Commentary

Assessing Cervical Cancer Risk in Hispanics

Gloria D. Coronado, 2 Thomas D. Koepsell, Beti Thompson, Stephen M. Schwartz, Rebecca S. Wharton, and Julia E. Grossman

Fred Hutchinson Cancer Research Center, Seattle, Washington 98109 [G. D. C., B. T., S. M. S., R. S. W., J. E. G.], and Departments of Epidemiology [G. D. C., T. D. K., S. M. S.] and of Health Services [B. T.], University of Washington, Seattle, Washington 98195-7234

Abstract

Population-based cancer registries rely on various methods to assign Hispanic ethnic identifiers to patients in the registry. The methods may result in misclassification of patient ethnic identities. Such misclassification may obscure the real incidence of cervical cancer among Hispanic women. This review summarizes previous literature on the accuracy of methods used to ascertain Hispanic ethnicity in numerator and denominator data for the calculation of cancer incidence. In addition, cancer registry ethnicity ascertainment methods were examined for six United States states (California, Florida, Illinois, New Mexico, New York, and Texas) that have a high proportion of Hispanics. The percentage of persons classified as Hispanic who self-identified as Hispanic (predictive value positive) in various reported studies ranged from 54 to 76% for women. The accuracy of ethnicity assignments based on either the United States census list or the Generally Useful Ethnic Search System (GUESS) program show slight differences in percentages of self-identified Hispanics who were classified as Hispanic (sensitivity among women: 62–80% for 1980 United States census list, 63–82% for GUESS program). Higher sensitivity and lower predictive value positive is achieved with a greater number of sources used. In conclusion, decisions about collecting racial and ethnicity information are influenced by demographic changes, immigration trends, changes in ethnic and racial identity, legislative needs, and public policies. The rapidly growing Hispanic population and the excess incidence of cervical cancer in this population requires improving the accuracy of ethnicity information.

Introduction

Research suggests that estimates of cancer incidence among Hispanics may be overestimated for some types of cancer and underestimated for others (1, 2). One key factor thought to be associated with the bias is the methods used by cancer registries and hospitals that contribute data to registries to collect and classify Hispanic ethnicity information.

The racial and ethnic coding system of the OMB 3 describes four racial categories (white, black, “American Indian” or Alaskan Native, and Asian or Pacific Islander) and two ethnic categories (Hispanic and non-Hispanic; 3). Thus, Hispanics may be from any race. The OMB, however, provides no guidelines for how to assign race or ethnicity codes to individuals (3). In enumerating ethnic-specific cervical cancer cases, the SEER program registry generally relies on standards set forth by the North American Association of Central Cancer Registries (NAACCR; Ref. 4). The SEER program collects cancer incidence and mortality data from 11 population-based registries and three supplemental registries and covers 14% of the United States population (5). To assign ethnic identity, the NAACCR standards call for the use of all available information sources, including stated ethnicity in the medical record, stated Hispanic origin on the death certificate, birthplace, information about life history and language spoken, and last name or maiden name appearing on a list of Spanish surnames. In cases in which ethnicity cannot be determined from the available information, codes are provided to record computed ethnicity. Computed ethnicity codes indicate whether the patient’s surname or maiden name matches a standard list of Spanish surnames. The standards allow for the exclusion from consideration as Hispanic, individuals who belong to racial categories with Spanish surnames but who generally are non-Hispanic (such as some Native Americans or Filipinos; Ref. 4).

In this report, we summarize the previous literature on the accuracy of methods used to ascertain Hispanic ethnicity in numerator and denominator data for the calculation of cancer incidence. Furthermore, registry ethnicity ascertainment methods are examined for six United States states that have a high proportion of Hispanics. The six states are California, Florida, Illinois, New Mexico, New York, and Texas. These data were collected by telephone inquiries made to representatives from the cancer registries.

