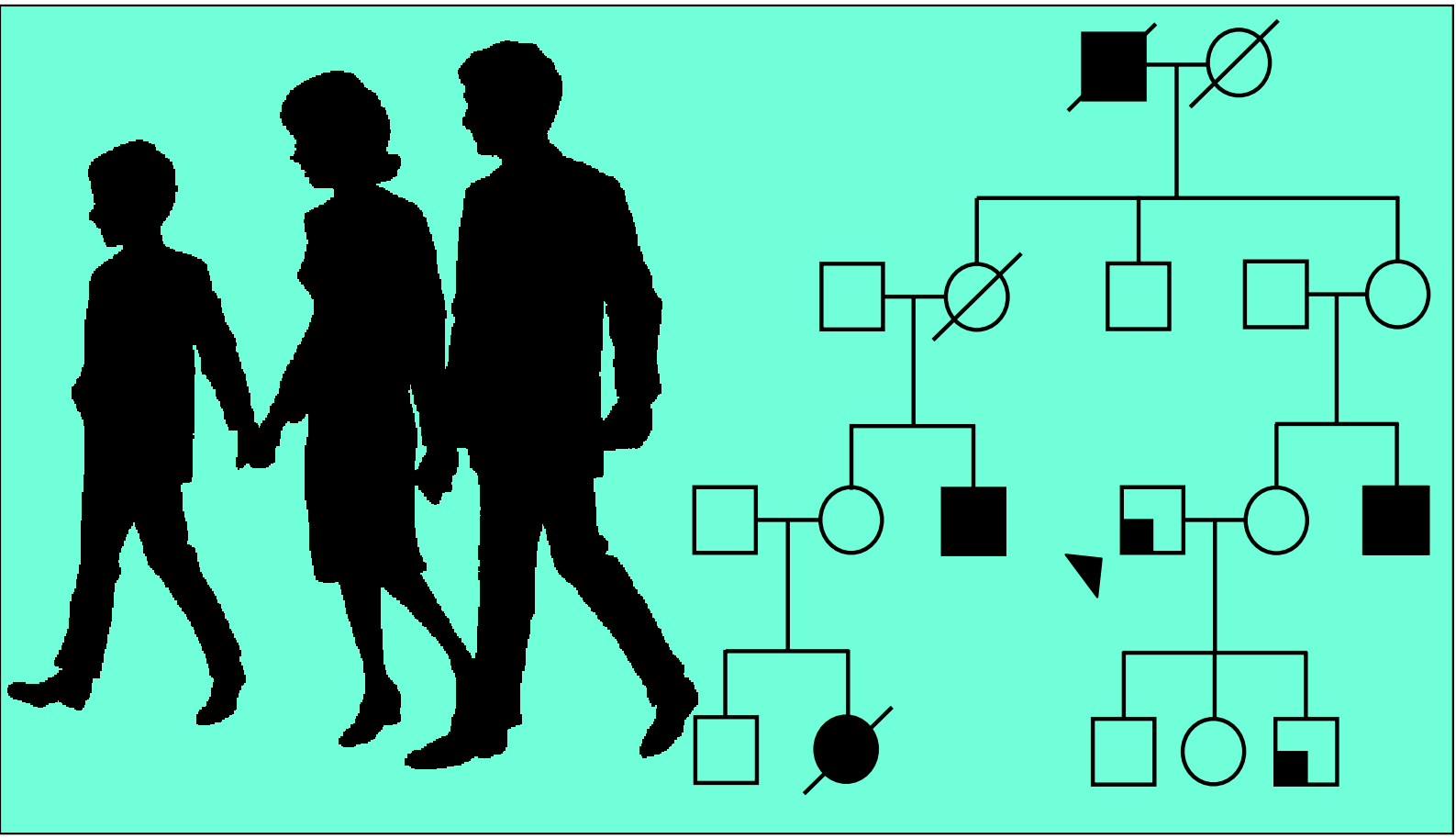
Family history as a risk factor for common, complex disease

An independent, epidemiological assessment of the evidence for familial risk of disease



Adam Butterworth

Public Health Genetics

2007

PHGU Project team

Adam Butterworth

Epidemiologist, Public Health Genetics Unit

Dr. Paul Pharoah

Honorary Consultant, Cambridge Genetics Knowledge Park
Cancer Research UK Senior Clinical Research Fellow

Public Health Genetics Unit
Strangeways Research Laboratory
2 Worts Causeway
Cambridge CB1 8RN
Tel & Fax +44 (0) 1223 740200

The report can be downloaded from the PHG Foundation website (formerly the Public Health Genetics Unit): **www.PHGFoundation.org**

The report is published by the Public Health Genetics Unit, a core facility of Cambridge Genetics Knowledge Park

Contents

Acknowledgements

We would like to thank the Association of British Insurers (ABI) for their financial support of this project. Whilst acknowledging their support, we remind readers that the project was commissioned as an independent, scientific piece of work in which the funders had no influence on the results. The ABI had no editorial control over content and the views and opinions expressed in this report do not necessarily reflect those of the ABI.

Special thanks should go to Leorita Stubbs, Gurdeep Sagoo and Helen Jones for retrieving full-text copies of articles, as well as Pamela Black, Iain Tatt and Edgar Hau for translating foreign language publications.

Thanks are also due to Julian Higgins for statistical advice and support.

Abbreviations

AR	Affected relative
BC	Breast cancer
CADASIL	Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy
CI	Confidence interval
CRC	Colorectal cancer
FAP	Familial adenomatous polyposis
FDR	First-degree relative
HNPCC	Hereditary non-polyposis colorectal cancer
IAR	Individual at-risk
ICD	International Classification of Diseases
LC	Lung cancer
MS	Multiple sclerosis
MHC	Major histocompatibility complex
OC	Ovarian cancer
ONS	Office of National Statistics
PC	Prostate cancer
PSA	Prostate-specific antigen
RR	Relative risk
SDR	Second-degree relative
TIA	Transient ischemic attack
TDR	Third-degree relative

Glossary

Absolute risk

an individual's observed or calculated risk without reference to a background or unexposed population e.g. a lifetime absolute risk of 10%.
(See *relative risk*)

Affected relative

the family member of the disease-free individual at-risk who has already been diagnosed with the disease
(See *individual at-risk*)

Case-control study

a study in which the presence of a risk factor is compared in a group of diseased individuals ('cases') and a comparison group ('controls'), usually comprised of 'healthy' or non-diseased individuals, in order to look for an association between the risk factor and disease.
(See *odds ratio; cohort study*; *cross-sectional study*).

Cohort study

a study in which a group of individuals is examined, either prospectively or retrospectively, for the presence of both a disease outcome and exposure to a potential risk factor to test for an association between the two.
(See *retrospective cohort study*; *prospective cohort study*; *case-control study*; *relative risk*)

Confidence interval

a measure of uncertainty around a statistical estimate. For example, a 95% confidence interval denotes the range in which the estimate would be expected to lie 95 out of every 100 times if the experiment was repeated.
(See *p value*)

Cross-sectional study

a study in which a group of individuals is simultaneously assessed for exposure to a risk factor and disease status allowing a test for association at that particular time.
(See *odds ratio*; *case-control study*)

First-degree relative

a person who is directly related to another i.e. parent, sibling or offspring a.k.a. 'close relative'
(See *second-degree relative*; *third-degree relative*)

Fixed effect meta-analysis

statistical method of pooling risk estimates where each study is assumed to be estimating the same underlying (or 'true') risk with no allowance for between-study heterogeneity.
(See *meta-analysis*; *random effects meta-analysis*)

Incidence

the number of new cases of a disease that appear in a particular population over a specified period of time
(See *population prevalence*)

Individual at-risk

the disease-free individual who may be risk of disease due to having a family member who is already affected
(See *affected relative*)

Life-table

a model that follows the experiences of a cohort through time, taking into account disease risk and mortality risk, including competing causes of death. In this case, it can be used to model the impact of an increased relative risk on a population exposed to a risk factor, in order to express their increased risk as an absolute risk over a specified time.
(See *relative risk*; *absolute risk*)

Meta-analysis

a statistical method of pooling the results of a number of independent studies to estimate an overall effect.
(See *meta-regression*)

Meta-regression

a specific type of regression model used within meta-analysis to explore the effects of study-level characteristics e.g. study design, on the results of a meta-analysis.
(See *meta-analysis*)

Odds ratio

a ratio of the odds of a diseased case being exposed to a risk factor to the odds of a non-diseased individual being exposed to the same factor. If the disease is rare, the odds ratio is virtually equivalent to the relative risk.
(See *relative risk*)

P value

the probability of obtaining a result as least as extreme as that seen, due to chance alone.
(See *confidence interval*)

Population prevalence

the amount of a disease (or a risk factor) in a specific population, usually taken at one particular point in time e.g. the prevalence of smoking in the UK in 2005 was 24%.
(See *incidence*)

Prospective cohort study

an association study design in which a group of disease-free people are assessed for exposure status and then followed over time to monitor disease status.
(See *retrospective cohort study*)

Random effects meta-analysis

a statistical method of pooling risk estimates in which heterogeneity between studies is allowed for. An underlying assumption is that the studies may be estimating different underlying risks.
(See *meta-analysis*; *fixed effect meta-analysis*)

Recurrence risk

the probability that a relative of a diseased individual will also have the disease, e.g. sibling risk, often denoted as λ_s, is the probability that the sibling of a disease case also carries the disease.

Relative risk

a ratio of the risk of an individual exposed to a risk factor being diseased compared to the risk of an unexposed individual being diseased.
(See *absolute risk*; *odds ratio*; *cohort study*)

Retrospective cohort study

cohort study in which the cohort is assessed for previous exposure to a potential risk factor. The prevalence of the risk factor in a diseased cohort is often compared to a background population rate to produce a relative risk estimate.
(See *cohort study*; *prospective cohort study*; *case-control study*)

Second-degree relative

a relative who is indirectly related via a first-degree relative e.g. grandparent, grandchild, aunt or uncle, niece or nephew.
(See *first-degree relative*; *second-degree relative*)

Systematic review

a rigorous and comprehensive review of a topic that uses explicit methods to find all relevant evidence for inclusion.

Third-degree relative

a 'distant' relative who is three relations from the individual at-risk e.g. great-grandparent, cousin
(See *first-degree relative*; *second-degree relative*)

1. Introduction

For decades it has been reported that many diseases tend to occur in multiple members of the same family. In very common diseases this can occur by chance, but for rarer diseases, it is very unlikely that more than one family member will be affected due to chance alone. There are two main reasons why a disease may aggregate in families; due to the genes that family members share or the common environmental factors that related individuals encounter, such as diet, exercise, or exposure to infectious agents. If any of these factors play a role in disease causation or increase the risk of developing the disease, all family members sharing this factor are at increased risk of disease.

In the second half of the twentieth century, as genetic technology improved, epidemiologists began to discover genes that cause diseases like cystic fibrosis and Huntington's disease. However, these diseases are (almost) wholly caused by single genetic mutations. For more common, complex diseases like cardiovascular disease, diabetes or many cancers, it has been harder to find the genetic factors involved in disease susceptibility. This is partly due to the polygenic nature of the diseases i.e. the fact that multiple genes interact to cause them. The picture is further clouded as many diseases are more likely to be caused by complex interactions between genetics and environment or may be caused by any one of a large number of genetic or environmental factors acting independently.

The majority of people with chronic diseases are sporadic cases i.e. they have no relatives who are affected with the same condition. Estimates of the proportion of familial cases (those with affected relatives) in complex diseases range from 5 to 30% (Scheuner *et al.*, 1997; Dong and Hemminki, 2001; Lynch and de la Chapelle, 2003). Although relatively few disease sufferers will have affected family members, the high incidence of some chronic diseases means that many unaffected individuals in the population may have a relative who is diseased and therefore be at higher risk than expected. In very few cases a familial syndrome may be present, where a number of family members are affected by the same (or related diseases), often at a younger age than most sufferers are diagnosed. Unaffected individuals in these families are at much higher risk than those with one or no affected family members, and are therefore usually excluded from epidemiological studies of familial risk. In this project we have excluded studies on defined genetic syndromes such as hereditary non-polyposis colorectal cancer (HNPCC) or CADASIL, a familial form of stroke.

Environmental factors have been studied for much longer than genetic factors as it is only recently that genetic variants have become as easy to measure. Vast numbers of environmental exposures have been postulated as risk factors for various diseases, from smoking, an established risk factor for lung and other cancers, to cholesterol, a risk factor for cardiovascular diseases, infections or viruses such as the human papilloma virus that leads to cervical cancer, or diet, components of which have been implicated in many complex diseases. Although there have been many more studies of the environment than genetics, the measurement of many of these factors is often less precise than the methods of determining the genetics of an individual that are used today.

In this project, we have estimated familial risks of common chronic diseases, such as breast cancer, prostate cancer and stroke. However, in doing so, we have not attempted to distinguish between the familial risk due to genetic factors, the risk due to environmental factors or the additional risk resulting from their interaction. In its simplest form, it is relatively easy to determine an accurate family history from an individual, especially for close relatives, and to use it to establish if the individual is at increased risk of a disease. This information is extremely useful in a public health context to ascertain people who may benefit from screening or interventions to try to prevent or

delay disease onset. It also has a use in the insurance industry where family history is commonly used to assess disease risk in order to underwrite critical illness and life insurance policies.

In both these settings, it is crucial that risk calculations are accurate and robust. There have been a wide range of study designs and sizes used to investigate the risk to those with affected family members. Some designs are less prone to bias and more reliable than others e.g. prospective studies, where the family history status is determined before the disease status. Larger studies are also more likely to provide more accurate estimates than smaller studies with lower power. Unsurprisingly, the variation in study designs and sizes leads to differences in risk estimate, making it difficult to determine which is the most accurate. The best method of determining this is to systematically review the results of all studies performed and use meta-analytic techniques to pool the risk estimates. This has the advantage of providing extra power and greater precision in estimating the relative risk, as well as providing the opportunity to examine the reasons for the variation in estimates. By reviewing the available evidence in the medical literature, we have used epidemiological and actuarial methods to obtain the most precise relative and absolute risk estimates possible.

2. Cancers

Since the early 20th century, epidemiologists have noted that cancers aggregate in families, suggesting that there is an inherited component to this group of diseases. Early case reports provided evidence of multiple case families, although these more commonly reported highly heritable, low-incidence forms of cancer (Warthin, 1925; Macklin, 1932; Cannon and Leavell, 1966). More recently twin studies, which are commonly used to evaluate the inherited nature of diseases, have also shown that many of the common cancers have a moderate familial component due both to heritable (or genetic) factors, and the shared environment which is present in the family (Ahlbom *et al.*, 1997; Lichtenstein *et al.*, 2000). However, the majority of familial studies in cancer epidemiology have looked for an increased incidence of cancer in relatives of affected family members compared to a disease-free 'control' population or background population levels. It is these association studies that estimate relative risks of disease for individuals with different family histories of disease and hence it is these studies that provided the evidence base for this work.

Due to the wide range of causes of cancer, it is a global disease with high incidence. The Office of National Statistics (ONS) recently estimated that roughly 1 in 3 people in the UK develop cancer and 1 in 4 die from cancer (ONS, 2006). With a group of diseases as common as this, large amounts of money and time have been invested in discovering risk factors and causal elements for cancers. This means that for easily measured exposures such as family history, there are many publications of risk estimates to include in systematic reviews, at least for the more common cancers. Another advantage for cancer epidemiologists of the high prevalence of cancer is that this led to the formation of national cancer registries, which record every confirmed case of cancer diagnosed in the UK. These registries provide highly accurate incidence and prevalence data for epidemiological studies such as this one.

3. Colorectal cancer

3.1 Background

Colorectal cancer (ICD10 C18-21), which includes cancers of the colon, rectum and anus, is common in both men and women, accounting for over 11% of all cancers in the UK. In 2003, over 35,000 new cases were diagnosed and over 16,000 people died primarily from the disease (ONS, 2006). The majority of cases are diagnosed after the age of 45 with a higher incidence seen in men than women as age increases.

Although the majority of cases are sporadic, there are two familial syndromes that are seen in a very small percentage of cases; hereditary non-polyposis colorectal cancer (HNPCC) and familial adenomatous polyposis (FAP). The risks for unaffected relatives of individuals with these diseases are much higher and are well described in other types of study. Because of this, studies of these conditions were excluded from this review.

Both genetic and environmental exposures have been well studied leading to the acceptance of genes such as *MLH1* and *MSH2* and dietary factors like a high intake of red meat as risk factors for developing colorectal cancer (Sandler, 1996; de la Chapelle, 2004). Both these exposures can lead to familial aggregation of the disease, which has also been investigated in a large number of studies. We reviewed this evidence and synthesised pooled relative risk estimates from these data.

3.2 Methodology

The methodology used is explained in much greater detail in a peer-reviewed paper in the European Journal of Cancer by the authors (Butterworth *et al.*, 2006), but a brief explanation is included in the following sections.

Databases of medical literature, such as Medline, Embase and Biosis, were searched using a keyword-based search strategy to attempt to find all the publications that contained appropriate data (see Appendix A of the paper). These were case-control, cohort or cross-sectional studies only. Studies based on screening programmes or familial cancer syndromes were excluded, as were studies that used individuals with cancer as control groups.

The relevant data, either the numbers of diseased and non-diseased, exposed and unexposed study participants, or a relative risk estimate and associated precision level, were extracted from the papers and collated in a database for statistical analysis. Relative risk estimates were pooled using both a fixed effect model with inverse variance weighting and a random-effects meta-analysis model, which weights studies according to the precision of their risk estimates whilst allowing for heterogeneity between studies (DerSimonian and Laird, 1986). Heterogeneity between studies was assessed using the I^2 statistic which quantifies statistical heterogeneity as a percentage, irrespective of the number of studies (Higgins and Thompson, 2002). Reasons for between-study heterogeneity were explored by pooling relative risk estimates for different subgroups within the dataset, and testing for differences between estimates using meta-regression or interaction tests (Sterne *et al.*, 2001).

Relative risk estimates for different age groups, both of the unaffected individual-at-risk and the affected relative, were calculated using the studies that had presented data stratified by age. To test the accuracy of these age-specific estimates, the mean age from each study (or each age-

specific stratum where applicable) was regressed against the natural logarithm of the relative risk using WinBUGS to generate an equation for the relative risk and age (Lunn *et al.*, 2000). If sufficient data were available to produce accurate relative risks, the age-specific estimates for both morbidity and mortality were input into life-tables.

Life-tables were constructed using the approach described by Chiang (Chiang, 1968). Age-specific incidence and mortality rates taken from population health data (ONS, 2004; ONS, 2006) were combined to produce a cumulative 'event' incidence for each year of life. Risks of mortality or morbidity were obtained by applying the incidence rates for individuals with a family history to cumulative survival probabilities. These risks were then summed to produce absolute cumulative risks over specific age ranges, and displayed graphically.

The pooled relative risk estimates for the general population (1, by definition), for individuals with at least one affected first-degree relative (parent, sibling or offspring), and for individuals with at least two affected first-degree relatives were entered into the life-table. Graphs were produced to show the absolute risks of developing and of dying from colorectal cancer over a 10-year period, a 20-year period, and a lifetime (taken to be until age 70). Crude 95% confidence intervals for these absolute risk estimates were produced by entering the upper and lower 95% confidence interval limits for the pooled relative risk estimates into the life-tables.

3.3. Relative risk estimation

The systematic review initially retrieved 4456 papers from 4 different databases (Medline, Embase, Dissertation Abstracts and Biosis). 4214 of these were eliminated through reading the title and/or abstract of the paper, leaving 242 papers for which the full article was obtained. From these 242 articles, and searching the references of relevant papers, we found data from 58 different studies that met our inclusion criteria. These studies provided data on nearly 40,000 colorectal cancer cases in total.

The pooled risk estimates were calculated for various characteristics of family history, study design and study participants (Figure 1/Table 1). The results under the fixed effect model showed that having at least one affected first-degree relative carried a relative risk of 2.11 (95% CI 2.02, 2.22), which increased slightly to 2.24 (95% CI 2.06, 2.43) under the random-effects model. As these estimates were similar and there was moderate heterogeneity between the studies (I^2 = 54%, 95% CI 36, 67), a random-effects model was used for further analyses.

