

The Collection and Use of Occupation and Industry Data by NAACCR Member Registries

John Fulton, Leanne Chiaverini, Steven Roffers, Dan Curran, Pam Parrish, Gayle Clutter, Irene Hall, Missy Jamison, Chris Newton

Abstract

Purpose

To evaluate the collection of industry and occupation (I/O) data by member registries, NAACCR's Data Quality Indicator (DQI) Subcommittee, a division of the NAACCR Data Evaluation and Publication Committee, surveyed the membership about the collection, coding, and use of I/O data.

Methods

Surveys were mailed to the directors of member registries. Responses were received from 41 (77%) of 53 U.S. state or territorial registries ("state" registries), the focus of this report.

Results

Seventy-six percent of the state registries that responded to the survey ("respondents") require I/O data in case reports, but only 11% report that their 1998 I/O data are ready to analyze. Only one respondent had ever used I/O data in an annual report. Twenty-seven percent have used I/O data at least once for epidemiological studies. Almost all get most of their I/O data from medical records. About half obtain these data from death certificates, as available. Many inadequacies of I/O data from hospital records were noted, including: frequent use of "unknown," frequent use of vague terms ("retired"), documentation of current rather than usual I/O, inconsistencies between data from hospital records and other I/O data, and the labor intensiveness of editing and coding the data, even when software is used which is capable of coding a proportion of cases.

Discussion

The U.S. government mandates the collection of I/O data by state registries funded by the National Program of Cancer Registries (NPCR). The labor invested in this process is considerable, and yet I/O data recorded by central cancer registries appear to have significant limitations. In the main, they are excluded from regularly published registry reports, although some epidemiological use of I/O data is reported. The findings detailed in this manuscript are relevant both to NPCR and to NAACCR; both organizations may wish to address this problem further.

Purpose

To evaluate the collection of data on industry (also known as kind of business or industry) and occupation (type of work or job) by member registries, NAACCR's Data Quality Indicator Subcommittee (DQI), a division of the NAACCR Data Evaluation and Publication Committee, surveyed the membership about the collection, coding, and use of industry and occupation (I/O) data.

Background

Prior to the advent of the National Program of Cancer Registries in 1992, few central cancer registries in North America mandated the inclusion of data on patients' industry of employment and occupation. The inadequacies of hospital medical records,¹⁻⁶ the primary source of cancer case reports, were too well known for most central cancer registries, struggling with minimal budgets, to pay much attention to them, much less mandate their inclusion in cancer case reports. For example, when the Rhode Island Cancer Registry was officially established by Rhode Island state law in 1985, the Rhode Island Department of Health conducted a study of the completeness of certain data in cancer case reports made to the Department in previous cancer registration efforts. These efforts,

although not sanctioned by law, had been serious and substantial, and had included sufficient resources to evaluate and improve cancer registration in the state. In a nutshell, the 1985 study found that industry and occupation data were available in less than 75% of the hospital charts from which most cancer case reports were abstracted, that industry data were sometimes reported by type of industry and sometimes by the name of the company, that many entries appeared to describe current employment instead of usual employment (the latter is preferred when a summary indicator is used), and that many case reports contained the categories "retired" and "housewife." Hospital cancer registrars were consulted about the findings of the study, and confirmed the inadequacies of medical records for the ascertainment of industry of employment and occupation. As a result, the Rhode Island Cancer Registry did not mandate the collection of these data in cancer case reports when it began receiving them in 1986.

