Report of the Ad Hoc Working Group on Cancer Surveillance and Data Systems

November 14-15, 1999
Houston, TX

A number of the facts in this report currently are being verified and additional changes and updates will be made to the document as needed.

Table of Contents

Executive Summary

The purpose of this meeting in Houston, TX, was to provide an overview of the state-of-the-science of cancer surveillance, particularly regarding data on population groups defined by sociodemographic and socioeconomic attributes such as race, ethnicity, place of residence, economic status and class, and culture. The meeting also involved the discussion of questions concerning improving surveillance among these populations that were raised at the Intercultural Cancer Council (ICC) Data Working Group meeting held September 25-26, 1999, in Washington, DC.

STATE-OF-THE-SCIENCE OF CANCER SURVEILLANCE

This portion of the meeting included presentations on the following topics:

- ✧ The National Cancer Institute Group on Medically Underserved.
- ✧ HHS data improvement initiatives.
- ✧ North American Association of Central Cancer Registries (NAACCR) data improvement initiatives, including NAACCR's Data Evaluation and Publication Committee (DEPC) and Collaborative Research Working Group (CRWG), and the status of the 1996 Hispanic Data Symposium recommendations.
- ✧ Regional and agency data improvement initiatives, including efforts of the Surveillance, Epidemiology and End Results (SEER) Program registries in New Mexico, Northern California, Hawaii, and Los Angeles; state registries in Kentucky and Texas; the Centers for Disease Control and Prevention (CDC); the Indian Health Service (IHS); the American Cancer Society (ACS); and the American College of Surgeons/Commission on Cancer.
- ✧ Report from the Bureau of the Census on the Impact of Changes in Racial/Ethnic Categories on Data Reporting.
- ✧ Conceptual Issues, Measurement and Data on Culture, Acculturation, and Cancer Control.
- ✧ The Role of Community Based Organizations (CBOs) in Data Collection, Use, Reporting, and Quality.
- ✧ Report from the National Center for Health Statistics, including Accessing Data Through the Resource Center, and Future Directions for Integrated Federal Health Data System and State/Regional Data.
- ✧ NCI Surveillance Activities Related to Sociodemographic Disparities in Cancer.

PRIORITIZATION OF THE ICC DATA WORKING GROUP RECOMMENDATIONS

Following these presentations on the state-of-the-science of cancer surveillance, participants were asked to prioritize the 13 recommendations developed by the ICC Data Working Group at the September meeting. The prioritization was accomplished in two phases. In the first phase, participants were asked to identify the most important recommendations. Five of the 13 recommendations were rated as the most important by the participants. In the second phase, participants were asked to prioritize these five

important recommendations by doability. Two of the five most important recommendations were identified as the most doable by the participants.

DEVELOPMENT OF ACTION PLANS TO IMPROVE DATA COLLECTED FOR USE IN MEASURING CANCER IN POPULATION SUBGROUPS

After prioritizing the recommendations, two breakout groups were formed to discuss actions that could be taken to improve cancer surveillance among ethnic minorities and the medically underserved. Action plans were developed for the following:

- Collecting SES information for cancer patients (Recommendation 1).
- Training epidemiologists and related professionals from minority and medically underserved populations (Recommendation 7).
- Promoting community involvement for culturally appropriate cancer control materials and instruments (Recommendation 12).
- Developing meaningful numerator/denominator statistics (Recommendation 2).
- Promoting collaboration among federal agencies collecting cancer data to provide comprehensive national data (Recommendation 11).

The actions plans do not represent the consensus of the meeting participants, rather they present a summary of the discussions that took place during that session. Participants' comments are captured in the meeting summary and should serve as an excellent basis for discussion of this topic at future meetings. The outcome of the November meeting was a key topic on the agenda of the ICC meeting, which was held in Washington, DC, on February 9-13, 2000.

The final session of the meeting was devoted to discussing how the various organizations involved in cancer surveillance could work together more effectively to implement the action plans.

Report of the Ad Hoc Working Group on Cancer Surveillance and Data Systems

November 14-15, 1999
Houston, TX

WELCOME AND PURPOSE OF THE MEETING

Dr. Gilbert Friedell, Chair of the meeting, welcomed the participants. He conveyed that the purpose of this meeting was to provide an overview of the state-of-the-science of cancer surveillance and to discuss the questions that were raised at the Intercultural Cancer Council (ICC) Data Working Group meeting held September 25-26, 1999, in Washington, DC. Dr. Friedell and **Dr. James Hampton** were responsible for organizing and chairing the September ICC meeting, and they worked closely with surveillance organizations, particularly the National Cancer Institute (NCI), to plan and organize this meeting in Houston. Dr. Friedell thanked **Dr. Brenda Edwards**, Associate Director of the Surveillance Research Program at NCI, for her assistance in planning and organizing the meeting. He also acknowledged the various organizations that shared the expenses of the meeting, including the NCI, ICC, Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS), American Cancer Society (ACS), and the American College of Surgeons (ACoS).

Dr. Friedell explained that, during the November meeting, several breakout groups were formed to discuss actions that could be taken to improve cancer surveillance among ethnic minorities and the medically underserved. Participants' comments are captured in this summary and should serve as an excellent basis for discussion of this topic at future meetings. The outcome of the November meeting was a key topic on the agenda of the ICC meeting, which was held in Washington, DC, on February 9-13, 2000.

STATE-OF-THE-SCIENCE OF CANCER SURVEILLANCE

Report From the July 1999 Meeting of the NCI Group on Medically Underserved

Dr. Robert Hiatt provided an update on the activities of the National Cancer Institute (NCI) to define medically underserved. One of the outcomes of the Institute of Medicine (IOM) report entitled *The Unequal Burden of Cancer* (1999), was a recognition that it was difficult to understand how much of NCI's research effort was devoted to the areas of ethnic minorities and the medically underserved. This difficulty is attributed, in part, to the lack of agreement on the definition of "medically underserved". Dr. Klausner (Director, NCI) asked Dr. Otis Brawley to convene a group of experts, which included representatives from the ICC, to study this issue. The first of several meetings of this group was held on July 23, 1999, and participants focused on questions such as: What is the concept? What is the best term? How do you measure it? What metrics should be used to track NCI's efforts in this area?

Dr. Hiatt indicated that the term "medically underserved" was used initially by federal agencies to understand medical care delivery in terms of level of insurance and level of access to care. It was used to differentiate between those who had the capability to access and receive care and those who did not. Dr. Hiatt pointed out that the definition of the medically underserved has broadened since then. During the July meeting, many difficulties associated with defining medically underserved were identified. The

group recognized that who is underserved changes for different diseases—cancer, heart disease, and diabetes—and that, for cancer, who is underserved changes depending on the stage in the cancer process. There was general agreement that medically underserved is associated with socioeconomic status (SES), and the definition of this concept is close to that for medically underserved. The group discussed potential measures for medically underserved. It was recognized that part of being underserved is a structural or systematic problem (i.e., the health care system has not made available certain resources to those in need); another part is that some individuals choose to be medically underserved (e.g., they choose not to get screened even though they have access to health services). Dr. Hiatt noted that both of these factors must to taken into consideration when defining medically underserved. He mentioned that there is time to provide input to this group before it meets again in January.

Another group was formed to focus on the coding of special populations, because the IOM report questioned the method used by the NCI for categorizing special populations. This group has prepared a draft report, which is not yet available. The report recommends that the NCI code not only the categorization of ethnic minorities, but also low literacy, low income, and rural or inner city status. These elements are similar to those that would be used to identify the medically underserved. Should we separate ethnic minorities and special populations from the medically underserved? The concept of medically underserved is not an independent variable; it is so dynamic and changeable that it is difficult to define some group as being underserved. Is medically underserved a useful categorization over time? The greater concern is defining subpopulations that are more at risk for unwanted outcomes. The group continues to grapple with this intellectually challenging task and is working to develop a definition that is practical and workable.

Discussion

Dr. Jerome Wilson suggested that it might be informative to define medically served. Such an effort may help the group to develop a clearer definition of what it means to be medically underserved. Dr. Marjorie Kagawa-Singer, who served on the NCI group defining medically underserved, mentioned two definitions: (1) those who cannot afford healthcare, and (2) those who encounter barriers to optimal healthcare that are not of their own choosing. She noted that the second definition, which is used by the ACS in California, is more inclusive than the first. Dr. Friedell mentioned that the term “medically underserved” has been useful for the population with which he has been working—rural, poor white people. This subpopulation does not have any distinguishing characteristics by which they can be aggregated, and it is larger than any ethnic minority population. Dr. Friedell noted that the term may become less useful in the future if SES and related information are collected routinely. He expressed some concern about “lumping” medically underserved with SES, because SES does not include the cultural differences of ethnic minorities. If the rural, poor white population could be considered an ethnic minority, the term medically underserved would not be required.

Report and Discussion From the HHS Data Improvement Initiatives

Dr. Brenda K. Edwards opened her presentation by acknowledging the life and work of Eleanor Nealon, a valued colleague at the NCI. There has been considerable discussion and activity with regard to health data at the Department of Health and Human Services (DHHS) level. She mentioned two major initiatives that have commanded the time and attention of many individuals during the past several years. The first is the HHS Initiative to Eliminate Racial and Ethnic Disparities in Health (including cancer). As a result of this initiative, two groups have been convened—the HHS Data Council’s Working Group on Racial and Ethnic Data and the Data Work Group for the HHS Initiative to Eliminate Racial and Ethnic Disparities in Health—to examine the data needs and how to measure the disparities, and to make recommendations that address the data and intervention components. The deliberations of these two

groups are being pooled in a report that will be available by January 2000. The other initiative at the federal level is Healthy People 2010. The issues that we are discussing at this meeting are not new, many committees have discussed and examined them, and many reports have been written that address racial and ethnic data problems, including:

- 1985 ***Report of the Secretary's Task Force on Black and Minority Health***.
- 1992 ***Improving Minority Health Statistics***, Public Health Service Task Force on Minority Health Data.
- 1992 ***Enhancing Health Statistics for the American Indian and Alaska Native Communities: An Agenda for Action***, Report to National Center for Health Statistics by the Indian Health Care Association.
- 1993 ***Recommendations to the Surgeon General to Improve Hispanic/Latino Health***, Surgeon General's Hispanic/Latino Health Initiative.
- 1995 ***Recommendations for Improving the Health of Asian Americans and Pacific Islanders***, Report by delegates to the First National Health Summit of Asian American and Pacific Island Health Organizational Leaders.
- 1996 ***Hispanic Agenda for Action: Improving Services to Hispanic Americans, A Report to the Secretary***, Departmental Working Group on Hispanic Issues.

These reports were examined by the two data groups mentioned above in an effort to synthesize and describe what has been learned in the past. Although the focus of their report will be implementation and operation, these two groups do not discount the role of policy in addressing the health disparity issue. In 1997, the DHHS issued an inclusion policy statement requiring, with few exceptions, that racial and ethnic data be collected for HHS-maintained data collection systems. Dr. Edwards mentioned that, despite this departmental policy, about 10 percent of all federal data systems do not collect racial and ethnic data at the present time. Other groups for which data need to be collected and reported (e.g., special populations based on age, gender, sexual orientation, and disabilities) are included under this policy statement. A group chaired by Olivia Carter-Pokras (Office of Minority Health) and Tony D'Angelo (Indian Health Service) is preparing a general report and a detailed implementation plan for the HHS to improve the collection and use of data for racial and ethnic minority populations. In preparing this report, a number of recurring issues have been identified, including:

- Funding.
- Infrastructure.
- Policy and legislation that support the collection and use of data.
- Cooperative efforts are required among federal, state, and local agencies, private groups, researchers, etc.
- OMB identifier and subgroups have always driven the minimum data requirements for data collection.
- Evaluation, methodology, and data quality are key factors.
- Oversampling or target surveys (how do you get data on small groups?).
- Content and relevance (what are you trying to measure and for what purpose?).
- More analysis, publication, dissemination, and use are needed.
- Training and technical assistance are needed to enhance data use.

