ELSEVIER

Methodologic Issues in Follow-Up Studies of Cancer Incidence Among Occupational Groups in the United States

THOMAS J. BENDER, PhD, COLLEEN BEALL, DrPH, HONG CHENG, PhD,
ROBERT F. HERRICK, SD, AMY R. KAHN, MS, ROBERT MATTHEWS, BS,
NALINI SATHIAKUMAR, MD, DrPH, MARIA J. SCHYMURA, PhD,
JAMES H. STEWART, PhD, AND ELIZABETH DELZELL, SD

PURPOSE: Incidence studies of occupational factors and cancer in the United States are problematic because the use of population-based registries to identify cases requires development of historical data on subjects' residences and often severely restricts the time period of follow up. This article describes procedures for addressing these challenges.

METHODS: We used data from studies of cancer incidence and mortality among microelectronics industry employees to assess various methods for developing residential histories and the relative informativeness of the two studies.

RESULTS: We developed residential histories for 98% of 99,229 mortality study subjects. Analyses making alternative assumptions about residential histories yielded standardized incidence ratios varying by at most 6%. Use of postemployment residential histories increased person-years by up to 62% and increased the observed number of cancers by up to 28%. The proportion of mortality study person-years included in the cancer incidence study ranged from 40% to 77% among work activity subcohorts. The number of observed cancer cases in the incidence study was 60% higher than the number of observed cancer deaths in the mortality study.

CONCLUSIONS: Assumptions about residential history had little impact on validity. Use of information sources with national coverage to develop residential histories increased the incidence study's precision. Despite geographic and temporal restrictions, incidence studies provide more data than mortality studies on cancers with good survival. However, the potential for selection bias in incidence studies may vary considerably among subcohorts, indicating the need for cautious interpretation of such research.

Ann Epidemiol 2006;16:170–179. © 2006 Elsevier Inc. All rights reserved.

KEY WORDS: Follow-Up Studies, Incidence, Occupational Diseases, Neoplasms, Registries, Residential Mobility, Retrospective Studies.

INTRODUCTION

Follow-up studies of cancer incidence are potentially more comprehensive and informative than mortality studies but may pose several challenges (1, 2). Most incidence studies use record linkage with cancer registries to identify cases because obtaining incidence data directly from large numbers of subjects is problematic (3–5). Reliance on registries to identify cases requires development of historical data on subjects' residences to compute person-time at risk accurately because cancer registry coverage in the United States (US) is temporally and geographically limited (6, 7). This article evaluates the completeness and accuracy of information sources used to develop residential histories, assesses the impact on validity and precision of procedures and assumptions used to develop residential histories, evaluates variation in the impact of follow-up restrictions among subcohorts specified on the basis of work activity, and describes the informativeness of a recent cancer-incidence study relative to its companion mortality study.

METHODS

Data came from a mortality study of International Business Machines (IBM) employees at three microelectronics facilities (8) and a cancer incidence study of employees at two of the three facilities (9). The latter two facilities were located in East Fishkill, New York (NY), and San Jose, California (CA). Electronic databases of personnel records were used to identify subjects and to develop work histories. We classified work histories according to manufacturing activity (15 "work groups" at East Fishkill and 19 at San

From the Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL (T.J.B, C.B., H.C., R.M., N.S., E.D.); Department of Environmental Health, Harvard School of Public Health, Boston, MA (R.F.H., J.H.S.); and New York State Cancer Registry, New York State Department of Health, Albany, NY (A.R.K., M.J.S.).

Address correspondence to: Thomas J. Bender, Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, 523 Ryals Building, 1665 University Blvd, Birmingham, AL 35294. Tel.: (205) 934-1200; fax: (205) 975-7058. E-mail: bender@uab.edu.

Received February 18, 2005; accepted July 21, 2005.

© 2006 Elsevier Inc. All rights reserved.

360 Park Avenue South, New York, NY 10010

1047-2797/06/$–see front matter
doi:10.1016/j.annepidem.2005.06.055

Selected Abbreviations and Acronyms

CA	= California
CI	= confidence interval
CMS	= Centers for Medicare and Medicaid Services
DMV	= Department of Motor Vehicles
IBM	= International Business Machine Corporation
MRR	= cancer mortality rate ratios
NDI	= national death index
NHL	= Non-Hodgkin Lymphoma
NY	= New York
RP	= registry period
SES	= socioeconomic status
SIR	= standardized incidence ratio
SIR_U	= SIR in an uncertainty analysis
SIR_M	= SIR in the main analysis
SMR	= standardized mortality ratio
SSA	= Social Security Administration
SSN	= social security number
UA	= uncertainty analysis
US	= United States
VR	= voter registration
YRS	= years worked
YSF	= years since first record of employment

Jose) (8, 9). IBM records and record linkage with several national and state databases provided information on vital status. Cause of death information came from death certificates and from the National Death Index (NDI). Subjects accumulated person-years of follow-up between the later of January 1, 1965 or the employee's facility hire date and the earliest of the date of loss to follow-up, death date, or December 31, 1999. Standardized mortality ratios (SMRs) compared the mortality rates of employees with the rates of the general population of the states where the facilities were located (10).

Subjects in the cancer incidence study were employees who were in the mortality study (8) and who (a) worked at East Fishkill between 1965 and 1999 and lived in NY at any time between 1976 and 1999; or (b) worked at San Jose between 1965 and 1999 and lived in CA at any time between 1988 and 1999 (9). These eligibility requirements were necessary because of procedures used to identify cancer cases, described later.