In 1992, the 102nd U.S. Congress passed Public Law 102-515, the Cancer Registries Amendment Act, "to support the operation of population-based, statewide cancer registries in order to collect, for each form of in-situ and invasive cancer (with the exception of basal cell and squamous cell carcinoma of the skin), data containing –

- (1) demographic information about each case of cancer;
- (2) information on the industrial or occupational history of the individuals with the cancers, to the extent such information is available from the same record;
- (3) administrative information, including date of diagnosis and source of information;
- (4) pathological data characterizing the cancer, including the cancer site, stage of disease (pursuant to Staging Guide), incidence, and type of treatment; and
- (5) other elements determined appropriate by the Secretary."⁷

With resources made available by annual congressional appropriations, the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention (CDC) currently supports 49 central cancer registries (45 state registries, three territorial registries, and one registry in the District of Columbia.)⁸ Because the federal funding of these registries is made under the auspices of the Cancer Registries Amendment Act of 1992, these 49 registries are required to mandate the inclusion of industry and occupation (I/O) data in the cancer case reports made to them from all sources. Most case reports, of course, are heavily dependent upon hospital chart abstracts, and thus, may be subject to the same inadequacies which inhibited the collection of I/O data in cancer case reports prior to the Cancer Registries Amendment Act.

In the main, central cancer registries in Canada do not routinely collect I/O data in cancer case reports. Although registries in Canada are not mandated to collect I/O data, they may be obtained by linking cancer case reports with other available Canadian data sets. Practices vary from one province to another, based on available resources and each registry's assessment of the benefits and costs of these data relative to other public health surveillance needs.

Role of NAACCR's Data Quality Indicator Subcommittee (DQI)

DQI is a Subcommittee of NAACCR's Data Evaluation, and Publication Committee (DEPC). DEPC has been given many responsibilities, among which is the improvement, monitoring, and reporting of the completeness and quality of cancer incidence data collected throughout North America.⁹ Some of this responsibility is delegated to DQI at the pleasure of DEPC, under the auspices and guidance of NAACCR's Board of Directors. Typically, DQI is assigned an evaluation of data *not centrally important* to NAACCR's current annual Call for Data (a task which is usually retained by DEPC or falls under the aegis of other NAACCR committees altogether), but which may have implications for future Calls. For example, DQI has performed exploratory evaluations of the reliability of stage of disease at diagnosis in cancer case reports made to selected central registries in the NAACCR membership.¹⁰⁻¹² At present, these data are not published by NAACCR, but because their importance is central to cancer control efforts, they may figure prominently in future Calls.

I/O data have not been requested to date in NAACCR's annual Call for Data, nor have they figured prominently in the annual reports of central cancer registries. Nonetheless, collecting I/O information for roughly one million cancer cases annually represents a substantial investment of time and resources, and were I/O data to be complete, accurate, specific, and representative of usual lifetime industry and occupation, their promise for epidemiological studies might be substantial. For these reasons, DQI was assigned the responsibility of evaluating the availability and condition of these data, focusing on member U.S. registries (but including member Canadian registries, as well). Given the scope of this task, DQI began by assessing the collection, processing, and use of these data at present, receiving permission from the NAACCR Board of Directors to query the membership with a short, straight forward mail survey.

Methods

Following advice and guidance from NAACCR's Board of Directors (the Board), a 4-page questionnaire was drafted to query the NAACCR membership about the collection, processing, and use of I/O data. The questionnaire was balanced with sufficient pre-categorized responses to assure ease of response, and sufficient open-ended questions to facilitate rich feedback from the membership.

The questionnaire was pre-tested by members of DQI who volunteered for this purpose, and was modified on the basis of pretest responses and comments. It was then submitted to the Board, accompanied by proposed methods of administration, for permission to mail to the membership. The Board approved the survey form and methods, contingent on specific modifications, which were effected immediately.

Questionnaires were mailed to the directors of all member registries (all member registries in the U.S., including those funded by NPCR and those funded by the Surveillance, Epidemiology, and End Results – SEER – System, and all member registries in Canada) in early October, 2001, to be completed by them and returned by mail. A self-addressed, stamped envelope was included for this purpose. Responses were treated as anonymous. Respondents were not asked to reveal self-identifying information on the survey form. However, to administer the survey efficiently, the identity of each registry was noted on the outside of the return envelope, and recorded in a file separate from the data themselves. After about three weeks, members who had not returned the questionnaires were emailed a second request, with the survey form in ".pdf" format as an attachment.

