

SHOULD WE OR SHOULDN'T WE COMPARE CANCER INCIDENCE RATES AMONG REGISTRIES?

Vivien W. Chen, Ph.D.

INTRODUCTION

"What is the cancer incidence rate in our area? Is it higher, lower, or comparable to the national average?"

These are the most frequently asked questions of a population-based cancer registry, often raised by local government officials, concerned citizens, and researchers. Can a registry make the appropriate comparative analysis? What is the national rate anyway? Until recently, cancer incidence statistics from the National Cancer Institute-funded SEER (Surveillance, Epidemiology and End Results) Program have been referred to as the national rates in the United States, although the population in the combined SEER areas reflects cancer rates for only about 14 percent of the U.S. population and the method for selecting the areas was designed to include unique and special populations within the U.S.

With more states establishing population-based registries, cancer incidence data outside the SEER areas are now becoming available. In 1994, the Data Evaluation and Publication Committee of the North American Association of Central Cancer Registries (NAACCR) published for the first time *combined* cancer incidence rates for the United States. The question remains, should a registry compare its incidence rates to the SEER combined rates or the combined rates for the U.S.? The answer depends on how comparable the data are. Before one performs a comparative analysis, several issues should be evaluated in order to determine the comparability of data and the appropriateness of the comparison group.

DEFINITION OF AN INCIDENT CASE OF CANCER

Incident cases constitute the numerator of a rate in cancer registry data. Unless registries use the same case definition, data collected may not be compatible or comparable.

Reportable Diagnoses

Most registries report all neoplasms with a behavior code of 2 or 3 in the International Classification of Diseases for Oncology, 2nd edition (ICD-0-2), diagnosed on or after the registry's reference date.^a Exceptions are basal cell and squamous cell carcinomas of non-genital skin and pre-invasive cervical neoplasia. Recurrent cancer cases are excluded and metastatic tumors are distinguished from primary lesions. If two registries use different reportable criteria, comparisons of incidence rates between them will be invalid unless the registry can extract a comparable subset of cases. While some registries abstract tumors that are benign or of uncertain behavior, these cases should be excluded in all computations of malignant neoplasms.

Ambiguous Terms

While pathologists, radiologists, and clinicians sometimes use ambiguous terms in making cancer diagnoses, each registry should set rules in interpreting those terms. The SEER Program rules mandate inclusion of cases with diagnostic terms such as *probable*, *suspected*, *suspicious*, *compatible with*, *most likely*, *consistent with* and exclusion of cases

^a A reference date is the date when case ascertainment of cancers cases begins or becomes effective for a specified population.

considered as *possible*, *suggestive*, and *equivocal*. If a registry intends to compare its rates to the SEER rates, it should adopt the same rules for interpreting ambiguous terms as the SEER program.

A registry which includes only definite diagnoses and excludes any diagnoses with the above ambiguous terms will have lower incidence rates than the combined SEER areas. A registry which reports cases with possible or suggestive cancers might have higher rates.

Single vs. Multiple Primaries

While the determination of how many primary tumors a patient has is a medical decision, operational rules for registries are needed in order to ensure consistency of reporting and comparability of data. The basic components in SEER's policy for reporting whether tumors are single or multiple primaries include: the site of origin, the date of diagnosis, histologic type, the behavior of neoplasm (*in situ* vs. invasive), and laterality. Registries not collecting these variables, or not following SEER rules, will likely report the number of primaries differently. Take, for example, a woman diagnosed with bilateral breast cancer simultaneously, two lesions (medullary carcinoma and lobular carcinoma) in the right breast and one lesion (lobular carcinoma) in the left breast. A SEER registry would record three primaries of breast cancer for the woman, a Canadian registry following the International Association of Cancer Registries (IACR) rules would record two, whereas a central registry not collecting laterality or following SEER or IACR rules of histologic type, but resolving apparent duplicate reports thoroughly, would record only one. As a result, these two registries would report different breast cancer incidence rates regardless of the risk of breast cancer in the two populations.