The relative risk increased with multiple affected first-degree relatives to 3.97 (95% CI 2.60, 6.06) when two or more are affected. The risk also varied according to the age of the unaffected individual at risk, with an individual under 50 years having a higher risk of 3.17 (95% CI 2.37, 4.25) compared to 1.90 (95% CI 1.59, 2.28) in older subjects, and with the age of diagnosis of the affected relative, with an affected relative under 50 conferring a relative risk of 3.55 (95% CI 1.84, 6.83) compared to a risk of 2.18 (95% CI 1.56, 3.04) with older affected relatives.

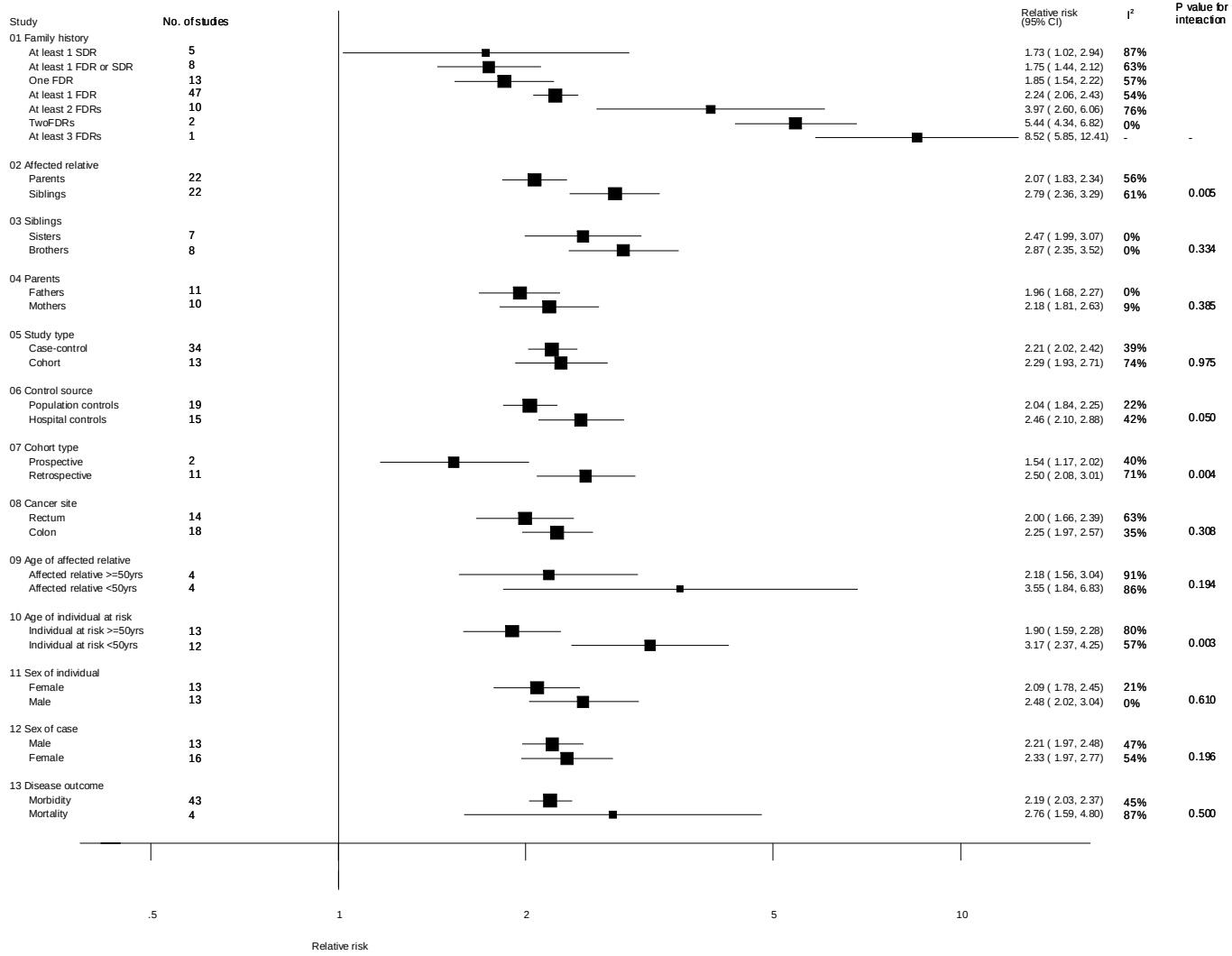
As these age-specific risk estimates were only based on age-stratified data from 13 separate studies, we attempted to model the age-specific relative risk using data on average age from each study through a meta-regression model. For an individual with at least one affected first-degree relative, the age-specific relative risk estimate was equal to $e^{(1.616 - 0.0125 \times age)}$. This model was a smoother fit of the age-specific relative risk, but due to the method employed, there was less precision in the estimate leading to wider confidence intervals and hence it was only used to

Disease	All studies	n	At least 1 FDR	n	Fixed effect	At least 2 FDRs	n	IAR cutoff	Younger IAR	n	Older IAR	n	AR cutoff	Younger AR	n	Older AR	n
Colorectal cancer	2.14 (1.98, 2.32)	58	2.24 (2.06, 2.43)	47	2.11 (2.02, 2.22)	3.97 (2.60, 6.06)	10	50	3.17 (2.37, 4.25)	12	1.90 (1.59, 2.28)	13	50	3.55 (1.84, 6.83)	4	2.18 (1.56, 3.04)	4
Prostate cancer	2.42 (2.25, 2.60)	59	2.42 (2.25, 2.60)	50	2.39 (2.30, 2.47)	4.27 (3.13, 5.84)	8	60	3.08 (2.43, 3.91)	9	2.28 (1.98, 2.63)	9	60	2.91 (2.12, 4.01)	4	1.88 (1.47, 2.40)	4
Breast cancer	-	-	1.80 (1.70, 1.91)	52	-	3.01 (2.46, 3.69)	-	40/60	2.91 (2.05, 4.13)	-	1.64 (1.36, 1.99)	-	40/60	2.22 (1.71, 2.87)	-	1.55 (1.38, 1.74)	-
Lung cancer	1.89 (1.70, 2.11)	29	1.83 (1.65, 2.03)	25	1.83 (1.75, 1.91)	2.54 (1.78, 3.63)	7	55	1.91 (1.23, 2.96)	8	1.61 (1.39, 1.87)	8	-	-	-	-	-
Ovarian cancer	2.92 (2.50, 3.41)	33	2.85 (2.41, 3.37)	26	2.70 (2.43, 3.00)	14.74 (5.78,37.60)	3	45	2.24 (1.27, 3.96)	6	2.36 (1.34, 4.16)	6	50	3.98 (2.53, 6.26)	6	2.90 (2.10, 4.01)	6
Stroke	1.76 (1.60, 1.95)	63	1.73 (1.52, 1.97)	36	1.63 (1.53, 1.73)	1.69 (1.16, 2.46)	2	55	3.33 (1.22, 9.08)	4	1.54 (1.02, 2.33)	4	55/70	3.53 (1.94, 6.43)	2	1.39 (0.94, 2.06)	2
Multiple sclerosis	11.25 (7.94, 15.94)	32	14.63 (11.1, 19.4)	20	13.51 (12.4, 14.7)	43.4 (5.4, 81.4)	1	-	-	-	-	-	-	-	-	-	-

Table 1. Summary of pooled relative risk estimates and 95% confidence intervals for different diseases and types of family history.

The first figure is the pooled relative risk estimate with 95% confidence intervals in parentheses below.
Figures are derived from random-effects meta-analysis models unless otherwise stated.
n = number of studies providing data on each subgroup.
Cut-off points for each disease are the ages at which the individuals at-risk (IAR) and affected relatives (AR) were stratified to obtain the age-specific risks.

Figure 1. Forest plot of the pooled relative risk estimates for subgroups of studies estimating the risk associated with a family history of colorectal cancer.



Each line represents a pooled subgroup risk estimate with the width of the horizontal line indicating the 95% CI, the position of the box representing the point estimate and the size of the box being proportional to the weight of each subgroup category.

check the age-stratified estimate was reliable. From the graph in Appendix 1, it can be seen that there were a range of participant mean ages from 35 to 85 with wider total age ranges. There appears to be a slight decrease in relative risk as mean age increases, adding weight to the age-specific estimates derived from the meta-analysis.

The I^2 statistic for the 47 studies that contained information on having at least one affected first-degree relative was 54%, indicating that there was moderate heterogeneity between the studies. Although none of the studies showed negative associations, there was a range of risk estimates from 1.2 to just over 9, and with nine of the studies not having confidence intervals that overlapped the pooled estimate, there were significant differences between risk estimates from different studies.

We tried to explore the reasons for this heterogeneity using subgroup analysis. All the pooled relative risk estimates were significantly higher than 1 emphasising the increased risk resulting from any familial history of the disease. Most of the subgroups we tested e.g. affected fathers versus affected mothers did not show a significant difference between the pooled estimates (i.e. the p value for the interaction test was much greater than 0.05). The only categories that were significantly different were parents (RR = 2.07) versus siblings (RR = 2.79) (p value = 0.005), and prospective cohort studies (RR = 1.54) versus retrospective cohorts (RR = 2.50) (p value = 0.004). Prospective studies are often thought to be more accurate when estimating familial risk, as they are not subject to recall bias where self-reported family history may be influenced by the individual's disease status. The risk estimates in this case suggest that the overall pooled risk of 2.24 may be inflated, as only 2 of the 59 included studies were prospective cohorts.

Another type of bias commonly seen in meta-analyses is publication bias, where the likelihood of a study being published depends on the significance of its results (Begg and Berlin, 1989). This often manifests as a study size effect, where smaller studies have more extreme results. Egger's test for publication bias was highly significant (p = 0.001) suggesting that there is a very strong relationship between study size and effect size. In an attempt to allow for this, we implemented the trim-and-fill method which recalculates the pooled relative risk estimate assuming there was no study size effect (Duval and Tweedie, 2000). In this case, the risk for having one affected first-degree relative was reduced to 2.07 under a random-effects model after the inclusion of data from 13 extra hypothetical studies. This again suggests that the true relative risk may be slightly lower than the estimates presented here.

3.4. Absolute risk estimation

The age-specific risk estimates derived from the meta-analysis were used as the basis for the absolute risk estimation, modelled using the life-table approach. Age-specific population data were collected from the ONS for colorectal cancer incidence and mortality, as well as all-cause mortality rates across England and Wales in 2003. A life-table was constructed to model a population cohort from age 0 to 85, and the relative risk estimates from the meta-analysis were used to model an 'at-risk' cohort to obtain absolute risks over the same ages. These are presented graphically for morbidity and mortality over 10 years, 20 years, and by age 70 (Figures 2 and 3). The probability of developing colorectal cancer over the next 10 years was less than 1% regardless of family history, until the age of 45, after which the risks increase up to age 75 to 2.5%, 4.7% (95% CI 4.0, 5.6) and 9.6% (95% CI 6.3, 14.2) for the general population, those with at least one affected first-degree relative and those with two or more affected first-degree relatives respectively (Figure 2). The risks by age 70 also remain constant until around

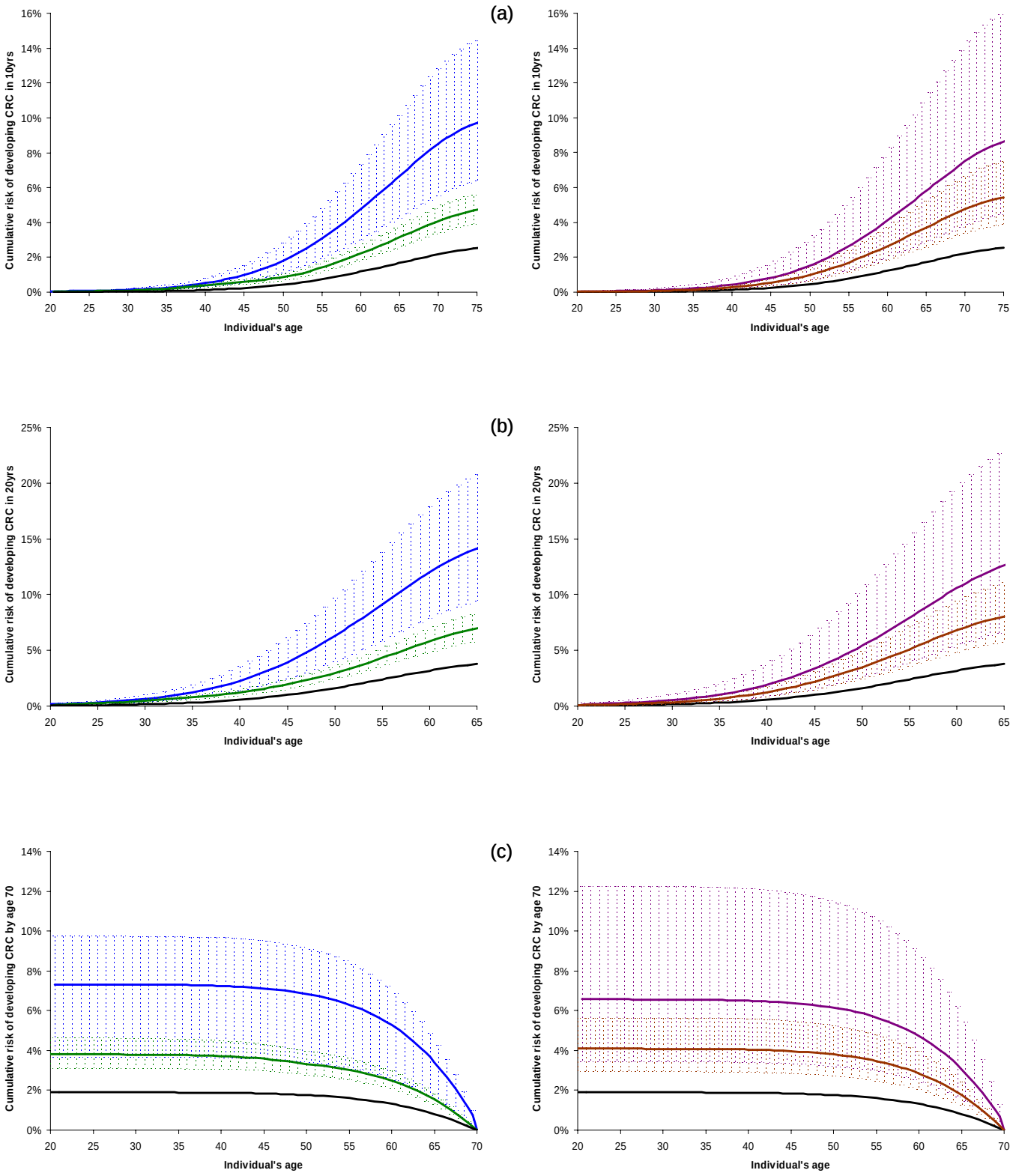


Figure 2. Cumulative absolute risks of developing colorectal cancer for different categories of family history and ages of affected relatives over (a) 10yrs, (b) 20yrs and (c) by age 70.

General population; One first-degree relative; Two first-degree relatives;

Affected relative <50 years; Affected relative >50 years.

Dashed vertical bars show 95% upper and lower confidence intervals for the risk estimates. Note: graphs have different Y-axes

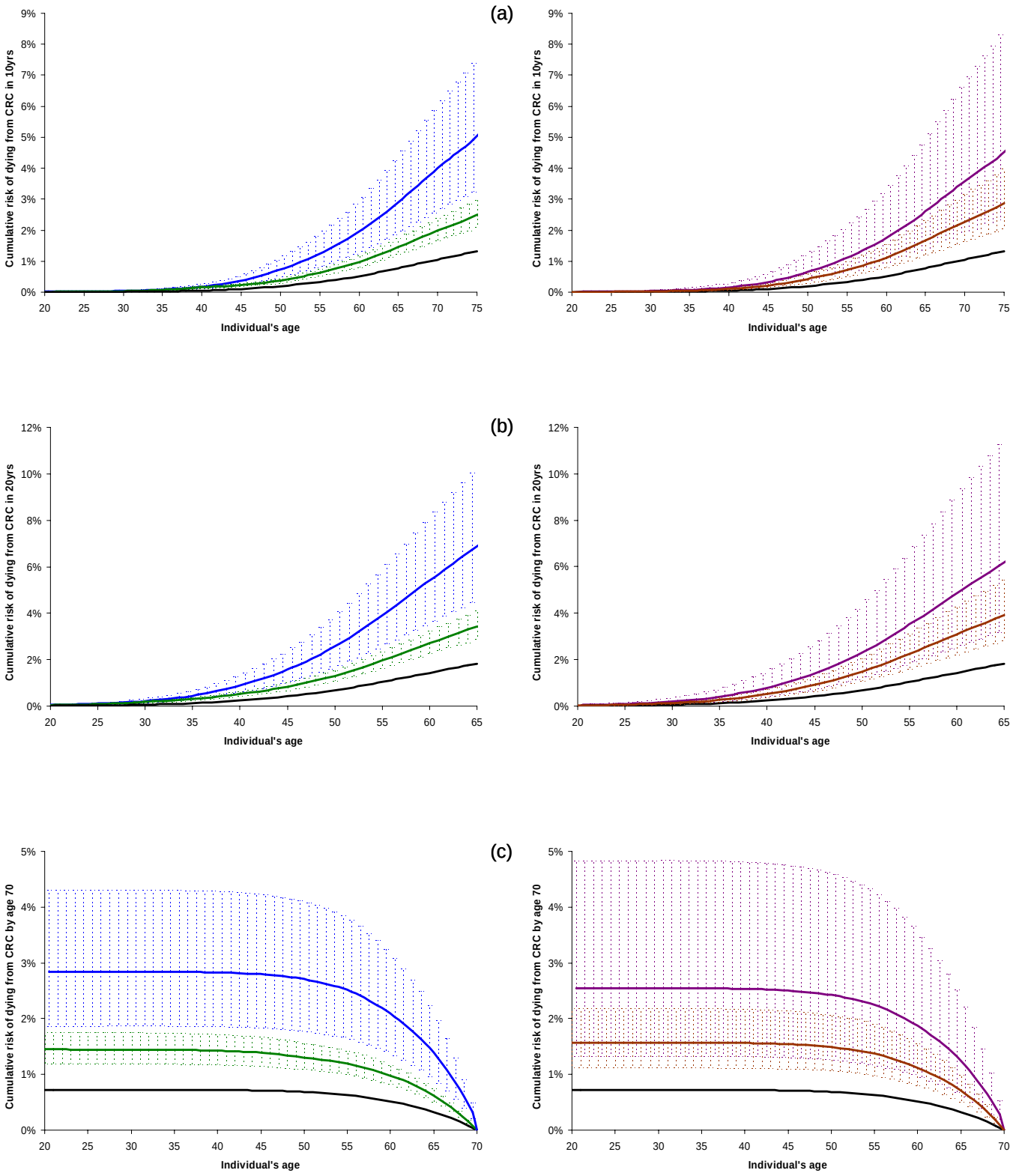


Figure 3. Cumulative absolute risks of dying from colorectal cancer for different categories of family history and ages of affected relatives over (a) 10yrs, (b) 20yrs and (c) by age 70.

General population; One first-degree relative; Two first-degree relatives;

Affected relative <50 years; Affected relative >50 years.

Dashed vertical bars show 95% upper and lower confidence intervals for the risk estimates. Note: graphs have different Y-axes

the age of 45 at 1.9% (~1 in 50) for the general population, 3.6% (~1 in 30) for those with at least one first-degree relative with colorectal cancer and 7.3% (~ 1 in 14) for people with at least two affected first-degree relatives. These values decrease throughout middle age reaching zero risk at age 70.

The absolute risks associated with having a first-degree relative affected below age 50 or at 50 and above were also calculated using the corresponding relative risk estimates from the meta-analysis. The risks were greater if the relative was diagnosed at a younger age, with a maximum 10-year morbidity risk of 8.6% (95% CI 4.6, 16.1) seen at age 75. The cumulative risk by age 70 began at 6.6% (~1 in 15) and remained above 4% (1 in 25) until the age of 60 if the relative was diagnosed at less than 50 years of age. For those with affected relatives diagnosed at 50 or above, the risks were lower, but were still higher than the general population risks.

The absolute risk curves for mortality from colorectal cancer were very similar to the morbidity curves, although with lower risks (Figure 3). Until the age of 45, the cumulative risk by age 70 was 0.72% (~1 in 140) for the general population, 1.4% (~ 1 in 70) for individuals with at least one affected first-degree relative and 2.8% (~ 1 in 35) for those with two or more first-degree relatives with colorectal cancer.