Pre-categorized survey responses were aggregated using frequency distributions, and are summarized in Table 1 and the Results section. Responses to open-ended questions were extracted, collated, and categorized before summarizing in the Results section. Questions focused on I/O data collected in the years 1995-2001. (See Table 1).

Results

Responses were received from 41 (77 percent) of 53 U.S. state or territorial registries ("state" registries), eight (40 percent) of 20 other central cancer registries in the U.S., and seven (58 percent) of 12 central registries in Canada. Thirty-nine of the 49 U.S. registries that responded to the survey are NPCR-funded; 10 are SEER-funded. Responses from Canadian registries revealed that most (six of seven respondents) do not collect I/O data in cancer case reports. This report focuses on the 41 U.S. *state* registries that responded to the questionnaire (hereafter referred to as "respondents"), including 36 NPCR-funded state registries and five SEER-funded state registries.

Summary of Collection, Processing, and Use (Table 1)

Mandating I/O data: About three-fourths of the 41 state registries that responded to the survey ("respondents") require I/O data in case reports. Only about half of the respondents who require I/O data in case reports, however, mandate the collection of I/O data in state law.

Collecting I/O data: More than half of the 41 respondents have I/O data for diagnosis years 1995 through 2001. Almost 90 percent of the respondents (37 of 41) have I/O data for diagnosis years 1998 through 2001. (Please note that one of the respondents collected data on occupation, but not on industry in the years 1995-2000.)

Obtaining or Abstracting I/O data: Of the 37 respondents who collect I/O data, all but one (97 percent) obtain them (at least in part) from medical records. About half the respondents obtain I/O data (at least in part) from death certificates. Two mentioned obtaining I/O data from other sources, like patient questionnaires and follow-up letters.

Assuring the completeness of I/O data: Only eight percent of the respondents who collect I/O information check on its completeness in cancer case reports, using visual checks or computer edits for the presence of I/O narrative, but 27 percent have tried (at least once) to promote completeness of reporting, by stressing the importance of complete data with hospital cancer registrars, by holding training sessions on the collection of I/O data, or by conducting evaluation studies. Responses to open-ended questions indicated that these data are very incomplete in medical records.

Coding I/O data: All 37 respondents who collect I/O information do so in narrative format. Thirty percent also receive pre-coded I/O data from hospital registries. Only 11 (30 percent) of the respondents code narrative I/O information centrally. Of these 11 respondents, five code the data manually (using standard U.S. government coding manuals), four code I/O information using computer software, and two code I/O data both manually and with computer software. Of the six respondents who use computer software, five verify computer coding with at least some manual coding. Many inadequacies of I/O data from hospital records were noted as barriers to proper coding, including: frequent use of "unknown," frequent use of vague terms for industry or occupation (e.g., "retired"), inconsistencies between I/O data from different sources (e.g., hospital records and death certificates), and the labor intensiveness of coding data, even when software is used.

Computerizing I/O data: Fifteen of 37 respondents who collect I/O data computerize narrative information, and of these, 10 (67 percent of 15) computerize coded information as well.

Editing I/O data: Very few of the respondents have I/O data ready for analysis for any of the years 1995 through 2001. The highest percentage of respondents with "ready" data in any one year is 14 percent (1999).

Using I/O data: Only one respondent had ever used I/O data in an annual report, although 27 percent (10/37) had used their I/O data at least once in a special epidemiological study.

Summary of Barriers to the Accuracy, Completeness, and General Usefulness of I/O Data from Cancer Case Reports

Barriers Inherent in Data Sources: I/O data in medical records are inconsistently recorded throughout the record. Generic occupational categories are used. The data are incomplete, not validated, and conflicting. I/O data from death certificates are generally more complete, but their accuracy is often not verified.

Barriers Inherent in Collecting Data: It is costly to obtain standardized I/O data. Hospitals do not have standard protocols for the collection of these data. They are not readily available to registrars in a standard format. The data are difficult to collect reliably.

