

# Reliability of Self-Reported Family History of Cancer in a Large Case–Control Study of Lymphoma

Ellen T. Chang, Karin Ekström Smedby, Henrik Hjalgrim, Bengt Glimelius, Hans-Olov Adami

**Background:** Case–control studies of familial cancer risk traditionally rely on self-reported family history of cancer, which may bias results due to differential recall between case patients and control subjects. To evaluate the reliability of self-reported data, we analyzed questionnaire and registry-based data on familial cancer from a population-based case–control study of malignant lymphoma. **Methods:** All 1508 lymphoma case patients and 1229 control subjects completed a telephone interview assessing cancer in family members. Participants were linked to the Swedish Multi-Generation Register and Cancer Register to identify confirmed cancer diagnoses in first-degree relatives. The sensitivity and specificity of self-reported familial cancer were calculated among case patients and control subjects and were compared using logistic regression. All statistical tests were two-sided. **Results:** Lymphoma case patients reported a family history of any cancer with statistically significantly higher sensitivity than control subjects (0.85, 95% confidence interval [CI] = 0.83 to 0.87 and 0.80, 95% CI = 0.77 to 0.82, respectively) but with marginally lower specificity (0.89, 95% CI = 0.87 to 0.91 and 0.92, 95% CI = 0.90 to 0.94, respectively). The sensitivity of self-reporting familial cancers by site ranged from less than 0.20 for rare malignancies to nearly 0.75 for more common types, whereas specificity was generally 0.98 or greater. For most sites, the reliability of self-report was similar in patients and control subjects. However, patients reported familial hematopoietic cancer with statistically significantly higher sensitivity (0.60, 95% CI = 0.57 to 0.62) than control subjects (0.38, 95% CI = 0.35 to 0.40). Odds ratios for the association between familial cancer and risk of non-Hodgkin lymphoma were consistently higher when based on self-reported, compared with registry data-based, family history of any cancer or of hematopoietic cancer. **Conclusions:** Reliability of self-reported family history of cancer varies between case patients and control subjects. Recall bias may thus produce biased results in case–control studies of familial cancer risk. [J Natl Cancer Inst 2006;98:61–8]

Family history of hematopoietic malignancy is an established risk factor for lymphoma (1–14). However, in case–control studies in which participants are asked to report their relatives' medical history, recall bias arises if awareness of family history differs systematically between case patients and control subjects. Patients with cancer may be more likely than control subjects to report a positive family history of cancer, leading to overestimated measures of association. Accuracy of reporting may also be affected by other factors, such as the age, sex, and education level of the subject and the amount of time since a family member was diagnosed with cancer. Nevertheless,

family history of cancer is almost always assessed by self-report via questionnaires in observational studies because medical records of family members are difficult or impossible to obtain. In contrast to observational studies, verified cancer diagnoses in relatives can sometimes be assessed reliably in registry-based studies; however, these studies lack the detailed, individualized questionnaire data that enrich interview-based observational studies.

A handful of case–control studies of several cancer types have described the accuracy of self-reported familial cancer (15–20). However, some of these and other validation studies of familial cancer examined only positive self-reports, rather than both positive and negative reports, and the sensitivity and specificity of self-reported data cannot be calculated from such studies. Of the three case–control studies that assessed the accuracy of negative reports of familial cancer (16,18,20), two found substantial false-negative reporting of events (18,20). Most previous studies also suggest considerable false-positive reporting of familial cancer (15–18,20), and some show differences in accuracy between individuals who had and had not been diagnosed with cancer (16,17).

To our knowledge, no previous study has examined the accuracy of reporting familial cancer among lymphoma case patients and control subjects. In Sweden, the unique national Multi-Generation Register (21) identifies all first-degree relatives of index individuals, and the Swedish Cancer Register (22) contains records of all cancer diagnoses in Swedish residents. In a population-based case–control study of malignant lymphoma in Sweden, we linked the participants to these registers to obtain verified data on confirmed cancer diagnoses and lack of diagnoses in participants' relatives. In addition, in a telephone interview, we asked all participants to report their family history of cancer. Thus, we were able to compare validated, registry-based data with self-reported data on familial cancer. Using these two data sources, we were able to evaluate the reliability of self-reported family history in case patients and control subjects and to quantify any bias resulting from the use of such data.

*Affiliations of authors:* Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden (ETC, KES, H-OA); Northern California Cancer Center, Fremont, CA (ETC); Department of Epidemiology Research, Danish Epidemiology Science Center, Statens Serum Institut, Copenhagen, Denmark (HH); Department of Oncology, Radiology and Clinical Immunology, University of Uppsala, Uppsala, Sweden; and Department of Oncology and Pathology, Karolinska University Hospital, Karolinska Institutet, Stockholm, Sweden (BG).

*Correspondence to:* Ellen Chang, ScD, Northern California Cancer Center, 2201 Walnut Ave., Suite 300, Fremont, CA 94538 (e-mail: ellen@nccc.org).

See "Notes" following "References."

DOI: 10.1093/jnci/djj005

© The Author 2006. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org.

## SUBJECTS AND METHODS

### Study Population

The Scandinavian Lymphoma Etiology (SCALE) study is a population-based case-control study of incident malignant lymphoma in Denmark and Sweden (23). For the present study, the source population was restricted to Swedish residents aged 18 to 74 years between October 1, 1999, and April 15, 2002. Eligible participants were required to speak Swedish and to have no history of organ transplantation, human immunodeficiency virus infection, or prior hematopoietic malignancy.