Employees' work histories included, for each job, a code indicating the state of employment, which we assumed was the state of residence. For retirees, work histories also contained records with address information for each year of retirement.

For employees who had separated without retiring, work histories provided residential history only during active employment. The postemployment residential histories of these employees came from state departments of motor vehicles (DMVs), voter registration records (VRs), and the EZFIND file from LexisNexis, a private vendor of residential information. Record linkage with DMVs and VRs used matching based on name and birth date. Linkage with LexisNexis records used Social Security number (SSN), name, and birth date. DMV and VR records provided one or more addresses and dates of activity (i.e., license issuance, registration). LexisNexis provided current and previous addresses with associated dates. For decedents, LexisNexis often provided only information on the state of residence at death.

For each subject, we compiled data from all sources into a chronological series of addresses and estimated the dates of entry into and exit from NY or CA (the "facility state"). We assumed that separated employees for whom we had no postemployment residential history left the facility state after their last date of employment (11).

We identified cancer cases through record linkage with the NY State and CA cancer registries. We counted a case if the diagnosis date was between the beginning and ending dates of follow-up for the incidence study, was after starting work at the facility, and occurred when the residential history indicated the subject was living in the facility state (9).

Person-year accumulation began on the latest of the cancer registry inception date (NY, January 1, 1976; CA, January 1, 1988) or the subject's facility hire date and ended on the earliest of the study closing date, the last date of residence in NY or CA, the date of loss to follow up, or the death date. Between these beginning and ending dates, subjects accrued person-time only while they lived in the facility state. We computed standardized incidence ratios (SIRs) to compare the cancer incidence rates of employees with rates of the facility state general population (10).

To assess the completeness of each external residential history source, we determined the proportion of subjects having a record and the median number of dated addresses in the record. To evaluate each source's accuracy, we compared states from work histories with those from external sources during periods of active employment or retirement. We evaluated the potential for each source to introduce selection bias by determining if the proportion of subjects having a record in the source differed by race, gender, vital status, age at the end of follow-up, socioeconomic status (SES), employment status, year of first work, year of separation, years worked, and years since first record of employment at the facility. We assigned each subject to one of three SES groups based on salary information for the job in which the subject worked longest (8).

Postemployment residential histories of separated employees had the greatest potential for inaccuracy and required the most effort to produce. We carried out five uncertainty analyses (UAs) to quantify the impact on validity and precision of different assumptions about these residential histories. The main analysis of the incidence study included the postemployment experience of separated workers; UA-A completely excluded this experience.

The main analysis assumed that subjects lived continuously in or outside the facility state from one transition to the next; UA-B expanded postemployment follow-up of separated employees to include every person-year except those specific years in which an address *outside* NY or CA occurred, and UA-C restricted postemployment follow-up of separated employees to those specific years in which an address *in* NY or CA occurred. The main analysis allowed subjects to exit and reenter the facility state multiple times; UA-D restricted follow-up to experience before the first NY or CA exit date. The main analysis did not use death certificates to specify residential histories for decedents; UA-E expanded follow-up to include the entire postemployment experience of all employees who died in NY or CA.

To evaluate the impact on validity of the temporal and geographic restrictions imposed by using cancer registries to identify cases, we partitioned the total mortality study person-time into three categories: (a) included in the cancer incidence study, (b) lost from the cancer incidence study—accrued before the registry period, and (c) lost—accrued during the registry period but while subjects were living outside NY or CA. We compared the distribution of mortality study person-time included in the cancer incidence study with the distribution of lost person-time according to selected demographic and employment characteristics. We used Poisson regression to compute cancer mortality rate ratios (MRRs) that compared rates for lost person-time with rates for included person-time. The MRRs provided an indirect assessment of potential bias, assuming that, if cancer mortality rates for lost person-time were similar to rates for included person-time, cancer incidence rates may also be similar for lost and included person-time. MRRs were adjusted for age, race, gender, SES, years worked, years since first record of employment, and, when possible, calendar time. We examined the variation in the proportion of mortality study person-time included in the cancer incidence study by work group.

To evaluate the impact of restrictions on precision, we compared expected number of cancer cases computed for the lost person-time with the expected number for included person-time. To determine expected numbers, we applied gender, race, age, calendar time, and state-specific cancer incidence rates to the corresponding distributions of person-time. To compute expected numbers for the time period before the state registries began, we used the earliest available rates (1976–1979 for NY and 1988–1989 for CA). Finally, to evaluate the informational gain from inclusion of nonfatal and fatal incident cases, we compared the observed number of cancer cases in the incidence study to the observed number of cancer deaths in the mortality study.

RESULTS

Assessment of Information Sources

Of the 99,229 employees in the mortality study, 96% had LexisNexis records (Table 1). LexisNexis records contained a median of six dated addresses for each subject. The proportion of subjects with DMV records was 47% overall, 60% for San Jose, and 32% for East Fishkill. DMV records contained a median of one dated address. Thirty-four percent of subjects had VR records, containing a median of one dated address. Only 388 subjects (< 1% of all subjects) had DMV or VR records and lacked LexisNexis records. The average proportion of a subject's postemployment address records from each source was 77% LexisNexis, 10% DMV, and 13% VR.

