JNCI Journal of the National Cancer Institute Advance Access published September 9, 2008

ARTICLE

Evaluation of Bias in Familial Risk Estimates: A Study of Common Cancers Using Swedish Population-based Registers

Monica Leu, Marie Reilly, Kamila Czene

Background Bias in estimates of familial cancer may result if population-based registers fail to identify relatives as affected when disease occurs before the start-up of registration (ie, "left-truncation" of family history).

Methods Apparent familial relative risks (among offspring of parents with cancer) of colorectal, lung, breast, and prostate cancers and melanoma in a Swedish cohort were compared with relative risks in a simulated population. The study cohort (approximately 7 million individuals) was based on the Swedish Multi-Generational Register linked to the Swedish Cancer Register for the period 1961–2002. A similar population of related individuals (approximately 7 million) with complete family information was simulated by using the R-package PopLab and used to estimate the sensitivity of the observed family history. This sensitivity was then used to calculate corrected age group–specific and overall risks, which were compared with the apparent familial risks of cancer in the cohort.

Result The apparent familial risks for colorectal, lung, breast, and prostate cancers and melanoma were 1.99 (95% confidence interval [CI] = 1.85 to 2.14), 2.05 (95% CI = 1.86 to 2.26), 1.84 (95% CI = 1.76 to 1.92), 2.33 (95% CI = 2.19 to 2.48), and 2.68 (95% CI = 2.35 to 3.07), with corresponding absolute rates of 3.69, 2.59, 16.05, 10.38, and 2.96 per 10 000 person-years, among offspring of parents diagnosed with the same cancer. Corrected age group–specific and overall estimates of the familial risks were close to these apparent risks for all studied cancers (all approximately 2.0), except for melanoma. For melanoma, the corrected estimate of 3.18 (95% CI = 2.73 to 3.64) was somewhat larger than the apparent estimate and was not included in the confidence interval for the apparent estimate. When the exposure of interest was a parent affected at a younger age, this bias was more pronounced; the apparent estimate for melanoma changed from 4.07 (95% CI = 3.21 to 5.16) to 5.67 (95% CI = 4.51 to 6.83) after correction.

Conclusions For common cancers, risk estimates from the Swedish MultiGenerational cohort do not generally appear to be biased by left-truncation.

J Natl Cancer Inst 2008;100:1318–1325

Estimates of familial aggregation provide important information in the study of disease etiology. The administrative population-based registers that are available in many countries offer tremendous potential for such studies. For example, in Sweden, the nationwide MultiGenerational Register (1) and Cancer Register (2) can be linked to create a cohort to study cancer. Similar resources for cancer and other diseases exist in other Scandinavian countries (3–7). Even though these registers have extensive coverage of the population members, the study cohort has incomplete exposure information due to failure to identify relatives as affected when disease occurs before the start-up of registration (ie, "left-truncation" of family history), failure to track all relatives, and various inclusion and exclusion criteria for registering individuals. With the exception of published estimates from some studies that used a specialized design that is robust to truncation bias (8,9), many of the published estimates of familial aggregation may be biased by an underestimate of the number of affected relatives (10,11).

Unbiased estimates of familial risk can be obtained by using one of several existing bias-correction methods, which require the implementation of sophisticated nonstandard statistical algorithms (10) and/or the estimation of the sensitivity and specificity of the observed exposure (12–15). These quantities can be estimated using a statistical model that defines misclassification (16) or by using validation samples, which are generally unavailable or costly (17–19).

Here, we examined the bias in familial risk estimates when exposure was defined as the parental history of disease in a study

Affiliation of authors: Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

Correspondence to: Kamila Czene, PhD, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Box 281, SE-171 77, Stockholm, Sweden (e-mail: kamila.czene@ki.se).

See "Funding" and "Note" following "References."

DOI: 10.1093/jnci/djn290

cohort obtained by merging the Swedish MultiGenerational Register (1) and the Swedish Cancer Register (2). We have previously shown (11) that the main source of the bias in familial aggregation estimates from such cohort is the left-truncation of parental disease in the register that is linked to the family register, ie, the misclassification of exposure from failing to identify relatives as affected when disease occurs before the start-up date of registration. In contrast, the missing familial links due to death in the Swedish MultiGeneration Register had little or no effect on familial risk estimates. Therefore, we corrected here for bias due to left-truncation, applying our previously developed methodology (20,21; M. Leu, K. Czene, M. Reilly, unpublished data, 2008) that produces estimates of the sensitivity of the observed exposure without using validation samples. Age group–specific and overall estimates of risk were investigated for the following cancers: colorectal, lung, female breast, prostate, and melanoma.

Subjects and Methods

Participants

We extracted the study cohort from the Swedish MultiGenerational Register linked to the Swedish Cancer Register in 2002. The MultiGenerational Register records offspring (index persons) who were born in Sweden since January 1, 1932, and were still alive on December 31, 1961, and their biological parents. The total number of individuals (offspring plus parents) is more than 10 million, organized in 3.2 million nuclear families. A substantial fraction (~50%) of the individuals who died before December 31, 1990, have one or both parents unidentified (1). To reflect the usual approach in analyses that use these data sources, we included in our analysis only offspring for whom both parents were identified (approximately 7 million).