Case patients were diagnosed with incident, morphologically verified non-Hodgkin lymphoma (NHL; International Classification of Diseases, 10th revision [ICD-10]) (24): C82–C85, C88.0, C91.3–5, C91.7), including chronic lymphocytic leukemia (ICD-10: C91.1), or Hodgkin lymphoma (ICD-10: C81). Prevalent Hodgkin lymphoma case patients diagnosed in 1999 were also included. Case patients were identified through a rapid case ascertainment network consisting of contact physicians from all hospital departments in which malignant lymphomas are diagnosed and treated in Sweden and also through continuous collaboration with the Swedish regional cancer registries, which have an estimated coverage of nearly 100% (22). All specimens from case patients were histopathologically evaluated and confirmed by a senior hematopathologist or cytologist affiliated with the study (23).

Control subjects were randomly sampled from the Swedish population every 6 months during the study period, using a continuously updated computerized population register. Control subjects were frequency-matched to the expected 10-year age group and sex distribution of all lymphoma case patients.

Within the defined study base, 85% of eligible lymphoma case patients ( $N = 2347$ ) and 75% of potential control subjects ( $N = 2001$ ) consented to participate in the study. The most common reason for nonparticipation of eligible case patients in our study was early death (6%), whereas the most common reason among control subjects was unwillingness to participate (16%). All participants granted informed consent before the interview. This study was approved by all regional ethics committees in Sweden.

### Registry Linkage

The Swedish Multi-Generation Register was created by Statistics Sweden in the early 1990s using national registration records; it includes individuals born in 1932 or later and those alive in 1961 or born thereafter (21). The Register contains information on individuals, referred to as index persons, and their parents, siblings, and children, with indicators for nonbiologic (e.g., adoptive) relationships. Siblings are identified indirectly through common linkages to parents. Parental information is nearly complete for subjects and their siblings who were alive in 1991 and about 50% complete for those who died between 1961 and 1991. The Multi-Generation Register currently includes more than 11 million individuals (approximately 3.2 million nuclear families).

The Swedish Cancer Register was established in 1958 and contains data on all newly diagnosed malignant tumors in Sweden (22). It is required by law that tumors be reported to the Cancer Register by both the diagnosing clinician and the responsible pathologist or cytologist. Although tumor registration was not complete during the Register's initial years, nearly 100% of newly diagnosed cancers are now recorded.

Based on their individually unique national registration numbers, we linked all participants in the SCALE study born in 1932 or later (1841 lymphoma case patients [78%] and 1498 control subjects [75%]) to the Multi-Generation Register to identify their parents and siblings. All participants and their relatives were then linked by national registration number to the Swedish Cancer Register to ascertain cancer diagnoses in all individuals. To ensure that parents and siblings and their registered cancer diagnoses were as completely identified as possible, we limited the study population to individuals who were linked to both of their parents (1508 lymphoma case patients and 1229 control subjects, or 82% of both groups). Patients were linked to a total of 2917 siblings, and control subjects to a total of 2442 siblings. We excluded familial cancers recorded in the Cancer Register that were diagnosed after or within 30 days before the date of interview of any participant (83 cancers among relatives of case patients and 42 among relatives of control subjects). We also excluded familial cancers that had been detected only at autopsy (21 cancers among relatives of case patients and 17 among relatives of control subjects).

### Classification of Cancer Type

Based on the International Classification of Diseases, 7th Revision (ICD-7) codes (25), diagnoses in the Cancer Register were grouped in the following organ-specific sites or areas: any site (ICD-7: 140–209); buccal cavity, pharynx, nasal cavity, or larynx (ICD-7: 140–148, 160–161); esophagus (ICD-7: 150); stomach (ICD-7: 151); small intestine, colon, or rectum (ICD-7: 152–154); liver or bile passages (ICD-7: 155–156); pancreas (ICD-7: 157); lung (ICD-7: 162–163); breast (ICD-7: 170); uterine cervix or corpus (ICD-7: 171–174); ovary (ICD-7: 175); prostate (ICD-7: 177); testis or other male genital organ (ICD-7: 178–179); kidney (ICD-7: 180); bladder or other urinary organ (ICD-7: 181); brain or other part of nervous system (ICD-7: 193); thyroid or other endocrine gland (ICD-7: 194–195); bone (ICD-7: 196); and hematopoietic system (ICD-7: 200–209), including lymphoma (ICD-7: 200–202, 204.1, 205), multiple myeloma (ICD-7: 203), or leukemia (ICD-7: 204, 206–207, excluding 204.1). Self-reported diagnoses in parents and siblings were also classified into the same groups. Some of these groupings (e.g., small intestine/colon/rectum or buccal cavity/larynx/nasal cavity/pharynx) were created to reflect a high frequency of poorly defined self-reported diagnoses (e.g., “intestinal cancer” or “throat cancer”), which indicated an anatomic area but not a precise affected organ or site. Among 1846 total self-reported diagnoses in relatives (693 maternal, 684 paternal, and 469 sibling cancers, including reports of multiple cancers in one family), 113 (6%) could not be classified by site due to unclear or uninterpretable responses and were therefore excluded from the main analysis. All registered and self-reported skin cancers were excluded from the analysis, because most subjects did not distinguish among the three main skin cancer types (malignant melanoma, squamous cell carcinoma, and basal cell carcinoma) and because basal cell carcinoma is not recorded in the Swedish Cancer Register.