- ✧ Representativeness of data.

The following 10 overarching, high-priority recommendations will be included in the report on improving collection and use of racial and ethnic data:

- ✧ A strategic plan for improving the collection and use of data on ethnic minorities is needed.
- ✧ Health Insurance Portability and Accountability Act standards.
- ✧ Healthy People 2010 is key to data analysis and collection.
- ✧ Registries (for cancer and other diseases) are important.
- ✧ State-based data collection that goes beyond and complements the federal effort is needed.
- ✧ Support for data analysis and research is needed.
- ✧ Accessibility to data is important.
- ✧ More reporting of national data is needed.
- ✧ Training is needed.
- ✧ Dissemination to racial/ethnic communities impacted by the data is needed.

With regard to the recommendation on data collection, the group stressed the following: geocoding is needed, data should be collected to examine the impact of program interventions, all federal agencies should adhere to the inclusion policy (i.e., all federal agencies should collect the minimum set of data), more national conferences are needed on data collection and use, sociocultural data should be collected (but the group did not fully define this), more culturally appropriate surveys should be conducted, use of electronic patient records should enhance data collection efforts, and surveys should include Puerto Rico and the U.S. Pacific Territories.

In the area of data analysis and interpretation, the group stressed the need for: culturally sensitive analysis; reporting, referral, and treatment patterns; sociocultural analyses; and reduction of misclassification of death rates by training or data linkage. The group also recommended addressing analysis and communication problems associated with the new age-adjustment standard; problems associated with the new OMB standards for maintaining, collecting, and presenting federal data on race and ethnicity; and problems associated with classifying death using ICD-10 (International Classification of Disease version 10).

For the data dissemination and use recommendation, the group emphasized the need for more data retrieval systems, more data centers, and Web sites such as the one for the HHS Initiative to Eliminate Racial and Ethnic Disparities in Health (**http://raceandhealth.hhs.gov**) to communicate data.

With regard to the data research and maintenance recommendation, the group stressed the following needs:

- ✧ Research on reporting of race and ethnicity.
- ✧ Continued work on the use of administrative records.
- ✧ Limitations of telephone survey methods should be addressed.
- ✧ More work in the area of questionnaire design.
- ✧ Improvement of death certificate accuracy.

- ✧ Record linkage.
- ✧ Post-censal estimates.
- ✧ Improved measures of income and wealth.

Dr. Edwards noted that public health research into the reasons for racial and ethnic health disparities has focused largely on differences in SES. Although lower SES is probably the most powerful single contributor to premature morbidity and mortality, the association between race/ethnicity and SES is complex and cannot fully explain differentials in health status. Dr. Edwards also mentioned that although minority health researchers urge attention to the influence of SES factors, especially income, they caution against using SES as a surrogate for race/ethnicity.

Steps have been taken to implement a number of the recommendations made in the reports mentioned earlier. Although many resources have been devoted to making improvements in racial and ethnic data, much of the progress remains undocumented or unnoticed. Dr. Edwards closed by acknowledging that issues still exist, much work is being done to address those issues, improvement is possible, and working together is essential. It must be a "mutual endeavor" and "shared response."

Discussion

One participant asked Dr. Edwards to expand on what efforts have been undocumented or unnoticed. She responded by providing two examples. The first example involved the inventory of data systems for racial/ethnic groups prepared by the Office of Minority Health (**http://www.omhrc.gov**). Many of these databases are not well known and are used only by a small group of individuals. The second example concerned the involvement of the NCI in the update of the National Longitudinal Mortality Study (NLMS) data set. NCI's involvement was the result of interaction between Dr. Edwards and Dr. Harry Rosenberg at a National Coordinating Council for Cancer Surveillance (NCCCS) meeting. This interaction also led to working together to update the Current Population Survey (CPS) linked to the National Death Index.

Another participant pointed out that although agencies such as the National Center for Health Statistics (NCHS) collect and store data on the minimum OMB categories of race/ethnicity, the agencies only publish data on white, black, other, and sometimes Hispanic populations. Because the detailed data on race/ethnicity are not published in reports, many assume that the data are not collected.

Dr. Edwards commented that there is a movement among the federal government for collecting data on individuals that will characterize their SES. She pointed out, however, that this is complicated in that there are legitimate counterforces against linking socioeconomic data with data on the medical services being received by the individual. More discussion on how this linkage can be implemented is needed. Dr. Friedell suggested that SES data collection efforts could start with education level. This information would be useful and its collection should be less controversial than collection of income data.

NAACCR Data Improvement Initiatives

The North American Association of Central Cancer Registries

Dr. Holly Howe described the quality and availability of the North American Association of Central Cancer Registries (NAACCR) data (**http://www.naaccr.org**). She explained that NAACCR was established in 1987 and its members include all Surveillance, Epidemiology and End Results (SEER) program (**http://seer.cancer. gov**) participants, nearly all National Program of Cancer Registries (NPCR) registries (**http://www.cdc.gov/CANCER/npcr/register.htm**), all Canadian provinces/territories, all

national cancer and registration organizations (e.g., NCI, CDC, ACS [**http://www.cancer.org**], ACoS [**http://www.facs.org**], Statistics Canada and Health Canada, and National Cancer Registrars Association [**http://www.ncra-usa.org**), and interested individuals. NAACCR establishes and maintains standards, provides education and training to those who collect and use data, certifies registries that meet standards of high quality, aggregates and publishes data, and promotes use of registry data. NAACCR is successful because cancer registry staff participate in developing and setting standards, training, and data aggregation and evaluation activities. Dr. Howe noted that this approach significantly improves compliance and participation.

NAACCR registries have information found in a medical record (e.g., medical information, age, sex, race, residence at diagnosis, stage of disease, and history at facility). They do not have information on income, education, genetic data, risk factors, residential history, or quality of life. NAACCR registries conduct surveillance activities and the data are used to identify cases for in-depth studies and high-risk populations for cancer control efforts, evaluate cancer interventions, and develop policies and program plans. These registries cannot do in-depth analyses, prevent cancer alone, or produce rates for all populations in all areas without working with partners. NAACCR registries work to define numerators, but they rely on partners to provide accurate and complete denominators.

In the past 5 years, the following information needs have been identified by NAACCR:

- ✧ Cancer rates by SES.
- ✧ Urban-rural variation in cancer.
- ✧ Cancer burden in all U.S. areas.
- ✧ Increase cancer information for more race and ethnic groups.
- ✧ Patterns of care, risk, quality of life.
- ✧ How to increase public access to cancer data.

Dr. Howe noted that some of the challenges faced by NAACCR include: confidentiality (need to balance between data use and protecting the individual's right to privacy), denominators for ethnicity and race (need accurate counts of populations at risk), and demonstrating quality of registry data. One of the more recent initiatives of NAACCR is to certify registries. NAACCR's Registry Certification Committee has established objective indicators for data outcomes to evaluate the quality of registry data. The Committee provides confidential feedback to registries (that participate voluntarily) so that they can continue to improve the quality of their data. In addition, NAACCR recognizes those registries that have demonstrated excellence. The certification criteria include:

- ✧ Completeness of case identification (Death Certificate Only cases and NAACCR method based on case fatality ratios).
- ✧ Completeness of key information (age, geography, race, and sex).
- ✧ Accurate record linkage and case consolidation.
- ✧ Accurate data.
- ✧ Timeliness.

In 1999, NAACCR certified 1996 data. To be certified, a registry must meet or exceed all six certification standards at the proscribed levels. Twenty-two registries in the United States and Canada met the gold standard of certification and nine additional registries met the silver standard of certification.

With regard to the individual criteria used in certification, a large number of registries met some of the criteria, but not all. For example, of the 50 registries having their 1996 data reviewed, 37 were 90 percent complete or higher, and 29 of them were at least 95 percent complete. Forty-five registries had fewer than two duplicate reports per 1,000 records, and 40 had fewer than 1 duplicate report per 1,000 records. The 42 U.S. registries were evaluated for missing race identifiers (this variable is not collected in Canada). A total of 38 had a race identifier that was at least 95 percent complete, while 34 were at least 97 percent complete.

Dr. Howe indicated that the charge to the NAACCR Data Evaluation and Publication Committee is to gather data from member registries and to review, evaluate, and compile the data for publication. This Committee is responsible for collecting and publishing statistics from registries that submit data in response to the call for data. The first publication, titled *Cancer Incidence in the USA*, was issued in 1993. This publication included 1988-1989 data from 25 U.S. registries and 1 Canadian registry. Incidence data for all races, whites, and blacks were reported; no combined rates were reported. In 1994, the publication's title was changed to *Cancer in North America* (CINA). This publication included 1988-1990 data from 35 U.S. and 12 Canadian central cancer registries. Incidence data for all races, whites, and blacks were reported; combined rates were generated by pooling data from 24 U.S. and 10 Canadian registries. (Selection for the combined rate was based on self-report of data quality indicators.) In 1995, the CINA publication included 1988-1991 data from 36 U.S. and 12 Canadian central cancer registries. Incidence data for all races, whites, and blacks were reported; combined rates were generated by pooling data from 22 U.S. and 9 Canadian registries (selection was based on meeting some objective data quality indicators). In 1996, the CINA publication included 1988-1992 data from 39 U.S. and 12 Canadian central cancer registries. Incidence data for all races, whites, and blacks were reported; combined rates were generated by pooling data from 25 U.S. and 11 Canadian registries (selection was based on meeting objective data quality indicators but not passing a standard set of edits). In 1997, the CINA publication included 1989-1993 data from 32 U.S. and 10 Canadian central cancer registries. It was released in two volumes—Volume I reported incidence data for all races, whites, and blacks; and Volume II reported mortality data for all races, whites, and blacks from the NCHS. Combined rates were generated by pooling data from 19 U.S. registries.

As part of the combined rates, rates for childhood cancers were published by age and race (all race, whites, and blacks). This year, the inclusion criteria were expanded to require that all records on the file pass a standard set of edits. To meet the criteria, the file had to be free of coding errors. These stringent criteria were difficult for many registries to meet in the first year, thus reducing the number of registries included in the combined rates.

In 1998, the CINA publication reported 1990-1994 data from 40 U.S. and 12 Canadian central cancer registries. Volume I reported incidence data for all races, whites, and blacks; Volume II reported mortality data for all races, whites, and blacks. Combined rates were generated by pooling data from 22 U.S. and 6 Canadian registries. In 1999, the CINA publication reported 1991-1995 data from 42 U.S. and 12 Canadian registries. Volume I reported incidence data for all races, whites, and blacks. A section on pediatric rates was repeated and became a standard part of the annual monograph. Proportional incidence rates for the top five cancer types for blacks, American Indian/Alaska Natives, Chinese, Japanese, Filipino, Korean, Vietnamese, white, and all races were reported for the first time. Volume II reported mortality data for all races, whites, and blacks. Combined rates were generated by pooling data from 25 U.S. and 8 Canadian registries. In 2000, the CINA publication will issue two monographs—one for 1992-1996 data and a second for 1993-1997 data. Volume I in each of the monographs will present incidence data, and Volume II will present mortality data in the same format as released in the 1999 monograph. Currently under consideration is whether the race-specific standardized incidence rates can be expanded to include rates for Asian/Pacific Islanders and American Indian/Alaska Natives and also to

include comparable rates for Hispanic populations. Availability of population data and accuracy of these race and ethnic identifiers in cancer registries are the variables that will affect this decision.

Dr. Howe concluded her presentation by reporting that the quality of CINA data is progressively improving and the majority of registries meet quality standards. CINA is producing more information each year to address questions of national significance, and the geographic coverage of CINA is growing.