The Swedish Cancer Register records primary cancers diagnosed since January 1, 1958, based on compulsory reports from health care providers. The completeness of this register and the accuracy of cytologically or histologically verified cancers are estimated to be close to 100% (22). Cancer diagnoses were defined according to the International Classification of Diseases (ICD-7): colorectal (ICD-7: 153–154), lung (ICD-7: 162–163), breast (ICD-7: 170), and prostate (ICD-7: 177) cancers and melanoma (ICD-7: 190) (2).

For each of the considered cancers, an independent virtual population register of related individuals (approximately 7 million) with complete family information was created for Sweden for the time interval 1961–2002 by use of the R-package PopLab (20,21). We used the actual Swedish age- and calendar year–specific mortality and fertility rates, which are available online from Statistics Sweden (23), and the actual calendar-, age-, and sex-specific incidence rates, which are available online from the Swedish National Board of Health and Welfare (24). Only female breast cancers were considered.

Evaluation of Bias

For each of the studied cancers, several steps were performed to evaluate the bias. First, from the real data extracted from the MultiGeneration Register and Swedish Cancer Register, we obtained the observed prevalence of exposure (ie, an affected parent) and the relative risk, which will be referred to as the apparent relative risk, in each of the strata defined by age at diagnosis and calendar time (Figure 1). Second, from the simulated data, we estimated the sensitivity of the observed exposure. Next, these three quantities were used to calculate the bias-corrected relative risks within each stratum, and these stratum-specific relative risks were then combined on the logarithmic scale to give an overall corrected estimate. As a last step, the overall bias-corrected relative risk was compared with the apparent relative risk to evaluate the bias.

Analyses of the Real Data. Follow-up for each individual started at birth or on January 1, 1961, whichever occurred later, and was terminated on diagnosis of first cancer, death, emigration, or the closing date of the study (December 31, 2002), whichever occurred first. Parental cancer entered the analysis as a time-dependent exposure: individuals were moved from the unexposed to the exposed group at the moment their parent developed cancer. Thus, cancer patients were considered to be exposed only if the parental cancer occurred before their cancer. Many individuals in the study cohort appear in different generations in the MultiGenerational Register, first as offspring and later as parents, and these were considered independently. Multiple affected offspring in the same family were also treated as independent events. For the three cancers that were not sex-specific (colorectal cancer, lung cancer, and melanoma), we assumed a familial risk model in which males and females experience an equal increase in disease risk from having either of their parents affected. We calculated the incidence rate ratios for a positive family history from Poisson regression models using the *glm* function R

CONTEXT AND CAVEATS

Prior knowledge

Bias in estimates of familial cancer risks may occur in population-based studies when family history is incomplete if relatives of study subjects are diagnosed with cancer before the beginning of cancer registration.

Study design

Corrected familial risks (due to a parent having cancer) of colorectal, lung, female breast, and prostate cancers and melanoma were calculated for 1961–2002 after comparing apparent risks among a Swedish cohort from the Swedish MultiGeneration Register linked to the national Cancer Register and risks from a simulated population, with complete family history, for this period.

Study contribution

Corrected familial risks were similar to the apparent risks for all cancers and were close to 2.0, except melanoma, for which the corrected value was approximately 3.2.

Implications

For most of the common cancers studied, the Swedish MultiGeneration Register provided unbiased familial risks.

Study limitations

The bias in familial risks of other types of cancer should be evaluated in this population and in other populations with national cancer registries before these results can be generalized to other types of cancer and to other populations.

