EXPLAINING GEOGRAPHIC VARIATION IN BREAST AND CERVICAL CANCER INCIDENCE RATES IN US HISPANIC WOMEN

Objective: This study examined geographic variation in incidence rates for two cancers common in US Hispanic women and considered some potential explanations, by using data from several high-quality cancer registries.

Methods: Age-standardized incidence rates (ASIRs) per 100,000 Hispanic women per year were analyzed for breast and cervical cancer in the population-based cancer registries of the Surveillance, Epidemiology and End Results (SEER) Program. The percentage potentially misclassified (PPM) as Hispanic, because of the frequent absence of maiden name, was estimated. Sociodemographic characteristics of the cancers and of the Hispanic population in eight areas were compared.

Results: The ASIRs varied from 80.9 to 113.3 for breast and 8.2 to 19.9 for cervix cancers. The PPM was higher for breast (16%) than for cervix (7%). Differences in PPM across SEER areas did not appear consistent with geographic variation in ASIRs, while some variation in ASIRs was consistent with differences in sociodemographic characteristics. Registry data on subgroups (defined by ancestry and birthplace) were too incomplete for analyses of ASIRs.

Conclusions: Some of the geographic variation appeared to be explained by sociodemographic factors, but improvements in cancer registries are needed to analyze ASIRs for Hispanic subgroups. (Ethn Dis. 2005;15:727–732)

Key Words: Breast Cancer, Cancer Surveillance, Cervical Cancer, Hispanic

INTRODUCTION

Geographic variation in cancer incidence rates has been reported in the US Hispanic/Latino population from incidence data from population-based cancer registries that differ in completeness of case ascertainment and overall quality.1-3 Matching surnames in the registry with a Spanish surname list may improve estimates of cancer incidence in Hispanics.4 Validation studies, however, suggest that misclassification of Hispanic ethnicity in registries is greater for women than men, due at least in part to frequent absence of maiden names.4-6 The extent of such misclassification could vary among cancer registries, thus producing apparent (but artificial) geographic variation in incidence rates for Hispanic women.

Another potential explanation for geographic variation in cancer incidence rates among Hispanics is geographic variation in socioeconomic status (SES) and other sociodemographic characteristics of the Hispanic populations. Hispanic subgroups differ in sociocultural factors, in medical care use, health behaviors such as use of tobacco, and genetic composition.5,7,8 Cancer incidence data for specific Hispanic subgroups have not been reported, however, because data on ancestry and birthplace in cancer registries have been judged to be too incomplete, but detailed data on completeness have not been presented.9,10 The purpose of the present study is to examine these two potential explanations for geographic variation in incidence rates for two cancers (breast and cervix) common in Hispanic women,1-5 by using data from selected high-quality cancer registries. These two cancers were selected, in part, because SES is positively associated with breast cancer risk and negatively associated with cervical cancer risk.11,12 While misclassification of patients as Hispanic should have a similar effect on rates for both cancers, the effects of geographic variation in SES should differ by cancer site.

METHODS

This study used data from the US National Cancer Institute’s Surveillance, Epidemiology and End Results (SEER) Program. The program’s standards for quality, such as completeness of reporting as assessed by independent audits of hospitals and proportion of cancers ascertained solely from death certificates, are more stringent than for other registries.13 The SEER program’s five states and six metropolitan areas cover ~14% of the entire US population and ~25% of the Hispanic population. The total population covered is not representative of the entire United States.5,14 However, SEER data have been the major source of national estimates of cancer incidence rates.10,15,16 A SEER public-use data file issued in 200317 was used to identify breast and cervical cancers diagnosed in 1995-

Anthony P. Polednak, PhD

Validation studies, however, suggest that misclassification of Hispanic ethnicity in registries is greater for women than men, due at least in part to frequent absence of maiden names.4-6
2000 or the same period used in a report from the North American Association of Central Cancer Registries (NAACCR) on cancer incidence in US Hispanics/Latinos. Cancers and cancer rates routinely reported from cancer registries involve numbers of cancers, not persons (or women) with cancer; some patients are diagnosed with two or more cancers simultaneously or during any specific time period. Some 10,623 invasive (malignant) breast cancers and 2441 invasive cervical cancers among Hispanic women were identified for all 11 SEER registries combined. “Hispanic” was defined (as in SEER reports) by using SEER codes 1–7 for the Spanish surname/origin variable. Codes 1–5 indicate specific Hispanic subgroup (ie, Mexican, Puerto Rican, Cuban, South or Central American except Brazil, and other specified origin including European). Code 6 is Spanish/Hispanic/Latino not otherwise specified (NOS). These codes are derived mainly from reports from hospitals or clinical laboratories to each of the SEER registries. The SEER registries use code 7 for cancers diagnosed since 1994 to indicate that the patient may be of Spanish surname/origin or Hispanic solely because the surname and/or maiden name matched with a list(s) of Spanish surnames. The 1980 or 1990 Census list of Spanish surnames is used by registries. In this report, records with code 7 are referred to as “SS-only.”