Barriers Inherent in Staffing: Central registries do not have adequate, properly trained staff to make the collection, processing, and use of I/O data efficient and worthwhile.

Barriers to Determining "Usual Occupation": In many cases it is unclear whether or not I/O data from medical records refer to current occupation or usual occupation, although it appears as if they are more likely to refer to current occupation. Terms such as "retired" are used frequently.

Barriers to Accurate and Precise Coding: A very large percentage of cases (approaching half in many hospitals) have "retired" listed for I/O. Narrative I/O data from medical records are so vague that they are difficult to code, either by hand or with computer software. Coding schemes are complex, thus difficult to

use without extensive training. When coding software is used, the results must be checked, at least on a sample basis, for coding reliability, and many cases must be coded by hand, anyway, because they contain data that are too vague for the software to handle.

Other Barriers: The data from medical records are so poor that these data are of the lowest priority for improvement and use. Significant resources at the hospital level as well as the central registry level would be needed to turn this situation around.

Summary of Suggestions from Respondents for Improving the Accuracy, Completeness, and General Usefulness of I/O data from Cancer Case Reports

Suggested Changes at the Hospital/Health Care Level: Advocate for policy regarding the collection of I/O data in health care facilities. Suggest a practical standard for the collection and storage of this information by health care facilities. The standard should include type of data to be collected, point and method of data collection, data format, and location of data storage. Advocacy should begin with acute care hospitals, beginning at the national level with the American Hospital Association and similar organizations.

Suggested Changes at the Central Registry Level: Assign I/O data concerns to one registrar at the central registry level. Provide the registrar with in-depth training about I/O collection, processing, and use, and with sufficient support to train hospital registrars throughout the jurisdiction on I/O collection. Create a manual on I/O data collection for use at the hospital level to support training and data collection efforts.

Suggested Changes in Overall Approach: Do not collect I/O data from medical records. Define special studies for the use of I/O data, and obtain the necessary data through patient/family questionnaires. Special studies might be defined at the national, state, or local level, or by researchers whose work goes beyond these jurisdictional boundaries. Create a budget line in NPCR awards for the support of special I/O studies by central cancer registries, beginning with studies defined at the national level. Create a team of epidemiologists at the national level who would work with I/O data pooled from central cancer registries, similar to the model of the Central Brain Tumor Registry of the U.S.

Discussion

The U.S. government mandates the collection of I/O data by state registries funded by the National Program of Cancer Registries (NPCR). The labor invested in this process is considerable, and yet I/O data recorded by central cancer registries appear to have significant limitations. In the main, they are excluded from regularly published registry reports, although some epidemiological use of I/O data is reported. The heart of the problem is haphazard data collection in health care facilities, not by registrars, but by a variety of professionals who may be collecting data for different reasons. I/O data may be collected as part of standard social histories, or to guide treatment. They may be collected for billing purposes, or less commonly, for research. I/O data may be recorded in various places and at various times, and with little concern for consistency. At the "hospital" level, registrars must locate the data, reconcile inconsistencies, and include them as best they can in cancer case reports. At the central registry level, registrars must work with vague, incomplete data, collected originally for a variety of unknown purposes according to unknown standards or no standards at all.

It is no wonder that most central registries do not spend much time and effort on these inadequate data. At present, they have no control over the root of the problem. Although central registries have some control over data abstraction, processing, and reporting of data at the facility level, they have little or no control over how data get into medical records in the first place. The only reason cancer registration works at all, given available resources, is because health care providers normally record information about patient demographics, diagnosis, and treatment in a normative way, supported in this endeavor by organizations such as the American College of Surgeons. Central registries build on this base to make sure that the data as originally collected are represented faithfully in composite case reports and aggregate statistics.