We chose to review the incidence of cervical cancer for several reasons. First, cervical cancer affects women of all ages; thus, incidence estimates are likely to reflect present population characteristics (such as rates of intermarriage) that impact ethnic classification. Compared with incidence of other cancers, cervical cancer incidence is relatively low; therefore, a small proportion of misclassified cases can potentially alter rates substantially. Cervical cancer occurs only in women, thus ethnicity assignments based on a match of the patient’s name with lists of Spanish surnames likely are misclassified more frequently and are more biased for women than for men. Finally, invasive cervical cancer ranks as the fourth most common type of cancer among Hispanic women in the United States (6).
Commentary: Cervical Cancer in Hispanics

from the National Cancer Institute’s SEER program registry show the incidence of invasive cervical cancer in the year 2000 to be 17.0/100,000 for Hispanics; nearly twice the 9.6/100,000 for non-Hispanic whites in the same year (5). Thus, the accuracy of calculated incidence is an issue of high public health importance.

Numerator Data

We reviewed a limited number of previous studies that examine the accuracy of methods for assigning ethnic identifiers to data used as the numerator in estimating cancer incidence. These methods include stated ethnicity in medical records, patient’s last name appearing on a list of Spanish surnames (United States Census Spanish surname lists), or letter combinations in a patient’s last name matching the letter combination in the GUESS.

The validity of these methods was examined using four measures: sensitivity, specificity, predictive value positive, and predictive value negative. For each ethnicity assignment method, sensitivity is defined as the percentage of self-identified Hispanics who were classified as Hispanic; specificity as the percentage of self-identified non-Hispanics who were classified as such. The predictive value positive is defined as the percentage of persons classified as Hispanic who self-identified as Hispanic, and predictive value negative gives the percentage of persons who were classified as non-Hispanic who self-identify as non-Hispanic. For each measure of accuracy, self-identified ethnicity serves as the criterion measure (“gold-standard”).

Medical Record

All of the state cancer registries that we reviewed relied on stated ethnicity in medical records to assign ethnicity codes to new cancer cases (7–15). They also assigned Hispanic ethnicity to patients whose surname visually matched a name on a list of Spanish surnames or whose surname was thought to be Spanish (determined by an educated guess by medical record abstractors). State registries in California, Illinois, New Mexico, New York, and Texas reported relying on medical record birthplace information (7–9, 11–14, 16). State registries in California, New Mexico, New York, and Texas also used other information, such as patient’s maiden name, parents’ surname, patient’s life history, or language spoken (7–9, 12–14, 16).

We report the concordance of Hispanic ethnicity assigned using medical record information with self-reported ethnicity information in one previous epidemiological study (Ref. 17; Table 1). The study matched ethnicity data from the San Francisco/Oakland SEER program to self-reported ethnicity as ascertained by the same questions used in the 1980 United States census. A total of 560 Hispanic and 594 non-Hispanic whites (men and women), ages 20–89, who were diagnosed with invasive or in situ cancers of the colon, lung, female breast, cervix, or prostate were included (17). The study found that among women who self-identified as Hispanic, 64% were assigned a Hispanic identifier (sensitivity = 64%). Among women who were assigned a Hispanic identifier in the registry, 76% self-identified as Hispanic (predictive value positive = 76%).

A portion of the inaccuracy of ethnicity assignments based on medical record information may be explained by missing ethnic identifiers in medical charts. Symposium proceedings of a conference on the accuracy of Hispanic-ethnicity reporting in state cancer registries found that race and ethnicity are specified with decreasing frequency on hospital admissions forms (on which ethnicity is likely to be self-reported); only 25% of medical chart abstractors interviewed in one study reported ethnicity to be consistently recorded in medical charts (9). The frequency with which ethnicity is recorded in other sections of the medical chart, such as the physicians’ or nurses’ notes (in which ethnicity information is likely to be based on provider appraisals) is unknown (9). When the stated ethnicity is unavailable in a given medical chart, undercounting of Hispanic patients occurs, and incidence based on these counts may underestimate true incidence.