Agreement with states listed in work histories during periods of active employment or retirement was 90% for LexisNexis states, 92% for DMV states, and 99% for VR states. Agreement was similar for East Fishkill and San Jose for LexisNexis and VR data but varied by facility for DMV data (83% for East Fishkill, 98% for San Jose subjects).

The proportion of subjects with LexisNexis records varied little by demographic and employment characteristics (Table 2). Having a LexisNexis record was more common among subjects who were alive (97%) at the end of follow-up for the mortality study than among subjects who were deceased (94%) or who had unknown vital status (80%). The proportion of subjects with a DMV record and the proportion with a VR record also were higher for subjects who were alive than for subjects who had unknown vital status or were deceased. In addition, DMV records were more likely to be available for Hispanics and Asians, for subjects under age 60, for production workers, for subjects who separated in the 1990s, and for subjects with 5+ years of employment. VR records were most likely to be available for subjects ages 40–59, for professionals and technicians, for active employees, for subjects with 5+ years of employment, and for subjects with 15+ years since first record of employment.

Uncertainty Analyses

The alternative assumptions used to specify residential histories for UAs resulted in modest to large changes in the numbers of person-years and cases as compared to the main analysis (Table 3). However, differences between the UAs and the main analysis with regard to SIRs for all types of cancer combined were small. The three analyses that restricted follow-up (Table 3: UA-A, C, and D) included fewer person-years and cases and produced slightly higher SIRs for most types of cancer. The postemployment person-time accrued by separated employees constituted 52% of the

TABLE 1. Number of subjects with records from LexisNexis, departments of motor vehicles (DMV), or voter registration (VR) records and time coverage provided by each source of residential history

	LexisNexis		DMV		VR	
	N	%	N	%	N	%
Any record						
Yes	95,432	96	47,029	47	33,702	34
No	3,797	4	52,200	53	65,527	66
Total	99,229	100	99,229	100	99,229	100
Number of dates or dated addresses[*]						
0	2,873	3	103	0	0	0
1–4	30,447	32	45,615	97	25,319	75
5–9	54,956	58	1,276	3	6,764	20
10+	7,156	7	35	0	1,619	5
Total	95,432	100	47,029	100	33,702	100
Median (Range)	6 (0–16)		1 (0–15)		1 (1–25)	
Number of calendar years in which dates occur[†]						
1	5,927	6	35,219	75	20,708	61
2	11,635	13	8,624	18	1,842	5
3+	74,997	81	3,080	7	11,152	33
Total	92,559	100	46,926	100	33,702	100
Median (Range)	4 (1–10)		1 (1–9)		1 (1–15)	
Time span of dates (years)[†]						
< 5	24,790	27	40,280	86	22,041	65
5–9	28,045	30	6,149	13	4,904	15
10–14	30,593	33	421	1	2,806	8
15+	9,131	10	76	0	3,951	12
Total	92,559	100	46,926	100	33,702	100
Median (Range)	8 (1–37)		1 (1–28)		1 (1–37)	
Median year of dates[†]						
< 1990	3,016	3	743	2	4,552	14
1990–1994	21,518	23	9,444	20	4,001	12
1995–1999	49,674	54	23,726	51	17,996	53
2000+	18,351	20	13,013	28	7,153	21
Total	92,559	100	46,926	100	33,702	100
Median (Range)	1997 (1969–2001)		1998 (1973–2002)		1998 (1953–2003)	

[*] Among those with records.
[†] Among those with 1 or more dates.

total follow-up at San Jose. In the UA that completely excluded this experience, the SIR (SIR$_{\mathrm{U}}$) for all cancers combined was 6% greater than the SIR in the main analysis (SIR$_{\mathrm{M}}$) (Table 3: A). At East Fishkill, where the postemployment person-time of separated employees constituted 28% of the total follow up, the same UA produced an SIR$_{\mathrm{U}}$ for all cancers combined that was only 1% greater than the SIR$_{\mathrm{M}}$. Of the analyses that expanded postemployment follow-up of separated employees, the analysis having the greatest impact on SIRs assumed that employees who died in the facility state spent their entire postemployment residential history in that state (Table 3: E). In this analysis SIRs were higher than those in the main analysis for many types of cancer.

Relative Informativeness of Cancer Incidence and Mortality Studies

Of the 99,229 mortality study subjects from the two facilities, 89,054 (90%) were eligible for the main analysis of the cancer incidence study. At East Fishkill, the cancer incidence study included 42,612 (94%) of the subjects and 61% of the person-time of the mortality study (Table 4). The person-time distribution of the mortality and cancer incidence studies were similar with regard to median values of calendar year, age, years worked, and years since first record of employment at the facility. Compared to the total mortality study, the cancer incidence study had a lower proportion of person-years among men and in the highest SES group. Most of the mortality study person-time lost from

TABLE 2. Number of subjects with records from LexisNexis, department of motor vehicles (DMV), or voter registration (VR) by selected demographic and employment characteristics and percent of total subjects in each category

