasise the scale of the task of identifying and screening this group. The inherent limitations of genetic risk assessment based on family history are further compounded by the considerable degree of under-reporting and inaccuracy in family history information observed in data obtained at interview.

5.3 Genetic Testing for Mismatch Repair Gene Mutations

Genetic testing for predisposition to colorectal cancer, excluding very rare cancer syndromes such as Familial Adenomatous Polyposis, has become available relatively recently, following the implication of mismatch repair gene mutations in the aetiology of a sub-set of colorectal cancer cases. As illustrated by the literature review described within this thesis, such genetic testing also has considerable advantages and disadvantages as a strategy for identifying people at increased genetic risk of colorectal cancer. Mismatch repair gene mutation carriers are at a very high risk of developing the disease compared to the rest of the population, and thus constitute an important subgroup. However, genetic testing for mismatch repair gene mutations is expensive and time-consuming at present, necessitating the targeting of such analyses to individuals at high risk of carrying a mutation. This drawback stems from the extreme heterogeneity of mismatch repair genes. The fact that interpretation of gene variants remains somewhat subjective is a further concern. Such specific issues relating to mismatch repair genes apply in addition to

more general considerations regarding the practical and ethical application of genetic testing at the population level.

5.4 The Evaluation of Cascade Genetic Testing

In light of the above limitations of family history, and the high risk associated with mismatch repair gene mutations, it is apparent that identifying carriers of such mutations and providing appropriate clinical screening in this sub-group of the population constitutes a key aspect of the prevention of colorectal cancer through targeting individuals at increased genetic risk. Presently, the cost and time-consuming nature of conducting mutation analysis, coupled with the remaining uncertainty surrounding the interpretation of gene variants, precludes conventional population genetic testing for mismatch repair gene mutations. Population genetic testing stratified by family history provides another approach, but substantial resources would still be required, and the limitations of family history information outlined above and considered throughout this thesis are also applicable in this context. As an alternative to these options, cascade genetic testing currently represents the most realistic approach to identifying mismatch repair gene mutation carriers in the Scottish population. The main advantage of this approach is that genetic testing is restricted to people with a high probability of carrying a mismatch repair gene mutation and relatives of known carriers.

The major barrier to the systematic implementation of cascade genetic testing is the lack of data to inform the planning of an appropriate programme. The computer model developed and analysed as part of this thesis provides some relevant data to

help address the gaps in current knowledge of cascade genetic testing. The methods employed, based on visual object-orientated modelling, were appropriate and effective, facilitating a smooth transition from conceptual model to functional model, and finally to a working version. Data obtained as part of the systematic literature review described in chapter 2 was crucial to the conceptual construction of a computer model, and provided quantitative estimates to inform the model inputs. The use of real data from the ongoing COGS study to provide other key input estimates helped ensure the realistic nature of the computer model.

At default settings, the computer model constitutes a realistic representation of a cascade genetic testing system for identifying mismatch repair gene mutation carriers in the Scottish population. Model outcomes predict that approximately 27% of carriers in the population will be identified at an asymptomatic stage during a twenty-year cascade genetic testing programme. The model also demonstrates the major effect that penetrance and population prevalence of mismatch repair genes are likely to have on outcomes of cascade genetic testing, further emphasising the need for further studies into the genetic epidemiology of mismatch repair gene mutations. Additionally, the model predicts that outcomes from cascade genetic testing will be heavily influenced by acceptance rates and the age criteria for recruiting colorectal cancer cases. These are vital considerations in the planning of any cascade genetic testing programme.

Outcomes generated by the computer model generally support the hypothesis that cascade genetic testing is a potentially useful and effective strategy for identifying

asymptomatic mismatch repair gene mutation carriers in the Scottish population. However, the model is subject to various limitations in scope and detail, as well as inaccuracies arising from the poor quality of data available to inform input estimates. Consequently, further work is required to address these limitations and ensure that the outcomes from the computer model are sufficiently accurate to inform future planning and implementation of cascade genetic testing.

Cascade genetic testing provides an ideal framework for integration of the dual approaches of family history and genetic testing as tools for identifying people at increased genetic risk of colorectal cancer. Individuals with a family history, whether ascertained through clinical cancer genetics services or as part of a population-based intervention, can potentially be incorporated into a cascade genetic testing programme as possible mutation carriers. Asymptomatic individuals presenting to clinical cancer genetics services who meet certain family history criteria are currently eligible for genetic testing. This provides a clear alternative route into a hypothetical cascade genetic testing programme, and incorporating this process into the computer model is a priority for future work.

Conversely, cascade genetic testing has the potential to contribute towards the optimisation of strategies aimed at people with a family history of colorectal cancer. This conclusion is based on the principle that identifying a pathogenic mutation in an individual and subsequently testing family members for that mutation will allow elucidation of risk on an individual basis within a family. Hence, mutation carriers can be offered the intensive clinical screening warranted by this status, and non-

carriers could be reassured as to their personal risk and removed from screening. Effectively, this type of approach is already in use in the clinical setting, and this process may be enhanced through the systematic application of cascade genetic testing.

The computer model is currently designed to evaluate cascade genetic testing in relative isolation, whereas in reality this strategy must be considered in the context of ongoing clinical cancer genetics services and possibly the presence of alternative strategies for colorectal cancer prevention such as population-wide faecal occult blood testing. Economic considerations relating to the cascade genetic testing programme and the impact of such a programme on clinical services such as colonoscopy are likely to dictate whether or not cascade genetic testing is feasible, and determine the precise programme applied. At present, the computer model does not have the capacity to consider these factors directly, although data obtained from the model can be used to inform such considerations. Crucially, however, the methods used to develop the computer model are highly flexible, ensuring that additional considerations can be included in the model at a later stage.

The flexible and adaptable nature of the computer model also ensures that future advances in understanding of the genetic epidemiology of colorectal cancer can be incorporated into future versions. The model could be adapted to consider other genes implicated in the aetiology of colorectal cancer, and improved knowledge of the influence of genetic and environmental modifiers of risk mediated by mismatch repair gene mutations could also be accommodated. Additional quality data

regarding the prevalence and penetrance of mismatch repair gene mutations, and other factors such as acceptance rates, can be directly added to the model via the user interface.

Although the model is designed to represent cascade genetic testing for mismatch repair gene mutations, it is also sufficiently adaptable to potentially represent other genes and other clinical situations. This raises the possibility of applying the computer model to any complex disease in which causative genetic mutations have been identified. The broader application of the computer model of cascade genetic testing is a secondary consideration in the context of this thesis, but constitutes a vital feature of future model development.

5.5 The Population Impact of Strategies for Colorectal Cancer Prevention in People at Increased Genetic Risk

The identification of mismatch repair gene mutation carriers and the provision of appropriate screening forms a key element of an overall approach for preventing colorectal cancer through targeting those at increased genetic risk. Devising and implementing an appropriate strategy for achieving this may confer significant health benefits to this subgroup of the population. At the population level, however, the identification and screening of mismatch repair gene mutation carriers would have only a very minor impact on the overall incidence of colorectal cancer in the Scottish population. Moreover, the majority of people at increased genetic risk will not have a mismatch repair gene mutation. Therefore, strategies based on accurately establishing family history of colorectal cancer, and providing clinical screening accordingly, are essential to meet the needs of such people.

Although the magnitude of risk associated with a family history of colorectal cancer is less than that associated with mismatch repair gene mutations, this increased risk applies to a relatively large sub-group of the population. Hence the population impact of interventions aimed at people with a family history may be comparatively large. However, the numerous limitations of family history information, including the practical difficulties and potential inaccuracies inherent in obtaining such data, are such that population-based strategies to identify people at increased genetic risk of colorectal cancer through assessment of family history may not be appropriate at present. Additionally, clinical cancer genetic services already address the needs of people who are concerned about their family history, and the benefits of identifying and screening additional people, who may not be aware of or concerned about their family history, are unclear.

Strategies for people at increased genetic risk form just one approach to the prevention of colorectal cancer. Absolute risk is the crucial consideration in the context of secondary prevention through the provision of appropriate clinical screening, and hence strategies aimed at older members of the general population are highly relevant. From a public health perspective, primary prevention of colorectal cancer achieved through the adoption of a healthy lifestyle and diet is an ideal means of reducing the burden of colorectal cancer in Scotland. Nonetheless, people at increased genetic risk constitute a large and important sub-group of the Scottish population, and the targeting of such people represents an important aspect of an overall strategy for colorectal cancer prevention.

5.6 Summary

The conclusion of this thesis is that cascade genetic testing for mismatch repair gene mutations merits further evaluation as a potential strategy for addressing the needs of people at increased genetic risk of colorectal cancer. Integration of this approach with ongoing clinical cancer genetics services based on family history currently represents an appropriate framework for developing and improving strategies for prevention of colorectal cancer in people at increased genetic risk. The research presented herein contributes to the information available to facilitate this goal, but further research relating to the molecular genetic epidemiology of colorectal cancer and the utility of family history is required to determine the optimal approach to identifying such people. Ultimately, targeting people at increased genetic risk of colorectal cancer has the potential to provide substantial health benefits to a small sub-group of the population, and thus to make a minor contribution to the overall prevention of colorectal cancer in Scotland.

REFERENCES

1. Aaltonen LA, Peltomaki P, Leach FS et al. Clues to the pathogenesis of familial colorectal cancer. Science 1993; 260:812-6.

2. Aaltonen LA, Salovaara R, Kristo P et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. New England Journal of Medicine 1998; 338:1481-7.

3. Aarnio M, Mecklin JP, Aaltonen LA et al. Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (HNPCC) syndrome. International Journal of Cancer 1995; 64:430-3.

4. Aarnio M, Sankila R, Pukkala E et al. Cancer risk in mutation carriers of DNA-mismatch-repair genes. International Journal of Cancer 1999; 81:214-8.

5. Aebi S, Fink D, Gordon R et al. Resistance to cytotoxic drugs in DNA mismatch repair-deficient cells. Clinical Cancer Research. 1997; 3:1763-7.

6. Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. American Statistician 1998; 52:119-26.

7. Aitken J, Bain C, Ward M et al. How accurate is self-reported family history of colorectal cancer? American Journal of Epidemiology 1995; 141:863-71.

8. Alberts DS, Martinez ME, Roe DJ et al. Lack of effect of a high-fiber cereal supplement on the recurrence of colorectal adenomas. Phoenix Colon Cancer Prevention Physicians' Network. New England Journal of Medicine 2000; 342:1156-62.

9. American Cancer Society. Cancer Facts and Figures 2002. www.cancer.org. 2002.

10. Armitage NC, Farrands PA, Mangham CM et al. Faecal occult blood screening of first degree relatives of patients with colorectal cancer. International Journal of Colorectal Disease 1986; 1:248-50.

11. Bai YQ, Akiyama Y, Nagasaki H et al. Predominant germ-line mutation of the hMSH2 gene in Japanese hereditary non-polyposis colorectal cancer kindreds. International Journal of Cancer 1999; 82:512-5.

12. Bapat BV, Madlensky L, Temple LKF et al. Family history characteristics, tumor microsatellite instability and germline MSH2 and MLH1 mutations in hereditary colorectal cancer. Human Genetics 1999; 104:167-76.

13. Baron JA, Cole BF, Sandler RS et al. A randomized trial of aspirin to prevent colorectal adenomas. New England Journal of Medicine 2003; 348:891-9.

14. Beck NE, Tomlinson IP, Homfray T et al. Genetic testing is important in families with a history suggestive of hereditary non-polyposis colorectal cancer even if the Amsterdam criteria are not fulfilled. British Journal of Surgery 1997; 84:233-7.

15. Bellacosa A, Genuardi M, Anti M et al. Hereditary nonpolyposis colorectal cancer: review of clinical, molecular genetics, and counseling aspects. American Journal of Medical Genetics 1996; 62:353-64.

16. Benamouzig R, Deyra J, Martin A et al. Daily soluble aspirin and prevention of colorectal adenoma recurrence: one-year results of the APACC trial. Gastroenterology 2003; 125:328-36.

17. Boardman LA, Couch FJ, Burgart LJ et al. Genetic heterogeneity in Peutz-Jeghers syndrome. Human Mutation 2000; 16:23-30.

18. Bodmer WF, Bailey CJ, Bodmer J et al. Localization of the gene for familial adenomatous polyposis on chromosome 5. Nature 1987; 328:614-6.

19. Boland CR. Molecular genetics of hereditary nonpolyposis colorectal cancer. Annals of the New York Academy of Sciences. 2000; 910:50-9.

20. Bond JH. Virtual colonoscopy: promising, but not ready for widespread use. New England Journal of Medicine 1999; 341:1540-2.

21. Bondy ML, Strom SS, Colopy MW et al. Accuracy of family history of cancer obtained through interviews with relatives of patients with childhood sarcoma. Journal of Clinical Epidemiology 1994; 47:89-96.

22. Bonelli L, Martines H, Conio M et al. Family history of colorectal cancer as a risk factor for benign and malignant tumours of the large bowel. A case-control study. International Journal of Cancer 1988; 41:513-7.

23. Borresen AL, Lothe RA, Meling GI et al. Somatic mutations in the hMSH2 gene in microsatellite unstable colorectal carcinomas. Human Molecular Genetics. 1995; 4:2065-72.

24. Boutron-Ruault MC, Senesse P, Meance S et al. Energy intake, body mass index, physical activity, and the colorectal adenoma-carcinoma sequence. Nutrition & Cancer 2001; 39:50-7.

25. Boutron MC, Faivre J, Quipourt V et al. Family history of colorectal tumours and implications for the adenoma- carcinoma sequence: a case control study. Gut 1995; 37:830-4.

26. Brewster D, Muir C, Crichton J. Registration of colorectal cancer in Scotland: an assessment of data accuracy based on review of medical records. Public Health 1995; 109:285-92.

27. Brewster DH, Crichton J, Harvey JC et al. Completeness of case ascertainment in a Scottish regional cancer registry for the year 1992. Public Health 1997; 111:339-43.

28. Brockton N, Little J, Sharp L et al. N-acetyltransferase polymorphisms and colorectal cancer: a HuGE review. American Journal of Epidemiology 2000; 151:846-61.

29. Bronner CE, Baker SM, Morrison PT et al. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature 1994; 368:258-61.

30. Brown ML, Kessler LG. The use of gene tests to detect hereditary predisposition to cancer: economic considerations. Journal of the National Cancer Institute 1995; 87:1131-6.

31. Brown ML, Kessler LG. Use of gene tests to detect hereditary predisposition to cancer: what do we know about cost effectiveness? International Journal of Cancer 1996; 69:55-7.

32. Bubb VJ, Curtis LJ, Cunningham C et al. Microsatellite instability and the role of hMSH2 in sporadic colorectalcancer. Oncogene 1996; 12:2641-9.

33. Buerstedde JM, Alday P, Torhorst J et al. Detection of new mutations in six out of 10 Swiss HNPCC families by genomic sequencing of the hMSH2 and hMLH1 genes. Journal of Medical Genetics 1995; 32:909-12.

34. Burke W, Petersen G, Lynch P et al. Recommendations for follow-up care of individuals with an inherited predisposition to cancer. I. Hereditary nonpolyposis colon cancer. Cancer Genetics Studies Consortium. Journal of the American Medical Association 1997; 277:915-9.

35. Calistri D, Presciuttini S, Buonsanti G et al. Microsatellite instability in colorectal-cancer patients with suspected genetic predisposition. International Journal of Cancer 2000; 89:87-91.

36. Canzian F, Ushijima T, Serikawa T et al. Instability of microsatellites in rat colon tumors induced by heterocyclic amines. Cancer Research 1994; 54:6315-7.

37. Carethers JM, Chauhan DP, Fink D et al. Mismatch repair proficiency and in vitro response to 5-fluorouracil. Gastroenterology 1999; 117:123-31.

38. Carpenter S, Broughton M, Marks CG. A screening clinic for relatives of patients with colorectal cancer in a district general hospital. Gut 1995; 36:90-2.

39. Carstensen B, Soll-Johanning H, Villadsen E et al. Familial aggregation of colorectal cancer in the general population. Internation Journal of Cancer 1996; 68:428-35.

40. Chapman PD, Burn J. Genetic predictive testing for bowel cancer predisposition: the impact on the individual. Cytogenetics and Cell Genetics 1999; 86:118-24.

41. Church JM. Prophylactic colectomy in patients with hereditary nonpolyposis colorectal cancer. Annals of Medicine 1996; 28:479-82.

42. Coogan PF, Rosenberg L, Louik C et al. NSAIDs and risk of colorectal cancer according to presence or absence of family history of the disease. Cancer Causes & Control 2000; 11:249-55.

43. Cotton SC, Sharp L, Little J et al. Glutathione S-transferase polymorphisms and colorectal cancer: a HuGE review. American Journal of Epidemiology 2000; 151:7-32.

44. Coughlin SS, Miller DS. Public health perspectives on testing for colorectal cancer susceptibility genes. American Journal of Preventive Medicine 1999; 16:99-104.

45. Cripps NP, Heald RJ. Family based colorectal cancer screening in a district hospital. Gut 1996; 38:421-5.

46. Croyle RT, Lerman C. Risk communication in genetic testing for cancer susceptibility. Journal of the National Cancer Institute Monographs 1999; 59-66.

47. Curia MC, Palmirotta R, Aceto G et al. Unbalanced germ-line expression of hMLH1 and hMSH2 alleles in hereditary nonpolyposis colorectal cancer. Cancer Research 1999; 59:3570-5.

48. de Jong MM, Nolte IM, te Meerman GJ et al. Low-penetrance genes and their involvement in colorectal cancer susceptibility. Cancer Epidemiology, Biomarkers & Prevention 2002; 11:1332-52.

49. de Leon MP, Pedroni M, Benatti P et al. Hereditary colorectal cancer in the general population: from cancer registration to molecular diagnosis. Gut 1999; 45:32-8.

50. Debniak T, Kurzawski G, Gorski B et al. Value of pedigree/clinical data, immunohistochemistry and microsatellite instability analyses in reducing the cost of determining hMLH1 and hMSH2 gene mutations in patients with colorectal cancer. European Journal of Cancer 2000; 36:49-54.

51. Desai DC, Lockman JC, Chadwick RB et al. Recurrent germline mutation in MSH2 arises frequently de novo. Journal of Medical Genetics 2000; 37:646-52.

52. Dieumegard B, Grandjouan S, Sabourin J-C et al. Extensive molecular screening for hereditary non-polyposis colorectal cancer. British Journal of Cancer 2000; 82:871-80.

53. Douglas FS, O'Dair LC, Robinson M et al. The accuracy of diagnoses as reported in families with cancer: a retrospective study. Journal of Medical Genetics 1999; 36:309-12.

54. Dowling DJ, St John DJ, Macrae FA et al. Yield from colonoscopic screening in people with a strong family history of common colorectal cancer. Journal of Gastroenterology & Hepatology 2000; 15:939-44.

55. Dunlop MG, British Society for Gastroenterology, The Association of Coloproctology for Great Britain and Ireland. Guidance on gastrointestinal surveillance for hereditary non-polyposis colorectal cancer, familial adenomatous polypolis, juvenile polyposis, and Peutz-Jeghers syndrome. Gut 2002; 51 Suppl 5:V21-V27.

56. Dunlop MG, British Society for Gastroenterology, The Association of Coloproctology for Great Britain and Ireland. Guidance on large bowel surveillance for people with two first degree relatives with colorectal cancer or one first degree relative diagnosed with colorectal cancer under 45 years. Gut 2002; 51 Suppl 5:V17-V20.

57. Dunlop MG, Farrington SM, Carothers AD et al. Cancer risk associated with germline DNA mismatch repair gene mutations. Human Molecular Genetics. 1997; 6:105-10.

58. Dunlop MG, Farrington SM, Nicholl I et al. Population carrier frequency of hMSH2 and hMLH1 mutations. British Journal of Cancer 2000; 83:1643-5.

59. Ellison AR, Lofing J, Bitter GA. Functional analysis of human MLH1 and MSH2 missense variants and hybrid human-yeast MLH1 proteins in Saccharomyces cerevisiae. Human Molecular Genetics 2001; 10:1889-900.

60. Eng C, Hampel H, de la CA. Genetic testing for cancer predisposition. Annual Review of Medicine 2001; 52:371-400.

61. Enger SM, Longnecker MP, Lee ER et al. Recent and past physical activity and prevalence of colorectal adenomas. British Journal of Cancer 1997; 75:740-5.

62. Farrington SM, Lin-Goerke J, Ling J et al. Systematic analysis of hMSH2 and hMLH1 in young colon cancer patients and controls. American Journal of Human Genetics 1998; 63:749-59.

63. Featherstone C. Aspirin for bowel cancer: an old friend finds a new role. Lancet 1997; 350:418.

64. Fenlon HM, Nunes DP, Schroy PC, III et al. A comparison of virtual and conventional colonoscopy for the detection of colorectal polyps. New England Journal of Medicine 1999; 341:1496-503.

65. Fernandez E, La Vecchia C, D'Avanzo B et al. Risk factors for colorectal cancer in subjects with family history of the disease. British Journal of Cancer 1997; 75:1381-4.

66. Fernandez E, La Vecchia C, Talamini R et al. Joint effects of family history and adult life dietary risk factors on colorectal cancer risk. Epidemiology 2002; 13:360-3.

67. Fidalgo P, Almeida MR, West S et al. Detection of mutations in mismatch repair genes in Portuguese families with hereditary non-polyposis colorectal cancer (HNPCC) by a multi-method approach. European Journal of Human Genetics 2000; 8:49-53.

68. Fink D, Aebi S, Howell SB. The role of DNA mismatch repair in drug resistance. Clinical Cancer Research. 1998; 4:1-6.

69. Fishel R, Lescoe MK, Rao MR et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 1993; 75:1027-38.

70. Fisher G, Armstrong B. Familial colorectal cancer and the screening of family members. Medical Journal of Australia 1989; 150:22-5.

71. Floderus B, Barlow L, Mack TM. Recall bias in subjective reports of familial cancer. Epidemiology 1990; 1:318-21.

72. Fornasarig M, Viel A, Valentini M et al. Microsatellite instability and MLH1 and MSH2 germline defects are related to clinicopathological features in sporadic colorectal cancer. Oncology Reports 2000; 7:39-43.

73. Fouchier SW, Defesche JC, Umans-Eckenhausen MW et al. The molecular basis of familial hypercholesterolemia in The Netherlands. Human Genetics 2001; 109:602-15.

74. Friedman LC, Webb JA, Richards CS et al. Psychological and behavioral factors associated with colorectal cancer screening among Ashkenazim. Preventive Medicine 1999; 29:119-25.

75. Froggatt NJ, Brassett C, Koch DJ et al. Mutation screening of MSH2 and MLH1 mRNA in hereditary non-polyposis colon cancer syndrome. Journal of Medical Genetics 1996; 33:726-30.

76. Froggatt NJ, Green J, Brassett C et al. A common MSH2 mutation in English and North American HNPCC families: Origin, phenotypic expression, and sex

specific differences in colorectal cancer. Journal of Medical Genetics 1999; 36:97-102.

77. Froggatt NJ, Koch J, Davies R et al. Genetic linkage analysis in hereditary non-polyposis colon cancer syndrome. Journal of Medical Genetics 1995; 32:352-7.

78. Fuchs CS, Giovannucci EL, Colditz GA et al. A prospective study of family history and the risk of colorectal cancer. New England Journal of Medicine 1994; 331:1669-74.

79. Fuchs CS, Giovannucci EL, Colditz GA et al. Dietary fiber and the risk of colorectal cancer and adenoma in women. New England Journal of Medicine 1999; 340:169-76.

80. Fuchs CS, Willett WC, Colditz GA et al. The influence of folate and multivitamin use on the familial risk of colon cancer in women. Cancer Epidemiology, Biomarkers & Prevention 2002; 11:227-34.

81. Fujieda S, Tanaka N, Sunaga H et al. Expression of hMSH2 correlates with in vitro chemosensitivity to CDDP cytotoxicity in oral and oropharyngeal carcinoma. Cancer Lett. 1998; 132:37-44.

82. Fujita S, Moriya Y, Sugihara K et al. Prognosis of hereditary nonpolyposis colorectal cancer (HNPCC) and the role of Japanese criteria for HNPCC. Japanese Journal of Clinical Oncology 1996; 26:351-5.

83. Gaglia P, Atkin WS, Whitelaw S et al. Variables associated with the risk of colorectal adenomas in asymptomatic patients with a family history of colorectal cancer. Gut 1995; 36:385-90.

84. General Register Office for Scotland. Projected Population of Scotland (2002 Based). General Register Office for Scotland . www.gro-scotland.gov.uk 2002.

85. Giacosa A, Franceschi S, La Vecchia C et al. Energy intake, overweight, physical exercise and colorectal cancer risk. European Journal of Cancer Prevention. 1999; 8 Suppl 1:S53-60.:S53-S60.

86. Giardiello FM, Hamilton SR, Krush AJ et al. Treatment of colonic and rectal adenomas with sulindac in familial adenomatous polyposis. New England Journal of Medicine 1993; 328:1313-6.

87. Giovannucci E. Insulin and colon cancer. Cancer Causes Control 1995; 6:164-79.

88. Giovannucci E. An updated review of the epidemiological evidence that cigarette smoking increases risk of colorectal cancer. Cancer Epidemiology Biomarkers & Prevention. 2001; 10:725-31.

89. Giovannucci E. Insulin, insulin-like growth factors and colon cancer: a review of the evidence. J.Nutr. 2001; 131:3109S-20S.

90. Giovannucci E, Colditz GA, Stampfer MJ et al. Physical activity, obesity, and risk of colorectal adenoma in women (United States). Cancer Causes & Control 1996; 7:253-63.

91. Glanz K, Grove J, Le Marchand L et al. Underreporting of family history of colon cancer: correlates and implications. Cancer Epidemiology Biomarkers & Prevention 1999; 8:635-9.

92. Goldgar DE, Easton DF, Cannon-Albright LA et al. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. Journal of the National Cancer Institute 1994; 86:1600-8.

93. Green SE, Chapman PD, Burn J et al. Clinical impact of colonoscopic screening in first-degree relatives of patients with hereditary non-polyposis colorectal cancer. British Journal of Surgery. 1995; 82:1338-40.

94. Greenberg ER, Baron JA, Freeman DH, Jr. et al. Reduced risk of large-bowel adenomas among aspirin users. The Polyp Prevention Study Group. Journal of the National Cancer Institute 1993; 85:912-6.

95. Groden J, Thliveris A, Samowitz W et al. Identification and characterization of the familial adenomatous polyposis coli gene. Cell 1991; 66:589-600.

96. Guillem JG, Forde KA, Treat MR et al. Colonoscopic screening for neoplasms in asymptomatic first-degree relatives of colon cancer patients. A controlled, prospective study. Diseases of the Colon & Rectum 1992; 35:523-9.

97. Hall NR, Finan PJ, Ward B et al. Genetic susceptibility to colorectal cancer in patients under 45 years of age. British Journal of Surgery 1994; 81:1485-9.

98. Han H-J, Yuan Y, Ku J-L et al. Germline mutations of hMLH1 and hMSH2 genes in Korean hereditary nonpolyposis colorectal cancer. Journal of the National Cancer Institute 1996; 88:1317-9.

99. Hara AK, Johnson CD, Reed JE et al. Detection of colorectal polyps with CT colography: initial assessment of sensitivity and specificity. Radiology 1997; 205:59-65.

100. Hardcastle JD, Chamberlain JO, Robinson MH et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 1996; 348:1472-7.

101. Hardman AE. Physical activity and cancer risk. Proceedings of the Nutrition Society 2001; 60:107-13.

102. Harris V et al. Cancer Registration Statistics Scotland 1986-1995. Edinburgh: ISD Scotland Publications, 1998.

103. Hawk E, Lubet R, Limburg P. Chemoprevention in hereditary colorectal cancer syndromes. Cancer 1999; 86:2551-63.

104. Heasman MA, Clarke JA. Medical record linkage in Scotland. Health Bulletin (Edinburgh) 1979; 37:97-103.

105. Heinimann K, Scott RJ, Buerstedde J-M et al. Influence of selection criteria on mutation detection in patients with hereditary nonpolyposis colorectal cancer. Cancer 1999; 85:2512-8.

106. Hemminki K., Lonnstedt I., Vaittinen P. et al. Estimation of genetic and environmental components in colorectal and lung cancer and melanoma. Genetic Epidemiology 12001; 20:107-16.

107. Hemminki A, Markie D, Tomlinson I et al. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. Nature 1998; 391:184-7.

108. Hill MJ, Morson BC, Bussey HJ. Aetiology of adenoma--carcinoma sequence in large bowel. Lancet 1978; 1:245-7.

109. Hodgson SV, Bishop DT, Dunlop MG et al. Suggested screening guidelines for familial colorectal cancer. Journal of Medical Screening 1995; 2:45-51.

110. Holloway S, Brock DJ. Cascade testing for the identification of carriers of cystic fibrosis. Journal of Medical Screening 1994; 1:159-64.

111. Houlston RS, Murday V, Harocopos C et al. Screening and genetic counselling for relatives of patients with colorectal cancer in a family cancer clinic. British Medical Journal 1990; 301:366-8.

112. House W, Sharp D, Sheridan E. Identifying and screening patients at high risk of colorectal cancer in general practice. Journal of Medical Screening 1999; 6:205-8.

113. Howe JR, Roth S, Ringold JC et al. Mutations in the SMAD4/DPC4 gene in juvenile polyposis. Science 1998; 280:1086-8.

114. HuGE Net. Human Genome Epidemiology Network. The Centers for Disease Control and Prevention (CDC) www.cdc.gov/genomics/hugenet/default.htm 2003.

115. Hunt LM, Rooney PS, Bostock K et al. Chemical and immunological testing for faecal occult blood in screening subjects at risk of familial colorectal cancer. Gut 1997; 40:110-2.

116. Hunt LM, Rooney PS, Hardcastle JD et al. Endoscopic screening of relatives of patients with colorectal cancer. Gut 1998; 42:71-5.

117. Hutter P, Couturier A, Membrez V et al. Excess of hMLH1 germline mutations in Swiss families with hereditary non-polyposis colorectal cancer. International Journal of Cancer 1998; 78:680-4.

118. IARC database. Globocan 2000; Cancer Incidence, Mortality and Prevalence Worldwide. WHO. www.who.int 2000.

119. IARC, WHO. Cancer Incidence in Five Continents. Lyon: IARC Scientific Publications, 1997.

120. Imperiale TF. Aspirin and the prevention of colorectal cancer. New England Journal of Medicine 2003; 348:879-80.

121. Imperiale TF, Wagner DR, Lin CY et al. Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings. New England Journal of Medicine 2000; 343:169-74.

122. Information and Statistics Division. The Scottish Record Linkage System (ISD information document). 2002.

123. Information and Statistics Division. Cancer Facts and Figures. ISD Scotland . www.show.scot.nhs.uk/isd/ 2003.

124. International Agency for Research on Cancer. IARC Handbooks of Cancer Prevention Vol.6: Weight Control and Physical Activity. Lyon: IARC, 2002.

125. International Agency for Research on Cancer. Nutrition and Lifestyle: Opportunities for Cancer Prevention. Lyon: IARC, 2002.

126. International Agency for Research on Cancer. IARC Handbooks of Cancer Prevention Vol.8: Fruit and Vegetables. Lyon: IARC, 2003.

127. International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. ICG-HNPCC mutation database. ICG-HNPCC. www.nfdht.nl/ 2001.

128. Ionov Y, Peinado MA, Malkhosyan S et al. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 1993; 363:558-61.

129. Jager AC, Bisgaard ML, Myrhoj T et al. Reduced frequency of extracolonic cancers in hereditary nonpolyposis colorectal cancer families with monoallelic hMLH1 expression. American Journal of Human Genetics 1997; 61:129-38.

130. Jarvinen HJ, Aarnio M, Mustonen H et al. Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. Gastroenterology 2000; 118:829-34.

131. Jarvinen HJ, Mecklin JP, Sistonen P. Screening reduces colorectal cancer rate in families with hereditary nonpolyposis colorectal cancer.[comment]. Gastroenterology 1995; 108:1405-11.

132. Jass JR. Diagnosis of hereditary non-polyposis colorectal cancer. Histopathology 1998; 32:491-7.

133. Java SMI. http://java.sun.com/ 2003.

134. Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. American Journal of Gastroenterology 2001; 96:2992-3003.

135. Johnson N, Lancaster T, Fuller A et al. The prevalence of a family history of cancer in general practice. Family Practice 1995; 12:287-9.

136. Karner-Hanusch J, Mittlbock M, Fillipitsch T et al. Family history as a marker of risk for colorectal cancer: Austrian experience. World Journal of Surgery 1997; 21:205-9.

137. Katballe N, Juul S, Christensen M et al. Patient accuracy of reporting on hereditary non-polyposis colorectal cancer-related malignancy in family members. British Journal of Surgery 2001; 88:1228-33.

138. Kee F, Collins BJ. How prevalent is cancer family syndrome? Gut 1991; 32:509-12.

139. Kendrick SW, Clarke JA. The Scottish Medical Record Linkage System. Health Bulletin (Edinburgh) 1979; 51:72-9.

140. Kerber RA, Slattery ML. Comparison of self-reported and database-linked family history of cancer data in a case-control study. American Journal of Epidemiology 1997; 146:244-8.

141. Kerber RA, Slattery ML, Potter JD et al. Risk of colon cancer associated with a family history of cancer or colorectal polyps: the diet, activity, and reproduction in colon cancer study. International Journal of Cancer 1998; 78:157-60.

142. Kewenter J, Brevinge H, Engaras B et al. Results of screening, rescreening, and follow-up in a prospective randomized study for detection of colorectal cancer by fecal occult blood testing. Results for 68,308 subjects. Scandinavian Journal of Gastroenterology 1994; 29:468-73.

143. Kim YI. Diet, lifestyle, and colorectal cancer: is hyperinsulinemia the missing link? Nutrition Review 1998; 56:275-9.

144. Knudson AG. Antioncogenes and human cancer. Proceedings of the National Academy of Sciences USA 1993; 90:10914-21.

145. Koch M, Hill GB. Problems in establishing accurate family history in patients with ovarian cancer of epithelial origin. Cancer Detection & Prevention 1987; 10:279-83.

146. Kohonen-Corish M, Ross VL, Doe WF et al. RNA-based mutation screening in hereditary nonpolyposis colorectal cancer. American Journal of Human Genetics 1996; 59:818-24.

147. Kolodner R. Biochemistry and genetics of eukaryotic mismatch repair. Genes & Development 1996; 10:1433-42.

148. Kong S, Amos CI, Luthra R et al. Effects of cyclin D1 polymorphism on age of onset of hereditary nonpolyposis colorectal cancer. Cancer Research 2000; 60:249-52.

149. Kotake K, Koyama Y, Nasu J et al. Relation of family history of cancer and environmental factors to the risk of colorectal cancer: a case-control study. Japanese Journal of Clinical Oncology 1995; 25:195-202.

150. Krawczak M, Cooper DN, Schmidtke J. Estimating the efficacy and efficiency of cascade genetic screening. American Journal of Human Genetics 2001; 69:361-70.

151. Kronborg O, Fenger C, Olsen J et al. Randomised study of screening for colorectal cancer with faecal-occult- blood test. Lancet 1996; 348:1467-71.

152. Kuijten RR, Bunin GR, Meadows T. Validation and accuracy of cancer reporting in first-degree relatives (SER abstract). American Journal of Epidemiology 1992; 136:1019.

153. Kune GA, Kune S, Watson LF. The role of heredity in the etiology of large bowel cancer: data from the Melbourne Colorectal Cancer Study. World Journal of Surgery 1989; 13:124-9.

154. La Vecchia C, Gallus S, Talamini R et al. Interaction between selected environmental factors and familial propensity for colon cancer. European Journal of Cancer Prevention 1999; 8:147-50.

155. La Vecchia C, Negri E, Decarli A et al. Diabetes mellitus and colorectal cancer risk. Cancer Epidemiology Biomarkers & Prevention 1997; 6:1007-10.

156. La Vecchia C, Negri E, Franceschi S et al. Aspirin and colorectal cancer. British Journal of Cancer 1997; 76:675-7.

157. La Vecchia C, Negri E, Franceschi S et al. Family history and the risk of stomach and colorectal cancer. Cancer 1992; 70:50-5.

158. Lamberti C, Kruse R, Ruelfs C et al. Microsatellite instability-a useful diagnostic tool to select patients at high risk for hereditary non-polyposis

colorectal cancer: a study in different groups of patients with colorectal cancer. Gut 1999; 44:839-43.

159. Le Marchand L, Wilkens LR, Hankin JH et al. Independent and joint effects of family history and lifestyle on colorectal cancer risk: implications for prevention. Cancer Epidemiology Biomarkers & Prevention 1999; 8:45-51.

160. Le Marchand L, Zhao LP, Quiaoit F et al. Family history and risk of colorectal cancer in the multiethnic population of Hawaii. American Journal of Epidemiology 1996; 144:1122-8.

161. Leach FS, Nicolaides NC, Papadopoulos N et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell 1993; 75:1215-25.

162. Leggatt V, Mackay J, Yates JR. Evaluation of questionnaire on cancer family history in identifying patients at increased genetic risk in general practice. British Medical Journal 1999; 319:757-8.

163. Leppert M, Dobbs M, Scambler P et al. The gene for familial polyposis coli maps to the long arm of chromosome 5. Science 1987; 238:1411-3.

164. Lerman C, Hughes C, Trock BJ et al. Genetic testing in families with hereditary nonpolyposis colon cancer. Journal of the American Medical Association 1999; 281:1618-22.

165. Lerman C, Marshall J, Audrain J et al. Genetic testing for colon cancer susceptibility: Anticipated reactions of patients and challenges to providers. International Journal of Cancer 1996; 69:58-61.

166. Levin TR, Palitz AM. Flexible sigmoidoscopy: an important screening option for average-risk individuals. Gastrointestinal Endoscopy Clinics of North America 2002; 12:23-40.

167. Liang KY, Beaty TH, Cohen BH. Application of odds ratio regression models for assessing familial aggregation from case-control studies. American Journal of Epidemiology 1986; 124:678-83.

168. Lichtenstein P, Holm NV, Verkasalo PK et al. Environmental and heritable factors in the causation of cancer-- analyses of cohorts of twins from Sweden, Denmark, and Finland. New England Journal of Medicine 2000; 343:78-85.

169. Lieberman DA, Weiss DG, Bond JH et al. Use of colonoscopy to screen asymptomatic adults for colorectal cancer. Veterans Affairs Cooperative Study Group 380. New England Journal of Medicine 2000; 343:162-8.

170. Lin KM, Shashidharan M, Thorson AG et al. Cumulative incidence of colorectal and extracolonic cancers in MLH1 and MSH2 mutation carriers of hereditary nonpolyposis colorectal cancer. Journal of Gastrointestinal Surgery 1998; 2:67-71.

171. Lindblom A, Tannergard P, Werelius B et al. Genetic mapping of a second locus predisposing to hereditary non- polyposis colon cancer. Nature Genetics 1993; 5:279-82.

172. Lindgren G, Liljegren A, Jaramillo E et al. Adenoma prevalence and cancer risk in familial non-polyposis colorectal cancer. Gut 2002; 50:228-34.

173. Liu B, Farrington SM, Petersen GM et al. Genetic instability occurs in the majority of young patients with colorectal cancer. Nature Medicine 1995; 1:348-52.

174. Liu T, Wahlberg S, Burek E et al. Microsatellite instability as a predictor of a mutation in a DNA mismatch repair gene in familial colorectal cancer. Genes, Chromosomes & Cancer 2000; 27:17-25.

175. Liu T, Wahlberg S, Rubio C et al. DGGE screening of mutations in mismatch repair genes (hMSH2 and hMLH1) in 34 Swedish families with colorectal cancer. Clinical Genetics 1998; 53:131-5.

176. Logan RF, Little J, Hawtin PG et al. Effect of aspirin and non-steroidal anti-inflammatory drugs on colorectal adenomas: case-control study of subjects participating in the Nottingham faecal occult blood screening programme. British Medical Journal 1993; 307:285-9.

177. Love RR, Evans AM, Josten DM. The accuracy of patient reports of a family history of cancer. Journal of Chronic Disease 1985; 38:289-93.

178. Love RR, Morrissey JF. Colonoscopy in asymptomatic individuals with a family history of colorectal cancer. Archives of Internal Medicine 1984; 144:2209-11.

179. Lovett E. Family studies in cancer of the colon and rectum. British Journal of Surgery 1976; 63:13-8.

180. Luce MC, Marra G, Chauhan DP et al. In vitro transcription/translation assay for the screening of hMLH1 and hMSH2 mutations in familial colon cancer. Gastroenterology 1995; 109:1368-74.

181. Luchtefeld MA, Syverson D, Solfelt M et al. Is colonoscopic screening appropriate in asymptomatic patients with family history of colon cancer? Diseases of the Colon & Rectum 1991; 34:763-8.

182. Lyddy P. Modelling 'The Central Dogma of Molecular Biology' using the Unified Modelling Language (UML). 2002. Thesis/Dissertation, Heriot-Watt University.

183. Lynch HT, Krush AJ. Cancer family "G" revisited: 1895-1970. Cancer 1971; 27:1505-11.

184. Lynch HT, Lemon SJ, Karr B et al. Etiology, natural history, management and molecular genetics of hereditary nonpolyposis colorectal cancer (Lynch syndromes): genetic counseling implications. Cancer Epidemiology, Biomarkers & Prevention 1997; 6:987-91.

185. Lynch HT, Shaw MW, Magnuson CW et al. Hereditary factors in cancer. Study of two large midwestern kindreds. Archives of Internal Medicine 1966; 117:206-12.

186. Lynch HT, Watson P, Shaw TG et al. Clinical impact of molecular genetic diagnosis, genetic counseling, and management of hereditary cancer. Part II: Hereditary nonpolyposis colorectal carcinoma as a model. Cancer 1999; 86:2457-63.

187. Macdonald ASWHR&WCTTSAJ22. The genetics of breast and ovarian cancer I: A model of family history. Scandinavian Actuarial Journal 2003; 1-27.

188. Macrae FA, St John DJ, Ambikapathy A et al. Factors affecting compliance in colorectal cancer screening. Results of a study performed in Ballarat. Medical Journal of Australia 1986; 144:621-3.

189. Maliaka YK, Chudina AP, Belev NF et al. CpG dinucleotides in the hMSH2 and hMLH1 genes are hotspots for HNPCC mutations. Human Genetics 1996; 97:251-5.

190. Mandel JS, Bond JH, Church TR et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. New England Journal of Medicine 1993; 328:1365-71.

191. Marchand LL. Combined influence of genetic and dietary factors on colorectal cancer incidence in Japanese Americans. Journal of the National Cancer Institute Monographs 1999; 101-5.

192. Markowitz S. TGF-beta receptors and DNA repair genes, coupled targets in a pathway of human colon carcinogenesis. Biochimica et Biophysica Acta 2000; 1470:M13-M20.

193. Martinez ME, Giovannucci E, Spiegelman D et al. Leisure-time physical activity, body size, and colon cancer in women. Nurses' Health Study Research Group. Journal of the National Cancer Institute 1997; 89:948-55.

194. Meagher AP, Stuart M. Colonoscopy in patients with a family history of colorectal cancer. Diseases of the Colon & Rectum 1992; 35:315-21.

195. Mecklin JP. Frequency of hereditary colorectal carcinoma. Gastroenterology 1987; 93:1021-5.

196. Medvidovic N. Modeling Software Architectures in the Unified Modeling Language. ACM Transaction on Software Engineering and Methodology 2002; 11:2-57.

197. Millar AL, Pal T, Madlensky L et al. Mismatch repair gene defects contribute to the genetic basis of double primary cancers of the colorectum and endometrium. Human Molecular Genetics 1999; 8:823-9.

198. Mitchell RJ, Brewster D, Campbell H, et al. Accuracy of Reporting of Family History of Colorectal Cancer. Gut 2004; 55:291-295.

199. Mitchell RJ, Farrington SM, Dunlop MG et al. Mismatch repair genes hMLH1 and hMSH2 and colorectal cancer: a HuGE review. American Journal of Epidemiology 2002; 156:885-902.

200. Miyaki M, Konishi M, Muraoka M et al. Germ line mutations of hMSH2 and hMLH1 genes in Japanese families with hereditary nonpolyposis colorectal cancer (HNPCC): Usefulness of DNA analysis for screening and diagnosis of HNPCC patients. Journal of Molecular Medicine 1995; 73:515-20.

201. Montera M, Resta N, Simone C et al. Mutational germline analysis of hMSH2 and hMLH1 genes in early onset colorectal cancer patients [letter]. Journal of Medical Genetics (Online) 2000; 37:E7.

202. Moslein G, Tester DJ, Lindor NM et al. Microsatellite instability and mutation analysis of hMSH2 and hMLH1 in patients with sporadic, familial and hereditary colorectal cancer. Human Molecular Genetics 1996; 5:1245-52.

203. Myrhoj T, Bisgaard ML, Bernstein I et al. Hereditary non-polyposis colorectal cancer: clinical features and survival. Results from the Danish HNPCC register. Scandinavian Journal of Gastroenterology 1997; 32:572-6.

204. Nakama H, Fukazawa K. Colorectal cancer risk in first-degree relatives of patients with colorectal adenomatous polyp. Hepato-Gastroenterology 2002; 49:157-9.

205. Negri E, Braga C, La Vecchia C et al. Family history of cancer and risk of colorectal cancer in Italy. British Journal of Cancer 1998; 77:174-9.

206. Newcomb PA, Taylor JO, Trentham-Dietz A. Interactions of familial and hormonal risk factors for large bowel cancer in women. International Journal of Epidemiology 1999; 28:603-8.

207. Newcombe HB. Handbook of Record Linkage. New York: OUP, 1988.

208. Nilsen TI, Vatten LJ. Prospective study of colorectal cancer risk and physical activity, diabetes, blood glucose and BMI: exploring the hyperinsulinaemia hypothesis. British Journal of Cancer 2001; 84:417-22.

209. Nomura S, Sugano K, Kashiwabara H et al. Enhanced detection of deleterious and other germline mutations of hMSH2 and hMLH1 in Japanese hereditary nonpolyposis colorectal cancer kindreds. Biochemical & Biophysical Research Communications 2000; 271:120-9.

210. Novakovic B, Goldstein AM, Tucker MA. Validation of family history of cancer in deceased family members. Journal of the National Cancer Institute. 1996; 88:1492-3.

211. Nystrom-Lahti M, Holmberg M, Fidalgo P et al. Missense and nonsense mutations in codon 659 of MLH1 cause aberrant splicing of messenger RNA in HNPCC kindreds. Genes, Chromosomes & Cancer 1999; 26:372-5.

212. Nystrom-Lahti M, Parsons R, Sistonen P et al. Mismatch repair genes on chromosomes 2p and 3p account for a major share of hereditary nonpolyposis colorectal cancer families evaluable by linkage. American Journal of Human Genetics 1994; 55:659-65.

213. Nystrom-Lahti M, Perrera C, Raschle M et al. Functional analysis of MLH1 mutations linked to hereditary nonpolyposis colon cancer. Genes Chromosomes.Cancer 2002; 33:160-7.

214. Nystrom-Lahti M, Sistonen P, Mecklin JP et al. Close linkage to chromosome 3p and conservation of ancestral founding haplotype in hereditary nonpolyposis colorectal cancer families. Proceedings of the National Academy of Sciences USA 1994; 91:6054-8.

215. Nystrom-Lahti M, Wu Y, Moisio A-L et al. DNA mismatch repair gene mutations in 55 kindreds with verified or putative hereditary non-polyposis colorectal cancer. Human Molecular Genetics 1996; 5:763-9.

216. Papadopoulos N, Nicolaides NC, Wei YF et al. Mutation of a mutL homolog in hereditary colon cancer. Science 1994; 263:1625-9.

217. Parkin DM. Studies of cancer in migrant populations: methods and interpretation. Rev.Epidemiol.Sante Publique 1992; 40:410-24.

218. Peltomaki P, Aaltonen LA, Sistonen P et al. Genetic mapping of a locus predisposing to human colorectal cancer. Science 1993; 260:810-2.

219. Pensotti V, Radice P, Presciuttini S et al. Mean age of tumor onset in HNPCC families correlates with the presence of mutations in DNA hismatch repair genes. Genes, Chromosomes & Cancer 1997; 19:135-42.

220. Petersen GM, Brensinger JD, Johnson KA et al. Genetic testing and counseling for hereditary forms of colorectal cancer. Cancer 1999; 86:2540-50.

221. Ponder BA. Costs, benefits and limitations of genetic testing for cancer risk. British Journal of Cancer 1999; 80 Suppl 1:46-50.

222. Poole CA, Byers T, Calle EE et al. Influence of a family history of cancer within and across multiple sites on patterns of cancer mortality risk for women. American Journal of Epidemiology 1999; 149:454-62.

223. Potter JD. Colorectal cancer: molecules and populations. Journal of the National Cancer Institute 1999; 91:916-32.

224. Quatrani T. Visual Modeling with Rational Rose and UML. Reading, Massachusetts: Addison-Wesley, 1998.

225. Ramesar RS, Madden MV, Felix R et al. Molecular genetics improves the management of hereditary non-polyposis colorectal cancer. South African Medical Journal 2000; 90:709-14.

226. Ransohoff DF, Lang CA. Screening for colorectal cancer with the fecal occult blood test: a background paper. American College of Physicians. Annals of Internal Medicine 1997; 126:811-22.

227. Reeve J, Owens G. Psychological Impact of Predictive Testing of Colonic Cancer. Journal of Health Psychology 2000; 5:99-108.

228. Registrar General for Scotland. 2001 Population Report Scotland. 2002. Edinburgh, General Register Office for Scotland.

229. Reitmair AH, Cai JC, Bjerknes M et al. MSH2 deficiency contributes to accelerated APC-mediated intestinal tumorigenesis. Cancer Research 1996; 56:2922-6.

230. Renkonen-Sinisalo L, Aarnio M, Mecklin JP et al. Surveillance improves survival of colorectal cancer in patients with hereditary nonpolyposis colorectal cancer. Cancer Detection & Prevention 2000; 24:137-42.

231. Rex DK, Cutler CS, Lemmel GT et al. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. Gastroenterology 1997; 112:24-8.

232. Rex DK, Vining D, Kopecky KK. An initial experience with screening for colon polyps using spiral CT with and without CT colography (virtual colonoscopy). Gastrointestinal Endoscopy 1999; 50:309-13.

233. Ricciardiello L, Roda E, Bazzoli F. Chemoprevention in colorectal neoplasias: what is practical and feasible? Digestive Diseases 2002; 20:70-2.

234. Rijnkels JM, Hollanders VM, Woutersen RA et al. Absence of an inhibitory effect of a vegetables-fruit mixture on the initiation and promotion phases of azoxymethane-induced colorectal carcinogenesis in rats fed low- or high-fat diets. Nutrition & Cancer 1998; 30:124-9.

235. Rodriguez-Bigas MA. Prophylactic colectomy for gene carriers in hereditary nonpolyposis colorectal cancer. Has the time come? Cancer 1996; 78:199-201.

236. Rodriguez-Bigas MA, Boland CR, Hamilton SR et al. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. Journal of the National Cancer Institute 1997; 89:1758-62.

237. Rose P et al. Screening for late onset genetic disorders: Colorectal cancer. National Electronic Library for Health website (NHS) . http://www.nelh.nhs.uk/screening/adult_pps/Rose_report.pdf 2002.

238. Roth S, Sistonen P, Salovaara R et al. SMAD genes in juvenile polyposis. Genes Chromosomes Cancer 1999; 26:54-61.

239. Rozen P, Fireman Z, Figer A et al. Family history of colorectal cancer as a marker of potential malignancy within a screening program. Cancer 1987; 60:248-54.

240. Ruschoff J, Wallinger S, Dietmaier W et al. Aspirin suppresses the mutator phenotype associated with hereditary nonpolyposis colorectal cancer by genetic selection. Proceedings of the National Academy of Sciences USA 1998; 95:11301-6.

241. Salovaara R, Loukola A, Kristo P et al. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. Journal of Clinical Oncology 2000; 18:2193-200.

242. Samowitz WS, Curtin K, Lin HH et al. The colon cancer burden of genetically defined hereditary nonpolyposis colon cancer. Gastroenterology 2001; 121:830-8.

243. Sandhu MS, Luben R, Khaw KT. Prevalence and family history of colorectal cancer: implications for screening. Journal of Medical Screening 2001; 8:69-72.

244. Sandler RS, Halabi S, Baron JA et al. A randomized trial of aspirin to prevent colorectal adenomas in patients with previous colorectal cancer. New England Journal of Medicine 2003; 348:883-90.

245. Sankila R, Aaltonen LA, Jarvinen HJ et al. Better survival rates in patients with MLH1-associated hereditary colorectal cancer. Gastroenterology 1996; 110:682-7.

246. Sasaki S, Tokino T, Miyatsu T et al. Mutational analysis of the hMLH1 gene using an automated two- dimensional DNA typing system. Human Mutation 1997; 9:164-71.

247. Schatzkin A, Lanza E, Corle D et al. Lack of effect of a low-fat, high-fiber diet on the recurrence of colorectal adenomas. Polyp Prevention Trial Study Group. New England Journal of Medicine 2000; 342:1149-55.

248. Scott RJ, McPhillips M, Meldrum CJ et al. Hereditary nonpolyposis colorectal cancer in 95 families: differences and similarities between mutation-positive and mutation-negative kindreds. American Journal of Human Genetics 2001; 68:118-27.

249. Scottish Cancer Group CGS2. Cancer Genetics Services in Scotland - guidance to support the implementation of genetics services for breast, ovarian and colorectal cancer predisposition. Scottish Health on the Web http://www.show.scot.nhs.uk/sehd/mels/HDL2001_24Guide.pdf 2002.

250. Selby JV, Friedman GD, Quesenberry CP, Jr. et al. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. New England Journal of Medicine 1992; 326:653-7.

251. Sharp L. Current trends in colorectal cancer: what they tell us and what we still do not know. Clinical Oncology 2001; 13:444-7.

252. Shike M. Body weight and colon cancer. American Journal of Clincial Nutrition 1996; 63:442S-4S.

253. Sijmons RH, Boonstra AE, Reefhuis J et al. Accuracy of family history of cancer: clinical genetic implications. European Journal of Human Genetics 2000; 8:181-6.

254. Slattery ML, Curtin K, Anderson K et al. Associations Between Cigarette Smoking, Lifestyle Factors, and Microsatellite Instability in Colon Tumors. Journal of the National Cancer Institute 2000; 92:1831-6.

255. Slattery ML, Edwards S, Curtin K et al. Physical activity and colorectal cancer. American Journal of Epidemiology 2003; 158:214-24.

256. Slattery ML, Kerber RA. Family history of cancer and colon cancer risk: the Utah Population Database. Journal of the National Cancer Institute 1994; 86:1618-26.

257. Slattery ML, Potter JD, Curtin K et al. Estrogens reduce and withdrawal of estrogens increase risk of microsatellite instability-positive colon cancer. Cancer Research 2001; 61:126-30.

258. Slattery ML, Potter JD, Ma KN et al. Western diet, family history of colorectal cancer, NAT2, GSTM-1 and risk of colon cancer. Cancer Causes & Control 2000; 11:1-8.

259. Slupska M.M., Baikalov C., Luther W.M. et al. Cloning and sequencing a human homolog (hMYH) of the Escherichia coli mutY gene whose function is required for the repair of oxidative DNA damage. Journal of Bacteriology 1996; **178**:3885-92.

260. St John DJ, McDermott FT, Hopper JL et al. Cancer risk in relatives of patients with common colorectal cancer. Annals of Internal Medicine 1993; 118:785-90.

261. Stanley AJ, Gaff CL, Aittomaki AK et al. Value of predictive genetic testing in management of hereditary non- polyposis colorectal cancer (HNPCC). Medical Journal of Australia 2000; 172:313-6.

262. Steinbach G, Lynch PM, Phillips RK et al. The effect of celecoxib, a cyclooxygenase-2 inhibitor, in familial adenomatous polyposis. New England Journal of Medicine 2000; 342:1946-52.

263. Stephenson BM, Murday VA, Finan PJ et al. Feasibility of family based screening for colorectal neoplasia: experience in one general surgical practice. Gut 1993; 34:96-100.

264. Strand M, Prolla TA, Liskay RM et al. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 1993; 365:274-6.

265. Super M, Schwarz MJ, Malone G. Screening for cystic fibrosis carriers. Lancet 1992; 340:490-1.

266. Super M, Schwarz MJ, Malone G et al. Active cascade testing for carriers of cystic fibrosis gene. British Medical Journal 1994; 308:1462-7.