Updated and additional absolute risk data will be made available at www.PHGFoundation.org.

4. Prostate cancer

4.1 Background

Prostate cancer (ICD10 C61) is a male, late-onset disease that can remain undetected for many years. In 2003, over 30,000 new cases were diagnosed, with the majority being in men over 70 (ONS, 2006). Despite the widespread availability of PSA screening in the UK, there were still 9,000 deaths from prostate cancer, although this is currently decreasing (Marugame and Mizuno, 2005).

Although there is thought to be a strong genetic component to prostate cancer, it has been hard to elicit likely candidate genes that give susceptibility to the disease. *ELCAC2*, *RNASEL* and *MSR1* have all been suggested as possible candidate genes, although many other loci have also been proposed (Schaid, 2004). Environmental factors have been easier to find, with sexual activity, sunlight exposure and green tea consumption all thought to provide some protection from prostate cancer (Bostwick *et al.*, 2004). Like colorectal cancer, there are many studies that have tried to estimate the familial risks.

4.2 Methodology

An identical methodology was used for prostate cancer as for colorectal cancer, with a slightly adapted search strategy designed to retrieve all potentially useful studies in the area of prostate cancer. The final list of included papers was also checked against those gathered in previous reviews (Bruner *et al.*, 2003; Johns and Houlston, 2003). Incidence and mortality data were again taken from the ONS data from England and Wales in 2003 (ONS, 2004; ONS, 2006).

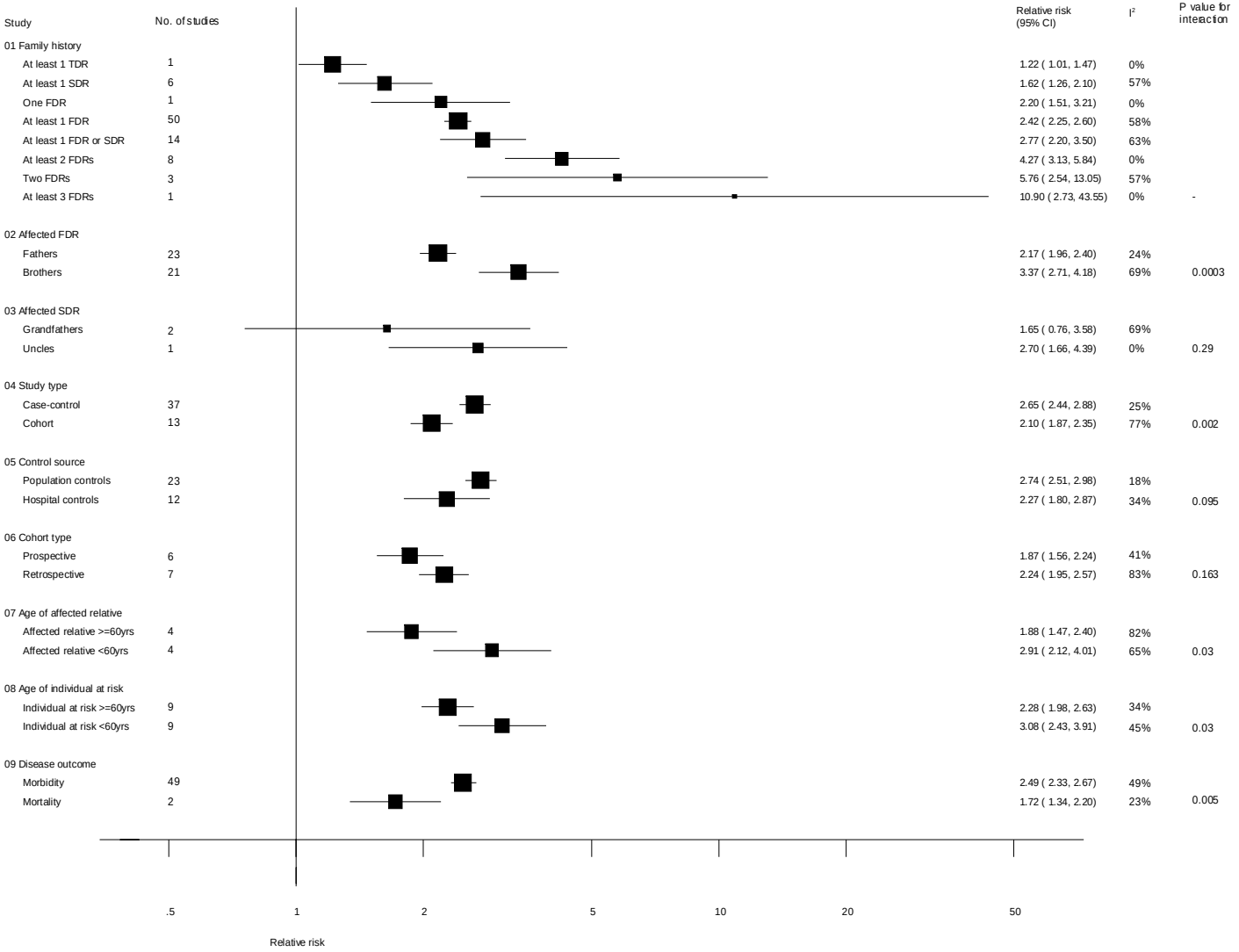
4.3 Relative risk estimation

A large number of articles were retrieved by the search strategy (3806) and nearly 20% of these were obtained in full (n = 759). Eventually 59 eligible studies were included in the meta-analysis, of which 50 contained data on the risk associated with having at least one first-degree relative affected. In total data were available from 40,000 prostate cancer cases and 25,000 disease-free controls. As prostate cancer tends to have a later age at onset than colorectal cancer, the age data were stratified at 60 instead of 55.

The overall and subgroup pooled relative risk estimates are shown in Figure 4/Table 1. As the risks with one affected first-degree relative were very similar under the fixed and random-effects models (fixed RR = 2.39, random RR = 2.42) in the presence of moderate heterogeneity (I^2 = 58%), the random-effects model was used to estimate further relative risks.

The results for prostate cancer were very similar to those seen in colorectal cancer, with significantly higher risks for individuals with multiple affected relatives, individuals at a younger age, and those with affected relatives diagnosed at younger ages. The relative risks were also significantly higher in individuals with affected brothers compared to affected fathers, in case-control studies compared to cohort studies and for morbidity compared to mortality. Again, the majority of the subgroup estimates were significantly higher than 1, with the exception of the risk for those with affected grandfathers (RR = 1.65, 95% CI 0.76, 3.58) though this was only based on data from 2 studies.

Figure 4. Forest plot of the pooled relative risk estimates for subgroups of studies estimating the risk associated with a family history of prostate cancer.



Each line represents a pooled subgroup risk estimate with the width of the horizontal line indicating the 95% CI, the position of the box representing the point estimate and the size of the box being proportional to the weight of each subgroup category.

The I^2 statistic was 58% (95% CI 42, 69) for the 'at least one first-degree relative' meta-analysis, implying that there was moderate heterogeneity between studies. With individual study estimates ranging from 1.3 to 9.7 and 12 studies in which the 95% confidence interval doesn't overlap the pooled RR estimate, it is clear that there is variation in study estimates. However, this doesn't appear to be solely explained by a study size effect as publication bias only has a marginally significant effect on the overall pooled risk estimate ($p = 0.05$). The trim-and-fill method only reduced the risk for having one affected first-degree relative from 2.42 to 2.40, through the addition of 3 'missing' studies.

The mean ages of participants were higher than for colorectal cancer as prostate cancer tends to be identified at a greater age (Appendix 1). However, there were still a range of participant ages from 20 to more than 90. A meta-regression of mean age of participants against relative risk estimate produced the equation $RR = e^{(1.616 - 0.0125 \times age)}$ for individuals with at least one affected first-degree relative, which again fitted the derived pooled estimates well, but had less precision than the random-effects model.

4.4. Absolute risk estimation

The absolute risk curves for morbidity of prostate cancer are shown in Figure 5. The curves were similar in shape to the colorectal cancer risk curves, but later, and steeper after 50 years, reflecting the background population incidence. For this reason, the graphs start at age 30 rather than 20 as for the other cancers.

The 10-year risk of developing prostate cancer for the general population was very low until age 40, and then rose rapidly, before levelling off at 5.8% (~1 in 17) at age 75. The risk with one affected first-degree relative at age 75 was, as we would expect, higher, with an absolute cumulative risk of 12.6% (95% CI 11.1, 14.4) and a much higher risk was seen with multiple affected first-degree relatives of 22.3% (95% CI 16.9, 29.2), nearly 1 in 4. The corresponding 20-year risks for a 65-year old were 9.0%, 19.3% and 32.8% respectively, and the lifetime risks were similarly higher than for colorectal cancer with plateaus throughout middle age at around 4% (1in 25), 9% (~1 in 11) and 16% (~1 in 6) for the general population, those with one affected first-degree relative and those with multiple affected first-degree relatives.

For mortality of prostate cancer, the risk curves were lower than those for morbidity, but again higher and steeper than those for colorectal cancer (Figure 6). The maximum 10-year risks were 2.7%, 6.0% (95% CI 5.2, 6.9) and 10.9% (95% CI 8.1, 14.6) at age 75 for the general population, one affected first-degree relative and multiple affected first-degree relatives respectively. Due to the higher proportion of prostate cancer deaths after age 70, the lifetime mortality risks for prostate cancer were actually lower than those for colorectal cancer, as the definition we have used for lifetime risk is 'by age 70'. The highest lifetime risks were 0.5% (1 in 20), 1.3% (~1 in 75) and 2.2% (~1 in 45) for the three categories.

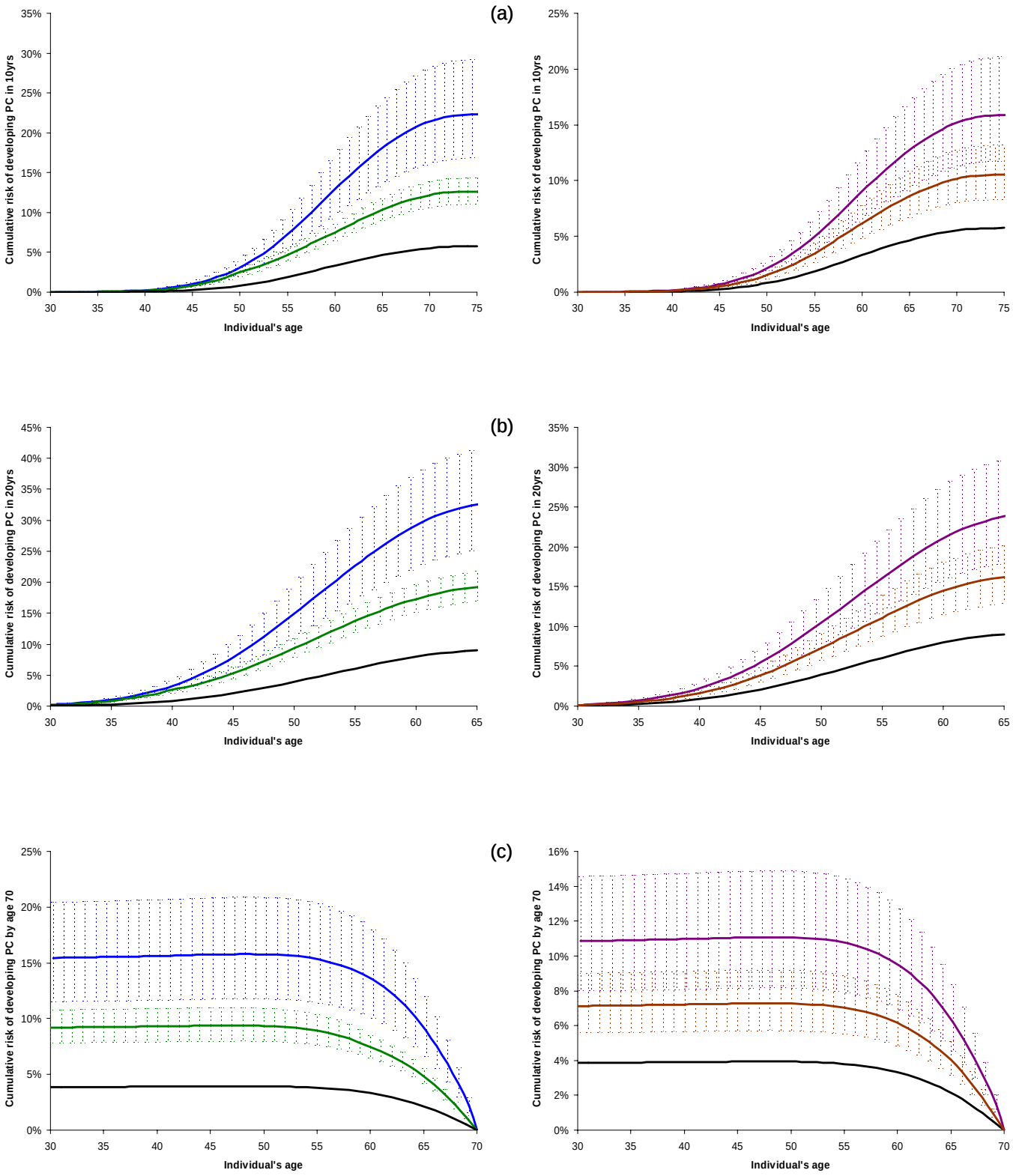


Figure 5. Cumulative absolute risks of developing prostate cancer for different categories of family history and ages of affected relatives over (a) 10yrs, (b) 20yrs and (c) by age 70.

General population; One first-degree relative; Two first-degree relatives;

Affected relative <60 years; Affected relative >60 years.

Dashed vertical bars show 95% upper and lower confidence intervals for the risk estimates. Note: graphs have different Y-axes

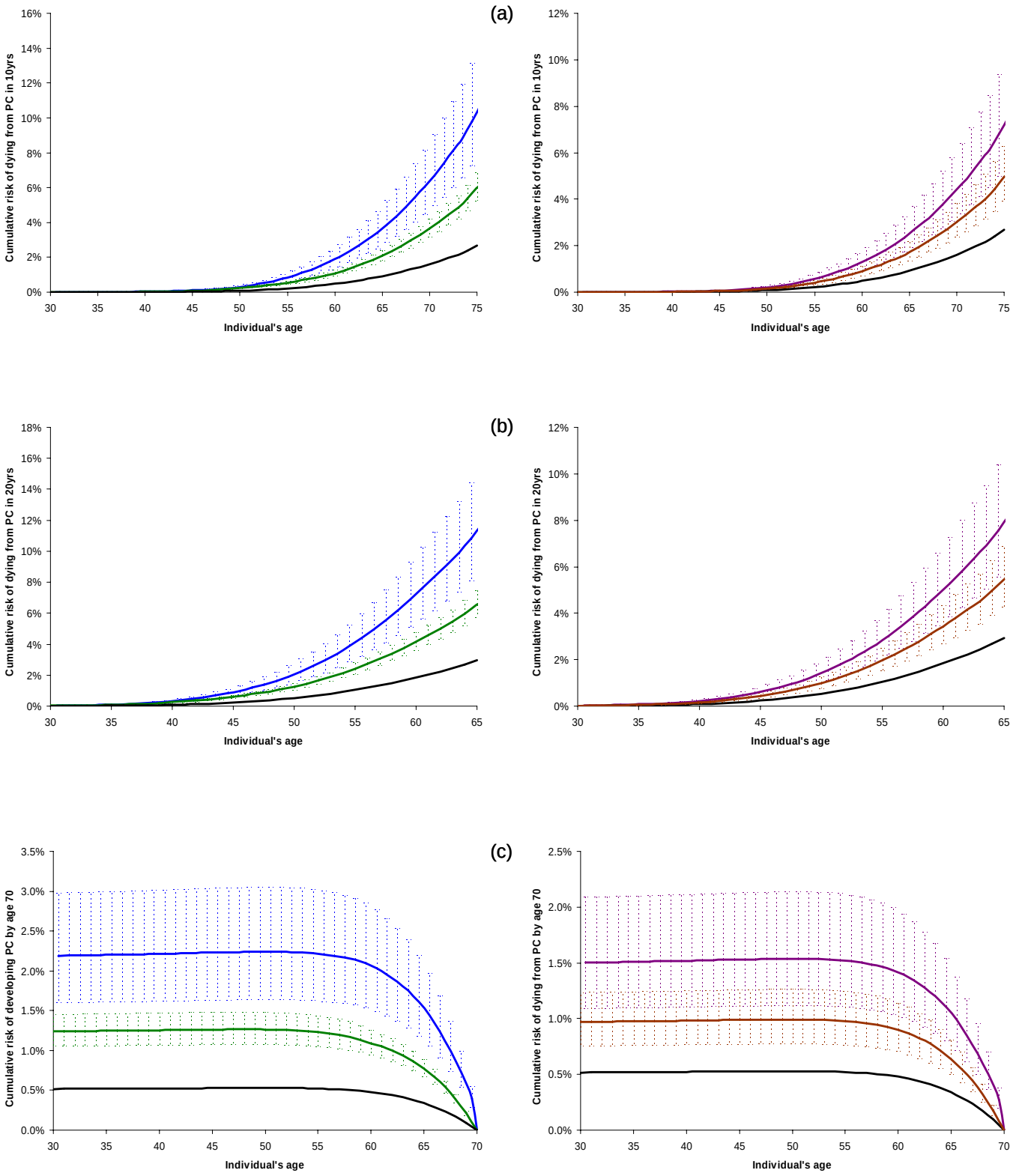


Figure 6. Cumulative absolute risks of dying from prostate cancer for different categories of family history and ages of affected relatives over (a) 10yrs, (b) 20yrs and (c) by age 70.

General population; One first-degree relative; Two first-degree relatives;

Affected relative <60 years; Affected relative >60 years.

Dashed vertical bars show 95% upper and lower confidence intervals for the risk estimates. Note: graphs have different Y-axes

5. Breast cancer

5.1. Background

Breast cancer (ICD10 C50) is the most common cancer amongst females in the UK, with over 44,000 new cases diagnosed in 2003 alone (ONS, 2006). It is also one of the most highly studied cancers with a wealth of published literature available. As it is so common, it was one of the first cancers to have a demonstrated familial effect with many case or cluster reports of families with multiple cases arising from the early 20th century (Pearson, 1912; Gardner and Stephens, 1950). Breast cancer was also the first polygenic cancer in which a strong candidate gene, *BRCA1*, was discovered (Miki *et al.*, 1994). Other highly penetrant genes like *BRCA2* and *ATM* have followed, although there are likely to be many more low-penetrance genes that affect breast cancer susceptibility (Dumitrescu and Cotarla, 2005).

5.2. Methodology

A study by the Collaborative Group on Hormonal Factors in Breast Cancer in 2001 attempted to collect all published familial risk studies with at least 100 cases and pool them to estimate the true relative risks (Collaborative Group on Hormonal Factors in Breast Cancer, 2001). Although the methodology used in this study was slightly different to those used for colorectal and prostate cancer (e.g. synthesis of risk estimates from individual participant data), the detailed age-specific risk estimates and narrow confidence intervals would be unlikely to be improved in an updated literature-based review and meta-analysis. As this review contains over 50 studies with over 58,000 affected cases, we decided not to undertake our own review, but to use the relative risk estimates from this study, as summarised in Table 1.

5.3. Relative risk estimation

With risks estimated from such a large number of participants, the Collaborative group were able to estimate risks for more subgroups, such as 'at least 3 first-degree relatives', and were able to subdivide into nine age categories. Again, there was an increase in risk with multiple affected first-degree relatives as well as a decrease in risk as the age of the individual at-risk increases, whether with one or multiple affected relatives.

As there were no available data on the age ranges in each study, we did not attempt to graph these against relative risk. However, there was a spread of mean ages from 33 to 67, with an apparent trend of decreasing relative risk with increasing age. The data on the mean ages of participants from all 52 studies were used to try and estimate the effect of age on the relative risk using meta-regression. The equation for the effect of age on the relative risk associated with having one first-degree relative was $RR = e^{(1.407 - 0.015 \times age)}$. This curve was an excellent fit to the stratified age-group estimates (data not shown), although again the confidence intervals generated by the model were much larger than those from the meta-analysis, so the stratified estimates were used for the absolute risk estimation. Some of the age-specific risk estimates that were entered into the life-tables are shown in Table 1 (Collaborative Group on Hormonal Factors in Breast Cancer, 2001).