COMPLETENESS OF CASE ASCERTAINMENT

Completeness of case ascertainment is defined as the extent to which all the incident cancers occurring in a defined population are included in the registry database. Ideally, completeness should be close to 100 percent, so that comparison of rates between registries reflects true differences in risk of cancer and not artifacts of the registration process. In most registries, incompleteness is often the result of less than ideal casefinding procedures, brought about by constraints posed by limited staff, resources, and support. It has been demonstrated that adequately funded registries, such as those in the SEER program and those which are mandated by law with an enforcement mechanism, have higher percent of case completeness. Registries which are totally dependent on passive reporting generally have a lower degree of case completeness.

Report Sources

Most population-based cancer registries in the U.S. use multiple report sources for casefinding. In addition to hospitals, private pathology laboratories and free-standing radiation centers are screened routinely. With health care reform and the shift of health care delivery towards ambulatory or outpatient services, it is expected that cancer cases diagnosed or treated in outpatient settings will continue to increase. Registries not capturing cases from outpatient sources are under-ascertaining cases. An important example is the treatment of cutaneous melanoma at dermatology clinics. A substantial proportion of cases (21 percent) has been shown to be missed when melanoma cases from these outpatient clinics are not ascertained or reported.¹

While active reporting of cancer from physicians is not required in many population-based registries, follow-back to physician offices is often needed for additional demographic information, residence, or date of diagnosis to determine the eligibility of cases identified from sources such as private pathology laboratories, dermatology clinics, or death certificates. Under-registration is most likely to occur when registries do not perform queries at physicians' offices.

Cases Diagnosed or Treated Outside Reporting Area

By definition, the cancer incidence rate for a specific population includes in the numerator, all incident cases diagnosed among *residents* of that population regardless of the place of diagnosis. It is, therefore, essential for registries to ascertain resident cases diagnosed outside the reporting area so that accurate incidence rates can be computed. Registries that do not have case-sharing agreements with neighboring states or areas are probably under-reporting incidence, especially if they are adjacent to large medical center complexes, comprehensive cancer centers, or specialized pediatric cancer centers. While some cases are recaptured at death (if death certificate clearance is performed by the central registry), the non-lethal cases will be missed.

Death Certificate Clearance

Death certificate clearance is a vital step in achieving complete case ascertainment in a population-based registry. It serves as a quality control check on completeness of casefinding from medical facilities in the reporting area and identifies missed cases. All deaths among residents with cancer mentioned on the death certificates should be linked to the registry data file. The unmatched cases are often followed back to institutions where the death occurred, to coroner's offices and nursing homes, or by contacting the certifying physicians.

Cases that remain unresolved are included in the registry as death-certificate-only cases. They may represent cancers not diagnosed before death or cases diagnosed and treated outside the reporting area but not ascertained by the registry. While these cases are recaptured, they often do not reflect the actual year of diagnosis.

Registries which do not perform death certificate clearance have under-reporting in their cancer statistics. On the other hand, registries that identify cases from death certificates but do not investigate vigorously and follow-back could inflate their rates by including prevalent cases from a previous year. This scenario is more common in registries that have only been in operation for a short time period.

Measurement of Completeness

Since the completeness in reporting cancer cases can affect incidence rates, the Data Evaluation and Publication Committee of NAACCR includes in this monograph several indicators of completeness of case ascertainment for each registry. In addition to the self-reported estimated percent of completeness, the following indicators of completeness are also presented: percent of death-certificate-only cases, the race- and site-specific incidence to mortality ratios, and the inclusion of death clearance (follow-back) as a routine operational procedure of the registry.

