American Journal of Epidemiology

Vol. 170, No. 1
DOI: 10.1093/aje/kwp091
Advance Access publication May 4, 2009

Practice of Epidemiology

Methodological Issues in a Retrospective Cancer Incidence Study

Jeanine M. Buchanich, Ada O. Youk, Gary M. Marsh, Zb Bornemann, Steven E. Lacey, Kathleen J. Kennedy, Roger P. Hancock, Nurtan A. Esmen, and Frank S. Lieberman

Initially submitted October 9, 2008; accepted for publication March 18, 2009.

The authors traced incidence of central nervous system cancer in a large occupational cohort of jet engine manufacturing workers from 1976 to 2004 in the 24 US states that comprised 95% of the cohort deaths. The cohort of approximately 224,000 employees was matched with cancer registry data; all central nervous system cancer matches were requested with their diagnostic data. This paper highlights the obstacles encountered while conducting this retrospective cancer incidence study. The authors spent approximately 700 hours completing applications and obtaining the cohort matches. Approximately 70% of the cases were identified in the state in which the facility of interest is located. In addition to the large amount of time involved, identified issues include complicated approval processes, high costs, temporal differences among the registries, and registry agency difficulty in performing the matching. Several states do not allow individual-level data to be used for research purposes. Researchers can gain important cancer incidence information by matching retrospective cohorts to multiple state cancer registries. However, they should carefully weigh the time and costs required and plan accordingly. Despite some serious obstacles, many of which are potentially resolvable, cancer incidence studies of retrospective cohorts using multiple cancer registries are feasible.

central nervous system neoplasms; cohort studies; data collection; epidemiologic methods; incidence; neoplasms; vital statistics

Abbreviations: CNS, central nervous system; HIPAA, Health Insurance Portability and Accountability Act; IRB, institutional review board; SEER, Surveillance, Epidemiology, and End Results.

Retrospective cohort studies have typically relied upon cause-of-death information to identify the occurrences of cancer among cohort members. Because the United States has the National Death Index, a well-established and thorough repository for death information, tracing vital status and obtaining cause-of-death information is relatively straightforward. However, some researchers have expressed concern about the accuracy and completeness of cancer data available through death certificates (1–4).

To date, most of the cancer incidence studies conducted on occupational cohorts have traced through only one state cancer registry (5–8) or used one registry and self-identified cases (9). However, doing so leads to loss of case ascertainment due to out-migration. One way to avoid these losses in retrospective occupational epidemiology studies is to trace through multiple cancer registries rather than only in the state in which the facility of interest is located. However, attempting to completely ascertain incident cases by tracing through multiple state cancer registries may lead to other methodological issues. We experienced the issues and obstacles discussed here while tracing the incidence of central nervous system (CNS) cancer through 24 state cancer registries of a large occupational cohort of jet engine manufacturing workers in the United States.

MATERIALS AND METHODS

Our cohort was traced for mortality from 1952 to 2004 and for CNS cancer incidence from 1976 to 2004. The analyses are ongoing; to date, total and cause-specific mortality

Correspondence to Dr. Jeanine M. Buchanich, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261 (e-mail: jeanine@pitt.edu).

results have been published (10, 11). Subsequent phases of the study include an analysis by CNS histology and work history factors and a nested case-control study of the CNS cases and matched cohort controls. We are conducting interviews with the cases or next of kin and thus asked the registries for permission to perform followback after obtaining the individual-level case data. A companion exposure assessment piece is under development and will be analyzed in later phases of the study.

To minimize loss of case ascertainment due to outmigration, we identified US states through which we needed to trace the cohort for cancer incidence based upon the states in which 95% of the deaths in the total cohort occurred. We reasoned that the pattern of the incident cancer cases' location of diagnosis would be similar to the places of death identified in the mortality tracing. Twenty-four states comprised 95% of the deaths in our cohort.