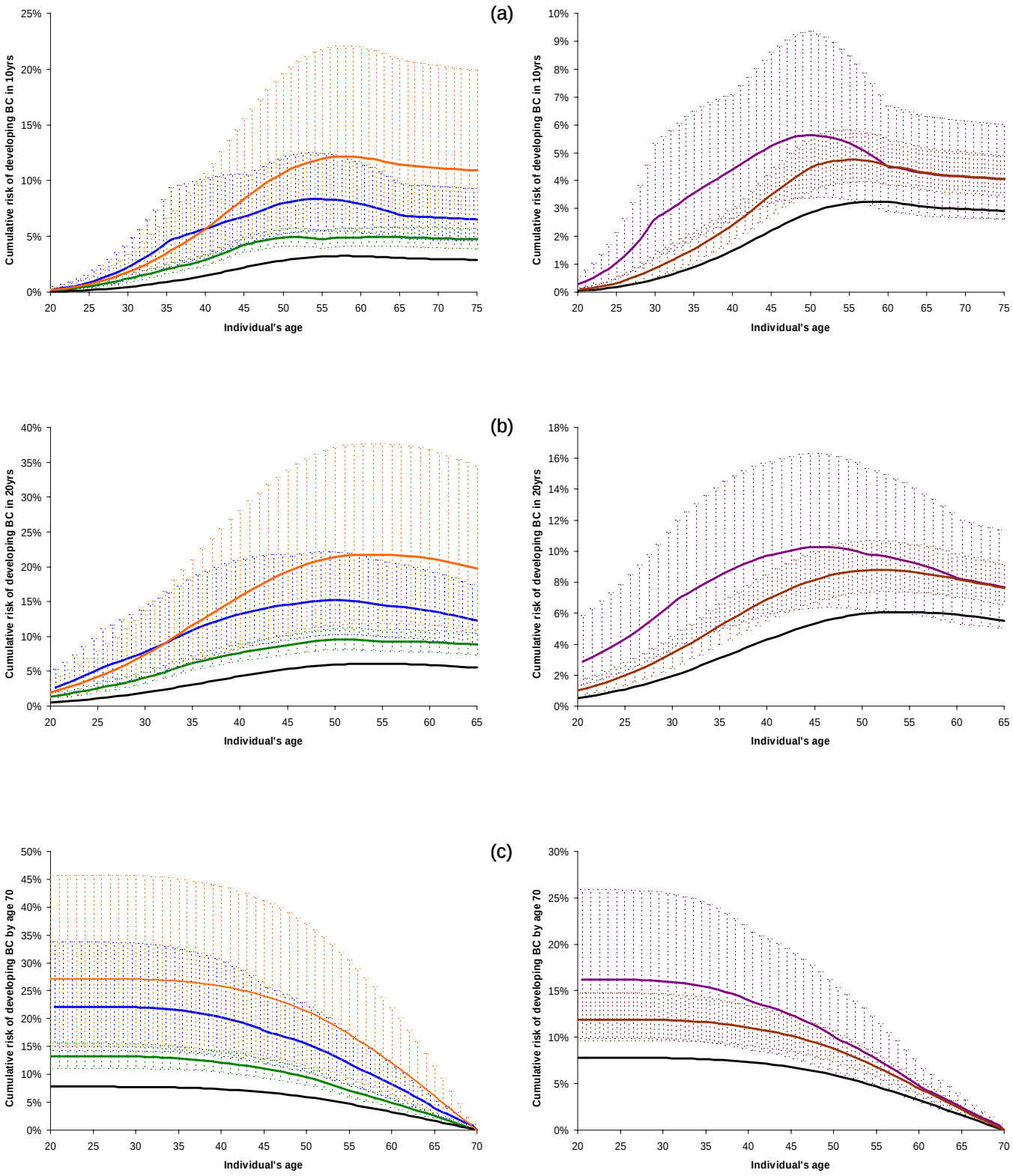


Figure 7. Cumulative absolute risks of developing breast cancer for different categories of family history and ages of affected relatives over (a) 10yrs, (b) 20yrs and (c) by age 70.

General population; One first-degree relative; Two first-degree relatives; At least 3 first-degree relatives; Affected relative <40 years; Affected relative >60 years.

Dashed vertical bars show 95% upper and lower confidence intervals for the risk estimates. Note: graphs have different Y-axes

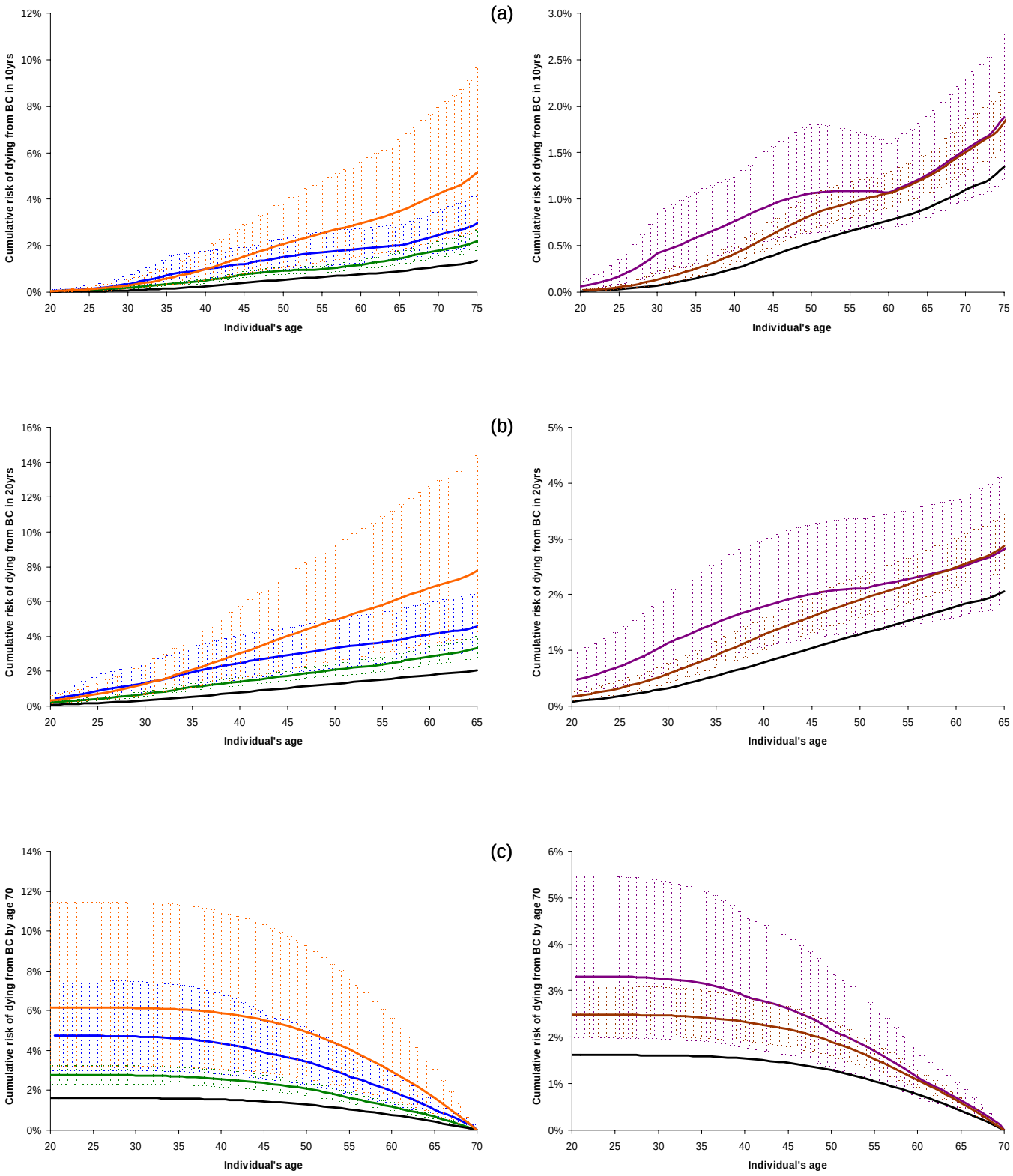


Figure 8. Cumulative absolute risks of dying from breast cancer for different categories of family history and ages of affected relatives over (a) 10yrs, (b) 20yrs and (c) by age 70.

General population; One first-degree relative; Two first-degree relatives;
At least 3 first-degree relatives; Affected relative <40 years; Affected relative >60 years.

Dashed vertical bars show 95% upper and lower confidence intervals for the risk estimates. Note: graphs have different Y-axes

5.4. Absolute risk estimation

The increase in relative risk seen with multiple affected relatives was mirrored in the cumulative absolute risks for all time periods (Figure 7). The absolute risk of developing breast cancer in the next 10 years increased until age 60, where it was 3.2% (~1 in 30) for the general population. This translates to a maximum 10-year risk of 4.9% (95% CI 4.2, 5.8) with one affected first-degree relative, 8.3% (95% CI 5.6, 12.4) with two affected first-degree relatives and 12.1% (95% CI 6.5, 22.0) with three or more.

The lifetime risks showed similar increases with number of affected relatives. The risk to age 70 for the general population decreased from a maximum of 7.8% (~1 in 12) at 30. (This is lower than the currently quoted UK lifetime risk of breast cancer of 1 in 9 as that is based on a risk to age 85, which is not presented in this report. However, our estimate of this risk was 11.2%, which is equivalent to a 1 in 9 chance). The maximum lifetime risk for those with one affected first-degree relative was 13.2% (~1 in 8), which increased to 22.0% (~1 in 5) with two affected relatives and 27.1% (~1 in 4) with three or more.

As the investigators conducting the previous meta-analysis had age data from all participants in each study, they were able to produce stratified age-specific relative risk estimates for a number of age groups, both of the individual at-risk and the affected relative. To look at the effect of the age of diagnosis of the affected relative, we took the two most extreme subgroups – less than 40 and above 60 – and produced risk curves for these categories (Figure 7). The risk for those with younger relatives was higher for the 10 and 20-year risks until the age of 60 where no difference was seen. The lifetime morbidity risk for those with relatives diagnosed below age 40 was 16.2% (~1 in 6) at age 20, which was higher than the risk with one affected first-degree relative if the relative was diagnosed at above 60 at 11.9% (~1 in 8).

The mortality risks of breast cancer were much lower than the morbidity risks as we would expect from the relatively high survival rates (Figure 8). The lifetime risk for the general population was below 2% at all ages, with the highest risk at 2.8% (~1 in 35) for those with one affected first-degree relative, 4.7% (~1 in 20) for those with two affected first-degree relatives and 6.2% (~1 in 16) with three or more. Again the risks stratified by age of affected relative showed a higher risk with younger relatives until around the age of 60.

6. Lung cancer

6.1 Background

Lung cancer (ICD10 C32-4) is the most common cancer in the world as well as one of the most common in the UK. In 2003 there were over 35,000 cases in men and women in the UK, and nearly the same number of deaths, making it the most common cause of death from cancer in both sexes (ONS, 2006). Since the 1960's, studies have shown evidence of a familial aggregation of cancer, although the case of lung cancer is not as straightforward as other cancers. In the 1950's, the pioneering work of Richard Doll and colleagues proved the link between smoking and developing lung cancer, which has been established as the major risk factor for lung cancer morbidity and mortality ever since (Alberg *et al.*, 2005).

As smoking also shows a strong familial clustering – i.e. having a relative who smokes increases an individual's likelihood of also smoking, it is hard to dissociate the hereditary effects of lung cancer from those caused by the shared smoking status. To combat this, some studies have only investigated familial aggregation in non-smokers or people who have never smoked, whilst many others adjust their risks according to the smoking status, either of the individual or the relatives. Although smoking is a risk factor already accounted for in medical underwriting, it may be useful in this instance, where there is such a strong environmental causative factor, to examine how the risks differ between smokers and non-smokers and also by adjustment for smoking.

6.2. Methodology

Matakidou and colleagues at the Institute of Cancer Research carried out a systematic review and meta-analysis of the relationship between family history and lung cancer risk, published in the British Journal of Cancer in October 2005 (Matakidou *et al.*, 2005). Their review found 32 separate studies that had published relative risks of lung cancer in those with a family history, four of which were cohort studies.

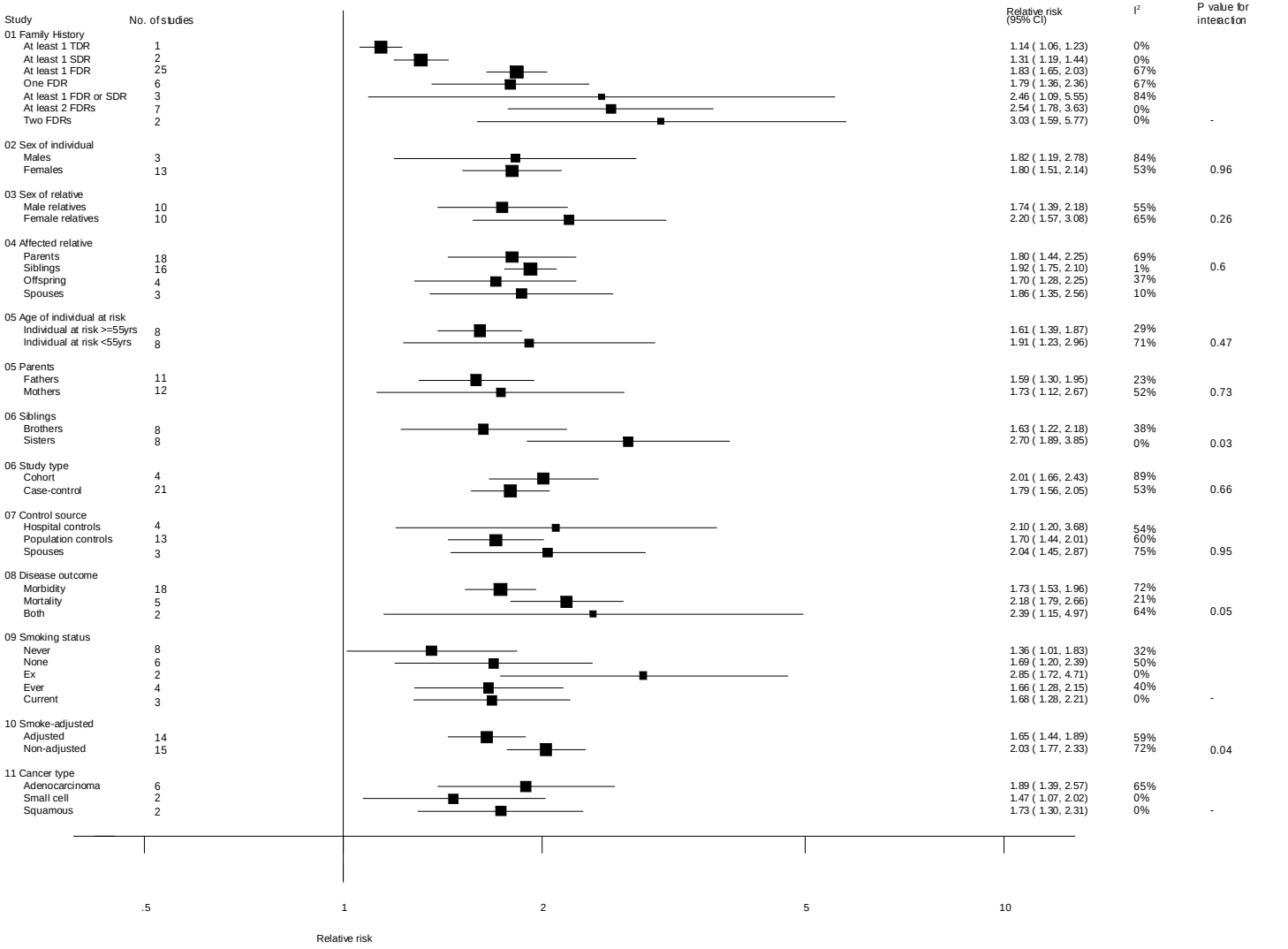
Due to differences in our methodologies, we obtained copies of all publications included in their review and evaluated them by our own criteria. We excluded two case-control studies as the controls included hospital patients who had malignant disorders which, due to the well-established associations between various cancers, may bias the results. A further study was excluded due to the absence of a disease-free population for comparison. Despite a brief literature search, we were unable to discover any extra studies suitable for inclusion.

As the Matakidou review had not attempted to produce age-stratified relative risks through meta-analysis, we also extracted the risk estimates, confidence intervals and age data from all the included studies to reanalyse the data using our own methodology.

6.3. Relative risk estimation

The new dataset included 26 case-control studies and 3 cohort studies from which data were extracted. In total these studies included over 17,000 cases, as well as more than 16,000 cancer-free controls. The pooled relative risk estimates can be seen in Table 1/Figure 9. Once again, an approximate doubling of risk was seen for those with an affected first-degree relative (RR = 1.83, 95% CI 1.65, 2.03), although the risk was closer to the risk of breast cancer than those of colorectal and prostate cancer which were greater than 2. This was identical to the risk under

Figure 9. Forest plot of the pooled relative risk estimates for subgroups of studies estimating the risk associated with a family history of lung cancer.



Each line represents a pooled subgroup risk estimate with the width of the horizontal line indicating the 95% CI, the position of the box representing the point estimate and the size of the box being proportional to the weight of each subgroup category.

the fixed effect model, although the random-effects model had slightly wider confidence limits. As risk estimates varied from 1.3 to 5.7, there was strong heterogeneity between the studies (I^2 = 67%, 95% CI 50, 78), so a random-effects model was used for the subgroup analysis to allow for heterogeneity.

The relative risks for familial lung cancer again showed an increase with the number of affected relatives and the proximity of the relationship (Table 1). The relative risk for an individual with two affected first-degree relatives was 2.54 (95% CI 1.78, 3.63), although unlike the previous cancer results, this was not statistically significantly higher than the risk associated with just one first-degree relative (p value for interaction = 0.2). The pooled relative risk estimate increased from just 1.14 (95% CI 1.06, 1.23) for an affected third-degree relative to 1.31 (95% 1.19, 1.44) for a second-degree relative to the first-degree risk of 1.83. Again, all the subgroup estimates were statistically significant, although many had wide confidence intervals due to the small number of studies included than in the other meta-analyses.

Unlike the other cancers, there was no statistically significant difference between the age-specific estimates stratified by the age of the individual at-risk. Age-specific data were available from eight studies from which a cut-off age of 55 was used. The risk to those aged under 55 with an affected first-degree relative was estimated to be 1.91 (95% CI 1.23, 2.96) compared to 1.61 (95% CI 1.39, 1.87) for older individuals. Unfortunately there were no studies that had stratified the risk by the age of the affected relative so we were not able to include this category in our meta-analysis or absolute risk estimation.

The graph of age ranges versus relative risk supports the lack of age-specific risk difference as little trend can be seen (Appendix 1). The mean ages for lung cancer are mainly between 40 and 70, although few age data were available. There is a wide spread of overall age ranges with participants from under 20 and over 80 both included.

Marginally significant differences were seen between the risk estimates for sisters (RR = 2.70, 95% CI 1.89, 3.85) and brothers (RR = 1.63, 95% CI 1.22, 2.18; p = 0.03) and estimates which were adjusted for smoking status (RR = 1.65, 95% CI 1.44, 1.89) compared to those that weren't (RR = 2.03, 95% CI 1.77, 2.33; p = 0.04). However the risk estimates did not differ significantly between other relatives, the sex of the individual or relative, the type of study, the source of controls, disease morbidity or mortality, or lung cancer type. There was no clear pattern in the risk estimates of different smoking groups, with 'never smokers' having a lower risk than non-smokers or smokers, whilst ex-smokers had a higher relative risk of 2.85 (95% CI 1.72, 4.71).

As there was little evidence of a study size effect (Egger's p = 0.065), the trim-and-fill method did not alter the random-effects estimate greatly. Only two hypothetical studies were added, which reduced the relative risk associated with one affected first-degree relative to 1.80 (95% CI 1.61, 2.00). A meta-regression model was unable to confirm an increase or decrease in relative risk with age of the individual at-risk with the 95% confidence interval for the gradient coefficient ranging from –0.00074 to 0.0145. This highlights the lack of a significant change in the risk with age.