	Total	LexisNexis		DMV		VR	
Demographic & employment characteristics	N	N	%	N	%	N	%
Total	99,229	95,432	96	47,026	47	33,702	34
Gender/race or ethnicity							
Men, total	65,125	62,498	96	30,599	47	21,841	34
White	45,153	43,284	96	19,235	43	15,862	35
Hispanic	4,489	4,313	96	2,684	60	1,455	32
Asian	10,715	10,318	96	6,411	60	3,240	30
African American	4,512	4,343	96	2,156	48	1,217	27
American Indian	172	163	95	71	41	43	25
Unknown	84	77	92	42	50	24	29
Women, total	34,104	32,934	97	16,427	48	11,861	35
White	20,130	19,407	96	8,777	44	7,515	37
Hispanic	4,010	3,894	97	2,486	62	1,275	32
Asian	5,480	5,285	96	3,160	58	1,643	30
African American	4,293	4,161	97	1,917	45	1,369	32
American Indian	138	135	98	71	51	44	32
Unknown	53	52	98	16	30	15	28
Vital status*							
Alive	89,388	86,791	97	44,995	50	33,156	37
Deceased	5,379	5,074	94	707	13	33	1
Unknown	4,462	3,567	80	1,324	30	513	11
Age*							
< 40	36,810	35,001	95	17,645	48	10,495	29
40–49	23,629	22,802	97	11,985	51	9,136	39
50–59	20,637	19,978	97	9,706	47	7,822	38
60+	18,153	17,651	97	7,690	42	6,249	34
SES group							
Professionals	38,045	36,456	96	17,257	45	13,800	36
Technicians	9,806	9,486	97	4,084	42	3,544	36
Production	51,378	49,490	96	25,685	50	16,358	32
IBM employment status*							
Active	20,367	19,505	96	9,621	47	8,991	44
Retired	19,305	18,919	98	8,412	44	7,045	36
Separated	59,557	57,008	96	28,993	49	17,666	30
Year first at facility							
< 1965–1969	19,874	19,118	96	7,806	39	6,465	33
1970–1979	18,392	17,722	96	8,418	46	6,988	38
1980–1989	32,397	31,168	96	16,006	49	12,104	37
1990–1999	28,566	27,424	96	14,796	52	8,145	29
Median	1984	1984		1984		1982	
Separation date							
1965–1979	11,201	10,469	93	3,882	35	2,814	25
1980–1989	23,722	22,680	96	10,021	42	7,319	31
1990–1999	64,306	62,283	97	33,123	52	23,569	37
Median	1993	1993		1993		1993	
Years worked at facility*							
< 1	35,730	33,910	95	16,282	46	10,033	28
1– < 5	29,361	28,261	96	13,074	45	8,610	29
5+	34,138	33,261	97	17,670	52	15,059	44
Median	2	2		2		4	
Years since first record of employment at facility*							
< 15	46,984	44,569	95	22,505	48	13,498	29
15+	52,245	50,863	97	24,521	47	20,204	39
Median	16	16		15		17	

*At the end of the mortality study.

TABLE 3. Number of person-years (PY) and cases included in each of several uncertainty analyses, the standardized incidence ratio (SIR) and 95% confidence interval (CI) for all types of cancer combined, and a summary of the changes in the uncertainty analysis SIR (SIR_U) compared to the main analysis SIR (SIR_M) for 26 specific types of cancer, by facility

				All types of cancer combined					Specific types of cancer with $SIR_U > SIR_M$		Specific types of cancer with $SIR_U \leq SIR_M$	
Facility, Analysis*	PY	Change in PY	Change in PY (%)	Cases	Change in cases	Change in cases (%)	SIR, 95% CI	Change in SIR (%)	Number of types	(+) Change in SIR (% Range)	Number of types	(−) Change in SIR (% Range)
Panel 1. East Fishkill & San Jose												
Main	861,520			2,860			84, 81–87					
A.	531,903	− 329,618	− 38%	2,233	− 627	− 22%	86, 83–90	+ 3%	16	2–21%	8	1–13%
B.	940,092	+ 78,571	+ 9%	2,923	+ 63	+ 2%	82, 79–85	− 2%	6	1–4%	18	1–11%
C.	644,910	− 216,611	− 25%	2,443	− 417	− 15%	85, 81–88	+ 1%	16	1–10%	8	0–9%
D.	832,744	− 28,777	− 3%	2,759	− 101	− 4%	84, 81–87	+ 1%	14	0–5%	10	0–7%
E.	863,054	+ 1,533	+ < 1%	2,927	+ 67	+ 2%	85, 82–89	+ 2%	15	1–10%	9	< 1%
Panel 2. East Fishkill												
Main	499,445			1,541			81, 77–85					
A.	359,560	− 139,885	− 28%	1,309	− 232	− 15%	82, 78–87	+ 1%	14	1–20%	10	0–22%
B.	566,377	+ 66,932	+ 13%	1,594	+ 53	+ 3%	79, 75–83	− 2%	6	1–6%	18	1–14%
C.	406,686	− 92,760	− 19%	1,393	− 148	− 10%	81, 77–86	+ 0%	11	1–11%	13	0–24%
D.	481,193	− 18,252	− 4%	1,490	− 51	− 3%	82, 78–86	+ 1%	16	0–5%	8	0–8%
E.	500,748	+ 1,303	+ < 1%	1,593	+ 52	+ 3%	83, 79–88	+ 3%	14	0–24%	10	0–1%
Panel 3. San Jose												
Main	362,076			1,319			87, 82–92					
A.	172,344	− 189,732	− 52%	924	− 395	− 30%	92, 86–98	+ 6%	20	1–36%	4	3–15%
B.	373,714	+ 11,638	+ 3%	1,329	+ 10	+ 1%	86, 82–91	− 1%	7	0–10%	17	1–4%
C.	238,225	− 123,851	− 34%	1,050	− 269	− 20%	90, 84–95	+ 3%	16	1–24%	8	0–16%
D.	351,551	− 10,525	− 3%	1,269	− 50	− 4%	87, 82–92	+ 0%	13	0–4%	11	0–7%
E.	362,306	+ 230	+ < 1%	1,334	+ 15	+ 1%	88, 83–93	+ 1%	8	1–11%	16	< 1%