267. Syngal S, Fox EA, Eng C et al. Sensitivity and specificity of clinical criteria for hereditary non-polyposis colorectal cancer associated mutations in MSH2 and MLH1. Journal of Medical Genetics 2000; 37:641-5.

268. Syngal S, Fox EA, Li C et al. Interpretation of genetic test results for hereditary nonpolyposis colorectal cancer: implications for clinical predisposition testing. Journal of the American Medical Association 1999; 282:247-53.

269. Syrigos KN, Charalampopoulos A, Ho JL et al. Colonoscopy in asymptomatic individuals with a family history of colorectal cancer. Annals of Surgical Oncology 2002; 9:439-43.

270. Tannergard P, Lipford JR, Kolodner R et al. Mutation screening in the hMLH1 gene in Swedish hereditary nonpolyposis colon cancer families. Cancer Research 1995; 55:6092-6.

271. Terdiman JP, Conrad PG, Sleisenger MH. Genetic testing in hereditary colorectal cancer: indications and procedures. American Journal of Gastroenterology 1999; 94:2344-56.

272. Thun MJ, Calle EE, Namboodiri MM et al. Risk factors for fatal colon cancer in a large prospective study. Journal of the National Cancer Institute 1992; 84:1491-500.

273. Thune I, Lund E. Physical activity and risk of colorectal cancer in men and women. British Journal of Cancer 1996; 73:1134-40.

274. Todora HM, Skinner CS, Gidday L et al. Perceptions of genetic risk assessment and education among first-degree relatives of colorectal cancer patients and implications for physicians. Family Practice 2001; 18:367-72.

275. Toft NJ, Winton DJ, Kelly J et al. Msh2 status modulates both apoptosis and mutation frequency in the murine small intestine. Proceedings of the National Academy of Sciences USA 1999; 96:3911-5.

276. Tomlinson IPM, Beck NE, Homfray T et al. Germline HNPCC gene variants have little influence on the risk for sporadic colorectal cancer. Journal of Medical Genetics 1997; 34:39-42.

277. Trojan J, Zeuzem S, Randolph A et al. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. Gastroenterology 2002; 122:211-9.

278. UK Government. The Data Protection Act. 1998.

279. UK National Screening Committee. The Criteria for Appraising the Viability, Effectiveness and Appropriateness of a Screening Programme. UK National Screening Committee. http://www.nsc.nhs.uk/pdfs/criteria.pdf 2003.

280. Umans-Eckenhausen MA, Defesche JC, Sijbrands EJ et al. Review of first 5 years of screening for familial hypercholesterolaemia in the Netherlands. Lancet 2001; 357:165-8.

281. van Kranen HJ, van Iersel PW, Rijnkels JM et al. Effects of dietary fat and a vegetable-fruit mixture on the development of intestinal neoplasia in the ApcMin mouse. Carcinogenesis 1998; 19:1597-601.

282. Vasen HF, Hartog Jager FC, Menko FH et al. Screening for hereditary non-polyposis colorectal cancer: a study of 22 kindreds in The Netherlands. American Journal of Medicine 1996; 86:278-81.

283. Vasen HF, Mecklin JP, Khan PM et al. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Diseases of the Colon & Rectum 1991; 34:424-5.

284. Vasen HF, Mecklin JP, Watson P et al. Surveillance in hereditary nonpolyposis colorectal cancer: an international cooperative study of 165 families. The International Collaborative Group on HNPCC. Diseases of the Colon & Rectum 1993; 36:1-4.

285. Vasen HF, Nagengast FM, Khan PM. Interval cancers in hereditary non-polyposis colorectal cancer (Lynch syndrome). Lancet 1995; 345:1183-4.

286. Vasen HF, Stormorken A, Menko FH et al. MSH2 mutation carriers are at higher risk of cancer than MLH1 mutation carriers: a study of hereditary nonpolyposis colorectal cancer families. Journal of Clinical Oncology 2001; 19:4074-80.

287. Vasen HF, Taal BG, Nagengast FM et al. Hereditary nonpolyposis colorectal cancer: results of long-term surveillance in 50 families. European Journal of Cancer 1995; 31A:1145-8.

288. Vasen HF, Wijnen J. Clinical implications of genetic testing of hereditary nonpolyposis colorectal cancer. Cytogenetics & Cell Genetics 1999; 86:136-9.

289. Vasen HF, Wijnen JT, Menko FH et al. Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis. Gastroenterology 1996; 110:1020-7.

290. Viel A, Genuardi M, Capozzi E et al. Characterization of MSH2 and MLH1 mutations in Italian families with hereditary nonpolyposis colorectal cancer. Genes, Chromosomes & Cancer 1997; 18:8-18.

291. Voskuil DW, Vasen HF, Kampman E et al. Colorectal cancer risk in HNPCC families: development during lifetime and in successive generations. National Collaborative Group on HNPCC. International Journal of Cancer 1997; 72:205-9.

292. Wahlberg SS, Nystrom-Lahti M, Kane MF et al. Low frequency of hMSH2 mutations in Swedish HNPCC families [letter]. International Journal of Cancer 1997; 74:134-7.

293. Wallace E et al. A cross sectional survey to estimate the prevalence of family history of colorectal, breast and ovarian cancer in a general practice population. Submitted (British Journal of Cancer) 2003.

294. Wang Q, Desseigne F, Lasset C et al. Germline hMSH2 and hMLH1 gene mutations in incomplete HNPCC families. International Journal of Cancer 1997; 73:831-6.

295. Wang Q, Lasset C, Desseigne F et al. Prevalence of germline mutations of hMLH1, hMSH2, hPMS1, hPMS2, and hMSH6 genes in 75 French kindreds with nonpolyposis colorectal cancer. Human Genetics 1999; 105:79-85.

296. Wang Y. Formal Description of the UML Architecture and Extendibility. L'object 2000; 6:1-20.

297. Wang Y, Friedl W, Lamberti C et al. A novel missense mutation in the DNA mismatch repair gene hMLH1 present among East Asians but not among Europeans. Human Heredity 1998; 48:87-91.

298. Warthin AS. Heredity with reference to carcinoma. Archives of Internal Medicine 1913; 12:546-55.

299. Weber TK, Chin H-M, Rodriguez-Bigas M et al. Novel hMLH1 and hMSH2 germline mutations in African Americans with colorectal cancer. Journal of the American Medical Association 1999; 281:2316-20.

300. Weinstein MC, O'Brien B, Hornberger J et al. Principles of good practice for decision analytic modeling in health- care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. Value & Health 2003; 6:9-17.

301. Welch HG, Burke W. Uncertainties in genetic testing for chronic disease. Journal of the American Medical Association 1998; 280:1525-7.

302. Weller DP, Owen N, Hiller JE et al. Colorectal cancer and its prevention: prevalence of beliefs, attitudes, intentions and behaviour. Australian Journal of Public Health 1995; 19:19-23.

303. Westerman AM, Entius MM, Boor PP et al. Novel mutations in the LKB1/STK11 gene in Dutch Peutz-Jeghers families. Human Mutation 1999; 13:476-81.

304. Wijnen J, van der KH, Vasen H et al. MSH2 genomic deletions are a frequent cause of HNPCC [letter]. Nature Genetics 1998; 20:326-8.

305. Wijnen J, Vasen H, Khan PM et al. Seven new mutations in hMSH2, an HNPCC gene, identified by denaturing gradient-gel electrophoresis. American Journal of Human Genetics 1995; 56:1060-6.

306. Wijnen JT, Vasen HF, Khan PM et al. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. New England Journal of Medicine 1998; 339:511-8.

307. Wildhagen MF, van Os TA, Polder JJ et al. Efficacy of cascade testing for fragile X syndrome. Journal of Medical Screening 1999; 6:70-6.

308. Wilson EB. Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association 1927; 22:209-12.

309. Wilson JMG, Junger G. Principles and Practice of Screening for Disease. Geneva: WHO, 1968.

310. Winawer S, Fletcher R, Rex D et al. Colorectal cancer screening and surveillance: Clinical guidelines and rationale-Update based on new evidence. Gastroenterology 2003; 124:544-60.

311. Winawer SJ. Natural history of colorectal cancer. American Journal of Medicine 1999; 106:3S-6S.

312. Winawer SJ, Fletcher RH, Miller L et al. Colorectal cancer screening: clinical guidelines and rationale. Gastroenterology 1997; 112:594-642.

313. Woodford-Richens K, Bevan S, Churchman M et al. Analysis of genetic and phenotypic heterogeneity in juvenile polyposis. Gut 2000; 46:656-60.

314. World Cancer Research Fund Panel (Potter JDC. Diet, Nutrition, and the Prevention of Cancer. WCRF. www.wcrf.org.uk 2001.

315. Wu AH, Shibata D, Yu MC et al. Dietary heterocyclic amines and MSI in colon adenocarcinomas. Carcinogenesis 2001; 22:1681-4.

316. Wu CS, Tung SY, Chen PC et al. The role of colonoscopy in screening persons with family history of colorectal cancer.[comment]. Journal of Gastroenterology & Hepatology 1995; 10:319-23.

317. Yamamoto H, Itoh F, Fukushima H et al. Overexpression of cyclooxygenase-2 protein is less frequent in gastric cancers with microsatellite instability. International Journal of Cancer 1999; 84:400-3.

318. Yamamoto M, Nakama H. Cost-effectiveness analysis of immunochemical occult blood screening for colorectal cancer among three fecal sampling methods. Hepato-Gastroenterology 2000; 47(32):396-9.

319. Yan H, Papadopoulos N, Marra G et al. Conversion of diploidy to haploidy. Nature 2000; 403:723-4.

320. Ye Z, Parry JM. Meta-analysis of 20 case-control studies on the N-acetyltransferase 2 acetylation status and colorectal cancer risk. Medical Science Monitor 2002; 8:CR558-CR565.

321. Ye Z, Parry JM. A meta-analysis of 20 case-control studies of the glutathione S-transferase M1 (GSTM1) status and colorectal cancer risk. Medical Science Monitor 2003; 9:SR83-SR91.

322. Yoon KA, Ku JL, Choi HS et al. Germline mutations of the STK11 gene in Korean Peutz-Jeghers syndrome patients. British Journal of Cancer 2000; 82:1403-6.

323. Yuan Y, Han HJ, Zheng S et al. Germline mutations of hMLH1 and hMSH2 genes in patients with suspected hereditary nonpolyposis colorectal cancer and sporadic early-onset colorectal cancer. Dis.Colon Rectum 1998; 41:434-40.

Appendix A1 Cancer Genetics Guidelines for Scotland: Risk Stratification for Colorectal Cancer

Low	Medium	High
Risk Stratification • anyone not fulfilling medium or high risk criteria	***Risk Stratification*** • One 1st degree relative affected by colorectal cancer when aged <45yrs • Two (one affect at less than 55yrs), one a 1st degree relative of subject • Three affected with colorectal or endometrial cancer who are 1st degree relatives of each other and one a first degree relative of subject • Two affected 1st degree relatives (one affected at lees than 55yrs)	***Risk Stratification*** • Gene carriers of HNPCC mutation • Untested 1st degree relatives of gene carriers • People with a family history compatible with HNPCC
Counselling Individuals deemed at low risk will be informed either by: • Telephone consultation with a genetic nurse associate, followed by a letter with a copy to GP, or • Face to face consultation with the genetic nurse associate and then by letter to the patient and the GP	***Counselling*** Individuals deemed to be at medium risk will be counselled by the genetic counsellor	***Counselling*** Individuals deemed to be at medium risk will be counselled by the clinical genetic physician
Management • Reassurance • Healthy lifestyle advice • Advise to report any changes • Return to GP care	***Management*** Screening: • Colonoscopy at 30-35 yrs, if findings are normal this need not be repeated until 55 yrs of age • Incomplete colonoscopy should be followed by a barium enema, preferably at the same hospital attendance	***Management*** Screening: • Colonoscopy every 2 yrs from age 30, of 5 yrs younger than the youngest affected, up until the age of 70 • Dicussion of prophylactic surgery, if recurrent polyps are identified • Consideration needs to be given to screening for other cancers that are part of the Hereditary Non-Polyposis Colorectal Cancer (HNPCC) spectrum Gene Testing: • Should ideally be available to all high-risk families

Appendix A2 Mismatch Repair Genes hMLH1 and hMSH2 and Colorectal Cancer: A HuGE Review

This paper has been reproduced in its published form with the permission of the publishers (Oxford University Press) and all co-authors.

American Journal of
EPIDEMIOLOGY

Volume 156
Number 10
November 15, 2002

Sponsored by the Society for Epidemiologic Research
Published by Oxford University Press

HUMAN GENOME EPIDEMIOLOGY (HuGE) REVIEWS

Mismatch Repair Genes *hMLH1* and *hMSH2* and Colorectal Cancer: A HuGE Review

R. J. Mitchell[1], S. M. Farrington[2], M. G. Dunlop[2], and H. Campbell[1]

[1] Department of Public Health Sciences, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, United Kingdom.
[2] Colon Cancer Genetics Group, MRC Human Genetics Unit, University of Edinburgh and Western General Hospital, Edinburgh EH4 2XU, United Kingdom.

Received for publication December 10, 2001; accepted for publication June 20, 2002.

Evidence to support a role for the mismatch repair genes human mutL homolog 1 (*hMLH1*) and human mutS homolog 2 (*hMSH2*) in the etiology of colorectal cancer has come from linkage analysis, segregation studies, and molecular biologic analysis. More recently, carriers of potentially pathogenic mutations in the *hMLH1/hMSH2* genes have consistently been shown to be at a greatly increased risk of developing colorectal cancer compared with the general population. When considered together, the available evidence shows a strong, consistent, and biologically plausible association between mismatch repair gene mutations and colorectal cancer. The penetrance of mutations in *hMLH1/hMSH2* is incomplete and is significantly higher in males (approximately 80%) than in females (approximately 40%). To date, evidence for gene-gene or gene-environment interactions is limited, although preliminary studies have revealed a number of avenues that merit exploration. Population screening for mutation carriers is not currently a feasible option, and mutation analysis remains restricted to either relatives of mutation carriers or colorectal cancer cases selected on the basis of phenotype.

colorectal neoplasms; epidemiology; genetic screening; germ-line mutation; *hMLH1*; *hMSH2*; penetrance; survival

Abbreviations: *hMLH1*, human mutL homolog 1; *hMSH2*, human mutS homolog 2; HNPCC, hereditary nonpolyposis colorectal cancer; MSI, microsatellite instability.

INTRODUCTION

The mismatch repair genes human mutL homolog 1 (*hMLH1*) and human mutS homolog 2 (*hMSH2*) are integral components of the DNA mismatch repair pathway. So far, over 200 allelic variants have been identified for each gene, and the majority of these have been reported to be pathogenic in terms of colorectal cancer. The primary objectives of this review are to describe what is known about *hMLH1* and *hMSH2* and their variants in different populations and to examine the evidence implicating these genes as risk factors in the development of colorectal cancer. Relevant Internet sites are listed in appendix 1.

Correspondence to Dr. Harry Campbell, Department of Public Health Sciences, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, United Kingdom (e-mail: Harry.Campbell@ed.ac.uk).

TABLE 1. Commonly observed pathogenic mutations in persons with colorectal cancer

Exon	Nucleotide change	DNA change	No. of kindreds	Founder effect?	No. from founder population
		*hMLH1**			
Exon 16		3.5-kilobase deletion	63	Finland	62
Exon 16	Deletion of AAG at nucleotide 1846	In-frame deletion in lysine codon 616	21		
IVS* 5	g-a at nucleotide 4541	Out-of-frame deletion in exon 6 codon 152182	18	Finland	15
Exon 16	AA-GC at nucleotide 1852	Lys618Ala	15		
Exon 4	C-T at nucleotide 350	Thr117Met	12		
Exon 19	G-A at nucleotide 2146	Val716Met	12		
Exon 13	Insertion of C at nucleotide 1490	Frameshift from codon 497	10		
Exon 4	T-G at nucleotide 320	Ile107Arg	7	Finland	7
Exon 13	C-T at nucleotide 1459	Arg487STOP	7		
Exon 17	C-T at nucleotide 1975	Arg659STOP	7		
Exon 19	G-A at nucleotide 2141	Trp714STOP	6		
Exon 8	C-T at nucleotide 676	Arg226STOP	6		
Exon 2	G-A at nucleotide 199	Gly67Arg	5		
Exon 2	C-T at nucleotide 184	Gln62STOP	5		
IVS 14	4-base-pair insertion/3-base-pair deletion at nucleotide 1667+2	Silencing of allele	5	Denmark	4
		*hMSH2**			
IVS 5	a-t at nucleotide 942+3	In-frame deletion in exon 5	46	Newfoundland	10
Exon 6	G-A at nucleotide 965	Gly322Asp	32		
Exon 12	Deletion of AAT at nucleotide 1786	In-frame deletion in asparagine codon 596	11		
Exon 7	C-T at nucleotide 1216	Arg406STOP	6		

* *hMLH1*, human mutL homolog 1; IVS, intervening sequence; *hMSH2*, human mutS homolog 2.

GENERAL METHODOLOGY

Search strategy

The MEDLINE (National Library of Medicine), EMBASE (Excerpta Medica), and CANCERLIT (National Cancer Institute) databases were searched for papers published before December 31, 2001, using the keywords *hMSH2* and *hMLH1*. Relevant papers were identified, critically appraised, and entered into a Reference Manager (ISI ResearchSoft, Berkeley, California) database. In addition, PubMed was searched via Reference Manager, by author name, for papers from research groups that had published several times on this subject. Finally, the database thus created was cross-referenced with papers cited in the International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer database of mutations (1).

For the "Gene variants" section, we considered a total of 109 papers, which were identified by the above strategy and fulfilled the following selection criteria: 1) complete mutation analysis had been performed on more than five patients with colorectal cancer and 2) there was sufficient detail on the molecular nature of the genetic alteration. Details on all gene variants described in these published papers are given in the first supplementary table, which is posted on the website of the Human Genome Epidemiology Network (http://www.cdc.gov/genomics/hugenet/default.htm), as well as on the *Journal*'s website (http://aje.oupjournals.org/).

For the "Associations" section, the above strategy led to the identification of eight studies that had conducted an analysis of the risk of developing colorectal cancer among carriers of mismatch repair gene mutations and 77 papers that included results of complete mutation analysis performed on more than five colorectal cancer patients selected on the basis of family history, microsatellite instability (MSI), or age of onset. These studies are summarized in table 2 and the second supplementary table, respectively. Many papers included information relevant to both gene variants and associations.

Classification of gene variants

For the purposes of this review, we classified gene variants into one of four categories. These categories are loosely based on the definitions given below, modified according to clinical observations.

1. Pathogenic mutation—generally frameshifts, nonsense mutations, and splice variants

2. Probable pathogenic mutation—generally nonconservative amino acid changes

3. Probable polymorphism—generally conservative changes, often observed in controls
4. Definite polymorphism—synonymous variants

GENE

hMSH2

The *hMSH2* gene is located at chromosome 2p21, an area initially identified as an important candidate region for genes involved in hereditary nonpolyposis colorectal cancer (HNPCC) syndrome by genetic linkage analysis within large affected families (2, 3).

The *hMSH2* protein product is a component of the DNA mismatch repair pathway, the role of which is well established in bacteria and yeast. *hMSH2* can form a heterodimer with one of two other mismatch repair proteins, *hMSH6* or *hMSH3*. This protein complex recognizes and binds any errors that may have occurred during DNA replication, and a larger protein complex is then recruited to excise the incorrect daughter sequence and replace it with the correct sequence, using the parental strand as a template. In *Escherichia coli*, mutS has been implicated in both short- and long-patch repair systems (4).

hMLH1

The *hMLH1* gene is located at chromosome 3p21–23, an area also identified by genetic linkage analysis as an important candidate region within large HNPCC families that are not connected with the chromosome region 2p21–22 (5, 6).

The *hMLH1* protein product is another component of the DNA mismatch repair pathway, and it has been shown to form a heterodimer with *hMLH3*, *hPMS2*, or *hPMS1*. The *hMLH1* protein has no known enzymatic activity and probably acts as a "molecular matchmaker," in that it recruits other DNA repair proteins to the mismatch repair complex. Again, the bacterial homolog of *hMLH1* has been implicated in both short- and long-patch repair (4).

GENE VARIANTS

One conclusion generated by early attempts to identify precise genetic alterations in *hMLH1* and *hMSH2* was that variants in these genes are extremely heterogeneous. All 16 exons of the *hMSH2* gene and 19 exons of the *hMLH1* gene have been found to contain pathogenic mutations.

At present, there are no standard criteria for classifying variants as pathogenic mutations or polymorphisms, and consequently there is considerable variation in interpretation by different researchers. In general, categorization of alterations is based on the predicted effect on protein, with segregation of the mutation with colorectal cancer in the kindred in question and/or analysis of control subjects for that specific mutation also being considered when possible. However, the functional consequences of many mutations are difficult to predict accurately. It has been suggested that even alterations that do not affect the amino acid sequence could lead to aberrant splicing, and that the position of the mutation may be more significant than the type (7). In vitro functional assays have been developed and applied to the task of determining the pathogenicity of missense mutations (8–10) and may eventually facilitate accurate classification of such changes.

The first supplementary table lists all of the gene variants identified as part of this review, illustrating the extreme range of mutations identified and the fact that the observed spectrum of mutation is not entirely uniform. Figures 1–4 summarize some features of this table. Figures 1 and 2 illustrate the distributions of unique gene variants that have been fully characterized at the molecular level in *hMLH1* and *hMSH2*, respectively, according to their position on the gene. Figures 3 and 4 are designed to show the actual numbers of families in which pathogenic mutations have been identified.

In total, 259 different pathogenic mutations, as defined above, have been identified in *hMLH1*, along with 45 polymorphisms. In *hMSH2*, 191 different pathogenic mutations and 55 polymorphisms have been characterized so far. This high degree of heterogeneity is similar to that found in the breast cancer genes *BRCA1* and *BRCA2*, in each of which over 400 gene variants have been reported. When considering the range and type of gene variants listed in the first supplementary table, there are several important sources of bias that merit consideration. Firstly, a significant publication bias is likely to exist in favor of apparently pathogenic alterations. Highly penetrant mutations are also likely to be overrepresented, since many studies involved conducting mutation analysis in patients selected on the basis of a strong family history of colorectal cancer. Secondly, genomic deletions in mismatch repair genes appear to occur relatively commonly, particularly in *hMSH2*, and such variants are not detected by many of the techniques commonly used for mutation analysis (11).

It is evident from the above figures that certain specific mutations have been identified in more than one kindred. Indeed, some mutations are found with a relatively high frequency. The most commonly observed mutations are summarized in table 1, which displays all mutations identified in more than four ostensibly independent kindreds.

The observed spectrum of gene variants may be largely due to genuine differences in the mutability of specific nucleotides or sequences within the gene, but in some cases variants identified in apparently unrelated kindreds can be traced to a common ancestor. Such "founder effects" have been identified in the Finnish population, where two specific founder mutations in *hMLH1* account for the vast majority of families in which mismatch repair gene mutations have been identified (12, 13). Another *hMLH1* founder effect is evident in the Danish population (14). The extent to which founder effects are responsible for other frequently detected alterations is not entirely clear from the data currently available, and it is likely that some of the kindreds included in the first supplementary table share a common ancestor. Interestingly, the intervening sequence 5 variant A-T at nucleotide 942+3 has been shown to occur as a founder mutation in Newfoundland (15), but another study found no evidence for a common haplotype in 10 carriers of this variant, of various origins, and concluded that the mutation also arises frequently de novo (16). This

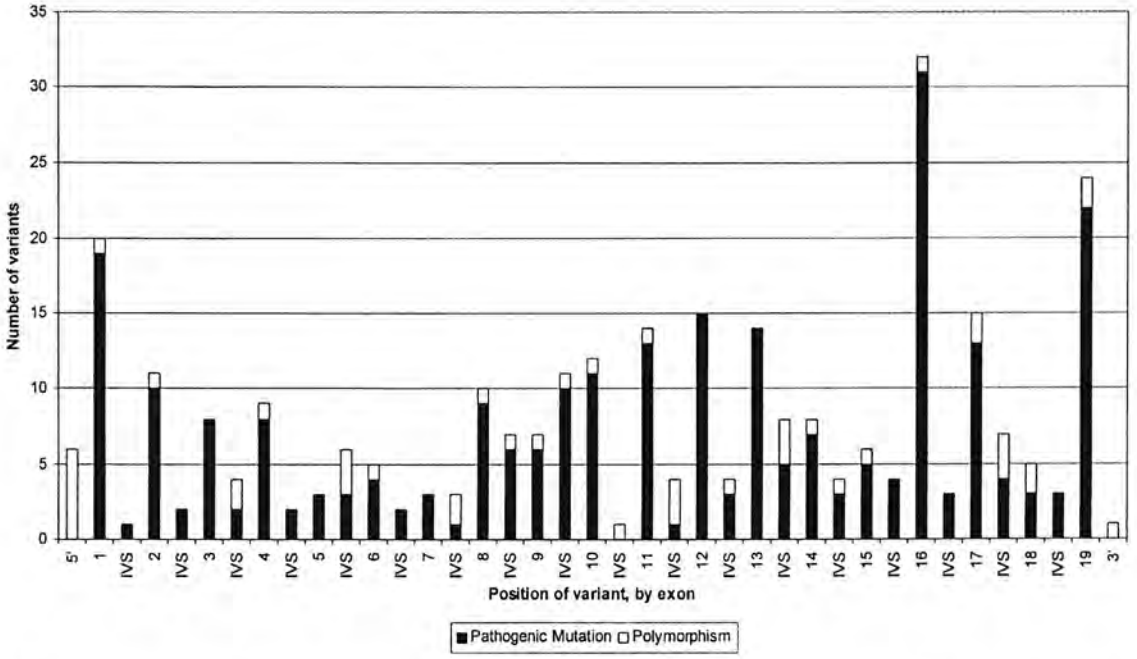


FIGURE 1. Distribution of unique gene variants in the mismatch repair gene *hMLH1*. The figure illustrates the distribution of all unique gene variants that have been identified and fully characterized in mutation analysis studies of colorectal cancer patients. Variants designated as categories 1, 1/2, 2, and 2/3 in the first supplementary table are considered to be pathogenic for the purpose of this summary figure, and all other variants are described as polymorphisms. Exon deletions in which the underlying molecular variant was not known were excluded. IVS, intervening sequence.

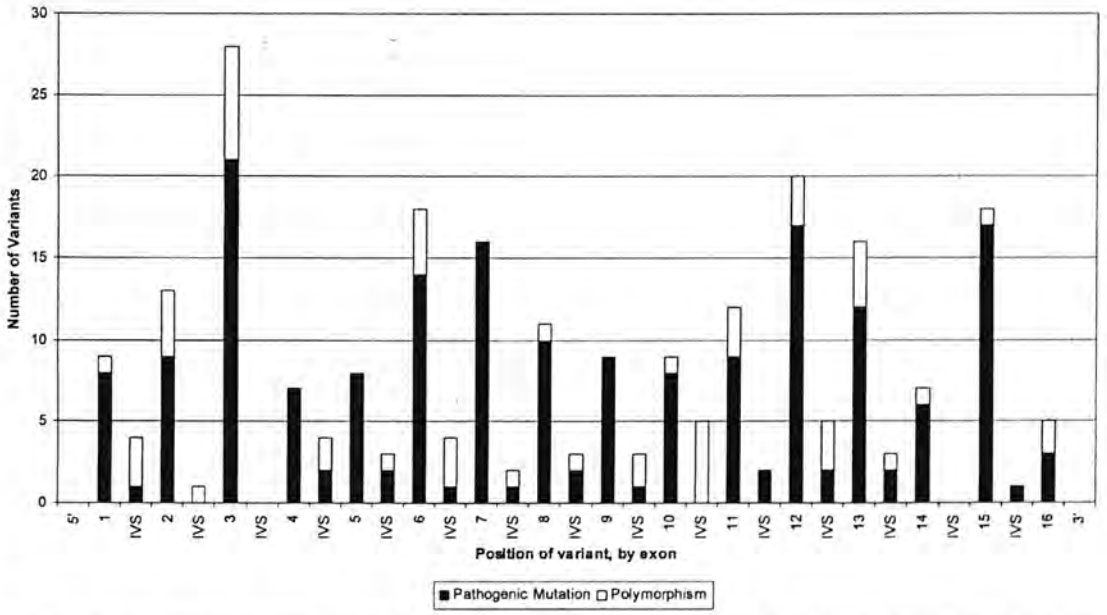


FIGURE 2. Distribution of unique gene variants in the mismatch repair gene *hMSH2*. The figure illustrates the distribution of all unique gene variants that have been identified and fully characterized in mutation analysis studies of colorectal cancer patients. Variants designated as categories 1, 1/2, 2, and 2/3 in the first supplementary table are considered to be pathogenic for the purpose of this summary figure, and all other variants are described as polymorphisms. Exon deletions in which the underlying molecular variant was not known were excluded. IVS, intervening sequence.

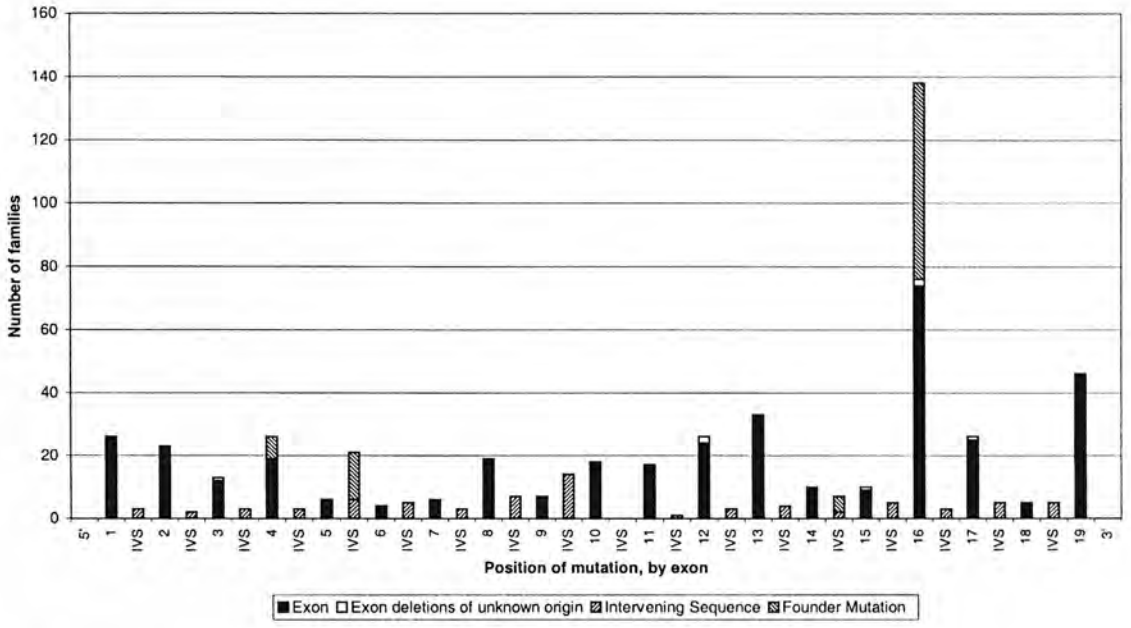


FIGURE 3. Distribution of gene variants in the mismatch repair gene *hMLH1* by the number of families affected. The figure illustrates the distribution of pathogenic mutations according to the actual number of families in which a pathogenic mutation has been identified. These figures include all pathogenic mutations as defined in figure 1, plus exon deletions of unspecified origin. Deletions of more than one exon were excluded. Families are deemed to have a "founder mutation" if they have a mutation which has been shown to have a founder effect in the same population.

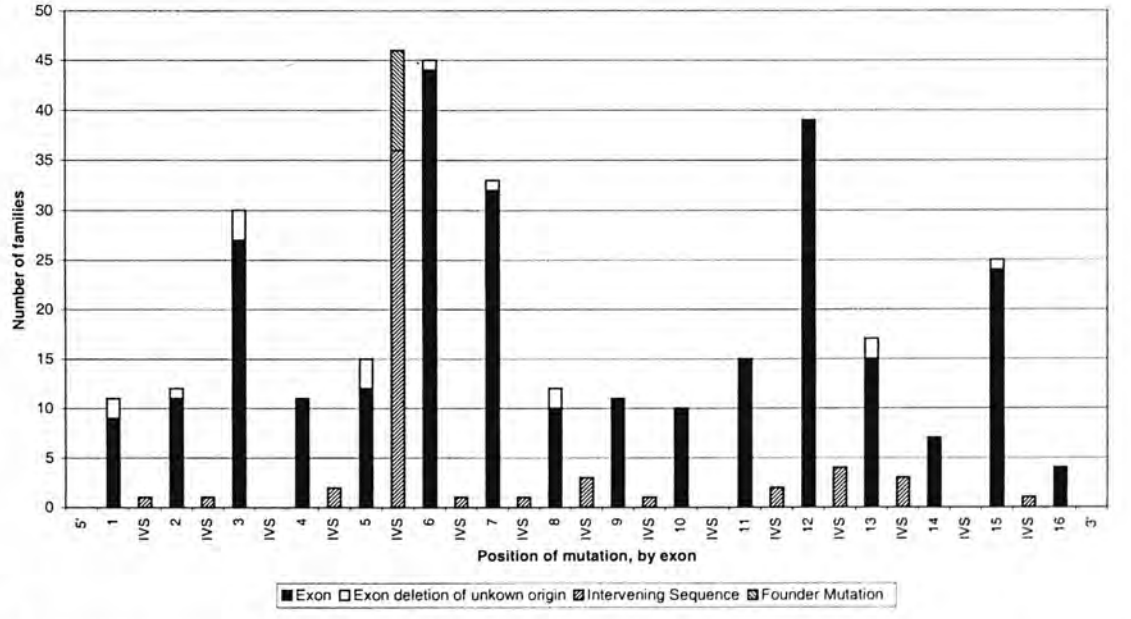


FIGURE 4. Distribution of gene variants in the mismatch repair gene *hMSH2* by the number of families affected. The figure illustrates the distribution of pathogenic mutations according to the actual number of families in which a pathogenic mutation has been identified. These figures include all pathogenic mutations as defined in figure 2, plus exon deletions of unspecified origin. Deletions of more than one exon were excluded. Families are deemed to have a "founder mutation" if they have a mutation which has been shown to have a founder effect in the same population.

example underlines the notion that observations of mutation frequency are the result of both the probability of a mutation at a given nucleotide and the demographic history of the population in question.

Overall, little ethnic or population variation is apparent from the available gene variant data. However, the current biases towards highly penetrant mutations are such that the effect of the identified mutation is likely to transcend any population differences. Clearly, there is a need for accurate and extensive population-based data before any population differences in the spectrum and frequency of mismatch repair gene variants become apparent.

There is no clear evidence to suggest that any specific mismatch repair gene mutation produces a specific phenotype of colorectal cancer, although it has been suggested that some differences exist between the spectrum of extracolonic cancers associated with *hMSH2* mutations in comparison with *hMLH1* mutations (17, 18).

DISEASE

Colorectal cancer is a major public health problem worldwide, with a current annual incidence approaching 950,000 cases (19). Colorectal cancer is more common in males than in females, and in both sexes the incidence rate increases with advancing age. Incidence rates vary globally and are about four times higher in developed countries than in developing countries (20). While incidence rates do vary according to ethnicity (21), there is compelling evidence that the observed variation between countries is primarily due to the role of environmental factors. This hypothesis is supported by the rising incidence of colorectal cancer in populations undergoing rapid economic development, with associated "westernization" of diet and lifestyle. Further evidence for a strong environmental influence comes from migrant data; despite the relatively low incidence of colorectal cancer in Japan, incidence rates in Hawaiian Japanese are among the highest in the world (22).

Considerable effort and resources have been expended with the aim of elucidating the precise dietary and other variables responsible for the observed environmental influences on colorectal cancer incidence. A report commissioned by the World Cancer Research Fund and the American Institute for Cancer Research concluded that evidence was sufficient to suggest that colorectal cancer risk could be substantially reduced by adhering to a diet high in vegetables and low in meat, together with regular physical activity and avoidance of alcohol (23). Other reviews have reached similar conclusions (24), but some studies have failed to provide evidence to uphold the hypothesis that dietary modification can prevent colorectal cancer. Clinical intervention studies (25, 26) and observational cohort studies (27), as well as studies utilizing animal models (28, 29), have shown no evidence of polyp prevention related to diet. Nonetheless, polyp prevention may not be the best endpoint, so results of further clinical studies with cancer prevention as the endpoint are awaited.

Both epidemiologic evidence and experiments utilizing murine models have suggested that nonsteroidal antiinflammatory drugs have antitumor properties that may prevent colorectal cancer. Sulindac has been shown to inhibit tumor growth in experimental systems and to reduce adenoma counts in humans with familial adenomatous polyposis (30), as has a recent study of the specific cyclooxygenase-2 inhibitor celecoxib (31).

A number of case-control and cohort studies have reported an association between hormone replacement therapy and colorectal cancer, with the majority of these providing evidence in favor of a protective effect (24). Accumulating evidence also implicates obesity as a risk factor for colorectal cancer (32), and a positive association may exist between colorectal cancer and diabetes (33, 34). The weight of evidence also suggests that smoking may be a significant risk factor (35).

Colorectal cancer is a multifactorial condition, and while environmental factors are clearly important in the etiology of the disease, there is a significant input from genetic factors. A recent study of twins provided evidence suggesting that about 35 percent of all colorectal cancer cases have a genetic component (36), and first-degree relatives of colorectal cancer patients are well-recognized to have a 2- to 4-fold increased risk of developing the disease themselves. The genetic factors involved are poorly understood and may include recessive genes, pathogenic mutations of low penetrance, and complex gene-gene and gene-environment interactions.

In addition to the less obvious genetic factors, two autosomally inherited cancer syndromes account for a significant minority of colorectal cancer cases. Familial adenomatous polyposis is a rare syndrome caused by mutations in the adenomatous polyposis coli gene and is characterized by the presence of multiple adenomas. In the HNPCC syndrome, affected kindreds have an unusually high occurrence of colorectal and certain extracolonic cancers, with a relatively early age of onset. HNPCC has traditionally been diagnosed on the basis of family history, and the various criteria used for defining HNPCC are summarized in appendix 2. For research purposes, the Amsterdam criteria are the most widely used, and by this definition of HNPCC, the syndrome may account for 2–5 percent of all colorectal cancer cases.

It has been established that a large proportion of families diagnosed with HNPCC harbor potentially pathogenic mutations in mismatch repair genes. Of the mutations identified so far, over 90 percent occur in *hMLH1* and *hMSH2*. HNPCC families in which mutations in *hMLH1* and *hMSH2* are not identified may harbor pathogenic mutations in other mismatch repair genes, such as *hMSH6* and *hPMS2*, or in genes as yet unidentified. Pathogenic mutations in *hMLH1* or *hMSH2* have also been identified in kindreds that do not meet the traditional criteria for diagnosis of HNPCC. This observation may be due to the inherent misclassification bias involved in diagnosing a condition on the basis of family history alone, particularly in small families.

ASSOCIATIONS

Evidence implying and supporting a causal role for *hMLH1/hMSH2* in colorectal cancer comes from both epidemiologic studies and laboratory-based molecular studies, as summarized in figure 5. Initially, linkage studies revealed

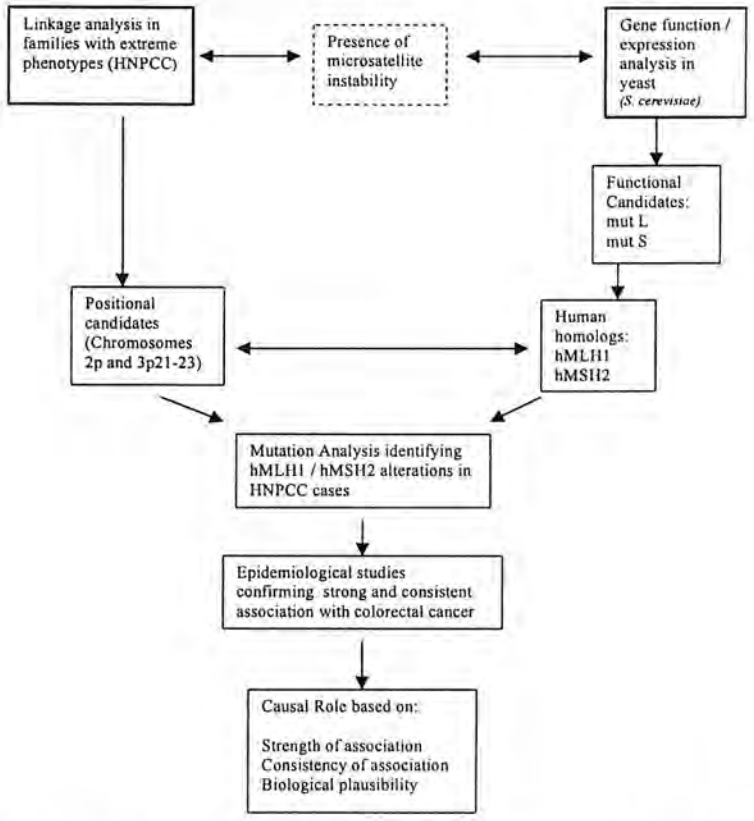


FIGURE 5. Pathways of epidemiologic and biologic research identifying and confirming the causal role of the mismatch repair genes *hMLH1* and *hMSH2* in colorectal cancer. HNPCC, hereditary nonpolyposis colorectal cancer; *hMLH1*, human mutL homolog 1; *hMSH2*, human mutS homolog 2.

that disease expression in a proportion of HNPCC kindreds was linked to either chromosome 2p21 (2, 3, 37) or chromosome 3p21–23 (5, 6, 37, 38).

The connection between the HNPCC syndrome and mismatch repair arose from the observation that the majority of tumors from HNPCC families exhibited a replication error phenotype, a feature resulting from instability of microsatellite repeats during replication that is found only in a minority of "sporadic" colorectal cancer cases (39, 40). Previous molecular studies in the yeast *Saccharomyces cerevisiae* had led to the identification of a group of genes, known as mismatch repair genes, that were involved in maintaining the fidelity of DNA replication. Defects in yeast mismatch repair genes led to MSI, prompting formulation of the hypothesis that human homologs of these genes were involved in the HNPCC syndrome (41). Subsequently, several such homologs were identified, and two of them, *hMLH1* and *hMSH2*, were shown to reside on chromosomes 3p21–23 and 2p21, respectively (2, 42–44). Further supportive evidence came from the observation that pathogenic mutations in *hMLH1* or *hMSH2* could be identified and shown to segregate with disease in a high proportion of kindreds that had shown linkage to the corresponding chromosome (2, 43, 44).

The heterogeneity of mutations in mismatch repair genes means that screening for mutations in these genes is a lengthy and complicated process. Consequently, for purely economic, practical, and ethical reasons, mutation analysis has been carried out almost exclusively among colorectal cancer patients, particularly those identified as being at high risk of harboring mutations. Only two studies identified in this review conducted mutation analysis among control subjects. Farrington et al. (45) found that none of 26 Scottish blood donors harbored previously identified mutations, although four variants of unknown significance were found. This was compared with the identification of potentially pathogenic mutations in 14 of 50 colorectal cancer patients diagnosed at less than 30 years of age. Similarly, no pathogenic mutations were reported in an analysis of 73 population controls from Utah (46).

Thus, the practical restrictions on mutation analysis, coupled with the low population prevalence of mismatch repair gene mutations and the fact that such mutations are found only in a minority of colorectal cancer patients, has

TABLE 2. Findings of risk analysis studies of colorectal cancer

Published reference	Area of study	Ascertainment of index cases	No. of index cases	Ascertainment of mutation carriers	No. of mutation carriers	Penetrance* in mutation carriers	Source of data for comparison	Risk in comparison group	Standardized incidence ratio/ relative risk
Aarnio et al. (47)	Finland	Members of HNPCC† kindreds previously shown to have an *hMLH1*† or *hMSH2*† gene mutation.	50	Test-positive or obligate carriers.	360	Males = 100%; females = 54% (to age 70 years)	Finnish Cancer Registry data, 1991–1995	N/A†	Females + males = 68 (95% CI†: 56, 81)
Dunlop et al. (56)	Scotland	Colorectal cancer cases aged ≤30 years identified through the Scottish National Cancer Registry between 1970 and 1993, excluding those with a family history fulfilling the Amsterdam criteria.	6	Relatives were traced, tested for mutation status where possible, and classified accordingly.	67	Males = 74%; females = 30% (to age 70 years)	United Kingdom cumulative incidence data published by EUCAN† (91)	Males = 2.53%; females = 1.67% (to age 70 years)	Males = 29‡ Females = 18‡
Aarnio et al. (92)	Finland	Families that fulfilled the Amsterdam criteria. In 24 of these, mutation analysis had demonstrated the segregation of *hMLH1* or *hMSH2*.	40	Cases of any cancer in relatives were identified and were included if adequate documentation was available "with the presumption that all tumor patients were HNPCC gene carriers."	293	Females + males = 78% (lifetime)	Finnish cumulative incidence data published by EUCAN (91)	Females + males = 2.6% (to age 75 years)	Females + males = 30‡
Vasen et al. (55)	Netherlands	Families that fulfilled the Amsterdam criteria, identified through the Netherlands HNPCC registry and found to have a mutation in *hMLH1* or *hMSH2*.	19	Relatives were traced and tested for mutation carrier status where possible.	210	Males = 92%; females = 83% (to age 75 years)	Netherlands cumulative incidence data published by EUCAN (91)	Males = 4.41%; females = 3.28% (to age 75 years)	Males = 21‡ Females = 25‡
Vasen et al. (54)	Netherlands and Norway	Kindreds registered with the Netherlands HNPCC registry (n = 193) or suspected HNPCC families from the Clinical Genetic Centre, Radium Hospital, Norway (n = 58).	34 *hMLH1* carriers; 40 *hMSH2* carriers	Mutation carrier status was assigned to one of three groups: 1) tested carriers; 2) relatives with colorectal or endometrial cancer (excluding those tested negative); and 3) obligate carriers.	362 *hMLH1* carriers; 301 *hMSH2* carriers	*hMLH1*: Males = 65%; females = 55%; females + males = 60% *hMSH2*: Males = 73%; females = 54%; females + males = 65% (to age 70 years)	Netherlands cumulative incidence data published by EUCAN (91)	Males = 2.81%; females = 2.17% (to age 70 years)	*hMLH1*: Males = 23‡ Females = 25‡ *hMSH2*: Males = 26‡ Females = 25‡
Froggatt et al. (93)	England	Families that fulfilled the Amsterdam criteria, with mutations in *hMLH1* or *hMSH2*.	8	Subjects with mutations were included in the analysis. No further details were given.	50 (*hMLH1*: n = 23; *hMSH2*: n = 27)	*hMLH1*: Females + males = 67% *hMSH2*: Females + males = 62%	United Kingdom cumulative incidence data published by EUCAN (91)	Females + males = 3.16%	*hMLH1*: 21‡ *hMSH2*: 20‡

Table continues

meant that traditional cohort and case-control study designs have not been feasible. However, despite this lack of conventional epidemiologic evidence, subsequent studies have provided convincing evidence to support the hypothesis that mismatch repair gene mutations cause a subset of colorectal cancer cases.

The most compelling supportive evidence comes from studies which demonstrate that mutation carriers are at greatly increased risk of developing colorectal cancer in comparison with the general population. Such studies are summarized in table 2. Aarnio et al. (47) calculated a standardized incidence ratio of 68 (95 percent confidence interval: 56, 81) for Finnish carriers of *hMLH1* or *hMSH2* mutations. In the other studies identified in table 2, researchers did not make a formal calculation of the standardized incidence ratio, but approximate estimates utilizing appropriate cancer registry data consistently show that the risk of colorectal cancer in mutation carriers is greatly in excess of the corresponding risk in the general population (see table 2). The relative risk of 8.1 (95 percent confidence interval: 3.5, 15.9) for first-degree relatives of mutation carriers observed by Millar et al. (48) is consistent with a risk

TABLE 2. Continued

Published reference	Area of study	Ascertainment of index cases	No. of index cases	Ascertainment of mutation carriers	No. of mutation carriers	Penetrance* in mutation carriers	Source of data for comparison	Risk in comparison group	Standardized incidence ratio/ relative risk
Millar et al. (48)	Canada	Women with both colorectal cancer and endometrial cancer before age 70 years, identified through the Ontario Cancer Registry and/or the tumor registry at Princess Margaret Hospital, Toronto, and harboring *hMLH1* or *hMSH2* mutations.	7	First-degree relatives were identified. Carrier status was *not* determined.	N/A	N/A	Ontario provincial cancer rate		First-degree relatives of mutation carriers 8.1 (95% CI: 3.5, 15.9); first-degree relatives of mutation-negative probands: 2.8 (95% CI: 1.7, 4.5)
Lin et al. (53)	United States	Kindreds were known to have mutations in *hMLH1* ($n = 2$) or *hMSH2* ($n = 2$). No further detail was given on how these kindreds were ascertained.	4	Mutation carriers were identified by testing ($n = 78$) or determined to be obligate carriers ($n = 27$).	105	*hMLH1*: Males = 94%; females = 63%; females + males = 84% *hMSH2*: Males = 96%; females = 39%; females + males = 71%	N/A		N/A

* Males: penetrance in males only; Females: penetrance in females only; Females + Males: penetrance in group comprising both sexes.

† HNPCC, hereditary nonpolyposis colorectal cancer; *hMLH1*, human mutL homolog 1; *hMSH2*, human mutS homolog 2; N/A, not applicable; CI, confidence interval; EUCAN, European Network of Cancer Registries.

‡ Where EUCAN data have been used for comparison, the estimate of the standardized incidence ratio is a crude one and does not take into account the age structure of the mutation carrier group. Because of the approximate nature of this comparison, we did not consider it appropriate to calculate confidence intervals for these estimates.

that is an order of magnitude greater in mutation carriers than in noncarriers.

The clinical presentation of colorectal cancer among mutation carriers appears to differ from that found among persons with sporadic cases in several respects, an observation that indirectly supports the hypothesis that mutations in mismatch repair genes account for a distinct subset of colorectal cancer cases. The most obvious clinical characteristic associated with colorectal cancer among mismatch repair gene mutation carriers is familial aggregation. Part *a* of the second supplementary table, which is available on the website of the Human Genome Epidemiology Network (http://www.cdc.gov/genomics/hugenet/default.htm) and the *Journal*'s website (http://aje.oupjournals.org/), provides details on mutation analysis studies conducted among patients selected on the basis of family history. The results of these studies are summarized in tables 3 and 4. The observed prevalence of potentially pathogenic mutations in individuals meeting the Amsterdam criteria is remarkably consistent across different populations (table 4).

MSI is evident in 12–15 percent of sporadic colorectal cancer cases, compared with over 90 percent of cases defined, according to the Amsterdam criteria, as being from HNPCC kindreds (49). MSI is currently thought to result from defective mismatch repair, although evidence to support this hypothesis is limited by two factors. Firstly, the vast majority of studies that examine mutations in MSI-positive patients concentrate on HNPCC families, introducing considerable bias. Secondly, few investigators look systematically for mutations in patients with MSI-negative tumors. Interestingly, when this has been done, there have been a few instances in which tumors from patients with identified mutations in *hMLH1* or *hMSH2* have not exhibited the MSI phenotype (45, 50, 51). Analysis of all published results from one research group showed that, among kindreds with suspected HNPCC, germline mutations could be detected in

TABLE 3. Association between the extent of family history of colorectal cancer and the prevalence of mismatch repair gene mutations

Family history criteria*	No. of studies	No. of index cases	*hMLH1* mutation carriers		*hMSH2* mutation carriers		Published references (ref. no. from current review)
			No.	%	No.	%	
Fulfillment of the Amsterdam criteria	27	534	145	27.2	87	16.3	(18, 50, 57, 93–116)
Strong family history not fulfilling the Amsterdam criteria	25	494	46	9.3	43	8.7	(18, 50, 57, 94, 96–101, 103–106, 108–118)

* See appendix 2.

16 out of 22 colorectal cancer patients with MSI-positive tumors, as compared with one out of 37 mutations in MSI-negative patients (52). The presence of mutations in MSI-negative cases may reflect mechanisms of tumorigenesis in people with mismatch repair gene mutations that do not require mutation instability. Mutation analysis studies involving patients selected on the basis of MSI are summarized in the second supplementary table, part *b*.

The association between early age of colorectal cancer onset and *hMLH1*/*hMSH2* gene mutations is often confounded by the fact that the selection criteria have included family history, but a few studies have performed mutation analysis on patients selected solely on the basis of early age of onset. As is illustrated in table 5, these studies demonstrate a trend towards a higher pathogenic mutation detection rate in individuals diagnosed at a relatively young age, an observation that is consistent with the hypothesis that these genes are involved in colorectal cancer tumorigenesis. Details on the studies considered in table 5 can be found in the second supplementary table, part *c*.

Penetrance

While it has become widely accepted that mutations in the mismatch repair genes *hMLH1* and *hMSH2* play a causal role in a subset of colorectal cancer cases, the precise pene-

TABLE 4. Results of mutation analysis in patients fulfilling the Amsterdam criteria* for colorectal cancer, by geographic origin

Country	No. of index cases	*hMLH1* mutation carriers	*hMSH2* mutation carriers	Published reference
		Asia		
Japan	15	1	8	Bai et al. (94)
Japan	11	5	0	Miyaki et al. (107)
Japan	4	0	1	Nomura et al. (109)
Korea	25	8	0	Han et al. (100)
Total	55	14 (25.5%)	9 (16.4%)	
		Europe		
Russia/Moldavia	7	1	3	Maliaka et al. (106)
Sweden	21	5	1	Tannergard et al. (112) and Wahlberg et al. (113)
Sweden	7	1	0	Liu et al. (104)
Switzerland	10	3	3	Buerstedde et al. (95)
Switzerland	15	6	4	Heinimann et al. (57)
Switzerland	14	10	0	Hutter et al. (101)
Italy	14	4	3	Pensotti et al. (110)
Italy	18	1	2	de Leon et al. (98)
Italy	17	5	2	Viel et al. (119)
Italy	17	2	3	Curia et al. (97)
Italy	13	3	3	Calistri et al. (96)
France	10	3	2	Dieumegard et al. (99)
France	3	2	0	Wang et al., 1997 (114)
France	22	11	3	Wang et al., 1999 (120)
Holland and Norway	92	25	16	Wijnen et al. (116)
Germany	57	11	4	Lamberti et al. (103)
England	17	3	5	Froggatt et al. (93)
Total	344	96 (27.9%)	54 (15.7%)	
		Australia		
Australia	18	4	2	Kohonen-Corish et al. (102)
Australia	33	11	9	Scott et al. (18)
Total	51	15 (29.4%)	11 (21.6%)	
		North America		
USA	12	4	2	Luce et al. (105)
USA	28	10	1	Syngal et al. (111)
Canada	14	2	5	Bapat et al. (50)
Total	54	16 (29.6%)	8 (14.8%)	

* See appendix 2.

TABLE 5. Association between age at onset of colorectal cancer and mismatch repair gene mutations

Age range (years)	No. of studies	No. of index cases	*hMLH1* mutation carriers		*hMSH2* mutation carriers		Published reference(s)
			No.	%	No.	%	
<30	1	50	7	14	7	14	Farrington et al. (45)
<40	1	12	1	8.3	1	8.3	Syngal et al. (111)
<45	1	38	1	2.6	2	5.3	Fornasarig et al. (121)
<50	6	135	6	4.4	6	4.4	Dieumegard et al. (99), Montera et al. (122), Tomlinson et al. (123), Wang et al., 1997 (114), Wang et al., 1999 (120), Weber et al. (124), and Yuan et al. (118)

trance of these mutations remains unknown. A number of studies, summarized in table 2, have addressed this issue. Results are presented differently for each study, so direct comparison is difficult. One consistent finding is that risk is higher among male mutation carriers (approximately 80 percent by age 70 years) than among females (approximately 40 percent by age 70 years), an observation with important implications for patient management and surveillance. Observed differences in penetrance between carriers of *hMLH1* or *hMSH2* mutations (53, 54) await confirmation in future studies.

A study by Aarnio et al. (47) classified relatives of clinically defined HNPCC cases as being at a 25 percent, 50 percent, or 100 percent risk of being mutation carriers and calculated the cumulative incidence of colorectal cancer up to age 70 years as being 100 percent and 54 percent for males and females, respectively. A potential source of bias in this particular study is the fact that the majority of the probands had one of the Finnish founder mutations. A similar study carried out in Amsterdam Dutch kindreds calculated risk of colorectal cancer among mutation carriers at age 75 years to be 92 percent in males and 83 percent in females (55).