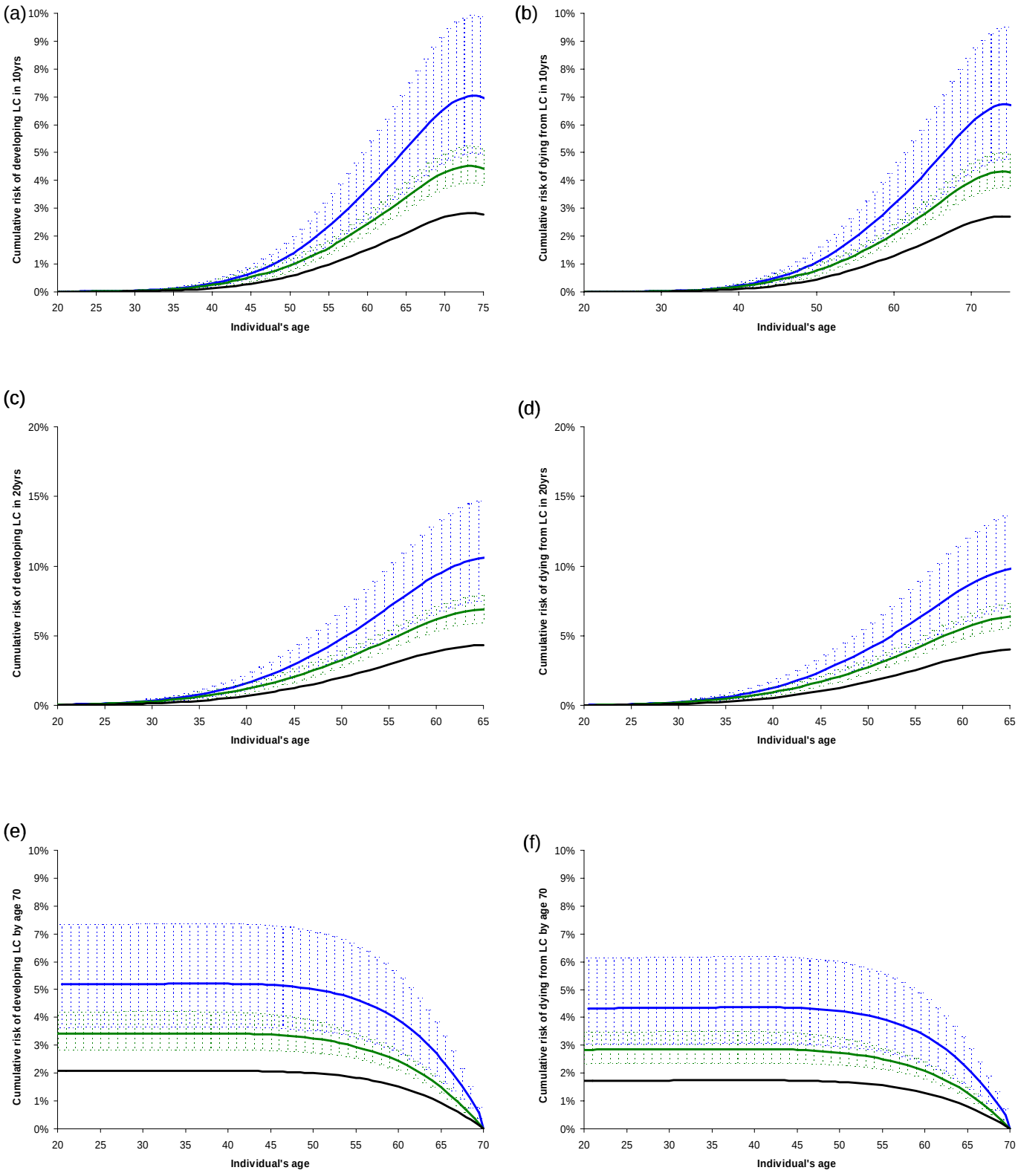


Figure 10. Cumulative absolute risks of developing and dying from lung cancer for different categories of family history and ages of affected relatives over (a,b) 10yrs, (c,d) 20yrs and (e,f) by age 70.

General population; One first-degree relative; Two first-degree relatives;

Dashed vertical bars show 95% upper and lower confidence intervals for the risk estimates. Note: graphs have different Y-axes

6.4. Absolute risk estimation

As there were no studies that stratified relative risks by the age of the affected relative, we were not able to produce absolute risk curves for these subgroups. Hence the risk curves for both morbidity and mortality over 10 years, 20 years and by age 70 are all presented in Figure 10.

The lung cancer risks for morbidity and mortality were very similar, due to the high mortality rates seen in the UK (ONS, 2004). Current estimates suggest the 5-year survival rate for lung cancer is just 15% and thus age-specific mortality rates are only marginally lower than incidence rates for the same age group, unlike the other cancers. Like prostate cancer, lung cancer showed a relatively late age of onset with the 10-year morbidity risk remaining under 1% until age 55 for the general population. This rose to 2.8% (~1 in 35) by the age of 75. For individuals with an affected first-degree relative the maximum risk was 4.5% (95% CI 3.9, 5.2) at age 73, which increased to 7.0% (95% CI 5.0, 9.9) for those with two or more affected first-degree relatives. The corresponding mortality risks for the same time period were 2.7%, 4.3% (95% CI 3.7, 5.0) and 6.7% (95% CI 4.8, 9.5) respectively, which were only marginally lower than the morbidity risks.

The risk curves for the next 20 years were similar in shape, with maximum absolute risks for morbidity and mortality of around 4%, 7% and 10% at age 65 for the general population, those with one affected first-degree relative and those with two or more.

Due to the late age of onset of lung cancer, the lifetime risks (to age 70) were low compared to other cancers and remained constant until around the age of 50. The population morbidity risk was 2% (1 in 50), which only increased to 3.4% (~1 in 30) and 5.2% (~1 in 20) with one and two or more affected first-degree relatives. Again the mortality risks were extremely similar, but slightly lower, reflecting the minimal difference in incidence and mortality rates.

7. Ovarian cancer

7.1. Background

There has been some debate about the familial aggregation of ovarian cancer (ICD10 C56), partly due to the association seen with breast cancer. As well as sharing common environmental causes e.g. hormonal and dietary factors, susceptibility to both breast and ovarian cancer is increased by mutations in the same genes, such as *BRCA1* and *BRCA2* (Welcsh and King, 2001). Although there have been many case and family studies describing families with multiple cases of ovarian and breast cancer, there are fewer studies that have tried to estimate the familial risk of ovarian cancer alone.

Despite this association with breast cancer, ovarian cancer is still a serious public health problem in its own right. With nearly 7,000 cases in UK women in 2003, it is the least common of the cancers we have studied, although with over 4,000 deaths per annum, it has a relatively high mortality rate (ONS, 2004). Regardless of the low incidence rate, it is important to identify those at high risk of developing ovarian cancer, as screening or even preventive measures can be used to manage this risk.

7.2. Methodology

A previous systematic review of familial risks was undertaken by Stratton and colleagues (Stratton *et al.*, 1998). We took the studies included in this review and measured them against our own inclusion criteria. We also searched the literature for more recent studies that would not have been included in this other review, studies that may have been missed by the previous review and papers containing updated information on studies previously included.

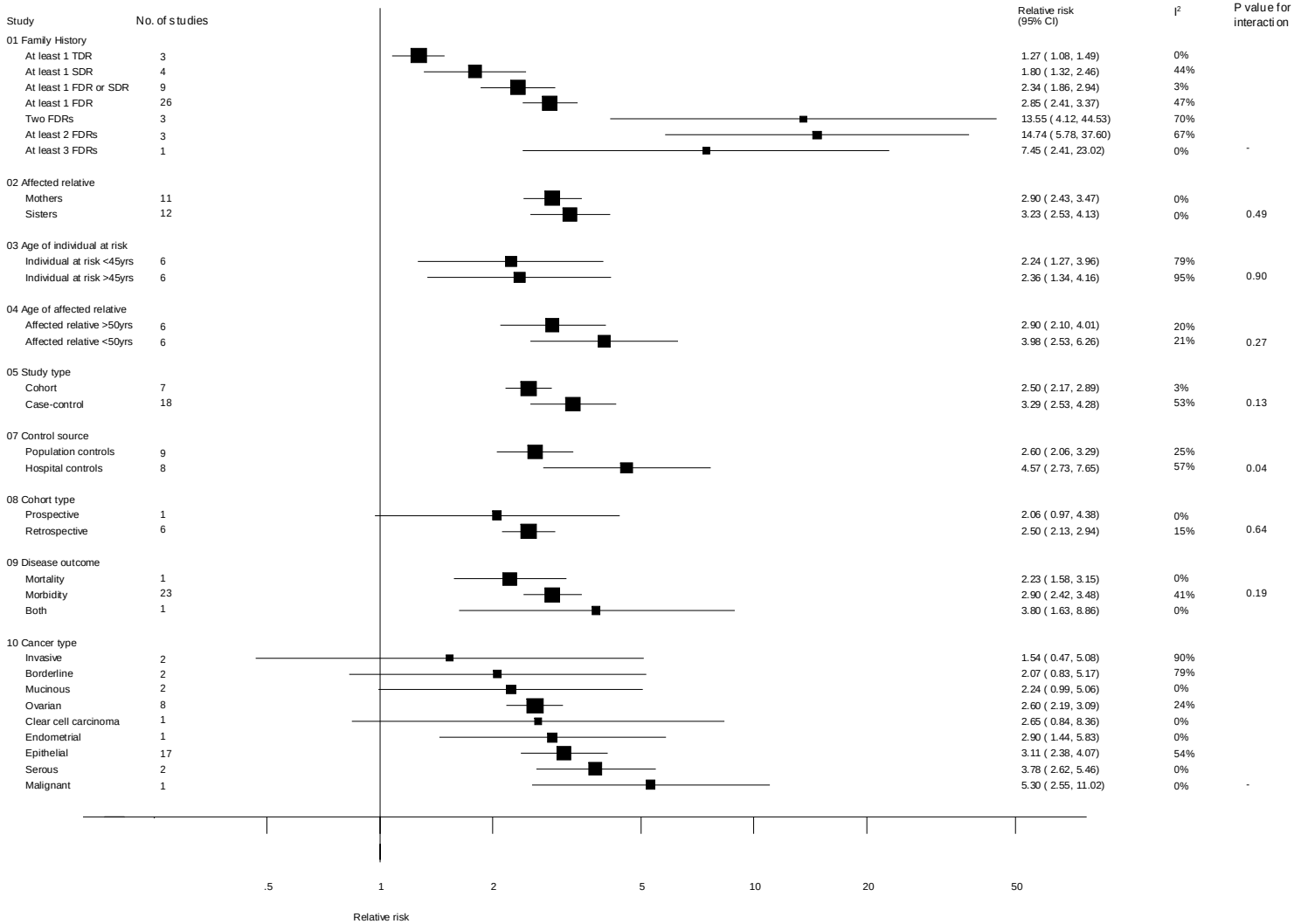
From the Stratton review, we included 11 of the studies, excluded 8 of the studies and found updated information for one study. We performed an additional literature search in the same databases as for other cancers which yielded a further 181 potential papers for inclusion. From these we included another 22 that were either published more recently or missing from the previous review. We used the same methodology as with the other cancers to extract the data from the studies and pooled the estimates using both fixed and random-effects meta-analysis models.

7.3. Relative risk estimation

Our final dataset included 33 studies published between 1981 and 2005 comprising 24 case-control studies and 9 cohort studies. In total, these studies included nearly 12,000 cases and over 25,000 disease-free controls. The pooled relative risk estimate from all 33 studies was 2.92 (95% CI 2.50, 3.41), which was higher than those of the other cancers (Table 1).

Of the 33 included studies, 26 presented data on the risk associated with having at least one affected first-degree relative. The fixed effect pooled estimate was 2.70 (95% CI 2.43, 3.00) which was not significantly different from that under a random-effects model (RR = 2.85, 95% CI 2.41, 3.37). As there was moderate heterogeneity between study results (I^2 = 47%, 95% CI 15, 67), a random-effects model was used for other pooled relative risk estimates.

Figure 11. Forest plot of the pooled relative risk estimates for subgroups of studies estimating the risk associated with a family history of ovarian cancer.



Each line represents a pooled subgroup risk estimate with the width of the horizontal line indicating the 95% CI, the position of the box representing the point estimate and the size of the box being proportional to the weight of each subgroup category.

The subgroup relative risk estimates are presented in Figure 11. Although there were only 3 studies that had investigated the risk of having multiple affected relatives, a highly significant risk estimate of 14.74 (95% CI 5.78, 37.60) was seen, although this had a high degree of heterogeneity (I^2 = 47%) reflecting the variation in the study estimates. This was far higher than the risk associated with multiple relatives in other cancers, although much more evidence is needed to estimate this risk more accurately.

There were slightly more studies (n = 6) that estimated age-specific relative risk estimates. The cut-off point for the age of the individual at-risk was drawn at 45 based on the available data, whilst 50 was used for the age of the affected relative. Unlike other cancers, no significant difference was seen between younger and older individuals (p value for interaction test = 0.9), although there was high heterogeneity in the study estimates. There was an increased risk associated with relatives affected below age 50 (RR = 3.98, 95% CI 2.53, 6.26) compared to older affected relatives (RR = 2.90, 95% CI 2.10, 4.01), although again this was not statistically significant (p = 0.27).

The only subgroup that showed significantly different risk estimates was the source of controls for case-control studies (Figure 11). The risk in the 9 studies that used population controls was 2.60 (95% CI 2.06, 3.29) whilst those that used hospital controls had a pooled estimate of 4.57 (2.73, 7.65). No significant differences were seen between the risks for different study types, relative types, cohort types or disease outcomes. Attempts to investigate risk differences in different ovarian cancer subtypes were hampered by a variety of listed subtypes and small numbers of studies on each.

The funnel plot for these 26 studies was asymmetric suggesting the presence of a study size effect, which was confirmed by Egger's test (p = 0.039). The trim-and-fill method added a further 4 'missing' studies to produce a symmetric funnel plot, which reduced the pooled RR estimate to 2.55 (95% CI 2.10, 3.10) suggesting that publication bias may be inflating our pooled estimates.

From the plot of mean age versus relative risk, there appears to be little association between age and risk estimate (Appendix 1). Although there is a wide range of mean and overall ages, there is little variation in relative risk (with 3 notable exceptions) making a trend hard to find. Despite this apparent lack of association, we implemented a meta-regression of study effect against mean age for confirmation. Using data from the 18 studies from which we were able to obtain a mean age gave an equation for RR of $e^{(0.0051 + 0.7885 \times age)}$. This suggests a slight increase in relative risk with increasing age although this was not significant.

7.4. Absolute risk estimation

The cumulative absolute risk estimates for morbidity of OC are shown in Figure 12. Due to the high risk associated with having multiple affected relatives, the maximum 10-year risk for 2 or more affected FDRs was much higher at 8.4% (95% CI 3.4, 20.3) than the risk for one affected FDR (1.4%, 95% CI 0.8, 2.5) or the general population risk of just 0.6%. The risk curves for 20-year risk showed a similar pattern.

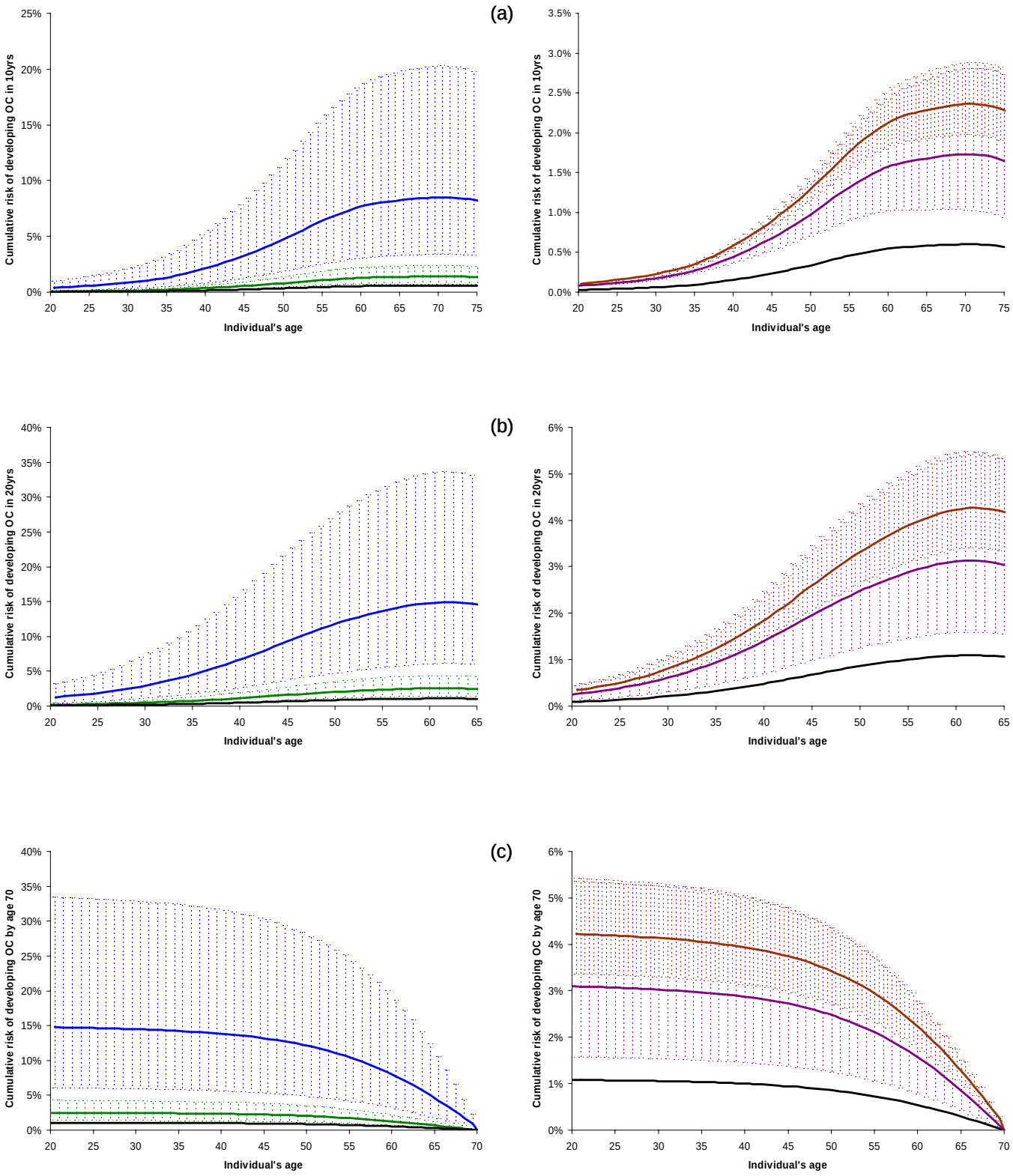


Figure 12. Cumulative absolute risks of developing ovarian cancer for different categories of family history and ages of affected relatives over (a) 10yrs, (b) 20yrs and (c) by age 70.

General population; One first-degree relative; Two first-degree relatives;

Affected relative <50 years; Affected relative >50 years.

