Coefficients of relationship by isonymy among registrations for five common cancers in Scottish males

Susan M Holloway, J A Sofar

Abstract

Study objective—The aim was to assess the relative importance of genetic factors in carcinoma of the stomach, colon, rectum, prostate, and bladder in Scottish males.

Design—Cancer cases and controls were compared in terms of the coefficient of relationship by isonymy (RI).

Setting—Surname distributions for cancer cases were derived from the Scottish Cancer Register for the years 1959–85. Control distributions were derived from all births, marriages and deaths in Scotland for 1976.

Subjects—Analysis was carried out on a total of 60 933 cancer registrations and 101 836 births, marriages, and deaths over the 12 local government regions of Scotland.

Main results—Comparisons of RI within and between regions indicated that inherited susceptibility was of greatest importance in carcinoma of the prostate and colon, of intermediate importance in carcinoma of the rectum and stomach, and of minimal importance in carcinoma of the bladder. Familial aggregation of cancers was most pronounced in Highland, Tayside, and Borders Regions. For Highland, this appeared to be the result of region-specific familial influences, while Tayside and the Borders shared genetic factors contributing to cancer aetiology with neighbouring regions in south east Scotland.

Conclusions—Surname analysis is a simple but useful tool for studying population genetic structure and its relationship to disease incidence.

J Epidemiol Community Health 1992; 46: 368–372

Cancer has a spectrum of aetologies. At one extreme are tumours such as retinoblastoma, a substantial proportion of which are hereditary. At the other are largely non-hereditary tumours, particularly those associated with occupational exposure, such as mesothelioma with asbestos and angiosarcoma of the liver with vinyl chloride. Almost all instances of carcinogenesis, however, probably involve a degree of interaction between genotype and the environment, comparable exposure to the same carcinogenic agent causing malignancy in some individuals but not in others. Indeed, 8% of known human genes appear to influence susceptibility or resistance to cancer. Examples of genotype-environment interaction occur among monogenic disorders known primarily for their association with neoplasia, for example increased susceptibility to sunlight induced skin cancer in patients with the autosomal recessive disorder xeroderma pigmentosum and their heterozygous relatives. Genetic variation in the response to environmental carcinogens may also be related to metabolic polymorphisms, as shown by the increased risk of smoking induced lung cancer in extensive metabolisers of debrisoquine. At the epidemiological level, however, because common genes and common environmental exposure to carcinogens among relatives can both cause familial aggregation of malignant disease, demonstration of a hereditary component in the absence of a known disorder or genetic marker can be demanding, both in terms of study design and resources.

Surnames provide a simple approach, using data that are readily available, which may help to shed light on the genetic contribution to cancer at different sites. This approach allows estimation of the degree of genetic relationship between groups of patients with the same disease in geographically distant locations, thus reducing the confounding of common genes and common environment that plagues conventional family studies. A disadvantage is that many surnames have multiple origins, but the magnitude of any bias can be assessed by a parallel analysis using only rare surnames, each of which is more likely to have a unique (monophyletic) origin.

Methods

Registrations among males for carcinoma of the stomach (ICD 151), colon (ICD 153), rectum (including rectosigmoid junction and anus, ICD 154), prostate (ICD 185), and bladder (ICD 188) were obtained from the Scottish Cancer Register (1959–85). Data for each case included an identification (ID) number, surname, date of birth, date of registration, date of death and local government region in which registration occurred. To maintain confidentiality, a key matching surname codes with surnames was held separately and used only when required. Codes for some of the earliest registrations referred to alphabetical groupings of surnames and cases with these codes were excluded. Duplicate registrations, defined as those with the same surname and ID number, or those with the same surname, date of birth, date of registration, and/or date of death, were also excluded. Females were not studied because maiden name was not generally available. As controls, surnames of all males who were born, married, or who died in each of the 12 local government regions of Scotland in 1976 were supplied by the Office of the Registrar General for Scotland. The numbers of controls and cancer
cases used in the analysis are listed by region in table I.