Discussion

Dr. LaMar McGinnis asked how NAACCR intends to encourage the use of data. Dr. Howe indicated that NAACCR's early efforts have been focused on setting standards to permit data exchange as well as the training of individuals to adhere to those standards. NAACCR now is doing more publishing and data dissemination; NAACCR also is working to increase accessibility to the data files. Dr. Jerome Wilson asked if the Canadian registries are collecting information on ethnicity because they are prohibited by law from collecting data on race. He believes we need to move away from race and towards ethnicity. Dr. Howe replied that the Canadians are looking to the United States to determine how to go forward. Dr. Tom-Orme asked why cancer rates are not reported for specific Asian groups as well as Pacific Islanders. Dr. Howe explained that NAACCR is limited to reporting rates for groups for which reliable denominators exist.

NAACCR's Data Evaluation and Publication Committee

Dr. Vivien Chen provided an overview of NAACCR's Data Evaluation and Publication Committee (DEPC) and some of its subcommittees that are conducting relevant work, particularly the Collaborative Research Working Group (CRWG). When the need to assess and evaluate the quality of certain data elements arises or when a gap in data is identified, a subcommittee or work group is formed to target those areas. Currently, the DEPC subcommittees include:

- CINA Monograph Production Subcommittee.
- Collaborative Research Working Group.
- Comparative Analysis of Incidence Rates (CAIR) Subcommittee.
- 5 other subcommittees—Data Quality Indicators, Canadian-U.S. Comparison of Rates, Public User Data File (including beta test and CINA Plus Online), and Staging.

The CRWG, which was "officially" formed in Spring 1998, was created to address the emerging need for pooled cancer statistics that best represent the cancer risk of the entire Nation. In addition, it was intended to facilitate comparative analyses of cancer rates with respect to race/ethnicity, SES, geographic variations, and urban/rural differences. The CRWG has 10 members including representatives from NAACCR, NCI, CDC, ACS, and state registries (SEER and NPCR). The overall goal of the CRWG is to provide estimates of cancer incidence statistics that are of high quality, inclusive of all racial/ethnic groups and geographic coverage, and are the best representation of the entire United States. The CRWG has five specific objectives:

- To identify the gaps in the currently available "national" estimates of cancer statistics, in particular among certain minority populations (such as Native Americans, rural African Americans, and Hispanics other than Mexican origin) and geographic areas (e.g., Appalachia and the South).
- To identify immediate and long-term needs and recommend future direction (and actions).
- To establish criteria for evaluating the comparability and the quality of data for the pooled data (criteria for inclusion in the CINA combined rates).

- To perform comparative analyses (in conjunction with the CAIR).
- To provide scientific advice and guidance to the DEPC subcommittees, especially for the CINA publication, regarding race/ethnicity, geographic region, and standard population.

The CRWG has identified gaps that exist in currently available cancer data as well as gaps associated with the CINA monograph. Although CINA covers a broader geographic area than the SEER program (approximately 40 percent of the U.S. population is covered in 1999 CINA), CINA reports no cancer incidence or mortality rates for populations other than all races, whites, and African Americans.

One of the immediate actions identified by the CRWG in 1998 was to expand the CINA monograph. The incidence volume was expanded to include a detailed description of ethnicity representation in the population covered by pooled registries. In addition, it would include the top five cancer sites (proportional incidence rate) by ethnic group (all, white, African American, Chinese, Japanese, Korean, Filipino, Vietnamese, Native American, and other races). The mortality volume was to be expanded to include cancer death rates by geographic regions (census division). These immediate actions were accomplished by April 1999.

The CRWG again identified further expansion of the CINA monograph as an immediate action in 1999. This would broaden the geographic coverage for the incidence volume and include cancer incidence and mortality rates for Asians/Pacific Islanders and Native Americans, in addition to all races, whites, and African Americans (this has not yet been done). The expanded volume also would identify the top five newly diagnosed cancer sites for all races, whites, African Americans, Chinese, Japanese, Korean, Filipinos, Vietnamese, Native Americans, Hispanics (newly added) and other races. This expanded 2000 CINA monograph was expected to be published in April.

In conjunction with CAIR, the CRWG is conducting in-depth analyses and special studies. Because there is a limit to what can be published in CINA, the results of these geographic variations of analyses are not routinely included. Some of the special analytical studies conducted by the CRWG/CAIR focus on lung cancer, pediatric cancer incidence rates in the United States, prostate cancer trends, breast cancer: stage distribution by race and geographic area, and colon cancer: subsite, stage, and race.

The CRWG has identified three specific special projects to address data gaps, including the Hispanic project, the SES-related project, and the urban/rural differences project. These projects involve follow-up or expansion of previous and ongoing work and they represent a collaborative, concerted effort to address the gaps in a more comprehensive, systematic, and focused manner. The Hispanic project includes representatives from NAACCR, NCI, CDC, Census Bureau, and state registries with large Hispanic populations originating from Mexico, Cuba, and Puerto Rico. This effort addresses comparability, misclassification, educational needs, operation standards, and the impact of the 2000 Census on cancer rates. The SES-related special project involves numerous studies conducted by various federal agencies that are led by NCI researchers. The projects include statistical modeling of cancer mortality trends by SES factors at the county level, linking of SEER incidence rates with aggregate SES data at the census tract level, and a pilot study on linking SEER registries to the CPS data files for SES variables at the individual level. This pilot project could result in a “model” that could be used for other registries and programs. The urban/rural special project examines the differences in cancer incidence rates by level of urbanization using a 10-category continuous measure. Fourteen states submitted cancer incidence data at the county level and were analyzed as a beta test project. The findings are expected to be presented in January 2000.

Implementation Status of the 1996 Hispanic Data Symposium Recommendations

Dr. Edward Trapido explained that, in 1996, NAACCR's Uniform Data Standards Committee determined that there was a need to examine the methodologic issues associated with cancer data among Hispanics and suggested that a subcommittee be formed to examine these issues. The Subcommittee on Methodologic Issues of Measuring Cancer Among Hispanics was established and the Hispanic data symposium in Atlanta was organized. The final report from the symposium has been completed and is posted on the NAACCR Web Site.

The symposium was attended by about 40 representatives from registries, NCHS, Census Bureau, researchers, and community-based organizations (CBOs). A number of recommendations resulted from the discussions at the symposium, including:

- Data should be evaluated at the local level. There should be surveys to identify items in medical records that were used for classification, surveys to determine how registrars make decisions, and a summary report that is available to the public.
- Registry audits are needed to determine the extent to which information is available on the medical record to base an ethnic designation.
- Surname lists should be evaluated (particularly at the local level).
- The usefulness of the GUESS (Generally Useful Ethnic Surnames Search) program should be determined. In general, this program uses a variety of variables (especially surname) to assign Hispanic ethnicity. Because it was designed for the Hispanic population of New Mexico, it must be adapted for use with Hispanics of other than Mexican origin.
- The Uniform Data Standards Committee should establish a classification standard that incorporates current sources of data; a future standard should be recommended as new sources of data are identified.
- If the ethnicity variable changes in the 2000 Census, the consistency with the prior variable should be assessed. Also, Hispanics should be oversampled using the "long" form.
- NAACCR, NCHS, ACS, CBOs, and non-hospital data sources should provide feedback on the Census variable and changes in Office of Management and Budget (OMB) Directive 15.
- NAACCR should determine the benefit of collecting Hispanic identifier information to hospitals and other sources of cancer cases.
- CBOs should work to educate hospitals and non-hospitals that are identifying cases as well as physicians about the importance of Hispanic identifiers.
- NCHS should oversample Hispanics, coordinate questions to those of other registries and the Census, continue to conduct studies of Hispanic identification on death certificates, and increase training of individuals who code death certificates with regard to these issues.
- ACS should provide professional training on this issue, integrate ideas and definitions in its publications, and provide support for consensus development.
- CBOs should increase awareness on the local level of the need for accurate information, encourage and facilitate changes in Hispanic identification in hospital data collection forms, encourage oversampling of Hispanics in state and federal surveys, and develop and provide to registries culturally appropriate information (e.g., terminology) and Hispanic identity.
- Non-hospital sources should be encouraged to collect relevant ethnicity data.
- Managed care providers should be approached for improvement of particular data items (for a minimal set of key variables). In addition, they should be educated about the importance of identifiers.

- ✧ Medical schools should train physicians about the importance of cancer data and demographic information.
- ✧ Researchers should be encouraged to develop and assess data strengths and limitations, and to work with registries on Hispanic identifiers.

Dr. Trapido pointed out that many issues have been identified and some progress has been made, but much work remains to be done.

Discussion

Did the committee make any recommendations on how to prevent the misclassification of American Indians who have Hispanic surnames? Dr. Trapido responded that the committee did not make any recommendations concerning this limitation of surname lists, although this is a concern. Dr. Charles Key mentioned that he deals with this issue extensively in New Mexico, where about one-third of the Pueblo Indians have Hispanic surnames. In those cases, race reported as American Indian takes precedence over the surname in classification. Dr. Manuel de la Puente noted that when the Census Bureau does edits to fill in missing data, it excludes American Indians when using Hispanic surname lists to assign Hispanic origin. Dr. Daniel Miller mentioned that the State of Washington has made arrangements with some tribal organizations to link their registry data with personal identifiers to the population registries. Many misclassifications in the cancer registry are detected and corrected through such a linkage. However, he noted that tribal organizations have different levels of comfort with permitting such linkages. Dr. Edwards added that in the Northwest (i.e., Washington, Oregon, and Idaho), tribal groups have agreed to allow their data to be linked to registry data only by a third-party who tabulates the aggregated data and then unlinks the data sets. Obviously, there are still some concerns among tribal groups about sharing their information.

Working Group for Calculation of National Cancer Incidence Rates for Latinos

Dr. Dee West described the activities of the working group for calculation of National Cancer Incidence Rates for Latinos (NCIRL). Working group members participated in a conference call to discuss the following questions:

- ✧ What has been done since the Atlanta symposium?
- ✧ Which registries collect Latino data?
- ✧ What classification methods are used?
- ✧ How are denominators obtained?
- ✧ How are rates now being calculated?
- ✧ What work is now being done for Latino classification?

The action items identified by the working group were to: re-contact all registries (the registries were surveyed prior to the Atlanta symposium), evaluate the Illinois method of classification, examine CDC data from registries, and summarize the Census 2000 plan and its effect on registries. The working group hopes to publish some data by 2001. The committee will seek funds as needed to continue its work and they eventually will develop an educational plan. The members of the committee also agreed to participate in conference calls/meetings over the next year or two to move this process forward.

During the conference call, a number of questions were raised and discussed by the committee, including: Is it possible to have uniform methods? Is it possible to adjust for misclassification? What

will be the impact of the 2000 census? What can we do in the short term (e.g., proportional rates)? What educational efforts may be effective?

Dr. West indicated that the Illinois Cancer Registry defines "Hispanic" as:

1. All cases identified by cancer registries as Hispanic.
2. Cases with heavily Hispanic surnames (last name for males, maiden name for females)
   - ✧ Not born in places with high Spanish surname and low Hispanic ethnicity.
   - ✧ Not American Indian, Filipino, or Hawaiian.
3. Birthplace associated with Hispanic ethnicity
   - ✧ Maiden name rarely Hispanic
   - ✧ Last name rarely Hispanic.

Discussion

Dr. Lovell Jones pointed out that cancer rates in Hispanic populations vary from region to region. He stressed the need for local as well as national data and cautioned against using SEER data for policy making with regard to Hispanics.