Dashed vertical bars show 95% upper and lower confidence intervals for the risk estimates. Note: graphs have different Y-axes

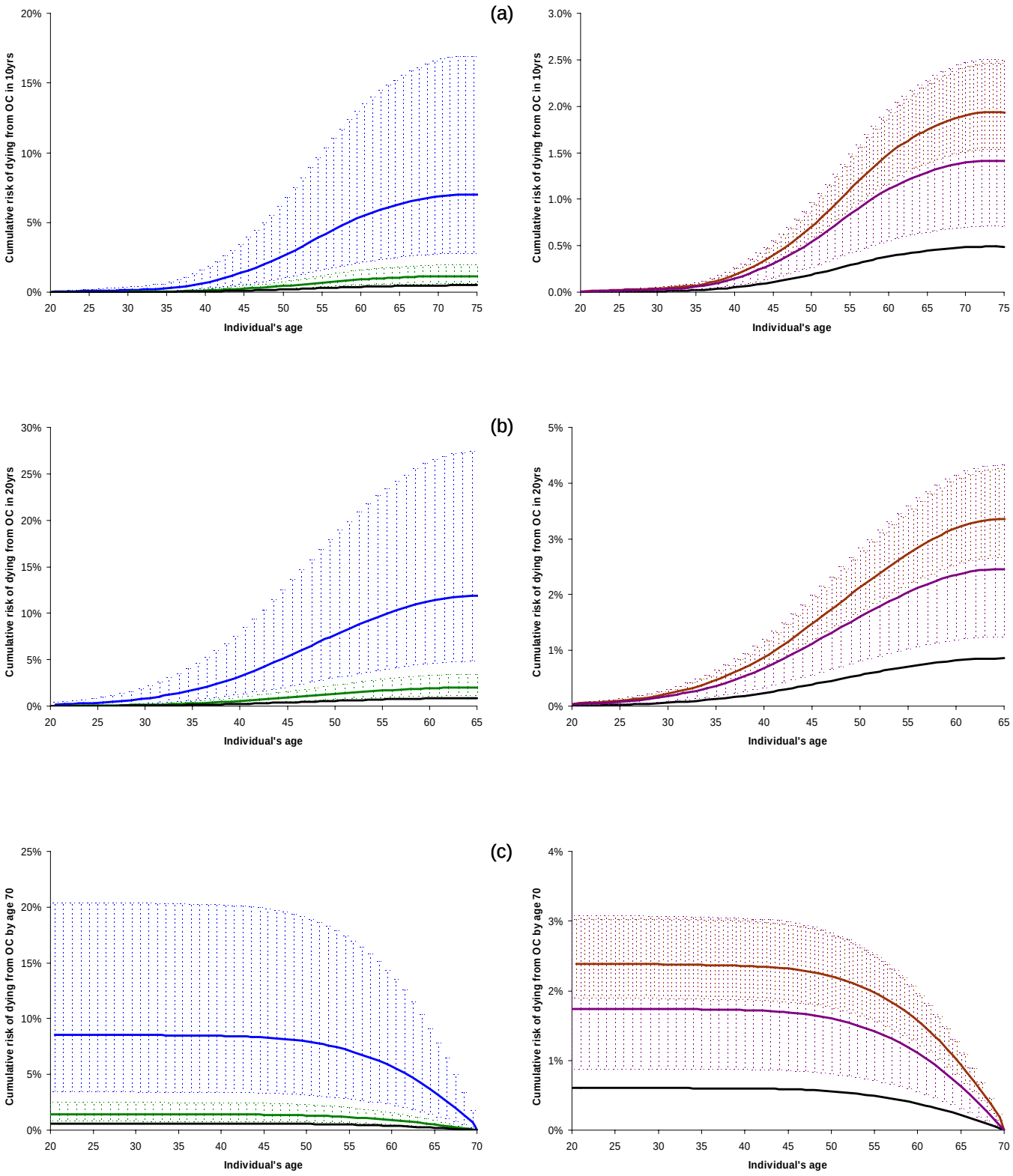


Figure 13. Cumulative absolute risks of dying from ovarian cancer for different categories of family history and ages of affected relatives over (a) 10yrs, (b) 20yrs and (c) by age 70.

General population; One first-degree relative; Two first-degree relatives;

Affected relative <50 years; Affected relative >50 years.

Dashed vertical bars show 95% upper and lower confidence intervals for the risk estimates. Note: graphs have different Y-axes

The general population lifetime risk was just 1.1% (~1 in 90) for a woman aged 30. This increased to 2.5% (1 in 40) with one affected first-degree relative and 14.5% (~1 in 7) for those with multiple affected first-degree relatives. For individuals with an affective relative diagnosed at less than 50, the lifetime risk was 4.1% (95% CI 2.6, 6.4), which was higher than that for those with older affected relatives at 3.1% (95% CI 2.2, 4.2).

Due to low survival rates in ovarian cancer, the absolute risk curves for mortality were very similar to those for morbidity, but only slightly lower (Figure 13). The maximum 10-year risk for the general population was 0.5% at age 75. This increased to 1.2% (95% CI 0.7, 2.0) for those with an affected first-degree relative and 7.0% (95% CI 2.8, 16.9) with two or more. Lifetime mortality risks peaked at 0.6% in the general population, 1.4% (~1 in 70) with one affected FDR and 8.5% (~1 in 12) with multiple affected relatives.

8. Cerebrovascular disease

8.1. Background

We have shown in a number of common cancers that the methodology we have used is a simple and effective, if labour-intensive way of generating absolute risk estimates for different family histories. However, it is more difficult to apply the same methodology to other complex diseases.

Cerebrovascular disease (ICD10 C60-69) is not as easy to define as cancer, where we only see a small number of subtypes in most cancers such as ovarian or lung cancer. There is a range of different types of cerebrovascular diseases which can be classified in different ways. This is shown by the large number of ICD headings that are incorporated under the umbrella term of 'stroke'. As there are also a number of ways of diagnosing stroke or cerebrovascular incidents, there are many different definitions used in epidemiological studies that have looked at family history. This causes heterogeneity between studies and makes it difficult to group studies into appropriate subgroups.

Another difficulty with non-cancerous conditions is the lack of available incidence data. When studying cancers, it is easy to find data on morbidity collected annually from national cancer registries and collated and reported by the ONS. Unfortunately for epidemiologists, no such registries exist for the majority of common, chronic diseases in the UK. Instead, we must rely on prospective studies of incidence, which are usually only regional due to the costs and difficulties of carrying out such a study nationwide. Whilst some of these studies may cover large areas and time periods, they are very unlikely to be as accurate at predicting disease incidence as registries with near complete ascertainment.

Like the cancers we have looked at, susceptibility to stroke is given by both genetic and environmental factors, as well as the interaction between them. Although smoking has been proven as a strong risk factor for stroke, there are no confirmed genes with strong causal effects on common stroke, although genes such as *MTHFR* and ACE have been proposed to confer moderate risk on individuals carrying rare polymorphisms within them (Casas *et al.*, 2004). As with lung cancer, it is likely that both smoking and genetic factors are likely to underlie any familial aggregation of stroke.

8.2. Methodology

Flossmann and colleagues carried out a recent systematic review and meta-analysis of familial risk of stroke (Flossmann *et al.*, 2004). They included 32 studies, 28 case-control and 4 prospective cohort studies, and found a pooled risk of 1.76 (95% CI 1.7, 1.9) for case-control studies and 1.30 (95% CI 1.2, 1.5) for cohort studies. However, they found a high degree of heterogeneity between the studies.

As the previous review used different methods from ours (e.g. no retrospective cohort studies, only including studies with raw data), we re-analysed the studies described in their paper, as well as carrying out a search to find additional studies that met our inclusion criteria or new studies that were published after their study.

Due to the lack of national stroke registries, we searched the literature for recent regional estimates of age-specific stroke incidence. The most promising estimates came from prospective

studies in Oxfordshire (Rothwell *et al.*, 2004) and South London (Wolfe *et al.*, 2002). These both had well stratified age-specific incidence estimates from relatively large populations. As both studies produced similar incidence rates, we used the data from the OXVASC study (Rothwell *et al.*, 2004) for the life tables as it contained results from a more recent analysis. Mortality data were obtained in the usual way from the ONS and life tables were constructed as for the cancers.

8.3. Relative risk estimation

The new dataset included a total of 63 studies, 45 case-control studies and 18 cohort studies, published between 1966 and 2005. In total these studies included over 25,000 cases, as well as more than 30,000 disease-free controls. The pooled relative risk estimates can be seen in Table 1/Figure 14. The risk for those with an affected first-degree relative was similar to most of the cancers (RR = 1.73, 95% CI 1.52, 1.97), although the risk was slightly lower than twice the population risk. This was slightly higher than the fixed effect estimate of 1.63 (95% CI 1.53, 1.73) which can be attributed to the high heterogeneity between the study estimates, hence the random-effects model was used for further analysis.

With relative risk estimates ranging from 0.9 to 4.2, there was some uncertainty about the 'true', underlying risk. Although only 3 of the studies had relative risk estimates of less than 1, a further 11 of the 36 studies had relative risks that were not significantly above the null hypothesis of no increased risk. This was highlighted by the I^2 score of 75%. Random-effects subgroup analysis was used to explore possible reasons for heterogeneity in study estimates.

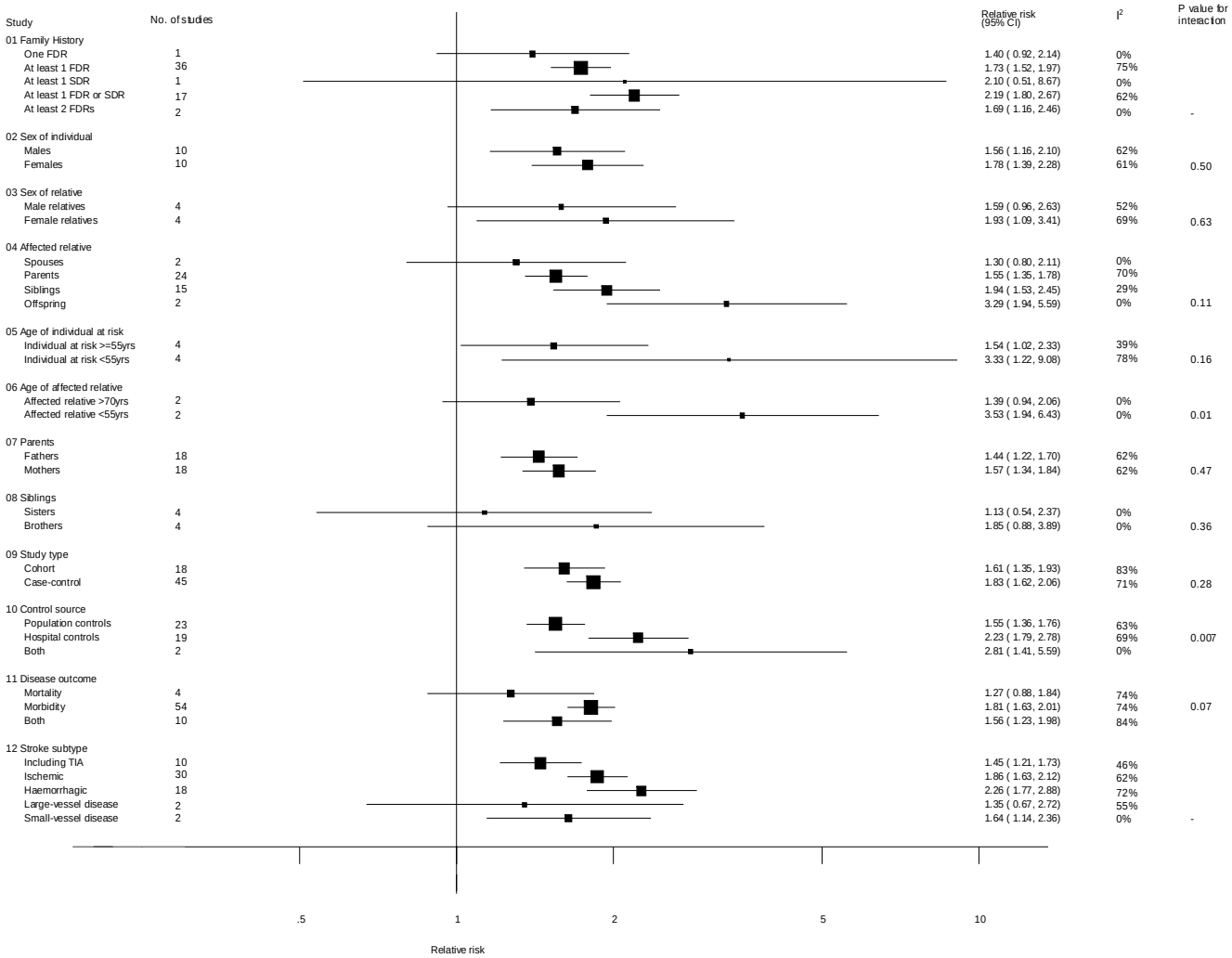
Unlike the cancer results, the risk of having multiple affected relatives was not greater than the risk with one affected first-degree relative (RR = 1.69, 95% CI 1.16, 2.46), although this was only based on data from two studies. Equally, there were too few studies with available data to show a decrease in risk with more distant relatives.

Although only four studies had stratified risks by the age of the individual at-risk, a difference in risk was seen, with a relative risk of 3.33 (95% CI 1.22, 9.08) for those under 55 and 1.54 (95% CI 1.02, 2.33) for older individuals (difference not statistically significant; p = 0.16). There were two further studies that stratified by age of individual at-risk, but they had divided at older ages (65 and 75) and it was therefore inappropriate to attempt to pool these data with the other studies. One of these studies also showed a higher risk for younger individuals, however another study found an opposing effect, albeit statistically insignificant (Khaw and Barrett-Connor, 1986).

There were also four studies that stratified by the age of the relative at diagnosis although again at different ages. As two studies split at 50/55 and two at 70, we investigated the risks to those with younger affected relatives (<50/55) and those with older relatives (>70). A significant difference was seen with higher risks for those with younger relatives (RR = 3.53, 95% CI 1.94, 6.43) than those with older relatives (RR = 1.39, 95% CI 0.94, 2.06; interaction p = 0.01).

The risks in various subgroups were generally lower than those seen in the cancers with many not being significantly higher than 1. The only subgroups that showed a significant difference in risk were the use of population controls versus hospital controls (p = 0.007), whereas no differences were found between the risks for different study types, affected relatives, disease outcomes or sex of individuals or relatives. There was some variation in the risks for different stroke subtypes e.g. ischemic versus haemorrhagic stroke, although whether this could account

Figure 14. Forest plot of the pooled relative risk estimates for subgroups of studies estimating the risk associated with a family history of stroke.



Each line represents a pooled subgroup risk estimate with the width of the horizontal line indicating the 95% CI, the position of the box representing the point estimate and the size of the box being proportional to the weight of each subgroup category.

for the high level of overall heterogeneity seen is unclear. Unfortunately it was not possible to find recent UK age-stratified incidence data on the different stroke types in order to convert these relative risk differences to absolute risk differences. As ischemic stroke makes up around 85% of all incident strokes in the UK, there are likely to be large absolute risk differences, regardless of the difference in familial relative risks (Rothwell *et al.*, 2005).

Although Egger's test for publication bias did not show a statistically significant study size effect ($p = 0.06$), the trim-and-fill method added 8 further studies to the affected first-degree relative risk producing a random-effects estimate of 1.50 (95% CI 1.30, 1.72), which was lower than our pooled estimate, suggesting that there maybe further data available that were unpublished or not found by the search strategy. For the whole dataset, this effect was even more extreme with a further 18 additional studies reducing the risk to 1.47 (95% 1.33, 1.64).

Thirty-one studies had age data for the category 'at least one affected first-degree relative' which were used for a meta-regression model to test the age-specific effect of the familial risks. The model produced a best-fit equation of $RR = e^{(1.466-0.01433 \times age)}$, which was a good fit to the stratified risks for the age of the individual at-risk although with wide confidence intervals. Any trend is difficult to see from the plot (Appendix 1), although there is a wide spread of mean ages, ranging from 34 to 85. The range of participants' ages covers most of adult life with individuals from 20 to over 90 included.

8.4. Absolute risk estimation

As there was little difference between the risk with one affected first-degree relative and the risk with multiple relatives, absolute cumulative risk curves are presented for just the general population and one affected first-degree relative (Figures 15 and 16). The 10-year morbidity risks for those with and without a family history increased up to the age of 75. Compared to the population risk of 9.3% at age 75, the risk for an individual with a family history was significantly higher at 13.9% (95% CI 9.4, 20.3). A similar pattern was seen for the 20-year risk curves, with a maximum of 18.8% (95% CI 12.9, 26.9) reached at age 65 in those with a family history.

The lifetime risks decreased from early middle age to 0 at age 70, with a maximum population risk of 6.1% increasing to 11.6% (~1 in 9) in those with a family history. This risk was greater if the relative was diagnosed under the age of 55 at 19.8% (~1 in 5), but lower for those with relatives diagnosed over 70 (8.3%).

Similar patterns were seen for the risks of mortality from stroke although with much lower risks. The maximum 10-year mortality risk was 4.1% (95% CI 2.7, 6.2) for those with a family history, which was higher in those with younger affected relatives (AR<55 = 9.2%, AR>70 = 3.7%). The lifetime mortality risk in the general population was only 1.7% at ages less than 50, reflecting the high survival rates from primary incident stroke. With a family history, this rose to 2.8% (~1 in 35), which increased again to 6.0% (~1 in 17) with a younger affected relative compared to 2.4% (~1 in 40) in those with relatives diagnosed above age 70, which was not significantly different from the general population risk.

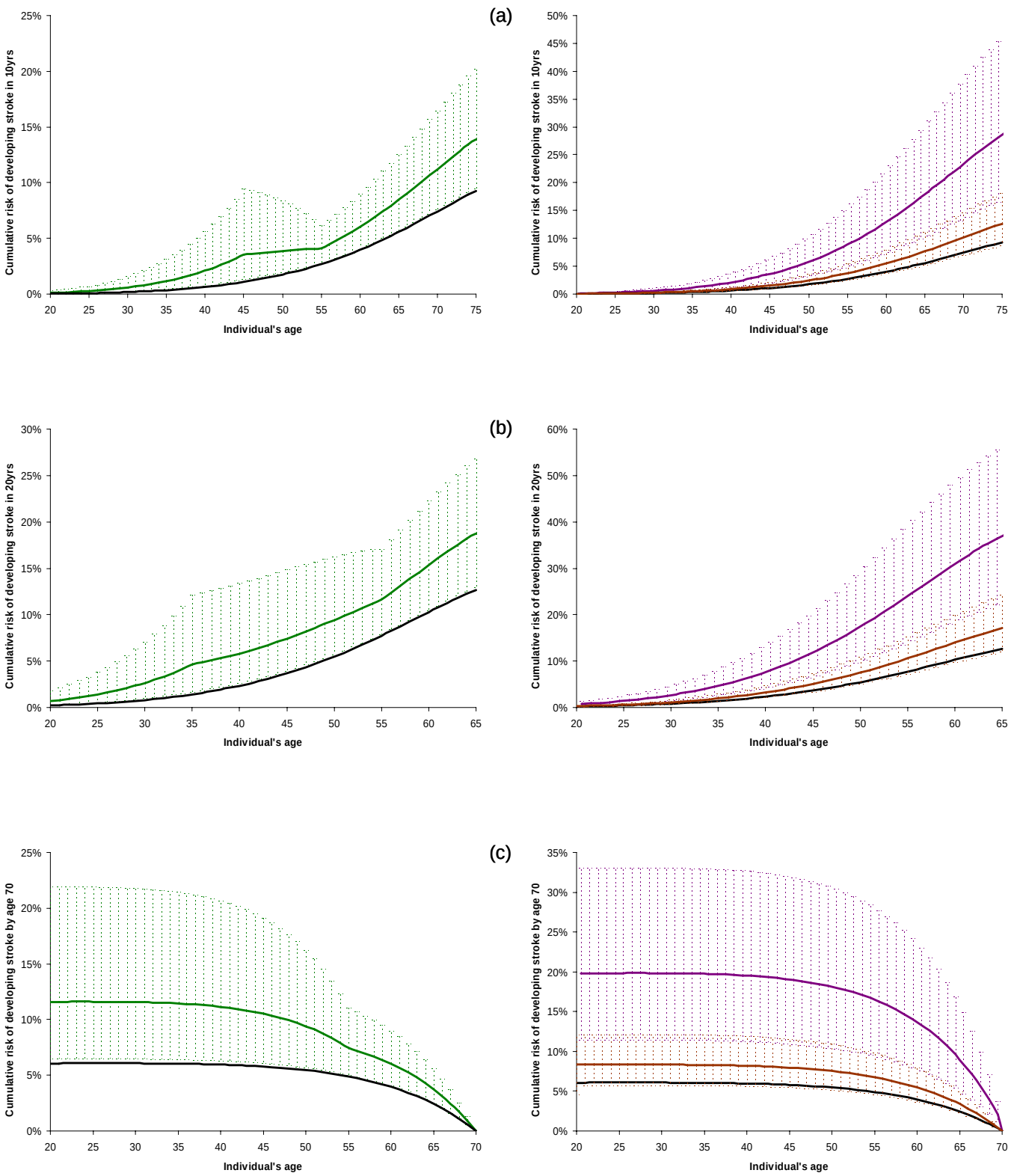


Figure 15. Cumulative absolute risks of developing stroke for different categories of family history and ages of affected relatives over (a) 10yrs, (b) 20yrs and (c) by age 70.

General population; One first-degree relative;

Affected relative <55 years; Affected relative >70 years.