Surname distribution in cancer cases and controls was compared using the coefficient of relationship by isonymy, Ri, which can be considered as half the probability that two people selected at random will have the same surname. The factor of $\frac{1}{2}$ is introduced so that the relationship corresponds with the expected proportion of shared autosomal genes on the assumption of monophyly of surnames. Values of Ri were calculated within and between groups of cancer cases and controls according to Lasker.$^{9,10}$

Within a group $R_i = \frac{S_i(S_i-1)/2n(n-1)}{n}$, where $S_i$ is the number of occurrences of surname $i$, and $n$ is the total number of individuals in the group, with summation over all surnames. Between groups $R_i = \frac{S_i1S_j2/2n1n2}{S_j1S_j2}$, where $S_j1$ and $S_j2$ are the numbers of occurrences of surname $j$ in groups 1 and 2, and $n1$ and $n2$ are the total numbers of individuals in groups 1 and 2, with summation over all surnames.

The significance of the difference between two $R_i$ values was assessed by the method of Fox and Lasker.$^{11}$ These authors showed that $R_i$ is effectively independent of sample size and, using empirical estimates of sampling variation for $R_i$, that $R1$ and $R2$, two independent values of $R_i$, differ significantly at the 5% level if:

$$R1 - R2 > \frac{\sqrt{(\text{var} R1 + \text{var} R2)}}{2}$$

or

$$\frac{R1 - R2}{\sqrt{(R1^2 + R2^2)}} > \frac{1}{2}$$

Values of $R_i$ were calculated for controls, within cancers, between cancers and for all cancers combined, both for Scotland as a whole and within each local government region. Values of $R_i$ were also calculated for controls, within cancers, and for all cancers combined between regions. Two sets of comparisons were made, first with all controls and second with control deaths only. Control deaths, which accounted for 32% of all controls, had an age distribution more similar to that of the cancer cases.

The Wilcoxon matched-pairs signed ranks test$^{12}$ was used to compare sets of the 12 within region or 66 between region $R_i$ values for each cancer with controls, and sets of the 12 within region $R_i$ values between cancers with controls. For each of these comparisons, the median difference ($R_i$ for cancer cases minus $R_i$ for controls) was calculated, added to the median control value, and the result expressed as a percentage of the median control value. To compare individual regions or pairs of regions with each other, $R_i$ for all cancer cases combined was expressed simply as a percentage of the corresponding control value. All calculations were repeated using only rare surnames, considered to be those with an occurrence of less than 1 per 1000 among births, marriages, and deaths for 1976 in Scotland as a whole.$^{13}$ Rare surnames accounted for 59% of controls and 56% of cancer cases. In expressing statistical significance, account was taken of the large number of comparisons through multiplying individual probabilities by the number of comparisons made at each stage.

Results

Calculations for Scotland as a whole, both within cancers and between cancers, did not reveal any significant increase of $R_i$ in cancer cases relative to either group of controls, either for all surnames or for rare surnames only. For individual regions, only Highland showed significant differences. In this region, using rare surnames only, $R_i$ values for carcinoma of the prostate, and for carcinoma of the prostate with carcinoma of the colon, were both significantly greater than $R_i$ for all controls.

The difference between $R_i$ for carcinoma of the colon and that for all controls was bordering on statistical significance.

Within region and between region $R_i$ values for the different cancers, expressed as percentages of their median control values, are given in table II. Within regions, using all controls and all surnames, the median difference between cancer cases and controls was significantly greater than zero for carcinoma of the stomach, colon, and prostate. Using rare surnames only, $R_i$ values for carcinoma of the prostate with carcinoma of the colon were both significantly greater than $R_i$ for all controls. Between regions, using all surnames, the median differences for all five cancers were significantly greater than zero using either all controls or control deaths only. Significance persisted for carcinoma of the colon, rectum, and prostate relative to all controls when the analysis was restricted to rare surnames, but there were no significant differences relative to control deaths when rare surnames were used.