To enact our protocol, we applied for permission to match our cohort of approximately 224,000 employees to the 24 cancer registries. We began applying to the state cancer registries in January 2003 and, in December 2008, submitted our latest revisions based on reviewer comments. We requested all CNS cancer matches with their diagnostic data, including date of diagnosis, histology, behavior, topography, residence at diagnosis, and last known vital status.

RESULTS

During the course of our cancer incidence substudy, we compiled detailed information on every state's cancer registry (Table 1), including the years in which cancer incidence reporting was mandated, complete state ascertainment began, and nonmalignant (benign) tumor reporting began. We determined whether individual-level data could be obtained from the registry and whether the registry contributes data to the Surveillance, Epidemiology, and End Results (SEER) Program.

We focused our tracing efforts on the 24 state cancer registries shown in Table 2. Table 2 also indicates the amount of time investigators spent securing the approval and matches, the costs involved, and the amount of time that elapsed between submitting our initial application and receiving approval.

As our tracing efforts proceeded, we recognized that there were many methodological obstacles to overcome. The 2 primary types are potentially resolvable and irresolvable.

Potentially resolvable obstacles in state cancer registry tracing

Registry application process. Most states have rigorous procedures in place to obtain institutional review board (IRB) or a similar level of approval. This approval is in addition to that required by the researchers' institution or university.

The approval process is different in each state (Table 1). In some (Connecticut, Maryland, Maine, New York, South Carolina, and Tennessee), the cancer registry is the agency to which investigators first apply, and the registry is responsible for getting the application through the IRB. In other states (Vermont and Florida), the cancer registry and the IRB are completely separate, each with its own applications. In Rhode Island and Virginia, the application goes directly to the IRB; the cancer registry has no approval authority at all. Massachusetts is the opposite—the cancer registry approves the application, and there is no IRB involvement.

Applications and submission requirements are not standardized across the registries. Researchers must complete multiple applications requesting similar, but not exactly the same, information; therefore, should one registry's review panel request modifications to the protocol, consent forms, or approach letters, those modifications must be sent through every other registry review board that has already approved the study.

One of the most difficult states from which to obtain approval has been New Jersey. In 2005, when we initially applied, New Jersey required approval from both the cancer registry and an IRB. The IRB application involved completing a certification course on human subjects' research separate from that required by our institution. They have requested multiple changes to our protocol, requiring additional reviews by our IRB, and have modified their application, requiring additional reviews by their boards. In August 2008, we learned that New Jersey has dissolved the IRB and that approval will be handled by the registry. However, to date, we have not received approval in New Jersey.

Time and expense of tracing. Obtaining cancer registry approval and matching is an extremely time-consuming process. Table 2 characterizes, for each state registry contacted for our study, the amount of time spent actively completing the application and the costs involved. We have spent approximately 400 hours working on state IRB and cancer registry applications, which ranged from a very reasonable 2 hours for several states to 120 hours to date on New Jersey alone. New Jersey and Washington combined accounted for 30% of the time actively spent on applications and have required 17 submissions between them.

We also spent about 2 hours per registry obtaining the application and information necessary to complete it. Additionally, the investigators spent at least 1 hour each week, while tracing was ongoing (5 years), discussing progress and issues. In total, we have spent about 700 hours over the past 5 years obtaining the approvals and cohort matches from these 24 registries. The fastest we received approval for tracing was 1 month for Arizona; the longest has been New Jersey ($\geq$36 months to date).

