Original Article

Registering Multiple Primary Tumors in Central Cancer Registries

Holly L. Howe, PhD[a]; Rachel Weinstein, PhD[b]; Tamara Schult, MPH[c]
Betsy Kohler, MPH, CTR[d]; Joellyn Hotes, MPH[1]

Abstract: Coding rules for multiple primary tumors are complex and may diminish data reliability. The purpose of this study was to assess the reliability and utility of reports of multiple primary cancers among breast cancer cases. A NAACCR dataset for tumors diagnosed from 1994 - 1998 was used. Within each registry, all tumors were linked by patient identification number to determine the history of primary tumors for individuals. Once linked, patients with one breast cancer and another primary tumor were extracted. For these cases, data recorded for sex, race, ethnicity, and sequence number were compared among the multiple primary tumor reports. A full 10% of the sample was lost by the omission of 2 registries with pervasive errors in sequence number assignment, resulting in 327,537 records of invasive breast cancers. Of the 62,394 multiple primary patients, 24,273 had multiple primaries diagnosed during the 5 year interval. Of these, 32 tumors had an unknown sequence number; 1,953 patients had the first tumor incorrectly coded as a single primary; and 158 records were sequenced incorrectly. Inconsistencies were also found in race, ethnicity, and sex identification, but these were few. It was concluded that some data quality problems existed, however, routine quality assurance registry operations produced a reasonably accurate and useful patient-linked file.

Key Words: *breast neoplasms, epidemiology, multiple primary tumors*

Introduction

Registering multiple primary tumors in cancer registries is difficult for many reasons. [1] The complexity of the US coding rules for multiple primary tumors may exacerbate the issues related to data reliability and validity. Bergfeldt and colleagues reported an analysis of data from registries known to be of high quality, but which contained sufficient errors in the multiple primaries to affect interpretation.[2] The purpose of this study was to assess the reliability and utility of reports of multiple primary cancers where breast cancer was one of the tumors. This project was one undertaken by a collaborative research group focused on several breast cancer studies. It was selected due to the interest in multiple primaries of the breast and the complexity of the multiple primary coding rules for this site. The source of the data was the North American Association of Central Cancer Registries' (NAACCR) large, population-based incidence dataset from 1994-1998 which had been aggregated from submissions of multiple registries.

Methods

The NAACCR dataset

The NAACCR dataset, 1994-1998, is comprised of registries meeting the NAACCR standards for high data quality. [3] These criteria include the following standards: first, registries met a standard of fewer than one (1) duplicate per 1,000 records; second, each registry had 90% or higher case ascertainment for all 5 years, after adjustment forduplicate reports; third, all files were evaluated using a standard edit program which found the file to be error-free for variables used in the computation of cancer rates; and ,finally, data for 5 years, 1994-1998, were submitted. Among the high-quality registries, the following registries provided consent to use their data in this study:

"Registering Multiple Primary Tumors in Central Cancer Registries"

Address correspondence to: Holly L. Howe, PhD, Executive Director, 2121 W White Oaks Dr, Suite C, Springfield, IL 62704; Telphone: 217.698.0800 ext 2; Fax: 217.668.0188; email: hhowe@naaccr.org

[a] Executive Director (Howe), Manager of Data Evaluation and Publication (Hotes), North American Association of Central Cancer Registries, 2121 W White Oaks Dr, Springfield, Illinois 62704

[b] Managing Director/Senior Biostatistician, Biostatistics Analysis Center, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, Pennsylvania

[c] Epidemiologist, Center for Chronic Disease Outcomes Research, Minneapolis VA Medical Center (152-2E), One Veterans Dr, Minneapolis, MN 55417

[d] Director, Cancer Epidemiology Services, New Jersey Department of Health and Senior Services, PO Box 369, Trenton, NJ 08625-0369

Arizona
Illinois
Pennsylvania
California: Los Angeles

Colorado
Iowa
Rhode Island
Delaware

Connecticut
Louisiana
Michigan: Metro Detroit
Utah
Washington
Greater San Francisco Bay Area