From the Editors

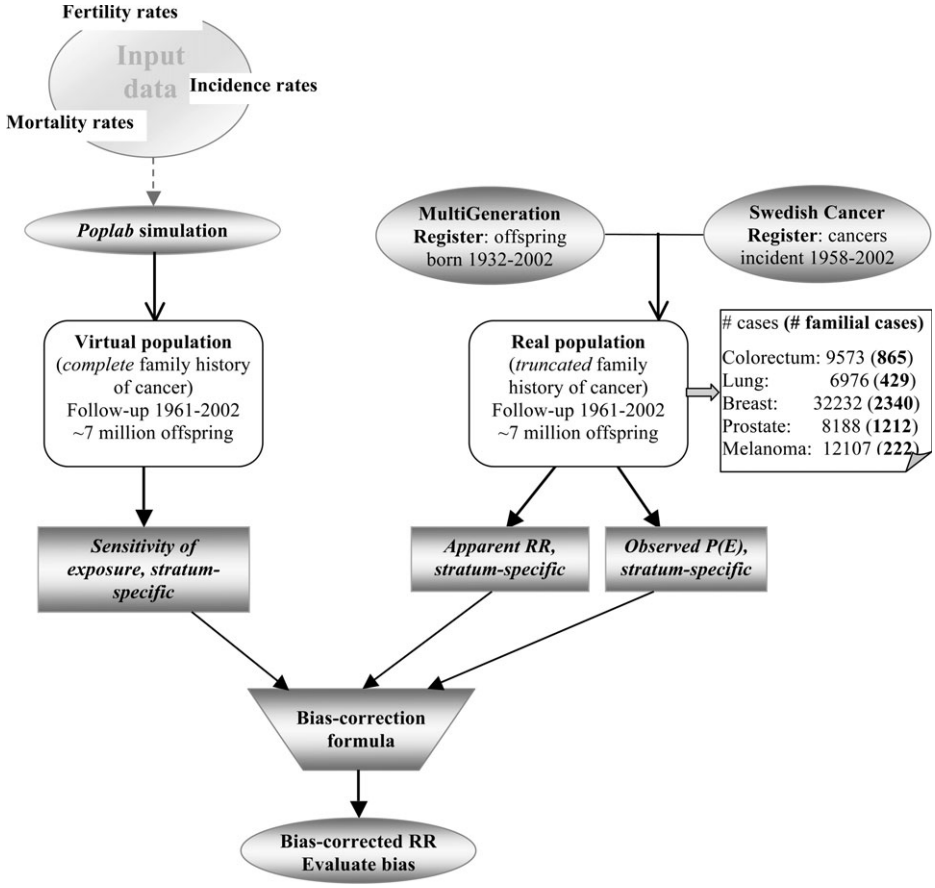

Figure 1. Schematic representation of the steps involved in the correction for and evaluation of the bias in the apparent relative risk for five common cancers. Bias corrections were performed within each strata defined by age group and calendar period, and the estimates were further combined to obtain overall corrected values. RR = relative risk; $P(E)$ = prevalence of exposure.

(http://www.r-project.org). The estimates were adjusted for age at diagnosis (in 5-year intervals) and calendar time (in decades).

Correcting the Apparent Relative Risk for Bias. Because the truncation of family history depends only on the start-up date of registration, we assumed that misclassification of exposure was nondifferential—that is, that there was a similar loss of family history for cancer patients and healthy individuals. To express the apparent relative risk as a function of directly estimable quantities from the available data and of the sensitivity of the observed exposure, we started from the general form of the relative risk

$$R = \frac{P(D|E)}{P(D|\overline{E})} = \frac{P(E|D)P(\overline{E})}{P(\overline{E}|D)P(E)}, \quad [1]$$

in which D denotes diseased, E exposed, and $\overline{E}$ unexposed. $P(E|D)$ and $P(E)$ represent the true prevalence of exposure among diseased (ie, the proportion of individuals in the diseased population with a positive family history) and the true overall prevalence, respectively, and $P(\overline{E}|D)$ and $P(\overline{E})$ are their complements.

Writing 1 for the apparent relative risk, $\hat{R}$, we obtain

$$\hat{R} = \frac{P(\hat{E}|D)P(\overline{\hat{E}})}{P(\overline{\hat{E}}|D)P(\hat{E})} = \frac{P(\overline{\hat{E}})}{P(\hat{E})}\frac{SP(E|D)}{1-SP(E|D)}, \quad [2]$$

in which $P(\hat{E}|D)$ and $P(\hat{E})$ are the observed prevalence of exposure among diseased individuals and the observed overall prevalence, respectively, and S is the sensitivity of the observed exposure, $P(\hat{E})/P(E)$, which is assumed to be equal for diseased and healthy individuals:

$$S = \frac{P(\hat{E})}{P(E)} = \frac{P(\hat{E}|D)}{P(E|D)} = \frac{P(\hat{E}|\overline{D})}{P(E|\overline{D})}, \quad [3]$$

where $\overline{D}$ denotes healthy individuals. By rearranging expression 2 and using the equalities in expression 3, the apparent relative risk can be written as:

$$\hat{R} = R(1 - SP(E))\frac{1}{\left(1 - S\right)RP(E) + P(\overline{E})}, \quad [4]$$

where R represents the true relative risk.

The derivation of the apparent relative risk from basic principles is presented elsewhere (13). Extracting R from equation 4 gives the expression for true relative risk, which will be denoted further as bias-corrected relative risk, R_{bc}:

$$R_{bc} = \hat{R}\frac{P(\overline{E})}{1+P(E)((\hat{R}-1)S-\hat{R})}. \quad [5]$$

Note that specificity of the observed exposure does not appear in this expression, as in the case of truncation this will be 100%.

The true prevalence of exposure and the sensitivity of exposure, as calculated below, were substituted in expression 5 to obtain the bias-corrected familial risk within each stratum defined by calendar period and age group. The available data were bootstrapped (100 times) to obtain independent realizations of these stratum-specific estimates and thus to obtain empirical estimates of their variances. Denoting the variance of the bias-corrected log RR (relative risk), $\beta_i = \log(R_{bc,i})$ in stratum i as $\mathrm{var}(\beta_i)$, the overall bias-corrected estimate for a specific age group was calculated as the weighted average of the appropriate calendar year–specific parameter estimates, taking as weights $(\mathrm{var}(\beta_i))^{-1}$.

Sensitivity of the Observed Exposure and True Prevalence of Exposure. We estimated the sensitivity of the observed exposure from the virtual population registers, in which each individual's family history of disease is complete. To construct a virtual population register, we initially created a simulated baseline population of related individuals for the first year of follow-up (1961), starting from 6 million unrelated founders (3 000 000 males and 3 000 000 females). For the creation of this baseline population, we used the first available incidence rates (1961). To evaluate the impact of increasing trend in the incidence rates before the baseline year, we also performed the simulation using half of these rates for breast cancer and melanoma. The 1961 population then evolved dynamically over time until 2002, with birth, death, and disease incidence events recorded on a yearly basis. Leu et al. (20) offer a detailed description of the simulation algorithm.