Dashed vertical bars show 95% upper and lower confidence intervals for the risk estimates. Note: graphs have different Y-axes

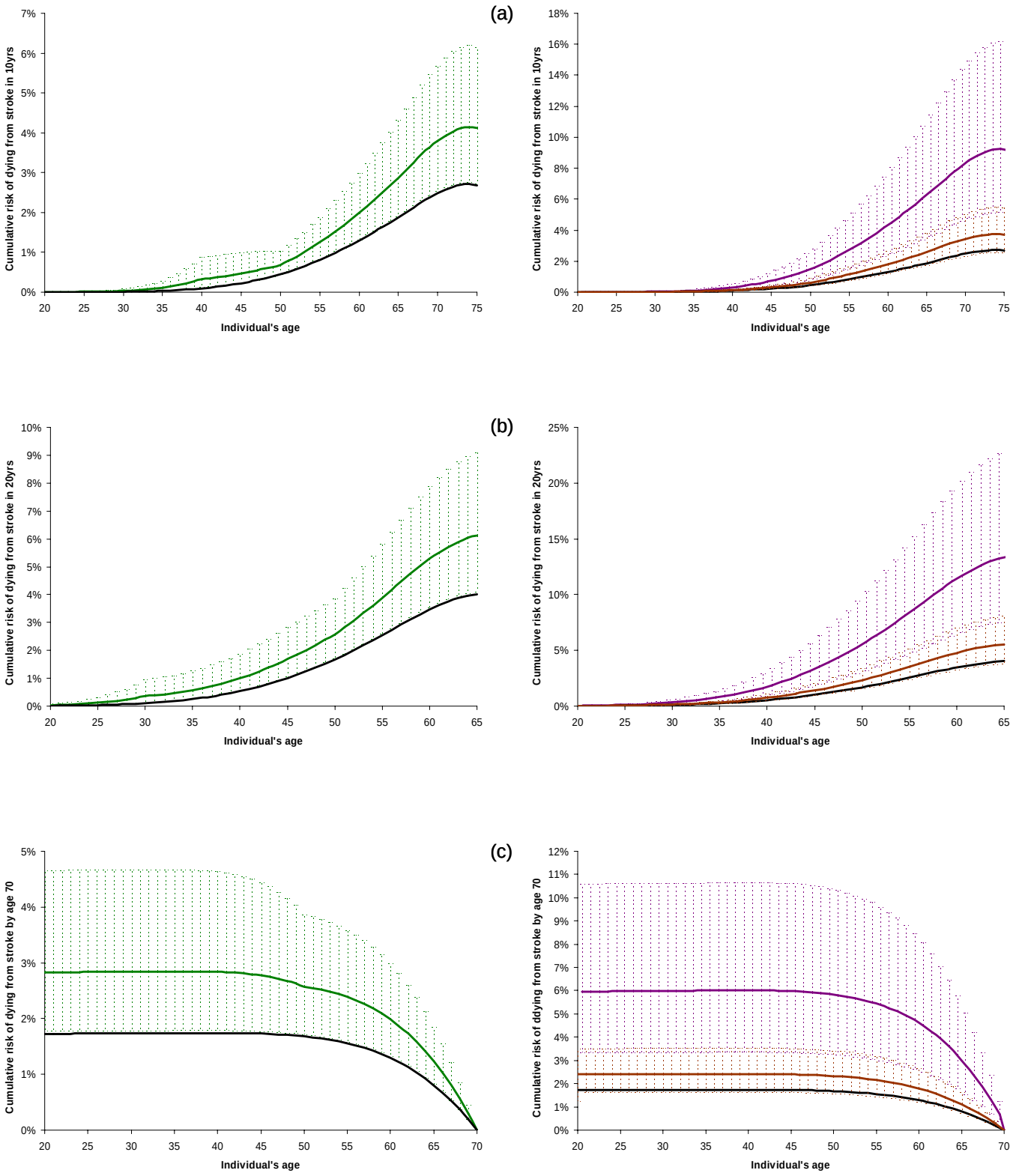


Figure 16. Cumulative absolute risks of dying from stroke for different categories of family history and ages of affected relatives over (a) 10yrs, (b) 20yrs and (c) by age 70.

General population; One first-degree relative;

Affected relative <55 years; Affected relative >70 years.

Dashed vertical bars show 95% upper and lower confidence intervals for the risk estimates. Note: graphs have different Y-axes

9. Multiple Sclerosis

9.1. Background

Multiple sclerosis (ICD10 G35) is a demyelinating disorder of the central nervous system that has a variety of chronic symptoms. It has a younger age of onset than most cancers or stroke, with the highest incidence seen between the ages of 20 and 50, and prevalence studies report more than double the number of cases in women as in men. An interesting factor in the epidemiology of MS is a well-studied 'latitude effect' which has hypothesised that populations at more extreme latitudes have higher incidence of MS (Ebers and Sadovnick, 1993). Whether this is due to measurement or disease classification differences, or is caused by an unidentified MS risk factor, it is clear that certain populations e.g. Scandinavians, have a much higher incidence of MS than others e.g. sub-Saharan Africans (Pugliatti *et al.*, 2002).

Like cerebrovascular disease, there is no national registry of MS cases in the UK. Instead, incidence estimates are based on regional prospective studies such as those in Cambridge (Robertson *et al.*, 1996), Leeds (Ford *et al.*, 2002) and Scotland (Rothwell and Charlton, 1998). As the total population incidence is very low compared to the other chronic diseases we have studied, these prospective studies need to be extremely large to accurately estimate incidence, particularly age-specific incidence. Due to the variation in incidence seen at different geographical locations, it is unreliable to extrapolate the results of any one study to an entire country and hence it is tricky to obtain accurate age-stratified incidence estimates for the UK.

MS is thought to be an autoimmune disorder with a number of potential causes and risk factors. Numerous environmental triggers have been proposed such as viruses like the herpes virus or Epstein-Barr virus, which are known to cause other demyelinating diseases, or heavy metals e.g. lead poisoning (Kantarci and Wingerchuk, 2006). It has been reported for many years now that MS has a tendency to cluster in families which is surprising given the extremely low incidence in the population. This has suggested that there are heritable factors involved in MS susceptibility. Despite a number of candidate gene loci being suggested by whole-genome scans of multiplex MS families (GAMES and the Transatlantic Multiple Sclerosis Genetics Cooperative, 2003), only the *MHC* region has shown repeated promise as a true susceptibility locus for the disease (Compston, 2000).

9.2. Methodology

For the retrieval of studies for the systematic review of familial MS risk, we used a similar search strategy to those used for the other diseases, with terms like "multiple sclerosis" and "demyelinating" in the disease section. From searching the usual databases, we retrieved over 5,000 studies, which were reduced to 450 studies through title and abstract reading and hand-searching of relevant references. However, this process was not as straightforward in MS as with the cancers. Searching the same databases with terms related to multiple sclerosis retrieved many older articles (including one from the 19th century!) which were harder to obtain full-text copies of, and also many more studies in foreign languages, particularly less common ones such as Polish, Hungarian and Russian which were difficult to translate, making the systematic review process more time-consuming than in other disease areas.

The situation with MS is also not as straightforward as with cancer due to the types of study that have been performed. For the other diseases, we commonly see 3 kinds of familial study; i) case-control studies, where family history is compared in diseased cases and non-diseased

controls, ii) prospective cohort studies, in which a group of non-diseased individuals has their family history status determined and is then followed up for disease status, or iii) retrospective cohort studies, where the disease incidence in a cohort of relatives of diseased cases is compared to that of a background population. However for MS, a disease with much lower incidence, there are complications with each of these.

In a typical case-control study with fewer than 1,000 cases, there are likely to be very few controls with affected family members to enable an accurate odds ratio to be calculated. Prospective cohorts are mostly used for more common diseases so that enough cases will develop over a reasonable follow-up period even in a small cohort. For a low-incidence disease like MS, a very large cohort or extremely long follow-up time will be necessary to generate enough cases for accurate relative risk calculation. With retrospective cohorts, the lack of valid population incidence data makes it difficult to find a comparison incidence group.

For these reasons, only 20 studies with published familial risk estimates were discovered in our systematic review. However, we also found a number of studies (n = 12) that have estimated both the recurrence risk in families i.e. the incidence of MS in relatives of cases without reference to a control population, as well as the population prevalence of disease. From these two measures, a relative risk and 95% CI were estimated using a Poisson distribution model.

Risk estimates were combined as with other diseases and heterogeneity statistics and P values from interaction tests were calculated in the usual way. As only 2 studies had attempted to stratify risk by age and these were both using differing stratifications, no pooling of age-specific risk estimates was performed (Warren *et al.*, 1991; Hemminki *et al.*, 2006). Presence of a general age trend was assessed using a WinBUGs model as described previously.

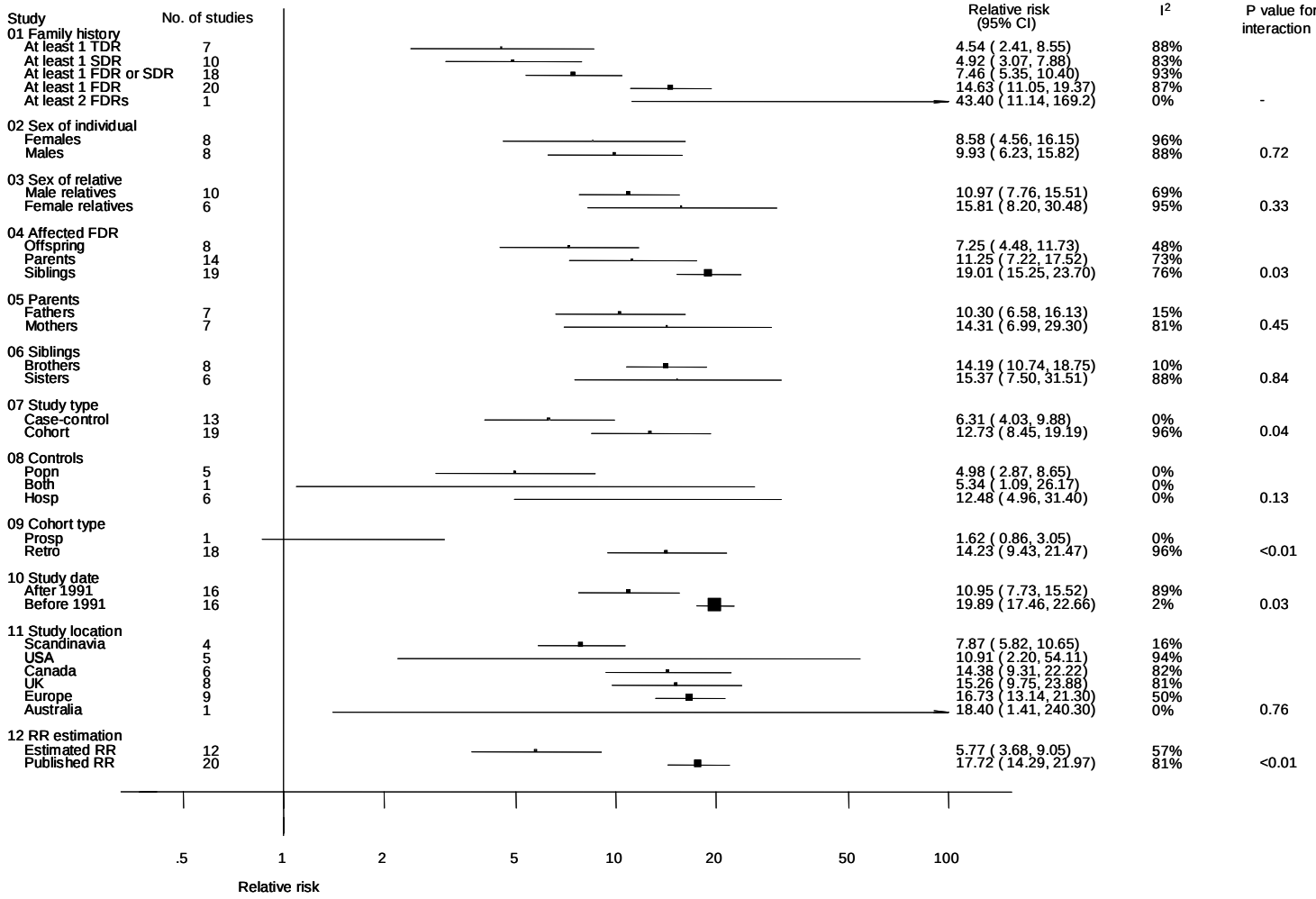
Due to the lack of national registry data for MS, data from regional registries were searched to find any age-specific prevalence or incidence data. A study in the Lothian and Borders region of Scotland had published age-specific prevalence and incidence data, which, when combined with population estimates from the relevant time period, were used to estimate age-specific incidence rates (Rothwell and Charlton, 1998). This is an area of high MS prevalence so these rates are likely to be overestimates for the UK as a whole. Mortality data were again obtained from the ONS and life tables were constructed in the usual way.

9.3. Relative risk estimation

32 relevant studies were found containing relative risk estimates that could be used in this analysis, of which 13 were case-control studies and 19 cohort studies, usually retrospective in design. The year of publication ranged from 1951 to 2006 with nearly half the papers published pre-1990, making this the 'oldest' dataset used in this project. There were over 20,000 cases of multiple sclerosis included in these studies, although over 12,000 of these came from the two largest studies (Nielsen *et al.*, 2005; Hemminki *et al.*, 2006).

The relative risk estimates associated with having at least one affected first-degree relative ranged from 1.6 to 49.5 with a pooled estimate of 14.63 (95% CI 11.09, 19.44) (Table 1/Figure 17). A random-effects model was deemed to be most appropriate due to the high degree of heterogeneity between the studies (I^2 = 87%) although the fixed effect estimate was lower (RR = 13.51, 95% CI 12.42, 14.70). With such high estimates, only 3 studies had confidence intervals that crossed the null (RR = 1) with the majority of estimates ranging from 7 to 25. There was no evidence of a study size effect with smaller studies having similar estimates to larger studies.

Figure 17. Forest plot of the pooled relative risk estimates for subgroups of studies estimating the risk associated with a family history of multiple sclerosis.



Each line represents a pooled subgroup risk estimate with the width of the horizontal line indicating the 95% CI, the position of the box representing the point estimate and the size of the box being proportional to the weight of each subgroup category.

As with other diseases, the risks were lower for individuals with more distant relatives and higher for those with multiple affected relatives, although there was only 1 study in this instance that had estimated the risk of having at least two first-degree relatives (Figure 17). Although there appeared to be no differences in risk according to the sex of either the proband or the affected relative, significantly higher relative risks were seen for siblings (RR = 19.01, 95% CI 15.25, 23.70) compared to parents (RR = 11.25, 95% CI 7.22, 17.52).

A difference in risk estimates was seen between the case-control studies and the cohort studies and between different types of cohort study, although there was only a single prospective follow-up study to compare with 18 retrospective studies. Other statistically significant differences were seen between studies which had published relative risk estimates (n = 20, RR = 17.72, 95% CI 14.29, 21.97) and those in which relative risk had to be estimated (n = 12, .RR = 5.77, 95%CI 3.68, 9.05) as well as between earlier studies (pre-1991) and more recent studies (post-1991). Despite large reported differences between the prevalence of MS in different countries (and particularly at different latitudes), there appeared to be little difference in the relative risk estimates between geographic locations, with the possible exception of the Scandinavian studies which appeared to have lower estimates.

Although it was not possible to estimate pooled age-stratified risk estimates, a meta-regression model on the studies that had relevant data showed a slight, non-significant increase in risk with age ($RR = e^{[1.81+0.016 \times age]}$). The mean ages of these studies were clustered in the 20-50 age range, although a small number of studies had age ranges extending outside these limits (Appendix 1).

9.4. Absolute risk estimation

As no age-stratified risks were estimated in the meta-analysis, the overall relative risk estimate associated with having at least one affected first-degree relative was used in the life tables to produce the absolute risk estimates. The absolute risk curves for morbidity and mortality are shown in Figure 18. The general population risk of developing MS in the next 10 years is extremely low, peaking at approximately 0.2% around age 30 With at least one affected FDR, the 10-year risk increases dramatically with a maximum of 2.8% (95% CI 2.1, 3.7) in the early 30s. For multiple relatives, this risk increases to 8.2%, (95% CI 1.0, 14.8) although there are wide confidence intervals due to the fact that the relative risk estimate is derived from only 1 study. The risk curves are similar but higher for the 20-year risks with maximum estimates of 0.4%, 5.2% (95% CI 4.0, 6.8) and 14.7% (95% CI 1.9, 25.9) for the general population, those with one FDR affected and those with two affected FDRs.

The lifetime risks (by age 70) of being diagnosed with MS start high, but then decline quickly through early and middle age. The population risk is 0.6% (~1 in 170), which decreases from age 25. This risk is 8.7% (~1 in 11) for those with an affected relative and 23.8% (~1 in 4) for multiple relatives. The mortality lifetime risks are lower, but remain constant for longer, showing the long duration of the disease in most patients. For the general population, the maximum lifetime risk of death of mortality is 0.1% (1 in 1000), which increases to 1.3% (~1 in 75) with one affected relative and 3.9% (~1 in 25) for those with multiple affected relatives.

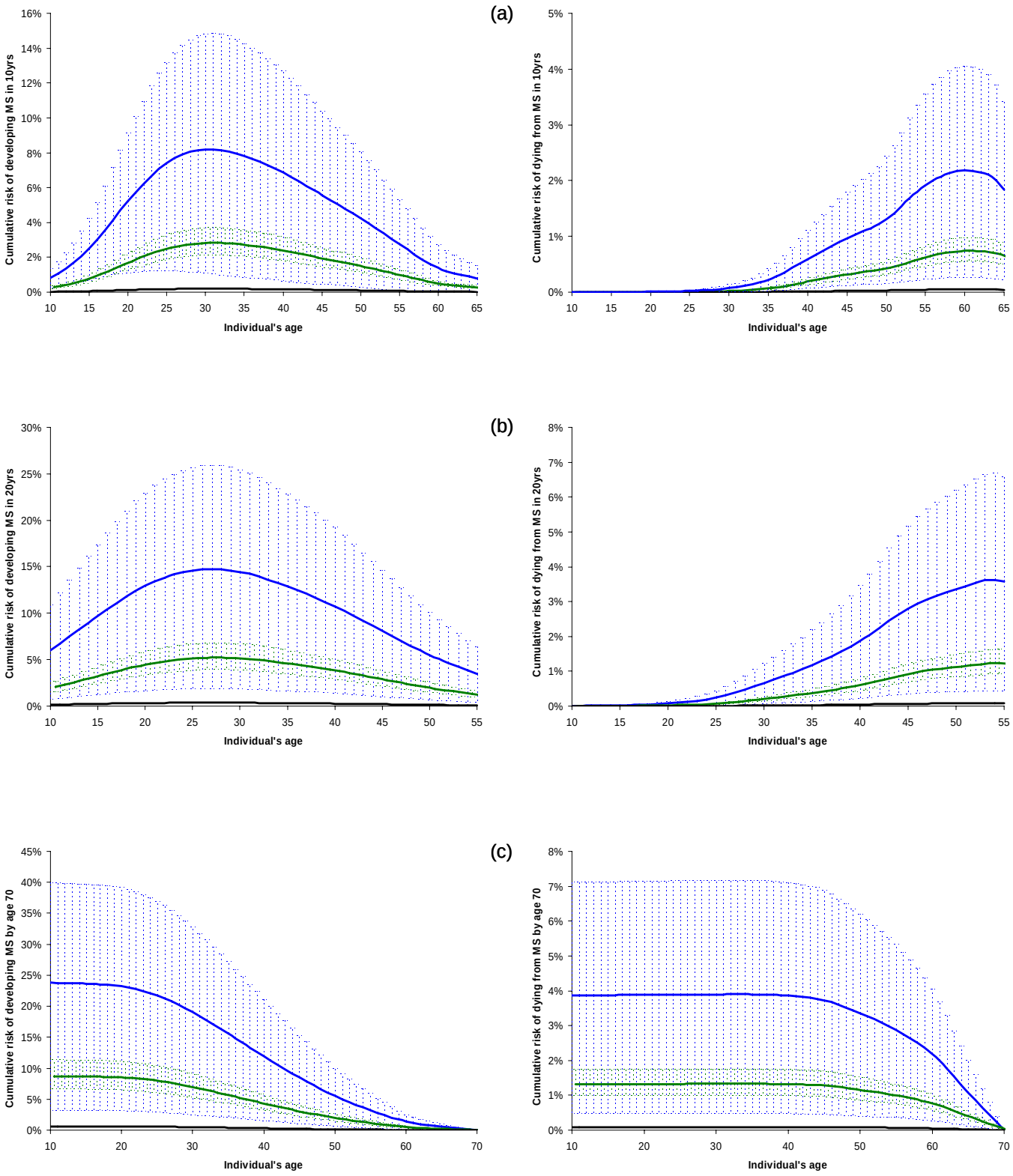


Figure 18. Cumulative absolute risks of developing (left) or dying from (right) multiple sclerosis for different categories of family history and ages of affected relatives over (a) 10yrs, (b) 20yrs and (c) by age 70.