Regional and Agency Data Improvement Initiatives

New Mexico Tumor Registry

Dr. Charles Key described some of his observations over the past 30 years as Medical Director for the New Mexico Tumor Registry (NMTR). The NMTR staff rely on a number of important resources, including: patients and physicians, hospitals and laboratories, cancer tumor registries, research scientists, public health officials and ACS volunteers, community leaders and health advocates, NCI staff and SEER colleagues, and funding from the NCI and the State of New Mexico. The NMTR's primary responsibility is to monitor cancer in New Mexico, but the study of cancer patterns and trends in the state's diverse population is a major secondary focus. A major challenge is to make sense out of imperfect information. Dr. Key summarized his perspective on race and ethnicity cancer statistics in the following statement from the report of the NAACCR Subcommittee on Methodologic Issues of Measuring Cancer Among Hispanics:

> "*... recognizing the limitations of existing enumeration and classification systems for both the incident cancer cases and the corresponding populations at risk, and the lack of consistency of individuals responding to questions about their race and ethnicity, the goal is to produce useful information that is reasonably reproducible and consistent over time (but never perfect).*"

NMTR data, which include incidence rates for American Indians and Hispanics since 1969, appear in the publication series of the International Agency for Research on Cancer (IARC) called "Cancer Incidence in Five Continents" (Volumes III, IV, V, VI, and VII). A bibliography and the text of selected publications and copies of recent graphical presentations are located on NMTR's new Web site at **http://hsc.unm.edu/epiccpro**. Dr. Key stressed the importance of the diversity of cancer occurrence within as well as between race/ethnic populations within New Mexico. He noted that the state's cancer data usually do not represent American Indian and Hispanic populations living in other parts of the United States.

The NMTR staff always are trying to improve the resources and tools available for gathering cancer data in a poor state with a 50 percent minority population. They pioneered the use of "circuit-riding" medical abstractors with laptop computers to serve widely dispersed, small rural hospitals. The NMTR currently participates in NCI's SEER program, Breast Cancer Surveillance Consortium (BCSC), and Cancer Genetics Network. For minority cancer data collection, the registry collaborates with Dr. Anne Lanier (Alaska Native Health Board), Sohail Khan (Cherokee Nation), Tim Flood (Arizona Cancer Registry), Dr. Nat Cobb and colleagues (Indian Health Service), and others. Dr. Key called attention to valuable reports from Roberta Paisano and Dr. Nat Cobb, Dr. Anne Lanier, Dr. William Athas (NMTR), and the November 1999 *American Journal of Public Health* report on cancer incidence among American Indians in Minnesota. A report containing American Indian/Alaska Native cancer data from all SEER sites (including Alaska, Arizona, New Mexico, Washington, and others) is in preparation. Dr. Key distributed copies of *Cancer in New Mexico 1970-1996: Changing Patterns and Emerging Trends* (1998).

The NMTR will continue to work with these colleagues to examine the data collected over time from the American Indian population in Arizona. The NMTR is concerned about the completeness of the data collected on this population. Dr. Key mentioned that there is a memorandum of agreement among New Mexico, Arizona, and the Indian Health Service to improve the collection of data on American Indians in Arizona. There also are plans for a publication that will include the New Mexico, Arizona, and Alaska data. Drs. Lillian Tom-Orme and Anne Lanier have been asked to assist with this effort.

Discussion

Dr. Lovell Jones asked if anyone had considered comparing the New Mexico and Arizona data with data on the plains American Indians. Dr. Key responded that the NMTR relies heavily on the publications of the mortality data that show huge differences between the American Indians in the southwest and the American Indians in the plains. He noted that the Oklahoma and South Dakota registries must be improved if they are going to provide the data needed to make comparisons between American Indians in the southwest and the plains. Dr. Edwards added that the NCI is working with South Dakota to obtain better data on the plains American Indians. Dr. Hampton noted that there are a number of northern states (e.g., Minnesota, Wisconsin) that have excellent data on American Indians. However, there has never been any attempt (outside of SEER) to pull together the data on American Indians from different state registries. Dr. Edwards commented that NAACCR has given some thought to this issue; however, the numbers in the registries that would meet the criteria are expected to be quite small. Dr. Daniel Miller mentioned that the North Dakota registry began collecting data in 1997, and the first report on North Dakota cancer statistics is in preparation. The South Dakota registry was funded approximately 1 year ago, and in that year, there has been a complete turnover of registry staff. Therefore, it will take some time before that registry meets the criteria.

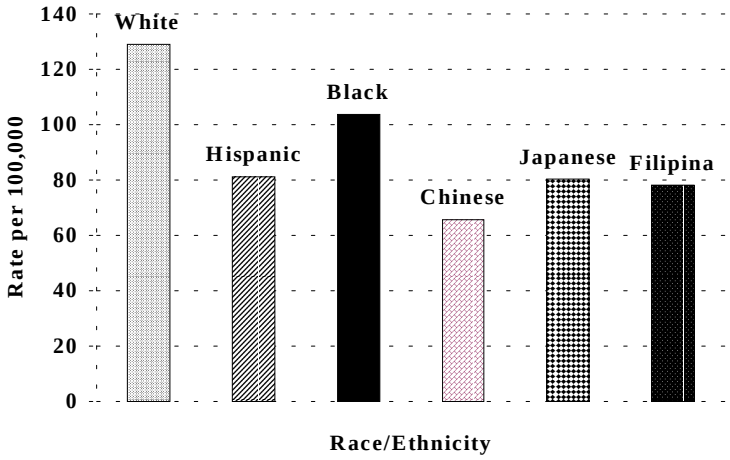
Race/Ethnic Studies From the SEER Program in Northern California

Dr. Dee West indicated that, according to a USA Today report, California has the largest Native American population in the United States. California has three SEER registries that cover nearly half of the state's population, but it often is difficult to correctly classify the individuals in the registry. After linking 170,000 records from the Indian Health Service with 570,000 records from the California Cancer Registry, 1,043 American Indian invasive cancer cases were identified. Eighty percent of these cases were not in the California Cancer Registry. Of the 634 cases in the California Cancer Registry, only 199 records (35 percent) were classified correctly as American Indians. Clearly, the registry is missing or misclassifying a large percentage of the American Indian cases. There is a need to develop better methods to identify Native Americans in urban populations.

Dr. West provided some data from the race/ethnic studies conducted by the Northern California Cancer Center (NCCC). In 1995, the NCCC published breast cancer incidence rates (see Figure 1) in the San Francisco Bay Area by race/ethnicity (1988-1992). The registry data indicate that breast cancer incidence was highest in the white population and lowest in the Asian population.

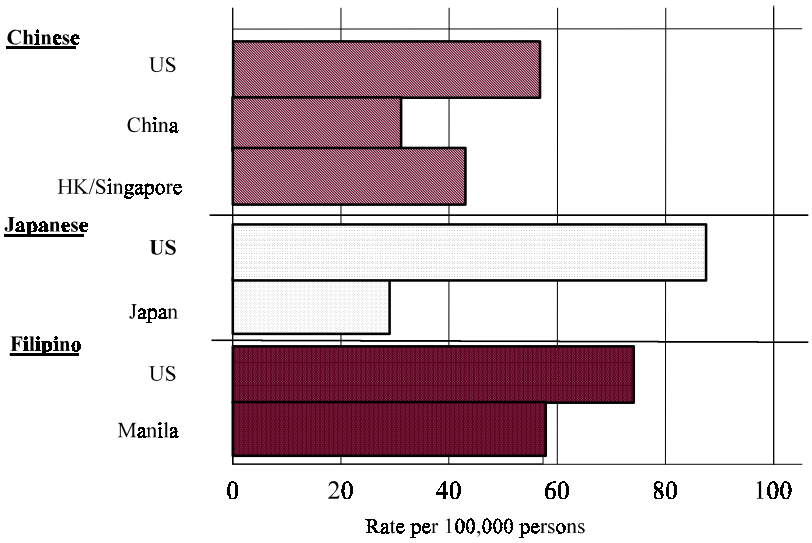
Figure 2 presents data on the international differences in breast cancer rates published in a 1999 NCCC report. This figure compares the incidence among Chinese in the United States, China, and Hong Kong/Singapore; the incidence among Japanese in the United States and Japan; and the incidence among Filipinos in the United States and Manila. The U.S. data were from the California registry and SEER combined, and the data for the other countries were from IARC's "Cancer Incidence in Five Continents." This comparison indicated that the incidence rates for Chinese in the United States are much higher than for Chinese in China or Hong Kong/Singapore. The incidence rate for Japanese in the United States is nearly three times higher than that for Japanese in Japan. However, the differences are not as great among Filipinos in the United States compared to those in Manila.

Figure 1. Breast Cancer Incidence Rates by Race/Ethnicity, 1988-1992



Source: NCCC, Cancer Incidence by Race/Ethnicity in the San Francisco Bay Area, 1995.

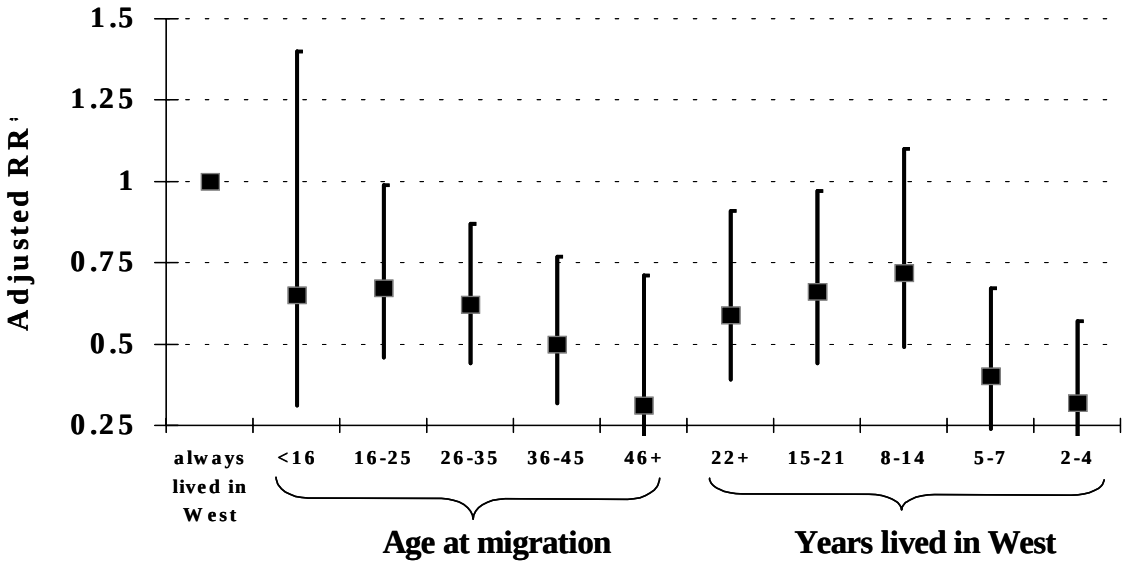
Figure 2. International Differences in Breast Cancer Incidence Rates, 1988-1992



Source: NCCC, Cancer Incidence in Chinese, Japanese, and Filipinos in the U.S. and Asia, 1999.

The relationship between migration patterns and breast cancer risk in Asians is depicted in Figure 3. The data indicate that Asians who migrate to the West at a younger age have higher relative risk than those who migrate at the age of 46+ years. In addition, the relative risk is higher for those Asians who have lived in the West for 8+ years. The relative risk was lower in Asians who migrated from rural areas.