Georgia: Metro Atlanta
Montana
Nebraska
West Virginia Wisconsin

Hawaii
New Hampshire
Wyoming
Idaho
New Jersey

On preliminary examination of the data file, obvious and pervasive errors in sequence number assignment were identified in 2 additional high-quality registries. Due to these errors, neither could be included in the multiple primary breast cancer studies. Omitting data from these 2 registries resulted in a loss of 36,264 breast cancer cases. All multiple primary tumors occurring in women with an invasive breast cancer (ICD-0-2 site code C50) were selected, as they were being used for several NAACCR breast cancer studies. No case was excluded based on the lack of microscopic confirmation (N=6744). A breast cancer that was diagnosed on autopsy or found only through a death certificate (N = 3137 women) was included. The dataset for 1994 through 1998 contained information regarding 327,537 cases of invasive breast cancer.

Multiple Primaries

Definition of Multiple Primary Cancers

The SEER multiple primary rules are the standard definitions used by cancer registries in the United States and were the rules used in this study. [4] Specifically, regarding breast cancer:

1. A single lesion of one histologic type is counted as a single primary.
2. A single lesion composed of multiple histologic types is counted as a single primary.
3. A subsequent cancer of the same histology as a previous cancer that is diagnosed within 2 months is counted as a single primary. A cancer that is stated to be a recurrent or metastatic cancer is not a new cancer.
4. Simultaneous (within 2 months) multiple lesions of the same histologic type (i.e., multi-focal) within the breast are counted as a single primary. Further, even when different lesions have different behavior codes (in situ or malignant), they are counted as a single primary with a malignant behavior code.
5. Simultaneous (within 2 months) bilateral involvement of the breast in which there is only one histology must have a determination of whether the patient has one or 2 independent primaries. If it cannot be determined, then it is counted as a single primary (and the laterality is recorded as bilateral).
6. Multiple lesions of the same histologic type occurring in the breast and a different site (e.g., ovary) are counted as multiple primary cancers unless stated to be metastatic.
7. Multiple lesions of different histologic types within the breast are counted as multiple primary cancers whether occurring simultaneously (within 2 months) or at different times. However, if combinations of ductal and lobular carcinoma occur within 2 months of each other, they are counted as a single primary with a combination histology (ICD-O-2 histology code of 8522).
8. Multiple lesions of different histologic types occurring in the breast and a different site are multiple primary cancers whether occurring simultaneously (within 2 months) or at different times.

Patient-linked Analyses

The data file of all tumors diagnosed from 1994-1998 was linked within each registry by patient identification number to determine the history of primary tumors for individuals. Once the records were linked, patients with one breast cancer and another primary tumor were extracted. For these cases, data recorded for sex, race, ethnicity, month and year of diagnosis, and sequence number were compared among the multiple primary tumor reports. The analyses of sequence number assignment of breast cancer cases included: identification of duplicate sequence numbers and consistency of sequence number assignment based on date of diagnosis for individual tumors. The accurate application of SEER multiple primary rules for designating bilateral breast tumors from multiple primary tumors could not be assessed since the rule is qualified by acceptance of a physician statement of multiple primary disease even when the rules would designate it as a bilateral single primary tumor. Thus, all tumors coded as a multiple primary that appeared to be a bilateral single primary based on dates and histology were accepted as a multiple primary assuming a doctor's statement was the basis for the decision.

Results

From the file of 327,537 records, 62,494 patients with multiple primary cancers involving at least one breast tumor were identified. One hundred records with a duplicate sequence number of zero (0) (indicating a single primary) had to be omitted before linking the tumors because the tumors could not be sequenced. Of the 62,394 patients, 24,273 had a combination of one breast tumor and a second multiple primary diagnosed within the 1994-1998 period. The remaining multiple primary diagnoses had another pri-

mary diagnosed outside the 5-year study period.

Sex, Race, and Ethnicity Inconsistencies

Among the 24,273 patients with information on 2 tumors, 74 (0.3%) had inconsistent codes for sex. Eighty-eight (0.4%) individuals had an inconsistent race code, of which some varied in terms of specificity (e.g., Asian NOS and Chinese). Hispanic ethnicity had the greatest number of inconsistencies with records for 205 (0.8%) individuals that did not agree.