These studies used family history as a selection criterion, an approach that introduces considerable ascertainment bias. Kindreds identified in this way will inherently have an unusually large number of colorectal cancer cases, and estimates of penetrance obtained in this way are likely to be falsely high. Dunlop et al. (56) used an alternative approach to identify mutation-carrying probands from the Scottish population, performing mutation analysis on colorectal cancer patients with a very early age of onset (<30 years). The cumulative incidence of colorectal cancer among relatives proven to be mutation carriers was found to be 74 percent in males and 30 percent in females at age 70 (56).

Note that the identification of families with mismatch repair gene mutations using any phenotypic selection criteria introduces ascertainment bias, and such kindreds may not be representative of all mutation-carrying families in the general population. Thus, there is a considerable need for estimates of penetrance based on systematically collected familial or population data.

Survival

Prior to the identification of mismatch repair genes, several studies suggested that the prognosis for patients with colorectal cancer due to HNPCC was more favorable than that for patients with sporadic colorectal cancer. Whether improved prognosis is specifically a feature of colorectal cancer in patients harboring mismatch repair gene mutations is not yet clear, although preliminary evidence suggests that this may be the case (57, 58).

A possible explanation for this phenomenon may be that the high frequency of mutations characteristic of mismatch repair-deficient tumors actually restricts tumor growth (58). However, kindreds included in survival analysis studies on the basis of a strong family history of colorectal cancer have, by definition, survived to produce a large family group for analysis. Therefore, these kindreds may not be representative of all mutation carriers, and there is a need for survival data from unselected, population-based cohort studies.

It has also been postulated that mismatch repair deficiency may have an effect on response to chemotherapy. Results are not entirely consistent, but several studies suggest an association between *hMLH1/hMSH2* deficiency in cell lines and resistance to chemotherapeutic agents (59–62).

INTERACTIONS

While the exact penetrance of specific mutations in *hMLH1* and *hMSH2* is unknown, it is not complete. Consequently, the age-related risk, pathologic features, and outcomes associated with such mutations are subject to modification by other genetic and environmental factors.

The body of epidemiologic data regarding modification of disease resulting from mismatch repair gene mutations is somewhat limited. The effects of known environmental risk factors for colorectal cancer in mutation carriers are largely unstudied, and much of the suggestive evidence for interactions comes indirectly from studies using MSI-positive or clinically defined HNPCC cases as a surrogate for mutation carriers. Furthermore, the apparent presence of a statistical interaction between mismatch repair gene mutations and other genetic or environmental factors does not necessarily imply the existence of a biologic or causal interaction. Therefore, the studies considered below do not constitute evidence for true interactions involving *hMLH1* and *hMSH2*,

although they may prove useful in terms of identifying potential interactions that merit further investigation.

Gene-environment interactions

Reports by Ruschoff et al. (63) and Yamamoto et al. (64) have suggested that treatment of *hMLH1*- or *hMSH2*-deficient cell lines with nonsteroidal antiinflammatory drugs leads to a significant reduction in the proportion of cells exhibiting MSI, indicating that this phenotypic manifestation of mismatch repair deficiency may be modified by these drugs.

Slattery et al. (65) have presented evidence suggesting that an interaction may exist between MSI and smoking. Compared with patients with MSI-negative tumors, patients with MSI-positive tumors were more likely to be heavy smokers: Odds ratios were 1.6 (95 percent confidence interval: 1.0, 2.5) in men and 2.2 (95 percent confidence interval: 1.4, 3.5) in women (65). These results are supported by those of another recent study (66), and the implication that smoking is specifically associated with a particular subset of colorectal cancer cases is consistent with the weak associations reported between smoking and sporadic colon cancer. It is possible that mismatch repair deficiency is involved in the observed association between smoking and MSI, but further studies involving known mutation carriers will be required to confirm this hypothesis.

Another recent paper by Slattery et al. (67) showed that the risk of MSI-positive colon cancer may be reduced by estrogens and increased by estrogen withdrawal.

Dietary heterocyclic aromatic amines are another risk factor that requires further evaluation. Wu et al. (66) found that patients with MSI-positive tumors had received a relatively high dietary exposure to heterocyclic aromatic amines, an observation that remained significant after adjustment for smoking and red meat intake. This finding is consistent with laboratory studies, which have shown that rats exposed to particular heterocyclic amines showed the trait of MSI (68).

Gene-gene interactions

Risk of colorectal cancer among female *hMLH1*/*hMSH2* mutation carriers is approximately half the risk in male mutation carriers (47, 56). In the absence of clear evidence of hormonal influence, the presence of a genetic modifier, X-linked or otherwise, remains a possibility.

The possibility of interaction between mismatch repair genes and other genes known to influence colorectal cancer susceptibility is an area that merits consideration. Initial studies have suggested that genes involved in carcinogen metabolism might modify the phenotypic expression of mismatch repair gene mutations. For example, Moisio et al. (69) demonstrated that a specific polymorphism in the gene encoding the xenobiotic enzyme *N*-acetyltransferase 1 was associated with a lower age of colorectal cancer onset in Finnish HNPCC kindreds with identified mutations in *hMLH1*. Similarly, an alteration in cyclin D1 has been associated with earlier age of onset in HNPCC cases; patients who harbor the mutant cyclin D1 allele develop cancer an average of 11 years earlier than patients with two wild-type alleles (70).

Murine studies have demonstrated that *MSH2* deficiency accelerates intestinal tumorigenesis in transgenic mice that are heterozygous for a germline mutation in the adenomatous polyposis coli gene (71). Similarly, Toft et al. (72) have used mice mutant for both *MSH2* and p53 to demonstrate interaction between these genes. Additionally, in-vitro studies have suggested that interactions may exist between mismatch repair genes and transforming growth factor-β receptor II (73). While these molecular studies demonstrate that gene-gene interactions may be worth further investigation, the above hypotheses have yet to be tested in human populations for relevance to cancer susceptibility.

LABORATORY TESTING

The heterogeneity of mutation types found in *hMLH1* and *hMSH2* has meant that many different techniques have been employed to test for mutations in these genes. A number of techniques are described below, along with their benefits and disadvantages.

In vitro synthesized protein assay

The in vitro synthesized protein assay technique uses an in vitro system to transcribe and translate a large polymerase chain reaction product containing several exons. The translated product is separated on a polyacrylamide gel electrophoresis system, and potential mutations are identified as truncated bands. These may represent a number of mutations that have the effect of altering splicing, therefore producing a translated fragment with certain exons deleted. Out-of-frame deletions or insertions, resulting in frameshifts or splice variants, will also be detected using this method.

In vitro synthesized protein assay does not detect missense mutations, in-frame deletions or insertions, large genomic deletions involving numerous exons, promoter mutations, or mutations that silence the gene. The assay also requires the use of mRNA for the production of a cDNA polymerase chain reaction product.

Genomic sequencing

cDNA sequencing also relies on mRNA being available. It will identify all mutation types except large genomic deletions, promoter mutations, and gene silencing mutations. Genomic sequencing detects even fewer changes than cDNA, but it does have the advantage of only requiring genomic DNA. Table 6 shows a comparison of the sensitivity of the two techniques, in vitro synthesized protein assay and genomic sequencing, as described by Farrington et al. (45).

DNA structure techniques

A number of techniques rely on changes in DNA structure created by a mutation. These include denaturing gradient gel electrophoresis (74), including the adaptation of using two-dimensional gel electrophoresis (75), single-strand confor-

TABLE 6. Sensitivity of mutation detection techniques

	Sensitivity (%)	Published reference
In vitro synthesized protein assay	69	Farrington et al. (45)
Genomic sequencing	80	Farrington et al. (45)
In vitro synthesized protein assay/genomic sequencing	93	Farrington et al. (45)
Denaturing gradient gel electrophoresis	>67	Fidalgo et al. (125)
Single-strand conformational polymorphism	>67	Fidalgo et al. (125)
Protein truncation test	50	Fidalgo et al. (125)
Heteroduplex analysis	19	Fidalgo et al. (125)
Two-dimensional DNA typing	—*	Sasaki et al. (75)

* Comparable to that of denaturing gradient gel electrophoresis.

mational polymorphism analysis (76), heteroduplex analysis, and denaturing high-performance liquid chromatography.

Table 6 summarizes the available information regarding the sensitivity of the above techniques. The use of various combinations of techniques may enhance sensitivity, but this is usually impractical. Recently, Yan et al. (77) demonstrated that the conversion of chromosomes from the diploid state to the haploid state, by fusion to a recipient rodent cell line, may facilitate improved sensitivity of current mutation detection techniques.

POPULATION TESTING

The population prevalence of *hMLH1/hMSH2* mutation carriers in the Scottish population aged 15–74 years has been estimated at 1 in 3,139 (78). A recent UK National Screening Committee workshop concluded that there is currently no case to offer population screening in an attempt to identify mutation carriers (Rose et al., UK National Screening Committee, unpublished data). Authors in the United States have reached similar conclusions, agreeing that more information regarding the prevalence and penetrance of mismatch repair gene mutations and more evidence of effective intervention strategies are essential prerequisites for implementing screening outside of the research context (79–84).

There are essentially two strategies that could be employed to search for mutations in the context of population screening: searching the entire gene(s) for mutations using the techniques considered above or looking for specific mutations. The latter option is far less expensive and labor-intensive and could be of particular benefit in countries where specific "founder" mutations are prevalent. It may also be possible to apply DNA pooling strategies in this context to enhance efficiency (85). However, this approach is not currently feasible because of the extreme heterogeneity of mismatch repair gene variants and the low allele frequency of individual mutations. The ethical issues inherent in genetic screening, coupled with the poor efficiency and high cost of detecting mutations using current technology, mean that population testing in any form is unlikely to be recommended in the near future.

Another approach to identifying mutation carriers is performing mutation analysis in colorectal cancer patients deemed to be at high risk of harboring mutations, and subsequently performing "cascade screening" of their relatives. The major issue in the context of a cascade screening program is that of how resources can be efficiently targeted towards the identification of kindreds with *hMLH1/hMSH2* mutations. This issue is considered in detail in an overview of findings from one research group (52), and the sensitivity and specificity of various clinical criteria are considered by Syngal et al. (86).

Currently, most mutation carriers are identified by referral of patients with a family history of colorectal cancer to cancer genetics services. Another option, under investigation in an ongoing program in Scotland, is to search for mismatch repair gene mutations among persons with early-onset colorectal cancer and subsequently perform cascade screening in the relatives of mutation carriers. The phenotypic features of age at onset, family history, and MSI are commonly used selection criteria in mutation analysis studies, as summarized for reference in the second supplementary table.

At the present time, there is no consensus regarding the most efficient approach to identifying mutation carriers. It is clear, however, that further understanding of the role of mismatch repair genes in colorectal cancer has important scientific and clinical implications.

REFERENCES

1. International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. ICG-HNPCC mutation database. Leiden, The Netherlands: International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer, 2001. (World Wide Web URL: http://www.nfdht.nl/).
2. Leach FS, Nicolaides NC, Papadopoulos N, et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell 1993;75:1215–25.
3. Peltomaki P, Aaltonen LA, Sistonen P, et al. Genetic mapping of a locus predisposing to human colorectal cancer. Science 1993;260:810–12.
4. Modrich P. Mechanisms and biological effects of mismatch repair. Annu Rev Genet 1991;25:229–53.
5. Lindblom A, Tannergard P, Werelius B, et al. Genetic mapping of a second locus predisposing to hereditary nonpolyposis colon cancer. Nature Genet 1993;5:279–82.
6. Nystrom-Lahti M, Sistonen P, Mecklin JP, et al. Close linkage to chromosome 3p and conservation of ancestral founding

haplotype in hereditary nonpolyposis colorectal cancer families. Proc Natl Acad Sci U S A 1994;91:6054–8.
7. Nystrom-Lahti M, Holmberg M, Fidalgo P, et al. Missense and nonsense mutations in codon 659 of *MLH1* cause aberrant splicing of messenger RNA in HNPCC kindreds. Genes Chromosomes Cancer 1999;26:372–5.
8. Ellison AR, Lofing J, Bitter GA. Functional analysis of human *MLH1* and *MSH2* missense variants and hybrid human-yeast *MLH1* proteins in *Saccharomyces cerevisiae*. Hum Mol Genet 2001;10:1889–900.
9. Nystrom-Lahti M, Perrera C, Raschle M, et al. Functional analysis of *MLH1* mutations linked to hereditary nonpolyposis colon cancer. Genes Chromosomes Cancer 2002;33: 160–7.
10. Trojan J, Zeuzem S, Randolph A, et al. Functional analysis of *hMLH1* variants and HNPCC-related mutations using a human expression system. Gastroenterology 2002;122:211–19.
11. Wijnen J, van der KH, Vasen H, et al. *MSH2* genomic deletions are a frequent cause of HNPCC. (Letter). Nature Genet 1998;20:326–8.
12. Nystrom-Lahti M, Wu Y, Moisio A-L, et al. DNA mismatch repair gene mutations in 55 kindreds with verified or putative hereditary nonpolyposis colorectal cancer. Hum Mol Genet 1996;5:763–9.
13. Salovaara R, Loukola A, Kristo P, et al. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. J Clin Oncol 2000;18:2193–200.
14. Jager AC, Bisgaard ML, Myrhoj T, et al. Reduced frequency of extracolonic cancers in hereditary nonpolyposis colorectal cancer families with monoallelic *hMLH1* expression. Am J Hum Genet 1997;61:129–38
15. Froggatt NJ, Green J, Brassett C, et al. A common *MSH2* mutation in English and North American HNPCC families: origin, phenotypic expression, and sex specific differences in colorectal cancer. J Med Genet 1999;36:97–102.
16. Desai DC, Lockman JC, Chadwick RB, et al. Recurrent germline mutation in *MSH2* arises frequently de novo. J Med Genet 2000;37:646–52.
17. Eng C, Hampel H, de la CA. Genetic testing for cancer predisposition. Annu Rev Med 2001;52:371–400
18. Scott RJ, McPhillips M, Meldrum CJ, et al. Hereditary nonpolyposis colorectal cancer in 95 families: differences and similarities between mutation-positive and mutation-negative kindreds. Am J Hum Genet 2001;68:118–27.
19. International Agency for Research on Cancer. GLOBOCAN 2000: cancer incidence, mortality and prevalence worldwide. Lyon, France: International Agency for Research on Cancer, 2000. (World Wide Web URL: http://www.who.int).
20. International Agency for Research on Cancer. Cancer incidence in five continents. Vol 7. (IARC Scientific Publication no. 143). Lyon, France: International Agency for Research on Cancer, 1997.
21. American Cancer Society. Cancer facts and figures 2002. New York, NY: American Cancer Society, 2002. (World Wide Web URL: http://www.cancer.org).
22. Parkin DM. Studies of cancer in migrant populations: methods and interpretation. Rev Epidemiol Sante Publique 1992; 40:410–24.
23. Potter JD (Chair). Food, nutrition and the prevention of cancer: a global perspective. London, United Kingdom: World Cancer Research Fund, 2001. (World Wide Web URL: http://www.wcrf.org.uk).
24. Potter JD. Colorectal cancer: molecules and populations. J Natl Cancer Inst 1999;91:916–32.
25. Alberts DS, Martinez ME, Roe DJ, et al. Lack of effect of a high-fiber cereal supplement on the recurrence of colorectal adenomas. Phoenix Colon Cancer Prevention Physicians' Network. N Engl J Med 2000;342:1156–62.
26. Schatzkin A, Lanza E, Corle D, et al. Lack of effect of a low-fat, high-fiber diet on the recurrence of colorectal adenomas. Polyp Prevention Trial Study Group. N Engl J Med 2000;342: 1149–55.
27. Fuchs CS, Giovannucci EL, Colditz GA, et al. Dietary fiber and the risk of colorectal cancer and adenoma in women. N Engl J Med 1999;340:169–76.
28. Rijnkels JM, Hollanders VM, Woutersen RA, et al. Absence of an inhibitory effect of a vegetables-fruit mixture on the initiation and promotion phases of azoxymethane-induced colorectal carcinogenesis in rats fed low- or high-fat diets. Nutr Cancer 1998;30:124–9.
29. van Kranen HJ, van Iersel PW, Rijnkels JM, et al. Effects of dietary fat and a vegetable-fruit mixture on the development of intestinal neoplasia in the ApcMin mouse. Carcinogenesis 1998;19:1597–601.
30. Giardiello FM, Hamilton SR, Krush AJ, et al. Treatment of colonic and rectal adenomas with sulindac in familial adenomatous polyposis. N Engl J Med 1993;328:1313–16.
31. Steinbach G, Lynch PM, Phillips RK, et al. The effect of celecoxib, a cyclooxygenase-2 inhibitor, in familial adenomatous polyposis. N Engl J Med 2000;342:1946–52.
32. Giacosa A, Franceschi S, La Vecchia C, et al. Energy intake, overweight, physical exercise and colorectal cancer risk. Eur J Cancer Prev 1999;8(suppl 1):S53–60.
33. La Vecchia C, Negri E, Decarli A, et al. Diabetes mellitus and colorectal cancer risk. Cancer Epidemiol Biomarkers Prev 1997;6:1007–10.
34. La Vecchia C, Negri E, Franceschi S, et al. Aspirin and colorectal cancer. Br J Cancer 1997;76:675–7.
35. Giovannucci E. An updated review of the epidemiological evidence that cigarette smoking increases risk of colorectal cancer. Cancer Epidemiol Biomarkers Prev 2001;10:725–31.
36. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78–85.
37. Nystrom-Lahti M, Parsons R, Sistonen P, et al. Mismatch repair genes on chromosomes 2p and 3p account for a major share of hereditary nonpolyposis colorectal cancer families evaluable by linkage. Am J Hum Genet 1994;55:659–65.
38. Froggatt NJ, Koch J, Davies R, et al. Genetic linkage analysis in hereditary nonpolyposis colon cancer syndrome. J Med Genet 1995;32:352–7.
39. Aaltonen LA, Peltomaki P, Leach FS, et al. Clues to the pathogenesis of familial colorectal cancer. Science 1993;260: 812–16.
40. Ionov Y, Peinado MA, Malkhosyan S, et al. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 1993;363:558–61.
41. Strand M, Prolla TA, Liskay RM, et al. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 1993;365:274–6.
42. Bronner CE, Baker SM, Morrison PT, et al. Mutation in the DNA mismatch repair gene homologue *hMLH1* is associated with hereditary nonpolyposis colon cancer. Nature 1994;368: 258–61.
43. Fishel R, Lescoe MK, Rao MR, et al. The human mutator gene homolog *MSH2* and its association with hereditary nonpolyposis colon cancer. Cell 1993;75:1027–38.

44. Papadopoulos N, Nicolaides NC, Wei YF, et al. Mutation of a mutL homolog in hereditary colon cancer. Science 1994;263:1625–9.
45. Farrington SM, Lin-Goerke J, Ling J, et al. Systematic analysis of *hMSH2* and *hMLH1* in young colon cancer patients and controls. Am J Hum Genet 1998;63:749–59.
46. Samowitz WS, Curtin K, Lin HH, et al. The colon cancer burden of genetically defined hereditary nonpolyposis colon cancer. Gastroenterology 2001;121:830–8.
47. Aarnio M, Sankila R, Pukkala E, et al. Cancer risk in mutation carriers of DNA-mismatch-repair genes. Int J Cancer 1999;81:214–18.
48. Millar AL, Pal T, Madlensky L, et al. Mismatch repair gene defects contribute to the genetic basis of double primary cancers of the colorectum and endometrium. Hum Mol Genet 1999;8:823–9.
49. Boland CR. Molecular genetics of hereditary nonpolyposis colorectal cancer. Ann N Y Acad Sci 2000;910:50–9.
50. Bapat BV, Madlensky L, Temple LKF, et al. Family history characteristics, tumor microsatellite instability and germline *MSH2* and *MLH1* mutations in hereditary colorectal cancer. Hum Genet 1999;104:167–76.
51. Debniak T, Kurzawski G, Gorski B, et al. Value of pedigree/clinical data, immunohistochemistry and microsatellite instability analyses in reducing the cost of determining *hMLH1* and *hMSH2* gene mutations in patients with colorectal cancer. Eur J Cancer 2000;36:49–54.
52. Liu T, Wahlberg S, Burek E, et al. Microsatellite instability as a predictor of a mutation in a DNA mismatch repair gene in familial colorectal cancer. Genes Chromosomes Cancer 2000;27:17–25.
53. Lin KM, Shashidharan M, Thorson AG, et al. Cumulative incidence of colorectal and extracolonic cancers in *MLH1* and *MSH2* mutation carriers of hereditary nonpolyposis colorectal cancer. J Gastrointest Surg 1998;2:67–71.
54. Vasen HF, Stormorken A, Menko FH, et al. *MSH2* mutation carriers are at higher risk of cancer than *MLH1* mutation carriers: a study of hereditary nonpolyposis colorectal cancer families. J Clin Oncol 2001;19:4074–80.
55. Vasen HF, Wijnen JT, Menko FH, et al. Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis. Gastroenterology 1996;110:1020–7.
56. Dunlop MG, Farrington SM, Carothers AD, et al. Cancer risk associated with germline DNA mismatch repair gene mutations. Hum Mol Genet 1997;6:105–10.
57. Heinimann K, Scott RJ, Buerstedde J-M, et al. Influence of selection criteria on mutation detection in patients with hereditary nonpolyposis colorectal cancer. Cancer 1999;85:2512–18.
58. Sankila R, Aaltonen LA, Jarvinen HJ, et al. Better survival rates in patients with *MLH1*-associated hereditary colorectal cancer. Gastroenterology 1996;110:682–7.
59. Aebi S, Fink D, Gordon R, et al. Resistance to cytotoxic drugs in DNA mismatch repair-deficient cells. Clin Cancer Res 1997;3:1763–7.
60. Carethers JM, Chauhan DP, Fink D, et al. Mismatch repair proficiency and in vitro response to 5-fluorouracil. Gastroenterology 1999;117:123–31.
61. Fink D, Aebi S, Howell SB. The role of DNA mismatch repair in drug resistance. Clin Cancer Res 1998;4:1–6.
62. Fujieda S, Tanaka N, Sunaga H, et al. Expression of *hMSH2* correlates with in vitro chemosensitivity to CDDP cytotoxicity in oral and oropharyngeal carcinoma. Cancer Lett 1998;132:37–44.
63. Ruschoff J, Wallinger S, Dietmaier W, et al. Aspirin suppresses the mutator phenotype associated with hereditary non-polyposis colorectal cancer by genetic selection. Proc Natl Acad Sci U S A 1998;95:11301–6.
64. Yamamoto H, Itoh F, Fukushima H, et al. Overexpression of cyclooxygenase-2 protein is less frequent in gastric cancers with microsatellite instability. Int J Cancer 1999;84:400–3.
65. 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;92:1831–6.
66. Wu AH, Shibata D, Yu MC, et al. Dietary heterocyclic amines and microsatellite instability in colon adenocarcinomas. Carcinogenesis 2001;22:1681–4.
67. Slattery ML, Potter JD, Curtin K, et al. Estrogens reduce and withdrawal of estrogens increase risk of microsatellite instability-positive colon cancer. Cancer Res 2001;61:126–30.
68. Canzian F, Ushijima T, Serikawa T, et al. Instability of microsatellites in rat colon tumors induced by heterocyclic amines. Cancer Res 1994;54:6315–17.
69. Moisio AL, Sistonen P, Mecklin JP, et al. Genetic polymorphisms in carcinogen metabolism and their association to hereditary nonpolyposis colon cancer. Gastroenterology 1998;115:1387–94.
70. Kong S, Amos CI, Luthra R, et al. Effects of cyclin D1 polymorphism on age of onset of hereditary nonpolyposis colorectal cancer. Cancer Res 2000;60:249–52.
71. Reitmair AH, Cai JC, Bjerknes M, et al. *MSH2* deficiency contributes to accelerated APC-mediated intestinal tumorigenesis. Cancer Res 1996;56:2922–6.
72. Toft NJ, Winton DJ, Kelly J, et al. *Msh2* status modulates both apoptosis and mutation frequency in the murine small intestine. Proc Natl Acad Sci U S A 1999;96:3911–15.
73. Markowitz S. TGF-beta receptors and DNA repair genes, coupled targets in a pathway of human colon carcinogenesis. Biochim Biophys Acta 2000;1470:M13–20.
74. Wijnen J, Vasen H, Khan PM, et al. Seven new mutations in *hMSH2*, an HNPCC gene, identified by denaturing gradient-gel electrophoresis. Am J Hum Genet 1995;56:1060–6.
75. Sasaki S, Tokino T, Miyatsu T, et al. Mutational analysis of the *hMLH1* gene using an automated two-dimensional DNA typing system. Hum Mutat 1997;9:164–71.
76. Bubb VJ, Curtis LJ, Cunningham C, et al. Microsatellite instability and the role of *hMSH2* in sporadic colorectal cancer. Oncogene 1996;12:2641–9.
77. Yan H, Papadopoulos N, Marra G, et al. Conversion of diploidy to haploidy. Nature 2000;403:723–4.
78. Dunlop MG, Farrington SM, Nicholl I, et al. Population carrier frequency of *hMSH2* and *hMLH1* mutations. Br J Cancer 2000;83:1643–5.
79. Brown ML, Kessler LG. The use of gene tests to detect hereditary predisposition to cancer: economic considerations. J Natl Cancer Inst 1995;87:1131–6.
80. Brown ML, Kessler LG. Use of gene tests to detect hereditary predisposition to cancer: what do we know about cost effectiveness? Int J Cancer 1996;69:55–7.
81. Coughlin SS, Miller DS. Public health perspectives on testing for colorectal cancer susceptibility genes. Am J Prev Med 1999;16:99–104.
82. Ponder BA. Costs, benefits and limitations of genetic testing for cancer risk. Br J Cancer 1999;80(suppl 1):46–50.
83. Terdiman JP, Conrad PG, Sleisenger MH. Genetic testing in hereditary colorectal cancer: indications and procedures. Am J Gastroenterol 1999;94:2344–56.
84. Welch HG, Burke W. Uncertainties in genetic testing for chronic disease. JAMA 1998;280:1525–7.

85. Amos CI, Frazier ML, Wang W. DNA pooling in mutation detection with reference to sequence analysis. Am J Hum Genet 2000;66:1689–92.
86. Syngal S, Fox EA, Eng C, et al. Sensitivity and specificity of clinical criteria for hereditary nonpolyposis colorectal cancer associated mutations in *MSH2* and *MLH1*. J Med Genet 2000; 37:641–5.
87. Vasen HF, Mecklin JP, Khan PM, et al. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 1991;34:424–5.
88. Bellacosa A, Genuardi M, Anti M, et al. Hereditary nonpolyposis colorectal cancer: review of clinical, molecular genetics, and counseling aspects. Am J Med Genet 1996;62:353–64.
89. Fujita S, Moriya Y, Sugihara K, et al. Prognosis of hereditary nonpolyposis colorectal cancer (HNPCC) and the role of Japanese criteria for HNPCC. Jpn J Clin Oncol 1996;26:351–5.
90. Rodriguez-Bigas MA, Boland CR, Hamilton SR, et al. A National Cancer Institute workshop on hereditary nonpolyposis colorectal cancer syndrome: meeting highlights and Bethesda guidelines. J Natl Cancer Inst 1997;89:1758–62.
91. European Network of Cancer Registries (EUCAN). Cancer incidence, mortality and prevalence in the European Union. Lyon, France: International Agency for Research on Cancer, 1999.
92. Aarnio M, Mecklin JP, Aaltonen LA, et al. Life-time risk of different cancers in hereditary nonpolyposis colorectal cancer (HNPCC) syndrome. Int J Cancer 1995;64:430–3.
93. Froggatt NJ, Brassett C, Koch DJ, et al. Mutation screening of *MSH2* and *MLH1* mRNA in hereditary nonpolyposis colon cancer syndrome. J Med Genet 1996;33:726–30.
94. Bai YQ, Akiyama Y, Nagasaki H, et al. Predominant germline mutation of the *hMSH2* gene in Japanese hereditary nonpolyposis colorectal cancer kindreds. Int J Cancer 1999;82: 512–15.
95. Buerstedde JM, Alday P, Torhorst J, et al. Detection of new mutations in six out of 10 Swiss HNPCC families by genomic sequencing of the *hMSH2* and *hMLH1* genes. J Med Genet 1995;32:909–12.
96. Calistri D, Presciuttini S, Buonsanti G, et al. Microsatellite instability in colorectal-cancer patients with suspected genetic predisposition. Int J Cancer 2000;89:87–91.
97. Curia MC, Palmirotta R, Aceto G, et al. Unbalanced germline expression of *hMLH1* and *hMSH2* alleles in hereditary nonpolyposis colorectal cancer. Cancer Res 1999;59:3570–5.
98. de Leon MP, Pedroni M, Benatti P, et al. Hereditary colorectal cancer in the general population: from cancer registration to molecular diagnosis. Gut 1999;45:32–8.
99. Dieumegard B, Grandjouan S, Sabourin J-C, et al. Extensive molecular screening for hereditary nonpolyposis colorectal cancer. Br J Cancer 2000;82:871–80.
100. Han H-J, Yuan Y, Ku J-L, et al. Germline mutations of *hMLH1* and *hMSH2* genes in Korean hereditary nonpolyposis colorectal cancer. J Natl Cancer Inst 1996;88:1317–19.
101. Hutter P, Couturier A, Membrez V, et al. Excess of *hMLH1* germline mutations in Swiss families with hereditary nonpolyposis colorectal cancer. Int J Cancer 1998;78:680–4.
102. Kohonen-Corish M, Ross VL, Doe WF, et al. RNA-based mutation screening in hereditary nonpolyposis colorectal cancer. Am J Hum Genet 1996;59:818–24.
103. Lamberti C, Kruse R, Ruelfs C, et al. Microsatellite instability—a useful diagnostic tool to select patients at high risk for hereditary nonpolyposis colorectal cancer: a study in different groups of patients with colorectal cancer. Gut 1999;44:839–43.
104. Liu T, Wahlberg S, Rubio C, et al. DGGE screening of mutations in mismatch repair genes (*hMSH2* and *hMLH1*) in 34 Swedish families with colorectal cancer. Clin Genet 1998;53: 131–5.
105. Luce MC, Marra G, Chauhan DP, et al. In vitro transcription/translation assay for the screening of *hMLH1* and *hMSH2* mutations in familial colon cancer. Gastroenterology 1995; 109:1368–74.
106. Maliaka YK, Chudina AP, Belev NF, et al. CpG dinucleotides in the *hMSH2* and *hMLH1* genes are hotspots for HNPCC mutations. Hum Genet 1996;97:251–5.
107. Miyaki M, Konishi M, Muraoka M, et al. Germ line mutations of *hMSH2* and *hMLH1* genes in Japanese families with hereditary nonpolyposis colorectal cancer (HNPCC): usefulness of DNA analysis for screening and diagnosis of HNPCC patients. J Mol Med 1995;73:515–20.
108. Moslein G, Tester DJ, Lindor NM, et al. Microsatellite instability and mutation analysis of *hMSH2* and *hMLH1* in patients with sporadic, familial and hereditary colorectal cancer. Hum Mol Genet 1996;5:1245–52.
109. Nomura S, Sugano K, Kashiwabara H, et al. Enhanced detection of deleterious and other germline mutations of *hMSH2* and *hMLH1* in Japanese hereditary nonpolyposis colorectal cancer kindreds. Biochem Biophys Res Commun 2000;271: 120–9.
110. Pensotti V, Radice P, Presciuttini S, et al. Mean age of tumor onset in hereditary nonpolyposis colorectal cancer (HNPCC) families correlates with the presence of mutations in DNA mismatch repair genes. Genes Chromosomes Cancer 1997; 19:135–42.
111. Syngal S, Fox EA, Li C, et al. Interpretation of genetic test results for hereditary nonpolyposis colorectal cancer: implications for clinical predisposition testing. JAMA 1999;282: 247–53.
112. Tannergard P, Lipford JR, Kolodner R, et al. Mutation screening in the *hMLH1* gene in Swedish hereditary nonpolyposis colon cancer families. Cancer Res 1995;55:6092–6.
113. Wahlberg SS, Nystrom-Lahti M, Kane MF, et al. Low frequency of *hMSH2* mutations in Swedish HNPCC families. (Letter). Int J Cancer 1997;74:134–7.
114. Wang Q, Desseigne F, Lasset C, et al. Germline *hMSH2* and *hMLH1* gene mutations in incomplete HNPCC families. Int J Cancer 1997;73:831–6.
115. Wang Y, Friedl W, Lamberti C, et al. A novel missense mutation in the DNA mismatch repair gene *hMLH1* present among East Asians but not among Europeans. Hum Hered 1998;48: 87–91.
116. Wijnen JT, Vasen HF, Khan PM, et al. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. N Engl J Med 1998;339:511–18.
117. Beck NE, Tomlinson IP, Homfray T, et al. Genetic testing is important in families with a history suggestive of hereditary nonpolyposis colorectal cancer even if the Amsterdam criteria are not fulfilled. Br J Surg 1997;84:233–7.
118. Yuan Y, Han HJ, Zheng S, et al. Germline mutations of *hMLH1* and *hMSH2* genes in patients with suspected hereditary nonpolyposis colorectal cancer and sporadic early-onset colorectal cancer. Dis Colon Rectum 1998;41:434–40.
119. Viel A, Genuardi M, Capozzi E, et al. Characterization of *MSH2* and *MLH1* mutations in Italian families with hereditary nonpolyposis colorectal cancer. Genes Chromosomes Cancer 1997;18:8–18.
120. Wang Q, Lasset C, Desseigne F, et al. Prevalence of germline mutations of *hMLH1*, *hMSH2*, *hPMS1*, *hPMS2*, and *hMSH6* genes in 75 French kindreds with nonpolyposis colorectal cancer. Hum Genet 1999;105:79–85.
121. Fornasarig M, Viel A, Valentini M, et al. Microsatellite instability and *MLH1* and *MSH2* germline defects are related to

clinicopathological features in sporadic colorectal cancer. Oncol Rep 2000;7:39–43.
122. Montera M, Resta N, Simone C, et al. Mutational germline analysis of *hMSH2* and *hMLH1* genes in early onset colorectal cancer patients. (Letter). J Med Genet (Online) 2000;37: E7.
123. Tomlinson IP, Beck NE, Homfray T, et al. Germline HNPCC gene variants have little influence on the risk for sporadic colorectal cancer. J Med Genet 1997;34:39–42.
124. Weber TK, Chin H-M, Rodriguez-Bigas M, et al. Novel *hMLH1* and *hMSH2* germline mutations in African Americans with colorectal cancer. JAMA 1999;281:2316–20.
125. Fidalgo P, Almeida MR, West S, et al. Detection of mutations in mismatch repair genes in Portuguese families with hereditary nonpolyposis colorectal cancer (HNPCC) by a multimethod approach. Eur J Hum Genet 2000;8:49–53.

APPENDIX 1

Internet Sites

The following Internet sites may be useful to investigators wishing to pursue further study of the above issues.

Database of Gene Variants and Summary of Mutation Analysis Studies (supplementary tables)

Human Genome Epidemiology Network	http://www.cdc.gov/genomics/hugenet/default.htm

Colorectal Cancer Statistics

International Agency for Research on Cancer	http://www-dep.iarc.fr/eucan/eucan.htm
Surveillance, Epidemiology, and End Results Program	http://seer.cancer.gov/

Genetic Information and Databases

National Centre for Biotechnology Information	http://www.ncbi.nlm.nih.gov/
Online Mendelian Inheritance in Man	http://www.ncbi.nlm.nih.gov/omim
ICG-HNPCC* database	http://www.nfdht.nl/
Cambridge Public Health Genetics Unit	http://www.medinfo.cam.ac.uk/phgu/

Patient Education and Support

World Cancer Research Fund	http://www.wcrf.org/
Genetic Health	http://www.genetichealth.com/
Medicine Online	http://www.meds.com/colon/colon.html
American Cancer Society	http://www.cancer.org/
Cancer Research Campaign	http://www.crc.org.uk/
International Union Against Cancer	http://www.uicc.org/
UK National Screening Committee	http://www.nsc.nhs.uk/

* ICG-HNPCC, International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer.

(Appendix 2 follows)

APPENDIX 2

Clinical Criteria for Diagnosis of Hereditary Nonpolyposis Colorectal Cancer

Name of criteria	Specific criteria	Published reference
Amsterdam	Three relatives with colorectal cancer, one of which is a first-degree relative of the other two; colorectal cancer affecting more than one generation; at least one colorectal cancer case diagnosed before age 50 years	Vasen et al. (87)
Modified Amsterdam*	Two colorectal cancer cases in first-degree relatives in very small families that cannot be expended further; colorectal cancer affecting more than one generation; at least one colorectal cancer case diagnosed before age 55 years	Bellacosa et al. (88)
	Two first-degree relatives affected by colorectal cancer, plus a third relative with an unusually early-onset neoplasm or endometrial cancer	
Japanese†	Three or more colorectal cancer cases among first-degree relatives	Fujita et al. (89)
	Two or more colorectal cancers among first-degree relatives and any of the following: diagnosis before age 50 years; right colon involvement; synchronous or metachronous multiple colorectal cancers; association with extracolonic malignancy	
Bethesda*	Individuals from families that fulfill the Amsterdam criteria	Rodriguez-Bigas et al. (90)
	Individuals with two HNPCC‡-related cancers, including synchronous and metachronous colorectal cancers or associated extracolonic cancers	
	Individuals with colorectal cancer, plus colorectal cancer and/or HNPCC-related extracolonic cancer and/or colorectal adenoma in a first-degree relative; at least one of the cancers diagnosed before age 45 years and the adenoma diagnosed before age 40 years	
	Individuals with colorectal or endometrial cancer diagnosed before age 45 years	
	Individuals with right-sided colorectal cancer with an undifferentiated histopathologic pattern (solid/cribiform) diagnosed before age 45 years	
	Individuals with signet-ring cell type colorectal cancer diagnosed before age 45 years	
	Individuals with colorectal adenomas diagnosed before age 40 years	

* Fulfillment of all criteria listed in any paragraph in this section is sufficient.

† Cases can be classified as fulfilling either the first set of criteria or the second set and can be diagnosed with hereditary nonpolyposis colorectal cancer if they fulfill either set of criteria.

‡ HNPCC, hereditary nonpolyposis colorectal cancer.

Appendix A3 Database of Gene Variants in hMLH1 and hMSH2

All the gene variants listed below have been published in the ICG-HNPCC database {1596}, or in other relevant papers. It should be noted, however, that the various sources differ considerably in the way alterations are presented, and the amount of detail provided. The authors of this review have interpreted the available information for presentation in the format shown, and in some cases have inferred details not specifically given in the original publication. In several instances, variants have been published which are inconsistent with the consensus sequence of hMLH1/hMSH2, and could not easily be re-interpreted. In such cases, variants have been entered as given in the publication, with possible alternative interpretations given afterwards in Italics.

Spaces within the table indicate that relevant information was not available.

The 'cancer' column lists the types of cancer found in the affected kindred(s). The extent to which such information is provided varies considerably, and in many cases very little clinical information is provided. Information in this column should thus be interpreted with particular care.

In the ID column, identification codes separated by commas refer to different families with identical alterations, whereas identification codes separated by a forward slash refer to the same kindred, which has been included in more than one publication under different IDs.

Classification of mutations, as given in the mutation/polymorphism (M/P) column, has been performed by the authors of the review for the purpose of summarising the information contained in this table, and is based on the predicted consequences of each variant, incorporating the results of functional assays where available. These classifications do not necessarily reflect the interpretation given in the original publications.

(a) Germline hMLH1 Alterations

ID	EXON	Nucleotide Change	cDNA change	Cancer	Origin	Frequency	Reference	ICG	M/P
Healthy controls	5'UTR	g-a nt 1-93				0.5	(37)	Y	4
13	5'UTR	c-t nt 1-42			Germany		(67)	N	4
14	5'UTR	a-g nt 1-28			Germany		(67)	N	4
Case 8	5'UTR	c-t, nt 1 -21			USA		(87)	N	4
1	5'UTR	t-c, nt 1-21			Germany		(34)	N	4
2	5'UTR	a-g, nt 1 –21 (17?)			Germany		(34)	N	4
	5'UTR	g-a nt 1-17			Switzerland		(32)	N	4
Family 2	Ex 1	G-A nt 1	loss of translation start site	co	Australia		(90)	N	2
72	Ex 1	T-A nt 2	Met1Lys	co	Germany		(102)	Y	2
3	Ex 1	T-A nt 2	Met1Lys	co	Germany		(34)	N	2
FF10, F34	Ex 1	T-G nt 2	Met1Arg	co/ut	French		(98)	N	2
F2	Ex 1	C-A nt 5	Ser2STOP		Italy		Piepoli, unpub.	Y	1
NL-205, NLB-1069	Ex 1	17bp del nt 17	frameshift from codon 5	co/HNPCC	Netherlands		(105)	Y	1
	Ex 1	G-A nt 39	Glu13Glu		Scotland	0.014	(23)	N	4
2841/ 4F	Ex 1	del G nt 63	frameshift from codon 20	co	USA		(66)	Y	1
39603	Ex 1	G-T nt 69 (*67*)	STOP		UK		Norbury, unpub.	Y	1
110	Ex 1	del A nt 73	frameshift from codon 24	co	Germany		(102)	Y	1
4	Ex 1	del A nt 73	frameshift from codon 24	co	Germany		(34)	N	1
Patient 377	Ex 1	del A nt 73	frameshift from codon 24		Czech Republic		Krepelova, unpub.	Y	1
166	Ex 1	A-T nt 73	Ile25Phe	co/re	USA African		(101)	Y	2/3
40547	Ex 1	T-C nt 74	Ile25Thr		UK		Norbury, unpub.	Y	2/3
	Ex 1	G-T codon 26 (*C-T nt 76*)	Glu26STOP (*Gln26STOP*)	co	Portugal		(17)	N	1
	Ex 1	del A nt 84	frameshift – STOP codon 35		Asia		Deffenbaugh, unpub	Y	1
52	Ex 1	C-T nt 83	Pro28Leu		Germany		(102)	Y	2/3
5	Ex 1	C-T nt 83	Pro28Leu		Germany		(34)	N	2/3
NLB-100	Ex 1	del AG nt 101	frameshift from codon 33	co/HNPCC	Netherlands		(105)	Y	1
3	Ex 1	T-G nt 104	Met35Arg	co/en/br	Sweden		(93)	Y	2
h9	Ex 1	ins AA nt 105	frameshift from codon 34		England		Bunyan, unpub.	Y	1
30	Ex 1	ins AA nt 105	frameshift from codon 34		England		(16)	N	1
477	Ex 1	T-G nt 107	Ile36Ser	co/ut	Finland		(1)	N	2
Family 3	Ex 1	A-G nt 112	Asn38Asp	co	Australia		(90)	N	2

MD815	IVS 1	g-a nt 116+1	splice defect	co	Scotland		(23)	N	1
Family 1	IVS 1	g-a nt 116+1	splice defect	co	Australia		(90)	N	1
MD985	IVS 1	g-t nt 116+1	splice defect	co	Scotland		Farrington/Dunlop, unpub.	Y	1
2	Ex 2	C-T nt 131	Ser44Phe	co/en/br/ga	Sweden		(10)	Y	2
15	Ex 2	C-T nt 131	Ser44Phe	co	Sweden		(93)	Y	2
6	Ex 2	ins G nt 151	frameshift from codon 51		Germany		(34)	N	1
Case 1 – BR	Ex 2	C-T nt184	Gln62STOP	co/en	USA		(52)	Y	1
91	Ex 2	C-T nt184	Gln62STOP		Germany		(102)	Y	1
127	Ex 2	C-T nt184	Gln62STOP	co	Germany		(47)	Y	1
234	Ex 2	C-T nt184	Gln62STOP	co	Germany		(47)	N	1
7	Ex 2	C-T nt184	Gln62STOP	co	Germany		(34)	N	1
N498	Ex 2	C-A nt184	Gln62Lys	co/HNPCC	Norway	0/100	(105)	Y	2
8	Ex 2	del AA nt 190	frameshift from codon 64		Germany		(34)	N	1
N2104	Ex 2	A-G nt 191	Asn64Ser	co/HNPCC	Norway	0/100	(105)	Y	2/3
MD1446	Ex 2	A-G nt 191	Asn64Ser	co	Scotland		Farrington, unpub.	N	2
	Ex 2	ins C nt 198	frameshift – STOP codon 78		Asia		Deffenbaugh, unpub.	Y	1
7	Ex 2	C-T nt 198	Thr66Thr		Germany	0.01	(102)	Y	4
Case 9	Ex 2	C-T nt 198	Thr66Thr		USA		(87)	N	4
MFI	Ex 2	G-A nt 199	Gly67Arg		USA		(33)	Y	2
JPN-28	Ex 2	G-A nt 199	Gly67Arg	co/HNPCC	Japan		(89)	Y	2
VS015	Ex 2	G-A nt 199	Gly67Arg		Switzerland		(35)	Y	2
1652	Ex 2	G-A nt 199	Gly67Arg		Switzerland		(32)	Y	2
OG/98-2989	Ex 2	G-A nt 199	Gly67Arg	co	Slovenia		(83)	N	2
69	Ex 2	G-C nt 199	Gly67Arg	co/en	Sweden		(93)	Y	2
Patient 304	Ex 2	G-C nt 199	Gly67Arg		Czech Republic		Krepelova, unpub.	Y	2
MD1759	Ex 2	G-C nt 199	Gly67Arg	co	Scotland		Farrington/Dunlop, unpub.	N	2
FF52	Ex 2	GGG-TGG	Gly67Trp	co/b/ly/leu/Rh	France		(99)	N	2
18	Ex 2	T-A nt 203	Ile68Asn	co/en/pa/ga	Sweden		(93)	Y	2
10	IVS 2	A-G nt 208-2	in frame del Ex 3 codon 70102	co/ga	Sweden		(93)	Y	1
Family 4	IVS 2	del 4bp, +1,+2,208,209	loss of intron/exon junction	co	Australia		(90)	N	1
	Ex 3	T-C or A nt 228	Cys77Arg	co/HNPCC	Italy		(73)	N	2
YKY	Ex 3	G-A nt 229	Cys77Tyr	co	England		(94)	N	2
IMS2	Ex 3	G-A nt 229	Cys77Tyr		Australia		Bennet, unpub.	Y	2
9	Ex 3	T-G nt 238	Phe80Val		Germany		(34)	N	3/4

1	Ex 3	T-G nt 238	Phe80Val		Germany		(67)	N	3/4
HK	Ex 3	1 bp del	frameshift codon 82	co/HNPCC	England		(5)	N	1
DF5	Ex 3	C-T nt 245	Thr82Ile	co/HNPCC	USA		(92)	N	2/3
10	Ex 3	A-G nt 250	Lys84Glu	co	Germany		(34)	N	2
171	Ex 3	A-G nt 250	Lys84Glu	co	Germany		(47)	Y	2
	Ex 3	A-G nt 277	Ser93Gly	co			(81)	Y	2/3
FF44, FF75	Ex 3	C-T nt 298	Arg100STOP	co/b/ut/urt	France		(99)	N	1
Case 1	Ex 3	C-T nt 298	Arg100STOP	co/sk	USA		(87)	N	1
61	Ex 3	G-A nt 304	Glu102Lys		USA		de la Chapelle, unpub.	Y	2
R-RM2	Ex 3	Ex 3 deleted	deletion Ex 3 codon 70102	co/ut/k	Italy		(95)	N	1
115	Ex 35	Ex 3-5 deleted	deletion Ex 35 codon 70151	co/en	Finland		(1)	N	1
9	IVS 3	g-a nt 306+1	inframe deletion exon 3	co/ut/lv	Germany/Czech Republic		(79)	N	1
11	IVS 3	g-a, 306+1	inframe deletion exon 3	co/ut/lv	Germany		(34)	N	1
H16	IVS 3	g-a nt 306+1	inframe deletion exon 3	co/ut	Iran		(80)	N	1
Case 10	IVS 3	c-g nt 306+15		co	USA		(87)	N	4
	IVS 3	c-a nt 307-29			Sweden	0.0038	(56)	Y	3/4
12	IVS 3	c-a nt 307-29			Germany		(34)	N	3/4
Case 11	IVS 3	c-a nt 307-29		co	USA		(87)	N	3/4
28,51,67,477	Ex 4	T-G nt 320	Ile107Arg	co/en/HNPCC	Finland		(72)	Y	2
614,661,811	Ex 4	T-G nt 320	Ile107Arg	co	Finland		(86)	N	2
355	Ex 4	C-T nt 332	Arg111Val	co	Japan		(70)	N	2
	Ex 4	C-G nt 341 (*350*)	Thr117Arg	co/HNPCC	Switzerland		(32)	N	2
0434*	Ex 4	C-G nt 350	Thr117Arg	co/b/br/ov	Switzerland		(12)	Y	2
LG	Ex 4	C-T nt 350	Thr117Met	co	USA		(53)	Y	2
5	Ex 4	C-T nt 350	Thr117Met	co/en/sa	Moldavia		(60)	Y	2
SB-2575	Ex 4	C-T nt 350	Thr117Met	co	Poland		StachowKurzawski, unpub.	Y	2
h83	Ex 4	C-T nt 350	Thr117Met		England		Bunyan, unpub.	Y	2
BES-SG624	Ex 4	C-T nt 350	Thr117Met		France		(22)	Y	2
84	Ex 4	C-T nt 350	Thr117Met		USA		de la Chapelle, unpub.	Y	2
26	Ex 4	C-T nt 350	Thr117Met		Germany		(82)	Y	2
Patient 431	Ex 4	C-T nt 350	Thr117Met		Czech Republic		Krepelova, unpub.	Y	2
Family 5	Ex 4	C-T nt 350	Thr117Met	co	Australia		(90)	N	2
FF39	Ex 4	C-T nt 350	Thr117Met	co/urt/ga/se	France		(99)	N	2
	Ex 4	C-T nt 350	Thr117Met				Deffenbaugh, unpub.	Y	2

Case2	Ex 4	C-T	Missense	co	Poland		(21)	N	2
Case 1	Ex 4	ins AA nt 352	frameshift from codon 118	co	Poland		(21)	N	1
1 kindred	Ex 4	A-G nt 375	Ala125Ala	co	Scotland		(23)	N	4
870	Ex 4	C-G nt 378	Tyr126STOP	co	Finland		(86)	N	1
FF53	Ex 4	del C	frameshift from codon 126	co/leu	France		(99)	N	1
Family 6	Ex 4	G-A nt 381 (*380 or A-G nt 381, exon 5*)	loss of intron/exon junction	co	Australia		(90)	N	1
34	Ex 411	-	in frame del Ex 411	co	Sweden		(96)	Y	1
SK1	IVS 4	t-a nt 380+2	splice defect		Slovakia		Bartosova & Fridrichova, unpub.	Y	1
1044	IVS 4	a-g nt 381-2	splice defect	co/en/ov/re	France		(62)	Y	1
1500/5	IVS 4	a-g nt 381-2	splice defect		Belgium/ Switzerland		(32)	Y	1
338	Ex 5	G-C nt 382	Ala128Pro		Italy		(77)	Y	2/3
12*,11*,10*,17*	Ex 5	G-T nt 397	Gly133STOP	co/ut/ov	Japan		(68)	N	1
MD1092	Ex 5	ins 1 bp	frameshift from codon 143	co	Scotland		Farrington/Dunlop, unpub.	Y	1
02	IVS 5	g-t nt 453+1	splice defect	co	Lithuania		Avizienyte, unpub.	Y	1
	IVS 5	g-t nt 453+1	splice defect		Latin American / Caribbean		Deffenbaugh, unpub.	Y	1
CC8	IVS 5	t-c nt 453+2	splice defect		Spain		Caldes, unpub.	Y	1
Finnish Kindreds	IVS 5	t-c nt 454-51			Finland	0.04	(72)	Y	3/4
Patient 478	IVS 5	t-c nt 454-51			Czech Republic		Krepelova , unpub.	Y	3/4
	IVS 5	a-g nt 454-16			Sweden		(96)	N	3/4
	IVS 5	a-g nt 454-13					Lindblom, unpub.	Y	3/4
102	IVS 5	g-a nt 454-1	out of frame del Ex 6 codon 152182		Sweden		(56)	Y	1
10 Kindreds	IVS 5	g-a nt 454-1	out of frame del Ex 6 codon 152182	co/HNPCC	Finland		(71)	Y	1
4 kindreds	IVS 5	g-a nt 454-1	out of frame del Ex 6 codon 152182	co	Finland		(86)	N	1
13	IVS 5	g-a nt 454-1	out of frame del Ex 6 codon 152182		Germany		(34)	N	1
274	IVS 5	g-a nt 454-1	out of frame del Ex 6 codon 152182	co	USA		(18)	N	1
62	IVS 5	g-t nt 454-1	out of frame del Ex 6 codon 152182, or splice defect	co/HNPCC	Finland		(1)	N	1

	Ex 6	C-T nt 479	Ala160Val				Deffenbaugh, unpub.	Y	3
33	Ex 6	ins A nt 504	frameshift from codon 168		USA		de la Chapelle, unpub.	Y	1
	Ex 6	ins AG nt 523	frameshift from codon 174	co			(81)	N	1
404	Ex 6	GG-AT	Glu177STOP	co/re/ut	USA African		(100)	N	1
FF7	Ex 6	A-G nt 544	Arg182Gly	co/b/liver	French		(98)	N	2
668	IVS 6	a-g nt 545+3	in frame del Ex 6 codon 152182		Italy		(77)	Y	1
16	IVS 6	a-g nt 546-2	out of frame del Ex 7 codon 183196	co/ga/k/sa/rp	Sweden		(93)	Y	1
163	IVS 6	a-g nt 546-2	out of frame del Ex 7 codon 183196	co/re/ut//ly	USA England		(101)	Y	1
DUV-SG135	IVS 6	a-g nt 546-2	out of frame del Ex 7 codon 183196	co/ov	France		(22)	Y	1
H17	IVS 6	a-g nt 546-2	out of frame del Ex 7 codon 183196	co/ov/br/ut/ga	Sweden		(80)	Y	1
OZ-1	Ex 7	T-G nt 554	Val185Gly	co/gb/bd/si/ov	Australia		(40)	Y	2
Family 7	Ex 7	T-G nt 554	Val185Gly	co	Australia		(90)	N	2
	Ex 7	T-G nt 554	Val185Gly		Europe		Deffenbaugh, unpub.	Y	2
	Ex 7	T-C nt 577	Ser193Pro	co	Italy		(73)	N	2
SMAL001	Ex 7	del A nt 583	frameshift from codon 195		Malta		Hutter, unpub.	Y	1
FF72	Ex 7	A-T nt 586	Lys196STOP	Co/ut/si/ga	France		(99)	N	1
Patient 494	IVS 7	g-a nt 588+5	splice defect?		Czech Republic		Krepelova, unpub.	Y	3
1 kindred	IVS 7	g-c nt 588+11			Scotland		(23)	N	4
207	IVS 7	a-g nt 589-2	out of frame del Ex 8 codon 197226	co/HNPCC	USA		(59)	Y	1
A-AV24	IVS 7	a-g nt 589-2	out of frame del Ex 8 codon 197226		Italy		(95)	Y	1
DF1751	IVS 7	a-g nt 589-2	out of frame del Ex 8 codon 197226	co/HNPCC	USA	1/180	(92)	Y	1
3003	Ex 8	G-C nt 595	Glu199Gln	co	Sweden		(96)	Y	2/3
24	Ex 8	G-A nt 637	Val213Met	co/HNPCC	Portugal		(24)	N	2/3
39356	Ex 8	G-A nt 637	Val213Met		Asia		Norbury, unpub.	Y	2/3
9	Ex 8	C-T nt 649	Arg217Cys	co	Japan		(64)	Y	2/3
SNU-H1006	Ex 8	C-T nt 649	Arg217Cys	co/HNPCC	Korea		(31)	Y	2/3
15*	Ex 8	C-T nt 649	Arg217Cys	co	Japan		(68)	N	2/3
GDLV-11#II-9	Ex 8	G-A nt 655	Val219Ile	co	Italy	0.06	(19)	N	3/4
78	Ex 8	A-G nt 655	Ile219Val	re	England		(94)	N	3/4
Patient 3	Ex 8	A-G nt 655	Ile219Val	co/es	England		(7)	N	3/4
12*	Ex 8	A-G nt 655	Ile219Val	co/ut	Japan		(68)	N	3/4
	Ex 8	A-G nt 655	Ile219Val		USA	0.13	(52)	Y	3/4