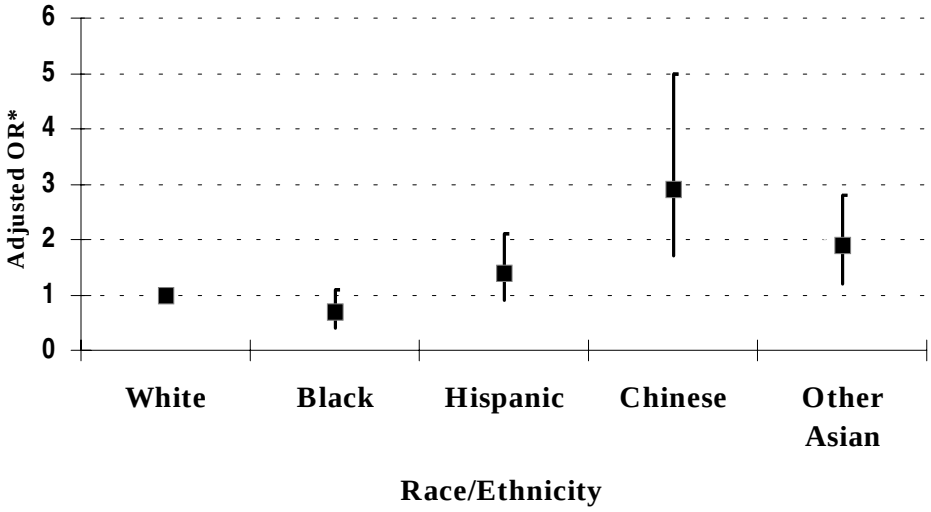
Figure 3. Migration Patterns and Breast Cancer Risk in Asians



Source: Ziegler RG, et al. JNCI, 1993.

Dr. West presented some data from a SEER patterns of care study on breast cancer treatment (see Figure 4). About 625 women diagnosed with local-stage breast cancer were involved in the study; approximately 40 percent of these women had a mastectomy. Asian women in the San Francisco Bay Area were two to three times more likely to have a total mastectomy than white women. This difference remains even after adjusting for demographic and clinical characteristics. Further examination is needed to determine why the treatment is very different among the racial/ethnic groups in the Bay Area.

Figure 4. Likelihood of Having a Mastectomy (vs. BCS) by Race/Ethnicity



* Adjusted for demographic and clinical factors

Source: Prehn AW, et al. Breast Cancer Treatment by Race/Ethnicity: Focus on Asian Women

The NCCC conducted a Vietnamese classification study to assess the accuracy of the rates generated by the registry for the Vietnamese population and to determine whether these individuals were correctly classified by the registry. Figure 5 indicates that 190 of the study subjects were classified by the registry as Vietnamese and by the subjects themselves; 806 were classified as not Vietnamese by the registry and by the subjects themselves. However, 68 were classified as Vietnamese by the registry and as not Vietnamese by the subjects themselves, and 20 were classified as not Vietnamese by the registry and as Vietnamese by the subjects themselves.

Figure 5. Cross-Classification of Study Subjects by Self-Identified and Registry Classified Race. Vietnamese Classification Study, Greater Bay Area Cancer Registry. *Unweighted n*.

	Self-Identified Vietnamese	Self-Identified Not Vietnamese
Registry classified Vietnamese	Correctly classified n=190 a	Incorrectly classified n=68 b
Registry classified not Vietnamese	Incorrectly classified n=20 c	Correctly classified n=806 d

Source: Swallen KC, Glaser SL, Stewart SL, West DW, Jenkins CNH, McPhee, SJ; Ethnicity & Disease; Volume 8, Spring 1998; pp 218-227.

How does this misclassification affect the incidence rates generated by the registry? Figure 6 indicates that the registry race was correct about 74 percent of the time; however, 26 percent of the time it was incorrect (the registry missed about 10 percent of those who identified themselves as Vietnamese). The most accurate approach for classifying Vietnamese subjects identified in the study was to use birthplace, surname, and the information from the registry race. If two of the three data elements indicated that the subject was Vietnamese, the subject was classified as Vietnamese. This approach resulted in correct classification of subjects about 78 percent of the time.

Figure 6. Predictive Value Positive and Sensitivity Vietnamese Race, for Various Classification Methods. Vietnamese Study, Greater Bay Area Cancer Registry (calculated using weights)

	Predictive Value Positive*	Sensitivity†
Registry race	73.6	89.6
Birthplace	70.0	82.9
1954 Surname list	66.6	55.4
1990 Surname list	68.8	80.0
Race + birthplace or Race + 1954 surname or Birthplace + 1954 surname	78.1	86.6
Race + birthplace or Race + 1990 surname or Birthplace + 1990 surname	78.3	91.2

* Predictive value positive, measured by a/(a+b), is the chance that a person classified as Vietnamese by the registry identified himself or herself as Vietnamese on interview.

† Sensitivity, measured by a/(a+c), is the chance that a person of Vietnamese race was correctly classified by the registry as Vietnamese.

Source: Swallen KC, Glaser SL, Stewart SL, West DW, Jenkins CNH, McPhee, SJ; Ethnicity & Disease; Volume 8, Spring 1998; pp 218-227.

Figure 7 presents the characteristics of the correctly classified Vietnamese (column 1), incorrectly classified Vietnamese (column 2), and incorrectly classified true non-Vietnamese (column 3) from the Vietnamese classification study. Those subjects who the registry incorrectly classified as not Vietnamese tended to be older and male; most had been in the United States longer. Those subjects who were incorrectly classified as Vietnamese by the registry, were primarily (98 percent) Chinese, they had been in the United States for a shorter time, and they tended to speak less English.

Figure 7. Characteristics of Correctly Classified Vietnamese, Incorrectly Classified True Vietnamese, and Incorrectly Classified True Non-Vietnamese from the Vietnamese Classification Study, Greater Bay Area Cancer Registry (percents are based on weighted sample)

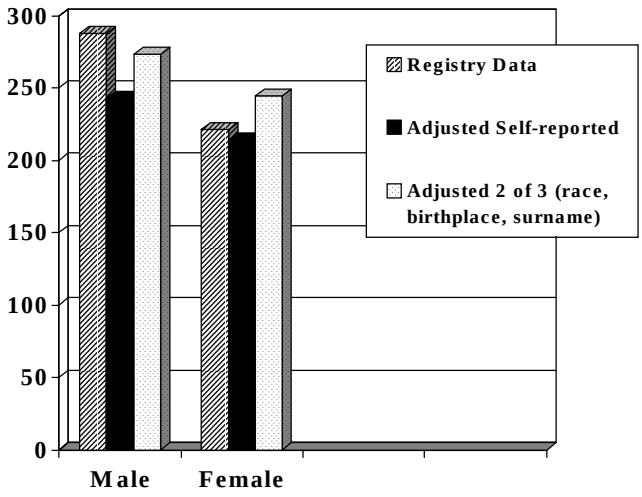
	Self ID Vietnamese Class Vietnamese (unweighted n=204)	Self ID Vietnamese Class Non-Viet. (unweighted n=6)	Self ID Non-Viet. Class Vietnamese (unweighted n=47)
Self-identified ethnicity			
Vietnamese	100.00	100.00	0.00
Chinese	0.00	0.00	97.96
Missing	0.00	0.00	2.04
% Male	37.06	67.96 *	38.45
Mean age (years)	55.23	67.49 *	56.60
% Dead	30.95	49.39	21.33
% Identified as Chinese-Vietnamese	74.20	65.78	48.36
% Born			
Vietnamese	97.94	83.13	85.69
China	0.94	16.87	12.27
Mean year US entry	1984.28	1978.12 *	1983.99
Mean years of education	8.74	7.93 *	5.79 *
% Speak English	67.96	67.47	40.37 *
% Speak Vietnamese	100.00	100.00	90.93 *
% Speak Chinese	12.83	49.39 *	95.91 *
Interview language			
English	16.10	32.53	8.18 *
Vietnamese	83.42	67.47	38.69 *
Cantonese	0.47	0.00	53.13 *

* Indicates different from correctly classified true Vietnamese at the 0.05 level.

Source: Swallen KC, Glaser SL, Stewart SL, West DW, Jenkins CNH, McPhee, SJ; Ethnicity & Disease; Volume 8, Spring 1998; pp 218-227.

Figure 8 depicts the impact of the misclassification on the cancer incidence rates. The first bar is from the registry data alone, the second bar is from self identification adjusted to the total population, and the third bar is the classification using the "two out of three rule." For males, the incidence using the "two out of three rule" for classification is slightly better than the incidence using the registry data alone.

Figure 8. Average Annual Age-Adjusted Cancer Incidence Rates and 95% Confidence Intervals for Vietnamese in Nine County Greater Bay Area, 1989-1992, Adjusted and Unadjusted, from the Vietnamese Classification Study, Greater Bay Area Cancer Registry

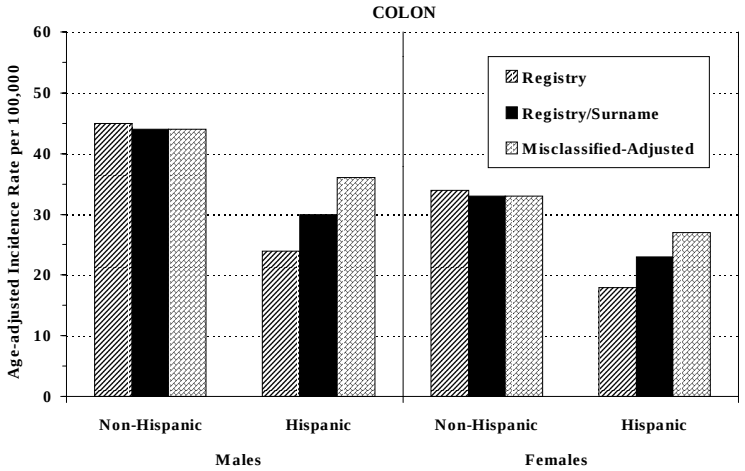


Source: Swallen KC, Glaser SL, Stewart SL, West DW, Jenkins CNH, McPhee, SJ; Ethnicity & Disease; Volume 8, Spring 1998; pp 218-227.

However, use of this rule results in over-reporting of incidence in both males and females. The study concluded that the incidence rates reported by the Greater Bay Area Cancer Registry for the Vietnamese population may be high, especially for females.

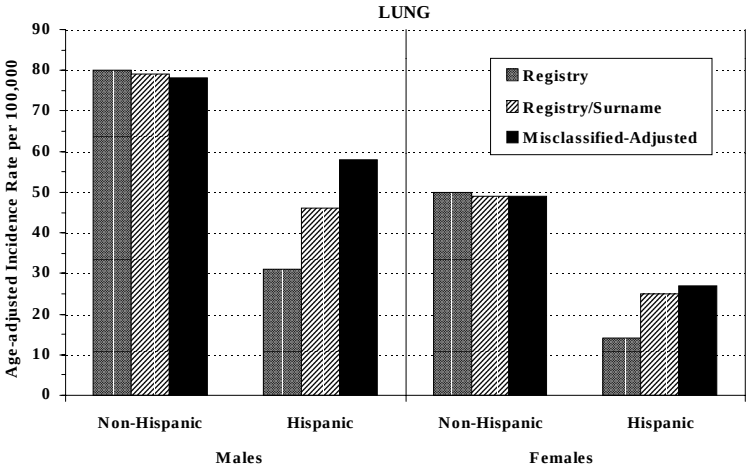
In a study to investigate misclassification of Hispanics and its effect on incidence rates, three classification methods were compared—registry data, registry data/surname, and self reported—for colon, lung, breast, cervical, and prostate cancer (see Figures 9-13). The incidence rates generated using registry data/surname are closer to the incidence resulting from self reporting. The registry concluded, therefore, that use of both registry race and surname was a better approach for classifying Hispanics.