Thus, we have the two rather radical solutions to the problem of inadequate I/O data offered by the central registries, themselves: either to recruit national organizations to provide a strong rationale for complete and accurate I/O data collection in health care facilities, standards to do so, and an infrastructure to support and assure a strong I/O data product, or to scrap the approach entirely, and concentrate efforts on other means of I/O data collection, such as focused patient surveys. The former is consistent with the letter of Public Law 102-515, but would require a substantial sum of money and even so, would take considerable time. The latter is not consistent with the letter of the law, but does address its purpose, and would require far less money and take far less time to accumulate useful I/O data. If NPCR registries were to adopt this approach to the collection of I/O, data it might be prudent to begin by focusing on those cancers that have been previously associated with occupational exposures, and to limit such studies to selected states rather than conducting them in all states. This strategy would protect scarce resources while maximizing the probability of identifying occupational-related cancers. Less radical solutions, such as those that would improve data abstracting, processing, and reporting without affecting the ways in which I/O data get into medical records in the first place, would merely increase the completeness of inadequate data. We would have more data, but they would still be difficult to interpret.

As we weigh the pros and cons of collecting and improving these data, we should consider the uses to which these data have been used in the past, as revealed in the public health literature. For example, I/O data from population-based cancer registries have been used to demonstrate higher-than-average cancer incidence among architects (renal cell carcinoma),¹³ among people who work in sedentary occupations (colon cancer),¹⁴ and among those who are exposed to wood dust, pesticides, radium, and asbestos in the workplace.¹⁵ Researchers who have been successful in using I/O data from population-based cancer registries have regarded their availability, inexpensiveness, volume, and coverage (across industries) as strengths.^{16,17}

In sum, central cancer registries in the United States and the organizations that set standards for them may wish to address problems with I/O data as currently collected and processed.

References

1. Swanson GM, Schwartz AG, Burrows RW. An assessment of occupation and industry data from death certificates and hospital medical records for population-based cancer surveillance. *Am J Public Health* 1984 May;74(5):464-7.
2. Schade WJ, Swanson GM. Comparison of death certificate occupation and industry data with lifetime occupational histories obtained by interview: variations in the accuracy of death certificate entries. *Am J Ind Med* 1988;14(2):121-36.
3. Brownson RC, Davis JR, Chang JC, DiLorenzo TM, Keefe TJ, Bagby JR Jr. A study of the accuracy of cancer risk factor information reported to a central registry compared with that obtained by interview. *Am J Epidemiol* 1989 Mar;129(3):616-24.
4. Kipen HM, Gelperin K, Tepper A, Stanbury M. Acute occupational respiratory diseases in hospital discharge data. *Am J Ind Med* 1991;19(5):637-42.
5. McDiarmid MA, Bonanni R, Finocchiaro M. Poor agreement of occupational data between a hospital-based cancer registry and interview. *J Occup Med* 1991 Jun;33(6):726-9.
6. Balmes J, Rempel D, Alexander M, Reiter R, Harrison R, Bernard B, Benner D, Cone J. Hospital records as a data source for occupational disease surveillance: a feasibility study. *Am J Ind Med* 1992;21(3):341-51.
7. U.S. Congress. Public Law 102-515, *Cancer Registries Amendment Act*. Washington, DC: U.S. Congress, October 24, 1992.
8. Centers for Disease Control and Prevention. *National Program of Cancer Registries*. 2001. Web site: <http://www.cdc.gov/cancer/npcr/index.htm>