*Procedure:
A: Completely excluded the postemployment experience of separated employees.
B: Expanded follow up to include each year of postemployment experience for separated employees unless they were known to have been living outside the facility state during a specific year.
C: Restricted postemployment followup of separated employees to years when they were specifically known to have been living in the facility state.
D: Restricted follow up to experience prior to employees' first facility state exit date.
E: Assumed that employees who died in the facility state spent their entire postemployment residential history living in the facility state.

the cancer incidence study occurred during the operational period of the cancer registry but while subjects were living outside NY. If incident cancer cases had been detectable for all mortality study person-time, 3005 cases would have been expected, 59% more than the expected number of cases (N =1892) in the cancer incidence study. Comparison of cancer mortality rates pertaining to person-time lost from the incidence study with rates for the included person-time yielded cancer MRRs at East Fishkill of 0.7 (0.5–1.0) for person-time that occurred before the registry period and 0.9 (0.8–1.1) for person-time that occurred during the registry period while subjects were living outside NY (Table 4).

At San Jose, the cancer incidence study included 46,912 (86%) of the subjects and 43% of the person-time of the mortality study (Table 4). The person-time distribution of the mortality and cancer incidence studies were similar with regard to years worked and years since first record of employment but differed with regard to median values of calendar year (1988 vs. 1994) and age (39 vs. 42). Compared to the mortality study, the cancer incidence study had a higher median value of calendar year and of age and a lower proportion among whites and men and in the highest SES group. Most of the lost person-time accrued before the registry period. If incident cancer cases had been detectable for all mortality study person-time, 2925 cases would have been expected, 76% more than the expected number of cases (N=1,496) computed for the cancer incidence study. With the included person-time as the referent, the MRR was 1.1 (0.9–1.3) for person-time that occurred before the registry period and 1.0 (0.9–1.2) for person-time that occurred during the registry period while subjects were living outside CA.

TABLE 4. Mortality study person-years included in and lost from the cancer incidence study because of temporal and geographic restrictions imposed by using cancer registries to identify incident cases, by facility

		Lost			
Facility, demographic & employment characteristics*	Included in incidence	During RP†, outside state	Before RP†	Total	Total in mortality
Panel 1. East Fishkill					
Person-years	496,049	212,794	105,118	317,911	813,961
Year	1989	1992	1971	1984	1988
Age	39	41	33	37	38
YRS	4	2	2	2	3
YSF	10	14	3	8	9
White	85%	82%	89%	85%	85%
Male	67%	74%	82%	76%	71%
SES 1	38%	47%	51%	48%	42%
Cancer cases:					
Observed	1,541				
Expected	1,892	964	149	1,113	3,005
SIR	81				
Cancer deaths:					
Observed	630	283	35	318	948
Expected	771	388	67	454	1,226
SMR (95% CI)	82 (75–88)	73 (65–82)	52 (37–73)	70 (63–78)	77 (73–82)
MRR (95% CI)	1.0 (referent)	0.9 (0.8–1.1)	0.7 (0.5–1.0)		
Panel 2. San Jose					
Person-years	354,097	102,222	360,576	462,798	816,895
Year	1994	1994	1980	1982	1988
Age	42	46	36	38	39
YRS	2	2	2	2	2
YSF	11	16	5	7	9
White	56%	78%	79%	79%	69%
Male	64%	73%	74%	74%	70%
SES 1	39%	53%	50%	51%	46%
Cancer cases:					
Observed	1,319				
Expected	1,496	553	876	1,429	2,925
SIR	88				
Cancer deaths:					
Observed	414	167	249	416	830
Expected	539	205	327	531	1,070
SMR (95% CI)	77 (7085)	82 (70–95)	76 (67–86)	78 (71–86)	78 (72–83)
MRR (95% CI)	1.0 (referent)	1.0 (0.9–1.2)	1.1 (0.9–1.3)		

*Number of person-years; median values and percentages for selected demographic and employment characteristics; observed and/or expected cancer cases, standardized incidence ratio (SIR); observed and expected cancer deaths, standardized mortality ratio (SMR); and cancer mortality rate ratio (MRR) comparing lost person-years to included person-years, adjusted for age, years worked (YRS), years since first record of employment (YSF), race, gender, socioeconomic status (SES), and, when possible, calendar year.
†Registry period.

At East Fishkill, the proportion of mortality study person-years included in the cancer incidence study varied by work group from 65% to 77% (Table 5). At San Jose, the proportion of person-years included varied by work group from 40% to 66%.

Comparison of observed numbers of cases in the incidence study with deaths in the mortality study indicated that the number of cases was less than or equal to the number of decedents for types of cancer having poor survival, including cancers of the lung (322 cases vs. 463 decedents), ovary (34 vs. 35), and central nervous system (55 vs. 82) (Table 6). Incident cases outnumbered decedents for types of cancer having good survival, including cancers of the breast (338 vs. 112), prostate (611 vs. 101), bladder (154 vs. 29), and thyroid (44 vs. 9).