Ethnic identity information in the medical record may be provided by the patient (on admissions forms) or assigned by a health care provider or hospital administrator. Hospital personnel often assign ethnicity based on their own personal appraisals of the patient’s last name, language spoken, or life experiences (such as birthplace or immigration status; Ref. 9). Assessments made by health care providers or hospital administrators are considered less accurate than those obtained directly from the patient and lead to an overcount or undercount of Hispanic cases.

Specific data items used to determine ethnicity from medical records varies by registry. Data from a survey of 26 regional cancer registries found that 3 registries used additional information (such as Spanish surnames) or general knowledge of cancer registry personnel (such as familiarity with common Hispanic last names or the geographic regions with a high proportion of Hispanics) to assign ethnic identity (9). Data from the same study showed that many abstractors relied on their own knowledge about the spelling and sound of a patient’s name to assign ethnicity rather than using the 1980 Census Hispanic surname list (9). Generally, these data are of limited accuracy; only one-third of medical record abstractors in one study reported finding place of birth (9).

Spanish Surname

Computer-generated matches of a patient’s surname to a standard list of Spanish surnames also are used to assign ethnic identifiers in cancer registries. Of the six state registries that we reviewed, all but Florida used either the 1980 or 1990 United States Census surname list to assign ethnicity codes (7–9, 11–14, 16). Two states, Illinois and New Mexico, also relied on the GUESS method for assigning Hispanic identifiers (11, 12). The use of Spanish surname as a proxy for self-reported ethnic identity holds an intuitive appeal; in New Mexico, 8 main surnames are thought to account for about 25% of the Hispanic population, and about 20 others cover almost all of the rest (11). Nevertheless, for certain subgroups of Hispanics, such as Puerto Ricans and Cubans, Spanish surname lists may inconsistently predict true ethnicity.

Previous research that has compared the correspondence of Spanish surname assignments with self-identified Hispanic ethnicity has found varying accuracy. Four studies compared ethnic identity assignments based on a match of the patient’s surname to a given United States Census surname list with self-reported ethnic identity (17, 19–21). One study (20) enrolled 1345 patients with Spanish surnames and 717 patients with non-Hispanic surnames as determined by the 1980 United States Census Spanish surname list. Patients were ages 35–74 and belonged to a health maintenance organization in Northern California. Telephone interviewing was conducted to determine the patients’ self-reported ethnicity (using the same response categories of the 1980 United States Census). When patients’ self-reported ethnic identity was compared with ethnicity as-

Downloaded from cebp.aacrjournals.org on July 9, 2017. © 2002 American Association for Cancer Research.
assignments generated from the surname list, only 70% of self-identified female Hispanics were assigned Hispanics identifiers. Furthermore, the surname list falsely assigned Hispanic identifiers to a large share of non-Hispanic whites (predictive value positive $= 56\%$).

Other studies that have compared the self-reported ethnic identity to ethnicity codes assigned using United States Census Spanish surname lists found sensitivity to range from 62 to 80% among women (17, 19, 21). This suggests that the use of Spanish surname lists alone failed to identify up to 38% of Hispanic women. In general, studies reporting both sensitivity and predictive value positive information show that ethnicity assignment methods that identify a large proportion of self-identified Hispanics (high sensitivity), falsely assign ethnicity codes to a large share of non-Hispanic whites (low predictive value positive).

Spanish surname methods may misclassify cases by assigning a Hispanic identifier to non-Hispanic members of racial groups known to have Spanish surnames [such as certain Asian groups (Filipinos) or Native American groups]. In fact, ethnicity assignments based on Spanish surname lists are reported to predict most accurately self-reported ethnicity in regions in which the Hispanic population is homogeneous or of predominant Mexican origin (17). The observed higher sensitivity of ethnicity assignments using United States Census Spanish surname lists among Hispanics in San Antonio and Albuquerque (in which the Hispanic population is relatively homogeneous), compared with assignments based on the same methods among Hispanics in Northern California and San Francisco (in which the Hispanic population is diverse and there is a large Asian population) supports this claim. Furthermore, data from one study that examined the race and ethnicity of non-Hispanic cases misclassified as Hispanic, showed Filipinos and non-Hispanic whites to represent the majority of those with a Spanish surname who were non-Latino/Hispanic (20). Howard et al. (19) showed that Italians represented a large percentage of persons misclassified as Hispanic when a given surname list was used. As a result, many state cancer registry programs eliminate from consideration as Hispanic, certain Native American groups and Filipinos. This practice may become problem-