9. North American Association of Central Cancer Registries. *Mission, Goals, and Objectives, 2001-2005*. Springfield, IL: North American Association of Central Cancer Registries, July 7, 2001.
10. Porter R, Correa CN, Fulton JP, Howe HL, Newton C, Nowak J, Roffers SD. Exploring the internal consistency of registry data on stage of disease at diagnosis. VI.1-12. In: Chen VW, Wu XC, Andrews PA (eds). *Cancer in North America 1990-1994. Volume One: Incidence*. Sacramento, CA: North American Association of Central Cancer Registries, April, 1998.
11. Parrish P, Fulton JP, Correa CN, Roffers SD, Jackson N, Newton C, O'Connor L, Perkins C, Porter R, Tucker T, Weir H. Exploring the internal consistency of registry data on stage of disease at diagnosis: Part II: cancer of the prostate. Pp. VI.1-5. In: Chen VW, Wu XC, Andrews PA (eds). *Cancer in North America 1991-1995. Volume One: Incidence*. Sacramento, CA: North American Association of Central Cancer Registries, April, 1999.
12. Fulton JP, Parrish P, Roffers SD, Correa CN, Jackson N, O'Connor L, Perkins C, Newton C, Clutter GG. Exploring the internal consistency of registry data on stage of disease at diagnosis, Part III: suggestions for an EDITS metafile. Pp. VI.1-7. In: Chen VW, Howe HL, Wu XC, Hotes JL (eds). *Cancer in North America 1993-1997. Volume One: Incidence*. Springfield, IL: North American Association of Central Cancer Registries, April, 2000.
13. Lowery JT, Peters JM, Deapen D, London SJ. Renal cell carcinoma among architects. *Am J Med* 1991;20(1):123-5.
14. Brownson RC, Zahm SH, Chang JC, Blair A. Occupational risk of colon cancer. An analysis by anatomic subsite. *Am J Epidemiol* 1989 Oct;130(4):675-87.
15. Hutton MD, Burnett CA. Occupation of cancer patients: A challenge to healthcare facilities. *J of AHIMA* 1996 Jun;67(6):64-68.
16. Burnett CA, Silverman DT, Lalich NR. A comparison of analyses of occupational bladder cancer: death certificate vs. population-based case-control interview data. *Am J Ind Med* 1994 May;25(5):677-88.
17. Brownson RC, Davis JR, Chang JC, DiLorenzo TM, Keefe TJ, Bagby JR Jr. A study of the accuracy of cancer risk factor information reported to a central registry compared with that obtained by interview. *Am J Epidemiol* 1989 Mar; 129(3):616-24.

Table 1. Affirmative responses to questions about industry and occupation (I/O) information in cancer case reports, expressed as percentages. "Does your registry...?"

	<u>Percent</u>		<u>N*</u>
	<u>Occ</u>	<u>Ind</u>	
Mandating I/O data			
Require I/O information in case reports?	76	73	41
Mandate the collection of I/O information in law?	34	32	41
Collecting I/O data			
Have I/O information for calendar year 1995?	59	56	41
Have I/O information for calendar year 1996?	71	68	41
Have I/O information for calendar year 1997?	83	80	41
Have I/O information for calendar year 1998?	88	85	41
Have I/O information for calendar year 1999?	88	85	41
Have I/O information for calendar year 2000?	88	85	41
Have I/O information for calendar year 2001?	85	85	41
Finding I/O data			
Obtain I/O information from medical records?	97		37
Obtain I/O information from death certificates?	51		37
Obtain I/O information from other sources?	5		37
Assuring the completeness of I/O data			
Check the completeness of I/O information on case reports?	8		37
Do anything to improve the completeness of I/O information?	27		37
Coding I/O data			
Collect I/O information in coded format?	30		37
Collect I/O information in narrative format?	100		37
Code narrative information?	30		37
Code narrative information manually?	46		11
Code narrative information with computer software?	36		11
Code narrative information both manually and with computer software?	18		11
Verify computer-coded information with manual coding?	83		6
Computerizing I/O data			
Computerize I/O information?	41		37
Computerize coded I/O information?	67		15
Finishing I/O data			
Have coded, computerized information for calendar year 1995 ready for analysis?	5		37
Have coded, computerized information for calendar year 1996 ready for analysis?	3		37
Have coded, computerized information for calendar year 1997 ready for analysis?	8		37
Have coded, computerized information for calendar year 1998 ready for analysis?	11		37
Have coded, computerized information for calendar year 1999 ready for analysis?	14		37
Have coded, computerized information for calendar year 2000 ready for analysis?	8		37
Have coded, computerized information for calendar year 2001 ready for analysis?	8		37
Using I/O data			
Ever use I/O information in annual reports?	3		37
Ever use I/O information or provide it for special epidemiological studies?	27		37

* Number of respondents who answered the specific question