### Telephone Interview

All participants in SCALE completed a detailed telephone interview assessing known and suspected risk factors for

lymphoma, including age, sex, education level, family background, and family history of cancer. Participants' education level was classified as  $\leq 9$ , 10–12, or  $\geq 13$  years of schooling; for participants' parents, education level was classified as  $< 7$ , 7–9, or  $\geq 10$  years. Family background included number of siblings (0, 1, 2, or  $\geq 3$ ); birth order (1st, 2nd, 3rd, or 4th or later); participants' birthplace (Sweden or elsewhere); and parents' birthplace (both in Sweden or not). Cut points for variables were determined by the distribution among the control subjects. Participants also provided each sibling's year of birth and age at death, if the sibling had died.

Participants were asked whether a parent or sibling had ever been diagnosed with cancer and, if so, which relative. They were also asked to report the age at diagnosis and type of cancer for each family member. Cancer type was reported as an open-ended, rather than fixed-choice, response. Based on self-reported year of birth and age at cancer diagnosis of family members, we calculated the year of diagnosis of all parent and sibling cancers. Cancers that were reported as being diagnosed before 1958, when the Cancer Register began, were excluded from the analysis (45 cancers among relatives of case patients and 35 cancers among relatives of control subjects). Nine of 10 self-reported familial cancers stated to have been diagnosed in 1958 were recorded in the Cancer Register.

## Statistical Methods

The sensitivity (i.e., true positive proportion) of self-reported family history of cancer was calculated as the proportion of individuals who reported familial cancer among all those who had familial diagnoses recorded in the Cancer Register. The specificity (i.e., true negative proportion) of self-reported family history of cancer was calculated as the proportion of individuals who reported no familial cancer among those who had no familial diagnoses recorded in the Cancer Register. Sensitivity and specificity were calculated separately for case patients and control subjects. For estimates of the reliability of self-reported sibling history of cancer, individuals who reported having no siblings were excluded. Large-sample 95% confidence intervals (CIs) for all proportions were calculated using the normal approximation to the binomial distribution.

We used logistic regression, controlling for age group (in 10-year categories), sex, education level, and birth order, to evaluate differences in sensitivity and specificity of self-reported family history of cancer between case patients and control subjects. We used the same approach, additionally adjusting for patient/control status, to evaluate whether sensitivity and specificity of self-reported familial cancer varied by age group, sex, own or parents' education level, number of siblings, birth order, own or parents' birthplace, years between first cancer diagnosis in a first-degree relative and interview (0–9, 10–19, 20–29, or  $\geq 30$  years), or time between diagnosis and interview among case patients ( $\leq 60$  or  $> 60$  days, based on a natural cutoff near the median of the distribution). All characteristics were entered as categorical variables in the statistical models, with cutoffs as defined above. To estimate the relative risk (approximated by the odds ratio [OR] with corresponding 95% confidence intervals) of NHL associated with a family history of cancer, we again performed logistic regression, adjusting for age group, sex, education level, and birth order. Chi-square tests were used to compare

values between groups;  $P$  values  $< .05$  (two-sided) were considered statistically significant.

We conducted secondary analyses excluding the 35 control subjects (2.9%) with a personal history of nonhematopoietic malignancy. Because information in the Multi-Generation Register is incomplete for individuals who died before 1991, we were concerned about not identifying some siblings who died before that year. Therefore, we also performed secondary analyses limited to individuals who did not report having a deceased sibling before 1991 (1343 case patients [89%] and 1114 control subjects [91%]). The statistical analyses were performed with SAS System Software, release 9.1 (SAS Institute Inc., Cary, NC, 1999–2001).

## RESULTS

### Reliability of Self-Reported Familial Cancer

The sensitivity and specificity of self-reported family history of cancer among lymphoma case patients and control subjects are presented in Table 1. For the analysis of individuals' family history of any cancer, parental and sibling cancers were considered separately; for individuals' family history of specific cancer sites or areas, parental and sibling cancers were counted together to enrich numbers. The sensitivity of self-reported family history of any cancer for both case patients and control subjects was higher for parental than sibling history of cancer, whereas specificity was approximately the same for parental and sibling history in both patients and control subjects. The sensitivity of self-reported familial cancer at individual sites or areas was quite low—mostly between 0.20 and 0.75—with generally highest sensitivity for cancers at more common sites, such as the lung, breast, and stomach (data not shown for less common cancer sites or areas). Sensitivity was generally, although not always, lower for particular cancer types, such as lymphoma (0.31, 95% CI = 0.29 to 0.32 overall), myeloma (0.12, 95% CI = 0.11 to 0.13 overall), or leukemia (0.50, 95% CI = 0.48 to 0.52 overall) than for broader areas, such as the hematopoietic system (0.53, 95% CI = 0.51 to 0.55). The specificity of self-reported familial cancer was mostly between 0.98 and 1.00 for individual cancer sites or areas but was lower (0.89 to 0.96) for all sites combined.