	Ex 8	A-G nt 655	Ile219Val		Italy	0.33	Piepoli, unpub.	Y	3/4
	Ex 8	A-G nt 655	Ile219Val		Sweden	0.34	(93)	Y	3/4
HNL-1,HNL-2	Ex 8	A-G nt 655	Ile219Val		Japan		(89)	N	3/4
	Ex 8	A-G nt 655	Ile219Val		Scotland	0.327	(23)	N	3/4
4 kindreds	Ex 8	A-G nt 655	Ile219Val		Switzerland		(32)	N	3/4
Family 9	Ex 8	A-G nt 655	Ile219Val		Australia		(90)	N	3/4
16	Ex 8	A-G nt 655	Ile219Val		Germany		(34)	N	3/4
	Ex 8	A-G nt 655	Ile219Val		USA	0.396	(18)	N	3/4
92	Ex 8	A-G nt 655	Ile219Val	co	Japan	0.038	(70)	N	3/4
0434*, 1033*, 1515*	Ex 8	A-G nt 655	Ile219Val	co/HNPCC	Switzerland		(12)	N	3/4
F90	Ex 8	A-C nt 655	Ile219Val		Italy		(65)	N	3/4
	Ex 8	A-C nt 655	Ile219Leu		USA/Ger	0.31	(66)	Y	3/4
	Ex 8	A-C nt 655	Ile219Leu		Finland		(52)	N	3/4
4082	Ex 8	del A nt 663	frameshift from codon 222		Uruguay		(88)	Y	1
FF37	Ex 8	del AGTC	frameshift from codon 225226	co/ut/ov	France		(99)	N	1
Family 8	Ex 8	del G nt 673	fameshift, STOP	co	Australia		(90)	N	1
88 or 12A	Ex 8	C-T nt 676	Arg226STOP	co	Germany		(66)	Y	1
12	Ex 8	C-T nt 676	Arg226STOP	co	Germany		(102)	Y	1
14,15	Ex 8	C-T nt 676	Arg226STOP	co	Germany		(34)	N	1
331	Ex 8	C-T nt 676	Arg226STOP	co	Japan		(70)	N	1
DF2842	Ex 8	C-T nt 676	Arg226STOP	co/HNPCC	USA		(92)	N	1
6	Ex 8	G-T nt 677	splice defect	co/st	Moldavia		(60)	Y	1
NL-20	Ex 8	G-A nt 677	Arg226Gln – splice defect	co/re/en/urt/ ov/b	Netherlands		(104)	Y	1
DF168	Ex 8	G-A nt 677	Arg226Gln – splice defect	co/HNPCC	USA		(92)	N	1
	Ex 8	G-A nt 677	Arg226Gln – splice defect		Europe		Deffenbaugh, unpub.	Y	1
	IVS 8	g-a nt 677+1	splice defect		Latin American / Caribbean		Deffenbaugh, unpub.	Y	1
NL-27	IVS 8	del g nt 677+1	splice defect	co/en/k	Netherlands		(104)	Y	1
NLB-296	IVS 8	ins t nt 677+5	splice defect	co/HNPCC	Netherlands		(105)	Y	1/2
Cases & Controls	IVS 8	a-g nt 677+39			USA		(87)	N	4
MD533	IVS 8	del ta nt 678-3	splice defect	co	Scotland		(23)	Y	1
77	IVS 8	del ta nt 678-3	in frame del Ex 910 codon 227295	co	USA		(53)	Y	1
F102	IVS 8	g-a	out of frame deletion exon 8	co	Italy		(65)	N	1
H38	IVS 8	g-t, splice acceptor site	out of frame deletion exon 9	co	Sweden		(80)	Y	1

SNU-H1010	Ex 9	del T nt 692	frameshift from codon 230	co/HNPCC	Korea		(31)	Y	1
GDLG-20#II1*	Ex 9	G-A nt 702	Glu234Glu	co	Italy	0/100	(19)	N	4
Control	Ex 9	G-A nt 702	Glu234Glu		USA		(87)	N	4
311	Ex 9	G-A nt 731	Gly244Asp	co	Italy		(77)	Y	2
H6	Ex 9	C-A nt 755	Ser252STOP	co	USA		(74)	Y	1
S1532, GE602	Ex 9	T-G nt 779	Leu260Arg	co	Italy		(65)	N	2
VS006	Ex 9	del ATC nt 784	in frame del Ile codon262	co	Portugal		(35)	Y	2
24*	Ex 9	C-T nt 790	splice defect	co	Hong Kong		(15)	N	1
GE042	IVS 9	g-c nt 790+1	in frame del Ex 910 codon 227295	co	Spain		(35)	Y	1
Patient 424	IVS 9	g-c nt 790+1	in frame del Ex 910 codon 227295		Czech Republic		Krepelova, unpub.	Y	1
21	IVS 9	g-a nt 790+1	splice defect	co	USA		(18)	N	1
19	IVS 9	g-a nt 790+1	splice defect		Germany		(34)	N	1
18	IVS 9	del g nt 790+1	probable splice defect		Germany		(34)	N	1
SNUH-H7	IVS 9	t-a nt 790+2	splice defect	co/pa	Korea		(30)	Y	1
JD	IVS 9	ins t nt 790+3	in frame del Ex 910 codon 227295	co	USA		(53)	Y	1
OZ-2	IVS 9	ins t nt 790+3	in frame del Ex 910 codon 227295	co/b/k	Australia		(40)	Y	1
67	IVS 9	a-g nt 790+4	splice defect	co	Germany		(102)	Y	2
17	IVS 9	a-g nt 790+4	splice defect	co	Germany		(34)	N	2
	IVS 9	a-g nt 790+10				0.01	(96)	Y	4
Patient 534	IVS 9	a-g nt 790+10			Czech Republic		Krepelova, unpub.	Y	4
1834	IVS 9	del ttagATCGT nt 791-4	frameshift from codon 264		Switzerland		Heinimann, unpub.	Y	1
Case 2	IVS 9	a-g nt 791-2	splice defect	co/ov/es/sp	USA		(87)	N	1/2
NL-6	IVS 9	g-c nt 791-1	splice defect	co/re	Netherlands		(104)	Y	1
DF357	IVS 9	g-t nt791-1	splice defect	co	USA		(92)	N	1
Family10	Ex 10	A-G nt 790 (*791*)	splice defect	co	Australia		(90)	N	1
33633	Ex 10	A-G nt 791	splice site mutation – unclear pathogenicity		UK		Norbury, unpub.	Y	1
	Ex 10	del ATCG nt 791	frameshift from codon 264	co			(76)	N	1
143	Ex 10	del ATCG nt 791	frameshift from codon 264	co	Germany		(47)		1
20	Ex 10	del ATCG nt 791	frameshift from codon 264	co	Germany		(34)	N	1
198	Ex 10	C-T nt 793	Arg265Cys	co	Sweden		(96)	Y	2/3
198	Ex 10	C-T nt 793	Arg265Cys		Portugal		Hutter, unpub.	Y	2/3
MD1535, MD1654	Ex 10	C-T nt 793	Arg265Cys	co	Scotland		Farrington/Dunlop, unpub.	N	2/3
A-PD1*	Ex 10	G-A nt 794	Arg265His	co/bl	Italy	0/196	(95)	Y	2/3
80	Ex 10	A-G nt 803	Glu268Gly	co	Sweden		(56)	Y	2

NL-30	Ex 10	C-G nt 806	Ser269STOP	co/HNPCC	Netherlands		(105)	Y	1
1805	Ex 10	del TCCTT nt 811-815	frameshift from codon 271		Switzerland		Heinimann, unpub.	Y	1
	Ex 10	C-T nt 842	Ala281Val				Deffenbaugh, unpub.	Y	3
Patient 519	Ex 10	del A nt 856	frameshift from codon 286		Czech Republic		Krepelova, unpub.	Y	1
4	Ex 10	A-C	Lys286Gln	co/HNPCC	England		(6)	Y	2
MD1435	Ex 10	del A nt 861	frameshift from codon 287	co	Scotland		Farrington/Dunlop, unpub.	N	1
NL-37	Ex 10	A-G nt 883	Ser295Thr splice defect?	co/HNPCC	Netherlands		(105)	Y	1
TF2	Ex 10	A-G nt 883	Ser295Thr splice defect?	co	France		(98)	N	1
Control	IVS 10	t-a nt 885-24			USA		(87)	N	3/4
SNU-H19	Ex 11	ins T nt 888	frameshift from codon 296		Korea		Park, unpub.	Y	1
FF2	Ex 11	G-T nt 889	Glu297STOP	co	France		(98)	N	1
A-AV2	Ex 11	C-T nt 901	Gln301STOP	Co/k/br	Italy		(95)	Y	1
SIF1	Ex 11	A-T nt 912	Asp304Val	co	Italy		(65)	N	2
NL-24	Ex 11	ins GC nt 923	frameshift from codon 307	co/en	Netherlands		(104)	Y	1
Case 3	Ex 11	ins A nt 938	frameshift from codon 313	co	USA		(87)	N	1
	Ex 11	del C nt 954	frameshift from codon 318	co	Italy		(73)	N	1
GDLG-31#III-11	Ex 11	del C nt 954	frameshift from codon 318	co	Italy		(19)	N	1
1515*	Ex 11	T-C nt 977	Val326Ala	co	Switzerland		(12)	Y	3/4
3273	Ex 11	T-C nt 977	Val326Ala	co/HNPCC	USA		(53)	Y	3/4
DAC	Ex 11	T-C nt 977	Val326Ala		France		(22)	N	3/4
CFS70,CFS136	Ex 11	T-C nt 977	Val326Ala	co	Italy		(25)	N	3/4
A-AV18	Ex 11	T-C nt 977	Val326Ala	co/br/ov/ut/re/ga	Italy		(29)	N	3/4
GDLG-26#II-4	Ex 11	T-C nt 977	Val326Ala	co	Italy		(19)	N	3/4
1515*	Ex 11	T-C nt 977	Val326Ala	co/HNPCC	Switzerland		(12)	N	3/4
	Ex 11	A-C nt 986	His329Pro	co	France		(98)	N	2/3
96	Ex 11	A-C nt 986	His329Pro	co	Germany		(47)	Y	2/3
2	Ex 11	A-C nt 986	His329Pro		Germany		(67)	N	2/3
21	Ex 11	A-C nt 986	His329Pro	co	Germany		(34)	N	2/3
RIE-SG125	Ex 11	del A nt 994	frameshift from codon 331		France		(22)	Y	1
SNUH-H4	Ex 11	del G nt 1005	frameshift from codon 335	co/en	Korea		(30)	Y	1
A-PD1*	Ex 11	del C nt 1010	frameshift from codon 337	co/bl	Italy		(95)	Y	1
VS003	Ex 11	ins C nt 1012	frameshift from codon 338	co	Switzerland		(35)	Y	1
Patient 339	Ex 11	ins G nt 1027	frameshift from codon 343		Czech Republic		Krepelova, unpub.	Y	1
22	IVS 11	c-t nt 1038+51			Germany		(34)	N	4
15	IVS 11	c-t nt 1038+51			Germany		(67)	N	4
	IVS 11	(TA)n-(T)n length polymorphism				0.73	(44)	Y	4
	IVS 11	ins a nt1039-7	None apparent		Scotland/USA	0.007	(23)	Y	3/4

39	IVS 11	g-a nt 1039-1	out of frame del Ex 12 codon 347470		Western Europe		(76)	Y	1
F16259	Ex 12	8bp del nt 1068	frameshift from codon 356	co	Germany		(47)	Y	1
23	Ex 12	8bp del nt 1068	frameshift from codon 356	co	Germany		(34)	N	1
GDLV-52#II-2	Ex 12	A-G nt 1090	Thr364Ala		Italy		(19)	N	2/3
F33,F4	Ex 12	G-A nt 1132/del TC nt 1133	frameshift from codon 378		Italy		Piepoli, unpub.	Y	1
319	Ex 12	T-A nt 1151	Val384Asp	co	Japan	0.038	(70)	N	2/3
	Ex 12	T-A nt 1151	Val384Asp		Japan	0.03	(43)	N	2/3
Family 11	Ex 12	ins 4bp nt 1164	stop codon	co	Australia		(90)	N	1
Case 12	Ex 12	C-T nt 1165	Arg389Trp	co	USA		(87)	N	2
MD1529	Ex 12	G-A nt 1166	Arg389Gln	co	Scotland		Farrington/Dunlop, unpub.	N	2
	Ex 12	G-A nt 1166	Arg389Gln		Latin American / Caribbean		Deffenbaugh, unpub.	Y	2
DP6	Ex 12	del T nt 1190	frameshift	co/ut	Canada		(63)	N	1
MD442	Ex 12	del T nt 1211	frameshift from codon 404	co	Scotland		Farrington/Dunlop, unpub.	N	1
CFS111	Ex 12	G-A nt 1217	Ser406Asn	co	Italy		(25)	N	2/3
	Ex 12	G-A nt 1217	Ser406Asn		Finland	0.015	(109)	Y	2/3
209	Ex 12	G-A nt 1217	Ser406Asn	co	USA		(18)	N	2/3
Control	Ex 12	G-A nt 1217	Ser406Asn		USA		(87)	N	2/3
HNPC-036	Ex 12	ins GTCAGCC nt 1217	frameshift from codon 406	co	England		(4)	Y	1
2	Ex 12	G-A nt 1321	Ala441Thr	co/br	Sweden		(93)	Y	2/3
Patient 466	Ex 12	G-A nt 1321	Ala441Thr	co	Czech Republic		Krepelova, unpub.	Y	2/3
83	Ex 12	G-A nt 1321	Ala441Thr	co	USA		(18)	N	2/3
NET	Ex 12	-	frameshift codon 450	co/HNPCC	England		(5)	N	1
DF1846	Ex 12	A-T nt 1381	Lys461STOP	co/HNPCC	USA		(92)	N	1
PB	Ex 12	2 bp del	frameshift codon 461	co/HNPCC	England		(5)	N	1
RA	Ex 12	Ex 12 deleted	deletion Ex 12 codon 347470	co/HNPCC	USA		(74)	N	1
Q1	Ex 12	Ex 12 deleted	deletion Ex 12 codon 347470	co/HNPCC	French Canada		(4)	N	1
NL67	Ex 12	genomic del of ~25kb	unknown		Netherlands		Wijnen, unpub.	Y	1
8	IVS 12	g-c nt 1409+1	out of frame del Ex 12 codon 347470		Finland		Holmberg, unpub.	Y	1
MD949	IVS 12	g-t nt 1409+1	out of frame del Ex 12 codon 347470	co/ut	Scotland		Farrington/Dunlop, unpub.	N	1
25	IVS 12	g-a nt 1406*(1410)*-14			Germany		(34)	N	3/4

MD817	IVS 12	1.2kb intronic del	in frame del Ex 13 codon 470520	co	Scotland	(23)	Y	1
MS	Ex 13	del AAAG nt 1410	frameshift from codon 470	co/HNPCC	USA	(53)	Y	1
24	Ex 13 *(Ex12)*	del AGAA(G) nt 1411 (*or AGAA nt 1408)*	frameshift from codon 471		Germany	(34)	N	1
CC22	Ex 13	del C nt 1420	frameshift from codon 474		Spain	Caldes Trinidad, unpub.	Y	1
DES-SG407	Ex 13	G-A nt 1421	Arg474Gln		Japan	(58)	Y	2
C009	Ex 13	del A nt 1448	frameshift from codon 483	co	England	(26)	Y	1
	Ex 13	T-A nt 1456	Asp485Glu	co/HNPCC	Switzerland	(32)	N	2/3
h5	Ex 13	C-T nt 1459	Arg487STOP		England	Bunyan, unpub.	Y	1
MAR-SG363	Ex 13	C-T nt 1459	Arg487STOP		France	(22)	Y	1
CC29	Ex 13	C-T nt 1459	Arg487STOP		Spain	Caldes Trinidad, unpub.	Y	1
Patient 470	Ex 13	C-T nt 1459	Arg487STOP		Czech Republic	Krepelova, unpub.	Y	1
2,8	Ex 13	C-T nt 1459	Arg487STOP	co/HNPCC	Portugal	(24)	N	1
MD1269	Ex 13	C-T nt 1459	Arg487STOP	co	Scotland	Farrington/Dunlop et al., unpub.	N	1
Family 15	Ex 13	del TGAT nt 1460	stop codon	co	Australia	(90)	N	1
1S or A4T	Ex 13	G-A nt 1474	Ala492Thr	co	USA	(66)	Y	2
5A or 17383	Ex 13	ins C nt 1490	frameshift from codon 497	co/re	USA	(66)	Y	1
CRC11	Ex 13	ins C nt 1490	frameshift from codon 497	co	Japan	(89)	Y	1
10,11,12,14	Ex 13	ins C nt 1490	frameshift from codon 497	co/ut/ov/ga/re	Germany/Czech republic	(79)	N	1
144	Ex 13	ins C nt 1490	frameshift from codon 497	co	Germany	(47)	Y	1
CH1	Ex 13	ins C nt 1490	frameshift from codon 497	co/HNPCC	Czech	(105)	Y	1
Patient 469	Ex 13	ins C nt 1490	frameshift from codon 497		Czech Republic	Krepelova, unpub.	Y	1
26	Ex 13	ins C nt 1490	frameshift from codon 497		Germany	(34)	N	1
	Ex 13	del G nt 1490	frameshift from codon 497	co	Finland	(76)	N	1
241-1	Ex 13	del G nt 1490	frameshift from codon 497	co/re/ga	USA England	(101)	Y	1
3118	Ex 13	T-C nt 1517	Val506Ala	co/HNPCC	USA	(53)	Y	2/3
DF260*	Ex 13	T-C nt 1517	Val506Ala	co/HNPCC	USA	(92)	N	2/3
	Ex 13	T-C nt 1517	Val506Ala		Native American	Deffenbaugh, unpub.	Y	2/3
Family 14	Ex 13	G-T nt 1535	stop codon	co	Australia	(90)	N	1
GC7237	Ex 13	ins T nt 1553	Glu519STOP	co/cx/b/br/ov/bc/rp/l	England	(41)	Y	1
Family 13	Ex 13	ins T nt 1554	stop codon	co	Australia	(90)	N	1
645	Ex 13-16	22.4kb deletion (Alu mediated)	deletion Ex 13-16 codon 470632	co/en	France	(62)	Y	1
A-MD10	IVS 13	g-t nt 1558+1	splice defect		Italy	(95)	Y	1
	IVS 13	g-a nt 1558+10			Spain	Trinidad, unpub.	Y	4
	IVS 13	g-a nt 1558+11		co/sk/br/d/pa	Switzerland	(36)	Y	4

Family 12	IVS13	t-a nt 1558+13	cryptic splice site	co	Australia		(90)	N	2/3
	IVS 13	g-a nt 1558+14			Sweden	0.06	(93)	Y	4
	IVS 13	g-a nt 1558+14			Germany	0.08	(102)	N	4
	IVS 13	g-a nt 1558+14			USA	0.042	(18)	N	4
	IVS 13	g-a nt 1558+14			Scotland/USA	0.014	(23)	N	4
	IVS 13	g-a nt 1558+14			USA		(87)	N	4
	IVS 13	g-a nt 1558+14			Italy		(78)	N	4
27	IVS 13	g-a nt 1558+14	NA		Germany		(34)	N	4
85	IVS 13	a-c nt 1559-2	out of frame del Ex 14 codon 20556		Finland		Holmberg, unpub.	Y	1
40	IVS 13	g-t nt 1559-1	out of frame del Ex 14 codon 520556	co/HNPCC	Finland		(72)	Y	1
39992	IVS 13	a-g nt 1580-2			UK		Norbury, unpub.	Y	1
	Ex 14	del GT nt 1572	frameshift – STOP codon 526				Deffenbaugh, unpub.	Y	1
Case 13	Ex 14	C-T nt 1590	Phe530Phe	co	USA		(87)	N	4
113	Ex 14	del C nt 1622	frameshift from codon 540	co	Germany		(47)	Y	1
28	Ex 14	del C nt 1622	frameshift from codon 540	co	Germany		(34)	N	1
SNUH-H2	Ex 14	A-T nt 1625	Gln542Leu	co/HNPCC	Korea		(30)	Y	2
5	Ex 14	T-A nt 1640	Leu547STOP	co	Germany		(47)	Y	1
29	Ex 14	T-A nt 1640	Leu547STOP	co	Germany		(34)	N	1
I-771	Ex 14	ins TTATA nt 1644	frameshift from codon 548		Italy		Radice and Bertario, unpub.	Y	1
SNU-H14	Ex 14	T-C nt 1646	Leu549Pro	co/HNPCC	Korea		(31)	Y	2/3
TF1	Ex 14	A-C nt 1652	Asp551Thr	co	French		(98)	N	2
VS012	Ex 14	A-C nt 1652	Asp551Thr		Switzerland		(35)	Y	2
30	Ex 14	C-T nt 1652*(3)*	Asp551AAT?		Germany		(34)	N	?
88	Ex 14 – 15	Ex 14-15	out of frame del Ex 14-15	co	Sweden		(96)	Y	1
12,18,21,31,41	IVS 14	4bp ins/ 3bp del nt 1667+2	silencing of allele	co/ut/am	Denmark		(38)	Y	1
5	IVS 14	3'splice site+7	aberrant transcript	co	England		(6)	N	2/3
	IVS 14	a-g nt 1668-19				0.38	Miller, unpub.	Y	4
	IVS 14	a-g nt 1668-19			Germany	0.5	(102)	Y	4
0434*,1033*, 1515*,1553*	IVS 14	a-g nt 1668-19			Switzerland		(12)	Y	4
	IVS 14	a-g nt 1668-19			Sweden	0.4	(93)	Y	4
	IVS 14	a-g nt 1668-19			Sweden	0.4	(52)	Y	4
CC-7	IVS 14	a-g nt 1668-19			Spain		Caldes Trinidad, unpub.	Y	4
	IVS 14	a-g nt 1668-19			Scotland/USA	0.063	(23)	N	4
	IVS 14	a-g nt 1668-19			USA	0.56	(18)	N	4

cases & controls	IVS 14	a-g nt 1668-19			USA	(87)	N	4
MAL-SG164	IVS 14	c-a nt 1668-3	deletion Ex 15 codon 556577		France	(22)	Y	1
FF48, FF70	Ex 15	G-T nt 1672	Glu558STOP	co	France	(99)	N	1
OG/97-4266	Ex 15	C-T nt 1684	Gln562STOP	co	Slovenia	(83)	N	1
1808	Ex 15	del CTCA nt 1690-1693	frameshift from codon 564		Switzerland	Heinimann, unpub.	Y	1
VD001	Ex 15	A-T nt 1693	Ile565Phe	co	Switzerland	(35)	Y	3
SNUH-H1	Ex 15	T-C nt 1721	Leu574Pro	co/ut/sk/re	Korea	(30)	Y	2/3
OZ-4	Ex 15	G-A nt 1731	out of frame del Ex 15 codon 556577	co/en/ov	Australia	(40)	Y	1
1801	Ex 15	G-A nt 1731	out of frame del Ex 15 codon 556577		Switzerland	Heinimann, unpub.	Y	1
13	Ex 15	G-A nt 1731	out of frame del Ex 15 codon 556577	co	Germany/Czech Republic	(79)	N	1
Family 16	Ex 15	G-A nt 1731	out of frame del Ex 15 codon 556577	co	Australia	(90)	N	1
EB	Ex 15	Ex 15 deleted	out of frame del Ex 15 codon 556577	co/HNPCC	USA	(53)	N	1
NL-25	IVS 15	g-a nt 1731+1	splice defect	co/en	Netherlands	(104)	Y	1
F197	IVS 15	g-a	out of frame del exon 15	co	Italy	(65)	N	1
D-105	IVS 15	a-t nt 1732-2	in frame del Ex 16 codon 578632	co/HNPCC	Denmark	(104)	Y	1
3, 5	IVS 15	a-t nt 1732-2	in frame del Ex 16 codon 578632	co/ut/bl/si	Denmark	(38)	Y	1
21	Ex 16	A-G nt 1733	Glu578Gly	co/pa	Sweden	(93)	Y	2
ENDO003	Ex 16	A-G nt 1733	Glu578Gly		USA	de la Chapelle, unpub.	Y	2
JPN-1	Ex 16	C-G nt 1744	Leu582Val	co	Japan	(30)	Y	2/3
31	Ex 16	T-C nt 1745	Leu582Pro		Germany	(34)	N	2/3
01	Ex 16	del TT nt 1747	frameshift from codon 583		Lithuania	Avizienyte – unpub	Y	1
4 kindreds	Ex 16	ins C nt 1756	frameshift from codon 586	co/pa/en/ov	Korea	(30, 31, 110)	Y	1
	Ex 16	ins C nt 1758	frameshift from codon 586 – STOP codon at 592		Asia	Deffenbaugh, unpub.	Y	1
	Ex 16	del T nt 1764	frameshift from codon 588	co	Korea	(110)	N	1
63	Ex 16	del T nt 1769	frameshift from codon 590	co/br	Sweden	(56)	Y	1
316 or 3F	Ex 16	del TAGA nt 1769	frameshift from codon 590	co	USA	(66)	Y	1
15980 or 13A	Ex 16	del TAGA nt 1769	frameshift from codon 590	co/en	Germany	(66)	Y	1
NL-28	Ex 16	del AG nt 1779	frameshift from codon 593	co/en	Netherlands	(104)	Y	1
Patient 531	Ex 16	del AG nt 1783-84	frameshift from codon 595		Czech Republic	Krepelova, unpub.	Y	1
CFS25	Ex 16	del AG	frameshift from codon 595	co	Italy	(25)	N	1
A-AV14	Ex 16	del AG	frameshift from codon 595	co/sk/ga/l	Italy	(29)	N	1

32	Ex 16	del TAGA nt 1790*(1769?)*	frameshift?		Germany	(34)	N	1
33	Ex 16	C-G nt 1802*(1807)?*	*Pro603Arg?*		Germany	(34)	N	?
3	Ex 16	C-G nt 1807?	*Pro603Arg?*		Germany	(67)	N	?
DF2722	Ex 16	A-T nt 1810	Lys604STOP	co	USA	(92)	N	1
MD770	Ex 16	G-C nt 1818	Gly606Ala		Scotland	Farrington/Dunlop, unpub.	N	3/4
14	Ex 16	T-A nt 1820	Leu607His	co/HNPCC	Portugal	(24)	N	2
h78	Ex 16	ins T nt 1822	frameshift from codon 607		England	Bunyan, unpub.	Y	1
MD329	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616	co	Scotland	(23)	Y	2
HNP-8	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616	co	Japan	(64)	Y	2
NL-29,NL-40,NL-59	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616	co/en/pa/st/re/HNPCC	Netherlands	(104)	Y	2
I-202	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616	co/en/urt	Italy	(104)	Y	2
42/43	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616	co	USA Germany	(101)	Y	2
R-RM3	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616	co/st/si	Italy	(95)	Y	2
GUE-SG618	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616		France	(22)	Y	2
1760	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616		Switzerland	Heinimann, unpub.	Y	2
JR or P4	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616	co/HNPCC	USA	(53)	Y	2
14	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616	co/TS	USA	(52)	Y	2
15980	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616		USA	(66)	Y	2
5 Kindreds	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616	co/bl/b/br/du/ut/si/sk/ga/m/urt/l/ov	France	(99)	N	2
DF336	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616	co	USA	(92)	N	2
GDLG-29#III-8	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616	co	Italy	(19)	N	2
32	Ex 16	del AAG nt 1846	in frame deletion Lys codon 616	co	England	(16)	N	2
34	Ex 16	del AG nt 1848*(7)*	frameshift from codon 616		Germany	(34)	N	1/2
Case 4	Ex 16	del AAG nt 1852-1854	in frame deletion lys codon 618	co	USA	(87)	N	2
GE111	Ex 16	A-T nt 1852	Lys618STOP	co	Switzerland	(35)	Y	1
MD696	Ex 16	AA-GC nt 1852	Lys618Ala	co	Scotland	(23)	Y	2/3
55-1	Ex 16	AA-GC nt 1852	Lys618Ala	co/re/rp/ly	USA/England	(100)	Y	2/3
756	Ex 16	AA-GC nt 1852	Lys618Ala	co/re	France	(62)	Y	2/3
16	Ex 16	AA-GC nt 1852	Lys618Ala		Netherlands	(93)	Y	2/3
CC12	Ex 16	AA-GC nt 1852	Lys618Ala		Spain	Caldes Trinidad, unpub.	Y	2/3
VD007	Ex 16	AA-GC nt 1852	Lys618Ala		Switzerland	Hutter, unpub.	Y	2/3
51**	Ex 16	AA-GC nt 1852	Lys618Ala		USA	de la Chapelle, unpub.	Y	2/3
Family 17	Ex 16	AA-GC nt 1852	Lys618Ala	co	Australia	(90)	N	2/3
28	Ex 16	AA-GC nt 1852	Lys618Ala	co/HNPCC	Portugal	(24)	N	2/3

DF260*	Ex 16	AA-GC nt 1852	Lys618Ala	co/HNPCC	USA	1/186	(92)	N	2/3
NLB-526	Ex 16	AA-GC nt 1852	Lys618Ala	co/HNPCC	Netherlands		(105)	Y	2/3
Case 14, control	Ex 16	AA-GC nt 1852	Lys618Ala		USA		(87)	N	2/3
	Ex 16	AA-GC nt 1852	Lys618Ala			0.99%	Deffenbaugh, unpub.	Y	2/3
4	Ex 16	AA-GC nt 1852	Lys618Ala		Germany		(67)	N	2/3
6	Ex 16	A-G nt 1852	Lys618Gln	co/HNPCC	England		(6)	N	2/3
18	Ex 16	A-G nt 1853	Lys618Arg	co	Germany/Czech Republic		(79)	N	2/3
	Ex 16	A-C nt 1853	Lys618Thr	co	Portugal		(17)	N	2/3
	Ex 16	A-C codon 618	Lys618Tyr (*Thr*)	co	Italy		(20)	N	2/3
F103	Ex 16	A-C nt 1853	Lys618Thr		Italy		(65)	N	2/3
US-6/4A	Ex 16	A-C nt 1853	Lys618Thr	co	USA		(30)	Y	2/3
35	Ex 16	AA-GC nt 1855/1856 *(1852?)*	*Lys618Ala?*		Germany		(34)	N	2/3
Case 5*	Ex 16	del T nt 1877	frameshift from codon 626	co/ut	USA		(87)	N	1
73	Ex 16	del TCTCTTT nt 1877	frameshift from codon 626	co/en/br	France		(62)	Y	1
EC	Ex 16	TTC.TCT-TCC.ACT	Phe.SerSer.Thr(626/627)	co/HNPCC	England		(5)	N	2
MTS-K14	Ex 16	del GGAAA nt 1884	frameshift from codon 628	MT/sa	Germany		(46)	Y	1
NL-3,NL-4,NL-204	Ex 16	G-A nt 1896	in frame del Ex 16 codon 578632	co/HNPCC	Netherlands		(104)	Y	1
	Ex 16	del G nt 1896	frameshift /splicing defect	co/HNPCC	Switzerland		(32)	N	1
NL-14	Ex 16	genomic del of 5.5kb			Netherlands		Wijnen, unpub.	Y	1
14 kindreds	Ex 16	3.5kb deletion	in frame del Ex 16 codon 578632	co/HNPCC	Finland		(71)	Y	1
20 kindreds	Ex 16	3.5kb deletion	in frame del Ex 16 codon 578632	co/HNPCC	Finland		(72)	N	1
9 Kindreds	Ex 16	3.5kb deletion	in frame del Ex 16 codon 578632	co/HNPCC	Finland		(1)	N	1
5 Kindreds	Ex 16	3.5kb deletion	in frame del Ex 16 codon 578632	co	Finland		(86)	N	1
7 Kindreds	Ex 16	3.5kb deletion	in frame del Ex 16 codon 578632	co/HNPCC	Finland		(74)	N	1
7,14,27,58,98, 131,142	Ex 16	3.5kb deletion	in frame del Ex 16 codon 578632		Sweden/Finland		(93)	Y	1
VS009	Ex 16	3.5kb deletion	in frame del Ex 16 codon 578632		Switzerland		(35)	Y	1
K-25/HNP85	Ex 16	Ex 16 deleted	deletion Ex 16 codon 578632	co/en	Japan		(2)	N	1
Case 2	IVS16	g-t in splice donor site	splice defect	co	Sweden		(69)	N	1
DF 397	IVS 16	g-a nt 1896+1		co/HNPCC	USA		(92)	N	1

1033*	IVS 16	del G nt 1896+1	frameshift /splicing defect	co	Switzerland		(12)	Y	1
1 kindred	Ex 17	G-A nt 1908	Leu636Leu		Scotland/USA		(23)	N	4
36	Ex 17	C-T nt 1943	Pro648Leu		Germany		(34)	N	2/3
5	Ex 17	C-T nt 1943	Pro648Leu		Germany		(67)	N	2/3
	Ex 17	T-G nt 1958	Leu653Arg		Europe		Deffenbaugh, unpub.	Y	2/3
37	Ex 17	G-T nt 1958 *(1959)*	Leu653Leu?		Germany		(34)	N	4
Family 18	Ex 17	G-T nt 1959	Leu653Leu	co	Australia		(90)	N	4
	Ex 17	G-T nt 1959	Leu653Leu		Sweden	0.0038	(52)	Y	4
1533*	Ex 17	G-T nt 1959	Leu653Leu	co/HNPCC	Switzerland		(12)		4
SKI	Ex 17	G-T nt 1959	Leu653Leu		Slovakia		Bartosova & Fridrichova, unpub.	Y	4
1 kindred	Ex 17	G-T nt 1959	Leu653Leu		Scotland/USA		(23)	N	4
CC-32	Ex 17	G-T nt 1959	Leu653Leu		Spain		Caldes Trinidad, unpub.	Y	4
36003	Ex 17	G-T nt 1959	Leu653Leu		UK		Norbury, unpub.	Y	4
	Ex 17	G-T nt 1959	Leu653Leu	co	USA	0.083	(18)	N	4
cases & controls	Ex 17	G-T nt 1959	Leu653Leu		USA		(87)	N	4
	Ex 17	G-T nt 1959	Leu653Leu		Italy		(78)	N	4
38469	Ex 17	C-T nt 1961	Pro654Leu		UK		Norbury, unpub.	Y	2/3
OG/97-3941	Ex 17	T-C nt 1964	Ile655Thr	co	Slovenia		(83)	N	2/3
38	Ex 17	T-C nt 1964	Ile655Thr		Germany		(34)	N	2/3
MD804	Ex 17	C-T nt 1975	Arg659STOP	co/ut	Scotland		(23)	Y	1
72,83	Ex 17	C-T nt 1975	Arg659STOP	co/en/HNPC	Finland		(72) & unpub.	Y	1
179	Ex 17	C-T nt 1975	Arg659STOP	co/ut	Finland		(1)	N	1/2
A262	Ex 17	C-T nt 1975	Arg659STOP	co	Finland		(57)		1
C219	Ex 17	C-T nt 1975	Arg659STOP	co	Finland		(109)	N	1
20198 or 5F	Ex 17	C-T nt 1975	Arg659STOP	co	USA		(66)	Y	1
OZ-3	Ex 17	del CG nt 1975	in frame del Ex 17 codon 633663	co/b/en/re	Australia		(40)	Y	1
Family 20	Ex 17	del GA nt 1975 (*1976*)	stop codon	co	Australia		(90)	N	1
7	Ex 17	G-C nt 1976	Arg659Pro	co/HNPCC	Finland		(72)	Y	2
NL-56	Ex 17	G-C nt 1976	Arg659Pro	co/HNPCC	Netherlands		Wijnen, unpub.	Y	2
51*	Ex 17	G-A nt 1976	Arg659Glu		USA		de la Chapelle, unpub.	Y	2
1	Ex 17	G-T nt 1976	Arg659Leu	co/HNPCC	Portugal		(24)	N	2
39	Ex 17	A-C nt 1985 *(1984)*	*Thr662Pro*		Germany		(34)	N	2/3
6	Ex 17	A-C nt 1984 (*?*)	*Thr662Pro*		Germany		(67)	N	2/3
CHI-SG277	Ex 17	A-G nt 1988	deletion Ex 17 codon 633663		France		(22)	Y	1
Family 19	Ex 17	G-A nt 1989	stop codon	co	Australia		(90)	N	1
I-337,I670	Ex 17	G-T nt 1989	splice defect		Italy		Radice and Bertario, unpub.	Y	1
FF66	Ex 17	G-T nt 1989	in frame del exon 17	co/br/bl/ga/o	France		(99)	N	1

				v/du					
SCA	Ex 17	-	del Ex 17		France		(22)	N	
156	IVS 17	g-t nt 1989+1	in frame del Ex 17 codon 633663	co	Germany		(47)	Y	1
40	IVS 17	g-t nt 1989+1	in frame del Ex 17 codon 633663	co	Germany		(34)	N	1
R-MD6	IVS 17	g-t nt 1989+5	in frame del Ex 17 codon 633663	co/ga/ut/br	Italy		(95)	Y	1
16	IVS 17	t-g nt1990-22			Germany		(67)	N	3/4
Patient 478	IVS 17	c-g nt 1990-3	splice defect?		Czech Republic		Krepelova, unpub.	Y	2
FF5	IVS 17	del nt 1990-16 --> -2	in frame del exon 18	co/si	France		(99)	N	1
SNU-YC32	Ex 18	C-T	Asp667Asp		Korea		(110)	N	4
Patient 386	Ex 18	G-T nt 2028	Leu676Leu		Czech Republic		Krepelova , unpub.	Y	4
C010	Ex 18	G-A nt 2041	Ala681Thr	co	England		(26)	Y	2/3
BT-1881/WH-821	Ex 18	G-A nt 2041	Ala681Thr		Poland		Kurzawski, unpub.	Y	2/3
Case 6	Ex 18	G-A nt 2041	Ala681Thr		Poland		(21)	Y	2/3
7	Ex 18	A-G nt 2066	Gln689Arg		Germany		(67)	N	2/3
SGE109	Ex 18	C-G nt 2093	Ser698STOP		Switzerland		Hutter, unpub.	Y	1
NL-9,NLB-35	IVS 18	g-a nt 2103+1	Splice defect	co/st/en/b/k/sk/HNPCC	Netherlands		(104, 105)	Y	1
1813	IVS 18	g-a nt 2103+1	splice defect		Switzerland		Heinimann, unpub.	Y	1
IRI-SG119	IVS 18	g-a nt 2103+3	splice defect		France		(22)	Y	2
1121*	IVS 18	a-t nt 2104-2	splice defect	co/b/ov/ur/re	Switzerland		(12)	Y	1
	Ex 19	A-T nt 2104	splice defect	co/HNPCC	Switzerland		(32)	N	1
DF1448	Ex 19	del AG nt 2104-2105	splice site loss	co/HNPCC	USA		(92)	N	1
	Ex 19	C-T nt 2118	Gly706Gly	co	Scotland	0.035	(23)	N	4
AM	Ex 19	G-A nt 2135	Trp712STOP	co/HNPCC	USA		(53)	Y	1
h44,h87	Ex 19	G-A nt 2135	Trp712STOP		England		Bunyan, unpub.	Y	1
29	Ex 19	G-A nt 2135	Trp712STOP	co	England		(16)	N	1
VS004,008,010	Ex 19	G-A nt 2141	Trp714STOP	co/ut/st/gl/br/sk	Switzerland		(35)	Y	1
	Ex 19	G-A nt 2141	Trp714STOP	co/HNPCC	Switzerland		(32)	N	1
C008	Ex 19	G-A nt 2141	Trp714STOP	co	England		(26)	Y	1
H50	Ex 19	G-A nt 2141	Trp714STOP	co	Sweden		(80)	Y	1
GE111*	Ex 19	G-A nt 2146	Val716Met	co	Switzerland		(35)	Y	2/3
CFS56,CFS144	Ex 19	G-A nt 2146	Val716Met	co	Italy		(25)	N	2/3
A-AV23	Ex 19	G-A nt 2146	Val716Met	co/ut	Italy		(29)		2/3
AUB	Ex 19	G-A nt 2146	Val716Met	co/l	France		(22)	N	2/3

DF1751	Ex 19	G-A nt 2146	Val716Met	co/HNPCC	USA		(92)	N	2/3
Family 22	Ex 19	G-A nt 2146	Val716Met	co	Australia		(90)	N	2/3
	Ex 19	G-A nt 2146	Val716Met		Netherlands	0/100	(108)	N	2/3
41	Ex 19	G-A nt 2146	Val716Met		Germany		(34)	N	2/3
8,9	Ex 19	G-A nt 2146	Val716Met		Germany		(67)	N	2/3
Case 15	Ex 19	G-A nt 2146	Val716Met	co	USA		(87)	N	2/3
CHO	Ex 19	C-T nt 2152	His718Tyr		France		(22)	N	2/3
Case 16	Ex 19	C-T nt 2152	His718Tyr	co	USA		(87)	N	2/3
	Ex 19	C-T nt 2152	His718Tyr		Scotland	0.021	(23)	N	2/3
3	Ex 19	T-A nt 2163	Tyr721STOP	co/HNPCC	Portugal		(24)	N	1
154	Ex 19	ins AAATAT nt 2167	in frame insertion codon 723	co	Germany		(47)	Y	1
42	Ex 19	ins AAATAT nt 2167	in frame insertion codon 723	co	Germany		(34)	N	1
24	Ex 19	del CA nt 2177	frameshift from codon 726	co/ph	France		(62)	Y	1
L7	Ex 19	del CACA nt 2177	frameshift from codon 726	co/HNPCC	USA		(74)	Y	1
	Ex 19	C-G nt 2185	Leu729Val		Scotland	0.021	(23)	N	2/3
HNP-004	Ex 19	ins AACA nt 2190	frameshift from codon 730	co	French Canadian		(4)	Y	1
1797	Ex 19	ins AAAC nt 2198	frameshift, codon 730 > truncation codon 733		USA		(85)	Y	1
DF2913	Ex 19	ins AAAC nt 2198	frameshift, codon 730 > truncation codon 733	co/HNPCC	USA		(92)	N	1
IMS1	Ex 19	del C nt 2224	frameshift from codon 742		Australia		(40)	Y	1
h112	Ex 19	C-A nt 2250	Tyr750STOP		England		Bunyan, unpub.	Y	1
DF171	Ex 19	C-G nt 2250	Tyr750STOP	co/HNPCC	USA		(92)	N	1
SNU-H1009	Ex 19	del AA nt 2251	frameshift from codon 751	co/HNPCC	Korea		(31)	Y	1
h18	Ex 19	del AA nt 2251	frameshift from codon 751		England		Bunyan, unpub.	Y	1
	Ex 19	A-G nt 2253	Lys751Arg		Scotland		Farrington/Dunlop, unpub.	N	3/4
Family 21	Ex 19	A-G nt 2253	Lys751Arg	co	Australia		(90)	N	3/4
17	Ex 19	ins AA nt 1853	frameshift	co	Germany/Czech Republic		(79)	N	1
FF74	Ex 19	A-T nt 2263	Arg755Trp	co/br	France		(99)	N	2
DF232	Ex 19	G-T nt 2265	Arg755TSer	co/HNPCC	USA		(92)	N	2
DG	Ex 19	ins TGTT nt 2268	frameshift from codon 756	co/HNPCC	USA		(53)	Y	1
L2516	Ex 19	ins TGTT nt 2270	frameshift from codon 756	co/HNPCC	USA		(74)	Y	1
A-AMD7	Ex 19	ins T nt 2270	STOP757Leu		Italy		(95)	Y	1
A-MD1	3' UTR	del ttc nt 2268+33			Italy	0.04	(95)	Y	4
166	3' UTR	del ttc nt 2268+33			USA		Weber unpub.	Y	4
GDLG-21#III2, GDLG-20#II1*	3' UTR	del ttc nt 2268+33		co	Italy	0.02	(19)	N	4
VS0005	3' UTR	del ttc nt 2268+33			Switzerland		Hutter, unpub.	Y	4

(b) Germline hMSH2 Alterations

ID	EXON	Nucleotide Change	cDNA change	Cancer	Origin	Frequency	Reference	ICG	M/P
MD1157	Ex 1	A-T nt 1	Met1Leu	co	Scotland		(23)	Y	2/3
344, 350	Ex 1	C-T nt 23	Thr8Met	co	Japan	0.01	(70)	N	2/3
Case 17, control	Ex 1	C-T nt 126	Phe42Phe		USA		(87)	N	4
DP5	Ex 1	del 29bp nt 134	frameshift codon 45	co/ut	Canada		(63)	N	1
33	Ex 1	C-G nt 138	His46Gln	co	Scotland		(11)	N	2
	Ex 1	del C nt 161	frameshift from codon 54 – STOP at codon 63				Deffenbaugh, unpub.	Y	1
8001	Ex 1	C-T nt 181	Gln61STOP		Uruguay		Sarroca et al., unpub.	Y	1
5	Ex 1	del G nt 201	frameshift from codon 67	co/en	Sweden		(97)	Y	1
38439	Ex 1	ins gg track ggggg nt 205	frameshift		UK		Norbury, unpub.	Y	1
NLB-50162	Ex 1	2.1kb genomic deletion	deletion Ex 1 codon 1-70	co/HNPCC	Netherlands		(106)	Y	1
621	Ex 1		deletion Ex 1 codon 1-70				(74)	N	1
NL-48, NL-53	Ex 1-6	2.5kb genomic deletion	deletion exon 1-6		Netherlands		(106)	Y	1
NLB-376	IVS 1	a-g nt 212-2	splicing defect	co/HNPCC	Netherlands		(105)	N	1
1	IVS 1	g-c nt 211+8			Germany		(34)	N	4
	IVS 1	c-g nt 211+9			Scotland	0.38	(11)	Y	4
	IVS 1	c-g nt 211+9			Scotland/USA	0.148	(23)	N	4
3 Kindreds	IVS 1	c-g nt 211+9			Switzerland		(32)	N	4
2	IVS 1	c-g nt 211+9			Germany		(34)	N	4
cases & controls	IVS 1	c-g nt 211+9			USA		(87)	N	4
	IVS 1	9th base pair			USA	0.12	(101)	N	4
	Ex 2	G-A nt 219	Lys73Lys		Scotland/USA	0.007	(23)	Y	4
352	Ex 2	del CT nt 223-224	frameshift	co	Japan		(70)	N	1
NLB-600	Ex 2	del AG nt 227	frameshift from codon 75	co/HNPCC	Netherlands		(105).	Y	1
1820	Ex 2	del TT nt 261-262	frameshift		Macedonia		Heinimann, unpub.	Y	1
I-334	Ex 2	del TT nt 278	frameshift from codon 93	co/HNPCC	Italy		(13)	Y	1
82	Ex 2	del TCT nt 279-281	frameshift from codon 93	co	Sweden		(96)	Y	1
	Ex 2	G-A nt 287	Arg96His		Netherlands	0.01	(107)	Y	3
130	Ex 2	ins 22bp nt 289	frameshift from codon 97	MT	Germany		(46)	Y	1
3	Ex 2	ins 22bp nt 289	frameshift from codon 97		Germany		(34)	N	1

4	Ex 2	del AGTTGA nt297?	codon 99/100 deleted?		Germany		(34)	N	2
523	Ex 2	G-C nt 319	Ala107Pro		Czech Republic		Krepelova, unpub.	Y	2/3
	Ex 2	A-G nt 329	Lys110Arg		USA	0.02	(101)	Y	3
	Ex 2	G-A nt 339	Lys113Lys		Czech Republic	0.0076	Krepelova, unpub.	Y	4
5	Ex 2	G-A nt 339	Lys113Lys		Germany		(34)	N	4
6	Ex 2	del A nt 344	frameshift from codon 115		Germany		(34)	N	1
NA-17/NA-86	Ex 2	5.4kb genomic deletion	deletion Ex 2 codon 71-122	co/HNPCC	Netherlands		(106)	Y	1
HNP-O-3	Ex 2-8	genomic deletion	frameshift del exons 2-8 codons 71-462	co/ut/ga/urt	UK		(4)	N	1
	IVS 2	del T from polyT tract			USA	0.05	(101)	Y	4
114-1-OL	Ex3	A-G nt 380	Asn127Ser		Nigeria		de la Chapelle, unpub.	Y	2/3
Case 5*,18,19	Ex3	A-G nt 380	Asn127Ser	co/ut	USA		(87)	N	2/3
MTS-K10(n)	Ex 3	del A nt 380	frameshift from codon 127	MT/pr	Germany		(46)	Y	1
1097He	Ex 3	del TC nt 387	frameshift from codon 129	co/TS	Switzerland		(12)	Y	1
	Ex 3	del nt 388-389	frameshift ->STOP codon 131	co/HNPCC	Switzerland		(32)	N	1
	Ex 3	C-T nt 399	Asp133Asp		African Am		Johnson, unpub.	Y	4
control	Ex 3	C-T nt 399	Asp133Asp		USA		(87)	N	4
HNP-1/ K-8/HNP10	Ex 3	del T nt 406	frameshift from codon 136	co/ut/urt/ov	Japan		(64)	Y	1
69	Ex 3	A-G nt 416	Asn139Ser	co	Scotland		(11)	N	2/3
338	Ex 3	T-G nt 435	Ile145Met		Czech Republic		Krepelova, unpub.	Y	3/4
	Ex 3	T-G nt 435	Ile145Met				Deffenbaugh, unpub.	Y	3/4
	Ex 3	G-A nt 446	Gly149Asp				Deffenbaugh, unpub.	Y	2/3
	Ex 3	C-T nt 459	Ser153Ser		USA		(66)	Y	4
175	Ex 3	C-A nt 471	Gly157Gly	co	Japan	0.028	(70)	N	4
	Ex 3	C-A nt 471	Gly157Gly	co	Japan	0.02	(43)	N	4
10	Ex 3	T-A nt 482	Val161Asp		Germany		(67)	N	1/2
4F	Ex 3	G-C nt 499	Asp167His	co	USA		(66)	N	2
control	Ex 3	G-C nt 499	Asp167His		USA		(87)	N	2
	Ex 3	A-G nt 505	Ile168Val		Asia		Deffenbaugh, unpub.	Y	3/4
K-79/HNP185	Ex 3	del ATAC codon 169-170	frameshift from codon 169		Japan		(3)	N	1
5	Ex 3	del G nt 513	frameshift from codon 171	co	Hong Kong		(15)	N	1
2	Ex 3	del T nt 518	frameshift from codon 173	co	Germany/Czech Republic		(79)	N	1
24/223	Ex 3	del T nt 518	frameshift from codon 173	co	Germany		(102)	Y	1
7	Ex 3	del T nt 518	frameshift from codon 173	co	Germany		(34)	N	1

8064	Ex 3	del AA nt 530	frameshift		Uruguay		Sarroca, unpub.	Y	1
4	Ex 3	C-T nt 547	Gln183STOP	co	Portugal		(24)	N	1
control	Ex 3	A-G nt 557	Asn186Ser		USA		(87)	N	3/4
HNPCC-8	Ex 3	del TGAGGCTCT nt 561	in frame del codon 187-190	co	Slovenia		(84)	Y	1/2
16	Ex 3	TC-CT nt 569/570	Leu190Pro	co	Germany/Czech Republic		(79)	N	2/3
8	Ex 3	CT-TC nt 569/570		co	Germany		(34)	N	2/3
	Ex 3	C-T nt573	Leu191Leu		Scotland/USA	0.021	(23)	N	4
Case 20	Ex 3	C-T nt 573	Leu191Leu		USA		(87)	N	4
199	Ex 3	A-G nt 593	Glu198Gly	co	Sweden		(96)	Y	2
D/19	Ex 3	T-C nt 595	Cys199Arg	gl/co/re	Hong Kong		(50)	Y	2
-	Ex 3	G-T nt 613	Glu205STOP	co	Poland		Kurzawski, unpub.	Y	1
HNP-055	Ex 3	C-T nt 643	Gln215STOP	co	France		(4)	Y	1
11	Ex 3(*4?*)	A-G nt 646	Ile216Val/possible splice defect		Germany		(67)	N	2/3
3 kindreds	Ex 3	2.2kb genomic deletion	Ex 3 deleted	co/HNPCC	Netherlands		(106)	Y	1
CO07	Ex 3		in frame del Ex 3 codon 123-215	co	England		(26)	N	1
HNP-048	Ex 3		in frame del Ex 3 codon 123-215	co/ga/ut/sk	UK		(4)	N	1
9	Ex 4	del GGTCC, ins TAAT nt 675?	frameshift?		Germany		(34)	N	1/2
HNP4/H114/K-43	Ex 4	ins A nt 688	frameshift from codon 228/230	co/ut/ov	Japan		(64)	Y	1
FF51	Ex 4		frameshift from codon 228	co/ov/urt	France		(99)	N	1
3	Ex 4	del TT nt 696	frameshift	co	Germany/Czech Republic		(79)	N	1
10	Ex 4	del TT nt 696	frameshift	co	Germany		(34)	N	1
DF268, DF951	Ex 4	del AA nt 704-705	frameshift --> Truncation, codon 235	co	USA		(92)	N	1
MD818	Ex 4	C-T nt 754	Gln252STOP	co/ut	Scotland		(51)	Y	1
MD968	Ex 4	C-T nt 754	Gln252STOP	co/ut	Scotland		Farrington/Dunlop, unpub.	N	1
3	Ex 4	G-T	STOP	co	Poland		(21)	N	1
FF17	Ex 4		in-frame del exon 4	co/ut	France		(99)	N	1
46	IVS 4	g-a nt 792+1	splice defect	co	USA		(18)	N	1
Case 21	IVS 4	a-g nt792+49			USA		(87)	N	4
MD1521	IVS 4	g-t nt 793-1	splice defect	co	Scotland		Farrington/Dunlop, unpub.	Y	1/2
	Ex 5	C-T nt 806	Ser 269Leu		Latin American / Caribbean		Deffenbaugh, unpub.	Y	2/3

MD825	Ex 5	del GTCT nt 808	in frame del Exon 5	co/ut/pa/st	Scotland		(51)	Y	1
DF597	Ex 5	C-T nt 815	Ala272Val	co	USA		(92)	N	3
Case 22	Ex 5	C-T nt 815	Ala272Val		USA		(87)	N	3
GALI	Ex 5	ins T nt 840	frameshift from codon 280	co/ut/bl/b	France		(48)	Y	1
BLA-SG101	Ex 5	ins T nt 840	frameshift from codon 280		France		(22)	Y	1
NL-10, NL-39	Ex 5	C-T nt 862	Gln288STOP	co/st/k/urt	Netherlands		(106)	Y	1
1	Ex 5	C-T nt 862	Gln288STOP	co/ut/ov/pa	Germany/Czech Republic		(79)	N	1
278(n)	Ex 5	C-T nt 862	Gln288STOP	MT/Hodgkins	Germany		(46)	Y	1
11	Ex 5	C-T nt 862	Gln288STOP	co	Germany		(34)	N	1
6	Ex 5	C-T codon 298 (?)		co	Sweden		(96)	Y	1
NL-38	Ex 5	G-A nt 913	Ala305Thr	co/HNPCC	Netherlands	0/100	(105)	Y	2/3
354	Ex 5	G-A nt 942	Splice mutation	co	Japan		(70)	N	1
	Ex 5	Ex 5 deleted	deletion Ex 5 cod 265-314	co			(75)	N	1
FF8	Ex 5	Ex 5 deleted	deletion Ex 5 cod 265-314	co/ut/leu	France		(98)	N	1
FF27	Ex 5	Ex 5 deleted	deletion Ex 5 cod 265-314	co/si/leu/br/l/chg	France		(99)	N	1
11,42	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co/ut/ov/urt/bo	Denmark		(38)	Y	1/2
119	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	Germany		(47)	Y	1/2
20681/2A	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co/HNPCC	USA		(66)	Y	1/2
3177	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co/HNPCC			(53)	Y	1/2
-	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	UK		(28)	Y	1/2
HNP7	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co/ut/b/sk	Japan		(64)	Y	1/2
667,349	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co/ut/pa/b	Italy		(77)	Y	1/2
C, MF, RB	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	Canada		(54)	Y	1/2
C017,C022	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	UK		(26)	Y	1/2
N-HS3, O-4, HNP-020, HNP-038	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co/ut/br/ov/bs/ga/urt/sk/th/l/bo	Norway, UK, Newfoundland		(4) & unpub.	Y	1/2
124	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co			unpub.	Y	1/2
TK1882	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	Poland		Kurzawski. unpub.	Y	1/2
K-39/HNP108	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	Japan		(58)	N	1/2
23	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	Hong Kong		(15)	N	1/2
	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co/HNPCC	Italy		(20)	N	1/2
SI505	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	Italy		(65)	N	1/2
42	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	Portugal		(24)	N	1/2
3 Kindreds	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	USA		(92)	N	1/2
10 Kindreds	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	Newfoundland		(27)	N	1/2