Having a family history of disease (ie, an affected mother for breast cancer, an affected father for prostate cancer, and an affected mother or father for colorectal cancer, lung cancer, or melanoma) increases the risk of cancer incidence for an individual from the year of incidence of the parent, and such individuals were categorized as exposed from that time. To reflect this relative risk model, we multiplied population age-specific rates of disease by a constant factor for exposed individuals. We simulated each virtual population with the apparent relative risk of that specific cancer. In addition, for individuals with cancer, we assigned an age-specific mortality that was five times that of the general population. This arbitrary value was selected for all studied cancers because we have previously shown that the mortality ratio has no impact on the bias due to left-truncation (11).

We imposed the effect of start-up of cancer registration on the complete virtual populations by recording in a separate variable whether family history was captured in the real data. Thus, any individual with a parental cancer occurring before the baseline year (1961) was classified as unexposed, and the new variable, the apparent exposure, corresponded to the family history recorded in the real Swedish Cancer Register data (2). Consequently, the ratio of apparent exposure and complete exposure as reflected in the simulated data provided an estimate of the sensitivity of observed exposure. The prevalence of exposure does depend on the familial risk (25), and thus we used the observed prevalence of exposure from the real data and the sensitivity from the simulation to obtain the true prevalence of exposure from expression 3.

Results

The number of patients observed in the offspring cohort extracted from the Swedish MultiGenerational Register linked to the Swedish Cancer Register (1961–2002) was lowest for lung cancer (6976) and highest for breast cancer (32 232) (Figure 1).

Sensitivity of the observed exposure did not depend substantially on the age group, but it did depend on cancer site, with prostate cancer exhibiting the highest sensitivity and melanoma the lowest (Figure 2, A). The calendar year–specific sensitivity

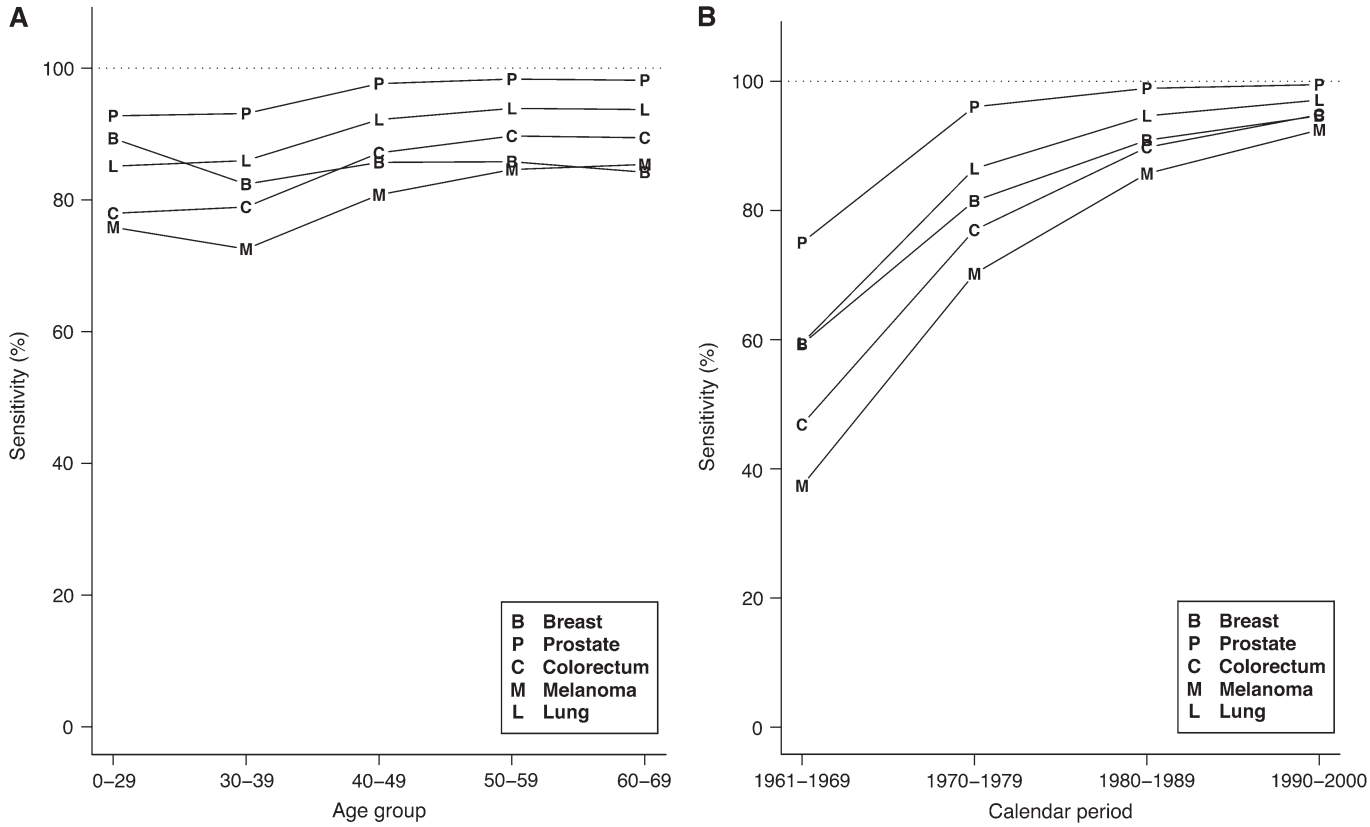

Figure 2. Sensitivity of the observed exposure for the five cancer sites studied. This quantity is calculated from the simulated populations as the ratio between the number of individuals recorded as exposed (ie, individuals whose parents were diagnosed with cancer after the start-up of registration) and the true number of exposed individuals. **A**) Sensitivity by age group for the entire follow-up period. **B**) Sensitivity by calendar period for all ages combined.

across all age groups increased over time, with a steep slope during the first two decades of follow-up (Figure 2, B). The calendar year–specific sensitivity also depended on cancer site, with melanoma having the lowest values. For breast cancer, colorectal cancer, and melanoma, the sensitivity exceeded 90% only after 30 years of follow-up.