Thirty-two (0.1%) patients had a tumor with an unknown sequence number (sequence number equal to 99). These records could not be corrected because they had inconsistent or incomplete information for the other tumors. These patients would have to be omitted from any analyses of breast multiple primaries. Of the 24,241 remaining patients, 1953 (8%) patients had the first tumor incorrectly coded to a zero sequence number. For all of these patients, the correct sequence number could be assigned based on month and year of diagnosis. Based on this information, the sequence of the other tumor (s) was corrected if the assigned number was affected by the correction. Similarly, based on month and year of diagnosis, 158 (0.6%) additional records were identified that were sequenced incorrectly. Using dates of diagnosis, the sequence numbers could be amended.

Discussion

The omission of the data from the 2 registries with pervasive errors in sequence number assignment reduced the sample size of all breast cancer patients by about 10%. Thus, the reported results of coding errors most likely underestimates the problem among all registries, since only those that met minimal data quality standards were even eligible for study inclusion.

Among the remaining registries that were able to be included in the analyses, results indicated that although some data quality problems existed in cancer incidence records for multiple primary tumors among registries designated as having high-quality data, these registries were able to produce a reasonably accurate patient-linked file. A thorough review of a dataset before analyses can identify, verify, and correct many errors, even without follow-back to reporting facilities; however, it would be preferable if registries incorporated a routine resolution process as a standard part of their record consolidation and data editing procedures. As noted by others,[2] these errors should be considered as sentinel of the data quality of multiple primary cases relative to indicators measuring more general aspects of quality cancer operations and registration.[5] If month and year of diagnosis are available on the data file, many errors can be corrected and the data can be included in subsequent analyses. Without this information on the analytic file, it would not be possible to correct the information, resulting in a loss of patients for analysis.

In this study, the hospital and registry source records were not assessed for validity of information. Despite the lack of a reabstracting effort, it was possible to evaluate the reliability of reported anonymized data. When studying rare events, such as specific combinations of multiple primary cancers, errors leading to the omission of cases can dramatically reduce cases available for analysis and introduce unknown biases. This may weaken the value of the cancer registry as a population-based data resource if too many cases are omitted.

Education and communication among registrars, abstractors, and physicians must be improved so that all are aware of uses of the multiple primary data and the importance of following coding rules precisely and documenting completely all relevant information in the medical record. Some of these uses are: the iatrogenic and carcinogenic effects of cancer treatment on the development of subsequent cancers; the increased risk of second tumors among cancer patients; the impact of multiple tumor diagnoses on survival rates; or the genetic predisposition or common risks that occur among persons with specific combinations of cancer diagnoses. It is an educational challenge to train registry employees to reliably apply multiple primary rules in a manner that is consistent across all cancer registries. The current rules used in the United States are complicated; using them to accurately classify and code multiple primaries can require years of experience.

Researchers and registries can easily identify inconsistencies in race, ethnicity, gender, and tumor sequence by the application of computerized logic checks. By using the other data in the record, or for the individual, many of these inconsistencies can be made internally consistent. Inter-record edit programs are available to cancer registries and all registries should routinely use them to identify errors and inconsistencies among multiple records for the same tumor or the same individual. In 2004, NAACCR will begin to require that a standard inter-record edits program be applied to data submissions.

However, the ambiguity surrounding certain aspects of the current multiple primary rules used in North America makes the verification of other data elements difficult. For example, the exception allowing a physician to call multiple breast tumors of the same histology occurring synchronously in different breasts either bilateral disease or multiple primaries prevents the application of logical edit checks to validate the data. Similarly the use of the rule defining tumors in the same organ and of the same histology diagnosed within 2 months, as a single primary, while those diagnosed more than 2 months apart are multiple primaries may cause misclassification of tumors. This rule, while easy to apply to most situations, causes data to be grouped in categories that may or may not have biological or clinical importance. Multiple primary rules need to have biological and clinical importance as well as being unambiguous in order to be meaningful in epidemiologic analysis. Misclassification of multiple primary status may obscure important patterns of the disease. While the international rules for categorizing multiple primaries are simpler, they were not developed as a reflection of the international thought on underlying biology or clinical significance of multiple cancers. They were designed as an analytic tool

whereby non-uniform definitions of multiple primaries could all be reduced to a lowest common denominator to enable world-wide or international comparisons among both developing and developed countries.