Figure 9. Age-Adjusted Invasive and In-Situ Colon Cancer Incidence Rates by 3 Classification Methods, San Francisco Bay Area, 1990



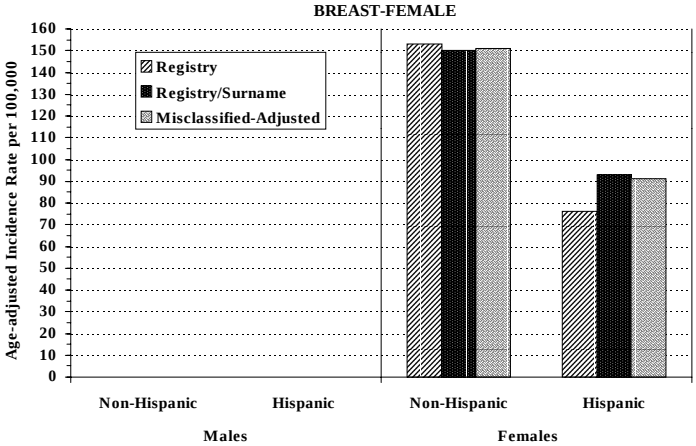
Source: Swallen KC, Glaser SL, Stewart SL, West DW, Jenkins CNH, McPhee, SJ; Ethnicity & Disease; Volume 8, Spring 1998; pp 218-227.

Figure 10. Age-Adjusted Invasive and In-Situ Lung Cancer Incidence Rates by 3 Classification Methods, San Francisco Bay Area, 1990



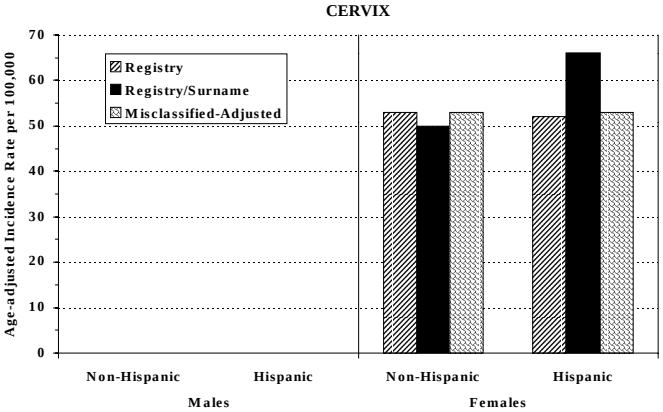
Source: Swallen KC, Glaser SL, Stewart SL, West DW, Jenkins CNH, McPhee, SJ; Ethnicity & Disease; Volume 8, Spring 1998; pp 218-227.

Figure 11. Age-Adjusted Invasive and In-Situ Breast Cancer Incidence Rates by 3 Classification Methods, San Francisco Bay Area, 1990



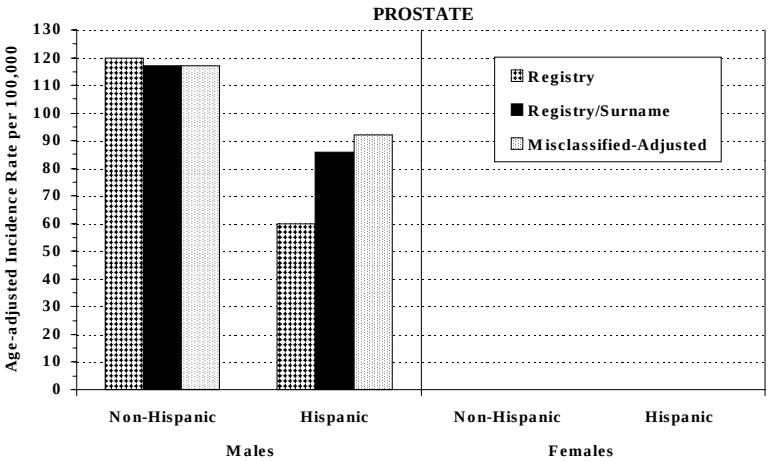
Source: Stewart SL, Am J Epidemiol 1999; 149:1063-71.

Figure 12. Age-Adjusted Invasive and In-Situ Cervical Cancer Incidence Rates by 3 Classification Methods, San Francisco Bay Area, 1990



Source: Stewart SL, Am J Epidemiol 1999; 149:1063-71.

Figure 13. Age-Adjusted Invasive and In-Situ Prostate Cancer Incidence Rates by 3 Classification Methods, San Francisco Bay Area, 1990



Source: Stewart SL, Am J Epidemiol 1999; 149:1063-71.

Dr. West closed with a description of the results of an NCCC hospital survey that was sent to hospital tumor registrars and hospital administrators. The purpose of the survey was to determine what information was available, where information was found, what was recorded, and how they determined race/ethnicity. The sources of information used to code birthplace, race, and ethnicity are presented in Figure 14, the frequency of recording this information is shown in Figure 15, and the methods used to determine if a surname is Spanish are listed in Figure 16. The responses to the question concerning ability to code race/ethnicity according to California Cancer Registry categories are provided in Figure 17, and the responses of registrars asked to code an individual who also had a Hispanic surname are presented in Figure 18. Dr. Key noted that about 12 percent of the registrars coded the Native American as Hispanic even though the coding rule states that an individual cannot be both Hispanic and Native American. Responses of the hospital administrators to some of the survey questions are presented in Figures 19-24. As indicated in Figure 19, most hospital administrators reported that race is recorded. Figure 22 indicates that birth, race, and ethnicity often are recorded differently in different parts of the medical record, so there are conflicting coding rules. Most hospital administrators indicated that Mexican ancestry is coded as Hispanic speaking, which does not indicate Mexican ancestry. For Chinese ancestry coding, about 74 percent of the hospital administrators indicate these individuals are coded as Asian, which provides no specific information about ancestry. Dr. Key concluded his presentation by stating that more instruction is needed to improve data collection at the source.

Figure 14. NCCC Hospital Survey of Tumor Registrars
Sources of Information Used to Code Birthplace, Race, and Ethnicity

	Birthplace %	Race %	Ethnicity %
Face Sheet	57.9	82.1	38.2
Physical Exam Report	87.5	93.0	87.7
Discharge Summary	61.8	70.9	63.6
Nurses' Notes	24.5	33.3	32.7
Death Certificates	72.7	68.5	59.3
Other*	31.0	37.9	34.5

* Includes Hospital Computer, Radiation Records, Insurance Forms, Progress Notes, MD Contact, Social Service Notes.

Figure 15. NCCC Hospital Survey of Tumor Registrars
Frequency of Information on Birthplace, Race, and Ethnicity is Available in the Hospital

	Birthplace %	Race %	Ethnicity %
All/Most Patients	34.5	77.6	25.8
Some Patients	15.5	17.2	27.6
Few or No Patients	50.0	5.1	46.6

Figure 16. NCCC Hospital Survey of Tumor Registrars
Methods Used to Determine if a Surname is Spanish

	N	%
Sound	27	49.1
Lists	4	7.5
Spelling	37	66.1
Medical Record	51	89.5

Figure 17. NCCC Hospital Survey of Tumor Registrars
Coding Race and/or Ethnicity According to California Registry Categories

With the information your hospital collects, are you able to code race and/or ethnicity according to California Cancer Registry categories?

	N	%
For All Patients	9	15.5
For Most Patients	31	53.4
For Some Patients	9	15.5
For Only a Few Patients	4	6.9
For No Patients	3	5.2
Don’t Know	2	3.4

Figure 18. NCCC Hospital Survey of Tumor Registrars
Coding Individuals Who Have Hispanic Surnames

How would you code the following if they also had a Hispanic Surname?

	Non-Hispanic	Hispanic	Unknown	Other	Dk
Black	32.8	22.4	34.5	10.3	
White	13.8	37.9	32.8	15.5	
Native American	34.5	12.1	43.1	8.6	1.8
Japanese	50.9	7.0	31.6	8.8	1.8
Filipino	54.4	7.0	28.1	10.5	
Guamanian	35.1	14.0	36.8	10.5	3.5

Figure 19. NCCC Hospital Survey of Hospital Administrators Collection of Information on Place of Birth, Race, and Ethnicity

Does hospital collect information on place of birth, race, and ethnicity?

	Birthplace %	Race %	Ethnicity %
Always	39.7	85.0	18.6
Sometimes	24.1	10.0	16.9
Never	32.8	5.0	55.9
Don't Know	3.4	0.0	8.5

Figure 20. NCCC Hospital Survey of Hospital Administrators Where Birthplace, Race, and Ethnicity Information is Found

Where in your hospital are birthplace, race, and ethnicity information found?
(number naming source)

	Birthplace	Race	Ethnicity
Face Sheet	27	50	12
Physical Exam Report	12	34	16
Discharge Surname	5	25	9
Nurses' Notes	3	15	8
Computer	3	2	0
Other	4	7	4

Figure 21. NCCC Hospital Survey of Hospital Administrators Source of Obtaining Birthplace, Race, and Ethnicity (number naming source)

	Birthplace	Race	Ethnicity
Patient	33	48	21
Family	29	44	18
Friend	18	31	12
Birthplace	NA	7	3
Physical Characteristics	NA	25	3
Language	4	17	6
Surname	2	11	6
Existing Medical Records	24	38	16

Figure 22. NCCC Hospital Survey of Hospital Administrators
Consistency of Coding Birthplace, Race, and Ethnicity

	Birthplace		Race		Ethnicity	
	N	%	N	%	N	%
Yes	9	36	25	51	9	41
No	10	40	16	33	3	14
Don't Know	6	24	8	16	10	46
TOTAL	25		49		22	

Figure 23. NCCC Hospital Survey of Hospital Administrators
How Mexican Ancestry is Coded

	Typical		Detailed	
	N	%	N	%
Mexican	7	13.7	11	21.6
Hispanic Speaking	39	76.5	10	19.6
Whites	2	3.9		
White & Mexican	1	2.0		
White & Hispanic Speaking	2	3.9		
Hispanic	1	2.0		
Don't Know			9	17.6
No Answer			20	39.2
TOTAL	51		51	

Figure 24. NCCC Hospital Survey of Hospital Administrators
Coding of Chinese Ancestry

	Typical		Detailed	
	N	%	N	%
Chinese	5	9.8	13	25.5
Asian	38	74.5	7	13.7
Oriental	6	11.8	2	3.9
Asian/Pacific Island	1	2.0		
SE Asian	1	2.0		
Don't Know			9	17.6
No Answer			20	39.2
TOTAL	51		51	

Discussion

Dr. Daniel Miller mentioned that Robert Hahn from the CDC has conducted several studies on information sources and has found there often is a mismatch among sources (e.g., many individuals are born white on the birth certificate but die black on the death certificate). Another survey indicated that self designation of ethnicity changes over time depending on who is asking the question and for what purpose. Dr. West added that there are plans to extend the Hispanic study to look at death certificates. He also mentioned that surveys are expensive and often incomplete. For example, only a 49 percent response rate was achieved for the Vietnamese classification survey because many of the study subjects moved frequently.

Hawaii Tumor Registry

Dr. Marc Goodman described the ethnic composition of Hawaii based on 1990 census data, and pointed out that there are no majority populations is Hawaii. Caucasians represent only about 25 percent of the state's population, followed closely by Hawaiians and Japanese. There also are substantial numbers of people from southeast Asia and some of the other Pacific Islands in Hawaii. About 20-22 percent of the people in Hawaii have mixed ancestry. Therefore, classifying these individuals poses a challenge to registrars.

The Hawaii Tumor Registry (HTR) has been in operation since 1960, and has been part of the SEER program since 1973. It has been involved in epidemiologic research examining the differences in cancer incidence and survival among the various ethnic groups as well as studies of cancer etiology, control, therapy, and quality of life. The registry also has developed a computer linkage program that is sensitive to Asian and Pacific Islander names. The HTR has more than 25 contracts with the National Institutes of Health and the Department of Defense to conduct multiethnic studies. The registry has participated in migrant studies since 1960 and it is very interested in the effect of generation. Approximately 18 percent of the individuals living in Hawaii were born outside of the United States (see Figure 25). Hawaii's population is primarily urban with most of the population living in Honolulu; however, there is disparity in urbanization in the neighboring islands (see Figure 26). Because most of the major hospital facilities are located in Honolulu, the individuals living on the other islands must fly to obtain medical care.