We compared the apparent and bias-corrected relative risks for the five studied cancer sites (Table 1). The incidence rates among offspring of parents with colorectal, lung, breast, and prostate cancers and melanoma were 3.69, 2.59, 16.05, 10.38, and 2.96 per 10 000 person-years, respectively. Corrected age group–specific and overall estimates of the familial risks for colorectal, lung, breast, and prostate cancers were close to the apparent risks, with overall values of 1.99 (95% confidence interval [CI] = 1.85 to 2.14), 2.05 (95% CI = 1.86 to 2.26), 1.84 (95% CI = 1.76 to 1.92), and 2.33 (95% CI = 2.19 to 2.48). For melanoma, however, the bias-corrected point risk was 3.18 (95% CI = 2.73 to 3.64) compared with an apparent risk of 2.68 (95% CI = 2.35 to 3.07), and the 95% confidence interval for the apparent relative risk did not contain the bias-corrected value.

We also investigated how the parental age at diagnosis would affect the magnitude of bias. For breast cancer, when exposure was defined as a mother who was diagnosed with cancer before the age of 50, we observed some underestimation in the relative risk: apparent RR = 2.1 and bias-corrected RR = 2.45 (Table 2). In contrast, no bias was observed when exposure was a mother who was diagnosed with cancer after 50 years of age. Similarly, there was some underestimation in the apparent RR of 2.9 compared with the bias-corrected RR estimate of 3.3 when exposure was a parent who was diagnosed with colorectal cancer before the age of 60. The most pronounced bias was seen when exposure was a parent who was diagnosed with melanoma before the age of 50, with the apparent estimate of familial risk in offspring changing from 4.07 (95% CI = 3.21 to 5.16) to 5.67 (95% CI = 4.51 to 6.83) after correction.

Discussion

We observed an age-dependent familial risk for colorectal, breast, and prostate cancers and also for melanoma, with relatively high risks at younger ages. For breast and prostate cancers, these observations are in agreement with numerous previous studies (26–29). For colorectal cancer and melanoma, the literature has focused on the modification of familial risk by age of the index case patient, although some studies of colorectal cancer provide indirect evidence that younger relatives of case patients are at higher risk (30,31). We found no differences in familial risks by age at onset for lung cancer, indicating a relatively low importance of genetic factors (as compared with the effect of smoking habits) on this cancer (32,33).

We observed little or no bias in the overall estimates of familial risk of cancer, with the exception of melanoma, for which a dilution of the time relative risk was found. The lack of bias for most of these cancers is due to the relatively low familial risk (RR

Table 1. Apparent and bias-corrected relative risks for offspring with parental history of concordant cancer stratified by the age of the person at risk*

Cancer	Age group, y†	No. of familial case patients	Apparent relative risk	Bias-corrected relative risk‡
Colorectal	30–39	52	4.79 (3.61 to 6.36)	4.72 (3.18 to 6.26)
	40–49	191	2.84 (2.44 to 3.30)	2.77 (2.35 to 3.20)
	50–59	398	1.96 (1.76 to 2.17)	1.80 (1.62 to 1.97)
	60–69	211	1.44 (1.25 to 1.66)	1.42 (1.19 to 1.65)
	Overall	865	1.99 (1.85 to 2.14)	2.10 (1.93 to 2.26)
Lung	40–49	68	1.89 (1.47 to 2.43)	1.90 (1.39 to 2.42)
	50–59	225	2.07 (1.81 to 2.38)	2.11 (1.81 to 2.41)
	60–69	128	2.12 (1.77 to 2.54)	2.10 (1.63 to 2.56)
	Overall	429	2.05 (1.86 to 2.26)	2.12 (1.89 to 2.35)
Breast	30–39	230	2.52 (2.20 to 2.88)	2.49 (2.14 to 2.85)
	40–49	865	1.99 (1.85 to 2.13)	2.00 (1.85 to 2.15)
	50–59	925	1.64 (1.53 to 1.76)	1.64 (1.52 to 1.76)
	60–69	293	1.69 (1.50 to 1.90)	1.62 (1.40 to 1.85)
	Overall	2340	1.84 (1.76 to 1.92)	1.85 (1.76 to 1.93)
Prostate	40–49	38	4.79 (3.37 to 6.81)	4.47 (2.92 to 6.01)
	50–59	530	2.71 (2.47 to 2.98)	2.62 (2.34 to 2.90)
	60–69	644	2.06 (1.89 to 2.24)	2.12 (1.91 to 2.34)
	Overall	1212	2.33 (2.19 to 2.48)	2.41 (2.24 to 2.59)
Melanoma	0–29	37	6.78 (4.87 to 9.43)	6.14 (4.18 to 8.10)
	30–39	52	2.84 (2.16 to 3.74)	2.96 (2.04 to 3.88)
	40–49	60	2.18 (1.69 to 2.81)	2.11 (1.50 to 2.71)
	50–59	52	2.15 (1.62 to 2.85)	2.83 (2.05 to 3.62)
	60–69	21	3.61 (2.34 to 5.56)	3.83 (2.04 to 5.62)
	Overall	222	2.68 (2.35 to 3.07)	3.18 (2.73 to 3.64)

* Relative risks are expressed as incidence rate ratios, with 95% confidence intervals in parentheses. Familial case patients are cancer patients among offspring of parents with cancer. The familial relative risk is the risk in those with the exposure (ie, an affected parent) compared with that in those without the exposure.