Table 2 also indicates the cost of tracing our cohort through the state cancer registries. Thirteen of the states completed the IRB approval and cohort matching without charging fees. Some states, such as Florida and California, had a very high tracing cost ($8,300 and $8,140, respectively) because of our large cohort, but we pursued the linkage because they were 2 of the most important states based on place of death (Florida was ranked second; California was ranked fifth). We also paid for approval and matching in New York ($4,500), another important study state. We did not pursue tracing in 3 of our eligible states (North Carolina, Nevada, and Alabama, ranked 11th, 23rd, and 24th, respectively) because of the extremely high cost of IRB approval and tracing for the extremely low return for our study. In

Table 1. Availability of Cancer Incidence Data for Each US State as of 2008

State	Mandated Reporting Started	Complete Ascertainment Started	Nonmalignant Tumor Reporting Started	Individual-level Data Available for Research?	Unique Requirements	Followback Allowed?	SEER Site?
Alabama	1996	1996	2004	Yes	IRB and state health council approvals are required.	Yes	No
Alaska[a]	1996	1996	2004	Unknown	None	No	Yes
Arizona	1992	1995	1992	Yes	None	Yes	Yes
Arkansas	1996	1997	1996	Yes	None	Yes	No
California	1947	1988	2001	Yes	None	Yes	Yes
Colorado	1968	1988	1988	Yes	None	Yes	No
Connecticut	1935	1973	1962	Yes	None	Yes	Yes
Delaware[b]	1980	1980	1980	No	None	No	No
District of Columbia	1985	1998	2004	Yes	None	Yes	No
Florida	1981	1997	2004	Yes	State and cancer registry IRBs are required.	Yes	No
Georgia[c]	1995	1995	2004	Yes	Cancer registry must approve the application before submitting it to the state IRB.	Yes	Atlanta + rural
Hawaii	1960	1960	2004	Yes	None	Yes	Yes
Idaho	1969	1971	1971	Yes	None	Yes	No
Illinois	1985	1985	2004	Yes	None	Yes	No
Indiana	1987	1987	2004	Yes	None	Yes	No
Iowa	1973	1982	2004	Yes	Registry must be contacted before beginning the application process.	Yes	Yes
Kansas	1984	2004	2004	Yes	None	Yes	No
Kentucky	1991	1994	2004	Yes	None	Yes	Yes
Louisiana	1974	1988	2004	No	None	No	Yes
Maine	1983	1995	2004	Yes	None	Yes	No
Maryland	1991	1996	2001	Yes	Agreement must be reached between several organizations, which takes months.	Yes	No
Massachusetts	1982	1982	1982	Yes	None	Yes	No
Michigan	1985	1985	2004	Yes	None	Yes	Detroit
Minnesota	1988	1988	1988	Yes	Proposals must be reviewed by peer reviewers first.	Yes	No
Mississippi	1996	1996	2004	Yes	None	Yes	No
Missouri[d]	1985	1999	2004	Yes	University of Missouri-Columbia and state IRB approval are required.	Yes	No
Montana	1981	1997	1997	Yes	None	Yes	No
Nebraska	1987	1987	2004	No	None	No	No
Nevada	1979	1979	2004	Yes	Active written consent is required.	Yes	No

Table continues

total, we spent $22,440 to obtain the approvals and matches from the 21 registries that we chose to pursue.

Data confidentiality. Cancer registries include personal medical information and therefore have very strict privacy and confidentiality rules. State regulations typically include limits on the use and disclosure of the cancer information reported to registries and may be quite restrictive about what information can be shared with researchers. There also seems to be some confusion among registrars regarding applicability of the Privacy Rule of the Health Insurance Portability and Accountability Act (HIPAA) to their data. Per the North American Association of Central Cancer Registries, HIPAA does not apply to cancer registries because they are not covered entities (12). However, several states invoked HIPAA regulations to restrict the data available for our study, even though HIPAA does not apply.