DISCUSSION

Investigators in the US have access to several information sources useful for developing residential histories. A source with national coverage, such as LexisNexis, should provide

TABLE 5. Number of person-years in the mortality and cancer incidence studies and the proportion of mortality study person-years included in the cancer incidence study, by facility and work group

	Person-years		
Facility & work group*	A. Mortality study	B. Cancer incidence study	% (B/A)
Panel 1. East Fishkill			
Semiconductor fab.	327,754	215,366	66
Masking	23,222	17,973	77
Packaging	173,830	125,387	72
Facilities/labs	87,613	57,724	66
Res. & dev.	78,541	51,039	65
Process equipment maintenance	64,096	42,752	67
Test/dice/probe[†]	111,868	73,350	66
Other manufacturing	65,453	47,340	72
Panel 2. San Jose			
Head fabrication	86,206	54,731	63
Disk manufacturing	70,870	40,869	58
Head wafer/tape head	52,505	34,692	66
Facilities/labs	61,013	26,972	44
Res. & dev.	54,042	24,374	45
Test/probe/dicing/slicing/die removal/wire bonding	75,783	36,027	48
Head suspension/head disk/ assembly/box	166,561	96,470	58
Other manufacturing	104,746	42,249	40
Assembly	134,152	63,432	47

*Work groups are not mutually exclusive.
[†]Test/probe/dicing/slicing/die removal/wire bonding.

TABLE 6. Observed number of cases of or deaths from specific types of cancer by facility

	East Fishkill		San Jose		Total	
Type of cancer	Cases	Deaths	Cases	Deaths	Cases	Deaths
All cancers	1,541	948	1,319	830	2,860	1,762
Oral cavity, pharynx	32	11	33	12	65	22
Esophagus	7	13	9	20	16	33
Stomach	18	26	23	28	41	54
Colorectum	184	102	148	96	332	196
Liver	8	21	12	18	20	39
Pancreas	37	65	20	42	57	107
Larynx	14	5	12	4	26	9
Lung	199	253	123	210	322	461
Melanoma of skin	45	32	71	14	116	45
Breast	185	56	162	56	347	111
Cervix	20	6	12	2	32	8
Endometrium*	29	5	17	1	46	3
Ovary	21	16	13	19	34	34
Prostate	277	48	334	53	611	100
Testis	17	3	13	3	30	6
Bladder	99	22	55	7	154	29
Kidney	55	28	29	19	84	45
Central nervous system	34	40	21	42	55	82
Thyroid	19	2	25	7	44	9
NHL[†]	74	59	60	42	134	100
Hodgkin's disease	25	5	9	5	34	10
Leukemia	35	37	37	34	72	70
Multiple myeloma	21	21	13	17	34	38
Other cancer	86	72	68	79	154	151

This table omits soft tissue sarcoma because the mortality study did not consider this type of cancer.
*Includes uterus, not otherwise specified.
[†]NHL, non-Hodgkin's lymphoma.

more complete residential history than state-specific DMVs or VRs, especially for people similar to our subjects who were relatively young, mobile, and of high SES. LexisNexis records contain more information than DMV and VR records, and the accuracy of that information is high and compares favorably with DMV and VR data. The accessibility, record linkage options, and record format also make LexisNexis more useful than DMV or VR data. The effort expended to obtain and process DMV and VR records was not justified in terms of the information yielded.

No residential history sources provided information for every postemployment year, and none provided much information that predated 1990. We are not aware of available data that would provide earlier residential history. Although 14% of subjects with VR records had addresses dated before 1990, VR records were available for a small proportion of our subjects. In any study that does not contact subjects directly, investigators must make assumptions about postemployment residential history that occurred between the last day worked and the earliest address from an external source.

Errors in the residential histories were unavoidable. Most dates of transition into and out of NY or CA were approximate. Inaccuracies in the external information could have occurred because of linkage errors or the external source's delay in recording when subjects moved. Other errors stemmed from assuming that subjects who lacked postemployment addresses did not live in NY or CA after their last date of active employment.

Although our residential histories had limitations, all UAs produced similar results. Alternative approaches that expand or restrict the inclusion of person-time and cases might have a greater impact in a future update or in another study with older subjects and more cancer cases. We would expect a larger difference between the main analysis and the UAs that completely excluded the postemployment experience of separated workers as the proportion of their contribution to total follow-up grows or that used death certificates as a source of residential history as the number of decedents grows.

Among the other investigations that used cancer registries to identify cases and described the development of residential histories (1, 12–20), only a few conducted UAs to evaluate assumptions about postemployment residential history (14, 17, 20). Those studies, like ours, found little impact on results.

Our comparison of cancer mortality rates for person-time included in the cancer incidence study with rates for lost person-time provided minimal evidence of an impact of temporal and geographic restrictions on validity for the overall cohorts from the two facilities. We expected this result at East Fishkill because the cancer incidence follow-up period included most of the mortality study follow-up period. At San Jose, where only 12 years were included in the cancer incidence follow-up, the included person-time and the total mortality study person-time differed with respect to several characteristics, suggesting an effect of the restricted observation period on validity of the incidence results. However, mortality rates for included and lost person-time were similar.