The program SEER*Stat was used to calculate average annual age-standardized incidence rates (ASIRs) per 100,000 Hispanic women per year and 95% confidence limits on these rates. Rates were directly standardized by applying age-specific rates (18 age groups) to the age distribution of the year 2000 US standard population. ASIRs are not shown separately for three SEER areas (Detroit, Mich; Hawaii; and Iowa) because the number of cancers was small (<100 breast and ≤20 cervix).

For use in interpreting the ASIRs, 1990 Census data are tabulated on the poverty rate (ie, the percentage of persons age ≥18 years with incomes below the federal poverty threshold) for the entire Hispanic population of each SEER area. Poverty rate is often used as an indicator of variation in SES in the populations at risk for cancer. The percentage of the Hispanic population that was foreign born was also tabulated for each SEER area. In addition, sociodemographic characteristics of Hispanic breast and cervical cancer cases are tabulated for each SEER registry. Included is the SEER birthplace variable, which was assigned a three-digit code for state or country of birth and 999 for unknown. The SEER marital status variable was coded as unknown for 442 (4.2%) of 10,623 breast and 115 (4.7%) of 2441 cervical cancers; others were recoded as ever vs never married. The proportion of SS-only cancers was multiplied by the proportion of SS-only that were coded as ever-married. This product was expressed as a percentage and referred to as the percentage potentially misclassified (PPM) as Hispanic because of the frequent absence of maiden names in cancer registries.

RESULTS

Average annual ASIRs per 100,000 Hispanic women per year by SEER area varied from 80.9 to 113.3 for breast and from 8.2 to 19.9 for cervix (Table 1). San Francisco-Oakland and San Jose-Monterey, Calif had breast cancer rates that were relatively high, while cervical cancer rates were relatively low; the Hispanic populations in these areas had the lowest poverty rates (12%–14%). In contrast, Los Angeles had a relatively low breast cancer ASIR and a relatively high cervical cancer ASIR; the Hispanic poverty rate was 20%. With Hispanic population poverty rates that were <20%, Atlanta and Seattle had relatively low Hispanic cervical cancer rates, while the breast cancer ASIR was relatively high in Atlanta. New Mexico had a low cervical cancer ASIR despite having a Hispanic population with the highest poverty rate of all SEER areas. Breast cancer ASIRs were highest in Connecticut and Utah, despite having populations with Hispanic poverty rates similar to that of Los Angeles (Table 2).

For all 11 SEER areas combined, 29.4% of breast and 15.6% of cervical cancers were classified as Hispanic by SS-only (Table 2). The percentage of SS-only breast cancers that were coded as ever-married varied showed limited variation by SEER area. The PPM was lower for cervix than breast in each SEER area, although the figures were close for New Mexico. The PPM varied from 20.5% in Los Angeles to 49.0% in New Mexico for breast, and from 9.7% to 46.4% for cervix. Geographic variation in PPM did not appear to explain variation in ASIRs. That is, Los Angeles had the lowest PPM but the highest cervical cancer rates. New Mexico had the highest PPM, but breast and cervical cancer rates were not relatively high.

Overall, 25.8% of the combined group of breast and cervical cancers were coded NOS subgroup of Hispanic rather than to a specific Hispanic subgroup (Table 2). The low proportion in New Mexico (Table 2) was largely explained by coding to the “other” Spanish surname/origin category (data not shown). Combining the 25.8% NOS with the 26.8% SS-only for the Spanish surname/origin variable (Table 2), 52.6% of all breast and cervical cancers in Hispanic women did not have Hispanic subgroup coded; this percentage was lowest in Los Angeles and highest in New Mexico.

Birthplace was coded as unknown for 34.9% of all breast and cervical cancers combined in Hispanic women although the rate of unknown birthplace was lower (22.2%) in Los Angeles than the >40% figure for each of the
other SEER areas tabulated (Table 2). The proportion of cancers coded as US-born was 25.2% and was highest in New Mexico (42.0%) and lowest in Atlanta (14.2%) (Table 2).