<table>
<thead>
<tr>
<th>Region</th>
<th>Controls</th>
<th>Stomach</th>
<th>Colon</th>
<th>Rectum</th>
<th>Prostate</th>
<th>Bladder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highland</td>
<td>3826</td>
<td>395</td>
<td>598</td>
<td>323</td>
<td>586</td>
<td>293</td>
</tr>
<tr>
<td>Grampian</td>
<td>8927</td>
<td>1116</td>
<td>1329</td>
<td>975</td>
<td>1622</td>
<td>1076</td>
</tr>
<tr>
<td>Tayside</td>
<td>7524</td>
<td>1138</td>
<td>1175</td>
<td>737</td>
<td>1246</td>
<td>900</td>
</tr>
<tr>
<td>Fife</td>
<td>6671</td>
<td>914</td>
<td>754</td>
<td>547</td>
<td>1083</td>
<td>858</td>
</tr>
<tr>
<td>Lothian</td>
<td>14327</td>
<td>2071</td>
<td>1855</td>
<td>1369</td>
<td>2578</td>
<td>1823</td>
</tr>
<tr>
<td>Borders</td>
<td>1842</td>
<td>298</td>
<td>300</td>
<td>239</td>
<td>405</td>
<td>164</td>
</tr>
<tr>
<td>Central</td>
<td>5090</td>
<td>586</td>
<td>527</td>
<td>392</td>
<td>764</td>
<td>642</td>
</tr>
<tr>
<td>Strathclyde</td>
<td>4068</td>
<td>6117</td>
<td>5701</td>
<td>3788</td>
<td>5663</td>
<td>5754</td>
</tr>
<tr>
<td>Dumfries*</td>
<td>2623</td>
<td>329</td>
<td>361</td>
<td>242</td>
<td>509</td>
<td>280</td>
</tr>
<tr>
<td>Orkney</td>
<td>399</td>
<td>35</td>
<td>29</td>
<td>31</td>
<td>76</td>
<td>1132</td>
</tr>
<tr>
<td>Shetland</td>
<td>430</td>
<td>22</td>
<td>50</td>
<td>35</td>
<td>63</td>
<td>28</td>
</tr>
<tr>
<td>Western Isles</td>
<td>629</td>
<td>86</td>
<td>76</td>
<td>42</td>
<td>123</td>
<td>54</td>
</tr>
<tr>
<td>Scotland</td>
<td>101836</td>
<td>13067</td>
<td>12755</td>
<td>8570</td>
<td>14712</td>
<td>11829</td>
</tr>
</tbody>
</table>

$^*$Dumfries and Galloway
<table>
<thead>
<tr>
<th>Region</th>
<th>Control births, marriages, and deaths</th>
<th>Control deaths only</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>Rare</td>
<td>All</td>
</tr>
<tr>
<td>Stomach—colon</td>
<td>121 120</td>
<td>108 97</td>
</tr>
<tr>
<td>Stomach—rectum</td>
<td>115 108</td>
<td>106 96</td>
</tr>
<tr>
<td>Stomach—prostate</td>
<td>118* 110</td>
<td>110 97</td>
</tr>
<tr>
<td>Stomach—bladder</td>
<td>116 107</td>
<td>95 92</td>
</tr>
<tr>
<td>Colon—rectum</td>
<td>124 124</td>
<td>110* 105</td>
</tr>
<tr>
<td>Colon—bladder</td>
<td>132T 131</td>
<td>114 105</td>
</tr>
<tr>
<td>Rectum—prostate</td>
<td>124 105</td>
<td>107 100</td>
</tr>
<tr>
<td>Rectum—bladder</td>
<td>119 119</td>
<td>107 96</td>
</tr>
<tr>
<td>Prostate—bladder</td>
<td>132T 134T</td>
<td>120* 111</td>
</tr>
<tr>
<td></td>
<td>111 127</td>
<td>112 100</td>
</tr>
</tbody>
</table>

Where the median difference was significantly greater than zero: \( p < 0.05 \), \( p < 0.01 \).