K-57/HNP143	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	Japan		(58)	N	1/2
GDLM-2#III-1	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	Italy		(19)	N	1/2
4	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co/re/lv	Germany/Czech Republic		(79)	N	1/2
3 kindreds	IVS 5	a-t nt 942+3	in-frame deletion Ex 5				Deffenbaugh, unpub.	Y	1/2
12	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	Germany		(34)	N	1/2
Case 6	IVS 5	a-t nt 942+3	in-frame deletion Ex 5	co	USA		(87)	N	1/2
	IVS 5	del 10bp from poly A tract			USA	0.032	(101)	Y	4
6	IVS 5	g-a nt 943-1	in-frame deletion Ex 6	co/ut/ga	Germany/Czech Republic		(79)	N	1/2
DF3002	Ex 6	G-T nt 945	Gly315Val	co	USA		(92)	N	2/3
SI524	Ex 6		Gly321Asp (*Gly322Asp*)	co	Italy		(65)	N	2/3
1	Ex 6	G-A nt 965	Gly322Asp	co/br	Russia		(60)	Y	2/3
JM	Ex 6	G-A nt 965	Gly322Asp	co	England		(94)	Y	2/3
Family 24	Ex 6	G-A nt 965	Gly322Asp	co	Australia		(90)	N	2/3
	Ex 6	G-A nt 965	Gly322Asp	co	Finland	.067	(86)	N	2/3
DF1448, DF1370	Ex 6	G-A nt 965	Gly322Asp	co	USA		(92)	N	2/3
HNPCC-1	Ex 6	G-A nt 965	Gly322Asp	co/xtracolon	England		(6)	N	2/3
CI-2556/SM-1820	Ex 6	G-A nt 965	Gly322Asp		Poland		Kurzawski, unpub.	Y	2/3
19	Ex 6	G-A nt 965	Gly322Asp	co	Portugal		(24)	N	2/3
9 Kindreds	Ex 6	G-A nt 965	Gly322Asp	co	Sweden		(55)	N	2/3
	Ex 6	G-A nt 965	Gly322Asp		Scotland / USA	0.02	(52)	Y	2/3
case 23,24, control	Ex 6	G-A nt 965	Gly322Asp		USA		(87)	N	2/3
	Ex 6	G-A nt 965	Gly322Asp		Norway	0.02	(8)	N	2/3
5 Kindreds	Ex 6	G-A nt 965	Gly322Asp		Czech Republic		Krepelova, unpub.	Y	2/3
	Ex 6	G-A nt 965	Gly322Asp		Spain		Caldes, unpub.	Y	2/3
	Ex 6	G-A nt 965	Gly322Asp		Germany	0.01	(102)	N	2/3
GDLM-9#II-2	Ex 6	G-A nt 965	Gly322Asp	co	Italy	0/100	(19)	N	2/3
13	Ex 6	G-A nt 965	Gly322Asp	co	Germany	0.01	(34)	N	2/3
186	Ex 6	CGC-GAC	Missense		Germany		(9)	Y	2/3
K-55/HNP137	Ex 6	C-T nt 968	Ser323Cys	co	Japan		(2)	N	2/3
331*	Ex 6	G-A nt 972	Gln324Gln	co	Japan	0.00	(70)	N	4
Case 18	Ex 6	G-C nt 982	Ala328Pro	co	USA		(87)	N	2/3
GDLM-8#II-1, GDLV-35#III-5	Ex 6	C-T nt 984	Ala328Ala	co	Italy	0/100	(19)	Y	4
	Ex 6	C-T nt 984	Ala328Ala		Sweden		(96)	Y	4
Family 25	Ex 6	C-T nt 984	Ala328Ala	co	Australia		(90)	N	4
Case 25,	Ex 6	C-T nt 984	Ala328Ala		USA		(87)	N	4

control									
114-1-OL*	Ex 6	T-C nt 997	Cys333Arg		Nigeria		de la Chapelle, unpub.	Y	2/3
IMS5	Ex 6	G-A nt 998	Cys333Tyr		Australia		Bennett, unpub.	Y	2/3
59	Ex 6	A-T nt 1000	Lys334STOP	co/urt/bl	USA - German		(101)	Y	1
	Ex 6	C-T nt 1006	Pro336Ser	co	Switzerland		(32)	N	2
Family 23	Ex 6	C-T nt 1009	Gln337STOP	co	Australia		(90)	N	1
	Ex 6	G-A nt 1012	Gly338Arg				Deffenbaugh et al., unpub	Y	2/3
NL-21	Ex 6	del AA nt 1016	frameshift from codon 339	co/pr/en/ov/urt	Netherlands		(107)	Y	1
GDLG-49#IV-2	Ex 6	G-A nt 1024	Val342Ile	co	Italy	0/100	(19)	N	3/4
h86	Ex 6	del G nt 1059	frameshift from codon 353		England		Bunyan, unpub.	Y	1
RH	Ex 6	173bp ins nt 1076-1092	frameshift from codon 359	co/HNPCC	USA		(53)	Y	1
	Ex 6	G-A, polyA region	-		USA	0.036	(101)	N	4
NL-33/NLB-50490	Ex 6	13kb genomic deletion	deletion Ex 6 codon 315-359	co/HNPCC	Netherlands		(106)	Y	
OZ-5	Ex 2-6	Ex 2-6 deleted	deletion Ex 2-6 codon 71-359	co/en/b	Australia		(40)	N	1
FF15	IVS 6	g-a nt 1076+1	Splice defect	co/bl/ut/ga	France		(99)	N	1
14	IVS 6	g-a nt 1076-35*(+35)*			Germany		(34)	N	4
	IVS 6	t-c nt 1077-10			Norway	0.05	(8)	Y	3/4
1 kindred	IVS 6	t-c nt 1077-10			Scotland		(23)	N	3/4
15	IVS 6	t-c nt 1076(*1077*)-10			Germany		(34)	N	3/4
cases & controls	IVS 6	t-c nt 1076–8			USA		(87)	N	3/4
30	Ex 7	ins A nt 1097	frameshift from codon 366, termination codon 388	co/en	Sweden		(56)	Y	1
H23	Ex 7	ins A nt 1097	frameshift from codon 366, termination codon 388	re/ut/ga	Sweden		(80)	N	1
2	Ex 7	ins TA nt 1127	frameshift from codon 376	co/en/br/sk	Russia		(60)	Y	1
NL-7	Ex 7	del T nt 1137	frameshift from codon 379	co/en/es/st	Netherlands		(107)	Y	1
He1587	Ex 7	C-T nt 1147	Arg383STOP	co/en	Switzerland		(12)	Y	1
6	Ex 7	C-T nt 1147	Arg383STOP	co/re/ut/br	Germany/Czech Republic		(79)	N	1
177	Ex 7	C-T nt 1147	Arg383STOP	co	Germany		(47)	N	1
	Ex 7	C-T nt 1147	Arg383STOP	co/HNPCC	Switzerland		(32)	N	1
	Ex 7	C-T nt 1165	Arg389STOP		French Canada		Foulkes, unpub.	Y	1
HNPCC-2	Ex 7	C-T nt 1165	Arg389STOP	co/xtracolon	England		(6)	N	1
DP7	Ex 7	C-T nt 1165	Arg389STOP	co/ ut	Canada		(63)	N	1
201	Ex 7	C-T nt 1165	Arg389STOP		Germany		(9)	Y	1
449	Ex 7	C-T nt 1168	Leu390Phe	co	USA - African		(100)	N	2/3

	Ex 7	C-T nt 1168	Leu390Phe		Japan	0.02	(43)	N	2/3
MD1466	Ex 7	A-T nt 1177	Lys393STOP	co	Scotland		Farrington/Dunlop, unpub.	N	1
C021	Ex 7	C-T nt 1189	Gln397STOP		UK		(26)	Y	1
CA135	Ex 7	ins CA	frameshift from codon 399	co	Italy		(65)	N	1
MD528	Ex 7	C-T nt 1216	Arg406STOP	co	Scotland		(51)	Y	1
8	Ex 7	C-T nt 1216	Arg406STOP	co/HNPCC	USA		(49)	Y	1
7	Ex 7	C-T nt 1216	Arg406STOP	co	Sweden		(96)	Y	1
24	Ex 7	C-T nt 1216	Arg406STOP	co/ut/ga/ph	Germany/Czech Republic		(79)	N	1
240-1	Ex 7	C-T nt 1216	Arg406STOP	co/re/ut	USA (Italian)		(101)	Y	1
IG-1296/4	Ex 7	C-T nt 1216	Arg406STOP		Poland		(21)	Y	1
17	Ex 7	ins T nt 1222	frameshift from codon 408		Germany		(34)	N	1
59	Ex 7	del AG nt 1226	frameshift from codon 409		Sweden		(56)	Y	1
330 or 14A	Ex 7	del AG nt 1226	frameshift from codon 409	co	Germany		(66)	Y	1
225	Ex 7	del AG nt 1226	frameshift from codon 409	co	Germany		(47)	N	1
HNPCC-MO-2	Ex 7	del CCTA nt 1243	frameshift from codon 414		Italy		(20)	Y	1
A-MD1	Ex 7	del CCTA nt 1243	frameshift from codon 414	co/ga/ut/b	Italy		(95)	Y	1
BR120	Ex 7	del G nt 1249	frameshift		Brazil		(14)	Y	1
2/H62	Ex 7	C-T nt 1255	Gln419STOP	co/en/ga	Japan		(64)	Y	1
	Ex 7	C-A	Gln419Lys		Japan	0.01	(43)	N	2/3
RV	Ex 7		out of frame del codon 359-426	co	USA		(53)	N	1
NL-5	Ex 7	6 kb genomic deletion	genomic deletion		Netherlands		Wijnen, unpub.	Y	1
167	Ex 7-10		del exon 7-10	co	Sweden		(96)	Y	1
3	IVS 7	g-c nt 1276+1	splice defect	co/en/br	Moldavia		(60)	Y	1
	IVS 7	c-a nt 1276+49			African Am		Johnson, unpub.	Y	4
NL-23/NL-220	Ex 8	C-T nt 1285	Gln429STOP	co/st/pr/cx/sk/b/ HNPCC	Netherlands		(107)	Y	1
90	Ex 8	del CT nt 1316-1317	frameshift from codon 439, STOP codon 441	co	USA		(18)	N	1
O3	Ex 8	T-C nt 1319	Leu440Pro		Lithuania		Avizienyte, unpub.	Y	2/3
MD1699	Ex 8	ins A nt 1320	frameshift from codon 440	co	Scotland		Farrington/Dunlop, unpub.	N	1
	Ex 8	T-G nt 1339	Phe447Val		Netherlands	0/100	(108)	N	3/4
DF1754	Ex 8	del AG nt 1352-1353	frameshift - truncation codon 451	co	USA		(92)	N	1
HNP-043	Ex 8	del 32bp at 1354	in frame del codon 452-462		Italy		(4)	Y	1
DP1	Ex 8	A-T nt 1358	creates novel splice site	co/ut	Canada		(63)	N	1
3106	Ex 8	T-G nt 1373	Leu458STOP	co/MT	USA		(54)	Y	1
28	Ex 8	T-G nt 1373	Leu458STOP	co	Sweden		(97)	Y	1

Family 26	Ex 8	complex inversion and substitution	nonsense segment	co	Australia		(90)	N	1
HNP-049	Ex 8	Ex 8 deleted	deletion Ex 8 cod 426-462		Newfoundland		(4)	N	1
HNP-047	Ex 8	Ex 8 deleted	deletion Ex 8 cod 426-462		Italy		(4)	N	1
19	IVS 8	g-a nt 1386+1			Germany		(34)	N	1
C54	IVS 8	g-t nt 1387-1	splice defect	co	Finland		(109)	Y	1
52	IVS 8	g-t nt 1387-1	splice defect	co	Finland		(1)	N	1
Case 26	IVS 8	g-t nt 1387–8		co	USA		(87)	N	3/4
Family 27	Ex 9	del G nt 1408	frameshift and STOP codon 470	co	Australia		(90)	N	1
20	Ex 9	del G nt 1408	frameshift and STOP codon 470		Germany		(34)	N	1
CFS137	Ex 9	C-G nt 1418	Ser473STOP	co/es	Italy		(25)	N	1
43	Ex 9	A-T nt 1435	Ser479Cys	co	Italy		(13)	N	2/3
I-219	Ex 9	ins A nt 1444	frameshift from codon 481	co/en/st/l	Italy		(107)	Y	1
27,18	Ex 9	del AATG nt 1455	frameshift from codon 486	co	Hong Kong		(15)	N	1
21	Ex 9	C-T nt 1477	Gln493Stop		USA		de la Chapelle, unpub.	Y	1
R-RMS15	Ex 9	del A nt 1497	frameshift from codon 499 – truncation	co/ut/leu	Italy		(29)	N	1
386	Ex 9	ins C nt 1501	frameshift from codon 501		Czech Republic		Krepelova, unpub.	Y	1
Family 28	Ex 9	G-C nt 1510	splice defect	co	Australia		(90)	N	1
1515*	IVS 9	g-t nt 1511-10		co/HNPCC	Switzerland		(12)	Y	3/4
	IVS 9	t-a nt 1511-9			Norway	0.2	(8)	Y	3/4
	IVS 9	t-a nt 1511-9			Scotland/USA	0.852	(23)	N	3/4
	IVS 9	t-a nt 1511-9			Spain		Caldes, unpub.	Y	3/4
21	IVS 9	t-a nt 1510*(1511)*-9			Germany		(34)	N	3/4
cases & controls	IVS 9	a-t nt 1511–9			USA		(87)	N	**3/4**
NLB-172	IVS 9	a-g nt 1501(*1511*)-2	splice defect	co/HNPCC	Netherlands		Wijnen, unpub.	Y	1
SNU-YC13	Ex 10	G-T nt 1516	Asp506Tyr	co	Korea		(31)	Y	2
38	Ex 10	del CA nt 1550	frameshift from codon 517	co/HNPCC	Finland		(72)	Y	1
80	Ex 10	del CA nt 1550	frameshift from codon 517	co	Sweden		(56)	Y	1
19*	Ex 10	C-T nt 1552	Gln518STOP	co	Portugal		(24)	N	1
	Ex 10	T-C nt 1563	Tyr521Tyr		Czech Republic		Krepelova, unpub.	Y	4
1846	Ex 10	del A nt 1575	frameshift from codon 525		Switzerland/Ital		Heinimann, unpub.	Y	1
MTS-K2(n)	Ex 10	del A nt 1576	frameshift from codon 525	MT/en	Germany		(46)	Y	1
h1	Ex 10	del C nt 1577	frameshift from codon 525		England		Bunyan, unpub.	Y	1
Case 1/H23/H8	Ex 10	del A, codon 529	frameshift --> termination codon 542	co	Sweden		(69)	N	1
NL-13	Ex 10	ins G nt 1595	frameshift from codon 531	co/sk/ov/st/cx	Netherlands		(107)	Y	1

R-RM5	Ex 10	ins GA	truncation, codon 546-547	co	Italy		(29)	N	1
F719	IVS10	a-g nt 1661+10			Italy		(65)	N	4
5 kindreds	IVS 10	a-g nt 1661+10		co/HNPCC	Switzerland		(12)	N	4
Family 29	IVS 10	a-g, 12 from exon 10		co	Australia		(90)	N	4
	IVS 10	a-g nt 1661+12			Scotland/USA	0.62	(23)	N	4
	IVS 10	g-a nt 1661+12			Netherlands	0.25	(103)	Y	4
	IVS 10	g-a nt 1661+12			Sweden	0.47	(97)	Y	4
	IVS 10	g-a nt 1661+12			Sweden		(96)	Y	4
	IVS 10	g-a nt 1661+12			Spain		Caldes, unpub.	Y	4
22	IVS 10	g-a nt 1661+12			Germany		(34)	N	4
cases & controls	IVS 10	g-a nt 1661+12			USA		(87)	N	4
1 Kindred	IVS 10	a-t nt 1662-25			Scotland		Farrington/Dunlop, unpub.	N	4
-	IVS 10	t-a nt 1662-9			Scotland/USA	0.07	(23)	N	3/4
5 Kindreds	IVS 10	t-a nt 1662-9			Switzerland		(32)	N	3/4
3 Kindreds	IVS 10	t-a nt 1662-9			USA		(18)	N	3/4
12	Ex 11*(10)?*	A-C *(nt 1660)?*	Ser554Arg or possible splice defect		Germany		(67)	N	1/2
	Ex 11	T-C nt 1666	Leu556Leu		Netherlands	0.01	(107)	Y	4
	Ex 11	T-C nt 1666	Leu556Leu		Germany	0.01	(102)	Y	4
Case 27	Ex 11	T-C nt 1666	Leu556Leu	co	USA		(87)	N	4
Family 30	Ex 11	T-C nt 1669 (*1666*)	Leu557 (*556*) Leu		Australia		(90)	N	4
	Ex 11	del T nt 1674	frameshift – STOP codon 559		Europe		Deffenbaugh, unpub.	Y	1
132	Ex 11	del T nt 1677	frameshift from codon 559	MT/ga/br	Germany		(46)	Y	1
27	Ex 11	del A nt 1683	frameshift from codon 561	co	Germany		(102)	Y	1
23	Ex 11	del A nt 1683	frameshift from codon 561	co	Germany		(34)	N	1
S11	Ex 11	A-T nt 1685	Glu562Val	co	England		(5)	N	2
NL-221	Ex 11	del AA nt 1696	frameshift from codon 565	co/HNPCC	Netherlands		(105)	Y	1
167	Ex 11	del AAACA nt 1699	frameshift from codon 566	MT/en	Germany		(45)	Y	1
24	Ex 11	del AAACA nt 1699	frameshift from codon 566	co	Germany		(34)	N	1
GC12845, GC12313	Ex 11	del AAACA nt 1699	frameshift from codon 566		UK		West, unpub.	Y	1
Case 7	Ex 11	del nt 1700-1704	frameshift from codon 567	co/ut/ly	USA		(87)	N	1
339	Ex 11	del GA nt 1705	frameshift from codon 568	co	Japan		(70)	N	1
K-70/HNP168	Ex 11	del GA, codon 568-569	frameshift		Japan		(3)	N	1
GDLX-39#III-7	Ex 11	A-G nt 1737	Lys579Lys	co	Italy	0/100	(19)	N	4
	Ex 11	A-G nt 1737	Lys579Lys		Netherlands	0.01	(107)	Y	4
Case 28	Ex 11	A-G nt 1737	Lys579Lys	co	USA		(87)	N	4

1642	Ex 11	G-T nt 1738	Glu580STOP	Co/HNPCC	Switzerland		(32)	Y	1
B/4	Ex 11	G-T nt 1738	Glu580STOP	gl/co	Hong-Kong		(15, 50)	Y	1
	Ex 11	T-C nt 1755	Ser585Ser		Finland	0.005	(97)	Y	4
	IVS 11	g-a nt 1759+1	Splice defect		Latin American / Caribbean		Deffenbaugh, unpub.	Y	1
1807	IVS 11	t-c nt 1759 + 2	Splice defect		Switzerland		Heinimann, unpub.	Y	1
C/2	Ex 12	del G nt 1760	frameshift from codon 588	gl/co	Hong-Kong		(15, 50)	Y	1
MD1325	Ex 12	93 bp deletion nt 1783-1875	in frame del codon 595-625	co	Scotland		Farrington/Dunlop, unpub.	N	1/2
MD666	Ex 12	del AAT nt 1786	in frame del Asn codon 596	co/ut	Scotland		(53)	Y	1
MD1332,MD1370	Ex 12	del AAT nt 1786	in frame del Asn codon 596	co/ut	Scotland		Farrington/Dunlop, unpub.	N	1
1898 or 1A	Ex 12	del AAT nt 1786	in frame del Asn codon 596	re	USA		(66)	Y	1
1383*	Ex 12	del AAT nt 1786	in frame del Asn codon 596	co/HNPCC	Switzerland		(12)	Y	1
	Ex 12	del AAT nt 1786	in frame del Asn codon 596	co/HNPCC	Switzerland		(32)		1
	Ex 12	del AAT nt 1786	in frame del Asn codon 596	co/re	Switzerland		(61)	Y	1
1342	Ex 12	del AAT nt 1786	in frame del Asn codon 596	co/pa	Norway		(8)	Y	1
N-534, N-554	Ex 12	del AAT nt 1786	in frame del Asn codon 596	co/HNPCC	Norway		(105)	Y	1
FF50	Ex 12	del AAT nt 1786	in frame del Asn codon 596	co/ut	France		(99)	N	1
HNPCC-MO-29	Ex 12		Missense codon 596		Italy		Ponz de Leon, unpub.	Y	2/3
R-MD4	Ex 12	A-G nt 1787	Asn596Ser		Italian	<0.004	(95)	Y	2/3
GC9160	Ex 12	C-T nt 1801	Gln601STOP		UK		West, unpub.	Y	1
MD814	Ex 12	C-T nt 1801	Gln601STOP	co/ut/MT	Scotland		(51)	Y	1
1	Ex 12	C-T nt 1801	Gln601STOP	MT	UK		(45)	Y	1
138	Ex 12	G-A nt 1808	Asp603Asn	co	Finland		(86)	N	3/4
133	Ex 12	del T nt 1809	frameshift from codon 602	co/MT/pr	Germany		(45)	Y	1
25	Ex 12	del T nt 1809	frameshift from codon 602	co	Germany		(34)	N	1
26	Ex 12	C-T nt 1826	Ala609Val		Germany		(34)	N	3/4
	Ex 12	del T nt 1827	634STOP				Deffenbaugh, unpub.	Y	1
K-41/H110	Ex 12	T-G nt 1857	Tyr619STOP	co	Japan		(58)	N	1
25	Ex 12	ins TG nt 1860	frameshift from codon 619	co/HNPCC	Finland		(72)	Y	1
110	Ex 12	C-T nt 1861	Arg621STOP	co/bl/urt	USA		(101)	Y	1
4	Ex 12	C-T nt 1861	Arg621STOP	co/en/st/ov	Russia		(60)	Y	1
J	Ex 12	C-T nt 1865	Pro622Leu		New Zealand		(49)	Y	2/3
-	Ex 12	G-C nt 1906	Ala636Pro		France		Bressac-de Pailerets, unpub.	Y	2/3
54	Ex 12	G-C nt 1906	Ala636Pro		USA		de la Chapelle, unpub.	Y	2/3
Cx10	Ex 12	C-T nt 1915	His639Tyr – creates splice site		USA		(49)	N	1/2
MD579	Ex 12	C-T nt 1915	del codons 588-820	co/ut/ov	Scotland		(23)	Y	1
MD435	Ex 12	C-T nt 1915	del codons 588-820	co/ut/ov/br	Scotland		Farrington/Dunlop, unpub.	N	1

JV	Ex 12	C-T nt 1915	out of frame del codon 638-669	co	USA		(54)	Y	1
12*,4*,8*,11*, 15*,17*	Ex 12	A-G nt 1916	His639Arg		Japan		(68)	N	3
	Ex 12	T-G nt 1921	Cys641Gly		Scotland/USA	0.007	(23)	N	3/4
4*,8*,15*	Ex 12	G-A nt 1939	Glu647Lys		Japan		(68)	N	2
OZ-6	Ex 12	C-G nt 1968	Try656STOP	co/ut	Australia		(40)	Y	1
Family 31	Ex 12	G-C (*C-G*) nt 1968	stop codon	co	Australia		(90)	N	1
2	Ex 12	del AG nt 1985	frameshift from codon 661	co/MT/en/ov/ ur/bl	UK		(42)	Y	1
h110	Ex 12	del AG nt 1985	frameshift from codon 661		England		Bunyan, unpub.	Y	1
DP4	IVS 12	del 11bp nt 2005+2	splice defect	co/ut	Canada		(63)	N	2/3
	IVS 12	t-c nt 2006-6			Norway	0.14		N	3/4
	IVS 12	t-c nt 2006-6			USA		(49)	Y	3/4
	IVS 12	t-c nt 2006-6			Germany	0.06	(102)	Y	3/4
	IVS 12	t-c nt 2006-6			Scotland/USA	0.063	(23)	N	3/4
	IVS 12	t-c nt 2006-6			Switzerland		(32)	N	3/4
	IVS 12	t-g nt 2006-6			Sweden	0.23	(97)	Y	3/4
27	IVS 12	t-c nt 2006-6			Germany		(34)	N	3/4
cases & controls	IVS 12	t-g nt 2006-6			USA		(87)	N	3/4
	IVS 12	g-a, -6 *(-4)* upstream of exon 13			Sweden		(96)	Y	3/4
1033*, 1121*	IVS 12	t-c –4 (*-6*) from exon 13 splice acceptor		co/HNPCC	Switzerland		(12)	Y	2/3
MTS-K8(n)	Ex 13	del T nt 2015	frameshift from codon 671	MT/pa	Germany		(46)	N	1
MD830	Ex 13	C-T nt 2038	Arg680STOP	co	Scotland		(23)	Y	1
MD1440	Ex 13	C-T nt 2038	Arg680STOP	co	Scotland		Farrington/Dunlop, unpub.	N	1
N-414	Ex 13	C-T nt 2038	Arg680STOP	co/HNPCC	Norway		(105)	Y	1
	Ex 13	C-T nt 2038	Arg680STOP				Deffenbaugh, unpub.	Y	1
403	Ex 13	C-T nt 2038	Arg680STOP	co	USA		(18)	N	1
Case 29	Ex 13	G-A nt 2047	Gly683Arg	co	USA		(87)	N	2
Case 30	Ex 13	G-T nt 2048	Gly683Val	co	USA		(87)	N	2/3
Family 33	Ex 13	duplication 66 bp	reiteration	co	Australia		(90)	N	2
92	Ex 13	G-A nt 2064	Met688Ile	co	Japan		(70)	N	3/4
SNU-H1024	Ex 13	G-A nt 2064	Met688Ile	co	Korea		(110)	N	3/4
Case 31	Ex 13	T-C nt 2072	Ile691Arg	co	USA		(87)	N	2
56	Ex 13	T-C nt 2090 (*2089*)	Cys697Arg		Germany		(82)	Y	2
FF46	Ex 13	T-C nt 2090 (*2089*)	Cys697 Arg	co/ut/leu	France		(99)	N	2

62	Ex 13	G-T nt 2090	Cys697Phe		Germany		(102)	Y	2/3
28	Ex 13	G-T nt 2090	Cys697Phe		Germany		(34)	N	2/3
21957/ 3A	Ex 13	del G nt 2113	frameshift from codon 704		USA		(66)	Y	1
	Ex 13	C-T nt 2131	Arg711STOP		Argentina		Fodde, unpub.	Y	1
	Ex 13	G-C nt 2139	Gly713Gly		Finland	0.04	(72)	Y	4
IMS6	Ex 13	C-T nt 2152	Gln718STOP		Australia		Bennett, unpub.	Y	1
	Ex 13	A-G nt 2154	Gln718Gln		Lithuania		Avizienyte, unpub.	Y	4
Family 32	Ex 13	A-G nt 2157	Gln719Gln	co	Australia		(90)	N	4
21907 or 4A	Ex 13	del T nt 2204	frameshift from codon 734		USA		(66)	Y	1
TM	Ex 13	Ex 13 deleted	out of frame del Ex 13 codon 669-737	co/HNPCC	USA/Finland		(54)	N	1
300	Ex 13	Ex 13 deleted	out of frame del Ex 13 codon 669-737	co/HNPCC	USA		(39)	N	1
	IVS 13	t-c nt2211-6*(+6 or IVS12)*		co	USA		(18)	N	3/4
GC15825	IVS13	a-t nt 2211-2	Splice site		UK		West, unpub.	Y	1
MD579	IVS 13	g-t nt 2211 - 1	Splice defect, with mutation in Exon 12 – del codon 588-820	co/ut/ov	Scotland		(23)	Y	1
MD435	IVS 13	g-t nt 2211-1	Splice defect, with mutation in Exon 12 – del codon 588-820	co/ut/ov/br	Scotland		Farrington/Dunlop, unpub.	N	1
BIL-IGR924	Ex 14	G-A nt 2251	Gly751Arg		France		(22)	Y	2/3
1827	Ex 14	del C nt 2261	frameshift from codon 754		Switzerland		Heinimann, unpub.	Y	1
I-357	Ex 14	del C nt 2294	frameshift from codon 764	co	Italy		(77)	Y	1
	Ex 14	A-G nt 2308	Ile770Val		Scotland/USA	0.07	(23)	N	3/4
NL-203	Ex 14	del C nt 2345	frameshift from codon 781	co/en/pr/bl/st	Netherlands		(107)	Y	1
96	Ex 14	del C nt 2345	frameshift from codon 781		Germany		(82)	Y	1
22	Ex 14	ins TT nt 2360	frameshift from codon 787	co	Portugal		(24)	N	1
HNP3/HNP129/ K-32	Ex 14	T-G nt 2432	Leu811STOP	co/ga/ov	Japan		(64)	Y	1
CFS69	Ex 15	C-T nt 2470	Gln824STOP	co/ut	Italy		(25)	N	1
A-AV17	Ex 15	C-T nt 2470	Gln824STOP	co/ut	Italy		(29)	N	1
PLO-SG644	Ex 15	ins 14bp nt 2484	frameshift from codon 828		France		(22)	Y	1
NL-57	Ex 15	G-A nt 2500	Ala834Thr	co/HNPCC	Netherlands		(105).	Y	2/3
	Ex 15	G-A nt 2500	Ala834Thr		Netherlands	0/100	(108)	N	2/3
CFS112	Ex 15	G-A nt 2500	Ala834Thr	co	Italy		(25)	N	2/3
h26	Ex 15	del CTAATTT nt 2501	frameshift from codon 833		England		Bunyan, unpub.	Y	1
Family 34	Ex 15	del TAATTTC nt 2502	stop codon	co	Australia		(90)	N	1

Case 31	Ex 15	A-G nt 2503	Asn835Asp	co	USA	(87)	N	3/4
SC	Ex 15	7-bp deletion	frameshift from codon 834	co	England	(5)	N	1
31	Ex 15	7-bp deletion	frameshift from codon 834	co	England	(16)	N	1
HNPCC-3	Ex 15		frameshift from codon 834	co	England	(6)	N	1
K-30/HNP130	Ex 15	del T nt 2506	frameshift from codon 835	co	Japan	(58)	N	1
PT	Ex 15		frameshift from codon 843	co	England	(5)	N	1
G726	Ex 15		Gly845STOP (*Lys845STOP*)	co	Italy	(65)	N	1
350*	Ex 15	A-G nt 2533	Lys845Glu	co	Japan	(70)	N	2/3
MAR-SG63	Ex 15	C-T nt 2536	Gln846STOP		France	Bressac-de Paillerets, unpub.	Y	1
MAT-SG417	Ex 15	del C nt 2543	frameshift from codon 847		France	Bressac-de Paillerets, unpub.	Y	1
FF16	Ex 15	C-T nt 2581	Gln861STOP	co/d/ga/urt	France	(99)	N	1
37735	Ex 15	del 5 bp nt 2592	frameshift		UK	Norbury, unpub.	Y	1
HNP5	Ex 15	del AG nt 2629	frameshift from codon 876	co	Japan	(64)	Y	1
HNP-054	Ex 15	del AG nt 2633	frameshift from codon 877		French Canada	(4)	Y	1
DP2, DP3	Ex 15	del AG nt 2633	frameshift from codon 877	co/ut	Canada	(63)	N	1
SNU-H1027	Ex 15	del GA	frameshift	co	Korea	(110)	N	1
610	Ex 15	Ex 15 deleted	out of frame del Ex 15 codon 820-878	co/HNPCC	USA	(59)	N	1
DH	Ex 8-15	Ex 8-15 deleted	out of frame del Ex 8-15 codon 426-878	co/HNPCC	USA/Finland	(54)	N	1
JG	IVS 15	g-a nt 2634+1	out of frame del Ex 15 codon 820-878	co/HNPCC	USA/Finland	(54)	Y	1
30/96	Ex 16	del A nt 2647	frameshift from codon 883		Italy	(78)	N	1
K2064	Ex 16	del C nt 2662	frameshift from codon 887	co	USA	(91)	N	1
C003	Ex 16	C-G nt 2714	Thr905Arg	co	England	(26)	Y	2/3
Case 32,33	Ex 16	T-C nt 2766	Phe922Phe	co	USA	(87)	N	4
	Ex 16	G-C nt 2800 (*2802*)	Thr933Thr (*934*)		Switzerland	(32)	N	4

* = more than one gene variant identified in the index case

Key to 'Cancer' Column: am = ampulla of vater, b = brain, bd=bile duct, bl = bladder, bo = bone, br = breast, bs = basal cell, co = colon, cx = cervix, d = duodenum, en = endometrium, es = esophagus, ga = gastric, gb = gall bladder, gl = glioma, je = jejunum, k = kidney, l = lung, leu = leukaemia, ly = lymphoma, lv = liver, MT = Muir-Torre, ov = ovarian, pa = pancreas, ph = pharynx, pr = prostate, rh = rhabdomyosarcoma, re = rectal, rp = renal pelvic, sa = sarcoma, se = seminoma, si = small intestine, sk = skin, sp = spine, th = thyroid,TS = Turcots', uc = ulcerative colitis, ur = urether, urt = urinary tract, ut = uterine,

Abbreviations:

del = deletion, Ex = exon, ins = insertion, HNPCC = Hereditary Non-Polyposis Colorectal Cancer, IVS = Intervening Sequence, N / A = Not Available, nt = mucleotide, unpub. = unpublished, UTR = Untranslated Region

References

1. Aaltonen LA et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease [see comments]. New England Journal of Medicine 1998;338:1481-7.
2. Akiyama Y et al. Frequent microsatellite instabilities and analyses of the related genes in familial gastric cancers. Japanese Journal of Cancer Research 1996;87:595-601.
3. Bai YQ et al. Predominant germ-line mutation of the hMSH2 gene in Japanese hereditary non-polyposis colorectal cancer kindreds. International Journal of Cancer 1999;82:512-5.
4. Bapat BV et al. Family history characteristics, tumor microsatellite instability and germline MSH2 and MLH1 mutations in hereditary colorectal cancer. Human Genetics 1999;104:167-76.
5. Beck NE et al. Use of SSCP analysis to identify germline mutations in HNPCC families fulfilling the Amsterdam criteria. Human Genetics 1997;99:219-24.
6. Beck NE et al. Genetic testing is important in families with a history suggestive of hereditary non-polyposis colorectal cancer even if the Amsterdam criteria are not fulfilled [see comments]. British Journal of Surgery 1997;84:233-7.
7. Beck NE et al. Frequency of germline hereditary non-polyposis colorectal cancer gene mutations in patients with multiple or early onset colorectal adenomas. Gut 1997;41:235-8.
8. Borresen AL et al. Somatic mutations in the hMSH2 gene in microsatellite unstable colorectal carcinomas. Hum.Mol.Genet. 1995;4:2065-72.
9. Brieger A et al. Identification of germline mutations in hereditary nonpolyposis colorectal cancer using base excision sequence scanning analysis. Clin.Chem. 1999;45:1564-7.
10. Bronner CE et al. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature 1994;368:258-61.
11. Bubb VJ et al. Microsatellite instability and the role of hMSH2 in sporadic colorectalcancer. Oncogene 1996;12:2641-9.
12. Buerstedde JM et al. Detection of new mutations in six out of 10 Swiss HNPCC families by genomic sequencing of the hMSH2 and hMLH1 genes. Journal of Medical Genetics 1995;32:909-12.
13. Calistri D et al. Microsatellite instability in colorectal-cancer patients with suspected genetic predisposition. International Journal of Cancer 2000;89:87-91.
14. Cassandra M.Corvello RAUBBMRAJGS. A novel germline mutation at exon 7 of the MSH2 gene (1249delG) in a large HNPCC Brazilian kindred. Human Mutation 1999;13:506.
15. Chan TL et al. Frequent microsatellite instability and mismatch repair gene mutations in young Chinese patients with colorectal cancer. Journal of the National Cancer Institute 1999;91:1221-6.
16. Coleman MG et al. Minisatellite instability is found in colorectal tumours with mismatch repair deficiency. British Journal of Cancer 2001;85:1486-91.
17. Cravo M et al. BAT-26 identifies sporadic colorectal cancers with mutator phenotype: A correlative study with clinico-pathological features and mutations in mismatch repair genes. Journal of Pathology 1999;188:252-7.
18. Cunningham JM et al. The frequency of hereditary defective mismatch repair in a prospective series of unselected colorectal carcinomas. Am.J.Hum.Genet. 2001;69:780-90.
19. Curia MC et al. Unbalanced germ-line expression of hMLH1 and hMSH2 alleles in hereditary nonpolyposis colorectal cancer. Cancer Research 1999;59:3570-5.
20. de Leon MP et al. Hereditary colorectal cancer in the general population: from cancer registration to molecular diagnosis. Gut 1999;45:32-8.
21. Debniak T et al. Value of pedigree/clinical data, immunohistochemistry and microsatellite instability analyses in reducing the cost of determining hMLH1 and hMSH2 gene mutations in patients with colorectal cancer. European Journal of Cancer 2000;36:49-54.
22. Dieumegard B et al. Extensive molecular screening for hereditary non-polyposis colorectal cancer. British Journal of Cancer 2000;82:871-80.
23. Farrington SM et al. Systematic analysis of hMSH2 and hMLH1 in young colon cancer patients and controls. American Journal of Human Genetics 1998;63:749-59.
24. Fidalgo P et al. Detection of mutations in mismatch repair genes in Portuguese families with hereditary non-polyposis colorectal cancer (HNPCC) by a multi-method approach. European Journal of Human Genetics 2000;8:49-53.
25. Fornasarig M et al. Microsatellite instability and MLH1 and MSH2 germline defects are related to clinicopathological features in sporadic colorectal cancer. Oncology Reports 2000;7:39-43.
26. Froggatt NJ et al. Mutation screening of MSH2 and MLH1 mRNA in hereditary non-polyposis colon cancer syndrome. Journal of Medical Genetics 1996;33:726-30.
27. Froggatt NJ et al. A common MSH2 mutation in English and North American HNPCC families: Origin, phenotypic expression, and sex specific differences in colorectal cancer. Journal of Medical Genetics 1999;36:97-102.
28. Froggatt NJ et al. A frequent hMSH2 mutation in hereditary non-polyposis colon cancer syndrome [letter]. Lancet 1995;345:727.

29. Genuardi M et al. MLH1 and MSH2 constitutional mutations in colorectal cancer families not meeting the standard criteria for hereditary nonpolyposis colorectal cancer. International Journal of Cancer 1998;75:835-9.
30. Han H-J et al. Genomic structure of human mismatch repair gene, hMLH1, and its mutation analysis in patients with hereditary non-polyposis colorectal cancer (HNPCC). Human Molecular Genetics 1995;4:237-42.
31. Han H-J et al. Germline mutations of hMLH1 and hMSH2 genes in Korean hereditary nonpolyposis colorectal cancer. Journal of the National Cancer Institute 1996;88:1317-9.
32. Heinimann K et al. Influence of selection criteria on mutation detection in patients with hereditary nonpolyposis colorectal cancer. Cancer 1999;85:2512-8.
33. Herfarth KK et al. Mutations in MLH1 are more frequent than in MSH2 in sporadic colorectal cancers with microsatellite instability. Genes, Chromosomes & Cancer 1997;18:42-9.
34. Holinski-Feder E et al. DHPLC mutation analysis of the hereditary nonpolyposis colon cancer (HNPCC) genes hMLH1 and hMSH2. J.Biochem.Biophys.Methods 2001;47:21-32.
35. Hutter P et al. Excess of hMLH1 germline mutations in Swiss families with hereditary non-polyposis colorectal cancer. International Journal of Cancer 1998;78:680-4.
36. Hutter P et al. Complex genetic predisposition to cancer in an extended HNPCC family with an ancestral hMLH1 mutation. J.Med.Genet. 1996;33:636-40.
37. Ito E et al. A core promoter and a frequent single-nucleotide polymorphism of the mismatch repair gene hMLH1. Biochem.Biophys.Res.Commun. 1999;256:488-94.
38. Jager AC et al. Reduced frequency of extracolonic cancers in hereditary nonpolyposis colorectal cancer families with monoallelic hMLH1 expression. American Journal of Human Genetics 1997;61:129-38.
39. Jeon HM et al. Mutation of the hMSH2 gene in two families with hereditary nonpolyposis colorectal cancer. Hum.Mutat. 1996;7:327-33.
40. Kohonen-Corish M et al. RNA-based mutation screening in hereditary nonpolyposis colorectal cancer. American Journal of Human Genetics 1996;59:818-24.
41. Kolodner RD et al. Structure of the human MLH1 locus and analysis of a large hereditary nonpolyposis colorectal carcinoma kindred for mlh1 mutations. Cancer Res. 1995;55:242-8.
42. Kolodner RD et al. Structure of the human MSH2 locus and analysis of two Muir-Torre kindreds for msh2 mutations. Genomics 1994;24:516-26.
43. Konishi M et al. Molecular nature of colon tumors in hereditary nonpolyposis colon cancer, familial polyposis, and sporadic colon cancer. Gastroenterology 1996;111:307-17.
44. Kowalski LD et al. Mutational analysis of MLH1 and MSH2 in 25 prospectively-acquired RER+ endometrial cancers. Genes, Chromosomes & Cancer 1997;18:219-27.
45. Kruse R et al. Is the mismatch repair deficient type of Muir-Torre syndrome confined to mutations in the hMSH2 gene? Human Genetics 1996;98:747-50.
46. Kruse R et al. Muir-Torre phenotype has a frequency of DNA mismatch-repair-gene mutations similar to that in hereditary nonpolyposis colorectal cancer families defined by the Amsterdam criteria [published erratum appears in Am J Hum Genet 1998 Oct;63(4):1252]. American Journal of Human Genetics 1998;63:63-70.
47. Lamberti C et al. Microsatellite instability-a useful diagnostic tool to select patients at high risk for hereditary non-polyposis colorectal cancer: a study in different groups of patients with colorectal cancer. Gut 1999;44:839-43.
48. Lazar V et al. Accumulation of multiple mutations in tumour suppressor genes during colorectal tumorigenesis in HNPCC patients. Hum.Mol.Genet. 1994;3:2257-60.
49. Leach FS et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell 1993;75:1215-25.
50. Leung SY et al. Microsatellite instability and mutation of DNA mismatch repair genes in gliomas. Am.J.Pathol. 1998;153:1181-8.
51. Liu B et al. Genetic instability occurs in the majority of young patients with colorectal cancer [see comments]. Nature Medicine 1995;1:348-52.
52. Liu B et al. Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. Nat.Genet. 1995;9:48-55.
53. Liu B et al. Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients.[comment]. Nature Medicine 1996;2:169-74.
54. Liu B et al. hMSH2 mutations in hereditary nonpolyposis colorectal cancer kindreds. Cancer Research 1994;54:4590-4.
55. Liu T et al. MSH2 codon 322 Gly to Asp seems not to confer an increased risk for colorectal cancer susceptibility [letter]. European Journal of Cancer . 1998;34:1981.
56. Liu T et al. DGGE screening of mutations in mismatch repair genes (hMSH2 and hMLH1) in 34 Swedish families with colorectal cancer. Clinical Genetics 1998;53:131-5.
57. Loukola A et al. Microsatellite instability in adenomas as a marker for hereditary nonpolyposis colorectal cancer. Am.J.Pathol. 1999;155:1849-53.
58. Lu SL et al. Loss or somatic mutations of hMSH2 occur in hereditary nonpolyposis colorectal cancers with hMSH2 germline mutations. Japanese Journal of Cancer Research 1996;87:279-87.
59. Luce MC et al. In vitro transcription/translation assay for the screening of hMLH1 and hMSH2 mutations in familial colon cancer. Gastroenterology 1995;109:1368-74.
60. Maliaka YK et al. CpG dinucleotides in the hMSH2 and hMLH1 genes are hotspots for HNPCC mutations. Human Genetics 1996;97:251-5.
61. Mary JL et al. Mutational analysis of the hMSH2 gene reveals a three base pair deletion in a family predisposed to colorectal cancer development. Hum.Mol.Genet. 1994;3:2067-9.

62. Mauillon JL et al. Identification of novel germline hMLH1 mutations including a 22 kb Alu- mediated deletion in patients with familial colorectal cancer. Cancer Res. 1996;56:5728-33.
63. Millar AL et al. Mismatch repair gene defects contribute to the genetic basis of double primary cancers of the colorectum and endometrium. Human Molecular Genetics 1999;8:823-9.
64. Miyaki M et al. Germ line mutations of hMSH2 and hMLH1 genes in Japanese families with hereditary nonpolyposis colorectal cancer (HNPCC): Usefulness of DNA analysis for screening and diagnosis of HNPCC patients. Journal of Molecular Medicine 1995;73:515-20.
65. Montera M et al. Mutational germline analysis of hMSH2 and hMLH1 genes in early onset colorectal cancer patients [letter]. Journal of Medical Genetics (Online) 2000;37:E7.
66. Moslein G et al. Microsatellite instability and mutation analysis of hMSH2 and hMLH1 in patients with sporadic, familial and hereditary colorectal cancer. Human Molecular Genetics 1996;5:1245-52.
67. Muller-Koch Y et al. Sixteen rare sequence variants of the hMLH1 and hMSH2 genes found in a cohort of 254 suspected HNPCC (hereditary non-polyposis colorectal cancer) patients: mutations or polymorphisms? European Journal of Medical Research 2001;6:473-82.
68. Nakahara M et al. Identification of concurrent germ-line mutations in hMSH2 and/or hMLH1 in Japanese hereditary nonpolyposis colorectal cancer kindreds. Cancer Epidemiology, Biomarkers & Prevention 1997;6:1057-64.
69. Nilbert M et al. Microsatellite instability is rare in rectal carcinomas and signifies hereditary cancer. European Journal of Cancer 1999;35:942-5.
70. Nomura S et al. Enhanced detection of deleterious and other germline mutations of hMSH2 and hMLH1 in Japanese hereditary nonpolyposis colorectal cancer kindreds. Biochemical & Biophysical Research Communications 2000;271:120-9.
71. Nystrom-Lahti M et al. Founding mutations and Alu-mediated recombination in hereditary colon cancer. Nat.Med. 1995;1:1203-6.
72. Nystrom-Lahti M et al. DNA mismatch repair gene mutations in 55 kindreds with verified or putative hereditary non-polyposis colorectal cancer. Human Molecular Genetics 1996;5:763-9.
73. Panariello L et al. hMLH1 mutations in hereditary nonpolyposis colorectal cancer kindreds. Mutations in brief no. 182. Online. Hum.Mutat. 1998;12:216-7.
74. Papadopoulos N et al. Mutation of a mutL homolog in hereditary colon cancer [see comments]. Science 1994;263:1625-9.
75. Parsons R et al. Mismatch repair deficiency in phenotypically normal human cells. Science 1995;268:738-40.
76. Peltomaki P et al. Mutations predisposing to hereditary nonpolyposis colorectal cancer: Database and results of a collaborative study. Gastroenterology 1997;113:1146-58.
77. Pensotti V et al. Mean age of tumor onset in hereditary nonpolyposis colorectal cancer (HNPCC) families correlates with the presence of mutations in DNA hismatch repair genes. Genes, Chromosomes & Cancer 1997;19:135-42.
78. Percesepe A et al. Molecular screening for hereditary nonpolyposis colorectal cancer: a prospective, population-based study.[comment]. Journal of Clinical Oncology 2001;19:3944-50.
79. Pistorius SR et al. Clinical consequences of molecular diagnosis in families with mismatch repair gene germline mutations. Int.J.Colorectal Dis. 2000;15:255-63.
80. Planck M et al. hMLH1, hMSH2 and hMSH6 mutations in hereditary non-polyposis colorectal cancer families from southern Sweden. International Journal of Cancer 1999;83:197-202.
81. Quaresima B et al. Hereditary nonpolyposis coloretal cancer: identification of novel germline mutations in two kindreds not fulfulling the Amsterdam criteria [Online]. Human Mutation 1998;12:433.
82. Raedle J et al. Evaluation of rapid microsatellite analysis of paraffin-embedded specimens in screening for hereditary nonpolyposis colorectal cancer. Mod.Pathol. 1999;12:485-91.
83. Ravnik-Glavac M, Potocnik U, Glavac D. Incidence of germline hMLH1 and hMSH2 mutations (HNPCC patients) among newly diagnosed colorectal cancers in a Slovenian population. Journal of Medical Genetics 2000;37:533-6.
84. Ravnik-Glavac M et al. A novel in frame deletion of codons 188-190 in the hMSH2 gene of a slovenian patient with hereditary non-polyposis colorectal cancer. Hum.Hered. 1998;48:285-7.
85. Risinger JI et al. Molecular genetic evidence of the occurrence of breast cancer as an integral tumor in patients with the hereditary nonpolyposis colorectal carcinoma syndrome. Cancer 1996;77:1836-43.
86. Salovaara R et al. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. Journal of Clinical Oncology 2000;18:2193-200.
87. Samowitz WS et al. The colon cancer burden of genetically defined hereditary nonpolyposis colon cancer. Gastroenterology 2001;121:830-8.
88. Sarroca C et al. Hereditary nonpolyposis colorectal cancer (Lynch syndrome II) in Uruguay. Dis.Colon Rectum 2000;43:353-60.

89. Sasaki S et al. Mutational analysis of the hMLH1 gene using an automated two- dimensional DNA typing system. Human Mutation 1997;9:164-71.
90. Scott RJ et al. Hereditary nonpolyposis colorectal cancer in 95 families: differences and similarities between mutation-positive and mutation-negative kindreds. Am.J.Hum.Genet. 2001;68:118-27.
91. Swensen J, Lewis CM, Cannon-Albright LA. Identification of a one-base germline deletion (Codon 888 del C) and an intron splice acceptor site polymorphism in hMSH2. Human Mutation 1997;10:80-1.
92. Syngal S et al. Interpretation of genetic test results for hereditary nonpolyposis colorectal cancer: implications for clinical predisposition testing.[comment]. JAMA 1999;282:247-53.
93. Tannergard P et al. Mutation screening in the hMLH1 gene in Swedish hereditary nonpolyposis colon cancer families. Cancer Research 1995;55:6092-6.
94. Tomlinson IPM et al. Germline HNPCC gene variants have little influence on the risk for sporadic colorectal cancer. Journal of Medical Genetics 1997;34:39-42.
95. Viel A et al. Characterization of MSH2 and MLH1 mutations in Italian families with hereditary nonpolyposis colorectal cancer. Genes, Chromosomes & Cancer 1997;18:8-18.
96. Wahlberg S et al. Various mutation screening techniques in the DNA mismatch repair genes hMSH2 and hMLH1. Genetic Testing 1999;3:259-64.
97. Wahlberg SS et al. Low frequency of hMSH2 mutations in Swedish HNPCC families [letter]. International Journal of Cancer 1997;74:134-7.
98. Wang Q et al. Germline hMSH2 and hMLH1 gene mutations in incomplete HNPCC families. International Journal of Cancer 1997;73:831-6.
99. Wang Q et al. Prevalence of germline mutations of hMLH1, hMSH2, hPMS1, hPMS2, and hMSH6 genes in 75 French kindreds with nonpolyposis colorectal cancer. Human Genetics 1999;105:79-85.
100. Weber TK et al. Novel hMLH1 and hMSH2 germline mutations in African Americans with colorectal cancer. Journal of the American Medical Association 1999;281:2316-20.
101. Weber TK et al. Genomic DNA-based hMSH2 and hMLH1 mutation screening in 32 Eastern United States hereditary nonpolyposis colorectal cancer pedigrees. Cancer Research 1997;57:3798-803.
102. Wehner M et al. Hereditary nonpolyposis colorectal cancer (HNPCC): Eight novel germline mutations in hMSH2 or hMLH1 genes. Human Mutation 1997;10:241-4.
103. Wijnen J, Fodde R, Khan PM. DGGE polymorphism in intron 10 of MSH2, the HNPCC gene. Hum.Mol.Genet. 1994;3:2268.
104. Wijnen J et al. Majority of hMLH1 mutations responsible for hereditary nonpolyposis colorectal cancer cluster at the exonic region 15-16. American Journal of Human Genetics 1996;58:300-7.
105. Wijnen J et al. Hereditary nonpolyposis colorectal cancer families not complying with the Amsterdam criteria show extremely low frequency of mismatch-repair-gene mutations. American Journal of Human Genetics 1997;61:329-35.
106. Wijnen J et al. MSH2 genomic deletions are a frequent cause of HNPCC [letter]. Nature Genetics 1998;20:326-8.
107. Wijnen J et al. Seven new mutations in hMSH2, an HNPCC gene, identified by denaturing gradient-gel electrophoresis. American Journal of Human Genetics 1995;56:1060-6.
108. Wijnen JT et al. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. New England Journal of Medicine 1998;339:511-8.
109. Wu Y et al. MSH2 and MLH1 mutations in sporadic replication error-positive colorectal carcinoma as assessed by two-dimensional DNA electrophoresis. Genes, Chromosomes & Cancer 1997;18:269-78.
110. Yuan Y et al. Germline mutations of hMLH1 and hMSH2 genes in patients with suspected hereditary nonpolyposis colorectal cancer and sporadic early-onset colorectal cancer. Dis.Colon Rectum 1998;41:434-40.

Appendix A4 Summary of Studies Conducting Analysis of hMLH1 / hMSH2 in Colorectal Cancer Cases

The aim of this table is to summarise the findings of studies that conducted mutation analysis of hMLH1 / hMSH2 in colorectal cancer cases. Therefore, for the purpose of this table, "mutations" have been defined as those variants considered significant by the authors of the original papers.