### Results According to Patient or Control Status

The sensitivity of self-reported family history of any cancer was statistically significantly higher ( $P = .02$ ) among lymphoma case patients (0.85, 95% CI = 0.83 to 0.87) than among control subjects (0.80, 95% CI = 0.77 to 0.82), controlling for age, sex, education level, and birth order (Table 1). Patients also reported parental history of any cancer with higher sensitivity than control subjects, but the difference was not statistically significant. However, the specificity of self-reported family history of any cancer was marginally higher ( $P = .07$ ) among control subjects (0.92, 95% CI = 0.90 to 0.94) than among case patients (0.89, 95% CI = 0.87 to 0.91). Specificities of self-reported parental history and sibling history of cancer were also higher among control subjects than among case patients but not statistically significantly so. The differences in reliability between case patients and control subjects were not markedly affected by further

**Table 1.** Sensitivity and specificity of self-reported history versus registered diagnoses among first-degree relatives of case patients with malignant lymphoma and control subjects

Cancer type (ICD-7 code) and case patient/control subject status*	Self-reported familial cancer†	Cancer diagnosis in Swedish Cancer Register‡		Sensitivity (95% CI)	P§	Specificity (95% CI)	P§
		Registered	Not registered				
<b>All (140–209)</b>							
In any first-degree relative(s)							
Control subjects	Yes	376	52	0.80 (0.77 to 0.82)		0.92 (0.90 to 0.94)	
	No	96	604				
Case patients	Yes	517	83	0.85 (0.83 to 0.87)		0.89 (0.87 to 0.91)	
	No	89	681				
In parent(s)							
Control subjects	Yes	335	31	0.78 (0.76 to 0.81)		0.96 (0.95 to 0.97)	
	No	93	697				
Case patients	Yes	448	52	0.82 (0.80 to 0.84)		0.94 (0.93 to 0.95)	
	No	99	816				
In sibling(s)							
Control subjects	Yes	55	41	0.67 (0.64 to 0.70)		0.96 (0.95 to 0.97)	
	No	27	950				
Case patients	Yes	82	70	0.68 (0.65 to 0.70)		0.94 (0.93 to 0.95)	
	No	39	1130				
Stomach (151)							
Control subjects	Yes	17	32	0.71 (0.68 to 0.73)		0.97 (0.96 to 0.98)	
	No	7	1168				
Case patients	Yes	30	52	0.70 (0.67 to 0.72)		0.96 (0.95 to 0.97)	
	No	13	1403				
Small intestine/colon/rectum (152–154)							
Control subjects	Yes	40	10	0.53 (0.50 to 0.55)		0.99 (0.99 to 1.00)	
	No	36	1137				
Case patients	Yes	46	9	0.48 (0.46 to 0.51)		0.99 (0.99 to 1.00)	
	No	49	1389				
Lung/bronchus (162–163)							
Control subjects	Yes	33	23	0.72 (0.69 to 0.74)		0.98 (0.97 to 0.99)	
	No	13	1150				
Case patients	Yes	48	24	0.81 (0.79 to 0.83)		0.98 (0.98 to 0.99)	
	No	11	1412				
Breast (170)							
Control subjects	Yes	61	13	0.72 (0.69 to .74)		0.99 (0.98 to 1.00)	
	No	24	1114				
Case patients	Yes	72	18	0.73 (0.70 to .75)		0.99 (0.98 to 0.99)	
	No	27	1373				
Prostate (177)							
Control subjects	Yes	52	6	0.60 (0.57 to .63)		0.99 (0.99 to 1.00)	
	No	35	1122				
Case patients	Yes	49	19	0.47 (0.44 to .49)		0.99 (0.98 to 0.99)	
	No	56	1366				
Hematopoietic system (200–209)							
Control subjects	Yes	18	16	0.38 (0.35 to .40)		0.99 (0.98 to 0.99)	
	No	30	1157				
Case patients	Yes	61	30	0.60 (0.57 to .62)		0.98 (0.97 to 0.99)	
	No	41	1360				

\*ICD-7 = International Classification of Diseases, 7th revision. CI = confidence interval.

†Excludes cancers reportedly diagnosed before 1958 (start date of Cancer Register).

‡Excludes cancers diagnosed after or up to 30 days before date of interview or at autopsy.

§P value (two-sided) for chi-square test for difference between case patients and control subjects, adjusting for age (in 10-year categories), sex, education level ( $\leq 9$ , 10–12, or  $\geq 13$  years), and birth order (1st, 2nd, 3rd, or  $\geq 4$ th).

||Excludes individuals who reported having no siblings.

adjustment for number of siblings, parental education level, or personal or parental birthplace (data not shown). Exclusion of control subjects with a personal history of nonhematopoietic malignancy did not change the results (data not shown). Also, there were no statistically significant differences in the sensitivity or specificity of self-reported sibling cancer among individuals without a sibling who died before 1991, compared with the overall study population (data not shown), although the specificity of self-reported sibling cancer was statistically significantly higher

among control subjects than among case patients in this subgroup ( $P = .02$ ).

Patients and control subjects demonstrated similar reliability of self-reported family history of most individual cancer sites and areas except prostate cancer, for which control subjects' self-reports were statistically significantly more specific than those of case patients, although the specificity rounded off to 0.99 among both case patients and control subjects. In addition, the sensitivity of self-reported family history of

hematopoietic malignancy was statistically significantly higher ( $P = .04$ ) among lymphoma case patients (0.60, 95% CI = 0.57 to 0.62) than among control subjects (0.38, 95% CI = 0.35 to 0.40). Among those who provided unclassifiable (non-site-specific) information about the type of cancer diagnosed in a family member, 23 of 29 control subjects' reports (79%) were verified as actual cancers in the Cancer Register, whereas 36 of 41 case patients' reports (88%) were verified ( $P = .14$  for difference by case status, adjusting for age, sex, education, and birth order).