Table 1. Continued

State	Mandated Reporting Started	Complete Ascertainment Started	Nonmalignant Tumor Reporting Started	Individual-level Data Available for Research?	Unique Requirements	Followback Allowed?	SEER Site?
New Hampshire	1986	1986	2004	Yes	State IRB approval is required.	Yes	No
New Jersey	1979	1978	2004	Yes	Registry must be contacted before beginning the application/IRB process.	Yes	Yes
New Mexico	1966	1969	1996	No	IRB does not want identifiable data leaving the registry.	No	Yes
New York	1940	1976	1976	Yes	Confidential data can be released in government or government-sponsored studies only.	Yes	No
North Carolina	1990	1999	2004	Yes	Data will be released only if patients will not be contacted (no followback).	No	No
North Dakota	1997	1997	2004	Yes	None	Yes	No
Ohio	1992	1992	2004	Yes	None	Yes	No
Oklahoma	1997	1997	2004	Yes	None	Yes	No
Oregon	1996	1996	2004	Yes	None	Yes	No
Pennsylvania	1982	1985	2004	Yes	None	Yes	No
Rhode Island	1986	1986	1997	Yes	Confidentiality requests must start with the state IRB.	Yes	No
South Carolina	1996	1996	2004	Yes	None	Yes	No
South Dakota	2005	2005	2005	No	Has no policy for individual data release.	No	No
Tennessee	1983	1983	2004	Yes	None	Yes	No
Texas	1986	1995	2004	Yes	None	Yes	No
Utah	1966	1973	2004	Yes	Must obtain approval from an internal advisory committee.	Yes	Yes
Vermont	1994	1995	2004	Yes	Patient consent is required for data release.	Yes	No
Virginia[e]	1990	1998	1990	Yes	Data will be released only if patients will not be contacted (no followback).	Yes	No
Washington	1992	1992	1992	Yes	Data will be released only if patients will not be contacted (no followback).	No	Seattle
West Virginia	1993	1993	2002	No	Deidentified data will be released to government agencies only.	No	No
Wisconsin	1976	1978	2004	No	Deidentified data are released. Identified data are for use only within the Wisconsin Department of Health.	No	No
Wyoming	1962	1977	2004	Yes	None	Yes	No

Abbreviations: IRB, institutional review board; SEER, Surveillance, Epidemiology, and End Results.
[a] No one has ever requested individual-level data. The state does release deidentified data.
[b] The state is working on regulations to allow individual data to be released, but currently it does not.
[c] Atlanta started in 1975, selected rural counties were added in 1978, statewide reporting started in 1995.
[d] Reporting was voluntary from 1972 to 1984. No data were released prior to 1996.
[e] Reporting of nonmalignant brain tumors was "encouraged" beginning in 1990 and was mandated in 2004.

Registries typically have different regulations in place depending on whether a case is alive or deceased, which can be a problem because not all cancer registries regularly trace cases in the registry for vital status. Therefore, the registry may not be aware of cases who have died since the last trace and may unknowingly withhold data from investigators. Many registries will release data on deceased cases directly to researchers because their medical information is no longer protected.

Regulations often require that living cases give consent before their data are released, a very time-consuming and unwieldy task. It often means that the registry must contact the physician who treated the case at the time of diagnosis, who then contacts the patient with the request. Registries

Table 2. Time[a] and Cost Associated With Matching the Jet Engine Manufacturing Cohort to 24 US State Cancer Registries