Between-cancer Ri values within regions, expressed as percentages of the median control values, are given in Table III. Using all controls and all surnames, the median differences between Ri values for pairs of cancers and for controls were significantly greater than zero for all pairs of cancers except carcinoma of the stomach and colon or rectum, and all pairs that included carcinoma of the bladder. One of the four significant differences persisted when rare surnames were used. Relative to control deaths only, two differences were significant for all surnames but none was significant for rare surnames.

Within region Ri values for all cancers combined, expressed as percentages of their control values, are given in Table IV. The results for rare surnames using control deaths only, as well as all those for Orkney, Shetland, and the Western Isles, were based on small sample sizes and so were individually ignored. In none of the individual regions was there a significant difference between Ri for all cancer cases and Ri for either set of controls using all surnames or rare surnames only. However, the median difference over all 12 regions was significantly greater than zero for all cancer cases relative to all controls or control deaths only, and for cancer cases relative to all controls with rare surnames only. Highland, Tayside, and Borders showed comparatively high Ri values relative to controls for all surnames and rare surnames only, indicating that both types of surname may be associated with cancer in these regions. For Grampian, Ri was low using all surnames but high for rare surnames only, suggesting that particular rare surnames may be associated with cancer in this region.

The 10 highest Ri values between regions for all cancers combined, expressed as percentages of their control values, are shown for the four categories of analysis in Table V. All regions with an individually high Ri value in table IV are represented in these pairs, although between region values involving Highland were relatively low. There was no significant difference from controls for any one pair of regions in any category. However, the median difference between cancer cases and controls over the 66 pairs of regions was significantly greater than zero for all four categories of analysis.

**Discussion**

Analysis was carried out for all surnames and for rare surnames, using either all controls or control deaths only. All controls provided a greater number of surnames for comparison, but mean age, taking fathers’ age in the case of births, was only 40 years. Mean age for control deaths was 67 years, compared with 67 to 74 years between year of birth and 1976 for cancer patients (some of whom died before 1976). Age is of relevance because within region Ri among controls was greater for deaths than for births and marriages, as previously reported and as can be inferred from table II. Comparisons using all controls might therefore have tended to overestimate the increase in Ri shown by cancer cases. On the other hand, there was also a tendency to underestimate any difference of Ri between cancer cases and controls because about 4% of the cancer patients died in 1976 and were consequently included in the Registrar General’s figures. However, for each cancer, patients who died in 1976 contributed only about 0.5% of the total number of controls and about 1.5% of control deaths. The effect of any net bias is likely to have been small but in any event should not have affected the validity of the analysis since this was aimed primarily at disclosing the relative importance of genetic factors in cancer at the different sites.

The results from rare surnames were generally similar to those based on all surnames, although rather less pronounced, despite the fact that rare surnames are expected to reflect genetic affinity more closely than all surnames, where multiple origins dilute the degree of relationship. The most likely explanation for this is that in many regions the number of cancer cases with rare surnames was small.

The lack of a significant difference between Ri values for cancer cases and controls in Scotland as a whole was probably the result of important genetic factors being associated with a much larger number of surnames in the total population than within regions, making it more difficult to assess family relationship between affected individuals.
Surnames among cancer registrations

Excluding comparisons involving both rare surnames and control deaths, where sample sizes were small, the pattern of median differences between the within region or between region Ri values for cancer cases and controls suggested that the five cancers fall into three different classes. For carcinoma of the colon and prostate, within region and between region differences were, respectively, 14–42% and 10–30% of their median control values. All differences were significantly greater than zero except those for within region Ri for carcinoma of the colon using control deaths or rare surnames. For carcinoma of the stomach and rectum, within region and between region differences were respectively 9–28% and 0–20%, of their median control values. Only one of the six within region differences but five of the six between region differences were significantly greater than zero. For carcinoma of the bladder, the within region and between region differences were, respectively, 0–18% and 5–25% of their median control values. None of the within region differences and two of the three between region differences were significantly greater than zero (table II). Furthermore, Ri between cancers within regions disclosed some familial association between all cancers except carcinoma of the bladder (table III). Isonymy among cases of bladder cancer was therefore not only least pronounced but showed relatively little overlap with that among patients suffering from cancer at the other sites.