### Table 1 Accuracy of methods of assigning Hispanic ethnic identity in epidemiological studies, compared with self-reported ethnicity

<table>
<thead>
<tr>
<th>Method and study</th>
<th>Sample size</th>
<th>Identified as</th>
<th>Identified as</th>
<th>Sensitivity$^a$</th>
<th>Specificity$^b$</th>
<th>PV$^+$$^c$</th>
<th>PV$^-$$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical record</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Francisco/Oakland (17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>171</td>
<td>185</td>
<td>49</td>
<td>99</td>
<td>80</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>389</td>
<td>409</td>
<td>64</td>
<td>98</td>
<td>76</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>1980 United States Census surname lists Albuquerque (19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>253</td>
<td>528</td>
<td>85</td>
<td>95</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>380</td>
<td>511</td>
<td>79</td>
<td>90</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Northern California (KPMCP) (20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>627</td>
<td>343</td>
<td>88</td>
<td>96</td>
<td>68</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>718</td>
<td>374</td>
<td>70</td>
<td>94</td>
<td>56</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>San Francisco/Oakland (17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>171</td>
<td>185</td>
<td>58</td>
<td>98</td>
<td>74</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>389</td>
<td>409</td>
<td>62</td>
<td>97</td>
<td>68</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>Northern California (21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>324</td>
<td>NA</td>
<td>89</td>
<td>NA</td>
<td>83</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>402</td>
<td>NA</td>
<td>80</td>
<td>NA</td>
<td>74</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>GUESS program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Francisco/Oakland (17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>171</td>
<td>185</td>
<td>58</td>
<td>97</td>
<td>60</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>389</td>
<td>409</td>
<td>63</td>
<td>95</td>
<td>54</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>Albuquerque (19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>253</td>
<td>528</td>
<td>87</td>
<td>92</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>380</td>
<td>511</td>
<td>82</td>
<td>85</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>1980 United States Census surname list and GUESS program Albuquerque (19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>253</td>
<td>528</td>
<td>90</td>
<td>97</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>380</td>
<td>511</td>
<td>84</td>
<td>91</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Medical record and 1980 United States Census surname list San Francisco/Oakland (17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>171</td>
<td>185</td>
<td>59</td>
<td>98</td>
<td>68</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>389</td>
<td>409</td>
<td>73</td>
<td>96</td>
<td>65</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>Medical record, 1980 United States Census surname list, and GUESS program San Francisco/Oakland (17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>171</td>
<td>185</td>
<td>62</td>
<td>96</td>
<td>57</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>389</td>
<td>409</td>
<td>75</td>
<td>94</td>
<td>54</td>
<td>97</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Sensitivity is the percentage of self-identified Hispanics who were classified as Hispanic.

$^b$ Specificity is the percentage of self-identified non-Hispanics who were classified as non-Hispanic.

$^c$ PV$^+$, predictive value positive is the percentage of persons classified as Hispanic who self-identified as Hispanic.

$^d$ PV$^-$, predictive value negative is the percentage of persons classified as non-Hispanic who self-identified as non-Hispanic.

KPMCP = Kaiser Permanente Medical Care Program.
atic in coming years because, although most Hispanics are racially white, a large percentage, particularly those from South America or the Caribbean, may be of Asian or African origin (20).