State	Match Performed?	Time Spent by Investigators	Cost	Time From First Submission to First Approval
Alabama	No	2 hours	$5,000 fee for review by IRB and the Alabama state health council; $3,500 for linkage	N/A (declined to pursue because of cost)
Arizona	Yes	2 hours	No fees	1 month
California	Yes	20 hours	$8,300 for linkage	5 months
Connecticut	Yes	5 hours	No fees	2 months
Florida	Yes	20 hours (separate applications to IRB and cancer registry)	$8,140 for linkage	IRB: 5 months (1 resubmission); cancer registry: 4 months
Georgia	Yes	10 hours	No fees	7 months (2 resubmissions)
Maine	Yes	20 hours	No fees	14 months (5 resubmissions)
Maryland	Yes	5 hours	No fees	5 months (1 resubmission)
Massachusetts	Yes	5 hours	No fees	6 months
Michigan	Yes	5 hours	No fees	3 months
Nevada	No	2 hours	~$1,800 for cancer registry to contact cases/next of kin	2 months (declined to pursue match because of cost)
New Hampshire	Yes	5 hours	No fees	6 months (1 resubmission)
New Jersey	No	120 hours	No fees	≥36 months (8 resubmissions, ongoing)
New York	Yes	10 hours	$4,500 for linkage	6 months
North Carolina	No	2 hours	~$1,500 for linkage	N/A (declined to pursue because of cost)
Ohio	Yes	5 hours	No fees	4 months (1 resubmission)
Pennsylvania	Yes	5 hours	No fees	4 months
Rhode Island	Yes	10 hours	No fees	12 months (2 resubmissions)
South Carolina	Yes	5 hours	No fees	6 months (1 resubmission)
Tennessee	Yes	5 hours	No fees	2 months
Texas	Yes	10 hours	No fees	3 months (1 resubmission)
Vermont	Yes	2 hours	No fees	1 month
Virginia	Yes	15 hours (separate applications to IRB and cancer registry)	$1,500 for linkage	IRB: 3 months (1 resubmission); cancer registry: 12 months
Washington	No	110 hours	No fees	≥18 months (9 resubmissions, ongoing)

Abbreviations: IRB, institutional review board; N/A, not applicable.
[a] Total time spent: 400 hours.

may find this step difficult because the physician may have relocated, retired, or died since the time of diagnosis.

For studies requiring followback, the process is complicated further because the data request also includes whether the investigators can contact the case or next of kin about the study (Table 1). Some states, such as Vermont, require patient contact and consent for all deceased cases or for deceased cases for which followback will be conducted. The North Carolina and Washington registries will not release incidence data for studies involving followback at all.

Several states, including Delaware, Louisiana, and West Virginia, will not release individual-level data for any study (Table 1). Alaska has never had a request for individual-level data so was uncertain about the terms of its release.

Case matching. Cancer registries are required to completely and accurately record data (13). Investigators rely on the registry to screen out duplicate incident reports from multiple sources. Researchers must also depend on the registry to match their file with the cohort file and accurately return all relevant cases. Most states did not provide the details of the case matching. However, 3 states sent us detailed information on how the matches were conducted. California used a multipass matching algorithm. Florida used a probabilistic scoring system and then manually checked matches with borderline scores. South Carolina used both SAS software (SAS Institute, Inc., Cary, North Carolina) and the Centers for Disease Control and Prevention linkage program, Link Plus (14). Maryland mentioned that they also used Link Plus but did not provide details.

We identified several problems with registry matching. The Massachusetts registry did not have the computer ability to match their registry file to our entire cohort, so we sent them a reduced cohort for matching (subjects known to have

Table 3. US States Through Which Cancer Registry Tracing Occurred, in Chronological Order by Start, With Relative and Cumulative Relative Frequencies

State	Registry Start Year	Relative Frequency	Cumulative Relative Frequency
Connecticut	1976	0.744	0.744
New York	1976	0.021	0.765
Florida	1981	0.054	0.819
Massachusetts	1982	0.064	0.883
Pennsylvania	1983	0.008	0.891
Tennessee	1984	0.004	0.896
Michigan	1985	0.004	0.900
New Hampshire	1986	0.011	0.911
Rhode Island	1987	0.004	0.915
California	1988	0.014	0.929
Virginia	1990	0.001	0.930
Arizona	1992	0.010	0.940
Maryland	1992	0.006	0.946
Ohio	1992	0.001	0.947
Vermont	1994	0.007	0.954
Maine	1995	0.038	0.992
Texas	1995	0.007	0.999
Georgia	1995	0.000	0.999
South Carolina	1996	0.001	1.0

lived or died in Massachusetts and those whose social security numbers were issued by Massachusetts). Pennsylvania initially matched on first and last name, resulting in some incorrect matches; we provided additional information to the registry to facilitate identification of correct matches. Originally, Maine did not return any CNS case matches, although cases were expected because of their rank regarding place of death (fifth); upon questioning, registry personnel realized that there was an error in the selection process, reran the file, and returned eligible CNS cases to us.