The proportion of mortality study person-years included in the cancer incidence study varied considerably by work group, particularly at San Jose, and the validity of the cancer incidence results also may vary considerably across these subcohorts. A companion paper by the present authors describes in more detail the impact on the incidence study of selection bias due to follow-up restrictions (9). In the latter paper, when we found that the relation between a work group and a particular cancer differed in the mortality and cancer incidence studies, we examined the overlap of subjects counted as deaths versus cases in the studies and determined that most of the differences in the results could be attributable to follow-up restrictions and consequent selection bias in the cancer incidence study (9). Thus, investigators should examine variation in the potential for selection bias due to restricted follow-up across cohort subgroups and should use this information in interpreting the results of cancer incidence studies that rely on cancer registries that do not cover the entire potential follow-up experience of the study group. At present, the only alternative to the reliance on such registries in retrospective studies of cancer incidence is to conduct a survey to identify cancer cases by directly contacting all cohort members. Although methodologically superior, an incidence survey is often not feasible because of (1) difficulties locating cohort members or surrogates and recruiting their participation, and (2) obtaining medical records for case confirmation. Thus, this approach has rarely been used (3–5).

Although cancer registration is being implemented in all states, most registries are of relatively recent origin. The methodologic issues discussed in this paper will persist as challenges for any study of cancer incidence with cohorts established before the inception date of one or more registries. Other issues may create conditions that challenge the feasibility of research dependent on record linkage with multiple registries. Because confidentiality laws vary by state, procedures for obtaining access to cancer registry records and conducting record linkage are not standardized. After record linkage is completed, some states may require the removal of information identifying individuals from the analytic file.

The number of cancer cases substantially exceeded the number of cancer deaths at both facilities. Although the follow-up period for the cancer incidence study at San Jose was quite limited compared to that for the mortality study, much of the person-time that was lost from the cancer incidence study was accrued before 1988 when most subjects were young and when there were few cancer deaths or expected cases. The enhanced precision of the incidence study compared to the mortality study depended in part on developing postemployment residential histories, and the UA that eliminated the postemployment experience of separated employees discarded hundreds of cases and thousands of person-years.

Research in this country and abroad consistently indicates that, for cancers for which survival is relatively long, cancer incidence studies are more informative than mortality studies (21–30). In studies with follow-up periods of similar length for mortality and cancer incidence, the number of cases of all types of cancer is 50%–100% greater than the number of cancer deaths (23–30). Even in studies with cancer incidence follow-up periods that are as much as 20 years shorter in duration than the mortality follow-up periods, cancer cases can exceed the number of cancer deaths (21, 22). Cancer incidence studies are also better in delineating occupational exposure with a certain subtype of cancer or specific histology (e.g., acute myeloid leukemia, B-cell lymphoma), as this information is often recorded by the registry but absent from death certificates (17).

Many of the challenges we described are specific to conducting cancer incidence studies in the US. In Canada and several European countries, relying on record linkage with cancer registries to identify cases does not require temporal restrictions on follow-up because registries have existed for many years (23–30). Furthermore, the national coverage of these foreign registries makes it unnecessary to develop detailed residential histories or to apply geographical restrictions.

In summary, development of optimal residential histories for cancer incidence studies in the US should use information sources with national coverage. Conducting UAs to examine the limitations of the residential histories and the impact of various assumptions is prudent until there is an accepted standard for developing residential histories. The cancer incidence study had much more precision than our mortality results for evaluating cancers associated with relatively long survival. The temporal and geographic restrictions on our cancer incidence study did not appear to affect the validity of the results for the overall analysis, but the potential for selection bias varied considerably by work group subcohort. The proportion of mortality study follow-up included in a cancer incidence study and the overlap of individuals counted as cancer deaths versus cases should be evaluated critically when interpreting the results for a cancer incidence study. The impact of cancer incidence follow-up restrictions may vary among studies, depending on the age distribution and mobility of the study groups and the relative survival of specific cancers of interest.

This research was funded by International Business Machines Corporation. Thomas J. Bender also received support from the Medical Scientist Training Program at the University of Alabama at Birmingham.