### Other Factors Affecting Reliability

To evaluate predictors of reliability other than a personal history of cancer, we examined the distribution of the sensitivity

and specificity of self-reported family history of cancer according to various subject characteristics, adjusting for age, sex, education level, birth order, and case patient/control subject status (Table 2). Younger age was statistically significantly associated with higher specificity of self-reported familial cancer. Females reported familial cancer with statistically significantly higher sensitivity and lower specificity than males. Earlier birth order was also statistically significantly associated, and higher education was marginally associated, with higher specificity. Individuals whose first familial diagnosis of cancer occurred 20 or more years ago reported a family history of cancer with statistically significantly higher sensitivity than those whose family member(s) were diagnosed more recently. Sensitivity and specificity of self-reported family history of cancer did not vary according to the parents' education level, number of siblings,

**Table 2.** Predictors of sensitivity and specificity of self-reported family history of cancer at any site\*

Characteristic	Sensitivity (95% CI)	$P^\dagger$	Specificity (95% CI)	$P^\dagger$
Age, y				
<50	0.85 (0.83 to 0.88)		0.96 (0.95 to 0.97)	
50–59	0.82 (0.79 to 0.84)		0.89 (0.87 to 0.91)	
≥60	0.83 (0.80 to 0.85)	.34	0.82 (0.79 to 0.85)	<.001
Sex				
Male	0.81 (0.79 to 0.83)		0.93 (0.92 to 0.94)	
Female	0.86 (0.84 to 0.88)	.05	0.87 (0.85 to 0.89)	<.001
Education level, y				
≤9	0.82 (0.79 to 0.84)		0.84 (0.81 to 0.87)	
10–12	0.82 (0.80 to 0.84)		0.92 (0.90 to 0.93)	
≥13	0.86 (0.83 to 0.89)	.33	0.93 (0.91 to 0.95)	.11
Maternal education level, y				
<7	0.81 (0.78 to 0.83)		0.86 (0.84 to 0.88)	
7–9	0.83 (0.80 to 0.86)		0.92 (0.89 to 0.94)	
≥10	0.89 (0.86 to 0.91)	.12	0.96 (0.94 to 0.97)	.66
Paternal education level, y				
<7	0.80 (0.78 to 0.83)		0.86 (0.84 to 0.88)	
7–9	0.85 (0.82 to 0.88)		0.93 (0.91 to 0.95)	
≥10	0.86 (0.83 to 0.88)	.36	0.94 (0.92 to 0.96)	.95
No. of siblings				
0	0.81 (0.76 to 0.86)		0.95 (0.93 to 0.98)	
1	0.86 (0.84 to 0.88)		0.95 (0.94 to 0.97)	
2	0.84 (0.81 to 0.87)		0.92 (0.90 to 0.94)	
≥3	0.80 (0.80 to 0.81)	.37	0.82 (0.81 to 0.83)	.52
Birth order				
1st	0.86 (0.84 to 0.89)		0.94 (0.92 to 0.96)	
2nd	0.83 (0.80 to 0.85)		0.94 (0.92 to 0.96)	
3rd	0.80 (0.76 to 0.84)		0.87 (0.83 to 0.90)	
4th or later	0.81 (0.80 to 0.81)	.30	0.72 (0.72 to 0.73)	<.001
Birthplace				
Sweden	0.83 (0.81 to 0.84)		0.90 (0.89 to 0.91)	
Other	0.76 (0.66 to 0.87)	.58	0.96 (0.91 to 1.00)	.25
Parents' birthplace				
Both in Sweden	0.83 (0.81 to 0.84)		0.90 (0.89 to 0.91)	
Other	0.81 (0.76 to 0.87)	.79	0.94 (0.91 to 0.97)	.64
Years between first cancer diagnosis in relative and date of interview‡				
0–9	0.77 (0.72 to 0.81)		NA	
10–19	0.82 (0.78 to 0.86)			
20–29	0.87 (0.82 to 0.91)			
≥30	0.89 (0.85 to 0.93)	<.001		
Days between lymphoma diagnosis and date of interview (among case patients only)				
0–60	0.85 (0.82 to 0.88)		0.89 (0.87 to 0.92)	
>60	0.86 (0.83 to 0.88)	.68	0.89 (0.87 to 0.91)	.62

\*CI = confidence interval; NA = not applicable.

† $P$  value (two-sided) from chi-square test for difference by subject characteristic, adjusted for patient/control status, age (in 10-year categories), sex, education level, and birth order.

‡Year of diagnosis according to Swedish Cancer Register.