Figure 25. Place of Birth, Hawaii, 1990

	Number	Percent
Total Population	1,108,229	
Born in U.S.	914,024	82.5
Hawaii	621,992	56.1
Other	292,032	26.4
Born outside U.S.	194,205	17.5
U.S. citizen/Territory	31,501	2.8
Foreign-born	162,704	14.7
Alien	72,721	6.6

Figure 26. Urban and Rural, Hawaii, 1990

	% Urban
Total	88.9
Honolulu	96.4
Hawaii	60.8
Kauai	55.2
Maui and Kalawao	77.9

The HTR obtains numerator data from hospitals, physicians, laboratories, and death certificates; the denominator data come from the census and the Health Surveillance Program—a program of the Hawaii Department of Health. The registry abstracts ethnicity information from the medical chart using a 2-digit code, which allows for the identification of two primary ethnicities. The HTR has more data available than is used and reported by NCI's SEER program.

Dr. Goodman indicated that there are a number of problems associated with the ethnicity coding used for the U.S. census (see Figure 27). In 1960, when the census was self administered, approximately 102,000 part-Hawaiians were identified. In 1970, when the census was administered as a mail survey, this number dropped to 72,000. Despite the census data, the population of part-Hawaiians had actually increased according to a number of other data sources. For mixtures of non-white/non-Hawaiian groups, the 1960 and 1970 census coded these individuals to their father's ethnicity; in 1980 the census began coding these individuals to their mother's ethnicity. Therefore, it will be difficult to develop some continuity with regard to ethnic identification, even if many of these problems are ameliorated with the 2000 census.

Figure 27. Ethnicity Coding at the U.S. Census

Mixture	Census	Code
White/Non-White	1960, 1970, 1980, 1990	Non-White
Part-Hawaiian	1960	Hawaiian
	1970, 1980, 1990	Hawaiian or other ethnicity
Non-White/Non-Hawaiian	1960, 1970	Father's ethnicity
	1980, 1990	Mother's ethnicity (or first mentioned race)

The Health Surveillance Program (HSP) is used to generate intercensal population estimates; it samples about 2 percent of households in the state each year. The ethnicity information is obtained from a personal interview and is coded using a four-digit code (two ethnicities for each parent of record). Final coding of Caucasian, Hawaiian, and other ethnic groups is the same as the census. Dr. Goodman noted that the data from this program usually are more accurate than the information provided by the census. Of the 3,252 cases examined, there was very reasonable agreement on ethnic classification between the HTR and the HSP for most ethnic groups (e.g., 85 percent for Native Hawaiians). There is a small

difference between men and women, particularly among Chinese and Filipinos (see Figure 28). This difference was attributed to the fact that these two groups are most likely to inter-marry. With regard to birthplace classification, there was good agreement between the HTR and the HSP for Caucasians, Japanese, and Filipinos.

Figure 28. Percent Agreement Between HTR and HSP Ethnic Classification Among 3,252 Cases

	% Agreement	
Ethnic Group	**Men**	**Women**
Caucasian	95	94
Japanese	99	99
Chinese	96	92
Filipino	99	90
Hawaiian	85	85
Korean	96	96

Figure 29. Percent Agreement Between HTR and HSP Birthplace Classification Among 3,252 Cases

	% Agreement
Caucasian	97
Japanese	95
Filipino	96

Dr. Goodman indicated that the HTR has conducted a number of patient interview and linkage studies, including: ecologic studies, case-control studies, cohort studies (e.g., Population Cohort, Japan-Hawaii Cancer Study, and Multiethnic Cohort), intervention studies (e.g., Women's Intervention Nutrition Study; 9-A-Day Study; Prostate, Lung, Colorectal, and Ovarian Study (PLCO); and Women's Health Initiative), and health insurance files (e.g., Hawaii Medical Services Association Study, which includes information on comorbidities and screening and treatment practices; and Health Care Financing Administration linkage). The multiethnic cohort was established between 1993 and 1996 (see Figure 30). It contains 215,000 people, based in both Los Angeles (102,000) and Hawaii (113,000), who completed a 26-page mail survey. The cohort contains a number of ethnic minorities (about 22 percent Latino, 18-19 percent African American, 26 percent Japanese, and 6-7 percent Hawaiian). Blood and urine samples will be collected from these individuals if the next phase of the project is funded.

Figure 30. Distribution of the Multiethnic Cohort by Sex and Ethnicity

Ethnic Group	Men	Women
Latino	22,818	24,620
African-American	12,851	22,256
Japanese	26,964	29,957
Hawaiian	6,120	7,851
Caucasian	22,857	26,502
Other	5,200	7,255
Total	96,810	118,441

With regard to cancer patient followup, the HTR is conducting a breast cancer patterns of care and outcomes study, examining the poor survival among Filipino and Hawaiian females even when adjusted for stage. Several studies, which have been funded through the SEER Special Studies mechanism, are assessing the quality of life in cancer patients. Another study is examining ethnic differences in attitudes and experiences related to alternative cancer treatments. The registry also is interested in how informed consent may be a barrier to participation in clinical trials.

Dr. Goodman provided information on the ethnic distribution of participants in both therapeutic and epidemiologic studies (see Figure 31). Of the 2,700 people who were enrolled in the treatment trials, 70 percent were from ethnic minority groups. Of the 2,600 people involved in the epidemiologic studies, 69 percent were from ethnic minority groups. In terms of training and community outreach, the HTR provides pre-doctoral and post-doctoral training. The HTR works with the Native Hawaiian Health Care System, Papa Ola Lōkahi, and Alu Liké. In addition, the HTR is a minority-based CCOP and it assists with CDC's National Breast and Cervical Cancer Early Detection Program (NBCCEDP).

Figure 31. Ethnic Distribution of Participants in Therapeutic and Non-Therapeutic Studies, 1998-1999

	% population > 50 years (N=264,442)	% cancer cases (N=18,665)	% treatment (N=2,717)	% non-treatment (N=2,569)
Japanese	38	31	38	32
Chinese	8	6	6	6
Filipino	14	11	10	7
Hawaiian	11	11	11	15
Other Non-Caucasian	7	7	5	8
Caucasian	22	33	30	31

The HTR plans to use the 2000 census data to improve the denominators for Hawaii's ethnic populations. There also are plans to add behavioral risk factor data (from SEER Special Studies efforts) to the HTR, and to improve cancer reporting for other Asians and Pacific Islanders. The HTR also has calculated some provisional cancer incidence rates for American Samoans, and has begun collecting cancer data for Samoans living is American Samoa, Hawaii, and Los Angeles (see Figure 32 and 33).

Figure 32. Age-Standardized (World) Incidence Rates per 100,000 Males, 1990-1995

Site	American Samoa	Los Angeles	Hawaii
Prostate	43	52	116
Lung	30	89	57
Stomach	12	32	54
Colorectum	17	27	34
Leukemia	1	17	27
Liver	14	13	24

Figure 33. Age-Standardized (World) Incidence Rates per 100,000 Women, 1990-1995

Site	American Samoa	Los Angeles	Hawaii
Breast	36	84	117
Lung	7	36	42
Corpus uteri	14	41	37
Stomach	6	23	37
Cervix uteri	4	20	22
Colorectum	6	7	25
Ovary	7	7	22

Los Angeles Cancer Surveillance Program

Dr. Lihua Liu indicated that the Los Angeles Cancer Surveillance Program (LACSP), which was established in 1972 by researchers from the University of Southern California (USC), is a population-based registry that registers more than 40,000 cases each year. The registry joined the California Cancer Registry in 1988 and became a SEER registry in 1992.

In 1995, the total population of Los Angeles County was 9,138,789, ranking above 42 of the 50 states. The population in the county is ethnically diverse. More than 13 percent of the U.S. Asian population, about 8 percent of the Pacific Islander population, and 15 percent of the Hispanic population reside in Los Angeles County. The Asian population in the county includes Chinese, Filipino, Japanese, Asian Indian, Korean, Vietnamese, Cambodian, Hmong, Laotian, Thai, and Other Asian. The Pacific Islander population in the county includes Hawaiian, Samoan, Tongan, Other Polynesian, Guamanian, Other Micronesian, and Melanesian. The Hispanic origin of Los Angeles County residents includes Mexican, Puerto Rican, Cuban, and Other Hispanic.

About 95 percent of the cases in the registry are obtained from hospitals; the remaining 5 percent (about 2,000 cases per year) are obtained from other sources, including physicians' offices, pathology laboratories, and death certificates. The ethnic-related variables in the CSP database that are obtained from hospitals and other sources include: racial/ethnic categories (there are 28 categories in the database and only 2-3 percent are unknown), Spanish/Hispanic origin codes (there are 9 codes in the database and only 2.5 percent unknown), and birthplace (33-37 percent unknown). In addition to the information provided to the CSP, the registry derives the following ethnic-related variables:

- CSP race (non-Spanish surname white, black, Spanish surname white, Chinese, Japanese, Filipino, Korean, Other).
- Computer-derived ethnicity (to distinguish Hispanics).
- California Cancer Registry race recode (non-Hispanic white, non-Hispanic black, Hispanic, non-Hispanic Asian/Pacific Islander, non-Hispanic American Indian).
- SEER race (uses surname lists to identify more Hispanics and to reallocate certain cases of Asian origin).

Dr. Liu identified the following four issues of concern: (1) absence of ethnic data in medical records, especially birthplace; (2) the role of the registry in collecting data without contacting patients and in protecting patient privacy; (3) absence of ethnic population data as denominators to calculate incidence rates; and (4) heterogeneity within major racial/ethnic groups (Hispanics and subgroups within whites). Dr. Liu also suggested that the following items be considered:

- ✧ Race and birthplace information should be included in the hospital admission sheet. This approach, which is the most cost-efficient and useful approach, should be compatible with the expanded ethnic categories in the 2000 census.
- ✧ Registry data should be supplemented with available data resources through record linkages (e.g., death certificate) and feedback from data users.
- ✧ The Census Bureau and local government agencies should provide population estimates with ethnic detail and proportional incidence rates should be used.
- ✧ Birthplace and surname lists should be used to identify subgroups.
- ✧ These issues should be tackled through a national network/partnership and not at the local level.

Kentucky Tumor Registry

Dr. Thomas Tucker opened his presentation by stating that some populations that are at risk for cancer, such as that of Appalachia, have specific cultures that are well documented, but they will never be identified by race or ethnicity. In the state of Kentucky, there are 15 Area Development Districts (ADDs) made up of clusters of counties. Each of the ADDs has a District Cancer Council, which includes physicians, representatives from the ACS, nurses, cancer patients, and others from the community. An effort has been made to include on the council representatives from the high risk groups (e.g., low literate individuals) in the ADD. The registry shares cancer data with the ADDs and works with the respective councils to identify specific cancer problems. The community groups represented on the council are asked to help develop a plan to reduce the cancer rates in areas where the rates are high. Working with the community, the plan is implemented, and the registry periodically evaluates the effectiveness of the process using registry and other data sources.

There is a disproportionate cancer burden in the Appalachia area of Kentucky, an area typically associated with poverty and low literacy. The incidence of lung and cervical cancer is very high in this area. In addition, there is a low rate of detecting breast cancer early and a high rate of detecting late stage breast cancer. This is a consequence of inadequate screening of the population. When this trend became evident in 1991, the registry worked with the District Cancer Councils to identify and recruit lay health advisors (e.g., women who themselves may be isolated by low literacy skills or poverty) to encourage the women in the ADD to get screened. Four years after the program had been implemented, the rate of detecting breast cancer early has increased and the rate of detecting breast cancer in a late stage has declined dramatically. Dr. Tucker mentioned that this process also has been applied to other ethnic groups such as African Americans. About 93 percent of women diagnosed early, survived 5 years disease free after diagnosis, but only half of the women diagnosed in late stage survived disease free after 5 years.