Analysis of multiple primary data, including analysis of data quality, is affected by the number of years of data that are available. Results from studies that include only short intervals may be different than those that have long-term survival and follow-up information; any risks for primaries that occur after longer intervals of time will not be captured in such analyses. An additional consideration is the length of time a registry has been in existence. Registries that started recently will have fewer second primaries (and fewer higher order primaries) than more mature registries. However, the pattern will stabilize once a registry has collected incidence data for about 15 years.[5]

The length of time a registry has been in existence determines the prevalence of multiple primaries (tumor sequence numbers should be a patient-based measure that includes all tumors over the life span, regardless of geographic and temporal bounds). In examining the distribution of sequence number by age in SEER data, about 90 % of those in the under-50 age group have only one primary.[5] However, by age 75, 25% of cancers are multiple primary cancers. As the population ages, there will be more cases of multiple primary cancers and these cancers will increase in public health importance. Overall, according to SEER data, approximately 13% of invasive cancer at all sites are second primary cancers.[5] The importance of data quality affecting multiple primary information will therefore continue to increase.

The focus of this study was on multiple primaries of the breast. The issues related to tumor sequencing may be different for other cancer sites which have different multiple primary coding rules, in addition to differences in survival, treatment, and age at diagnosis.

All registries need to conduct routine processes to assure that vendors and reporters are following standard code assignment for multiple primary tumors. This practice will eliminate many potential errors, but not all of them. Additional steps to check the reliability of diagnosis dates, and race, sex, and ethnicity assignment among all the tumor reports must be used to identify and correct inconsistencies.

References

1. Howe HL, Weinstein R, Hotes J, Kohler B, Roffers SD, Goodman MT. Multiple primary cancers of the ovary in the United States, 1992-1997. *Cancer*. May 15, 2003;97/10(suppl):2660-75.

2. Bergfeldt K, Silferswärd C, Einhorn S, Hall P. Overestimated risk of second primary malignancies in ovarian cancer patients. *Eur J Cancer*. 2000;36:100-5.

3. Howe HL, Chen VW, Hotes J, Wu XC, Correa C, Fulton JP, eds. *Cancer in North America, 1994-1998, Vol One: Incidence*. Springfield, LI: North American Association of Central Cancer Registries; April 2001.

4. *SEER Program Code Manual, 3rd edition*. Surveillance, Epidemiology, and End Results (SEER) Program, National Cancer Institute, National Institutes of Health (NIH); SEER Publication No. T-894; 1998 NIH Publication No. 98-1999.

5. Howe HL, ed. *A Review of the Definition for Multiple Primary Cancers in the United States: Workshop Proceedings from* Princeton, New Jersey, December 4-6, 2002. Springfield, LI: North American Association of Central Cancer Registries; May 2003.

Acknowledgments:

"Supported in part by the Centers for Disease Control and Prevention under cooperative agreement U75/CCU515998 to the North American Association of Central Cancer Registries."

Did You Know . . . you can earn continuing education credit by reading this article?

All YOU need to do is carefully read the quiz article, pp 113-116, then turn to page 142 and complete the 10 question quiz. Pick only ONE best answer for each question. Mail your completed quiz answer sheet, along with your processing fee ($25 for NCRA members, and $35 for non-members) to NCRA. With a passing score of 70% or better, you will receive one CEU. Within 4-6 weeks after submitting the answer sheet each participant will receive notification of completion of the CE process from the Executive Office of NCRA. Quizzes may not be retaken nor can NCRA staff respond to questions regarding answers. The NCRA Executive Office will date each quiz when it is submitted, which will determine the CE cycle year. NCRA is very pleased to offer registrars this alternative method of self-study for earning continuing education. Please direct all inquiries regarding the new JRM CE Quiz to the NCRA Executive Office at (703) 299-6640.