A registry with no death-certificate-only cases is probably not using death certificates as a source of reporting and is therefore under-reporting incidence. A registry with a high percent of death-certificate-only cases, however, could be either under-reporting or over-reporting. The death-certificate-only cases may represent missed cases from other sources, such as outpatient clinics or cases diagnosed and treated outside the reporting area, that are recaptured at death. It is therefore reasonable to assume that there are other missed cases not recaptured at death, in particular, non-lethal cancers. On the other hand, a high percent of death-certificate-only cases may reflect ascertaining cancer cases by means of death certificates without clearance and follow-back, therefore over-reporting incidence by including prevalent cases.

The incidence to mortality ratio (*I/M*) is often used as a crude measure for completeness of casefinding of a registry. Assuming the SEER program as the "gold" standard, the *I/M* of SEER is applied to the mortality rate of the registry's target population to calculate the expected incidence and estimate completeness. However, there are other factors which can affect the *I/M* ratio. The most predominant one is the detection of asymptomatic cancers when a screening program or test is first introduced that might result in an increase in incidence. A recent example is that of the prostate-specific antigen (PSA) test and the concomitant increased incidence of prostatic cancer. Registries covering geographic areas where there are active cancer screening programs will have higher *I/M* ratio than areas without such programs regardless of the completeness of case ascertainment.

Quality control audits on casefinding are another measure. The audit is usually labor intensive and costly. It is often performed in hospitals to assure that cancer cases diagnosed in all possible sources within an institution are ascertained. Registries which have routine quality control programs in place are more likely to be complete in their reporting system.

The Data Evaluation and Publication Committee is in the process of identifying some indicators that will objectively measure completeness of reporting in a central registry.

RESOLUTION OF DUPLICATE CASES AND CASES WITH UNKNOWN DATA

Duplicate Cases

As the health care practice in the U.S. shifts towards cost-containment, a cancer patient is likely to be admitted to multiple medical facilities for diagnosis and treatment, e.g. diagnosed in a private clinic, admitted for surgery in a hospital, and received radiotherapy at a free-standing radiation center. A comprehensive network of reporting systems will ascertain the same patient three times in the registry. A properly functioning registry should have record linkage mechanisms in place to resolve duplicates of a single cancer case from multiple report sources so that over-estimation of incidence does not occur. A substantial number of duplicates in a registry can sometimes mask missing cases, giving an apparent impression of high degree of completeness, especially when evaluated by the I/M ratio.

Cases with Missing or Unknown Data

Cases with unknown or missing data are often ascertained from non-hospital reporting sources, particularly when follow-back is not performed. While unknown or missing data on primary site, age or race are often used as indicators for the quality of data in a registry, they can influence incidence rates as well.

Registries with a high percent of unknown primary site will have lower site-specific rates. Therefore, a deficit in the comparison of site-specific incidence rates does not necessarily reflect a lower risk of a specific cancer in the population. Likewise, unknown sex or race could result in under-estimation of incidence since the rates are generally computed by sex and by race. Age, another key variable in cancer registration, is needed for age-standardization of rates as well as for presentation of age-specific rates. Registries may vary in their procedures for handling unknown age in the computation of incidence rates. If a registry excludes cases of unknown age, age-adjusted incidence rates will be underestimated. Some registries proportionately distribute these cases in the computation of age-adjusted rates, which may or may not be appropriate. Therefore, before one performs comparisons, one needs to have some knowledge of the percent of unknowns in essential variables that affect incidence rates, as well as information on duplicates in the registries and the mechanism for duplicate resolution.

COMPARISONS OF PUBLISHED STATISTICS

Published cancer statistics (incidence and mortality) are available for comparison. The National Cancer Institute's SEER Program publishes cancer incidence, mortality, and survival, with additional special topics (e.g., time trend analyses, cancer experience of minority populations) featured every year. Incidence rates and survival are compiled for the combined SEER areas, whereas mortality rates are often provided for both SEER areas and the entire nation. The rates are age-adjusted to the U.S. 1970 standard population, and they are sex- and race-specific. Only invasive cancers are included in the computation of incidence rates, with the exception of bladder cancer which includes *in situ* cases. Registries that want to compare their rates with SEER rates should use the same criteria to generate the incidence rates.