Definitions of commonly used family history criteria can be found in table 1.1

(a) Mutation Analysis in Patients with a Family History of Colorectal Cancer

(i) Asia

Area of Study ; Recruitment period	Study Population and selection criteria	Total Number	Extent of Family History	Select Number	HMLH1 mutation carriers	hMSH2 mutation carriers	Method	Reference
Japan	Japanese kindreds	37	Fulfilment of Amsterdam Criteria	15	1 (6.7%)	8 (53.3%)	SSCP ⇒ Sequencing	(3)
			Fulfilment of Japanese Criteria	22	0	3 (13.6%)		
Japan	Japanese families that include multiple patients with CRC	11	Fulfilment of Amsterdam Criteria	11	5 (45.5%)	0	SSCP ⇒ Sequencing	(39)
Japan	Unrelated patients from laboratories in Japan (n=19), Korea (n=9) and the USA (n=6)	34	Fulfilment of Amsterdam Criteria	34	8 (24%)	N / A	SSCP ⇒ Sequencing	(23)
Japan	Japanese Kindreds	15	Fulfilment of Amsterdam Criteria	4	0 (Plus two missense mutations)	1 (25%)	RNA ⇒ Long RT-PCR ⇒ Sequencing	(44)
			Fulfilment of Japanese Criteria (1 group A, plus 10 group B)	11	1 (9.1%)	4 (36.4%) (Plus three missense mutations)		
Korea	Kindreds registered with the Korean Hereditary Colorectal Cancer Registry as having HNPCC or suspected HNPCC.	42	Group 1: Fulfilment of Amsterdam Criteria	25	8 (32%)	0	SSCP ⇒ Sequencing	(24)
			Group 2: Strong family history, but Amsterdam criteria not fulfilled	17	4 (23.5%)	0		

Korea	Kindreds registered with the Korean Hereditary Colorectal Cancer Registry. NB - This study was an extension of Han et al., (1996), and includes the 17 families previously analysed	31	Strong family history, but Amsterdam criteria not fulfilled	31	5 (16.1%)	2 (6.5%)	SSCP ⇒ Sequencing	(77)
Korea	Kindreds registered with the Korean Hereditary Colorectal Cancer Registry.	88	Fulfilment of the Amsterdam Criteria (n=33), or a history suggestive of HNPCC (n=55)	88	18 (20.5%)	2 (2.3%)	SSCP ⇒ Sequencing	(47)

(ii) Europe

Area of Study ; Recruitment period	Study Population and selection criteria	Total Number	Extent of Family History	Select Number	HMLH1 mutation carriers	hMSH2 mutation carriers	Method	Reference
Russia / Moldavia	Kindreds from Russia (n=3) and Moldavia (n=8)	11	Group 1: Fulfilment of Amsterdam Criteria	7	1 (14.3%)	3 (42.9%)	Sequencing	(37)
			Group 2: Clustering of colorectal and endometrial cancer in consecutive generations	4	1 (25%)	1 (25%)		
Finland	Finnish kindreds	55	Group 1: Fulfilment of Amsterdam Criteria	35	29 (82.9%)	1 (2.9%)	RT-PCR ⇒ 2-D DNA electrophoresis ⇒ Sequencing	(45)
			Group 2: An average of 4 affected members of each family, but not meeting the Amsterdam Criteria	20	5 (25%)	1 (5%)		

Sweden	Kindreds referred to the clinic for familial cancer at the Karolinska Hospital, Stockholm. Three families were of Finnish origin, while the rest were Swedish	39	Group 1: Fulfilment of Amsterdam Criteria	21	5 (23.8%)	N / A	DGGE ⇒ Sequencing	(62)
			Group 2: Family history of CRC	18	3 (16.7%)	N / A		
Sweden	Kindreds referred to the clinic for familial cancer at the Karolinska Hospital, Stockholm. Three families were of Finnish origin, while the rest were Swedish	39	Group 1: Fulfilment of Amsterdam Criteria	21	N / A	1 (4.8%)	DGGE ⇒ Sequencing	(67)
			Group 2: Family history of CRC	18	N / A	1 (5.6%)		
Sweden	Families recruited from N. Sweden and the Stockholm region through cancer family clinics at the Karolinska Hospital and Umea University Hospital	34	Group 1: Fulfilment of Amsterdam Criteria	7	1 (14.2%)	0	DGGE ⇒ Sequencing	(34)
			Group 2: More than one member affected with CRC or associated tumours, plus four families with at least one case with onset <35 years	27	1 (3.7%)	3 (11.1%)		
Switzerland	Kindreds	10	Fulfilment of Amsterdam Criteria	10	3 (30%)	3 (30%)	Direct Sequencing	(6)
Switzerland	Kindreds referred to University Hospital, Basel, from all parts of Switzerland, due to familial aggregation of CRC cases	26	Fulfilment of Amsterdam Criteria	15	6 (40%)	4 (27%)	SSCP ⇒ Sequencing	(25)
			Fulfilment of criteria extended to include extra-colonic tumours	11	0	0		
Switzerland	Index cases were referred to a familial cancer clinic with suspected HNPCC	23	Group 1: Fulfilment of Amsterdam Criteria	14	10 (71.4%)	0	IVSP ⇒ Sequencing	(27)
			Group 2: "HNPCC-like" but Amsterdam criteria not fulfilled	9	1 (11.1%)	0		

Italy	Families identified through probands hospitalised at the National Tumour Institute of Milan	16	Group 1: Fulfilment of Amsterdam Criteria	14	4 (28.6%)	3 (21.4%)	Combined strategies, including: IVSP, SSCP and direct sequencing	(49)
			Group 2: Strong family history, but Amsterdam criteria not fulfilled	2	0	1 (50%)		
Italy	Italian colorectal cancer patients identified through the specialised Modena CRC Registry.	36	Group 1: Fulfilment of Amsterdam Criteria	18	1 (5.6%)	2 (11.1%)	RT-PCR + SSCP ⇒ Sequencing	(14)
			Group 2: "suspected HNPCC", Amsterdam criteria not fulfilled.	18	0	0		
Italy	Italian kindreds	17	Fulfilment of Amsterdam Criteria	17	5 (29.4%)	2 (11.8%)	RT-PCR + SSCP ⇒ Sequencing	(66)
Italy	CRC patients identified through the Institute of Pathology and the Department of Clinical Physiopathology of the University of Florence, and from the Register of the Regina Elena Cancer Institute in Rome.	30	Group 1: Fulfilment of Amsterdam Criteria	17	2 (11.8%)	3 (17.6%)	IVSP ⇒ Sequencing	(13)
			Group 2: "HNPCC families", but Amsterdam criteria not fulfilled	13	2 (15.4%)	0		
Italy	CRC patients. Some of these patients had previously been reported by Pensotti et al.,	45	Group 1: Fulfilment of Amsterdam Criteria	13	3 (23.1%)	3 (23.1%)	SSCP ⇒ Sequencing	(7)
			Group 2: Incomplete HNPCC families, of patients with a strong family history, plus one patient with multiple primary tumours	11	1 (9.1%)	1 (9.1%)		

Italy, April 1994 – March 1996	CRC patients identified in three institutions – Aviano, Modena and Rome with putative HNPCC, but from families not meeting the standard Amsterdam criteria	32	At least one of the following criteria was met: - A family history suggestive of HNPCC, but one or more of the requirements of the Amsterdam criteria not fulfilled (n=29) - Age of onset < 40 years (n=15) - Multiple tumours in the index case(n=4)	32	3 (9.4%) NB - including two missense mutations of uncertain significance	3 (9.4%)	SSCP ⇒ Sequencing	(22)
France	CRC patients	17	Group 1: Fulfilment of Amsterdam Criteria	10	3 (30%)	2 (20%)	SSCP ⇒ Sequencing	(16)
			Group 2: Strong family history, but at least one Amsterdam criterion <u>not</u> fulfilled	7	2 (28.6%)	1 (10%)		
France	Families with suspected HNPCC, after exclusion of FAP	17	At least one family member who developed CRC at <50 years Index patient had at least one first degree relative with CRC or a tumour in the HNPCC spectrum At least 2 successive generations affected	17	5 (29.4%)	N / A	RT-PCR ⇒ Sequencing	(38)
France + Turkey	French families (n=10) were recruited through genetic consultations at the Centre Leon Berard and Hopital Edouard Herriot. Two Amsterdam +ve families were identified in Turkey.	12	Group 1: Fulfilment of Amsterdam Criteria	3	2 (66.6%)	0	IVSP + HA + Sequencing	(68)
			Group 2: "incomplete HNPCC syndrome", in which one of the Amsterdam criteria items was missing	9	3 (33.3%)	1 (11.1%)		

France	Kindreds were selected after genetic consultations at the Centre Leon Berard, Hopital Edouard Herriot, Centre Hospitalier Lyon-Sud, Centre Hospitalier Jean Minjoz and Centre Rene Gauducheau. NB – 17 kindreds were previously described (Wang et al., 1997)	75	Group 1: Fulfilment of Amsterdam Criteria	22	11 (50%)	3 (13.6%)	RNA + DNA based screening strategy, utilising HA, RT-PCR and sequencing	(69)
			Group 2: "incomplete HNPCC syndrome", in which one of the Amsterdam criteria items was missing	33	6 (18.2%)	2 (6.1%)		
			Group 3: Kindreds with one case of CRC and at least one case of an extra-colonic tumour belonging to the HNPCC spectrum	8	2 (25%)	2 (25%)		
Holland	Dutch Kindreds (n=30) were recruited from various clinical centres, largely through the Netherlands Foundation for the Detection of Hereditary Tumours. The remainder of the kindreds studied were Italian (n=3) and Danish (n=1) in origin.	34	Fulfilment of Amsterdam Criteria	34	12 (35.3%)	2 (20.6%)	DGGE ⇒ Sequencing	(65) NB – These same results are also included in subsequent papers: Wijnen,J. et al., 1995 and Wijnen et al., 1996
Holland + Norway etc.	The 34 kindreds from previous papers (Vasen et al., 1996) are again included in this analysis. In total, 97 Dutch kindreds were included, largely recruited through the Netherlands Foundation for the Detection of Hereditary Tumours. Other kindreds included were Norwegian (n=23), Italian (n=3), Danish (n=1) and Czech (n=1)	125	Group 1: Fulfilment of Amsterdam Criteria	86	25 (29.1%)	17 (19.8%)	DGGE ⇒ Sequencing	(73)
			Group 2: Family history of CRC, but at least one of the Amsterdam criteria items was missing	39	1 (2.6%)	2 (5.1%)		

Holland + Norway etc.	Some of these kindreds had been studied previously. Overall, 67 kindreds were recruited through the Netherlands Foundation for the Detection of Hereditary Tumours. Further kindreds from the Netherlands (n=56) and from Norway (n=56) were recruited by clinicians or clinical genetics centres. Other kindreds included were Norwegian (n=23), Italian (n=3), Danish (n=1) and Czech (n=1)	184	Group 1: Fulfilment of Amsterdam Criteria	92	25 (27.2%)	16 (17.4%)	DGGE ⇒ Sequencing	(75)
			Group 2: Family history of CRC, but at least one of the Amsterdam criteria items was missing	92	3 (3.3%)	3 (3.3%)		
Holland + Norway etc.	This paper re-examines all samples in which no MMR mutations were detected in previous studies, using a different technique to search for genomic deletions. The families studied were Dutch (n=86) and Norwegian (n=51)	137	Group 1: Fulfilment of Amsterdam Criteria	51	N / A	6 (11.7%)	Southern-blot analysis of genomic DNA	(74)
			Group 2: Family history of CRC, but at least one of the Amsterdam criteria items was missing	86	N / A	2 (2.3%)		
Holland + Norway etc.	Families registered with the Netherlands HNPCC registry between January and July 2000 (n=193, or which 116 met Amsterdam criteria I or II); plus suspected HNPCC families from the Clinical Genetic Centre Radium Hospital, Oslo, Norway (n=58)	251	Suspected HNPCC, including families meeting the Amsterdam criteria	251	34 (13.5%)	40 (15.9%)	DGGE ⇒ Sequencing	(64)

Germany	Families with suspected HNPCC	69	Group 1: Fulfilment of Amsterdam Criteria	57	11 (19.3%)	4 (7.0%)	SSCA + HA +PTT (where possible) ⇒ Sequencing	(29)
			Group 2: Fulfilment of a looser criteria, extended to include extra-colonic tumours	12	0	2 (16.7%)		
Germany	Families with suspected HNPCC	29	Most families included fulfilled the Amsterdam criteria (n=27). Two other samples were included.	29	6 (20.7%)	3 (10.3%)	SSCA + HA ⇒ Sequencing	(72)
France	Families, in which no alteration of MMR genes had been previously detected by classical methods	19	Fulfilment of Amsterdam Criteria (n=13) and partial fulfilment of these criteria (n=6)	19	0	3 (15.8%)	Multiplex PCR	(9)
England	Kindreds with clustering of colorectal and/or endometrial cancers in consecutive generations.	10	Family history present, but Amsterdam criteria not fulfilled	10	3 (30%)	3 (30%)	SSCP⇒ Sequencing	(5)
England	Kindreds with suspected HNPCC, presenting to Southampton General Hospital, UK	6	Family history present, but Amsterdam criteria fulfilled in only 4 of the 6 families	6	3 (50%)	1 (16.7%)	Sequencing	(10)
England	HNPCC families	17	Fulfilment of Amsterdam Criteria	17	3	5	mRNA analysis ⇒ IVTT ⇒ Sequencing	(21)

Portugal	Portuguese HNPCC families	20	Kindreds either fulfilled the Amsterdam criteria (n=16) or a modification of the criteria, designed to include those families in which three or more relatives have a tumour within the HNPCC spectrum (n=4)	20	7 (35%)	4 (20%)	PTT + SSCP + HA + DGGE	(19)

(iii) Australia

Area of Study ; Recruitment period	Study Population and selection criteria	Total Number	Extent of Family History	Select Number	HMLH1 mutation carriers	hMSH2 mutation carriers	Method	Reference
Australia	Families ascertained through three centres; The Royal Melbourne Hospital, The Princess Margaret Hospital for Children, and the John Curtin School of Medical Research	18	Fulfilment of Amsterdam Criteria	18	4 (22.2%)	2 (11.1%)	A combination of RNA-based and DNA based methods	(28)
Australia	Families with a history of colorectal cancer	95	Fulfilment of Amsterdam Criteria	33	11 (33.3%)	9 (27.3%)	DGGE⇒ Sequencing	(58)
			Fulfilment of Bethesda Criteria	62	6 (9.7%)	6 (9.7%)		

(iv) North America

Area of Study ; Recruitment period	Study Population and selection criteria	Total Number	Extent of Family History	Select Number	HMLH1 mutation carriers	hMSH2 mutation carriers	Method	Reference
USA	Kindreds from the USA (n=26), Finland (n=2) and New Zealand (n=1)	29	Fulfilment of Amsterdam Criteria	29	N / A	10 (34.5%)	Linkage analysis ⇒ IVSP ⇒ Sequencing	(33)
USA	Patients with a family history of colorectal and/or endometrial cancer were identified from a consecutive series of CRC patients referred to the Gastrointestinal Cancer Prevention Clinic of the Dept. of Internal Medicine at the University of Michigan	19	Group 1: Fulfilment of Amsterdam Criteria	12	4 33.3%)	2 (16.7%)	IVTT ⇒ Sequencing	(36)
			Group 2: Amsterdam criteria not entirely fulfilled	7	0	0		
USA + Germany	Patients with a family history of CRC were identified through the Dept. of Medical Genetics at the Mayo clinic, Minnesota, USA, or through the Dept. of Surgery of the University of Dusseldorf.	39	Group 1: Fulfilment of Amsterdam Criteria	20	4 (20%)	5 (25%)	Sequencing	(41)
			Group 2: Index patient plus at least one other family member with CRC	19	3 (15.8%)	0		
USA (?) + Finland	Kindreds with suspected HNPCC. Origin is not clear, but at least 7 of the kindreds were Finnish	28	Group 1: Fulfilment of Amsterdam Criteria, and linkage previously shown to 3p markers	10	9 (90%)	N / A	RT-PCR ⇒ Sequencing	(46)
			Group 2: Fulfilment of Amsterdam Criteria (kindreds too small for linkage analysis)	18	1 (5.6%)	N / A	IVTT ⇒ Sequencing	

USA	Colorectal cancer kindreds	11	Kindreds did not necessarily meet the criteria for HNPCC	11	N / A	1 (9.1%)	Sequencing	(60)
USA	Families self-referred, or referred by health care providers to a cancer genetics program on the basis of multiple CRC cases, early onset or familial association of CRC with other HNPCC-associated tumours	58	Group 1: Fulfilment of Amsterdam Criteria	28	10 (35.7%)	1 (3.6%)	Sequencing	(61)
			Group 2: Families fulfilling a modified Amsterdam Criteria	11	0	2 (18.2%)		
			Group 3: HNPCC variant – families with a history suggestive of HNPCC, but not fulfilling the above criteria	19	0	3 (15.8%)		
USA	HNPCC pedigrees registered with the Roswell Park Familial Cancer Registry and the Vermont Cancer Centre Familial Cancer Program	32	All but one of the kindreds met the Amsterdam Criteria. In the kindred that did not, all affected individuals were confined to the same generation.	32	5 (15.6%)	3 (9.4%)	SSCP⇒ Sequencing	(71)
Canada	Patients were identified through the Familial GI Cancer registry at Mount Sinai Hospital, Toronto. All but one families were of European descent, the other being Chinese.	3	Group 1: Fulfilment of Amsterdam Criteria	14	2	5	RT-PCR + PTT ⇒ Sequencing	(4)
			Group 2: Families fulfilling a modified Amsterdam criteria, referred to as the Mount Sinai Hospital Criteria.	19	1	4		

(v) Meta-analysis

Area of Study ; Recruitment period	Study Population and selection criteria	Total Number	Extent of Family History	Select Number	HMLH1 mutation carriers	hMSH2 mutation carriers	Method	Reference
Various	Data was collected from survey forms mailed to members of the ICG-HNPCC. Data were gathered from eight institutions in seven countries. Amsterdam +ve families were excluded from the analysis, and the famililies included were classified according to the criteria for families with suspected HNPCC, as defined by the ICG-HNPCC at the 8th annual meeting in Buffalo, USA	123	Criteria I: at least two first-degree relatives affected with colorectal cancer with at least one of the following: development of multiple colorectal tumors including adenomatous polyp, at least one colorectal cancer case diagnosed before the age of 50, and occurrence of a hereditary nonpolyposis colorectal cancer extracolonic cancer (endometrium, urinary tract, small intestine, stomach, hepatobiliary system, or ovary) in family members.	67	12 (17.9%)	7 / 66 (10.6%)	Various	(48)
			Criteria II: one colorectal cancer patient with at least one of the following: early age of onset (<40 years); endometrial, urinary tract, or small intestine cancer in the index patient or a sibling (one aged <50 years); and two siblings with other integral hereditary nonpolyposis colorectal cancer extracolonic cancers (one aged <50 years).	56	4 (7.1%)	1 (1.8%)		

(b) Mutation Analysis in Patients with MSI +ve Tumours

(i) Asia

Area of Study ; Recruitment period	Study Population and ascertainment criteria	Age Range	Total Number	Number MSI +ve	Criteria for MSI +ve status	Number selected for mutation analysis	HMLH1 mutation carriers	hMSH2 mutation carriers	Method	Reference
Hong Kong, 1991 - 1997	Patients with adenocarcinoma removed at colectomy at Queen Mary Hospital, Hong Kong	<46 years old	59	19 (32%)	RER phenotype at >40% of loci analysed	15	1 (7%)	7 (47%)	IVSP ⇒ sequencing	(8)
Japan	Colorectal cancer patients undergoing surgery at the Jichi Medical School. No patients had a family history of CRC in a first degree relative		102	16 (15.7%)	RER evident in at least 1/6 loci	16	0	0	SSCP ⇒ Sequencing	(59)
Japan, 1992-1997	CRC patients over a five year period at the Cancer Institute Hospital		129	N / A	RER phenotype in at least 2 of 4 loci	8	1 (12.5%)	N / A	Automated 2D DNA typing system	(57)
Japan 1980-1995	Patients treated at Hiroshima University School of Medicine, who fulfilled the Japanese registry's clinical diagnostic criteria for HNPCC. Five of these fulfilled the Amsterdam criteria		29	11 (37.%)	RER phenotype in >6 of 12 loci	11	4 (36.4%)	2 (18.2%)	SSCP ⇒ Sequencing	(42)
Japan	Sporadic CRC cases undergoing surgery at Nihon University School of Medicine, Tokyo.		110	31 (28.2%)	RER phenotype in at least 1 of 8 loci	31	0	0	SSCP ⇒ Sequencing	(2)

(ii) Europe

Area of Study ; Recruitment period	Study Population and ascertainment criteria	Age Range	Total Number	Number MSI +ve	Criteria for MSI +ve status	Number selected for mutation analysis	HMLH1 mutation carriers	hMSH2 mutation carriers	Method	Reference
Scotland 1970-1993	Index patients identified from the Scottish National Cancer Registry. Patients referred specifically because they fulfilled HNPCC criteria were excluded.	<35 years old	27	13 (48.1%)	RER phenotype in at least 50% of at least 4 loci	13	1 (7.7%)	5 (38.5%)	IVSP ⇒ SSCP ⇒ Sequencing	(17)
Poland	Late onset sporadic CRC cases, with no family history of HNPCC associated cancers, and without synchronous or metachronous cancers.	>40 years old	43	4 (9.3%)	RER phenotype in at least 2 loci from a panel of 5, or in at least 3 of 10 loci	4	0	0	Sequencing	(15)
Slovenia, 1996-1998	Patients diagnosed with CRC from clinics all over Slovenia		300	29 (MSI-H)	RER phenotype in at least > 40% of up to 12 loci	29	3 (10.3%)	1 (3.4%)	PCR + single / double conformational analysis ⇒ Sequencing	(54)
Germany	Group A: some relatives affected by tumours of the HNPCC spectrum, with out meeting the Amsterdam Criteria	<50 years old	45	26 (57.8%)	RER phenotype in at least > 40% of up to 10 loci	17	4 (23.5%)	1 (5.9%)	SSCA + HA +PTT (where possible) ⇒ Sequencing	(29)
	Group B: Apparently sporadic colorectal cancer		46	16 (34.8%)		10	2 (20%)	1 (10%)		
Germany / Czech Republic	Index cases meeting the Bethesda guidelines, recruited from the Department of Visceral, Thoracic and Vascular Surgery, University of Dresden and from other departments in Germany and the Czech Republic.		72	38 (52.8%)	RER phenotype at >30% of at least five markers assessed	38	8 (21%)	9 (23.7%)	Sequencing	(52)

Finland, May 1994 – April 1996	Consecutive cases of sporadic colorectal cancer treated in nine regional hospitals in SE Finland		509	63 (12.4%)	RER phenotype in at least 2/7 loci, or 1 of 7 plus at least 1 from an additional panel	63	9* (14.3%) *Mostly founder mutations	1 (1.6%)	Founder mutation analysis ⇒ Sequencing	(1)
Finland, March 1996 – June 1998	Consecutive cases of sporadic colorectal cancer treated in nine regional hospitals in SE Finland	29 – 91 years, mean age = 67	535	66 (12.3%)	DNA was studied for MSI using the BAT26 and TGF-βRII poly-A markers	66	17* (26%) *Mostly founder mutations	1 (1.5%)	Founder mutation analysis ⇒ DGGE ⇒ Sequencing	(55)
Finland, 1994-1998	Colorectal adenoma specimens collected at colonoscopy or surgery in nine large regional hospitals in SE Finland.	23 - 90 years old, mean age = 66	378	6 (1.6%)	DNA was studied for MSI using the BAT26 and TGF-βRII poly-A markers	6	5 (83.3%)	0	Founder mutation analysis ⇒ Sequencing	(35)
Finland	Colorectal tumour samples that had previously been shown to be unstable at 1 or 2 of 6-14 loci, but had not met the criteria for RER positivity		49	15	RER phenotype in at 1 or more of an additional six markers	11 'mild' RER tumour specimens	0	0	RT-PCR ⇒ IVSP ⇒ Sequencing	(51)
Finland	Tumour DNA was from primary colorectal carcinomas previously found to have an RER+ve phenotype	36 – 89 years mean age = 70	N / A	33	N / A	33	1 (3%)	1 (3%)	2-D Gel Electrophoresis	(76)
Sweden, 1996-1998	Consecutive pre-operative biopsy specimens from rectal cancers in patients referred to the Dept. of Oncology, University Hospital, Lund, Sweden	24 - 92 years median age = 65	165	3 (1.8%)	RER phenotype in at least 1 of at least 4 loci (always including BAT25, BAT26 and BAT40)	3	1 (33.3%)	1 (33.3%)	DGGE ⇒ Sequencing	(43)

Sweden	Kindreds meeting the extended Amsterdam criteria for HNPCC (n=14), plus kindreds in which the proband presented at an early age with metachronous colorectal / endometrial cancer (n=2). Origin of the kindreds is not given; most were Swedish although at least one was from the Czech republic, and one from Iran.		N / A	16	According to Boland et al., (1998)	16	5 (31.3%)	2 (12.5%)	DGGE ⇒ Sequencing	(53)
Portugal	Patients presenting with apparently sporadic colorectal cancer. Cases were excluded if they matched the criteria for a diagnosis of HNPCC, FAP, IBD or if they had undergone preoperative radio / chemotherapy.	34 – 83 years mean age =64	62	9 (14.5%)	RER phenotype in at least 2/7 poly(CA) loci	8	2 (25%)	0	DGGE ⇒ Sequencing	(11)
Italy	Italian colorectal cancer patients prospectively ascertained through the Modena CRC Registry.		336	28 (8.3%)	MSI in >30% of markers	12 **NB**: eight were randomly chosen, four were chosen on the basis of clinical features	0	1 (8.3%)	SSCP ⇒ Sequencing **NB**: Some sequenced directly	(50)

(iii) North America

Area of Study ; Recruitment period	Study Population and ascertainment criteria	Age Range	Total Number	Number MSI +ve	Criteria for MSI +ve status	Number selected for mutation analysis	HMLH1 mutation carriers	hMSH2 mutation carriers	Method	Reference

USA + Scotland	Scotland: A subset of 115 CRC registrations between 1970 and 1993, that met practical criteria for analysis (n=17) USA: (i) Patients with stage II or III CRC undergoing curative surgical resections at the Johns Hopkins Hospital between 1986 and 1990 (n=145) (ii) Randomly selected patients in whom surgery was performed between 1978 and 1985 (n=65) All patients were excluded if they met the Amsterdam criteria for HNPCC, or had evidence of FAP or IBD	Scotland: <30 years old USA: (i) unknown (ii) <50 years old	189 (38 of the 227 samples were not of adequate quality)	37 (19.6%)	RER phenotype in at least 2 of 4 or 5 loci	12	2 (16.7%)	3 (25%)	RT-PCR ⇒ IVSP ⇒ Sequencing	(30)
USA	Consecutive cases of sporadic colorectal cancer, treated at the Johns Hopkins Hospital (1984-1990)		181	24 (13.3%)	N / A	10	0	1 (10%)	RT-PCR ⇒ IVSP ⇒ Sequencing	(31)
USA + New Zealand + Europe	Kindreds meeting the Amsterdam criteria for HNPCC were included in the study. The origin of the kindreds were USA (n=50), New Zealand (n=12) and Europe (n=12) NB: 19 of the 31 mutation+ve patients had been identified through previous studies		74	68 (91.9%)	RER phenotype in at least 2 of 4 or more loci	48	16 (33.3%)	15 (31.3%)	RT-PCR ⇒ IVSP ⇒ Sequencing	(32)
USA	Patients with MSI +ve colorectal tumours identified by a previous study, classed as sporadic (n=7) CRC cases		N / A	7	N / A	7	1 (14.3%)	0	Sequencing	(41)

USA	Colorectal cancer patients were recruited through the Washington University School of Medicine, St. Louis, Missouri	38-83 years mean age = 65.6	61	13 (21.3%)	RER phenotype in at least 2 of 7 loci	12	1 (8.3%)	0	RT-PCR + SSCP ⇒ Sequencing	(26)
USA	Unselected prospective series of patients presenting to the Mayo clinic, Minnesota, between December 1995 and April 1997, and consenting to participate	29–91 years, mean age = 69	257	51 (19.8%)	Instability evident in more than 30% of seven markers analysed	51	4 (7.8%)	3 (5.8%)	Sequencing	(12)
USA	Colon cancer patients from Utah and California, originally enrolled as part of a wider epidemiological study. Patients with rectal cancer were excluded.	30-79 years	1066	171 (16.0%)	MSI in >30% of 10 tetranucleotide repeats, or instability of BAT-26 or TGF-βRII	130	5 (3.8%)	3 (2.3%)	Sequencing	(56)

(c) Mutation Analysis in Patients with Early Onset Colorectal Cancer

Area of Study ; Recruitment period	**Study Population and selection criteria**	**Age**	**Number selected for mutation analysis**	**HMLH1 mutation carriers**	**hMSH2 mutation carriers**	**Method**	**Reference**
Scotland	Scottish sporadic CRC patients diagnosed at <30 years, identified retrospectively from cancer registrations since 1970	<30 years old	50	7 (14%)	7 (14%)	Sequencing plus complementary IVSP	(18)

England	CRC cases in young patients with no family history of CRC	<45 years old	50	2 (4%)	1 (2%)	SSCP⇒ Sequencing	(63)
France	**Caucasian CRC patients in a clinic-based population, selected solely on the basis of age at onset of CRC.**	<50 years old	7	0	0	SSCP⇒ Sequencing	(16)
France	Patients recruited through genetic consultations at the Centre Leon Berard and Hopital Edouard Herriot.	<50 years old	7	0	0	IVSP + HA + Sequencing	(68)
France	Patients selected after genetic consultations at the Centre Leon Berard, Hopital Edouard Herriot, Centre Hospitalier Lyon-Sud, Centre Hospitalier Jean Minjoz and Centre Rene Gauducheau.	<50 years old	12	0	0	RNA + DNA based screening strategy, utilising HA, RT-PCR and sequencing	(69) NB – Some patients may have been described previously (Wang et al., 1997)
Italy	Patients were recruited "from various surgical and clinical units". Family history was not used as an entry criterion.	<50 years old	54	5 (9.2%)	4 (7.4%)	SSCP + PTT ⇒ Sequencing	(40)
Italy	Patients without a family history suggestive of hereditary CRC, treated at the Centro di Riferimento Oncologico, Aviano, Italy (~1995-1999)	<45 years old	38	1 (2.6%)	2 (5.3%)	SSCP⇒ Sequencing	(20)
Korea	Data on early-onset CRC without family history were collected through the Dept. of Surgery, Seoul National University Hospital.	<50 years old	22	0	1 (4.5%)	SSCP⇒ Sequencing	(24)

Korea	Data on early-onset CRC cases without family history were collected through the Dept. of Surgery, Seoul National University Hospital.	<50 years old	45	0	1 (2.2%)	SSCP⇒ Sequencing	(77) NB Continuation of above study
USA	Patients without a family history fulfilling the strict, or the modified, Amsterdam criteria	<40 years old	12	1 (8.3%)	1 (8.3%)	Sequencing	(61)
USA	African Americans enrolled with the Roswell Park Family Cancer Registry	<50 years old	10	1 (10%)	1 (10%)	SSCP⇒ Sequencing	(70)

Abbreviations: CRC = Colorectal Cancer, DGGE = Denaturing Grade Gel Electrophoresis, FAP = Familial Adenomatous Polyposis, HA = Heteroduplex Analysis, HNPCC = Hereditary Non-Polyposis Colorectal Cancer, IBD = Inflammatory Bowel Disease, ICG HNPCC = International Collaborative Group on HNPCC, IVTT = In Vitro Transcription-Translation, IVSP = In Vitro Synthesised Protein assay, MSI = Microsatellite Instability, PCR = Polymerase Chain Reaction, PTT = Protein Truncation Test, RER = Replication Error, RT-PCR = Reverse Transcription Polymerase Chain Reaction, SSCP = Single Stranded Conformational Polymorphism

References

1. Aaltonen LA et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease [see comments]. New England Journal of Medicine 1998;338:1481-7.
2. Abe Y, Masuda H. Genetic alterations of sporadic colorectal cancer with microsatellite instability, especially characteristics of primary multiple colorectal cancers. Journal of Surgical Oncology 2000;74:249-56.
3. Bai YQ et al. Predominant germ-line mutation of the hMSH2 gene in Japanese hereditary non-polyposis colorectal cancer kindreds. International Journal of Cancer 1999;82:512-5.
4. Bapat BV et al. Family history characteristics, tumor microsatellite instability and germline MSH2 and MLH1 mutations in hereditary colorectal cancer. Human Genetics 1999;104:167-76.
5. Beck NE et al. Genetic testing is important in families with a history suggestive of hereditary non-polyposis colorectal cancer even if the Amsterdam criteria are not fulfilled [see comments]. British Journal of Surgery 1997;84:233-7.
6. Buerstedde JM et al. Detection of new mutations in six out of 10 Swiss HNPCC families by genomic sequencing of the hMSH2 and hMLH1 genes. Journal of Medical Genetics 1995;32:909-12.
7. Calistri D et al. Microsatellite instability in colorectal-cancer patients with suspected genetic predisposition. International Journal of Cancer 2000;89:87-91.
8. Chan TL et al. Frequent microsatellite instability and mismatch repair gene mutations in young Chinese patients with colorectal cancer. Journal of the National Cancer Institute 1999;91:1221-6.
9. Charbonnier F et al. Detection of exon deletions and duplications of the mismatch repair genes in hereditary nonpolyposis colorectal cancer families using multiplex polymerase chain reaction of short fluorescent fragments. Cancer Research 2000;60:2760-3.
10. Coleman MG et al. Minisatellite instability is found in colorectal tumours with mismatch repair deficiency. British Journal of Cancer 2001;85:1486-91.
11. Cravo M et al. BAT-26 identifies sporadic colorectal cancers with mutator phenotype: A correlative study with clinico-pathological features and mutations in mismatch repair genes. Journal of Pathology 1999;188:252-7.

12. Cunningham JM et al. The frequency of hereditary defective mismatch repair in a prospective series of unselected colorectal carcinomas. Am.J.Hum.Genet. 2001;69:780-90.
13. Curia MC et al. Unbalanced germ-line expression of hMLH1 and hMSH2 alleles in hereditary nonpolyposis colorectal cancer. Cancer Research 1999;59:3570-5.
14. de Leon MP et al. Hereditary colorectal cancer in the general population: from cancer registration to molecular diagnosis. Gut 1999;45:32-8.
15. Debniak T et al. Value of pedigree/clinical data, immunohistochemistry and microsatellite instability analyses in reducing the cost of determining hMLH1 and hMSH2 gene mutations in patients with colorectal cancer. European Journal of Cancer 2000;36:49-54.
16. Dieumegard B et al. Extensive molecular screening for hereditary non-polyposis colorectal cancer. British Journal of Cancer 2000;82:871-80.
17. Dunlop MG et al. Cancer risk associated with germline DNA mismatch repair gene mutations. Hum.Mol.Genet. 1997;6:105-10.
18. Farrington SM et al. Systematic analysis of hMSH2 and hMLH1 in young colon cancer patients and controls. American Journal of Human Genetics 1998;63:749-59.
19. Fidalgo P et al. Detection of mutations in mismatch repair genes in Portuguese families with hereditary non-polyposis colorectal cancer (HNPCC) by a multi-method approach. European Journal of Human Genetics 2000;8:49-53.
20. Fornasarig M et al. Microsatellite instability and MLH1 and MSH2 germline defects are related to clinicopathological features in sporadic colorectal cancer. Oncology Reports 2000;7:39-43.
21. Froggatt NJ et al. Mutation screening of MSH2 and MLH1 mRNA in hereditary non-polyposis colon cancer syndrome. Journal of Medical Genetics 1996;33:726-30.
22. Genuardi M et al. MLH1 and MSH2 constitutional mutations in colorectal cancer families not meeting the standard criteria for hereditary nonpolyposis colorectal cancer. International Journal of Cancer 1998;75:835-9.
23. Han H-J et al. Genomic structure of human mismatch repair gene, hMLH1, and its mutation analysis in patients with HNPCC. Human Molecular Genetics 1995;4:237-42.
24. Han H-J et al. Germline mutations of hMLH1 and hMSH2 genes in Korean hereditary nonpolyposis colorectal cancer. Journal of the National Cancer Institute 1996;88:1317-9.
25. Heinimann K et al. Influence of selection criteria on mutation detection in patients with hereditary nonpolyposis colorectal cancer. Cancer 1999;85:2512-8.
26. Herfarth K et al. Mutations in MLH1 are more frequent than in MSH2 in sporadic colorectal cancers with microsatellite instability. Genes, Chromosomes & Cancer 1997;18:42-9.
27. Hutter P et al. Excess of hMLH1 germline mutations in Swiss families with hereditary non-polyposis colorectal cancer. International Journal of Cancer 1998;78:680-4.
28. Kohonen-Corish M et al. RNA-based mutation screening in hereditary nonpolyposis colorectal cancer. American Journal of Human Genetics 1996;59:818-24.
29. Lamberti C et al. Microsatellite instability-a useful diagnostic tool to select patients at high risk for hereditary non-polyposis colorectal cancer: a study in different groups of patients with colorectal cancer. Gut 1999;44:839-43.
30. Liu B et al. Genetic instability occurs in the majority of young patients with colorectal cancer [see comments]. Nature Medicine 1995;1:348-52.
31. Liu B et al. Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. Nat.Genet. 1995;9:48-55.
32. Liu B et al. Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients.[comment]. Nature Medicine 1996;2:169-74.
33. Liu B et al. hMSH2 mutations in hereditary nonpolyposis colorectal cancer kindreds. Cancer Research 1994;54:4590-4.
34. Liu T et al. DGGE screening of mutations in mismatch repair genes (hMSH2 and hMLH1) in 34 Swedish families with colorectal cancer. Clinical Genetics 1998;53:131-5.
35. Loukola A et al. Microsatellite instability in adenomas as a marker for hereditary nonpolyposis colorectal cancer. Am.J.Pathol. 1999;155:1849-53.
36. Luce MC et al. In vitro transcription/translation assay for the screening of hMLH1 and hMSH2 mutations in familial colon cancer. Gastroenterology 1995;109:1368-74.
37. Maliaka YK et al. CpG dinucleotides in the hMSH2 and hMLH1 genes are hotspots for HNPCC mutations. Human Genetics 1996;97:251-5.
38. Mauillon JL et al. Identification of novel germline hMLH1 mutations including a 22 kb Alu- mediated deletion in patients with familial colorectal cancer. Cancer Res. 1996;56:5728-33.
39. Miyaki M et al. Germ line mutations of hMSH2 and hMLH1 genes in Japanese families with HNPCC: Usefulness of DNA analysis for screening and diagnosis of HNPCC patients. Journal of Molecular Medicine 1995;73:515-20.
40. Montera M et al. Mutational germline analysis of hMSH2 and hMLH1 genes in early onset colorectal cancer patients [letter]. Journal of Medical Genetics (Online) 2000;37:E7.
41. Moslein G et al. Microsatellite instability and mutation analysis of hMSH2 and hMLH1 in patients with sporadic, familial and hereditary colorectal cancer. Human Molecular Genetics 1996;5:1245-52.
42. Nakahara M et al. Identification of concurrent germ-line mutations in hMSH2 and/or hMLH1 in Japanese HNPCC kindreds. Cancer Epidemiology, Biomarkers & Prevention 1997;6:1057-64.
43. Nilbert M et al. Microsatellite instability is rare in rectal carcinomas and signifies hereditary cancer. European Journal of Cancer 1999;35:942-5.
44. Nomura S et al. Enhanced detection of deleterious and other germline mutations of hMSH2 and hMLH1 in Japanese hereditary nonpolyposis colorectal cancer kindreds. Biochemical & Biophysical Research Communications 2000;271:120-9.
45. Nystrom-Lahti M et al. DNA mismatch repair gene mutations in 55 kindreds with verified or putative HNPCC. Human Molecular Genetics 1996;5:763-9

46. Papadopoulos N et al. Mutation of a mutL homolog in hereditary colon cancer [see comments]. Science 1994;263:1625-9.
47. Park JG et al. Gene-environment interaction in hereditary nonpolyposis colorectal cancer with implications for diagnosis and genetic testing. International Journal of Cancer 1999;82:516-9.
48. Park JG et al. Suspected hereditary nonpolyposis colorectal cancer: International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC) criteria and results of genetic diagnosis. Diseases of the Colon & Rectum 1999;42:710-5.
49. Pensotti V et al. Mean age of tumor onset in hereditary nonpolyposis colorectal cancer (HNPCC) families correlates with the presence of mutations in DNA hismatch repair genes. Genes, Chromosomes & Cancer 1997;19:135-42.
50. Percesepe A et al. Molecular screening for HNPCC: a prospective, population-based study.[comment]. Journal of Clinical Oncology 2001;19:3944-50.
51. Percesepe A et al. Mismatch repair genes and mononucleotide tracts as mutation targets in colorectal tumors with different degrees of microsatellite instability. Oncogene 1998;17:157-63.
52. Pistorius SR et al. Clinical consequences of molecular diagnosis in families with mismatch repair gene germline mutations. Int.J.Colorectal Dis. 2000;15:255-63.
53. Planck M et al. hMLH1, hMSH2 and hMSH6 mutations in HNPCC families from southern Sweden. International Journal of Cancer 1999;83:197-202.
54. Ravnik-Glavac M, Potocnik U, Glavac D. Incidence of germline hMLH1 and hMSH2 mutations (HNPCC patients) among newly diagnosed colorectal cancers in a Slovenian population. Journal of Medical Genetics 2000;37:533-6.
55. Salovaara R et al. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. Journal of Clinical Oncology 2000;18:2193-200.
56. Samowitz WS et al. The colon cancer burden of genetically defined hereditary nonpolyposis colon cancer. Gastroenterology 2001;121:830-8.
57. Sasaki S et al. Mutational analysis of the hMLH1 gene using an automated two- dimensional DNA typing system. Human Mutation 1997;9:164-71.
58. Scott RJ et al. Hereditary nonpolyposis colorectal cancer in 95 families: differences and similarities between mutation-positive and mutation-negative kindreds. Am.J.Hum.Genet. 2001;68:118-27.
59. Senba S et al. Clinicopathologic and genetic features of nonfamilial colorectal carcinomas with DNA replication errors. Cancer 1998;82:279-85.
60. Swensen J, Lewis CM, Cannon-Albright LA. Identification of a one-base germline deletion (Codon 888 del C) and an intron splice acceptor site polymorphism in hMSH2. Human Mutation 1997;10:80-1.
61. Syngal S et al. Interpretation of genetic test results for hereditary nonpolyposis colorectal cancer: implications for clinical predisposition testing.[comment]. JAMA 1999;282:247-53.
62. Tannergard P et al. Mutation screening in the hMLH1 gene in Swedish hereditary nonpolyposis colon cancer families. Cancer Research 1995;55:6092-6.
63. Tomlinson IPM et al. Germline HNPCC gene variants have little influence on the risk for sporadic colorectal cancer. Journal of Medical Genetics 1997;34:39-42.
64. Vasen HF et al. MSH2 mutation carriers are at higher risk of cancer than MLH1 mutation carriers: a study of hereditary nonpolyposis colorectal cancer families. J.Clin.Oncol. 2001;19:4074-80.
65. Vasen HF et al. Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis [published erratum appears in Gastroenterology 1996 Nov;111(5):1402]. Gastroenterology 1996;110:1020-7.
66. Viel A et al. Characterization of MSH2 and MLH1 mutations in Italian families with hereditary nonpolyposis colorectal cancer. Genes, Chromosomes & Cancer 1997;18:8-18.
67. Wahlberg SS et al. Low frequency of hMSH2 mutations in Swedish HNPCC families [letter]. International Journal of Cancer 1997;74:134-7.
68. Wang Q et al. Germline hMSH2 and hMLH1 gene mutations in incomplete HNPCC families. International Journal of Cancer 1997;73:831-6.
69. Wang Q et al. Prevalence of germline mutations of hMLH1, hMSH2, hPMS1, hPMS2, and hMSH6 genes in 75 French kindreds with HNPCC. Human Genetics 1999;105:79-85.
70. Weber TK et al. Novel hMLH1 and hMSH2 germline mutations in African Americans with colorectal cancer. Journal of the American Medical Association 1999;281:2316-20.
71. Weber TK et al. Genomic DNA-based hMSH2 and hMLH1 mutation screening in 32 Eastern United States HNPCC pedigrees. Cancer Research 1997;57:3798-803.
72. Wehner M et al. Hereditary nonpolyposis colorectal cancer (HNPCC): Eight novel germline mutations in hMSH2 or hMLH1 genes. Human Mutation 1997;10:241-4.
73. Wijnen J et al. Hereditary nonpolyposis colorectal cancer families not complying with the Amsterdam criteria show extremely low frequency of mismatch-repair-gene mutations. American Journal of Human Genetics 1997;61:329-35.
74. Wijnen J et al. MSH2 genomic deletions are a frequent cause of HNPCC [letter]. Nature Genetics 1998;20:326-8.
75. Wijnen JT et al. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. New England Journal of Medicine 1998;339:511-8.
76. Wu Y et al. MSH2 and MLH1 mutations in sporadic replication error-positive colorectal carcinoma as assessed by two-dimensional DNA electrophoresis. Genes, Chromosomes & Cancer 1997;18:269-78.
77. Yuan Y et al. Germline mutations of hMLH1 and hMSH2 genes in patients with suspected HNPCC and sporadic early-onset colorectal cancer. Dis.Col. Rect. 1998;41:434-40.

Appendix A5 Accuracy of Reporting of Family History of Colorectal Cancer

This paper has been reproduced in its published form with the permission of the publishers (BMJ Publishing Group Ltd) and all co-authors.

COLON CANCER

Accuracy of reporting of family history of colorectal cancer

R J Mitchell, D Brewster, H Campbell, M E M Porteous, A H Wyllie, C C Bird, M G Dunlop

Gut 2004;**53**:291–295. doi: 10.1136/gut.2003.027896

See end of article for authors' affiliations

Correspondence to: Professor M G Dunlop, Academic Coloproctology, University of Edinburgh, Edinburgh EH4 2XU, UK; Malcolm.Dunlop@hgu.mrc.ac.uk

Accepted for publication 30 September 2003

Background and aims: Family history is used extensively to estimate the risk of colorectal cancer but there is considerable potential for recall bias and inaccuracy. Hence we systematically assessed the accuracy of family history reported at interview compared with actual cancer experience in relatives.

Methods: Using face to face interviews, we recorded family history from 199 colorectal cancer cases and 133 community controls, totalling 5637 first and second degree relatives (FDRs/SDRs). We linked computerised cancer registry data to interview information to determine the accuracy of family history reporting.

Results: Cases substantially underreported colorectal cancer arising both in FDRs (sensitivity 0.566 (95% confidence interval (CI) 0.433, 0.690); specificity 0.990 (95% CI 0.983, 0.994)) and SDRs (sensitivity 0.271 (95% CI 0.166, 0.410); specificity 0.996 (95% CI 0.992, 0.998)). There was no observable difference in accuracy of reporting family history between case and control interviewees. Control subjects similarly underreported colorectal cancer in FDRs (sensitivity 0.529 (95% CI 0.310, 0.738); specificity 0.995 (95% CI 0.989, 0.998)) and SDRs (sensitivity 0.333 (95% CI 0.192, 0.512); specificity 0.995 (95% CI 0.991, 0.995)). To determine practical implications of inaccurate family history, we applied family history criteria before and after record linkage. Only two of five families reported at interview to meet surveillance criteria did so after validation, whereas only two of six families that actually merited surveillance were identified by interview.

Conclusions: This study has quantified the inaccuracy of interview in identifying people at risk of colorectal cancer due to a family history. Colorectal cancer was substantially underreported and so family history information should be interpreted with caution. These findings have considerable relevance to identifying patients who merit surveillance colonoscopy and to epidemiological studies.

People who have relatives affected by colorectal cancer have an increased personal risk of the disease compared with the general population. The degree of personal risk relates to the extent of family history and age of onset of affected relatives.[1] Thus family history is used in the clinical setting to inform decisions regarding the use of colonoscopic surveillance. Because of the increasing awareness of the genetic contribution to colorectal cancer, in the UK, elsewhere in Europe, and in the USA there has been a rapid increase in colonoscopy workload where family history is the primary concern. Guidelines based on degree of family history have been devised to determine when surveillance should be recommended.[2 3] This empiric approach inherently places considerable importance on the accuracy of family history information. Accuracy is also an important consideration in the context of the epidemiological studies that inform the guidelines for offering surveillance. In both situations, information on family history is usually gathered by interview with a family member. This approach is potentially subject to inaccuracy on the part of the interviewee. Underreporting of family history has being observed in previous studies[4 5] and there is evidence that systematic recall bias may arise from the fact that people with raised awareness of a particular cancer may be more likely to report a positive family history.[6] Furthermore, the social stigma associated with bowel cancer may mean that this condition is discussed less readily within families, and this factor could particularly affect reporting of family history.

Accuracy of reporting cancer in the family has been addressed in previous studies of people referred to genetics departments because of a cancer family history,[7–9] people with a personal history of cancer,[5 6 10–14] or close relatives of cancer cases.[4 15 16] However, only a few studies have related specifically to colorectal cancer cases[4 5 11 12 14] or to community based consultands who have not been referred to a genetics clinic.[5 10] Another limitation of the published literature is that validation of the interviewee's report is often only attempted for relatives reported to have had cancer. In such studies no information can be obtained regarding the sensitivity, specificity, or negative predictive value of reports, and the question of underreporting cannot be addressed.

In this study, information obtained at interview from colorectal cancer cases and community controls was linked systematically to Scottish Cancer Registry data in order to investigate the true accuracy of reporting of a family history of colorectal cancer. We determined the cancer experience of 5637 relatives, irrespective of the reporting of cancer by the interviewee, and so we were able to determine overall accuracy, including underreporting of cancer in relatives. We also evaluated the effect of any inaccuracies on clinical interpretation of family history with respect to recommending surveillance colonoscopy. The findings have considerable relevance to the methods used to validate family history and also have practical implications for surveillance guidelines.

METHODS

A genetics nurse conducted face to face interviews with cases and controls to obtain their reported family history. A total of 199 consecutive colorectal cancer cases were ascertained from Edinburgh Royal Infirmary, Western General Hospital, Edinburgh, and St Johns Hospital, Livingston. For community controls, our initial strategy was to recruit spouses of

Abbreviations: FDR, first degree relatives; SDR, second degree relatives; ISD, Information and Statistics Division

Table 1 Consultand knowledge of all cancer types in relatives

Interviewee group	Relative group	No of relatives	No (%) for whom interviewee could supply health information	No of relatives with confirmed cancer*	No (%) of affected relatives in which cancer was reported	Total No of cancers†	No (%) of cancers accurately reported
Cases	FDR	1322	1250 (95%)	215	152 (71%)	240	106 (44%)
Cases	SDR	1968	713 (36%)	274	84 (31%)	293	42 (14%)
Controls	FDR	1037	991 (96%)	113	76 (67%)	124	51 (41%)
Controls	SDR	1310	671 (51%)	189	77 (41%)	202	36 (18%)

*This column refers to the total number of relatives in a particular group found by ISD linkage to have had cancer.
†This column describes the total number of primary cancers occurring in relatives, including multiple primary cancers.
FDR, first degree relatives; SDR, second degree relatives.

cases. However, this approach proved impractical, and only 25 controls were identified by this means. A further 108 age and sex matched controls were ascertained from general practice lists in North West Edinburgh. Details of all first and second degree relatives (FDRs/SDRs), as reported by the interviewee, were recorded in a structured proforma. A comprehensive manual search of records of births, deaths, and marriages held at the General Register Office for Scotland was performed, in order to verify, correct, and extend pedigree information reported at interview in preparation for record linkage.

Data for all relatives were systematically linked to Scottish Cancer Registry data held by the Information and Statistics Division (ISD) of the Scottish Executive. The Scottish Record Linkage System links all records relating to hospital discharge, cancer registration, and cause of death for each individual, and represents a comprehensive resource for identifying cancer incidence in a given population group. Using techniques based on the principles of "probability matching" developed by Newcombe,[17] such records are linked via patient specific identifying information with a false positive rate of less than 1%.[18–20] Our own internal assessment of colorectal cancer ascertainment is that the false negative rate is also of this order. The same methodology can be applied to linking research data containing personal identifiers with the health information held by ISD. Surname, forename, sex, date of birth, and postcode are commonly used to match records, and our data set contained all but the latter of these. Record linkage served not only to validate reports of cancer but also to identify previously unidentified cases. Confidence intervals (CI) were calculated using an approximation based on inverting an appropriate score test statistic,[21] which compares favourably with exact methods for our data.[22]

Ethics approval for the recruitment and interview of patients and controls was granted by the Lothian Local Research Ethics Committee. The record linkage process was subject to approval from the Privacy Advisory Committee responsible for advising ISD on the release of patient identifiable data. All linked data remained at ISD throughout the analyses to ensure confidentiality.

RESULTS

Mean age of the 199 colorectal cases at the time of interview was 64.0 years. There were 86 females and 113 males, who had a total of 3290 relatives included in the database. One hundred and ten relatives were reported to be resident outside Scotland, and the nurse constructing the pedigrees classified a further 251 as "untraceable". Mean age of the 133 controls was 64.2 years at the time of interview. There were 60 females and 73 males with a total of 2347 relatives. In all there were 107 relatives who were reported to be resident outside Scotland and 91 were deemed to be untraceable. Individuals who have neither died nor developed cancer will not be matched through record linkage, and so it is impossible to distinguish these individuals from those who cannot be traced. Hence all 3290 relatives of cases and 2347 relatives of controls were included in the subsequent record linkage and analysis, regardless of apparent "traceability".

Knowledge of family members' health and occurrence of all types of cancer

Interviewees were asked to state their knowledge as to whether a given relative was alive, and regarding the medical history of relatives, including any history of cancer. The proportion of relatives for which the interviewees were able to provide any health related information is shown in table 1. Table 1 also details the responses given by interviewees for all relatives found to have any type of cancer by linking with central records.

In the majority of instances where a cancer was not correctly reported, the interviewee either had no knowledge of the health of the relative in question or was unaware that they had developed any type of cancer. However, in some cases a cancer was reported but the site was incorrect or unknown. An indication of the extent to which this occurred is provided by the sixth column in table 1, which states the proportion of affected relatives reported to have had any form of cancer.

Reporting of colorectal cancer cases

There were a total of 148 confirmed cases of colorectal cancer in FDRs or SDRs, of which 62 were reported correctly by the interviewee. Mean age at onset of cases that were correctly reported was 63.3 years (95% CI 60.5, 66.1), a value significantly different from the mean age of 70.2 years (95% CI 67.8, 72.5) for cases that were not correctly reported. This observation is not unexpected as cancer affecting more elderly relatives is less likely to be discussed within families. The suggestion that early onset cases are more likely to be reported accurately at interview is of clinical interest as such cases are more significant in terms of indicating increased genetic risk. A separate trend towards more accurate reporting in recent years was evident, although not statistically significant. Summary statistics associated with the accuracy of reporting of colorectal cancer in relatives are presented in table 2.

The data in table 2 demonstrate substantial underreporting of colorectal cancer in relatives. In both cases and controls, sensitivity of reporting in FDRs is approximately 50–60%, implying that a large proportion of cancers in FDRs go unreported. The poor sensitivity of reporting is even more striking in SDRs, with the majority of cases in SDRs of cases and controls not being reported at interview. The very high estimates of specificity and negative predictive value primarily reflect the fact that in absolute terms colorectal cancer affects only a small proportion of the population. However, even small effects on these parameters may have important implications for genetic risk assessment and resource allocation. For all relative groups, estimates of positive predictive value were in the range 60–70%, indicating that approximately one third of reports of individual colorectal

Table 2 Summary statistics for reporting of colorectal cancer

Group	Relative group	No of cases reported at interview	No of cases found by record linkage	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)
Cases	FDR (n = 1322)	43	53	0.566 (0.433, 0.690)	0.990 (0.983, 0.994)	0.698 (0.549, 0.814)	0.982 (0.973, 0.988)
Cases	SDR (n = 1968)	21	48	0.271 (0.166, 0.410)	0.996 (0.992, 0.998)	0.619 (0.409, 0.792)	0.982 (0.975, 0.987)
Controls	FDR (n = 1037)	14	17	0.529 (0.310, 0.738)	0.995 (0.989, 0.998)	0.643 (0.388, 0.837)	0.992 (0.985, 0.996)
Controls	SDR (n = 1310)	16	30	0.333 (0.192, 0.512)	0.995 (0.991, 0.995)	0.667 (0.417, 0.848)	0.985 (0.976, 0.990)

Sensitivity, the proportion of colorectal cancer cases identified by record linkage that were reported at interview; Specificity, the proportion of relatives confirmed by record linkage to be unaffected that were reported to be colorectal cancer free at interview; Positive predictive value, the proportion of reported colorectal cancer cases that were confirmed by record linkage; Negative predictive value, the proportion of individuals reported to be free of colorectal cancer at interview that were confirmed as unaffected by record linkage.
FDR, first degree relatives; SDR, second degree relatives; 95% CI, 95% confidence interval.

cancer cases are not confirmed using cancer registry data. The sensitivity of reporting of colorectal cancer compared with other common cancers is shown in table 3. As no differences were observed between cases and controls in terms of the accuracy of family history reporting, all consultands have been grouped together.

Estimates of sensitivity for colorectal cancer were broadly comparable with the other common cancer types listed in table 3, although numbers were small. However, it is noteworthy that breast cancer was more frequently reported than the other internal cancers in FDRs. This may reflect the more enigmatic presentation of visceral malignancy and the social stigma associated with bowel cancer in particular.

Practical implications of inaccurate or incomplete reporting of family history

From a clinical perspective it is important to determine the validity of interviewee reporting as a means of identifying families that are eligible for colonoscopic surveillance and/or genetic testing. Various guidelines exist to help determine the extent of family history that warrants such interventions, but for illustrative purposes we have applied family history criteria adopted by the British Society of Gastroenterology and the Association of Coloproctology of Great Britain and Ireland (two FDRs with colorectal cancer, or one FDR diagnosed under 45 years).[3] Using these family history criteria, we identified a group of interviewees who merited colonoscopic surveillance. We then re-evaluated the risk categorisation of these individuals based on validated family history data following record linkage.

Again, cases and controls were considered together. In order to gauge the overall impact of inaccurate or incomplete reporting on surveillance recommendations, cases and controls were considered simply as consultands, rather than cases meriting postsurgical surveillance following their own personal history of colorectal cancer. At interview, five of the interviewees reported a family history that met criteria indicating a need for surveillance. However, only two of these five families were confirmed by record linkage to meet these criteria, giving an overall positive predictive value of 0.400 (95% CI 0.118, 0.769). In addition, four further consultands who did not report a family history of colorectal cancer fulfilling criteria actually did have such a family history based on record linkage data. Therefore, only two of six consultands who should have been recommended for surveillance were identified at interview, suggesting that the sensitivity of interview in terms identifying appropriate individuals for surveillance is 0.333 (95% CI 0.097, 0.700).

DISCUSSION

This study has quantified the accuracy of reported family history of cancer in two important groups of people—namely, those with colorectal cancer and those from the general population. Because we confirmed cases reported to have colorectal cancer and also identified cases that had not been reported by the interviewee, we have been able to systematically assess overall accuracy of reported family history of large bowel malignancy.

Using this approach we have determined the accuracy of reporting of colorectal cancer in a large data set comprising 332 interviewees and 5637 first and second degree relatives. We showed conclusively that substantial underreporting of cancer family history is evident in reports made at interview. In this study, the family history documentation was optimal as a trained genetics nurse conducted interviews during a lengthy consultation at the interviewee's home. Reporting inaccuracies may be more extreme where family history is taken in a busy gastroenterology, surgical, or general practice clinic.