**Table 3.** Odds ratios (ORs) and 95% confidence intervals (CIs) for association between family history of cancer and risk of non-Hodgkin lymphoma (NHL), based on self-reported family history or on Swedish Cancer Register data

Family history of malignancy based on registry data	Self-reported questionnaire data*			Swedish Cancer Register data†			Difference between ORs (%‡)
	Control subjects (N = 1229)	NHL case patients (N = 1209)	OR (95% CI)‡	Control subjects (N = 1229)	NHL case patients (N = 1209)	OR (95% CI)‡	
Cancer in family							
No	747	604	1.0 (referent)	677	585	1.0 (referent)	
Yes	439	544	1.3 (1.1 to 1.6)	488	562	1.2 (1.0 to 1.4)	14
Cancer in parent(s)							
No	830	734	1.0 (referent)	738	662	1.0 (referent)	
Yes	377	450	1.2 (1.0 to 1.5)	440	496	1.1 (0.9 to 1.4)	8
Cancer in sibling(s)							
No	1121	1056	1.0 (referent)	1120	1065	1.0 (referent)	
Yes	97	145	1.3 (1.0 to 1.8)	84	114	1.2 (0.9 to 1.7)	9
Hematopoietic cancer in family							
No	1191	1121	1.0 (referent)	1175	1107	1.0 (referent)	
Yes	35	83	2.1 (1.3 to 3.2)	49	94	1.7 (1.1 to 2.4)	25
Hematopoietic cancer in parent(s)							
No	1201	1148	1.0 (referent)	1184	1131	1.0 (referent)	
Yes	26	57	1.7 (1.0 to 2.9)	41	73	1.5 (0.9 to 2.2)	18
Hematopoietic cancer in sibling(s)							
No	1218	1180	1.0 (referent)	1220	1182	1.0 (referent)	
Yes	10	28	2.7 (1.3 to 5.7)	8	23	2.6 (1.2 to 6.0)	4
Lymphoma in family							
No	1217	1177	1.0 (referent)	1199	1142	1.0 (referent)	
Yes	12	32	2.5 (1.2 to 5.1)	28	62	1.9 (1.2 to 3.2)	30
Multiple myeloma in family							
No	1229	1203	1.0 (referent)	1217	1189	1.0 (referent)	
Yes	0	6	Undefined	11	20	1.3 (0.6 to 2.8)	Undefined
Leukemia in family							
No	1210	1165	1.0 (referent)	1219	1193	1.0 (referent)	
Yes	17	39	1.6 (0.9 to 3.0)	8	14	1.6 (0.6 to 4.4)	2

\*Excludes cancers reportedly diagnosed before 1958 (start date of Cancer Register).

†Excludes cancers diagnosed after or up to 30 days before date of interview or at autopsy.

‡Adjusted for age (in 10-year categories), sex, education level ( $\leq 9$ , 10–12, or  $\geq 13$  years), and birth order (1st, 2nd, 3rd, or  $\geq 4$ th).

§Percent difference between odds ratio based on Swedish Cancer Register data and odds ratio based on self-reported data (using odds ratios calculated to two decimal places).

subject's birth country, parents' birth country, or time interval between diagnosis and interview among case patients only.

### Degree of Bias in Estimates of Association

To quantify the bias in estimates of associations based on self-reported rather than registry-validated family history of cancer, we calculated the odds ratios for the association between family history of cancer and risk of NHL, using both interview data and Cancer Register data, controlling for age, sex, education level, and birth order (Table 3). Additional adjustment for number of siblings, parental education, or personal or parental birthplace did not change the associations (data not shown). For associations between risk of NHL and a family history of any cancer or of hematopoietic cancers in particular, odds ratios based on self-reported data were uniformly higher (as much as 30% higher) than those based on Cancer Register data. Family history of cancer at other sites, whether self-reported or from registry data, was not associated with NHL risk (data not shown).

### DISCUSSION

By comparing self-reported with validated (i.e., registry-reported) family history of cancer, we found that self-reported data are far from perfectly reliable and that reliability varies between case patients and control subjects. Overall, the sensitiv-

ity of self-reported familial cancer was generally less than 0.80 and was substantially lower for specific cancer sites, especially rare ones. The sensitivity of self-reported familial cancer, especially hematopoietic cancer, was statistically significantly higher and specificity was statistically nonsignificantly lower in lymphoma case patients compared with control subjects. These differences remained after controlling for age, sex, educational level, and birth order, which were associated with reliability. As a result, measures of association between familial cancer and risk of NHL were substantially overestimated.

In contrast to the low sensitivity of self-reported familial cancer, specificity was generally greater than 0.90, with higher values for specific cancer sites. The reliability of negative self-reports was expected to be high, because cancer is a relatively uncommon disease, especially when individual cancer sites are considered. Inaccuracy of self-reported familial cancer may have been due to misinformation or lack of awareness about a relative's cancer, confusion between benign and malignant conditions, failure to recall cancer in remission, lack of communication within a family, or unwillingness to report a relative's cancer history. Low sensitivity of site-specific data, especially for rare or typically metastatic cancer sites, was probably due to the general population's limited knowledge about anatomy and differences between primary and metastatic cancer.

Our finding of higher sensitivity and lower specificity of self-reported familial hematopoietic cancer among lymphoma case

patients than among control subjects has important implications. These results suggest that in case-control studies of a specific cancer type, case patients are more likely than control subjects to report both true-positive and false-positive family histories of their particular cancer, resulting in inflated estimates of the relative risk associated with a family history of that malignancy. Indeed, we found that the odds ratios for the association between a family history of any cancer or of hematopoietic cancer in particular and risk of NHL were biased upward by up to 30% when based on self-reported rather than validated registry-based data.