Within regions, a significantly higher Ri for cancer cases can be taken as evidence of familial factors in the aetiology of malignant disease. Between regions, it suggests that these familial factors are likely to be genetic rather than environmental, the larger number of significant comparisons reflecting the difference in size between the sets of within region and between region Ri values. The results therefore suggest that genetic factors have a role in the aetiology of carcinoma of the colon and prostate and, probably to a lesser extent, in carcinoma of the stomach and rectum. Inherited susceptibility appears to be least important for carcinoma of the bladder.

Clee$^4$ has also used surname analysis to investigate the contribution of genetic factors to cancer aetiology. For all five cancers studied here, he found some surnames which occurred more frequently than expected, the greatest number of such surnames occurring in carcinoma of the colon and prostate. Certain surnames were also associated with cancer at more than one site, in keeping with the significantly increased Ri between some cancers within regions found in the present study (table III).

Using a related but more rigorous approach, Hill$^5$ was able to link 16 820 cases on the Utah Cancer Registry to the computerised Mormon genealogy records, comparing Malécot coefficients of kinship among cancer cases with those among matched controls chosen at random from the genealogy records. Kinship was greater among individuals with cancer at each of several sites, the ranking by site in order of decreasing kinship being colon, breast, gastrointestinal, and urogenital.

Several heritable syndromes are associated with a high risk of bowel cancer but these account for only a small proportion of cases. Relatives of patients with so-called sporadic disease have a two- to threefold increased risk compared with the general population$^{16}$ but it is likely that diet is of particular importance. High fat ingestion is currently thought to be one of the most relevant factors, raising the possibility of inherited differences of susceptibility through genetic variation in fat metabolism.$^3$ Familial aggregation of carcinoma of the prostate, thought to have at least a partially genetic basis, has been known for several years.$^{17, 18}$ However, since persons migrating from low risk countries to the high risk USA have shown intermediate incidence, environmental factors must also be involved.$^{19}$

Familial aggregation of gastric carcinoma is rare and this tumour has fallen in incidence by 60% from 1950 to the 1980s in the USA.$^{20}$ This fall, together with the observation that populations migrating from high to low risk countries show intermediate incidence, demonstrates the importance of environmental influences. On the other hand, the weak but significant association between carcinoma of the stomach and blood type A, first reported more than 35 years ago,$^{21, 22}$ indicates that inherited susceptibility contributes to the disease. Autosomal dominant transmission of carcinoma of the bladder has been reported in five families but this is rare and the risk to a relative of the usual 60–70 year old patient can be considered negligible.$^{23}$ Environmental factors such as cigarette smoking and employment in the dye, rubber, leather, paint, and organic chemical industries are known to increase the risk of developing the disease.$^{24}$

The results of the present study are therefore in agreement with previous findings using other methods. In addition, within region Ri for all cancers combined indicated that familial factors were most important in Highland, Tayside, and Borders regions. In particular, in Highland Region, there was evidence for the importance of familial factors in the aetiology of carcinoma of the prostate and colon. There was also evidence that particular rare surnames may be associated with cancer in Grampian Region (table IV). Pairs of regions with high region Ri (table V) were those with individually high within region values (table IV), suggesting that pairs of regions with high values are those likely to share genetic factors important in cancer aetiology. However, the high within region but less pronounced between region Ri values may reflect particular familial confounds to this region.

As in the study of Hill,$^5$ it is not possible to use the present results to draw inferences of relative risk in the families of cancer patients. It may, however, be feasible to investigate this matter further, either using empirical data from common disorders with known modes of genetic transmission or through computer simulation. Future investigations into the possible contribution of social class differences in surname distribution to isonymy among cancer cases would also seem to be indicated.

The authors are grateful to the Information and Statistics Division of the Scottish Health Service Common Services Agency and to the General Register Office for Scotland for making the data available.