Table 3 Sensitivity of interview as a means of identifying familial cancer cases, by site

Site	Relative group	No of cases*	No of cases correctly reported	Sensitivity of interviewee report (95% CI)
Colorectal	FDR	70	39	0.557 (0.441, 0.668)
Colorectal	SDR	78	23	0.295 (0.205, 0.404)
Breast	FDR	28	21	0.750 (0.566, 0.873)
Breast	SDR	37	11	0.297 (0.175, 0.458)
Bronchus and lung	FDR	66	37	0.561 (0.441, 0.674)
Bronchus and lung	SDR	67	10	0.149 (0.083, 0.253)
Stomach	FDR	30	12	0.400 (0.246, 0.577)
Stomach	SDR	64	11	0.172 (0.099, 0.282)

*Where more than one primary cancer occurred at the same site, it was not possible to determine whether the interviewee was aware of both of these tumours. Therefore, where metachronous primary cancers occurred, only the first is considered.
FDR, first degree relatives; SDR, second degree relatives; 95% CI, 95% confidence interval.

A comparable approach to assessing accuracy of reporting of colorectal cancer, which includes identification of unreported cases as well as checking the accuracy of cases reported at interview, has been employed in one previous study.[5] This study estimated the sensitivity of reporting a family history of colorectal cancer in FDRs as 0.65 (95% CI 0.39, 0.85) for colon cancer cases and 0.81 (95% CI 0.54, 0.95) for controls, and the authors concluded that subjects were able to accurately report family history.[5] However, this previous study did not consider SDRs, and no information is provided regarding the total number of relatives involved. Furthermore, the focus of this paper was on validation of an epidemiological study. The observed values for sensitivity of reporting may be less acceptable for genetic risk assessment where the objective is to determine the need for clinical intervention, particularly given the wide confidence intervals.

In general, there is a distinct lack of quality data regarding the accuracy of reporting of family history of colorectal cancer at interview, and the impact of inaccuracy and underreporting on genetic risk assessment has not been evaluated. The current study is thus highly relevant, particularly given the current increase in public demand for information on genetic risk.

We did not observe any difference in the accuracy of family history reporting in cases compared with controls. Similarly, age and sex of interviewee had no significant effect on accuracy. Clearly, the accuracy of reporting of family history by colorectal cancer cases is an important consideration as cancer occurrence is frequently the first point of contact with a particular family. This study addresses the hypothesis that individuals who have had colorectal cancer may be more likely than controls to provide false positive reports of the condition in their relatives. However, we found no evidence to support this hypothesis as there were 21 false positive reports among 199 cases compared with 11 false positive reports among 133 interviewed controls.

Table 1 shows that interviewees could provide no useful information for approximately half of all SDRs but did have some knowledge of the health status of all but approximately 5% of FDRs. This consistent disparity suggests that many instances in which cancer in SDRs goes unreported are due to lack of contact with relatives, rather than ignorance of diagnosis in a known family member. The observation that positive predictive value is similar in FDRs and SDRs lends further support to this notion. Clearly, one would expect that interviewees would have greater knowledge about FDRs, and would be more likely to receive and maintain knowledge of a cancer diagnosis from such close family. Disparity between FDRs and SDRs is evident throughout this study, and is consistent with findings from other published studies.[7–10]

There is some potential for bias within this study but we feel that the effect of such bias is minimal. The total proportion of potential participants who declined to take part in the study, or did not respond to a letter of invitation, was less than 20%. False positive and false negative rates were low for the record linkage process that we used, emphasising the overall validity of our approach. Spouses of cases may be more aware of their own family history of colorectal cancer than the general population, although any such effect would only apply to a small proportion of control subjects. Some mismatching may have occurred, and a proportion of relatives, probably approximately 10%, may have been untraceable. This latter effect would theoretically lead to an underestimation of the positive predictive value. However, no cases and only one control subject reported colorectal cancer in a relative reported to live abroad or deemed to be untraceable, and consequently this effect will have little influence on the reported results.

The accuracy and completeness of cancer registry data itself is a crucial consideration for any study that uses such a resource to validate or confirm diagnoses. The Scottish Cancer Registry was initiated in 1958, and ascertainment was considered to be suboptimal prior to 1968. Although ascertainment of any registry is unlikely to reach 100%, methods of ascertainment have steadily improved since this time, and the Scottish Cancer Registry is considered to be reasonably complete in recent years and to compare favourably with other registries.[23] An evaluation of the accuracy of colorectal cancer registration data found that while misclassifications do occur at a low level, such data exhibit a high degree of accuracy.[24] Colorectal cancer cases occurring prior to the availability of an effective cancer registry were only identified by this study if this malignancy was recorded as a cause of death. Again, this is unlikely to introduce systematic bias, but may have resulted in a slight underestimation of the positive predictive value. Overall, therefore, we consider record linkage with the Scottish Cancer Registry to constitute a reliable and valid means of determining the actual cancer experience of our study subjects. The intermediate use of central records to confirm or correct reported information and to extend knowledge of pedigrees was essential to ensure that study data were of sufficiently high quality for record linkage.

From a clinical perspective, the information provided about the family as a whole is more important than the accuracy of individual reports. The observation in this study that only two of six families who actually met surveillance criteria were identified at interview is a particular concern, implying that reliance on interview data in a clinical context could result in many families who actually meet criteria for significant family history being overlooked. Conversely, of five families reported at interview to meet the chosen criteria, only two were confirmed by record linkage to meet this classification. In practice, such an effect could lead to surveillance being

applied unnecessarily. While the numbers involved are too low to provide a conclusive assessment of the clinical utility of family history information reported at interview, our results illustrate that incomplete or inaccurate interviewee reporting could have a substantial impact on genetic risk assessment.

The appropriate family history criteria for offering genetic counselling, colonoscopic surveillance, or genetic testing is the subject of much current debate, and is likely to remain so. The findings of our study are highly relevant to this discussion, as they suggest that family history information obtained by interview may be misleading, and that verification of both positive and negative interviewee reports should be conducted whenever possible.

ACKNOWLEDGEMENTS

We acknowledge with thanks the considerable expertise of Norma Brown in interviewing the subjects in this study and the follow up genealogy by Rhona de Mey and Alison Fordyce of MRC Registry. We also thank Susan Frame and James Boyd for carrying out the record linkage in ISD, and Niall Anderson for guidance regarding statistical methods. RJM holds a Scottish Health Department Chief Scientist's Office PhD studentship. The work was supported by Cancer Research Campaign (now Cancer Research UK) Program Grant to AHW and CCB–SP1370/0501.

Authors' affiliations

R J Mitchell, H Campbell, Public Health Sciences, Department of Community Health Sciences, University of Edinburgh, Edinburgh, UK
D Brewster, Information and Statistics Division (ISD), Common Services Agency for NHS Scotland, Trinity Park House, Edinburgh, UK
M E M Porteous, Department of Clinical Genetics, University of Edinburgh, Western General Hospital, Edinburgh, UK
A H Wyllie, C C Bird, Department of Pathology, University of Edinburgh, Edinburgh, UK
M G Dunlop, Colon Cancer Genetics Group, University of Edinburgh, Division of Clinical and Molecular Medicine and MRC Human Genetics Unit, Western General Hospital, Edinburgh, UK

REFERENCES

1 **Johns LE**, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. *Am J Gastroenterol* 2001;**96**:2992–3003.
2 **Dunlop MG**, British Society for Gastroenterology, The Association of Coloproctology for Great Britain and Ireland. Guidance on gastrointestinal surveillance for hereditary non-polyposis colorectal cancer, familial adenomatous polypolis, juvenile polyposis, and Peutz-Jeghers syndrome. *Gut* 2002;**51**(suppl 5):V21–7.
3 **Dunlop MG**, British Society for Gastroenterology, The Association of Coloproctology for Great Britain and Ireland. Guidance on large bowel surveillance for people with two first degree relatives with colorectal cancer or one first degree relative diagnosed with colorectal cancer under 45 years. *Gut* 2002;**51**(suppl 5):V17–20.
4 **Glanz K**, Grove J, Le Marchand L, *et al*. Underreporting of family history of colon cancer: correlates and implications. *Cancer Epidemiol Biomarkers Prev* 1999;**8**:635–9.
5 **Kerber RA**, Slattery ML. Comparison of self-reported and database-linked family history of cancer data in a case-control study. *Am J Epidemiol* 1997;**146**:244–8.
6 **Floderus B**, Barlow L, Mack TM. Recall bias in subjective reports of familial cancer. *Epidemiology* 1990;1:318–21.
7 **Love RR**, Evans AM, Josten DM. The accuracy of patient reports of a family history of cancer. *J Chronic Dis* 1985;**38**:289–93.
8 **Sijmons RH**, Boonstra AE, Reefhuis J, *et al*. Accuracy of family history of cancer: clinical genetic implications. *Eur J Hum Genet* 2000;**8**:181–6.
9 **Douglas FS**, O'Dair LC, Robinson M, *et al*. The accuracy of diagnoses as reported in families with cancer: a retrospective study. *J Med Genet* 1999;**36**:309–12.
10 **Koch M**, Hill GB. Problems in establishing accurate family history in patients with ovarian cancer of epithelial origin. *Cancer Detect Prev* 1987;**10**:279–83.
11 **Katballe N**, Juul S, Christensen M, *et al*. Patient accuracy of reporting on hereditary non-polyposis colorectal cancer-related malignancy in family members. *Br J Surg* 2001;**88**:1228–33.
12 **Aitken J**, Bain C, Ward M, *et al*. How accurate is self-reported family history of colorectal cancer? *Am J Epidemiol* 1995;**141**:863–71.
13 **Novakovic B**, Goldstein AM, Tucker MA. Validation of family history of cancer in deceased family members. *J Natl Cancer Inst* 1996;**88**:1492–3.
14 **Kee F**, Collins BJ. How prevalent is cancer family syndrome? *Gut* 1991;**32**:509–12.
15 **Bondy ML**, Strom SS, Colopy MW, *et al*. Accuracy of family history of cancer obtained through interviews with relatives of patients with childhood sarcoma. *J Clin Epidemiol* 1994;**47**:89–96.
16 **Kuijten RR**, Bunin GR, Meadows T. Validation and accuracy of cancer reporting in first-degree relatives (SER abstract). *Am J Epidemiol* 1992;**136**:1019.
17 **Newcombe HB**. *Handbook of record linkage*. New York: OUP, 1988.
18 **Heasman MA**, Clarke JA. Medical record linkage in Scotland. *Health Bull (Edinb)* 1979;**37**:97–103.
19 **Kendrick SW**, Clarke JA. The Scottish Medical Record Linkage System. *Health Bull (Edinb)* 1979;**51**:72–9.
20 **ISD (Information and Statistics Division of the Scottish Executive)**. The Scottish Record Linkage System (ISD information document). Edinburgh: NHS Scotland, 2002.
21 **Wilson EB**. Probable inference, the law of succession, and statistical inference. *JASA* 1927;**22**:209–12.
22 **Agresti A**, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. *Am Stat* 1998;**52**:119–26.
23 **Brewster DH**, Crichton J, Harvey JC, *et al*. Completeness of case ascertainment in a Scottish regional cancer registry for the year 1992. *Public Health* 1997;**111**:339–43.
24 **Brewster D**, Muir C, Crichton J. Registration of colorectal cancer in Scotland: an assessment of data accuracy based on review of medical records. *Public Health* 1995;**109**:285–92.

Appendix A6 Contingency Tables for Reporting of Colorectal Cancer

(i) First degree relatives of Cases

		Confirmed		
		Yes	No	
Reported	Yes	30	13	43
	No	23	1256	1279
		53	1269	1322

Sensitivity = 0.566 (95% CI = 0.433,0.690)
Specificity = 0.990 (95% CI = 0.983,0.994)
Positive Predictive Value = 0.698 (95% CI = 0.549,0.814)
Negative Predictive Value = 0.982 (95% CI = 0.973,0.988)

(ii) Second degree relatives of Cases

		Confirmed		
		Yes	No	
Reported	Yes	13	8	21
	No	35	1912	1947
		48	1920	1968

Sensitivity = 0.271 (95% CI = 0.166,0.410)
Specificity = 0.996 (95% CI = 0.992,0.998)
Positive Predictive Value = 0.619 (95% CI = 0.409,0.792)
Negative Predictive Value = 0.982 (95% CI = 0.975,0.987)

(iii) First degree relatives of Controls

		Confirmed		
		Yes	No	
Reported	Yes	9	5	14
	No	8	1015	1023
		17	1020	1037

Sensitivity = 0.529 (95% CI = 0.310, 0.738)
Specificity = 0.995 (95% CI = 0.989,0.998)
Positive Predictive Value = 0.643 (95% CI = 0.388,0.837)
Negative Predictive Value = 0.992 (95% CI = 0.985,0.996)

(iv) Second degree relatives of Controls

		Confirmed		
		Yes	No	
Reported	Yes	10	6	16
	No	20	1274	1294
		30	1280	1310

Sensitivity = 0.333 (95% CI = 0.192,0.512)
Specificity = 0.995 (95% CI = 0.991,0.995)
Positive Predictive Value = 0.667 (95% CI = 0.417,0.848)
Negative Predictive Value = 0.985 (95% CI = 0.976,0.990)

Appendix A7 Conceptual Model of Cascade Genetic Testing

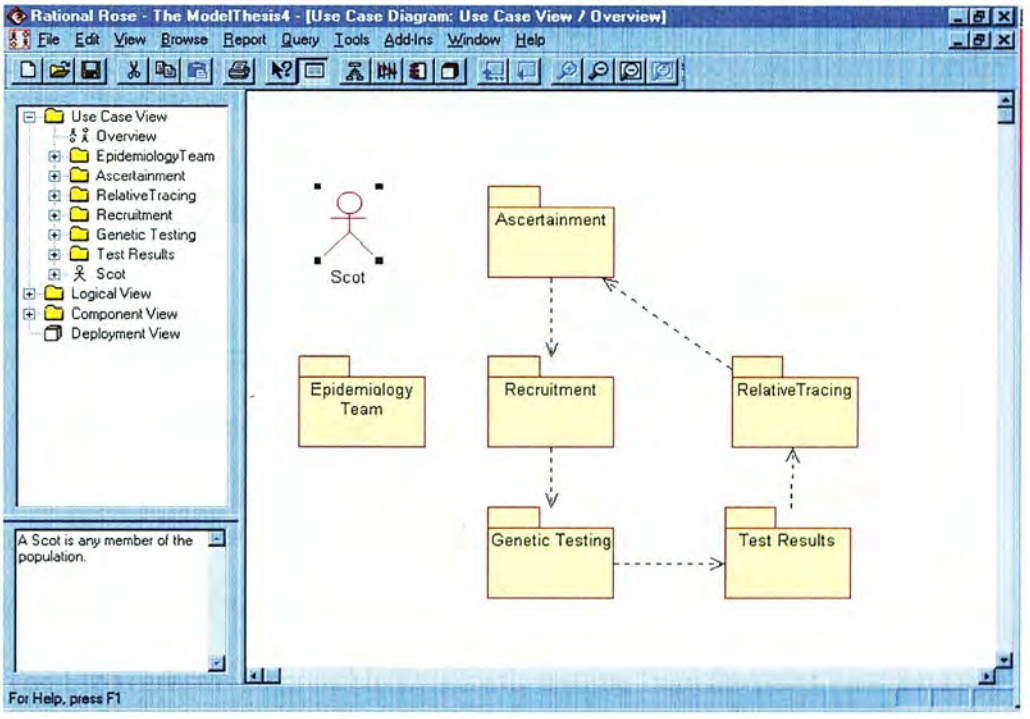
The objective of this appendix is to present the conceptual model in the form of the diagrams created and associated text. It is not feasible to display every element of the conceptual model in this manner, but the following appendix does represent an accurate and largely complete version of the conceptual model.

A7.1 Use Case View

A7.1.1 Overview

The use case view has been split into several "packages" in order to facilitate the organisation and logical presentation of the model. The screenshot shown in figure A7.1 illustrates the packages created, both in the browser view and in diagrammatic form.

Figure A7.1 Overview of Packages Created in Use Case View

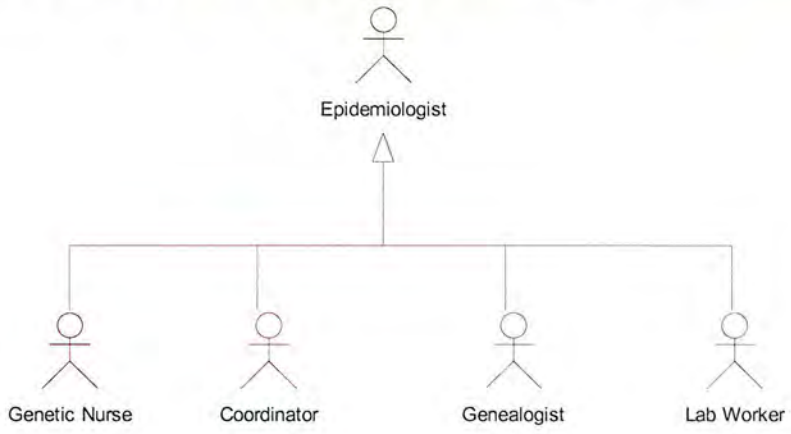


The "Scot" actor is the only actor that is not included in a specific package, and is therefore shown in the overview diagram. This actor represents any member of the Scottish population that interacts with the cascade genetic testing system.

A7.1.2 Epidemiology Team Package

This package is designed to define the actors that are involved in actually running the cascade genetic testing system.

Figure A7.2 Use Case Diagram: Epidemiology Team



A7.1.2.1 Actor: Epidemiologist

This actor is a generalised form of any member of the epidemiology team. The epidemiology team is defined as anyone involved the ascertainment of potential index cases, the genetic testing process, or the tracing and ascertainment of relatives of index cases.

A7.1.2.2 Actor: Coordinator

The "coordinator" has the ultimate responsibility for overseeing the process of ascertainment and genetic testing. It is the coordinator to whom at-risk individuals will be referred, and to whom test results should be provided.

A7.1.2.3 Actor: Genetic Nurse

This actor is responsible for the initial interview (appointment) with an at-risk individual, and also for taking a blood sample and sending it to the LabWorker.

A7.1.2.4 Actor: Lab Worker

The LabWorker is responsible for genetic testing, from the point they receive a sample to the point where they provide the test results to the coordinator.

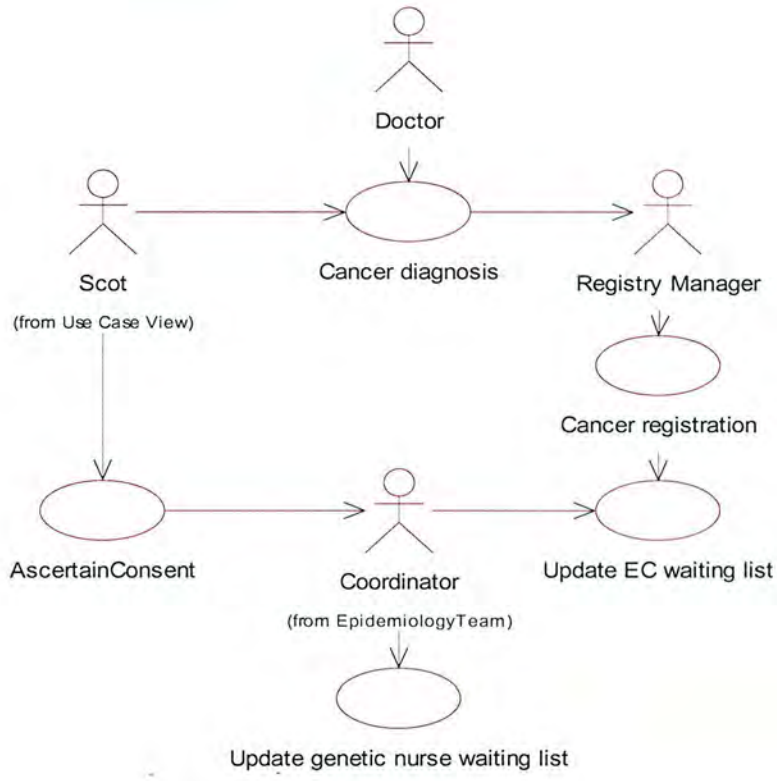
A7.1.2.5 Actor: Genealogist

Not a real person, but provides a useful way of modelling the process of tracing relatives, thus preventing the tasks of the coordinator and genetics nurse from becoming too complicated.

A7.1.3 Ascertainment Package

The ascertainment package contains the actors and use cases involved in the process of identifying Scots who are eligible to participate in the cascade genetic testing system, and including them in the system subject to consent.

Figure A7.3 Use Case Diagram: Ascertainment



A7.1.3.1 Actor: Doctor

This actor is representative of all health professionals involved in the identification of colorectal cancer cases. In reality, this would comprise GPs, Coloproctologists, Pathologists etc, but for the purposes of the model system one actor is sufficient. The main tasks of this actor are to diagnose new colorectal cancer cases and provide the relevant details of these cases to the registry manager.

A7.1.3.2 Use Case: Cancer Diagnosis

Documentation: This use case is initiated when a "Scot" is diagnosed with colorectal cancer. Details of this diagnosis are passed on to the Registry Manager. The actual process of cancer diagnosis could be modelled (and indeed was in early versions), but this is not necessary for the purposes of the model.

Figure A7.4 Sequence Diagram: Cancer Diagnosis

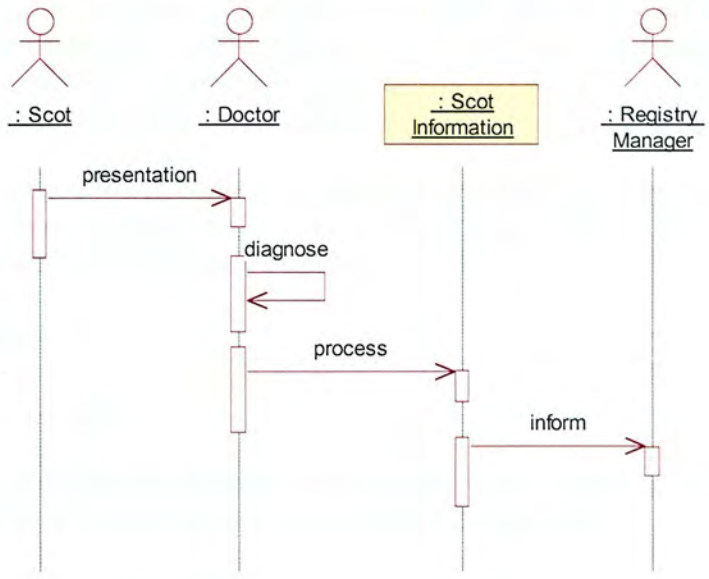
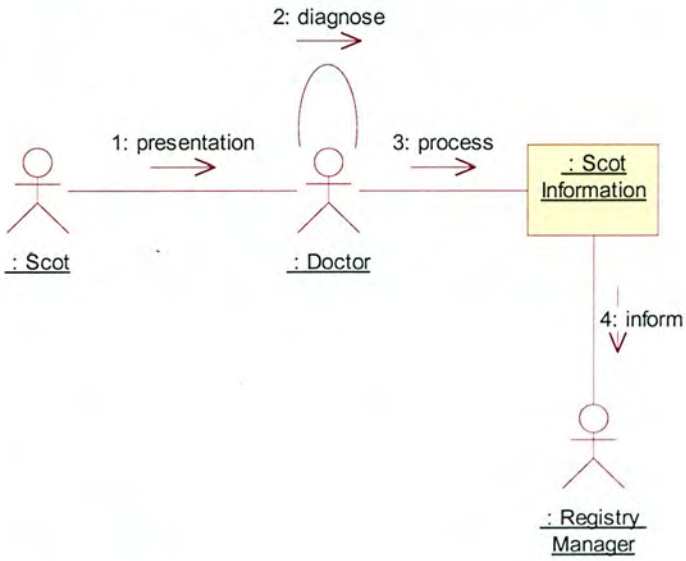


Figure A7.5 Collaboration Diagram: Cancer Diagnosis:



Flow of Events: Cancer Diagnosis

1.1 ***Preconditions***

A colorectal cancer must have developed in the "Scot" in question. This will occur randomly, based on probabilities determined by the age, sex and mutation status of the individual.

1.2 ***Main Flow***

The Use Case begins when a Scot presents to the healthcare system with colorectal cancer. This may be a symptomatic self-referral to a GP, an emergency presentation at a hospital, or the outcome of a screening program such as Faecal Occult Blood Testing; the detail is not relevant to the model. Similarly, diagnoses made at post-mortem are included at this stage, since all tumours should be reported to the cancer registry regardless of the clinical situation.

The "Doctor" actor will process the personal and clinical information relating to the case and will make an official diagnosis of colorectal cancer. This information will then be passed on to the registry manager.

1.3 ***Sub-flows***

1.4 ***Alternative flows***

N.B. Since the actual occurrence of colorectal cancer is essential to initiate this use case, alternative diagnoses are not considered in the model.

E-1: The tumour is mis-diagnosed. The individual in question will not be registered as having colorectal cancer at this point, although they may re-present at a later stage.

E-2: The tumour is diagnosed correctly, but details are not passed on to the registry manager. The individual in question will not be registered as having colorectal cancer and will take no further part in the model system.

A7.1.3.3 Actor: Registry Manager

The Registry Manager is the hypothetical person responsible for maintaining the cancer registry. This is sufficient for the model system, but is an over-simplification of reality, in which the process of maintaining and updating the Scottish Cancer Registry is complex and involves many different individuals and groups.

A7.1.3.4 Use Case: Cancer Registration

Documentation: This use case describes the creation of a colorectal cancer registry (beginning with no entries since all "Scots" are assumed to be cancer-free when the cascade genetic testing system is initiated). The registry will record new cases, and the Epidemiology Coordinator will have access to this resource to identify individuals who are aligible for cascade genetic testing.

This process ensures that the epidemiology coordinator becomes aware of each potential index case that is eligible for the cascade genetic testing programme, regardless of whether the patient themselves have any contact with the system.

Figure A7.6 Sequence Diagram: Cancer Registration

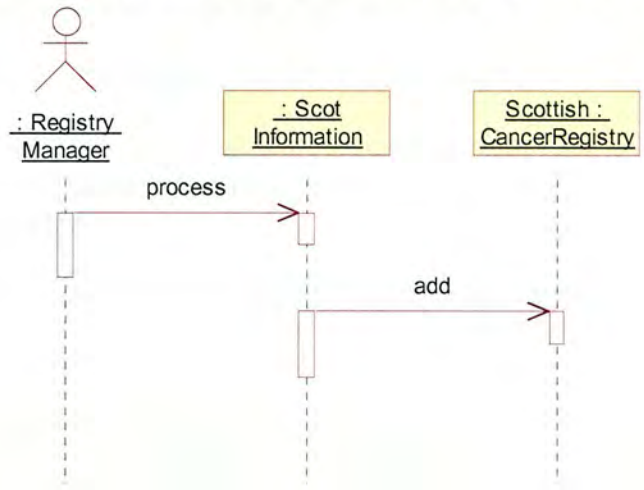
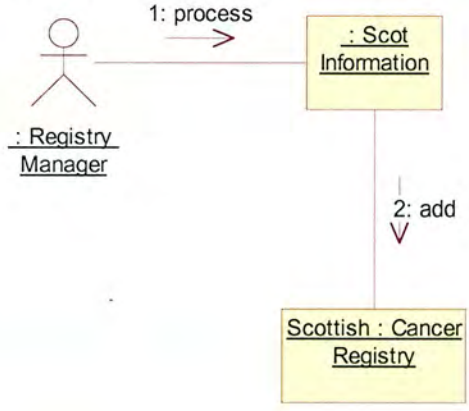


Figure A7.7 Collaboration Diagram: Cancer Registration



Flow of Events: Cancer Registration

1.1 ***Preconditions***

The "CancerDiagnosis" Use Case must have successfully completed.

1.2 ***Main Flow***

Information relating to a new colorectal cancer case is processed by the registry manager. This data is then added to the Scottish Cancer Registry.

1.3 ***Sub-flows***

1.4 ***Alternative flows***

E-1: The relevant information is not added to the Scottish Cancer Registry, or is entered wrongly. This would be an extremely rare occurrence due to clerical error. In

this event, the colorectal cancer case would not be recorded and the individual concerned would take no further part in the system.

A7.1.3.5 Use Case: Update EC Waiting List

Documentation: The epidemiology coordinator will have access to the Scottish Cancer Registry and will update the waiting list of eligible colorectal cancer cases from this source.

Figure A7.8 Sequence Diagram: Update EC Waiting List

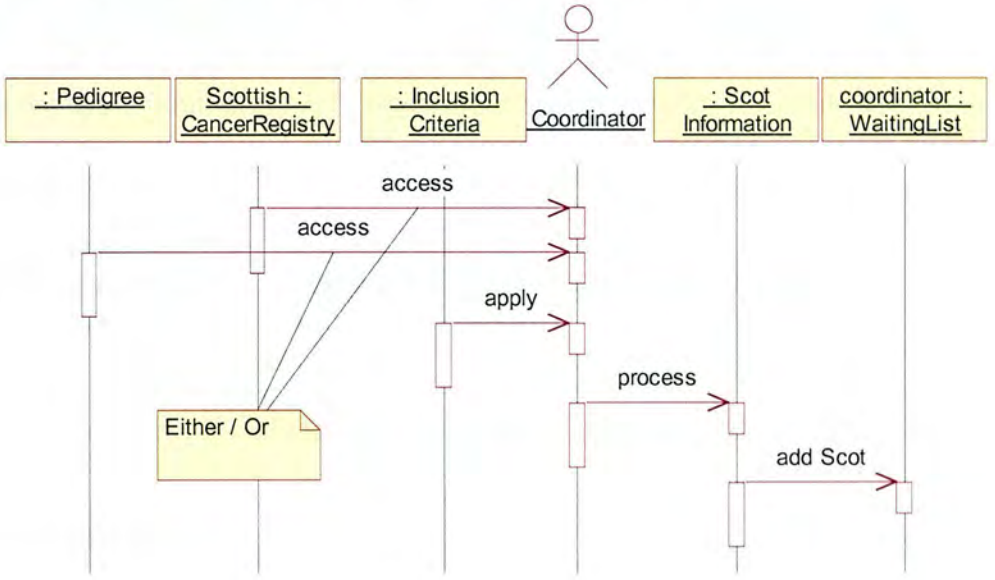
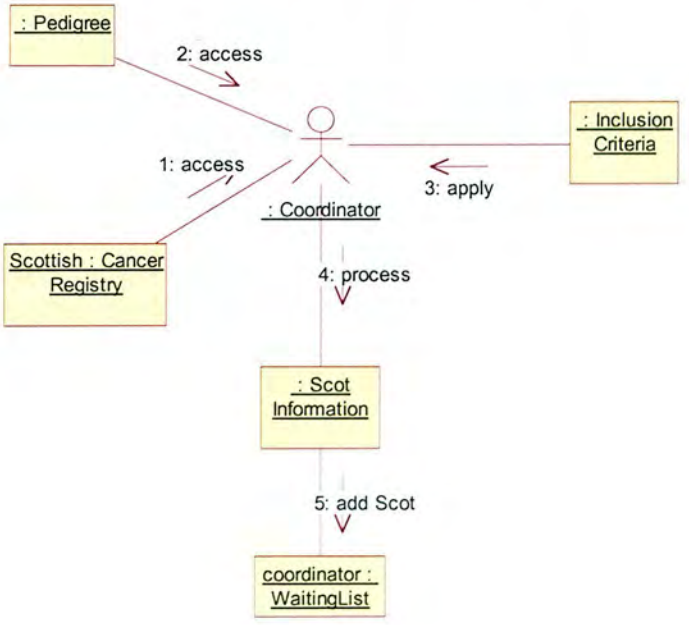


Figure A7.9 Collaboration Diagram: Update EC Waiting List



Flow of Events: Update EC Waiting List

1.1 ***Preconditions***

The Cancer Registration use case must be working correctly, and a new colorectal cancer case must have occurred.

1.2 ***Main Flow***

The Coordinator will access the Scottish Cancer Registry and obtain information on each new colorectal cancer case. They will apply the age limits specified for the cascade genetic testing system. If a colorectal cancer case is identified as falling within these limits, the S-1: Add to Waiting List sub-flow is initiated. If the case is above the upper limit, the S-2: Exclude From System sub-flow is followed.

1.3 ***Sub-flows***

S-1: Add to Waiting List
Personal and clinical data relating to this individual are collated and they are added to the EC waiting list.

S-2: Exclude From System
The case in question is not added to the EC waiting list, and takes no further part in the system.

1.4 ***Alternative flows***

A7.1.3.6 Use Case: Ascertainment Consent

Documentation: The "Consent" use case is a vital one, and can be adapted to several similar situations in the model. At various stages in the model system, the "Scot" in question will be asked to give their consent to proceed to the next stage. The probability that this consent will be forthcoming will be defined by an "Acceptance" look-up table. The principal stages at which consent must be obtained are as follows:

(i) Ascertainment: Consent to participate in the system by attending an interview with the genetic nurse
(ii) Recruitment: Consent to undergo genetic testing
(iii) Tracing Relatives: Consent to allow family members to be contacted in connection with the cascade genetic testing programme

Figure A7.10 Sequence Diagram: Ascertainment Consent

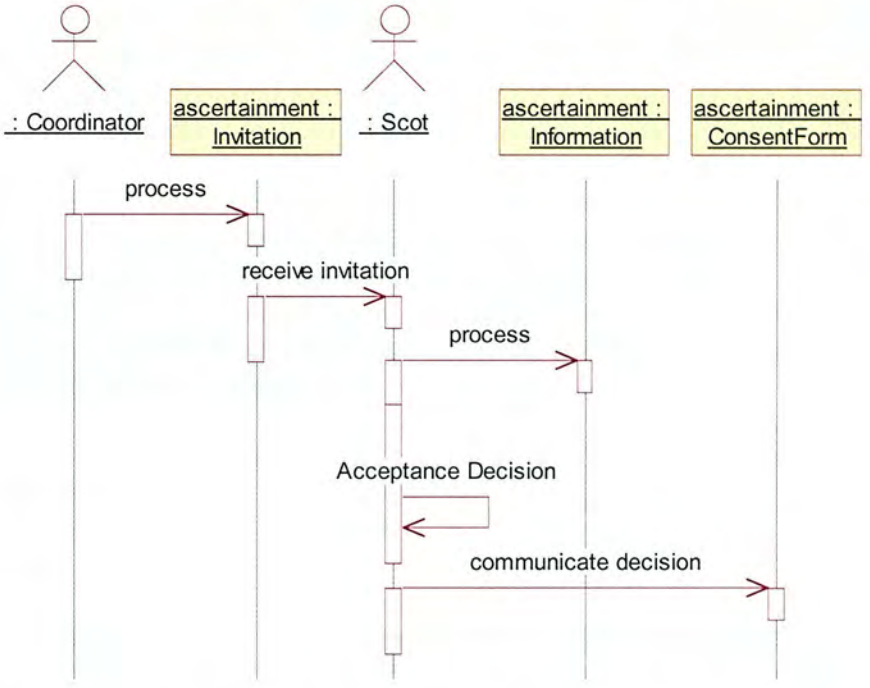
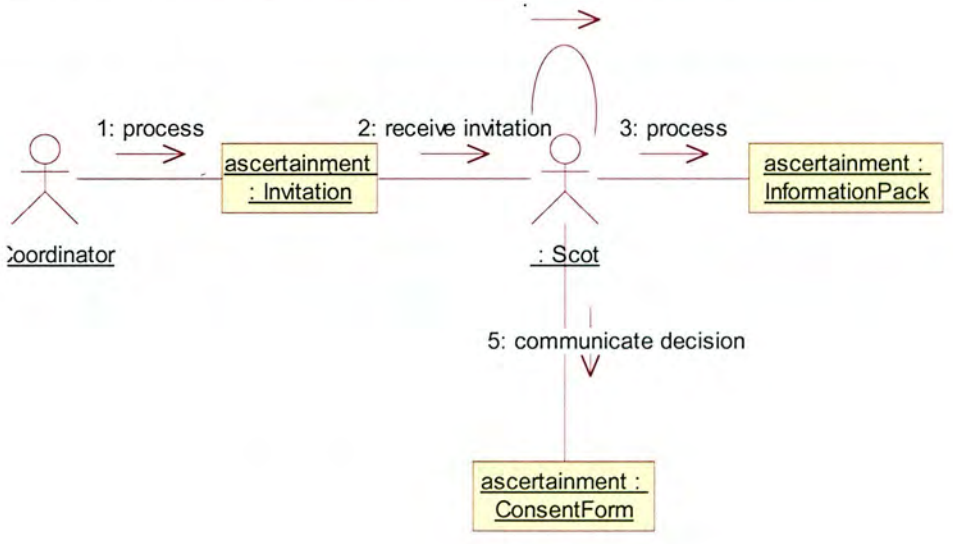


Figure A7.11 Collaboration Diagram: Ascertainment Consent



Flow of Events: Ascertainment Consent

1.1 ***Preconditions***

The "Ascertain Invitation" use case must have been completed successfully.

1.2 ***Main Flow***

This use case is begun by a "Scot" who has developed colorectal cancer, upon receipt of an invitation letter from the Coordinator asking them to participate in the cascade genetic testing programme by attending an interview with a genetic nurse. The Scot in question will study the information pack and then consider whether to take part. They will then complete a "consent form" specifying their decision. If they agree to take part the S-1: Accept sub-flow will be applied, if they do not the S-2: Decline sub-flow will be used.

This protocol (invitation – information – consideration – decision) is written on the assumption that the approach will be made by mail, but it is equally valid if contact is made through health care professionals (e.g. research nurses). In either case, the "consent form" represents an expression of desire, or otherwise, to participate, and is not necessarily a formal or legal form.

1.3 ***Sub-flows***

S-1: Accept

The Scot is prepared to participate in the next stage of the cascade genetic testing system (i.e. attend a genetic nurse interview), and completes a consent form to that effect.

S-2: Decline

The Scot does not wish to participate in the next stage of the cascade genetic testing system, and completes a consent form to that effect.

1.4 ***Alternative flows***

E-1: The Scot does not respond to the invitation. This can occur at any stage, as there is no guarantee that a definitive decision will be reached at interview. In this scenario the individual in question is assumed to have declined the invitation and accordingly takes no further part in the system.

A7.1.3.7 Use Case: Update Genetic Nurse Waiting List

Documentation: This use case is initiated by the epidemiology coordinator. When an eligible colorectal cancer patient accepts an invitation to participate in the cascade genetic testing system (this acceptance does not necessarily include genetic testing), the epidemiology coordinator will add their name and details to the genetic nurse waiting list. The genetic nurse will have access to this list and will use it to identify interviewees and arrange appointments with them.

Figure A7.12 Sequence Diagram: Update Genetic Nurse Waiting List

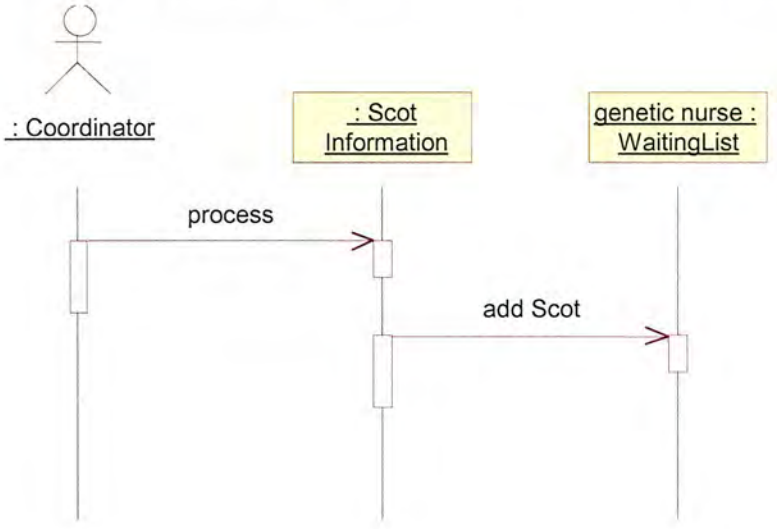
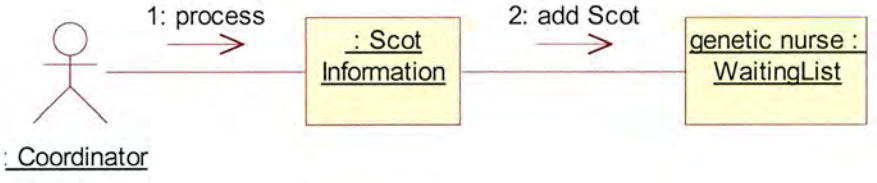


Figure A7.13 Collaboration Diagram: Update Genetic Nurse Waiting List



Flow of Events: Update Genetic Nurse Waiting List

1.1 ***Preconditions***

The "Consent" Use Case relating to the ascertainment of colorectal cancer cases must have executed; i.e. the Scot in question must have given consent to attend an appointment with a genetic nurse

1.2 ***Main Flow***

This use case begins when the Epidemiology Coordinator identifies a colorectal cancer case that meets the criteria for genetic testing, and has consented to attend a genetic nurse interview. Relevant data on this individual, including name, date of birth, date of diagnosis etc. must be collated and processed, and then recorded in the genetic nurse waiting list.

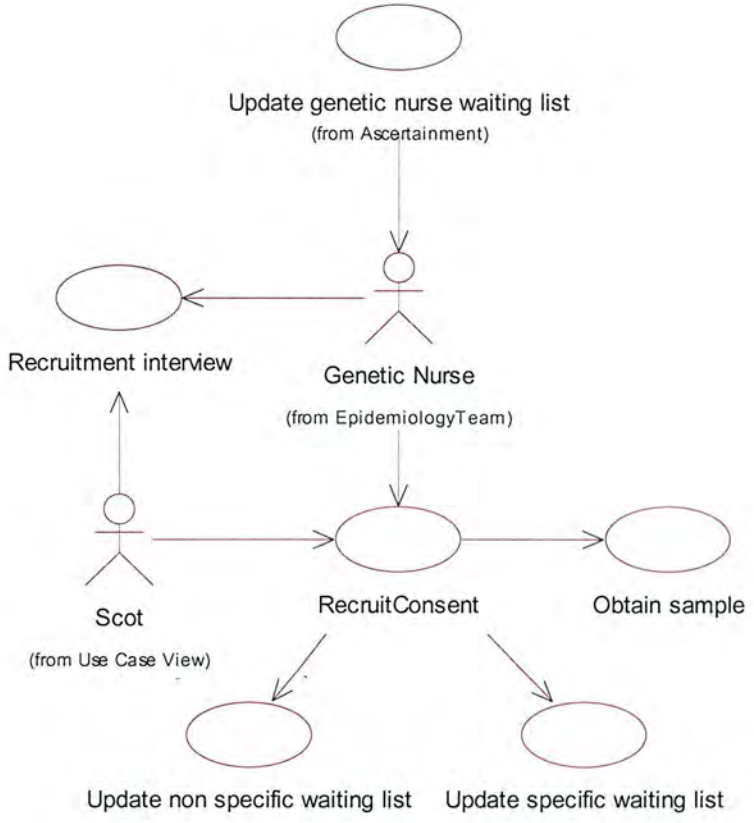
1.3 ***Sub-flows***

1.4 ***Alternative flows***

A7.1.4 Recruitment Package

The recruitment package includes all use case and actors involved with the process of recruiting an eligible Scot to the cascade genetic testing programme, and offering genetic testing.

Figure A7.14 Recruitment: Use Case Diagram



A7.1.4.1 Use Case: Recruitment Interview

Documentation: This use case describes the process of an interview with a Scot who is eligible for genetic testing, as conducted by a GeneticsNurse. The Scot may be a potential index case or a relative of a known mutation carrier.

Figure A7.15 Sequence Diagram: Recruitment Interview

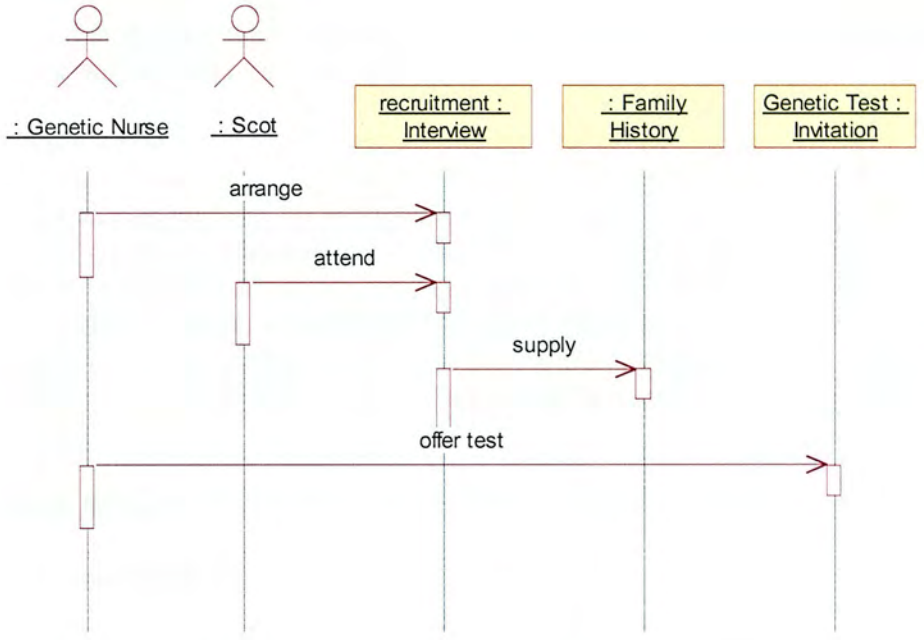
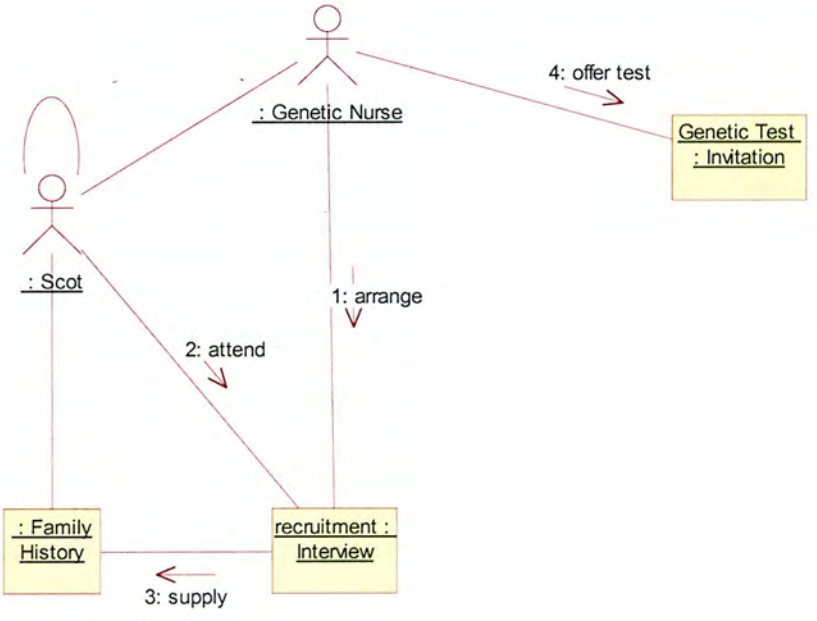


Figure A7.16 Collaboration Diagram: Recruitment Interview



Flow of Events: Recruitment Interview

1.1 ***Preconditions***

The Update genetic nurse waiting list use case must be working correctly, and there must be at least one new entry on this list.

1.2 ***Main Flow***

The use case is initiated by the genetic nurse. Details of an eligible Scot are obtained from the genetic nurse waiting list. The genetic nurse will then arrange a time and venue for the interview, and the Scot in question will attend this appointment. During the interview, more information will be provided to the Scot about the cascade genetic testing programme, and they will be invited to participate further by undergoing a genetic test. The Scot will also provide some information about their family history at this stage.

1.3 ***Sub-flows***

1.4 ***Alternative flows***

E-1: The Scot does not attend the interview. Another appointment will usually be made, and if that is not attended the individual in question will be removed from the genetic nurse waiting list and will take no further part in the system.

E-2: The Scot decides not to attend the interview and informs the genetic nurse of this. Either another appointment will be made, or they will be removed from the genetic nurse waiting list, according to their wishes.

A7.1.4.2 Use Case: Recruitment Consent

Documentation: The "Consent" use case is a vital one, and can be adapted to several similar situations in the model. At various stages in the model system, the "Scot" in question will be asked to give their consent to proceed to the next stage. The probability that this consent will be forthcoming will be defined by an "Acceptance" look-up table. The principal stages at which consent must be obtained are as follows:

(i) Ascertainment: Consent to participate in the system by attending an interview with the genetic nurse
(ii) Recruitment: Consent to undergo genetic testing
(iii) Tracing Relatives: Consent to allow family members to be contacted in connenction with the cascade genetic testing programme

Figure A7.17 Sequence Diagram: Recruitment Consent

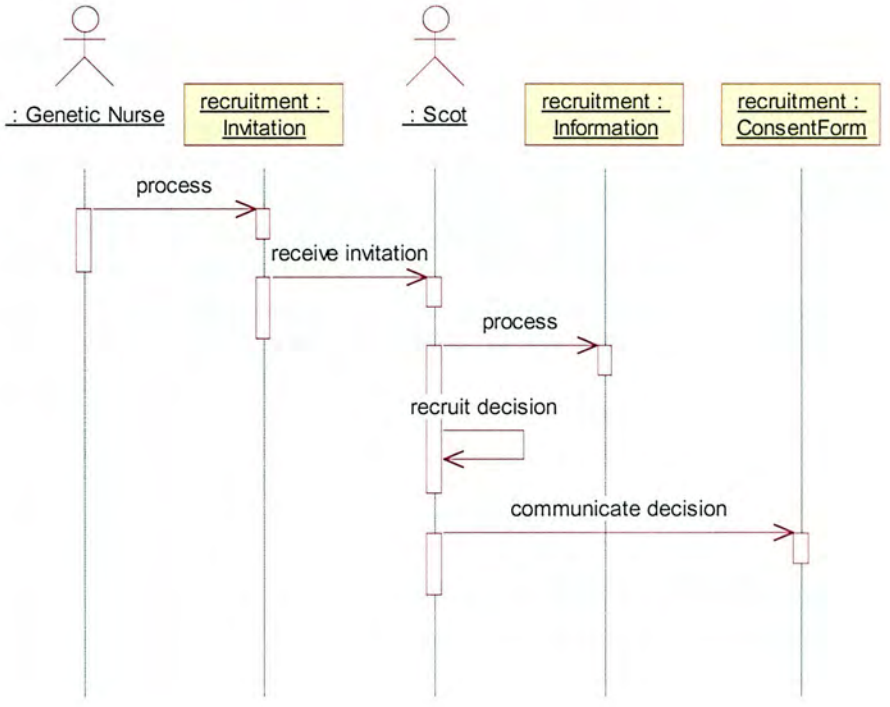
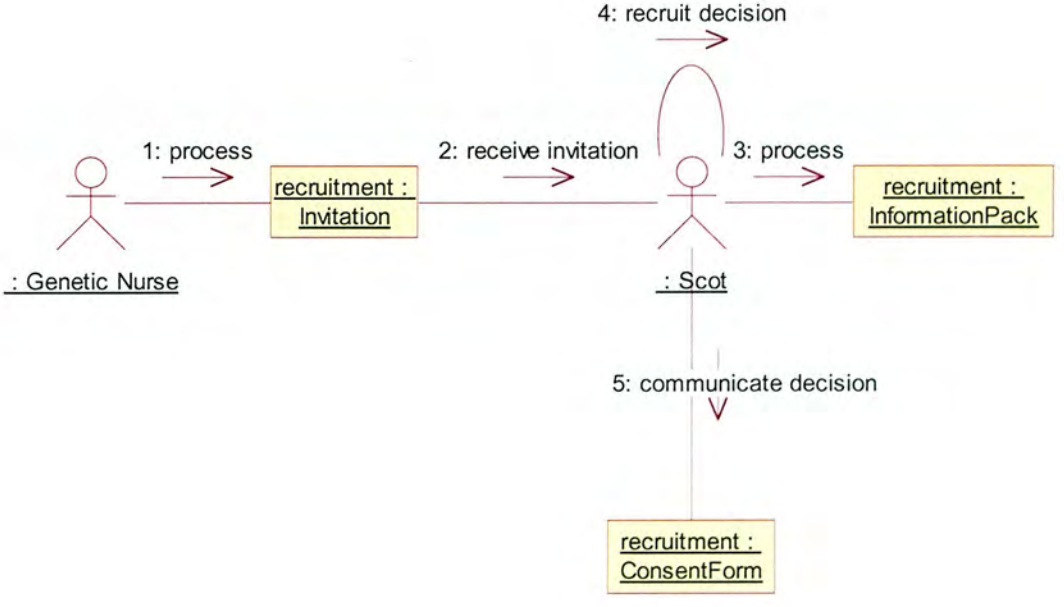


Figure A7.18 Collaboration Diagram: Recruitment Interview



Flow of Events: Recruit Consent

1.1 ***Preconditions***

The "Recruit Invitation" use case must have been completed successfully.

1.2 ***Main Flow***

This use case is begun by a "Scot" who has developed colorectal cancer, upon being invited to undergo genetic testing by a genetic nurse. The Scot in question will study the information pack (i.e. obtain and process the relevant information from the genetic nurse) and then consider whether to take part. They will then complete a "consent form" specifying their decision. If they agree to take part the S-1: Accept sub-flow will be applied, if they do not the S-2: Decline sub-flow will be used. The "consent form" represents an expression of desire, or otherwise, to participate, and is not necessarily a formal or legal form.

1.3 ***Sub-flows***

S-1: Accept

The Scot is prepared to participate in the next stage of the cascade genetic testing system (i.e. undergo genetic testing), and completes a consent form to that effect.

S-2: Decline

The Scot does not wish to participate in the next stage of the cascade genetic testing system, and completes a consent form to that effect.

1.4 ***Alternative flows***

E-1: The Scot does not respond to the invitation. This can occur whether the invitation is made by post or in person, as there is no guarantee that a definitive decision will be reached at interview. In this scenario the individual in question is assumed to have declined the invitation and accordingly takes no further part in the system.

A7.1.4.3 Use Case: Obtain Sample

Documentation: this use case describes the process of obtaining a blood sample from a recruit, in order to conduct genetic testing. The DNA sample is originally obtained at the recruitment stage by the genetic nurse. It is then passed on to the Lab Worker, who will perform the appropriate genetic test.

Figure A7.19 Sequence Diagram: Obtain Sample

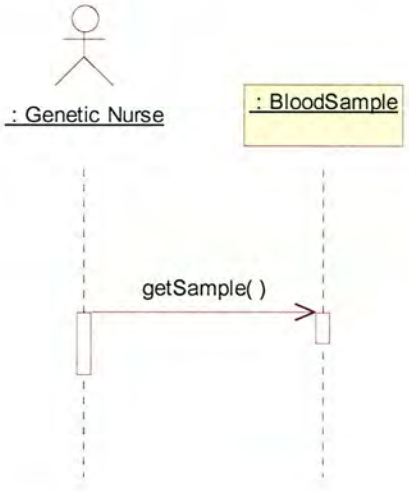
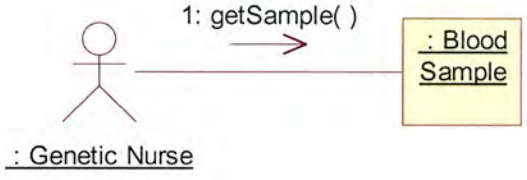


Figure A7.20 Collaboration Diagram: Obtain Sample



Flow of Events: Obtain Sample

1.1 ***Preconditions***

The S-1: Consent to Testing sub-flow of the RecruitInterview use case must have executed, and the S-1 Accept sub-flow of the RecruitConsent use case must also have been successfully completed.

1.2 ***Main Flow***

The Scot will supply a blood sample for the purpose of genetic testing.

1.3 ***Sub-flows***

1.4 ***Alternative flows***

A7.1.4.4 Use Case: Update Non-Specific Waiting List

Documentation: This use case is initiated by the genetic nurse. When a potential index case (i.e. a colorectal cancer patient who meets the specified criteria for inclusion in the cascade genetic testing system) accepts the invitation to undergo genetic testing,

their details will be placed on this waiting list and the blood sample that they provided willbe sent to the Lab Worker. The non-specific test, or "mutation analysis" will then be performed in due course, with the time-scale dependent on the specified time needed to perform such a test and the resources available.