Irresolvable limitations in state cancer registry tracing

Geographic and temporal restrictions. One unavoidable issue with cancer incidence tracing is the geographic and temporal restrictions of the state cancer registries (15). As shown in Table 1, Connecticut has the earliest registry start, 1935, whereas the South Dakota registry did not begin until 2005. Table 1 also indicates that many years can elapse between the registry start and the mandated reporting date for some state registries. The California registry, for example, began in 1947, but statewide reporting was not required until 1988.

One way to overcome this limitation is to restrict cohort incidence analyses to the time periods for which registry data are available. As shown in Table 3, only Connecticut and New York had complete incidence ascertainment in 1976, the year in which our incidence study of jet engine workers began. Thus, we were sure of complete ascertainment during the entire 1976–2004 time period for only Connecticut and New York. This limitation can lead to inflated expected values and artificially reduced standardized incidence ratios by including person-years for cohort members who did not have a chance of being diagnosed in the state during that period because they were living in other states.

Table 3 also shows the relative and cumulative frequencies for the 20 states from which we were able to receive incident cases. Connecticut, the state in which our study facilities are located, provided approximately 74% of the incidence matches. Florida, Massachusetts, and Maine each provided 5%–10% of the total number of cases. Four states (California, New Hampshire, New York, and Vermont) provided 1%–5% of the total number of cases; the other states provided less than 1% each of the cases.

Reporting of nonmalignant tumor cases. Data collection for nonmalignant tumor cases was mandated in 2004, and many states did not collect information on these cases until it was required (Table 1). In some instances, reporting agencies thought that the information on a particular nonmalignant histology was so important that it warranted collection even though it was not mandated. For example, SEER has always coded pilocytic astrocytomas (a nonmalignant type of tumor) as malignant because they occur so frequently in children. After the law mandating nonmalignant tumor reporting was passed, pilocytic astrocytomas were still coded as malignant although data on them were collected and they could legitimately be reported as nonmalignant. Otherwise, the historical perspective for these tumors would be changed so dramatically that it would be detrimental for reporting of childhood brain tumors in general. In our current study, we are fortunate to depend primarily on the Connecticut registry, which has collected data on nonmalignant incident cases since the early 1960s. The lack of information on nonmalignant cases may have precluded such a study in another state.

Obtaining standard population incidence rates. Calculation of expected numbers of cases is complicated by the temporal differences between the registries. SEER can be used to obtain rates that approximate a national comparison, but state or local comparisons may be more difficult. However, even using standardized data such as those from SEER can cause complications. Over time, knowledge about the histology and behavior of cancers has grown, and the way in which some cancers are categorized has changed dramatically, which is especially true for lympho-hematopoietic and CNS cancers. For example, relatively rare CNS cancer results must often be reported in broad categories including many codes (e.g., glioma), but neuropathologists may disagree about which subcodes should be included in the "standard" aggregated groups (16).

Other methods for identifying incident cancer cases. One way in which we attempted to identify cases, especially in states in which we did not or could not trace, was by using CNS cases identified through death certificates. While death certificates have some inherent limitations, as discussed above, we obviated these problems by contacting the next of kin and requesting the release of medical records to confirm the diagnosis histologically. We have identified 4 additional cases to date by using this method, although it is

extremely time consuming and not recommended in most instances.

Obtaining the counts of cases in incidence studies at the grouped-data level is also a possibility. Doing so avoids issues with researchers acquiring personal identifiers and followback; however, it limits the types of analyses that can be conducted. Because investigators are obtaining counts of cases by only age group, race, and sex, for example, individual employees cannot be linked to personal work histories or exposure profiles. This method may be an option for investigators wanting to conduct a broad analysis of multiple types of cancer in a cohort rather than a more detailed investigation of a specific cancer, but it is also complicated by the relatively recent start dates of many state registry programs and the dissimilar years of registry operation across states (Table 1).