We also evaluated whether characteristics other than a personal history of cancer were predictors of the reliability of self-reported familial cancer. Subjects who were younger, male, better educated, or first-born reported a family history of cancer with higher specificity, whereas females and those with a relative diagnosed 20 or more years ago reported familial cancer with higher sensitivity. The latter finding is consistent with the observation that parental cancers—which probably occurred longer ago on average than sibling cancers—were reported with higher sensitivity than sibling cancers. This difference in the sensitivity of self-reported parental versus sibling cancer could be attributed to individuals' closer parental than sibling ties or to greater care-taking responsibilities for children, especially first-borns, than siblings of cancer patients. Age, sex, education level, or family structure and responsibilities, as reflected by birth order, may also affect an individual's decision to participate in a study. Because these demographic and personal factors also affect the reliability of self-reported family history of cancer, differences in the accuracy of such self-reports between case patients and control subjects may bias estimates of association with familial cancer if not properly accounted for. Even after adjusting for age, sex, education level, and birth order, however, we found that substantial recall bias remained.

The only other two studies, both of small size, in which sensitivity and specificity of self-reported familial cancer were calculated found generally results similar to ours. In a study by Mitchell et al. (20) of 199 colorectal cancer case patients and 133 control subjects, the sensitivity of self-reported cancer in a first-degree relative was 0.71 among case patients and 0.67 among control subjects. The sensitivity was substantially lower for specific cancers other than breast cancer, whereas specificity was generally greater than 0.99. In another case-control study of colorectal cancer, by Aitken et al. (16), with 74 case patients and 163 control subjects, the sensitivity of self-reported familial colorectal cancer was 0.82 among case patients and 0.76 among control subjects; the specificity in both groups overall was 0.99. Other case-control studies that validated only positive self-reports of familial cancer also found that approximately 70%–90% of such reports were accurate, with the accuracy typically being lower for site-specific cancers, especially those at uncommon sites or those that were not a focus of the investigation (15,17,18).

Our study has some limitations, mainly imposed by the starting dates of the Multi-Generation Register and the Swedish Cancer Register. The Multi-Generation Register includes only individuals born in 1932 or after and contains incomplete family data for individuals who died before 1991. To increase the likelihood of recording cancer history for all family members of case patients and control subjects, we restricted our study to participants whose parents could both be identified. These restrictions

substantially reduced the size of our study population and also prevented us from studying index individuals older than age 70 years. Also, because the Swedish Cancer Register began recording cancer diagnoses in 1958 and was not complete for several years thereafter, it is likely that some self-reported familial cancers were unconfirmed because they were not registered, rather than because they were inaccurate.

However, these limitations are unlikely to have biased our results. Because an equal proportion of case patients and control subjects had identifiable parents and an equal proportion of both groups were nonimmigrants, the differential reliability of self-reported data between case patients and control subjects was unlikely to have been due to discrepancies in ascertainment of familial cancer history. Also, although the Multi-Generation Register does not completely identify siblings who died before 1991, we found that the results did not change when the analysis was limited to individuals who did not report having a deceased sibling before 1991. To avoid underestimating the reliability of positive self-reports, we excluded self-reported cancers that were stated to have been diagnosed before 1958. Furthermore, we found that 90% of familial cancers said to have been diagnosed in 1958 were indeed recorded in the Cancer Register. Although incomplete registration may have caused us to underestimate specificity, it is unlikely to have introduced differential bias in our results, because completeness should not have differed between case patients and control subjects.

One potential limitation that we were unable to account for was that of dissimilar reasons for nonparticipation between case patients and control subjects. That is, the most common reason for nonparticipation of eligible case patients in our study was early death, which could be associated with familial cancer history. In contrast, most eligible control subjects did not participate due to unwillingness. However, the effect of this differential nonparticipation, if any, on specificity and sensitivity is unclear.

In summary, this study is, to our knowledge, by far the largest case-control study of cancer to validate both positive and negative self-reported family history of cancer in participants. Furthermore, our ability to link our study participants and their relatives to the Swedish Cancer Register conferred an important advantage over manually searching for relatives' medical records, which are often missing or otherwise unobtainable. Our results suggest that estimates of the association between familial cancer and risk of malignancy may be biased in case-control studies that rely on self-reported family history. Although the extent of this bias is likely to vary by study population, specific type(s) of cancer assessed, and time period and although our results strictly apply only to our study population, we believe that the reliability of self-reported familial cancer is likely to be poorer in many other study populations than ours—a well-educated Swedish population with a high awareness of and willingness to discuss cancer. With a growing scientific emphasis on the study of genetic determinants of cancer development, it is increasingly important to obtain an accurate assessment of family history of cancer. Therefore, future observational studies should make any possible effort to validate both positive and negative self-reports of cancer in family members.

## REFERENCES

- (1) Razis DV, Diamond HD, Craver LF. Familial Hodgkin's disease: its significance and implications. *Ann Intern Med* 1959;51:933–71.