REFERENCES

1. Demers PA, Vaughan TL, Checkoway H, Weiss NS, Heyer NJ, Rosenstock L. Cancer identification using a tumor registry versus death certificates in occupational cohort studies in the United States. Am J Epidemiol. 1992;136:1232–1240.
2. Veys CA. Towards causal inference in occupational cancer epidemiology—II.: Getting the count right. Ann Occup Hyg. 1993;37:181–189.
3. Beall C, Delzell E, Rodu B, Sathiakumar N, Myers S. Cancer and benign tumor incidence among employees in a polymers research complex. J Occup Environ Med. 2001;43(10):914–924.
4. Steenland K, Whelan E, Deddens J, Stayner L, Ward E. Ethylene oxide and breast cancer incidence in a cohort study of 7576 women (United States). Cancer Causes Control. 2003;14:531–539.
5. Ward E, Carpenter A, Markowitz S, Roberts D, Halperin W. Excess number of bladder cancers in workers exposed to ortho-toluidine and aniline. J Natl Cancer Inst. 1991;83:501–506.
6. Howe HL. Population-based Cancer Registries in the United States. In: Howe HL, ed. Cancer Incidence in North America, 1988–1990. Springfield, IL: North American Association of Central Cancer Registries; 1994:VI-1–VI-10.
7. Wingo PA, Jamison PM, Hiatt RA, Weir HK, Gargiullo PM, Hutton M, et al. Building the infrastructure for nationwide cancer surveillance and control—A comparison between the National Program of Cancer Registries (NPCR) and the Surveillance, Epidemiology, and End Results (SEER) Program (United States). Cancer Causes Control. 2003;14(2): 175–193.
8. Beall C, Bender TJ, Cheng H, Herrick RF, Kahn AR, Matthews R, et al. Mortality among semiconductor and storage device manufacturing workers In Press. J Occup Environ Med. 2005.
9. Bender TJ, Beall C, Cheng H, Herrick RF, Kahn AR, Matthews R, et al. Cancer incidence among semiconductor and storage device manufacturing workers. Submitted for publication 2005.
10. Marsh GM, Youk AO, Stone RA, Sefcik S, Alcorn C. OCMAP-PLUS: A program for the comprehensive analysis of occupational cohort data. J Occup Environ Med. 1998;40(4):351–362.
11. Breslow NE, Day NE. Statistical methods in cancer research. Volume II—The design and analysis of cohort studies. IARC Sci Publ. 1987;82:49.
12. Meigs JW, Marrett LD, Ulrich FU, Flannery JT. Bladder tumor incidence among workers exposed to benzidine: A thirty-year follow-up. J Natl Cancer Inst. 1986;76(1):1–8.
13. Bond GG, Austin DF, Gondek MR, Chiang M, Cook RR. Use of a population-based tumor registry to estimate cancer incidence among a cohort of chemical workers. J Occup Med. 1988;30:443–448.
14. Leet T, Acquavella J, Lynch C, Anne M, Weiss NS, Vaughan T, et al. Cancer incidence among alachlor manufacturing workers. Am J Ind Med. 1996;30(3):300–306.
15. Acquavella JF, Riordan SG, Anne M, Lynch CF, Collins JJ, Ireland BK, et al. Evaluation of mortality and cancer incidence among alachlor manufacturing workers. Environ Health Perspect. 1996;104(7):728–733.
16. Olsen GW, Gondek MR, Flannery J, Cartmill JB, Bodner KM. A thirty-eight year comparison of cancer incidence and mortality among employees at a Connecticut chemical plant. Conn Med. 1997;61(2):83–89.
17. Huebner WW, Chen VW, Friedlander BR, Wu XC, Jorgensen G, Bhojani FA, et al. Incidence of lymphohaematopoietic malignancies in a petrochemical industry cohort: 1983-94 follow up. Occup Environ Med. 2000;57:605–614.
18. MacLennan PA, Delzell E, Sathiakumar N, Myers SL, Cheng H, Grizzle W, et al. Cancer incidence among triazine herbicide manufacturing workers. J Occup Environ Med. 2002;44:1048–1058.
19. Dement J, Pompeii L, Lipkus IM, Samsa GP. Cancer incidence among union carpenters in New Jersey. J Occup Environ Med. 2003;45(10):1059–1067.
20. Tsai SP, Chen VW, Fox EE, Wendt JK, Cheng Wu X, Foster DE, et al. Cancer incidence among refinery and petrochemical employees in Louisiana, 1983–1999. Ann Epidemiol. 2004;14:722–730.
21. Spinelli JJ, Band PR, Svirchev LM, Gallagher RP. Mortality and cancer incidence in aluminum reduction plant workers. J Occup Med. 1991;33:1150–1155.
22. Ott MG, Zober A. Cause specific mortality and cancer incidence among employees exposed to 2,3,7,8-TCDD after a 1953 reactor accident. Occup Environ Med. 1996;53:606–612.
23. Lundberg I, Milatou-Smith R. Mortality and cancer incidence among Swedish paint industry workers with long-term exposure to organic solvents. Scand J Work Environ Health. 1998;24:270–275.
24. Littorin M, Attewell R, Skerfving S, Horstmann V, Moller T. Mortality and tumour morbidity among Swedish market gardeners and orchardists. Int Arch Occup Environ Health. 1993;65:163–169.
25. Evanoff BA, Gustavsson P, Hogstedt C. Mortality and incidence of cancer in a cohort of Swedish chimney sweeps: an extended follow up study. Br J Ind Med. 1993;50:450–459.
26. Bulbulyan MA, Margaryan AG, Ilychova SA, Astashevsky SV, Uloyan SM, Cogan VY, et al. Cancer incidence and mortality in a cohort of chloroprene workers from Armenia. Int J Cancer. 1999;81:31–33.
27. Svensson BG, Mikoczy Z, Stromberg U, Hagmar L. Mortality and cancer incidence among Swedish fishermen with a high dietary intake of persistent organochlorine compounds. Scand J Work Environ Health. 1995;21:106–115.
28. Jakobsson K, Mikoczy Z, Skerfving S. Deaths and tumours among workers grinding stainless steel: A follow up. Occup Environ Med. 1997;54:825–829.
29. Lewis RJ, Schnatter AR, Drummond I, Murray N, Thompson FS, Katz AM, et al. Mortality and cancer morbidity in a cohort of Canadian petroleum workers. Occup Environ Med. 2003;60:918–928.
30. Gun RT, Pratt NL, Griffith EC, Adams GG, Bisby JA, Robinson KL. Update of a prospective study of mortality and cancer incidence in the Australian petroleum industry. Occup Environ Med. 2004;61: 150–156.