DISCUSSION

Occupational cohort studies typically use mortality data to ascertain cancer cases. However, Engel et al. (3) found that 12 of 83 cancer deaths (14%) were not attributed as such on the death certificate. Percy et al. (4) found that the cancer site recorded on the death certificate matched the information found on the SEER record only 86% of the time.

Retrospective cancer incidence studies have matched cases to only one state's registry (7, 8, 17). Had we used that methodology for our incidence tracing, we would have identified only 74% of our incident cases (Table 3) from Connecticut, the state in which our facilities are located. We would have lost 26% of our CNS cases to out-migration, a much higher proportion than that assumed by other investigators (5–8). Our new methodology enabled us to identify many more incident cases than we would have by using a more traditional method.

However, our multistate system revealed other problems, many of which may be potentially resolvable. The most onerous and time-consuming task was obtaining the approval and matches from the registries. It took us approximately 700 hours to have our protocol approved by 22 state cancer registries, primarily because of the different application form and process necessary for each registry. Although some of these applications and approvals were understandable and efficient, the processes in several states were very unwieldy, difficult, and time consuming. New Jersey and Washington, in particular, required more than 100 hours each, and those approvals are still outstanding.

We found that states in which the cancer registry website had application forms and submission directions available were much more user-friendly than those that did not. Clear instructions also greatly facilitated the application process. The use of different applications by the various states led to a great deal of repetition. One relatively quick solution to assist researchers would be for state cancer registries to adopt standardized forms for their applications. While investigators would still have to undergo multiple cancer registry reviews, standardized application forms would ease the time-consuming task of answering slightly reworded questions multiple times.

We also faced obstacles due to limitations of the registries themselves to match to our large cohort and to extremely costly, and in some cases cost-prohibitive, tracing. Our cohort is extremely large, probably larger than most registries had ever matched before; perhaps, understandably, we experienced problems because registries did not have the computing capability to match to such a large cohort. More disturbing was that some registries did not perform the match correctly or supplied us with incorrect matches. It may help registries to adapt a standard linkage program, such as the Centers for Disease Control and Prevention's Link Plus, to ensure the most accurate matching possible.

Some of the problems involved in retrospective cancer incidence tracing are unavoidable. We recognized the limitations inherent in the operating years of the various state cancer registries before we began our tracing. Although difficult, these differences can be quantified in person-year sensitivity analyses for the cohort. The use of detailed residential history can overcome this limitation. Bender et al. (17) developed residential histories for 98% of 99,229 cohort members. They performed an uncertainty analysis examining the impact on standardized incidence ratios of changing residential assumptions and found that, even with the most conservative assumption (a 52% reduction in person-years), the standardized incidence ratio increased by only 6%. Bender et al. concluded that ''despite geographic and temporal restrictions, incidence studies provide more data than mortality studies'' (17, p. 170) but mentioned the need for careful interpretation of incidence study results. In our ongoing study of jet engine workers, we are performing sensitivity analyses of our CNS standardized incidence ratios by comparing the results using the total person-year count with those obtained when person-years are very conservatively stopped at the date of termination.

Other difficulties occurred when defining and obtaining standard population rates. As registries collect data over longer periods of time, some of these unavoidable obstacles will become less critical. Cohort entrance periods will start in the 1980s and 1990s instead of the 1970s or earlier, as in our study. The periods of interest for performing retrospective cancer incidence studies eventually will correspond to the availability of the registry data.

In conclusion, researchers can gain important cancer incidence information by matching retrospective cohorts to multiple state cancer registries. However, they should carefully weigh the time and costs required and plan accordingly. Despite some serious obstacles, many of which are potentially resolvable, cancer incidence studies of retrospective cohorts using multiple cancer registries are feasible.