† For each cancer site, age at risk ranged from 0 to 69 years, but, only those age groups with at least 20 exposed patients are shown.

‡ The apparent relative risk, observed prevalence of exposure, and the sensitivity of the observed exposure are used to calculate bias-corrected age group–specific and overall corrected relative risks of cancer.

Table 2. Apparent and bias-corrected relative risks for offspring with parental history of concordant cancer, stratified by the age group of the affected parent*

Cancer	Age group, y	No. of familial case patients	Apparent relative risk	Bias-corrected relative risk†
Colorectal	<60	237	2.90 (2.16 to 3.90)	3.33 (2.14 to 4.53)
	≥60	628	1.89 (1.68 to 2.12)	1.81 (1.57 to 2.08)
Lung	<60	93	2.31 (1.83 to 2.92)	2.40 (1.83 to 2.97)
	≥60	336	2.01 (1.62 to 2.45)	2.02 (1.58 to 2.52)
Breast	<50	373	2.10 (1.89 to 2.33)	2.45 (2.19 to 2.72)
	≥50	1967	1.78 (1.70 to 1.86)	1.76 (1.67 to 1.84)
Prostate	<60	48	3.26 (2.42 to 4.38)	3.33 (2.32 to 4.33)
	≥60	1164	2.32 (2.18 to 2.47)	2.36 (2.18 to 2.54)
Melanoma	<50	73	4.07 (3.21 to 5.16)	5.67 (4.51 to 6.83)
	≥50	149	2.33 (1.98 to 2.73)	2.64 (2.16 to 3.11)

* Relative risks are expressed as incidence rate ratios, with 95% confidence intervals in parentheses. Familial case patients are cancer patients among offspring of parents with cancer.

† The apparent relative risk, observed prevalence of exposure, and the sensitivity of the observed exposure are used to calculate bias-corrected age group–specific and overall corrected relative risks of cancer.

approximately 2) and/or relatively low incidence in the population, combined with a reasonably high sensitivity of the observed family history. Because sensitivity depends on age at onset, it is not surprising that the lowest sensitivity was observed for melanoma, a cancer with relatively young age at onset. This poor sensitivity, combined with the relatively large value of familial risk, would be expected to show the most biased relative risk (11), as we observed here. This bias was worst for exposure defined as young age at onset in a parent, in which case the apparent relative risk in offspring of parents who were diagnosed with cancer before the age of 50 substantially underestimated the true (bias-corrected) risk. Such dependence on age is to be expected because registration start-up will result in more severe left-truncation of (young) parental cancers, resulting in poor sensitivity. A similar effect was observed for breast cancer. For cancers due to rare, highly penetrant mutations, the low incidence should result in less serious bias (11). Although we cannot identify cancers due to mutations in specific genes, we investigated breast cancer with young age at onset in both mother and daughter (mother younger than 50 years and daughter younger than 40 years), for which specific genes (eg, *BRCA1* and *BRCA2*) have been identified, and found that the bias was not more pronounced than that reported for the overall breast cancer risk.

The strengths of our method are that the truncation bias in familial risks from standard cohort designs can be assessed and corrected without the need for validation samples or specialized study designs that are robust against truncation bias (8). With simple population vital statistics and disease incidence rates, familial relative risks can be corrected for the unavoidable bias due to registration start-up. By performing these corrections within the strata defined by age and calendar time, we could accommodate the age and secular trends in cancer incidence. The corrected relative risk estimates and their standard errors obtained from our method are in close agreement with the bias-corrected odds ratios obtained from the methods of Chu et al. (34), reassuring us of the validity of the results from the current study (data not shown).

A potential limitation of our study reflects the fact that in creating the simulated baseline population we used the first available incidence rates in the run-in simulation. Because most cancers have a rising incidence with calendar time, these early estimates will overestimate the true incidence rates experienced by the population before the baseline year. We investigated the impact of the baseline incidence rates on the bias-corrected estimates by simulating, for breast cancer and melanoma, baseline populations with age-specific incidence rates that were half the true 1961 rates. These two cancers were chosen because they exhibit very different incidence profiles: breast cancer rates increase moderately with calendar time, whereas melanoma incidence shows a steep increase. We found that for both of these cancers, the age group–specific and the overall bias-corrected estimates were similar to those obtained for the populations, in which the baseline was simulated with the 1961 rates (data not shown). Another potential limitation of our method was the assumption, for all cancers, of an age-specific mortality rate five times higher than the mortality rate in the general population. This is likely to overestimate the mortality in all but the oldest age groups or most fatal cancers (35). However, we have shown elsewhere (11) that the bias in the familial risk is essentially independent of the mortality rate ratio. Finally, we did not use age-specific familial risks in the simulation but instead used an average value of familial risk across all ages for every simulated cancer. We do not expect this assumption to influence our results, because we have found in previous work (M. Leu, K. Czene, M. Reilly, unpublished data, 2008) that the sensitivity does not depend on the value of familial risk used in simulations.