The American Cancer Society publishes *Cancer Facts and Figures* annually, providing the estimated numbers of new cancer cases and cancer deaths by state. In addition, a death rate, adjusted to the U.S. 1970 population, is generated for each state. The reported death rate is the total cancer mortality rate, i.e., for all races and both sexes

combined. While this summary statistic has taken into account the differences in age distribution of population in each state (by age adjustment), it does not consider the differences in racial distribution. Since race is a main determinant of cancer mortality, the racial composition of the population in a state can influence its overall rate. For example, in 1994 the District of Columbia had a total cancer mortality rate 46% higher than the state of Iowa (230/100,000 *cf.* 158/100,000, respectively). Although the disparity reflects a difference in cancer mortality between these two areas, it does not take into consideration that the difference is partly due to a higher proportion of African Americans in the population of the District of Columbia. African Americans have higher cancer mortality rates than U.S. whites. Traditionally, local government officials and concerned citizens request a single summary cancer rate for comparison. Registries should be cautious in performing such comparative analyses and in interpreting the results.

Cancer Incidence in Five Continents (CI5) has been published by IACR during the past three decades. While IACR includes only invasive cancers in their statistics, their rules for determining single versus multiple primaries are slightly different from those of SEER. Only one primary is recorded for an organ or pair of organs or tissue, defined by the first three digit codes of ICD-O-1st edition. Therefore, cancer statistics published by IACR will reflect lower incidence rates than those by the SEER Program, in particular, cancers of the breast and colon. Because of the global participation of registries in CI5, the incidence rates are age-adjusted to the World Standard Population. Since SEER uses the U.S. 1970 population for age-adjustment, direct comparison of published statistics between SEER and IACR is, therefore, not meaningful and cannot be used to produce an accurate interpretation of regional variations.

COMPARABILITY OF POPULATIONS

While comparability of data between registries is important, the appropriateness of the comparison group is vital as well. The SEER combined rates have often been referred to as *the U.S. national rates* and used by many registries for comparison. However, the populations covered by the SEER registries include some of the unique and special populations in the U.S. The inclusion of Utah (a state whose population is predominantly Mormon and experiences very low cancer rates), Iowa (a rural state), Connecticut (an affluent state with a high socioeconomic status) could result in a lower cancer incidence rate for SEER whites. African Americans in the SEER areas are from predominantly metropolitan areas and may have different cancer rates than African Americans in non-metropolitan areas. Therefore, it is important that a registry not only compare its rates with the SEER combined rates, but also compare its rates with those of individual registries or a group of registries with known high quality, high case reporting completeness, and with similar population characteristics.

Finally, geographic areas with active cancer screening programs (e.g. screening mammography and breast cancer, PSA and prostatic cancer), will detect early cancers among asymptomatic individuals, resulting in higher incidence rates, at the start of screening programs. These higher rates do not necessarily reflect a higher risk of disease.

SUMMARY

Should we compare cancer incidence rates among registries? Yes, if we are comparing rates that are based on the same definitions and methods of computation; no, if the definitions or methods are different. All registries which wish to evaluate their incidence rates relative to national or international data should collect data according to established standards so that valid and appropriate comparisons can be made. Both NAACCR and IARC have published such standards.^{2,3}

REFERENCES

1. Karagas MR, Thomas DR, Roth GJ, Johnson LK, Weiss NS. The effect of changes in health care delivery on the reported incidence of cutaneous melanoma in western Washington State. *Am J Epidemiol* 1991; 133:58-62.
2. Seiffert, JE (ed). Standards for Cancer Registries Vol. III. Standards for Completeness, Quality, Analysis and Management of Data. Sacramento, CA: North American Association of Central Cancer Registries, 1994.
3. Parkin DM, Chen VW, Ferlay J, Galceran J, Storm HH, Whelan SL. (eds). Comparability and Quality Control in Cancer Registration. Lyon, France: IARC Technical Report No. 19, 1994.