Spanish surnames may misclassify a patient’s true ethnicity by either assigning or failing to assign Hispanic ethnicity to cases in which a person has changed a surname (usually through marriage). Data from the United States census show that the number of interracial couples has increased in recent decades, from 157,000 in 1960 to more than 1,000,000 in 1992 (22). It is likely that the number of interethic couples has also increased. The implications of these unions are thought to be extensive. A study that assessed the racial and ethnic classifications of 229,000 children of Native American and white unions found that nearly one-half identified with the race of their mother and one-half identified with the race of their father (22). Notably, these data demonstrated the greater predictive value of Spanish surname in identifying Hispanic men than women. The literature that we reviewed provides support for this claim. The difference between men and women is likely related to increasing interethnic marriages, and the greater tendency in these cases for a woman’s married name to misclassify her ethnicity than for a man’s.

The accuracy of ethnicity assignments based on Spanish surnames depends on the list of surnames used to make assignments. In general, a Hispanic identifier will be assigned if the patient’s last name appears on the list of Spanish surnames developed by the United States Census or letter combinations in the patient’s last name appear in the GUESS program (19). Our review of the accuracy of ethnicity assignments based on the United States census list or the GUESS program show slight differences in sensitivity or predictive value positive when either method is used by itself (sensitivity among women: 79% when the 1980 United States census list used to make assignments, 82% when GUESS program is used; predictive value: 90% when the 1980 United States census list was used to make assignments, 85% when GUESS program is used). Higher sensitivity is achieved when ethnic identity assignments are based on both the United States census list and GUESS program.

Data from a study that interviewed participants selected from the Polk Directory serving Albuquerque compared self-reported ethnic identity to ethnicity codes generated by the 1980 United States Census list of Spanish surnames and the GUESS program. Participants who were listed as students in the Polk Directory were excluded from recruitment and those who identified as Native American, black, or Asian were omitted from the analysis. Data from the study show that the percentage of self-identified Hispanics who were accurately identified in the registry, to be ~80% for women when either method was used. Sensitivity rose when both the United States Census list and the GUESS program were used to assign ethnicity codes (19). Data from a separate New Mexico study that compared ethnicity assignments made by the GUESS program with ethnicity recorded on death certification showed that the GUESS program missed 12.6% of females and 6.7% of males identified as of “Spanish origin” on the death record (9). The GUESS program is thought to be tailored to the New Mexico population, and may not generalize to Hispanics in other areas.

**Combined Methods**

The accuracy of Hispanic identification based on multiple information sources is reported in few previous studies. We reviewed two studies (11, 17) that report on the accuracy of ethnicity assignments based on a single method and on a combination of methods. Data from a study of cases from the San Francisco/Oakland registry showed a trend toward greater sensitivity and lower predictive value positive with a greater number of sources used (17). This was apparent when medical record and 1980 United States census surname lists, and medical record and 1980 United States census surname lists and the GUESS program assignments were compared with assignments based solely on medical records. The same trend was evident in the findings of an Albuquerque study that compared ethnicity assignments based solely on United States census surname lists or solely on the GUESS program to assignments based on the combination of the United States census lists and the GUESS program (11).

The accuracy of methods used to assign ethnic identity to cervical cancer cases specifically and the implications of ethnic misclassification has been reported in few previous studies (11, 17). One study recruited a sample of 7132 patients ages 20–89 in a regional SEER registry who were diagnosed with cancers of the colon, lung, female breast, cervix, or prostate in 1990 to examine the accuracy of methods to assign ethnic identity. Data from the study show that the percentage of persons who self-identified as Hispanic who were classified as such was 75% when medical records were used to make the assignment, 78% when a patient’s last name appeared in the 1980 United States census list of Spanish surnames (surname method), 86% when letter combinations of the patient’s last name appeared in the GUESS program (GUESS method), and 88% when a combination of methods was used (11). Similarly, the percentage of persons classified as Hispanic who self-identified as such was 83% when medical records were used to make the assignment, 73% when the surname method was used, 66% when the GUESS program was used, and 64% when a combination of methods was used.