Figure A7.21 Sequence Diagram: Update Non-Specific Waiting List

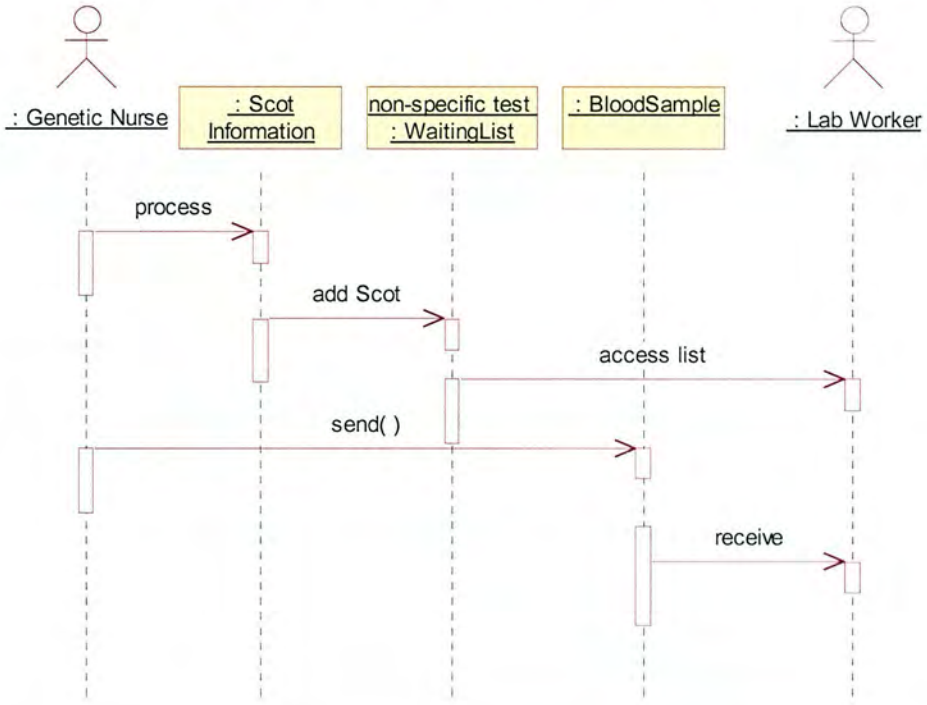
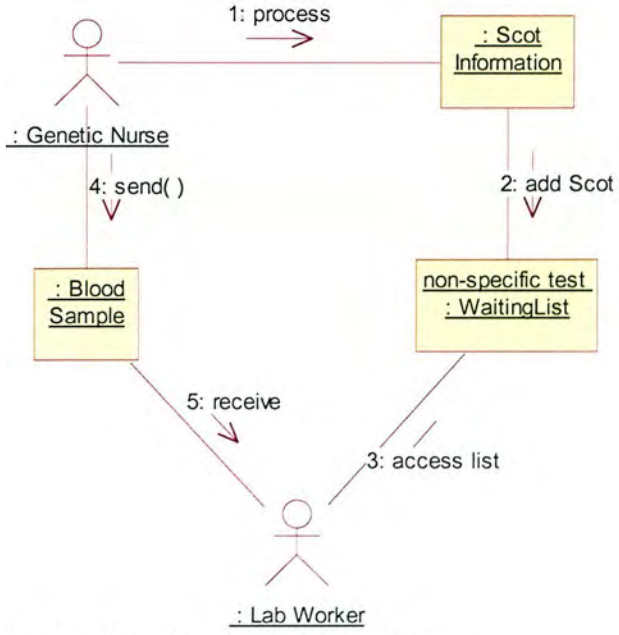


Figure A7.22 Collaboration Diagram: Update Non-Specific Waiting List



Flow of Events: Update Non-Specific Waiting List

1.1 ***Preconditions***

The S-1: Accept sub-flow of the Consent use case must have executed in the context of recruitment of a potential index case by a genetic nurse (i.e. a Scot with colorectal cancer must have consented to undergo genetic testing). The Obtain Sample use case must also have completed.

1.2 ***Main Flow***

The Genetic Nurse will update the relevant ScotInformation with information obtained at the recruitment interview. The updated information will then be added to the non-specific waiting list. The blood sample will also be sent to the LabWorker.

1.3 ***Sub-flows***

1.4 ***Alternative flows***

N.B. Since this use case is controlled entirely within the cascade genetic testing system it will be assumed to operate correctly.

A7.1.4.5 Use Case: Update Specific Waiting List

Documentation: This use case is initiated by the genetic nurse. When a relative of a known mutation carrier accepts the invitation to undergo genetic testing, their details will be placed on this waiting list and the blood sample that they provided will be sent to the Lab Worker. The specific test will then be performed in due course, with the time-scale dependent on the specified time needed to perform such a test and the resources available.

Figure A7.23 Sequence Diagram: Update Specific Waiting List

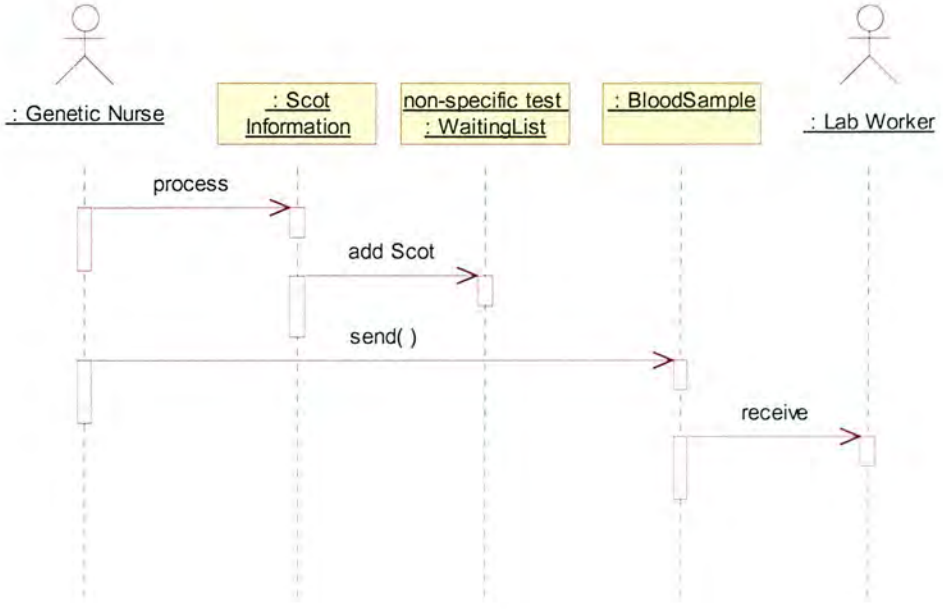
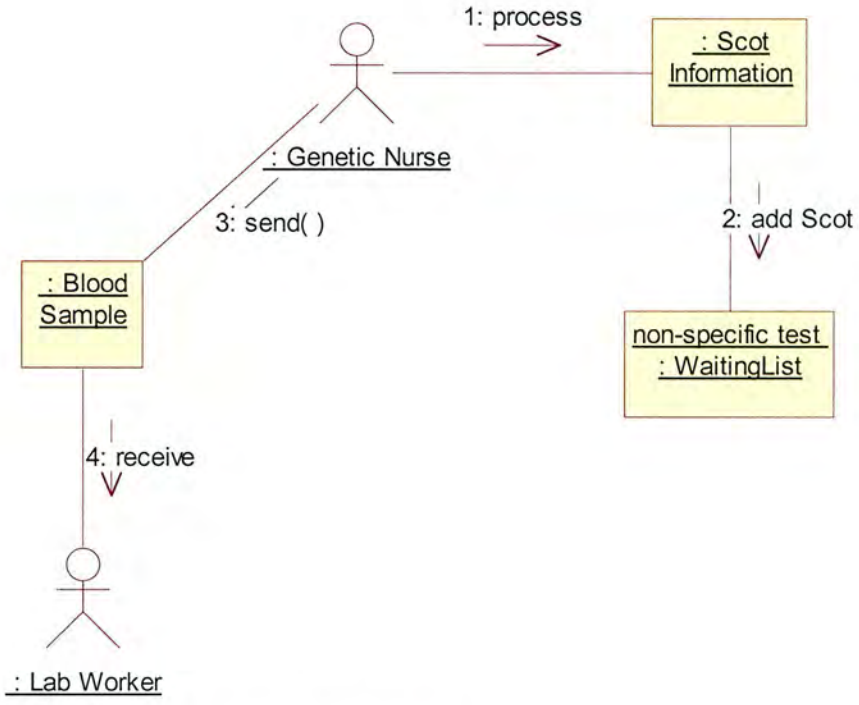


Figure A7.24 Collaboration Diagram: Update Specific Waiting List



Flow of Events: Update Specific Waiting List

1.1 ***Preconditions***

The S-1: Accept sub-flow of the Consent use case must have executed in the context of recruitment of a relative of a known mutation carriers by a genetic nurse (i.e. such a relative must have consented to undergo genetic testing). The Obtain Sample use case must also have completed.

1.2 ***Main Flow***

The Genetic Nurse will update the relevant ScotInformation with information obtained at the recruitment interview. This updated information will then be added to the specific test waiting list. The blood sample will then be sent to the LabWorker.

1.3 ***Sub-flows***

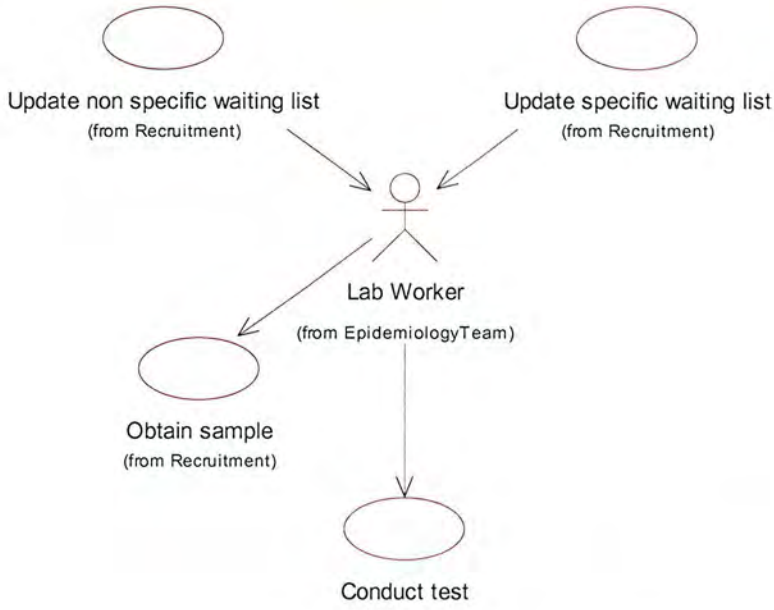
1.4 ***Alternative flows***

N.B. Since this use case is controlled entirely within the cascade genetic testing system it will be assumed to operate correctly.

A7.1.5 Genetic Testing Package

This package contains all use cases and actors involved in the process of conducting a genetic test.

Figure A7.25 Use Case Diagram: Genetic Testing Overview



A7.1.5.1 Use Case: Conduct Test

Documentation: The process of conducting the required molecular analyses. For relatives of known carriers this will consist of a test for a specific mutation. For suspected index cases this will consist of the analysis protocol defined as part of the model. These two options will have very different test characteristics (sensitivity etc.), and should probably be recorded in the flow of events as two distinct sub-flows. Details of these analyses will vary according to the variables set by the model.

Figure A7.26 Sequence Diagram: Conduct Test

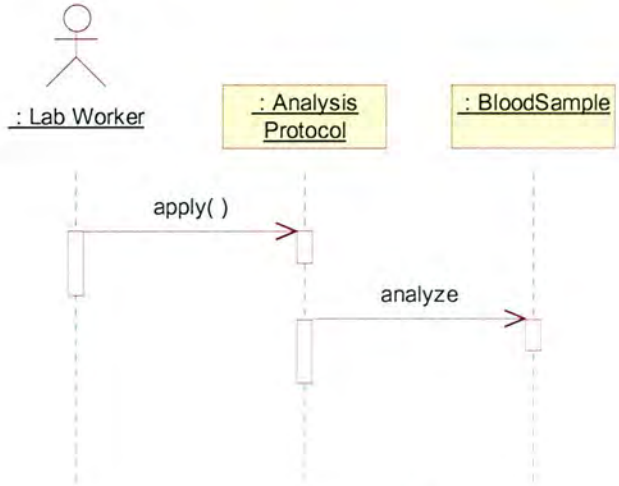
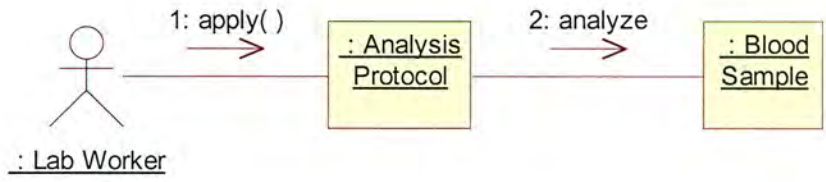


Figure A7.27 Collaboration Diagram: Conduct Test



Flow of Events: Conduct test

1.1 ***Preconditions***

The main flow of the Send Sample use case must have completed

1.2 ***Main Flow***

The molecular analyses will be undertaken by the LabWorker. When the BloodSample in question has come from a potential index case, the S-1 "Mutation Analysis" sub-flow will be performed. When it has come from a relative of a known index case the S-2 "Specific Test" subflow will be performed.

1.3 ***Sub-flows***

S-1: Mutation Analysis

When the BloodSample is from a suspected index case with no known carrier relatives, mutation analysis will be carried to investigate whether or not a mutation is present. The required AnalysisProtocol, and the associated test parameters, will be defined by the user of the model.

S-2: Specific Test

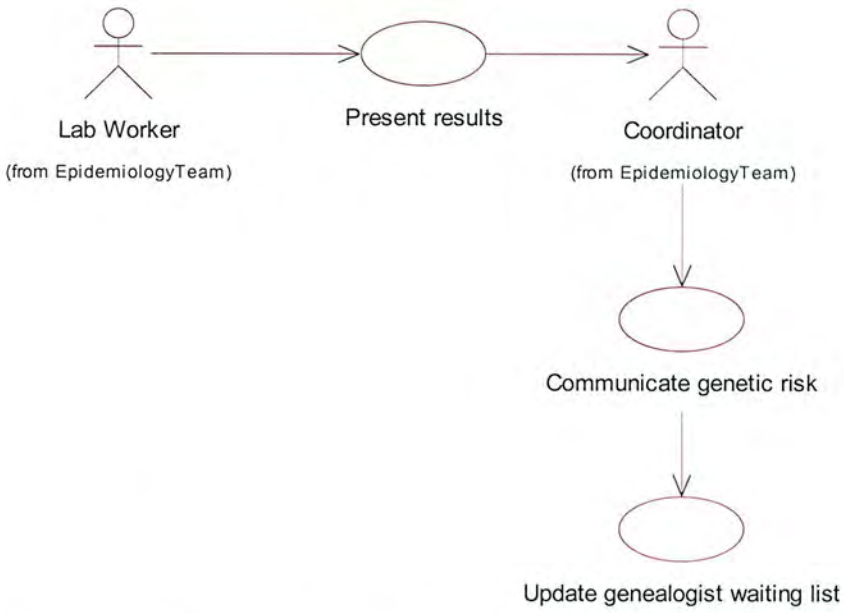
When the Blood Sample is from a relative of a known mutation carrier, the only test required will be a simple PCR-based test to determine if the relative has the same mutation. If they do not it will not be necessary to conduct more detailed mutation analysis: the individual will be classified as test negative. Because of the simple nature of this test, sensitivity and specificity will be very high, but flexibility in this respect should be built into the model regardless.

1.4 Alternative flows

A.7.1.6 Test Results Package

This packages contains actors and use cases relating to the processing and communication of genetic test results

Figure A7.28 Use Case Diagram: Test Results



A7.1.6.1 Use Case: Present Results

Documentation: In this use case, the results referred to are the actual product of molecular analysis. This includes the actual sequence of DNA, plus the results of any other relevant tests that may be applied in the specific protocol used by the model system. Obviously, the most important information will be details of any potential mutations.

Figure A7.29 Sequence Diagram: Present Results

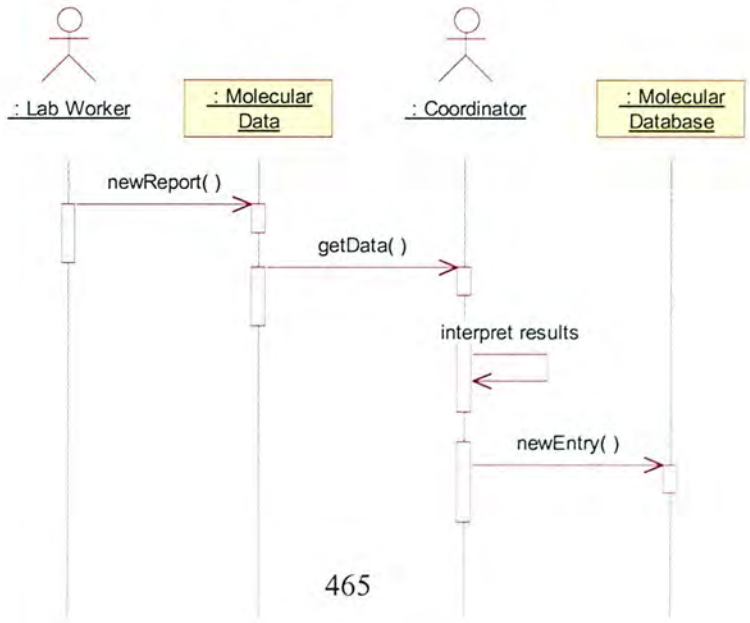
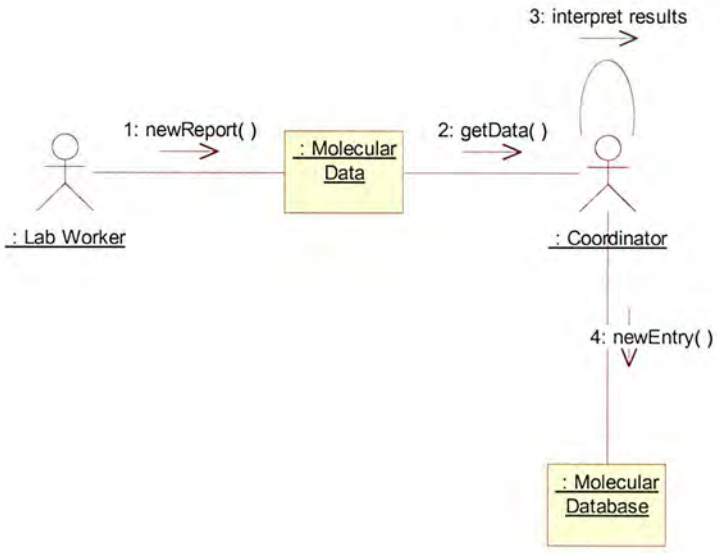


Figure A7.30 Collaboration Diagram: Present Results



Flow of Events: Present Results

1.1 ***Preconditions***

One of the sub-flows of the Conduct Test use case must have executed (i.e. a Scot must have undergone genetic testing).

1.2 ***Main Flow***

The molecular data arising from the genetic test will be collated, processed, and presented to the Co-ordinator. The co-ordinator will then interpret the results and will record the molecular data in a Molecular Database.

1.3 ***Sub-flows***

1.4 ***Alternative flows***

A7.1.6.2 Use case: Communicate Genetic Risk

Documentation: This use case describes the communication of the genetic testing results to the tested individual, their coloproctologist, and the genealogist responsible for tracing relatives, if appropriate.

Figure A7.31 Sequence Diagram: Communicate Genetic Risk

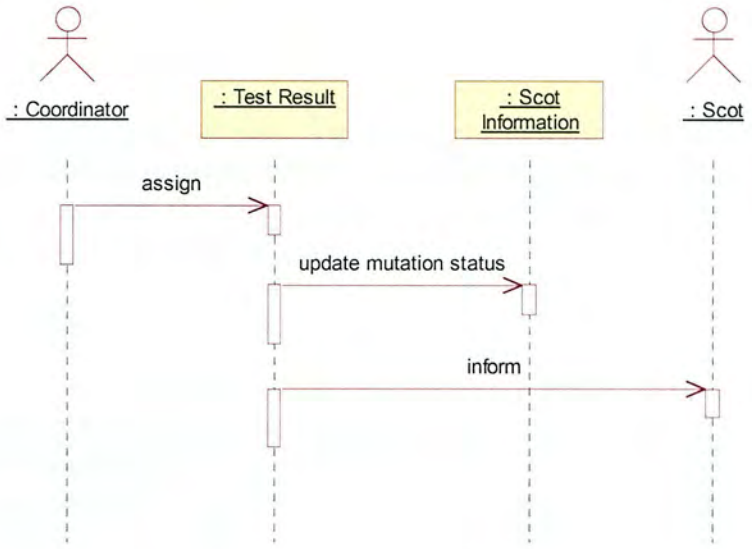
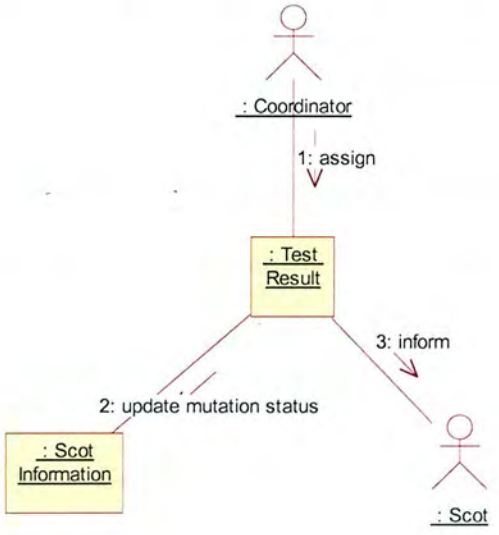


Figure A7.32 Collaboration Diagram: Communicate Genetic Risk



Flow of Events: Communicate Genetic Risk

1.1 ***Preconditions***

The Present Results use case must have been successfully completed.

1.2 ***Main Flow***

The coordinator will study the molecular data arising from a genetic test, and will assign a test result on this basis. If a pathogenic mismatch repair gene mutation is

evident, the S-1 "assign carrier status" subflow will be performed. If no such mutation is found, the S-2 "reassure" sub-flow will be performed.

1.3 ***Sub-flows***

S-1: Assign Carrier Status

The molecular data indicates that the Scot in question carries a MMR gene mutation. The Scot will be informed of this and offered genetic counselling and clinical surveillance. They will also be eligible for the next stage of the cascade genetic testing system, i.e. relative tracing.

S-2: Reassure

The molecular data does not indicate the presence of a MMR gene mutation. The Scot in question will therefore be reassured in this regard. It should be emphasized that failure to find a mutation does not necessarily mean that they, or their family, are not at increased genetic risk. Family history information etc. must also be taken into consideration when making decisions regarding clinical surveillance. However, the Scot will be considered mutation-free for the purposes of cascade genetic testing, and will take n further part in the system.

1.4 ***Alternative flows***

E-1: The test may be inconclusive, in which case it will be repeated or the S-2 sub-flow will be applied with added emphasis on the possibility of the Scot being at increased genetic risk.

A7.1.6.3 Use Case: Update Genealogist Waiting List

Documentation: This use case is initiated by the epidemiology coordinator. When a MMR gene mutation carrier is identified (positive genetic test), the epidemiology coordinator will add this information to the genealogist waiting list. The genealogist will have access to this list and will use it to identify interviewees and arrange appointments with them.

Figure A7.33 Sequence Diagram: Update Genealogist Waiting List

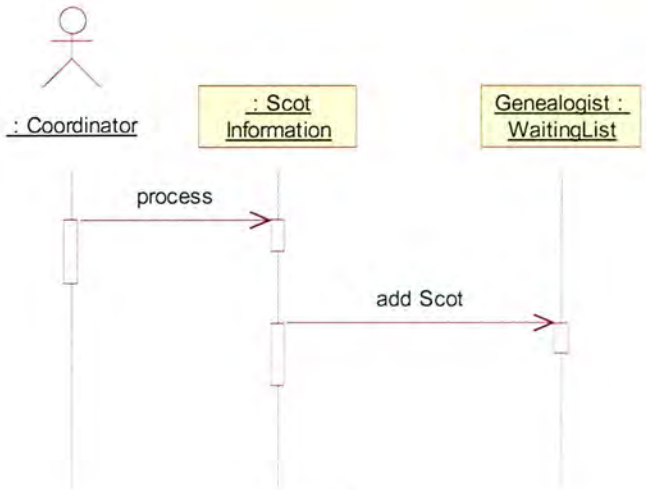
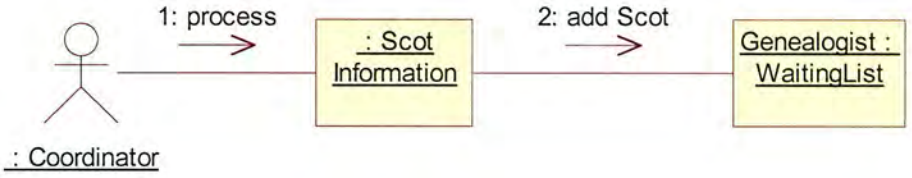


Figure A7.34 Collaboration Diagram: Update Genealogist Waiting List



Flow of Events: Update Genealogist Waiting List

1.1 ***Preconditions***

The S-1: Assign Carrier Status sub-flow of the Communicate Genetic Risk use case must have been successfully completed.

1.2 ***Main Flow***

This use case begins when a new mutation carrier has been identified, and is initiated by the Coordinator. Relevant data on this carrier, including name, date of birth, details of mutation etc. are collated and processed, and then recorded in the genealogist waiting list.

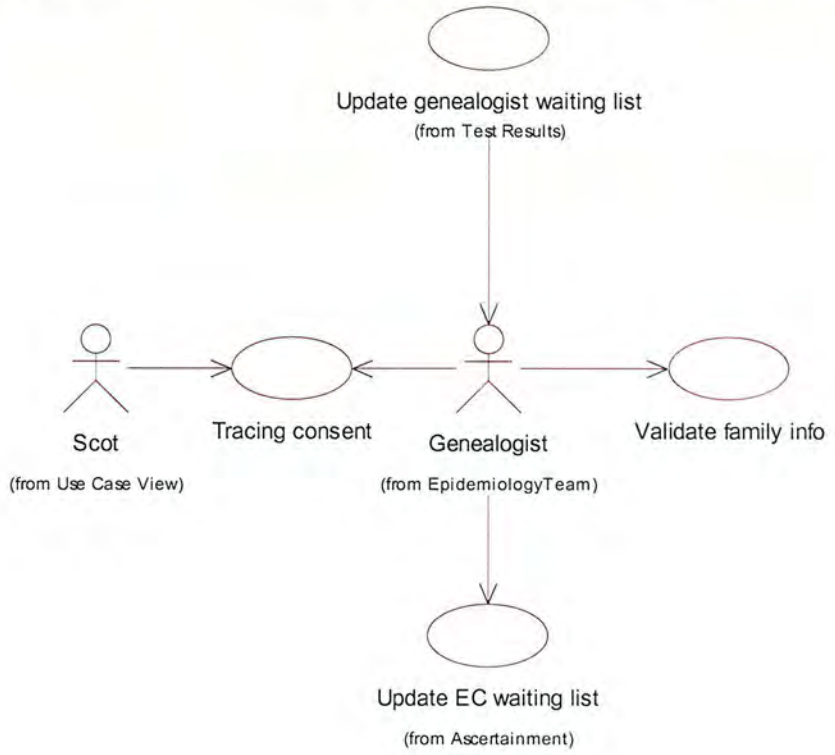
1.3 ***Sub-flows***

1.4 ***Alternative flows***

A7.1.7 Relative Tracing Package

This package is concerned with process of identifying relatives of a newly identified mutation carrier who are eligible for genetic testing.

Figure A7.35 Use Case Diagram: Relative Tracing



A7.1.7.1 Use Case: Trace Consent

Documentation: The "Consent" use case is a vital one, and can be adapted to several similar situations in the model. At various stages in the model system, the "Scot" in question will be asked to give their consent to proceed to the next stage. The probability that this consent will be forthcoming will be defined by an "Acceptance" look-up table. The principal stages at which consent must be obtained are as follows:

(i) Ascertainment: Consent to participate in the system by attending an interview with the genetic nurse
(ii) Recruitment: Consent to undergo genetic testing
(i) Tracing Relatives: Consent to allow family members to be contacted in connenction with the cascade genetic testing programme

Figure A7.36 Sequence Diagram: Trace Consent

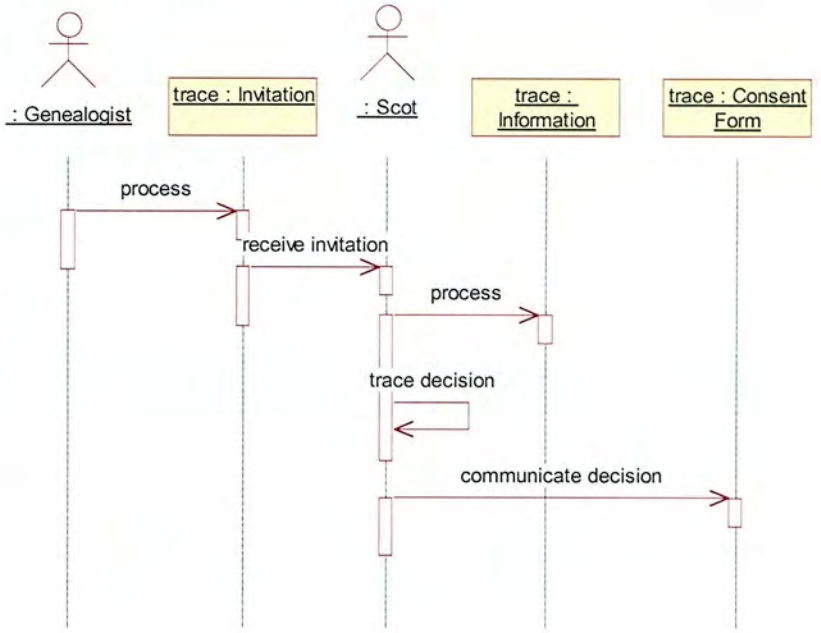
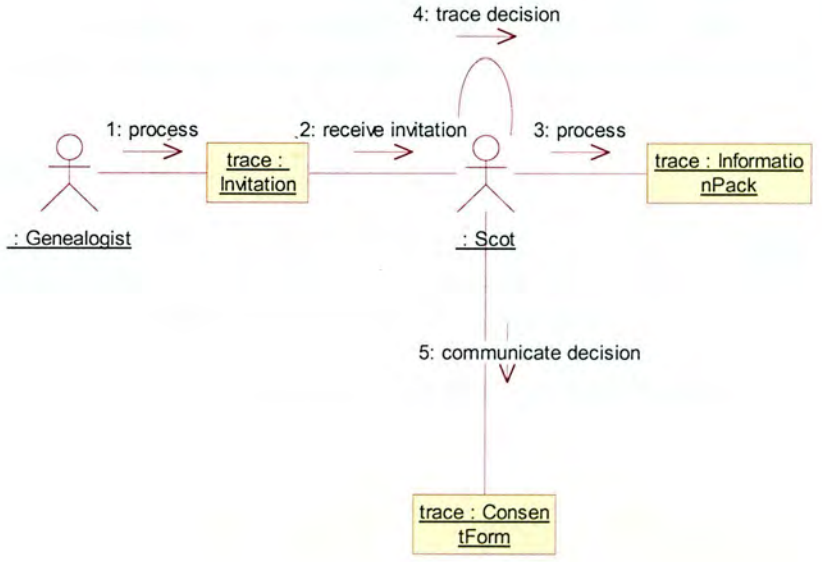


Figure A7.37 Collaboration Diagram: Trace Consent



Flow of Events: Tracing Consent

1.1 ***Preconditions***

The "Recruit Invitation" use case must have been completed successfully.

1.2 ***Main Flow***

This use case is begun by a "Scot" who has developed colorectal cancer, upon being invited to undergo genetic testing by a genetic nurse. The Scot in question will study the information pack (i.e. obtain and process the relevant information from the genetic nurse) and then consider whether to take part. They will then complete a "consent form" specifying their decision. If they agree to take part the S-1: Accept sub-flow will be applied, if they do not the S-2: Decline sub-flow will be used. The "consent form" represents an expression of desire, or otherwise, to participate, and is not necessarily a formal or legal form.

1.3 ***Sub-flows***

S-1: Accept

The Scot is prepared to participate in the next stage of the cascade genetic testing system (i.e. allow contact with relatives), and completes a consent form to that effect.

S-2: Decline

The Scot does not wish to participate in the next stage of the cascade genetic testing system, and completes a consent form to that effect.

1.4 ***Alternative flows***

E-1: The Scot does not respond to the invitation. This can occur whether the invitation is made by post or in person, as there is no guarantee that a definitive decision will be reached at interview. In this scenario the individual in question is assumed to have declined the invitation and accordingly takes no further part in the system.

A7.1.7.2 Use Case: Validate Family Info

Documentation: This use case is concerned with the further investigation and validation of relative information, using such death records, registry entries and medical notes as may be used under ethical regulations.

Figure A7.38 Sequence Diagram: Validate Family Info

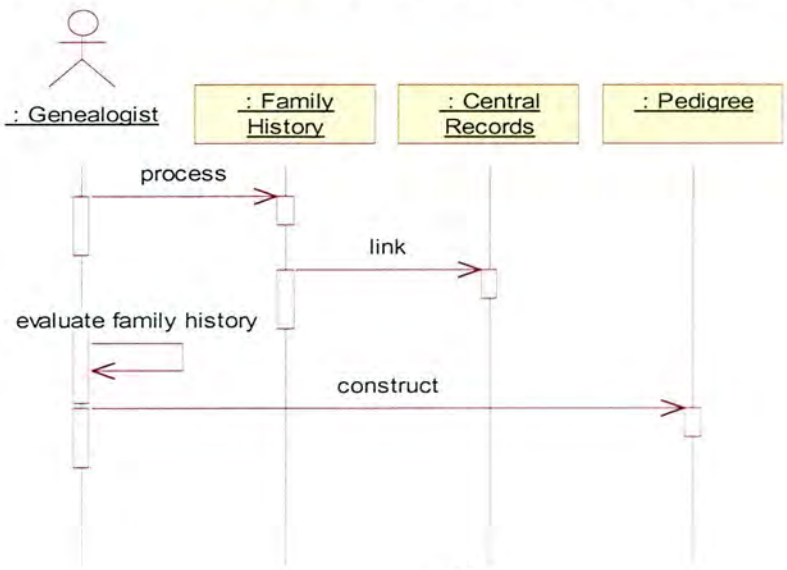
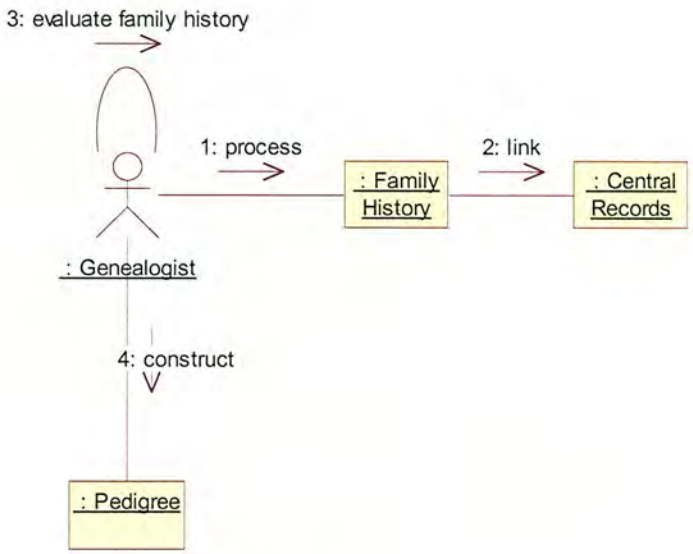


Figure A7.39 Collaboration Diagram: Validate Family Info



Flow of Events: Validate Family Info

1.1 ***Preconditions***

The S-1: Assign Carrier Status sub-flow of the communicate genetic risk use case must have been completed, and the Update Genealogist Waiting List use case must be working correctly (i.e. a new mutation carrier must have been identified and information on this carrier must have been passed to the genalogist).

1.2 ***Main Flow***

Using the central records such as records of births, deaths and marriage, the genealogist will undertake to extend and verify the family history information provided by the Scot in question at the genetic nurse interview. Based on all information sources available under ethical regulations, the genealogist will then construct a pedigree for that individual.

1.3 ***Sub-flows***

1.4 ***Alternative flows***

A7.2 Logical (Class) View

A7.2.1 Class Diagrams

Figure A7.40 Overview of Logical View

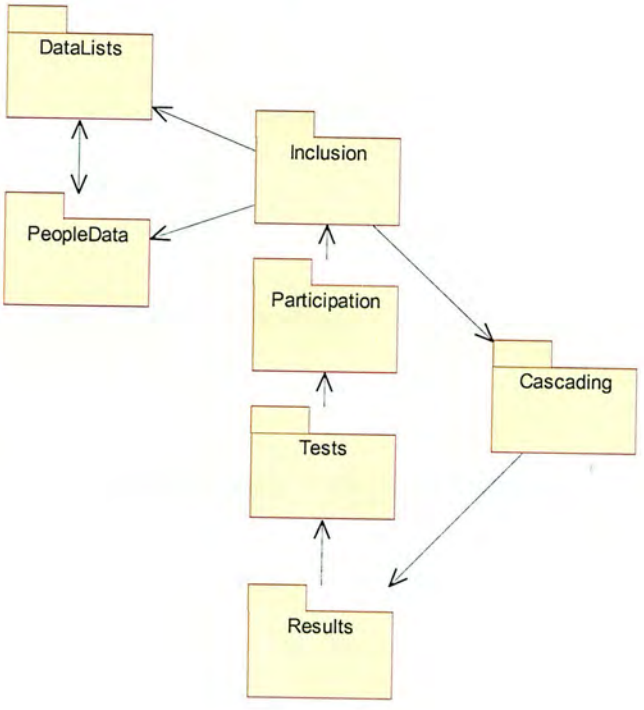


Figure A7.41 Class Diagram: DataLists Package

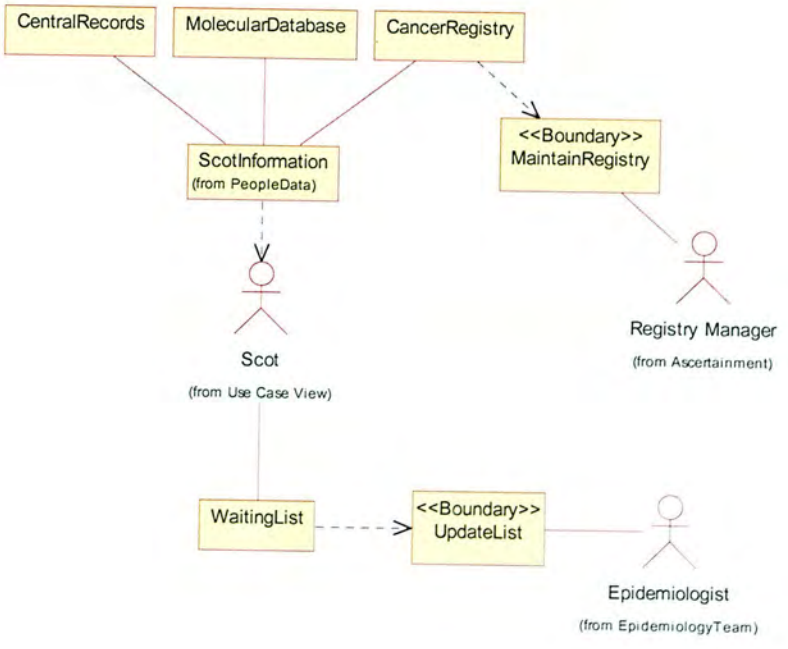


Figure A7.42 Class Diagram: PeopleData Package

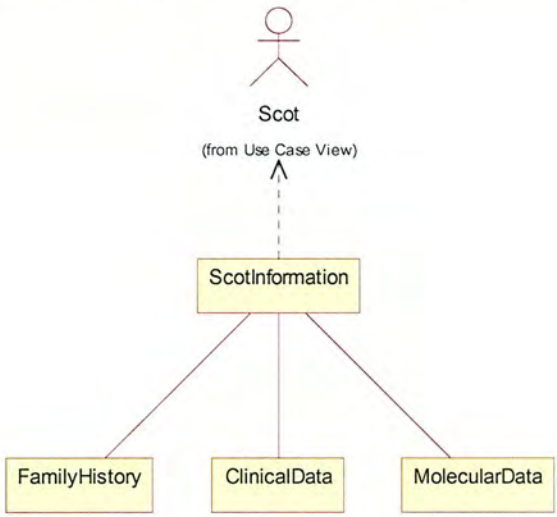


Figure A7.43 Class Diagram: Inclusion Package

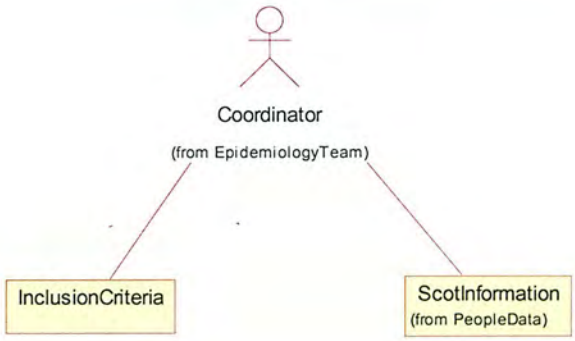


Figure A7.44 Class Diagram: Participation Package

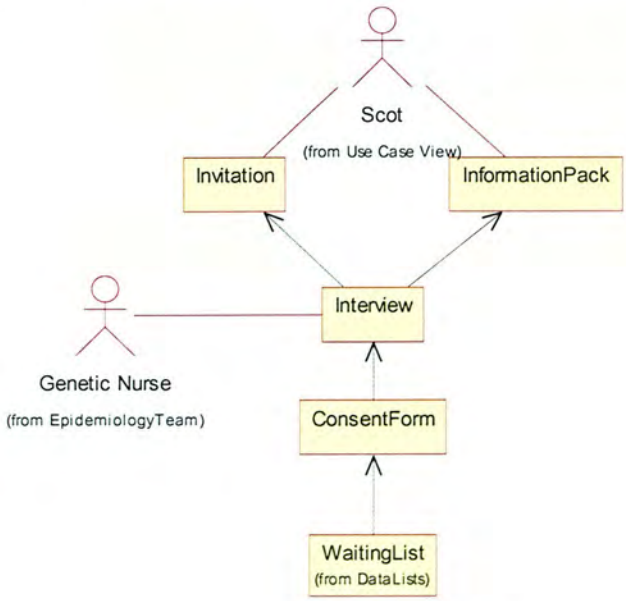


Figure A7.45 Class Diagram: Tests Package

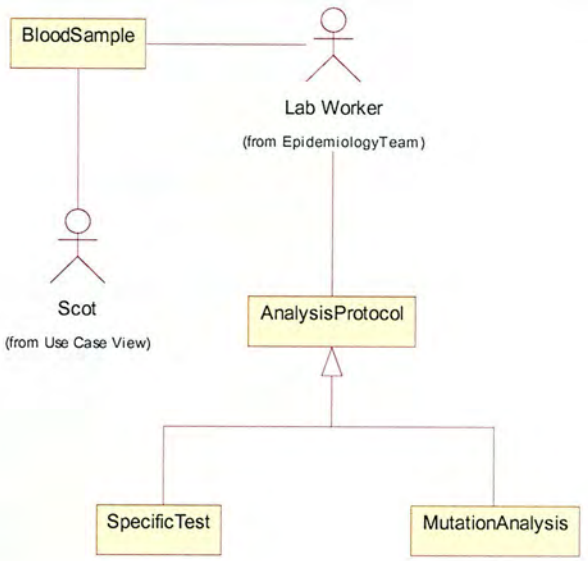
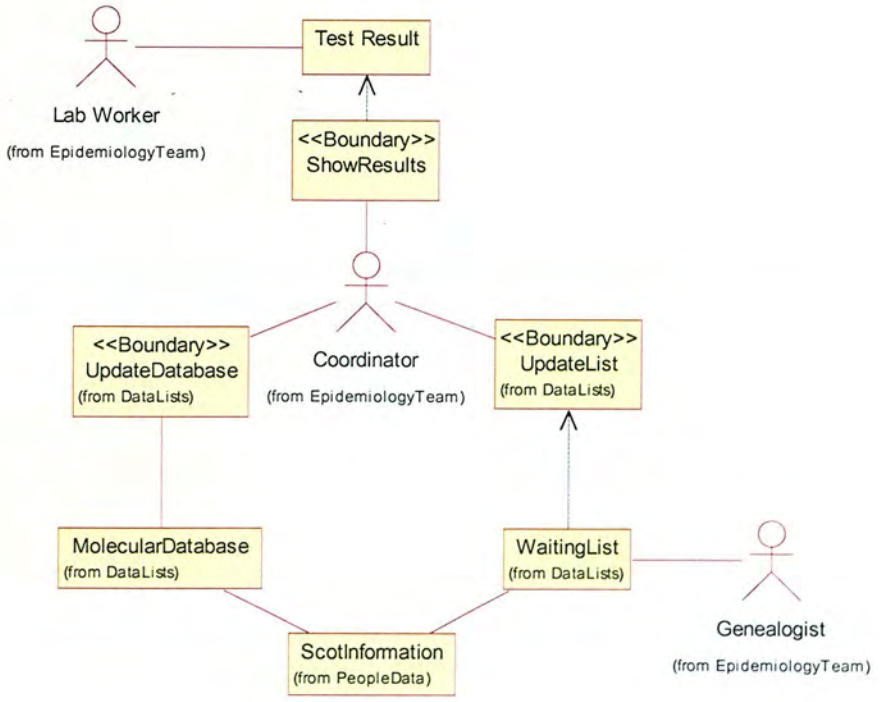


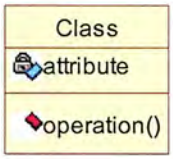
Figure A7.46 Class Diagram: Cascading Package



A7.2.2 Classes

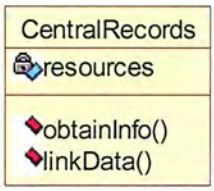
The classes used in the conceptual model, along with their attributes, operation and documentation, are presented below. Documentation was written for many attributes and operations, but was not used consistently, and is therefore not presented in this appendix.

Key for Class Notation:



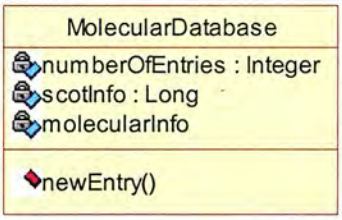
A7.2.2.1 Classes in Data Lists Package

CentralRecords



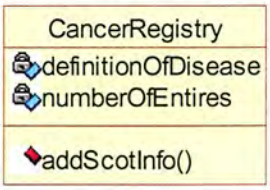
Documentation: This class refers to all centrally held records in Sotland, including details of births, marriages, deaths and health-related information. Because these databases can be linked by ISD they can be treated as one souce of information, or considered as their component parts.

MolecularDatabase



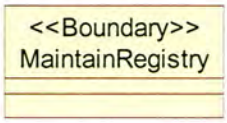
Documentation: This database contains all molecular results of genetic testing, linked to personal and clinical information.

CancerRegistry



Documentation: This class refers to the Scottish Cancer Registry, which will provide information about new colorectal cancer registrations to the system.

MaintainRegistry

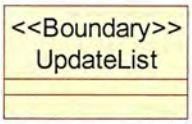


Documentation: This is a boundary class, that has the capability to maintain the cancer registry. This is a vital feature of the model, but in real life will take place outwith the cascade genetic testing system.

WaitingList

Documentation: This class is a template for all waiting lists used as part of the cascade genetic testing system.

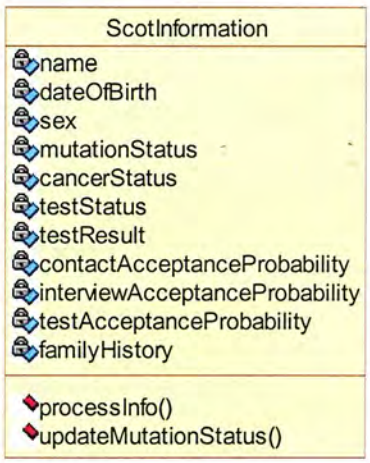
UpdateList



Documentation: This class provides the capacity to update waiting lists and present information on waiting lists and times to the system for display in the GUI.

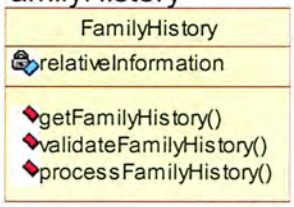
A7.2.2.2 Classes in People Data Package

ScotInformation



Documentation: This class represents the personal information on all "Scots" who are involved in the cascade genetic testing system.

FamilyHistory

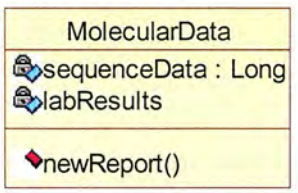


Documentation: This class contains information on the family history of any Scot.

ClinicalHistory

Documentation: This class contains information on the clinical history of any Scot.

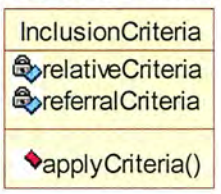
MolecularData



Documentation: This class contains, for each genetic test performed, the resulting molecular data. This may include sequence data, or the results of various other tests.

A7.2.2.3 Classes in Inclusion Package

Inclusion Criteria



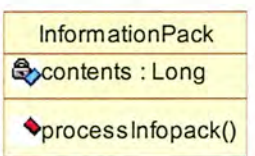
Documentation: This class contains the specifications for inclusion criteria for the following groups: 1. Referrals, 2. Relatives

A7.2.2.4 Classes in Participation Package

Invitation

Documentation: This class represents an invitation to participate in the next stage of the cascade genetic testing programme.

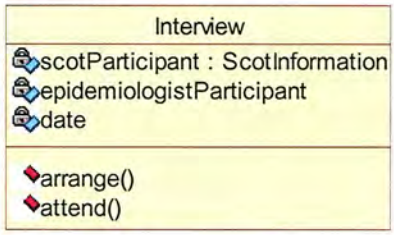
InformationPack



Documentation: This class represents the information provided for participants about the cascade genetic testing programme. In reality, this will be provided at counselling

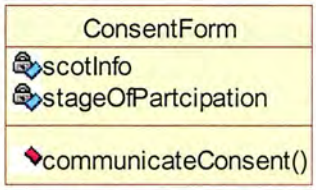
sessions as well as printed material, but for the purposes of the model it can be considered as one class.

Interview



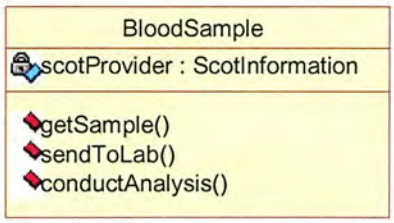
Documentation: This class represents any interview that takes place during ascertainment, recruitment or genetic testing, between a Scot and an Epidemiologist.

ConsentForm



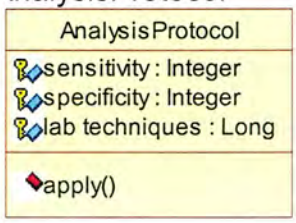
Documentation: A member of the consent form class will represent the communication of a Scot's decision regarding whether or not to progress to the next stage of the cascade genetic testing system. This may be done as an actual form, or by letter, or verbally. Written consent would be required for genetic testing.

BloodSample



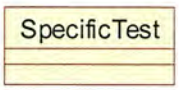
Documentation: This class comprises all the blood sample objects relating to each person who consents to a genetic test.

AnalysisProtocol



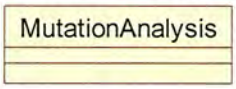
Documentation: This class contains all relevant information on the parameters of the test being applied. These include sensitivity and specificity, and will be entered via the user interface at each 'run' of the model.

SpecificTest



Documentation: This class is the subset of AnalysisProtocol that is concerned with testing for specific mutations in relatives of known carriers.Attributes and Operations are defined at the higher AnalysisProtocol level.

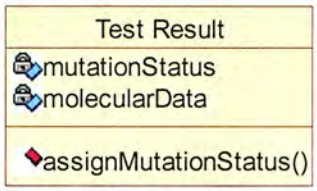
Mutation Analysis



Documentation: This class is the subset of AnalysisProtocol that is concerned with mutation analysis in suspected index cases. Attributes and Operations are defined at the higher AnalysisProtocol level.

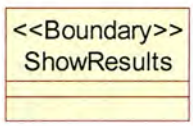
A7.2.2.5 Classes in Results Package

TestResult



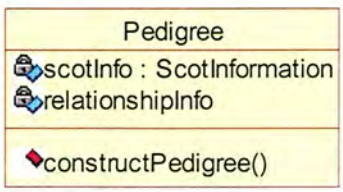
Documentation: This class contains relevant information on results of a genetic test.

ShowResults



Documentation: This boundary class enables the results of genetic testing to be displayed by the system.

A7.2.2.5 Classes in Cascading Pakage



Documentation: This class contains the pedigree information for each individual who has undergone genetic testing. By definition of the cascade screening process, these pedigrees will overlap considerably, so it may not be necessary to have a specific file for every carrier. However, the relevant data should be obtainable.

Appendix A8 Acceptance Rates Look-up Table

Age group	Sex	Relationship	Probability of Allowing Contact	Probability of Attending Genetics Nurse Interview	Probability of Accepting Pre-symptomatic test	Probability of Allowing Relative Contact
0-4	M	Patient	0.90	0.90	0.90	0.90
5-9	M	Patient	0.90	0.90	0.90	0.90
10-14	M	Patient	0.90	0.90	0.90	0.90
15-19	M	Patient	0.90	0.90	0.90	0.90
20-24	M	Patient	0.90	0.90	0.90	0.90
25-29	M	Patient	0.90	0.90	0.90	0.90
30-34	M	Patient	0.90	0.90	0.90	0.90
35-39	M	Patient	0.90	0.90	0.90	0.90
40-44	M	Patient	0.90	0.90	0.90	0.90
45-49	M	Patient	0.90	0.90	0.90	0.90
50-54	M	Patient	0.90	0.90	0.90	0.90
55-59	M	Patient	0.90	0.90	0.90	0.90
60-64	M	Patient	0.90	0.90	0.90	0.90
65-69	M	Patient	0.90	0.90	0.90	0.90
70-74	M	Patient	0.90	0.90	0.90	0.90
75-79	M	Patient	0.90	0.90	0.90	0.90
80-84	M	Patient	0.90	0.90	0.90	0.90
>85	M	Patient	0.90	0.90	0.90	0.90
0-4	F	Patient	0.90	0.90	0.90	0.90
5-9	F	Patient	0.90	0.90	0.90	0.90
10-14	F	Patient	0.90	0.90	0.90	0.90
15-19	F	Patient	0.90	0.90	0.90	0.90
20-24	F	Patient	0.90	0.90	0.90	0.90
25-29	F	Patient	0.90	0.90	0.90	0.90
30-34	F	Patient	0.90	0.90	0.90	0.90
35-39	F	Patient	0.90	0.90	0.90	0.90
40-44	F	Patient	0.90	0.90	0.90	0.90
45-49	F	Patient	0.90	0.90	0.90	0.90
50-54	F	Patient	0.90	0.90	0.90	0.90
55-59	F	Patient	0.90	0.90	0.90	0.90
60-64	F	Patient	0.90	0.90	0.90	0.90
65-69	F	Patient	0.90	0.90	0.90	0.90
70-74	F	Patient	0.90	0.90	0.90	0.90
75-79	F	Patient	0.90	0.90	0.90	0.90
80-84	F	Patient	0.90	0.90	0.90	0.90
>85	F	Patient	0.90	0.90	0.90	0.90
0-4	M	FDR	0.90	0.90	0.90	0.90
5-9	M	FDR	0.90	0.90	0.90	0.90
10-14	M	FDR	0.90	0.90	0.90	0.90
15-19	M	FDR	0.90	0.90	0.90	0.90
20-24	M	FDR	0.90	0.90	0.90	0.90
25-29	M	FDR	0.90	0.90	0.90	0.90
30-34	M	FDR	0.90	0.90	0.90	0.90
35-39	M	FDR	0.90	0.90	0.90	0.90
40-44	M	FDR	0.90	0.90	0.90	0.90
45-49	M	FDR	0.90	0.90	0.90	0.90
50-54	M	FDR	0.90	0.90	0.90	0.90
55-59	M	FDR	0.90	0.90	0.90	0.90