**DISCUSSION**

The large geographic variation in breast and cervical cancer incidence rates (Table 1) is consistent with the findings from a study² of ASIRs per 100,000 Hispanic women per year from the 1970s to 1992. In a NAACCR report on ASIRs for Hispanics in 1995–2000 for 17 areas covering 85% of the US Hispanic population, SEER areas were represented at or near both extremes (low and high) for both types of cancer.³ Included were some SEER areas (but not Connecticut, Seattle-Puget Sound, Utah, or Atlanta) as well as non-SEER areas (eg, Florida, New Jersey, and New York).³ The ASIRs per 100,000 Hispanic women per year were 89.2 for breast and 16.0 for cervical cancer;³ these rates were similar to the ASIRs for all 11 SEER areas combined in the present study (Table 1).

Percentage potentially misclassified (PPM) as Hispanic, because maiden name was frequently absent in cancer registries, was <22% (except in New Mexico), and variation in PPM (Table 2) did not appear to explain the geographic variation in ASIRs (Table 1). However, the actual proportion of cancers diagnosed among patients with missing maiden name in each registry was unknown because the SEER variable “computed ethnicity” was not included on the SEER public-use data file¹⁷; this variable indicates whether or not maiden name was examined or was missing for each cancer.¹⁸ Future studies should consider this variable for both SEER and non-SEER registries. In the NAACCR Hispanic algorithm, for registries with a large (but unspecified) percentage of missing maiden names, no maiden names are searched, and the surname alone is used to classify Hispanic ethnicity for records of all patients not reported as Hispanic by hospitals.³ Validation studies suggest that absence of maiden name in the cancer registry results in some overestimation of ASIRs for Hispanic women.⁶¹⁹²⁰ Any misclassification of Hispanic ethnicity did not obscure the finding that some geographic variation in incidence rates

---

**Table 1. Average annual age-standardized invasive breast and cervical cancer incidence rates (ASIRs) per 100,000 Hispanic women in 1995–2000, in the SEER (Surveillance, Epidemiology and End Results) Program, by SEER area**

<table>
<thead>
<tr>
<th>Hispanic Population*</th>
<th>Breast</th>
<th>Cervix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poverty Rate %</td>
<td>Foreign-born %</td>
<td>N</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>11 Areas Combined³</td>
</tr>
<tr>
<td>Los Angeles County Calif</td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Francisco-Oakland Calif</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Mexico</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicago, Ill</td>
<td>24%</td>
<td>10%</td>
</tr>
<tr>
<td>Connecticut</td>
<td>21%</td>
<td>17%</td>
</tr>
<tr>
<td>Utah</td>
<td>20%</td>
<td>16%</td>
</tr>
<tr>
<td>Atlanta Ga</td>
<td>16%</td>
<td>48%</td>
</tr>
<tr>
<td>Seattle-Puget Sound Wash</td>
<td>15%</td>
<td>20%</td>
</tr>
</tbody>
</table>

* Poverty rate and proportion foreign-born for the Hispanic population in each SEER area, from the 1990 Census.¹
† Rates shown are age-standardized using the age distribution of the 2000 US standard population (see text).
‡ Data for three areas (Detroit, Mich; Hawaii; and Iowa) of the total of 11 SEER areas are not shown separately, due to small numbers of cancers.
§ Rate is lower than that for 11 areas combined and 95% confidence intervals do not overlap.
|| Rate is higher than rate for 11 areas combined and 95% confidence intervals do not overlap.
CI = confidence interval.
exists, consistent with variation in SES, such as the contrasting ASIRs for Los Angeles vs San Francisco-Oakland or San Jose-Monterey (Table 1).

Comparison of ASIRs by Hispanic subgroups was precluded by the incompleteness of data in the SEER registries (Table 2). The high proportions for both unspecified Hispanic ancestry and unknown birthplace probably result in bias in the distributions of known ancestry and birthplace in each SEER area. However, the low percentage of foreign-born in the Hispanic population (Table 1) and among breast and cervical cancers (Table 2) in New Mexico could explain the low ASIR for cervical cancer, despite the high Hispanic poverty rate in the state’s Hispanic population (Table 1). Pap screening rates are lower, and cervical cancer mortality rates are higher, among foreign- than US-born women. Population rates of hysterectomy also affect cervical cancer incidence rates (because the tissue at risk is removed) and are highest in the South, but Atlanta, Ga, was the only SEER area located in the South. Childbearing rates are lower for young US women of Puerto Rican than Mexican origin, which could increase the risk of breast cancer in areas (such as Connecticut) with predominantly Puerto-Rican-origin Hispanics.