Although we have focused on bias due to truncation of disease events, this is only one of the potential sources of incompleteness in any real population. For family studies, missing parental information will result in subjects whose exposure information (ie, family history) is missing. These subjects will usually be dropped from the analysis, and only complete families will be analyzed (36,37). When the missing parental links depend on calendar time (eg, improvement in completeness of family registration over time), as is the case in the Swedish MultiGenerational Register (1), we have shown that substantial bias occurs only when there is very strong differential mortality between familial and nonfamilial cancer patients (11). We are aware of no such differential mortality for the five types of cancer in this study; even *BRCA1*- and *BRCA2*-related

breast cancer has been shown to have a similar prognosis to sporadic breast cancer (38–40).

Consistent with the study design used in a large number of publications assessing familial risk in which aging offspring become parents during follow-up, the individuals in our study were considered independently first as offspring and then as parents. Having these individuals appear in different generations as independent study subjects may result in confidence intervals of the risk estimates that are somewhat narrow, but this is likely to be a minor artifact because concordant cancers in three generations are extremely rare (41).

In this study, a positive family history was defined as having an affected parent. Because a sibling is also a first-degree relative, an affected sibling provides important information about genetic susceptibility. Although some studies have investigated the risk due to an affected sibling (42) or any affected first-degree relative (43), studies of parental relative risk are predominant in the literature. This is understandable because cancer is generally a disease of older people, so the parental generation provides more complete information about the disease profile in a family. However, siblings will be less subject to left-truncation due to their younger age relative to parents, so we would expect minimal bias in the estimates of sibling relative risks based on the findings presented here.

In conclusion, the work presented here confirms the large body of literature that has used the Swedish and other Scandinavian registers to estimate overall familial risks for common cancers. However, when the exposure of interest is early age of onset in a parent, which is commonly considered to be an indication of genetically determined cancer, estimates may be biased, especially when familial risk is high. The investigation of such bias in other diseases with high familial risks and/or high incidence rates is an area worthy of further study.