**Effects on Incidence**

Two studies compared Hispanic invasive cervical cancer incidence calculated using ethnicity assignments computed using self-identified ethnicity (11, 17). The findings of both studies suggest that the use of a single method may result in an underestimate of incidence. An analysis of one study that used data from the San Francisco/Oakland SEER registry showed that, compared with self-reported data, ethnicity codes assigned using only medical record information only slightly underestimated Hispanic cervical cancer incidence (invasive plus in situ incidence: 52/100,000 versus 53/100,000; Ref. 17). When a match of the patient’s last name to the 1980 United States census list of Spanish surnames was used in addition to the medical record, a substantial (but nonsignificant) overestimation in incidence was found (66/100,000 versus 53/100,000). The study concluded that the use of medical record information in combination with the Spanish surname provided the least amount of bias in overall incidence estimates, although the bias varied by cancer type.

A second study used data from the Illinois Department of Public Health and compared incidence estimates using three methods of assigning ethnic identity (11). One method (method 1) uses medical record information (stated ethnicity, birthplace, and surname). A second method (method 2) uses ethnicity computed by a match of the patient’s last name to a list of Spanish surnames (with exclusion of cases belonging to certain racial groups or having been born in regions outside the United States considered to be probable non-Hispanic). The third
method is the union of method 1 and method 2. When cervical cancer incidence among Hispanics was calculated using the three methods for the years 1989–1994, the lowest incidence was reported when method 1 was used (14.6/100,000 in 1994). Only a slight increase in incidence was reported when method 2 was used (14.9/100,000). The highest incidence was found when the combination of methods was used (17.1/100,000). The difference in incidence might suggest that method 1 and method 2 undercount cases of Hispanic ethnicity or that the combined method overcounts Hispanic cases.

Denominator Data
Aside from misclassification of ethnicity status and potential bias in numerator estimates, enumeration of the size of the population at risk might also be under- or overestimated. State- and ethnic-specific United States census population estimates were used as denominators for the calculation of incidence in the six state registries that we reviewed. The 1980 United States census sampling and interviewing procedures were heavily criticized for discouraging Hispanics response. Specifically, reliance on mail questionnaires (rather than in-person interviews), the limited availability of Spanish forms (by special telephone request only), an insufficient number of Spanish-speaking enumerators for callback interviews among those requesting Spanish forms, the complicated nature of the forms, and the inability of enumerators to overcome the distrust on the part of some Hispanics that census information might be used against them, were considered problematic (23). Recent research has suggested that census figures may underestimate the true size of the Hispanic population. Data from a study that assessed the accuracy of 1990 census population counts suggests that the number of Hispanics in the United States was undercounted by 410,221, representing 5% of the total number of Hispanics (24). Among the six states that we reviewed, data from this study showed the percentage underestimation to range from 2.5% for Hispanics in Illinois to 6.2% for Hispanics in Florida. Those who migrate from region to region, live in temporary housing, or have no established residence are least likely to be counted. Hispanics who cannot read English may avoid completing census forms. Undocumented Hispanics, estimated to number between four and twelve million (23), may refuse to complete government forms because of fears of deportation.

The scheme for assigning or recording ethnic identity on United States census forms has changed at various times in the past 40 years, limiting the accuracy of comparisons made from 1 decade to the next. Before 1970, the United States Census relied on indirect measures based on Spanish surname, Spanish speaking, birthplace, or birthplace of parents to identify a person of Hispanic origin (22, 25). In subsequent years, the coding system was revised: first in 1970 to allow self-identification of ethnicity and then in 1980 to allow ethnicity to be collected separately from race (22, 25). The revisions enhanced the flexibility of the coding scheme, but the inconsistency of the scheme, compared with schemes used in previous years, hindered the ability to make longitudinal population comparisons.