The SEER Program was expanded in 2001 to include four new geographic areas (ie, New Orleans, La; Kentucky; New Jersey; and the remainder of California not already covered), resulting in coverage of 40% of the US Hispanic population. Future analyses of the SEER database should be able to provide more accurate estimates of cancer incidence and mortality among US Hispanic women.

Table 2. Percentage coded as Hispanic based only on Spanish-surname (SS-only), percentage of SS-only coded as ever married, percentage of all cancers potentially misclassified (PPM) as Hispanic,* percentage coded as Hispanic not otherwise specified (NOS), and percentage coded as unknown birthplace, among incident breast and cervical cancers diagnosed in 1995–2000 coded as Hispanic

<table>
<thead>
<tr>
<th></th>
<th>Breast</th>
<th>Cervix</th>
<th>Breast and Cervix‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SS-Only</td>
<td>All</td>
<td>Married</td>
</tr>
</tbody>
</table>
| N‡ | %§ | %|| | % | N‡ | %§ | %|| | % | % | % | % | %%
| 3126 | 29.4 | 55.6 | 16.3 | 11 SEER Areas Combined‡ | 26.8 | 25.8 | 24.7 | 34.9 |
| 1116 | 20.5 | 54.5 | 11.2 | Los Angeles County Calif | 18.0 | 25.4 | 22.2 | 22.2 |
| 463 | 35.3 | 58.7 | 20.7 | San Francisco-Oakland Calif | 33.4 | 28.9 | 45.2 |
| 289 | 29.9 | 58.1 | 17.4 | San Jose-Monterey Calif | 28.0 | 26.6 | 28.1 | 43.1 |
| 759 | 49.0 | 54.4 | 26.7 | New Mexico | 48.8 | 16.4 | 42.0 | 52.9 |
| 231 | 41.6 | 51.1 | 21.3 | Connecticut | 39.6 | 24.5 | 15.2 | 47.3 |
| 64 | 28.2 | 57.8 | 16.3 | Utah | 27.0 | 51.1 | 24.1 | 61.5 |
| 59 | 34.7 | 62.7 | 21.8 | Atlanta Ga | 31.4 | 36.6 | 14.2 | 68.6 |
| 63 | 37.7 | 55.6 | 21.0 | Seattle-Puget Sound Wash | 37.4 | 30.0 | 26.8 | 54.7 |

* Percentage of SS-only that were coded as ever married (see text).
‡ For total numbers of cancers, see Table 1.
§ Percentage of all cancers (see Table 1 for total cancers coded as Hispanic).
|| Percentage of SS-only cancers with marital status coded as ever married (see text).
¶ Data for three areas (Detroit, Mich; Hawaii; and Iowa) are not shown separately, due to small numbers of cancers.

Percentage potentially misclassified (PPM) as Hispanic, because maiden name was frequently absent in cancer registries, was <22% (except in New Mexico) . . .

Ethnicity & Disease, Volume 15, Autumn 2005
to include Hispanic populations in these areas.

In conclusion, the study findings indicate that some of the geographic variation in breast and cervical cancer ASIRs is consistent with sociodemographic variation. Other findings indicate that educational programs are needed for hospital staff on the importance of collecting information on patients’ self-reported Hispanic ethnicity, ancestry, birthplace, and maiden name.27,28 Ideally, the developing nationwide system for cancer surveillance and control, which has produced cancer incidence data for White and Black race,29 should eventually include incidence rates for all Hispanics and for Hispanic subgroups for each geographic area. This method would provide data that could facilitate tailoring interventions to specific subgroups and specific geographic areas.30

The elimination of disparities in cancer by race/ethnicity is a major goal of various US agencies and organizations, and SEER incidence data have been used in assessing disparities.10,26,30,31 However, Hispanic subgroups differ in all-cause mortality rates,32 cancer mortality rates,33–36 prevalence of various cancer risk factors10,30,37 and use of cancer screening tests.38 Improving the cancer registry databases would produce better data on the burden of cancer among Hispanics by including ASIRs by Hispanic subgroups. These data would also enhance the interpretation of geographic variation in cancer rates for the heterogeneous US Hispanic population.

ACKNOWLEDGMENTS
This work was supported in part by Contract N01-PC-356133 between the National Cancer Institute and the Connecticut Department of Public Health.

REFERENCES
30. Huerta EE. Cancer statistics for Hispanics, 2003: good news, bad news, and the need for...


AUTHOR CONTRIBUTIONS

Design and concept of study: Polednak

Acquisition of data: Polednak

Data analysis and interpretation: Polednak

Manuscript draft: Polednak

Statistical expertise: Polednak

Acquisition of funding: Polednak

Administrative, technical, or material assistance: Polednak

Supervision: Polednak