- (2) Grufferman S, Cole P, Smith PG, Lukes RJ. Hodgkin's disease in siblings. *N Engl J Med* 1977;296:248–50.
- (3) Cuttner J. Increased incidence of hematologic malignancies in first-degree relatives of patients with chronic lymphocytic leukemia. *Cancer Invest* 1992;10:103–9.
- (4) Linet MS, Pottern LM. Familial aggregation of hematopoietic malignancies and risk of non-Hodgkin's lymphoma. *Cancer Res* 1992;52:5468s–73s.
- (5) Shpilberg O, Modan M, Modan B, Chetrit A, Fuchs Z, Ramot B. Familial aggregation of haematological neoplasms: a controlled study. *Br J Haematol* 1994;87:75–80.
- (6) Mack TM, Cozen W, Shibata DK, Weiss LM, Nathwani BN, Hernandez AM, et al. Concordance for Hodgkin's disease in identical twins suggesting genetic susceptibility to the young-adult form of the disease. *N Engl J Med* 1995;332:413–8.
- (7) Paltiel O, Schmit T, Adler B, Rachmilevitz EA, Polliack A, Cohen A, et al. The incidence of lymphoma in first-degree relatives of patients with Hodgkin disease and non-Hodgkin lymphoma: results and limitations of a registry-linked study. *Cancer* 2000;88:2357–66.
- (8) Hemminki K, Czene K. Attributable risks of familial cancer from the Family-Cancer Database. *Cancer Epidemiol Biomarkers Prev* 2002;11:1638–44.
- (9) Chiu BC, Weisenburger DD, Zahm SH, Cantor KP, Gapstur SM, Holmes F, et al. Agricultural pesticide use, familial cancer, and risk of non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev* 2004;13:525–31.
- (10) Chatterjee N, Hartge P, Cerhan JR, Cozen W, Davis S, Ishibe N, et al. Risk of non-Hodgkin's lymphoma and family history of lymphatic, hematologic, and other cancers. *Cancer Epidemiol Biomarkers Prev* 2004;13:1415–21.
- (11) Goldin LR, Pfeiffer RM, Gridley G, Gail MH, Li X, Mellekjær L, et al. Familial aggregation of Hodgkin lymphoma and related tumors. *Cancer* 2004;100:1902–8.
- (12) Domingo-Domenech E, Benavente Y, Alvaro T, Hernandez M, de Sevilla AF, de Sanjose S. Family clustering of blood cancers as a risk factor for lymphoid neoplasms. *Haematologica* 2005;90:416–8.
- (13) Altieri A, Bermejo JL, Hemminki K. Familial risk of non-Hodgkin lymphoma and other lymphoproliferative malignancies by histopathologic subtype from the Swedish family-cancer database. *Blood* 2005;106:668–72.
- (14) Chang ET, Ekström Smedby K, Hjalgrim H, Porwit-MacDonald A, Roos G, Glimelius B, et al. Family history of hematopoietic cancer and risk of lymphoma. *J Natl Cancer Inst* 2005;97:1466–74.
- (15) Koch M, Hill GB. Problems in establishing accurate family history in patients with ovarian cancer of epithelial origin. *Cancer Detect Prev* 1987;10:279–83.
- (16) Aitken J, Bain C, Ward M, Siskind V, MacLennan R. How accurate is self-reported family history of colorectal cancer? *Am J Epidemiol* 1995;141:863–71.
- (17) Parent ME, Ghadirian P, Lacroix A, Perret C. Accuracy of reports of familial breast cancer in a case-control series. *Epidemiology* 1995;6:184–6.
- (18) Kerber RA, Slattery ML. Comparison of self-reported and database-linked family history of cancer data in a case-control study. *Am J Epidemiol* 1997;146:244–8.
- (19) Garbers V, Toniolo PG, Taioli E. Changes in self-reported family history of breast cancer with change in case-control status. *Eur J Epidemiol* 2001;17:517–20.
- (20) Mitchell RJ, Brewster D, Campbell H, Porteous ME, Wyllie AH, Bird CC, et al. Accuracy of reporting of family history of colorectal cancer. *Gut* 2004;53:291–5.
- (21) Statistics Sweden. The Multi-Generation Register. Örebro (Sweden): Statistics Sweden; 2004.
- (22) Swedish Cancer Register. Cancer Incidence in Sweden 2001. Stockholm (Sweden): Centre for Epidemiology; 2003.
- (23) Ekström Smedby K, Hjalgrim H, Melbye M, Torrång A, Rostgaard K, Munksgaard L, et al. Ultraviolet radiation exposure and risk of malignant lymphomas. *J Natl Cancer Inst* 2005;97:199–209.
- (24) International statistical classification of diseases and related health problems. 10th rev. Geneva (Switzerland): World Health Organization; 1992.
- (25) Report of the International Conference for the seventh revision of the international lists of diseases and causes of death. Geneva (Switzerland): World Health Organization; 1955.

## NOTES

Supported by the National Institutes of Health (R01 CA069269-01 to M. Melbye) and the Stockholm Cancer Foundation (Cancerföreningen to B. Glimelius). The funding agencies did not contribute to the design or operation of the study, nor to the writing or submission of this manuscript.

We thank Leila Nyrén (Karolinska Institutet) for coordinating the project; Marie Reilly (Karolinska Institutet) for her statistical advice and expertise with using Multi-Generation Register data; and Anna Porwit-MacDonald (Karolinska University Hospital), Göran Roos (Umeå University Hospital), Christer Sundström (Akademiska Hospital, Uppsala), Edneia Tani (Karolinska University Hospital), Måns Åkerman (Lund University Hospital), and Åke Öst (Medilab, Stockholm) for their extensive histopathologic and cytologic review of tumor material. We are indebted to all of the contact doctors and nurses throughout Sweden who participated in our rapid case ascertainment system.

Manuscript received August 1, 2005; revised October 18, 2005; accepted November 7, 2005.