Ambiguous definitions of race or ethnicity in United States census forms are thought to result in inconsistently or incorrectly coded data. The 1980 and 1990 Census ethnicity coding schemes were thought to be poorly understood by Hispanics respondents. The confusion was reflected in the high percentage of Hispanics who checked “other” in the race category: 40% in the 1980 Census and 42–43% in the 1990 Census (about 9.8 million; Refs. 3, 22).

Undercount of Hispanics by the United States Census may decrease the denominator and erroneously inflate population disease rates. When numerator and denominator data related to race or ethnicity use different data collection methods, questions that differ in format or content, or rely on different definitions of race or ethnicity, it is difficult to know whether a change in rates is caused by a true change or by a change in the method of enumeration.

Published incidence for all ethnic groups ignores the fact that women who have had a hysterectomy (in which the cervix has been removed) technically are no longer at risk for cervical cancer. Thus, the denominators overestimate the true size of the population at risk, and the resulting rates are artificially low. When the proportion of Hispanic women who have had a hysterectomy matches the proportion of non-Hispanic white women who have had one, the incidence for each group would be equally underestimated. If a hysterectomy is performed less frequently among Hispanics than among non-Hispanic whites, the net result is that the true difference in cervical cancer incidence in the two groups is less than we report.

Use of Self-Reported Ethnicity as the Gold Standard
One final consideration in assessing issues of Hispanic ethnicity assignment is that self-reported ethnic identification may be influenced by local terminology. The terminology to define “Hispanic” is complex and has changed in the past decade in response to political and secular trends. The term Hispanic was originally defined by the federal OMB to include persons of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish culture or origin, regardless of race (6). The terms “Latino,” “Chicano,” “Raza,” “Mexican,” “Mexican-American,” “Mestizo,” “Puerto Rican,” “Spanish-American,” and “Cuban” have also been used, sometimes interchangeably, to mean “Hispanic.” Nevertheless, the term is believed to be controversial, because it emphasizes Spanish origin and ignores Indian (Mayan, Aztecan, and so forth) descent. The terminology is thought to vary by region (9).

The difficulty of ascertaining Hispanic ethnic identity also is influenced by a phenomenon called “ethnic flux.” The phenomenon suggests that self-identification may shift in response to political trends and personal preferences. Persons of mixed heritage, for example, at times may decide to mark two or more racial or ethnic categories on census or health surveillance instruments, and at other times may choose fewer. Research data from a United States Census Current Populations Survey found that 34.3% of households whose members were interviewed for 2 consecutive years reported having different ethnic identities from year to year (3). Data from the United States Census show that 5–11% of persons who report Spanish origin in a Census or survey will report a non-Spanish origin when reinterviewed (3). Certain demographic characteristics are shown to be associated with consistent reporting of Hispanic ethnicity, such as non-United States place of birth, and classification as Mexican, Puerto Rican, or Cuban origin versus Central American, South American, or other Hispanic origin (23).

Conclusions
In general, decisions about how racial and ethnicity information is collected is influenced by demographic changes, immigration trends, changes in ethnic and racial identity, legislative needs, and public policies (22). The rapidly growing size of the Hispanic population as well as the excess incidence of cervical
cancer in this population suggest that improving the accuracy of ethnicity information will become increasingly important. To improve ethnicity assignments in cancer surveillance systems, various steps are recommended. These include: (a) require self-reported ethnicity information on medical records using coding schemes to be consistent with the coding scheme of the United States Census. This approach would improve the accuracy of ethnic identity assignments made using medical records; (b) when medical record information is unavailable or incomplete, standardize the methods for ascertaining ethnic identity, such as using specific data sources or defining the decision rules for when a given assignment takes precedence; (c) provide comprehensive training to the appropriate hospital staff that record ethnicity information. This recommendation is considered essential for accurate data reporting (2, 9); and (d) standardize formats for reporting ethnic identity in scientific journals to facilitate the use of standard formats for the collection of ethnicity information.

References
10. MacKinnon. Florida Data System, Division of Cancer Prevention and Control; Sylvester Comprehensive Cancer Center, University of Miami School of Medicine. 2000.