References

1. Multi-Generation Register 2003. A description of contents and quality (Swedish: Bakgrundsfakta till Befolknings- och Vdlfdrdsstatistik). *Statistics Sweden*. 2004;7. http://www.scb.se. Accessed January 30, 2008.
2. Information Regarding the Swedish Cancer Register. http://www.socialstyrelsen.se/en/Statistics/statsbysubject/Cancer+Registry.htm. Accessed January 30, 2008.
3. Hjortshøj TD, Grønskov K, Rosenberg T, Brøndum-Nielsen K, Olsen JH. Risk for cancer in patients with Bardet-Biedl syndrome and their relatives. *Am J Med Genet A*. 2007;143(15):1699–1702.
4. Dalton SO, Laursen TM, Mellemkjaer L, Johansen C, Mortensen PB. Risk for cancer in parents of patients with schizophrenia. *Am J Psychiatry*. 2004;161(5):903–908.
5. Nielsen NM, Westergaard T, Rostgaard K, et al. Familial risk of multiple sclerosis: a nationwide cohort study. *Am J Epidemiol*. 2005;162(8):774–778.
6. Goldin LR, Landgren O, McMaster ML, et al. Familial aggregation and heterogeneity of non-Hodgkin lymphoma in population-based samples. *Cancer Epidemiol Biomarkers Prev*. 2005;14(10):2402–2406.
7. Skjaerven R, Vatten L, Wilcox AJ, Rønning T, Irgens LM, Lie RT. Recurrence of pre-eclampsia across generations: exploring fetal and maternal genetic components in a population based cohort. *BMJ*. 2005;331(7521):877.
8. Pfeiffer RM, Goldin LR, Chatterjee N, et al. Methods for testing familial aggregation of diseases in population-based samples: application to Hodgkin lymphoma in Swedish registry data. *Ann Hum Genet*. 2004;68(pt 5):498–508.
9. Goldin LR, Pfeiffer RM, Gridley G, et al. Familial aggregation of Hodgkin lymphoma and related tumors. *Cancer*. 2004;100(9):1902–1908.
10. Andersen EW, Andersen PK. Adjustment for misclassification in studies of familial aggregation of disease using routine register data. *Stat Med*. 2002;21(23):3595–3607.
11. Leu M, Czene K, Reilly M. The impact of truncation and missing family links in population-based registers on familial risk estimates. *Am J Epidemiol*. 2007;166(12):1461–1467.
12. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of releative risk. *Am J Epidemiol*. 1977;105(5):488–495.
13. Flegal KM, Brownie C, Haas JD. The effects of exposure misclassification on estimates of relative risk. *Am J Epidemiol*. 1986;123(4):736–751.
14. Szatmari P, Jones MB. Effects of misclassification on estimates of relative risk in family history studies. *Genet Epidemiol*. 1999;16(4):368–381.
15. Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int J Epidemiol*. 2005;34(3):680–687.
16. Höfler M, Lieb R, Wittchen HU. Estimating causal effects from observational data with a model for multiple bias. *Int J Methods Psychiatr Res*. 2007;16(2):77–87.
17. Holcroft CA, Spiegelman D. Design of validation studies for estimating the odds ratio of exposure-disease relationships when exposure is misclassified. *Biometrics*. 1999;55(4):1193–1201.
18. Ziogas A, Anton-Culver H. Validation of family history data in cancer family registries. *Am J Prev Med*. 2003;24(2):190–198.
19. Rauscher GH, Sandler D. Validating cancer histories in deceased relatives. *Epidemiology*. 2005;16(2):262–265.
20. Leu M, Czene K, Reilly M. Population Lab: the creation of virtual populations in Genetic Epidemiology Research. *Epidemiology*. 2007;18(4):433–440.
21. Leu M, Czene K, Reilly M. Population Lab website. http://cran.at.r-project.org/. Accessed January 30, 2008.
22. Ekström AM, Signorello LB, Hansson LE, Bergström R, Lindgren A, Nyren O. Evaluating gastric cancer misclassification: a potential explanation for the rise in cardia cancer incidence. *J Natl Cancer Inst*. 1999;91(9):786–790.
23. Statistics Sweden. http://www.scb.se. Accessed January 30, 2008.
24. CANCERMondial Statistical Information System. http://www-dep.iarc.fr/. Accessed January 30, 2008.
25. Hemminki K, Sundquist J, Bermejo J. How common is familial cancer? *Ann Oncol*. 2008;19(1):163–167.
26. Hemminki K, Vaittinen P. Familial breast cancer in the family-cancer database. *Int J Cancer*. 1999;77(3):386–391.
27. Collaborative Group on Hormonal Factors in Breast Cancer. 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*. 2001;358(9291):1389–1399.
28. Tulinius H, Sigvaldasson H, Olafsdóttir G, Tryggvadóttir L, Bjarnadóttir K. Breast cancer incidence and familiality in Iceland during 75 years from 1921 to 1995. *J Med Genet*. 1999;36(2):103–107.
29. Rebora P, Czene K, Reilly M. Timing of familial breast cancer in sisters. *J Natl Cancer Inst*. 2008;100(10):721–727.
30. Planck M, Anderson H, Bladström A, Möller T, Wenngren E, Olsson H. Increased cancer risk in offspring of women with colorectal carcinoma: a Swedish register-based cohort study. *Cancer*. 2000;89(4):741–749.
31. Stefansson T, Moller PH, Sigurdsson F, Steingrimsson E, Eldon BJ. Familial risk of colon and rectal cancer in Iceland: evidence for different etiologic factors? *Int J Cancer*. 2006;119(2):304–308.
32. Jonsson S, Thorsteinsdottir U, Gudbjartsson DF, et al. Familial risk of lung carcinoma in the Icelandic population. *JAMA*. 2004;292(24):2977–2983.
33. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers—a different disease. *Nat Rev Cancer*. 2007;7(10):778–790.
34. Chu H, Wang Z, Cole SR, Greenland S. Sensitivity analysis of misclassification: a graphical and a Bayesian approach. *Ann Epidemiol*. 2006;16(11):834–841.
35. Damber L, Grönberg H, Damber JE. Familial prostate cancer and possible associated malignancies: nation-wide register cohort study in Sweden. *Int J Cancer*. 1998;78(3):293–297.

36. Hemminki K, Vaittinen P. National database of familial cancer in Sweden. *Genet Epidemiol.* 1998;15(3):225–236.
37. Chang ET, Smedby KE, Hjalgrim H, et al. Family history of hematopoietic malignancy and risk of lymphoma. *J Natl Cancer Inst.* 2005;97(19):1466–1474.
38. Verhoog LC, Brekelmans CT, Seynaeve C, et al. Survival and tumour characteristics of breast-cancer patients with germline mutations of BRCA1. *Lancet.* 1998;351(9099):316–321.
39. Verhoog LC, Brekelmans CT, Seynaeve C, et al. Survival in hereditary breast cancer associated with germline mutations of BRCA2. *J Clin Oncol.* 1999;17(11):3396–3402.
40. Robson M, Gilewski T, Haas B, et al. BRCA-associated breast cancer in young women. *J Clin Oncol.* 1998;16(5):1642–1649.
41. Lindstrom L, Pawitan Y, Reilly M, Hemminki K, Lichtenstein P, Czene K. Estimation of genetic and environmental factors for melanoma onset using population-based family data. *Stat Med.* 2006;25(18):3110–3123.
42. Hemminki K, Chen B. Familial risks for cervical tumors in full and half siblings: etiologic apportioning. *Cancer Epidemiol Biomarkers Prev.* 2006;15(7):1413–1414.
43. Czene K, Adami HO, Chang ET. Sex- and kindred-specific familial risk of non-Hodgkin's lymphoma. *Cancer Epidemiol Biomarkers Prev.* 2007;16(11):2496–2499.

Funding

Swedish Cancer Society (contracts 030635 to M.R. and K.C. and 050444 to M.R.); Swedish Foundation for Strategic Research (grant A3 02:129 2006 to J.P.). The funding agencies had no role in the design of this study, the data analysis, data interpretation, the writing of the report, or in the authors' decision to submit the paper for publication.

Note

Manuscript received February 11, 2008; revised June 23, 2008; accepted July 17, 2008.