ACKNOWLEDGMENTS

Author affiliations: Center for Occupational Biostatistics and Epidemiology, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania (Jeanine M. Buchanich, Ada O. Youk, Gary M. Marsh, Zb Bornemann); Division of Environmental

and Occupational Health Sciences, School of Public Health, University of Illinois at Chicago, Chicago, Illinois (Steven E. Lacey, Kathleen J. Kennedy, Roger P. Hancock, Nurtan A. Esmen); and Department of Neuro-Oncology, University of Pittsburgh Medical Center, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania (Frank S. Lieberman).

This work was supported by Pratt & Whitney, East Hartford, Connecticut.

Conflict of interest: none declared.

REFERENCES

1. Moriyama I. Problems in measurement of accuracy of cause-of-death certificates. *Am J Public Health*. 1989;79(10):1349–1350.
2. Percy C, Stanek E, Gloeckler L. Accuracy of death certificates and its effect on cancer mortality statistics. *Am J Public Health*. 1981;71(3):242–250.
3. Engel LW, Strauchen JA, Chiazze L Jr, et al. Accuracy of death certificates in an autopsied population with specific attention to malignant neoplasms and vascular diseases. *Am J Epidemiol*. 1980;111(1):99–112.
4. Percy CL, Miller BA, Gloeckler Ries LA. Effect of changes in cancer classification and the accuracy of cancer death certificates on trends in cancer mortality. *Ann N Y Acad Sci.* 1990; 609:87–97.
5. Bond GG, Austin DF, Gondek MR, et al. Use of a population-based tumor registry to estimate cancer incidence among a cohort of chemical workers. *J Occup Med*. 1988;30(5):443–448.
6. Wasserman SL, Berg JW, Finch JL, et al. Investigation of an occupational cancer cluster using a population-based tumor registry and the National Death Index. *J Occup Med*. 1992; 34(10):1008–1012.
7. Demers PA, Vaughan TL, Checkoway H, et al. Cancer identification using a tumor registry versus death certificates in occupational cohort studies in the United States. *Am J Epidemiol*. 1992;136(10):1232–1240.
8. Zeka A, Eisen EA, Kriebel D, et al. Risk of upper aerodigestive tract cancers in a case-cohort study of autoworkers exposed to metalworking fluids. *Occup Environ Med*. 2004; 61(5):426–461.
9. Beall C, Delzell E, Rodu B, et al. Cancer and benign tumor incidence among employees in a polymers research complex. *J Occup Environ Med*. 2001;43(10):914–924.
10. Marsh G, Buchanich J, Youk A, et al. Long-term health experience of jet engine manufacturing workers: I. Mortality patterns from central nervous system neoplasms. *J Occup Environ Med*. 2008;50(10):1099–1116.
11. Marsh G, Buchanich J, Youk A, et al. Long-term health experience of jet engine manufacturing workers: II. Total and cause-specific mortality excluding central nervous system neoplasms. *J Occup Environ Med*. 2008;50(10): 1117–1129.
12. The North American Association of Central Cancer Registries. *Frequently Asked Questions and Answers About Cancer Reporting and the HIPAA Privacy Rule*. Springfield, IL: NAACCR; 2003. http://apps.nccd.cdc.gov/StateCancerFacts/.)
13. Seiffert J. *Standards for Cancer Registries*. Sacramento, CA: North American Association of Central Cancer Registries; 1997.
14. National Program of Cancer Registries. Link Plus. Atlanta, GA: Department of Health and Human Services, Centers for Disease Control and Prevention. (http://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm).
15. Wingo PA, Jamison PM, Hiatt RA, 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.
16. McCarthy BJ, Surawicz T, Bruner JM, et al. Consensus Conference on Brain Tumor Definition for Registration. November 10, 2000. *Neuro Oncol*. 2002;4(2):134–145.
17. Bender TJ, Beall C, Cheng H, et al. Methodologic issues in follow-up studies of cancer incidence among occupational groups in the United States. *Ann Epidemiol*. 2006;16(3): 170–179.