General population; One first-degree relative; Two first-degree relatives;

Dashed vertical bars show 95% upper and lower confidence intervals for the risk estimates. Note: graphs have different Y-axes

10. Discussion

Familial patterns of disease have interested geneticists and epidemiologists for many decades leading to a wealth of literature in this area. From early reports of multicase families to the modern multiplex family studies, the ways in which families with multiple disease cases are studied to determine the causes and risk factors for a wide range of diseases have evolved over time. However, having a family member affected with a disease has remained a risk factor for many diseases throughout this period.

In this independent, epidemiological project we have systematically reviewed the literature estimating the familial risks of a number of common, complex diseases and attempted to synthesise accurate, pooled risk estimates from the data. For common, well-studied diseases such as prostate, colorectal and breast cancer, there are large numbers of studies and accurate pooled relative risk estimates can be produced. Even for rarer cancers, such as ovarian cancer, it is still possible to estimate risks for basic family history categories, such as the presence or absence of affected first-degree relatives. However, with fewer background data to work from, it is more difficult to precisely estimate age-specific risks and the risks associated with multiple relatives. It is likely that for cancers with lower incidence that have been less well studied, similar difficulties would be encountered until large, age-stratified familial studies are conducted and published.

For all the diseases studied here it was possible to gather enough evidence to accurately estimate the relative risk associated with a family history of disease. In the case of the cancers, the risks ranged from 1.8 to 2.9 with narrow confidence intervals, particularly for the more common cancers. The risk for stroke was slightly lower (yet still significantly above 1), whereas for multiple sclerosis, a much higher relative risk was seen. The relative risks associated with multiple relatives were generally higher, especially for ovarian cancer and multiple sclerosis, but not significantly for lung cancer and stroke. There were far fewer data available on age-specific relative risks giving less power to detect significant differences in risk. Significantly higher pooled estimates were seen for younger individuals in colorectal cancer, prostate cancer and breast cancer, but the decrease in risk with age was non-significant for lung cancer and stroke, where the fewest data were available. With more age-specific data, it will be easier to clarify the effect that age has on the familial relative risks.

Despite the wealth of evidence we found on simple familial risks in common disease areas and our confidence in the sensitivity of our search strategy, it is possible that we may have missed some studies that would be suitable for inclusion. As we didn't limit our reviews to studies in which estimating familial risk was the main aim, the majority of our evidence comes from studies investigating other exposures. This helped to limit the amount of publication bias in our systematic reviews as there is little reason for non-significant familial risk estimates to remain unpublished when they are not the main focus of the publication. However, it can be difficult to discover these studies as terms related to family history are rarely included in the titles, abstracts or keywords of articles and most bibliographic databases only allow these areas to be searched. In the future, the ability to search the full-text of articles would greatly enhance our ability to accumulate evidence on familial risks, regardless of disease area.

Through the use of life tables, relative risks can be converted into absolute risks for varying time periods. This is straightforward in the case of the cancers where registry data are freely available on incidence and mortality rates. By applying familial relative risks to these population rates, cumulative absolute risks can be estimated. Outside the field of cancer, population data are less

readily available due to the lack of national registries. In the cases of stroke and multiple sclerosis, the best data are obtained from regional prospective studies, which can be evaluated and extrapolated to whole populations. However, there may be geographical and temporal variation in these data which makes the methodology prone to error. The absolute risk estimates in these cases are less reliable than those of the cancers.

Where there are sufficient data to accurately estimate relative familial risks, the absolute risk curves have narrow confidence intervals. However for categories with fewer available data e.g. risk associated with multiple affected relatives, the confidence intervals are much wider, as seen with ovarian cancer. Despite this, it is useful for decision-making purposes to see the familial risks in absolute terms, as the doubling of relative risk seen with an affected first-degree relative has little impact on the absolute risk if incidence or mortality is low. For example, a relative risk of 2.85 for having a first-degree relative affected with ovarian cancer only increases an individual's lifetime risk from 1 in 90 to 1 in 40.

Whilst we have comprehensively reviewed the evidence for family history as a risk factor for 7 common, complex diseases, there is further work that could be done in future to build on this project. We have shown that for the more common cancers, there are sufficient risk and incidence data available for implementing this methodology. As well as investigating other cancers e.g. bladder cancer, it would also be useful to study some of the more common, complex diseases present in the UK population such as coronary heart disease or diabetes, incidence data permitting. A key element to systematic reviews is to update them as more evidence is available. Not only would we advocate re-using this methodology in future on the diseases in this report, but through the use of novel techniques e.g. increased availability of full-text searching or searching further databases, it may be possible to find additional data to increase the evidence base. Use of these extra data to re-estimate familial risks will help to refine these risk estimates and produce a more robust evidence base on which to make decisions.

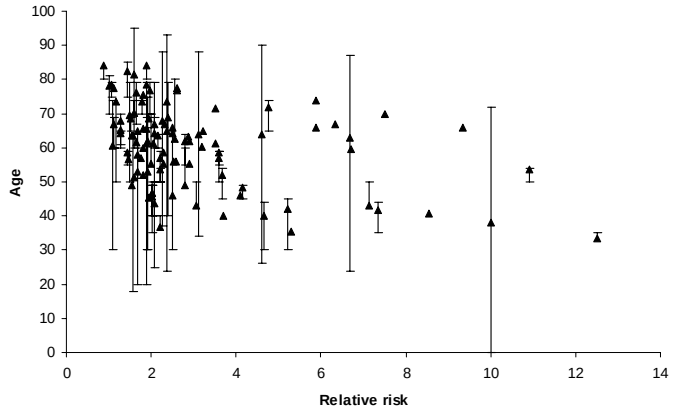
Additionally, it is known that some diseases such as ovarian and breast cancer tend to cluster in the same families. In this project we have only investigated the risk of disease when relatives are affected with the same disease. It would be interesting and clinically useful to investigate the relationship between familial risks of different diseases where there is available evidence.

To put these risk estimates into a decision-making context, it would be useful to compare the risks associated with a family history with other well-known risk factors e.g. smoking and age. In order to build accurate risk score models for complex diseases such as the Gail Risk model for breast cancer or the Framingham Risk Score model for heart disease, it is necessary to incorporate the best available evidence on each of the individual risk factors and to compare them to assess which are the best predictors of disease. It would be interesting to use the familial risk estimates produced here to consider the utility of family history as a predictor of disease in family members compared to other established risk factors.

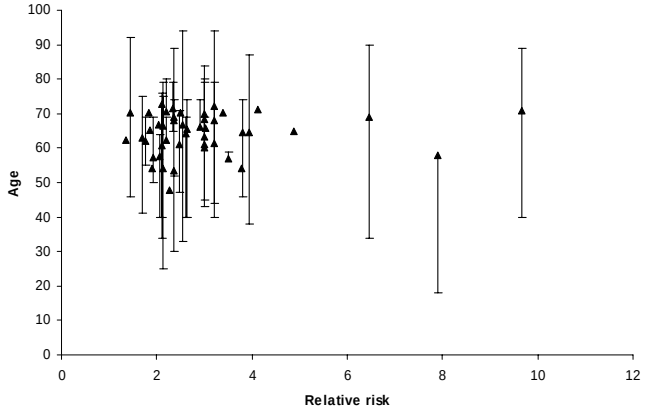
11. Conclusions

- For common cancers, there are sufficient data available to estimate accurate pooled relative risks.
- In general, the relative risk for those with at least one affected first-degree relative approximates two, which increases with more affected relatives.
- The relative risk is lower for those with more distant affected relatives.
- There are fewer age-specific data available so estimates are less accurate.
- For some cancers, higher relative risks are seen with lower ages of onset.
- Individuals who have relatives affected by stroke have lower relative risks than those with cancers.
- Multiple sclerosis has a much higher familial risk, however accurate risk estimates are harder to calculate due to lack of data and strong heterogeneity between studies.
- Where there are sufficient studies it is possible to use life-tables to convert population data and relative risk estimates into accurate absolute risk estimates for any time period up to age 85.
- For other diseases (stroke and multiple sclerosis), there are no national incidence registries and so incidence data are less reliable.

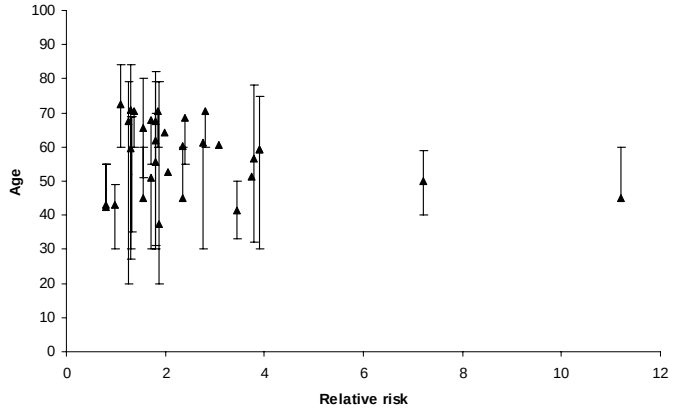
CRC



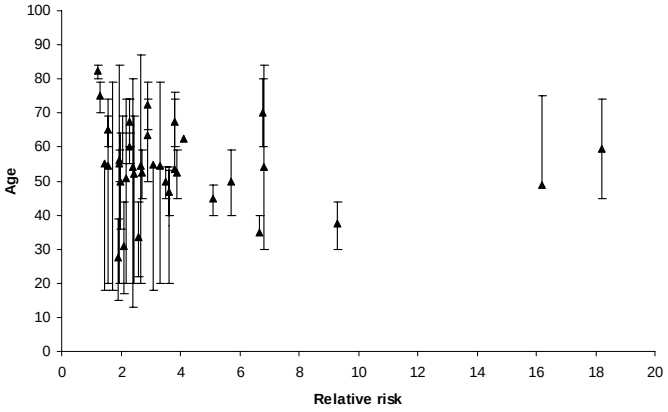
PC



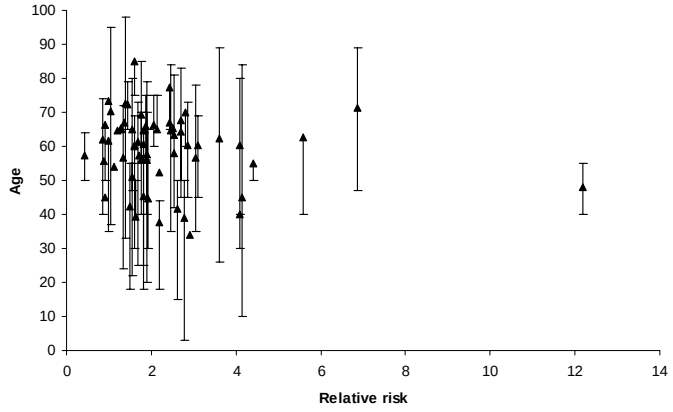
LC



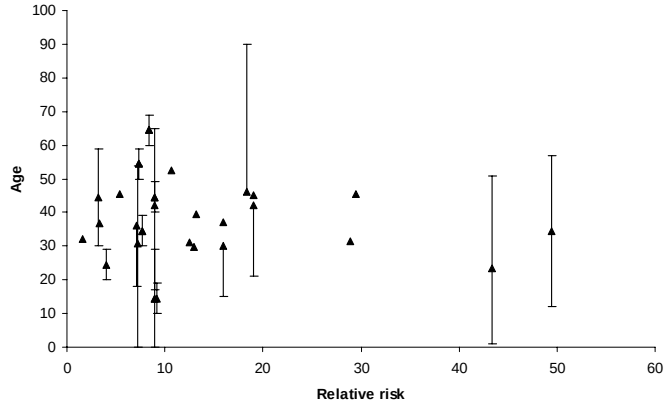
OC



Stroke



MS



Appendix 1. Graphs of study or strata age data versus relative risk estimate for 6 diseases. (Also available at www.PHGFoundation.org.uk)

Triangles mark the mean age and relative risk of a whole study or age-stratified stratum, error bars denote the range of ages of participants in the group.

Reference List

Ahlbom A, Lichtenstein P, Malmstrom H, *et al.* (1997). Cancer in twins: genetic and nongenetic familial risk factors. *J Natl Cancer Inst* **89**(4): 287-293.

Alberg AJ, Brock MV and Samet JM (2005). Epidemiology of lung cancer: looking to the future. *J Clin Oncol* **23**(14): 3175-85.

Begg CB and Berlin JA (1989). Publication bias and dissemination of clinical research. *J Natl Cancer Inst* **81**(2): 107-115.

Bostwick DG, Burke HB, Djakiew D, *et al.* (2004). Human prostate cancer risk factors. *Cancer* **101**(S10): 2371-2490.

Bruner DW, Moore D, Parlanti A, *et al.* (2003). Relative risk of prostate cancer for men with affected relatives: systematic review and meta-analysis. *Int J Cancer* **107**(5): 797-803.

Butterworth AS, Higgins JP and Pharoah P (2006). Relative and absolute risk of colorectal cancer for individuals with a family history: A meta-analysis. *Eur J Cancer* **42**(2): 216-227.

Cannon MM and Leavell BS (1966). Multiple cancer types in one family. *Cancer* **19**(4): 538-40.

Casas JP, Hingorani AD, Bautista LE, *et al.* (2004). Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. *Arch Neurol* **61**(11): 1652-61.

Chiang CL (1968). The life-table and its construction. New York, Wiley.

Collaborative Group on Hormonal Factors in Breast Cancer (2001). Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* **358**(9291): 1389-99.

Compston A (2000). The genetics of multiple sclerosis. *J Neurovirol* **6**(Suppl 2): S5-9.

de la Chapelle A (2004). Genetic predisposition to colorectal cancer. *Nat Rev Cancer* **4**(10): 769-80.

DerSimonian R and Laird N (1986). Meta-analysis in clinical trials. *Control Clin Trials* **7**(3): 177-188.

Dong C and Hemminki K (2001). Multiple primary cancers of the colon, breast and skin (melanoma) as models for polygenic cancers. *Int J Cancer* **92**(6): 883-7.

Dumitrescu RG and Cotarla I (2005). Understanding breast cancer risk - where do we stand in 2005? *J Cell Mol Med* **9**(1): 208-21.

Duval S and Tweedie R (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* **56**(2): 455-63.

Ebers GC and Sadovnick AD (1993). The geographic distribution of multiple sclerosis: a review. *Neuroepidemiology* **12**(1): 1-5.

Flossmann E, Schulz UG and Rothwell PM (2004). Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke* **35**(1): 212-27.

Ford HL, Gerry E, Johnson M*, et al.* (2002). A prospective study of the incidence, prevalence and mortality of multiple sclerosis in Leeds. *J Neurol* **249**(3): 260-5.

GAMES and the Transatlantic Multiple Sclerosis Genetics Cooperative (2003). A meta-analysis of whole genome linkage screens in multiple sclerosis. *J Neuroimmunol* **143**(1-2): 39-46.

Gardner EJ and Stephens FE (1950). Breast cancer in one family group. *Am J Hum Genet* **2**(1): 30-40.

Hemminki K, Li X, Johansson SE*, et al.* (2006). Re: "Familial risk of multiple sclerosis: a nationwide cohort study". *Am J Epidemiol* **163**(9): 873-4.

Higgins JP and Thompson SG (2002). Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**(11): 1539-1558.

Johns LE and Houlston RS (2003). A systematic review and meta-analysis of familial prostate cancer risk. *BJU Int* **91**(9): 789-794.

Kantarci O and Wingerchuk D (2006). Epidemiology and natural history of multiple sclerosis: new insights. *Curr Opin Neurol* **19**(3): 248-54.

Khaw KT and Barrett-Connor E (1986). Family history of stroke as an independent predictor of ischemic heart disease in men and stroke in women. *Am J Epidemiol* **123**(1): 59-66.

Lichtenstein P, Holm NV, Verkasalo PK*, et al.* (2000). Environmental and heritable factors in the causation of cancer - analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* **343**(2): 78-85.

Lunn DJ, Thomas A, Best N*, et al.* (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* **10**(4): 325-337.

Lynch HT and de la Chapelle A (2003). Hereditary colorectal cancer. *N Engl J Med* **348**(10): 919-32.

Macklin MT (1932). The Hereditary Factor in Human Neoplasms. *The Quarterly Review of Biology* **7**(3): 255-81.

Marugame T and Mizuno S (2005). Comparison of prostate cancer mortality in five countries: France, Italy, Japan, UK and USA from the WHO mortality database (1960-2000). *Jpn J Clin Oncol* **35**(11): 690-1.

Matakidou A, Eisen T and Houlston RS (2005). Systematic review of the relationship between family history and lung cancer risk. *Br J Cancer* **93**(7): 825-33.

Miki Y, Swensen J, Shattuck-Eidens D, *et al.* (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**(5182): 66-71.

Nielsen NM, Westergaard T, Rostgaard K, *et al.* (2005). Familial risk of multiple sclerosis: a nationwide cohort study. *Am J Epidemiol* **162**(8): 774-8.

ONS (2004). Mortality statistics: cause. *Series DH2 no. 30*.

ONS (2006). Focus on Health - Cancer. *http://www.statistics.gov.uk/cci/nugget.asp?id=1332*.

ONS (2006). ONS Cancer Statistics: Registrations Series MB1.

Pearson K (1912). IV. On "Cancer Houses", from the Data of the late Th. Law Webb, M.D. *Biometrika* **8**: 430-435.

Pugliatti M, Sotgiu S and Rosati G (2002). The worldwide prevalence of multiple sclerosis. *Clin Neurol Neurosurg* **104**(3): 182-91.

Robertson N, Deans J, Fraser M, *et al.* (1996). Multiple sclerosis in south Cambridgeshire: incidence and prevalence based on a district register. *J Epidemiol Community Health* **50**(3): 274-9.

Rothwell PM and Charlton D (1998). High incidence and prevalence of multiple sclerosis in south east Scotland: evidence of a genetic predisposition. *J Neurol Neurosurg Psychiatry* **64**(6): 730-5.

Rothwell PM, Coull AJ, Giles MF, *et al.* (2004). Change in stroke incidence, mortality, case-fatality, severity, and risk factors in Oxfordshire, UK from 1981 to 2004 (Oxford Vascular Study). *Lancet* **363**(9425): 1925-33.

Rothwell PM, Coull AJ, Silver LE, *et al.* (2005). Population-based study of event-rate, incidence, case fatality, and mortality for all acute vascular events in all arterial territories (Oxford Vascular Study). *Lancet* **366**(9499): 1773-83.

Sandler RS (1996). Epidemiology and risk factors for colorectal cancer. *Gastroenterol Clin North Am* **25**(4): 717-35.

Schaid DJ (2004). The complex genetic epidemiology of prostate cancer. *Hum Mol Genet* **13**(Spec No 1): R103-R121.

Scheuner MT, Wang SJ, Raffel LJ, *et al.* (1997). Family history: a comprehensive genetic risk assessment method for the chronic conditions of adulthood. *Am J Med Genet* **71**(3): 315-24.

Sterne JA, Egger M and Smith GD (2001). Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ* **323**(7304): 101-105.

Stratton JF, Pharoah P, Smith SK, *et al.* (1998). A systematic review and meta-analysis of family history and risk of ovarian cancer. *Br J Obstet Gynaecol* **105**(5): 493-9.

Warren S, Cockerill R and Warren KG (1991). Risk factors by onset age in multiple sclerosis. *Neuroepidemiology* **10**(1): 9-17.

Warthin AS (1925). The further study of a cancer family. *J Cancer Res* **9**: 279-286.

Welcsh PL and King MC (2001). BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum Mol Genet* **10**(7): 705-13.

Wolfe CD, Rudd AG, Howard R, *et al.* (2002). Incidence and case fatality rates of stroke subtypes in a multiethnic population: the South London Stroke Register. *J Neurol Neurosurg Psychiatry* **72**(2): 211-6.