The Kentucky registry has conducted record linkage studies with Medicare. These studies show that the treatment cost varies substantially between early and late stage diagnosis. Through efforts to encourage screening and early detection, about 446 women who would not have been diagnosed until late stage were diagnosed early. This resulted in a cost savings of $10,600 per woman in the first 2½ years following diagnosis—a total cost savings of $4.7 million.

Dr. Tucker pointed out that the registry may not be able to collect all of the information about the SES, racial/ethnic, and cultural differences needed. However, it may be able to use the data that are available in an effective way to make a real difference in people's lives if representatives from the subpopulations are invited to participate in planning, developing, and implementing programs for these subpopulations.

Texas Cancer Registry

Dr. Lucina Suarez indicated that the Texas Cancer Registry has made limited progress in improving data for special populations in the state because of the lack of adequate funding. The registry is grossly underfunded and it is struggling to get timely and complete data for the entire state. Therefore, there are no resources to improve or assess the quality of the data for any ethnic group. The two largest minority groups in Texas are Mexican Americans (about one-fifth of the state's population) and African Americans (about 10 percent of the state's population).

The registry has started to use a combined source of data for the identification of Hispanics using the medical record and Census list of Spanish surnames. Dr. Suarez mentioned that the registry no longer uses the GUESS program. She also noted that there are no resources available to do validity or reliability audits on the Hispanic classification in the registry. Texas recently received funding from the CDC to examine the completeness of the data for prostate cancer in African Americans.

Centers for Disease Control and Prevention

Dr. Daniel Miller provided an overview of the activities of CDC's Division of Cancer Prevention and Control (DCPC) with regard to improving information about cancer and risk factors as they pertain to ethnic minority populations. The DCPC has grown significantly since 1991, when its budget was approximately $2 million. The current budget for the DCPC is about $175 million per year, much of which is allocated to the NBCCEDP. The DCPC has expanded to four Branches—the Program Services Branch, the Cancer Surveillance Branch, the Health Services Research Branch, and the Behavioral Science and Health Communications Branch. Dr. Miller described some of the activities of each of these four Branches.

The NBCCEDP is administered by the Program Services Branch. The purpose of this program is to promote screening and diagnostic exams for minority and uninsured/underinsured women in the United States. The program has funded approximately 2 million screening tests (e.g., mammography, Pap smears) and diagnostic tests for uninsured/underinsured and minority women. The NBCCEDP also emphasizes public education, the need for early detection in various communities, health care provider education, community outreach, and surveillance. In addition, it evaluates and replicates community interventions. Health services data on every woman who is screened in the program is recorded and reported to the CDC. As a result, the CDC has a database of about 1 million records that include demographic information, results of mammography and Pap smear screening tests, the diagnostic procedures (and results) that were performed, and the date of initiation of treatment. These data are distributed to the breast and cervical cancer programs in the states to provide them feedback on the quality and completeness of the tests being performed as well as interpretations of what the data mean for cancer outreach programs and the need to expand their services to different areas within their state.

The Cancer Surveillance Branch administers the NPCR, which was established by a Congressional mandate in 1992 and received its first funding in 1994. The purpose of the program is to establish and enhance cancer registry operation. The NPCR currently is active in 45 states (including all of the non-SEER states), the District of Columbia, Puerto Rico, Virgin Islands, and Palau. The program also undertakes operations research, program development, and surveillance and health services research. Under operations research and program development, the NPCR provides grant supplements to improve

the completeness, timeliness, and quality of the data collected by the states. In 1999, $1.8 million in noncompeting supplements were provided to 48 of the 49 grantees. The NPCR did not mandate how the funds were to be spent; instead, the program allowed the grantees to develop approaches appropriate for their local circumstances.

An example of the NPCR's surveillance and health services research is its collaboration with NCI, ACS, NCHS, and NAACCR in preparing the annual report to the Nation on cancer. They currently are working on the third edition of this annual report. Last year's report provided information on cancer by race and ethnicity in the United States by whites, blacks, Asians/Pacific Islanders, Hispanics, and American Indians, etc. The CDC also generates routine surveillance summaries for breast, cervix, prostate, and colorectal cancers. One on breast and cervical cancer mortality will be published soon. In addition, the CDC has projects related to the BRFSS data, including a recent publication on breast and cervical cancer screening in the United States. Although there are limited resources, the DCPC also has a number of special projects to examine special populations and health disparities. For example, the CDC is funding four sites through the Association of Teachers of Preventive Medicine and the Association of Schools of Public Health to assess the racial and socioeconomic disparities in cancer incidence and mortality. There are plans to fund patterns of cancer care studies focused on how the cancer care provided compares to consensus standards. In addition, the CDC has several data enhancement initiatives, including one that is facilitating linkage of state data to the National Death Index; these linkages will enable the CDC to generate better statistics from the NPCR registries with regard to survival. There also are efforts to link state data with Census data to append SES indicators to ecological analyses of NPCR data. Other initiatives involve improving geocoding and medical claims linkages.

The data from the NPCR registries can be used to monitor trends in cancer incidence and mortality and to guide cancer control planning and the allocation of health resources. In addition, the data can be used to evaluate cancer prevention and control activities. Dr. Miller noted that there is a need to evaluate CDC's efforts. Nearly $1 billion have been spent on screening through the NBCCEDP. The data from the cancer registries can be used to evaluate the impact of this program in the communities. NPCR data also can be used to advance clinical, epidemiological, and health services research.

The Health Services Research Branch has an active program in conjunction with the NCI for the recruitment of African Americans into the PLCO trial (focused on Detroit and Alabama); evaluation of successful recruitment strategies of African Americans into prevention trials; and knowledge, attitude, and belief surveys of African American men related to prostate cancer and screening.

The Behavioral Science and Health Communications Branch works on the development of national public and provider education campaigns through the media and through formative research. The active areas of research include informed decision making in prostate cancer for African American men and barriers to colorectal cancer screening in African Americans and Hispanics.

Additional CDC activities to improve data on race and ethnicity, include the analysis of national surveys on adolescent and adult smoking behaviors being conducted by the Office on Smoking and Health; the Youth Risk Behavior Survey (YRBS), which includes data on adolescent and youth smoking; and CBO grants, four of which are focused on African Americans and one on Vietnamese, awarded by the Division of Adult and Community Health.

Dr. Miller pointed out that there are a number of challenges ahead, including:

- ✧ Resources—the states are underfunded by 40-60 percent to do the effort that will bring them to the minimum levels of completeness, timeliness, and quality.

- Coordination—this is critical because there are limited resources and data linkages provide valuable information.
- Communication—meetings such as this will help the organizations involved in surveillance to learn and understand what is being done in this area; these meetings also will help us identify ways to coordinate and work together to address the challenges.
- Data completeness and quality as written in the medical record—physicians and others in hospitals need to be educated about the importance of recording the information that is needed by registries.
- Local perspectives versus national perspectives—the solutions to improving the problems concerning data on race, ethnicity, and SES lie in the local areas. The decision making and problem solving are facilitated by local partnerships.

Indian Health Service

Ms. Roberta Paisano presented the Indian Health Service's (IHS) response to the recommendations developed by the Data Working Group at the September 1999 meeting. With regard to the first recommendation concerning the collection of SES information for cancer patients, the IHS responded that insurance status is the only item captured in the medical record. Ms. Paisano noted that some data may be captured through CDC's NBCCEDP. In response to the second recommendation on the development of meaningful numerator/denominator statistics, the IHS is aware of miscoding and that its data is only as good as the U.S. Census data. She noted that the issue of improving numerators has been addressed in the IHS publication, "Adjusting for Miscoding of Indian Race on State Death Certificates."

The IHS had no comment on the third recommendation on developing methods for overcoming data collection communication barriers. Concerning the fourth recommendation to collect nativity data, Ms. Paisano pointed out that the tribal designation should not be shared without prior approval from every tribe mentioned in any report. The IHS had nothing to report in response to the fifth recommendation on collecting cancer data in a disaggregated manner. In response to the sixth recommendation regarding community involvement in data collection, Ms. Paisano mentioned that there is a Network for Cancer Control Research Among American Indian and Alaska Native Populations that serves as a conduit between tribal communities and researchers. She also noted that the IHS IRB looks at cultural issues before approval is granted and that the IHS Epidemiology Centers work closely with tribal health boards and community members (the IHS works through these Centers to collect data).

In response to the seventh recommendation involving the training of epidemiologists and related professionals, Ms. Paisano indicated that there is a Native Researchers training course that has trained approximately 45 students to date. In addition, there is the EIS training through CDC/IHS, student internships at the IHS Epidemiology program, and discussions are underway with the CDC for training IHS/Tribal health professionals in epidemiology. In response to the eighth recommendation on risk factor collection and analysis, Ms. Paisano mentioned the collaboration of the IHS with the CDC to conduct health risk surveys. She noted that all of the IHS Epidemiology Centers use risk behavior surveys and she mentioned that the IHS is collaborating with the Bureau of Indian Affairs to conduct youth risk assessments. With regard to the ninth recommendation on SEER expansion, Ms. Paisano indicated that the IHS agrees that current SEER data do not represent all of Indian Country and supports expanded cancer activities, albeit without funds.

Ms. Paisano responded to the tenth recommendation regarding CDC's improvement of data at state registries by noting that the IHS has matched data with the following state tumor registries: California, New Mexico, Arizona, Minnesota, Wyoming, and Montana. This matching has been a slow process and may have been done only on a one-time basis in some instances. Ms. Paisano suggested that the CDC

assign an individual to the IHS to work on improving the quality of Native American data at the state level. In response to the eleventh recommendation regarding the collaboration of federal agencies to provide comprehensive national data, Ms. Paisano replied that the IHS would cooperate with such a collaborative effort. With regard to the twelfth recommendation on involving persons from the community being studied to provide advice on culturally appropriate cancer control materials and instruments, Ms. Paisano indicated that tribal consultation is occurring. In response to the thirteenth recommendation on studying special populations with low incidence of cancer to determine "protective factors," Ms. Paisano replied that the IHS agrees with this recommendation as long as it is carried out with the consent of the communities.

American Cancer Society

Dr. Phyllis Wingo explained that the ACS is not in the data collection business. The role of the ACS regarding the uses of cancer statistics is one of translation and interpretation for the public, the media, and a large segment of the clinical community. The ACS makes extensive use of the data collected by the SEER program, the CDC (mortality statistics, risk factor surveys), and the NAACCR. The ACS publishes *Cancer Facts & Figures* in January each year for these audiences. This publication includes:

- A section on cancer in minority populations (appears every year).
- A banner section on a specific cancer site that includes incidence and mortality statistics according to five Census population groups; the 1999 *Cancer Facts & Figures* banner section provided statistics about cancers of the colon and rectum.
- The first banner section in 1997 was on cancer and risk factors by race and ethnicity; this banner section topic will be repeated in 2002.
- The 5-year survival statistics for whites and blacks (appear every year).

The ACS prepares *Cancer Statistics*, which is published in the January/February issue of the journal *CA—A Cancer Journal for Clinicians*. This publication is targeted primarily to clinical and public health audiences. It was recognized in 1997 as one of the most cited publications in the epidemiologic literature and contains many of the same statistics as *Cancer Facts & Figures.* Every other year, the ACS publishes *Cancer Facts & Figures for African Americans*, which was first published for 1996. This publication is targeted to the same audiences as *Cancer Facts & Figures*, and it includes the estimated numbers of new cancer cases and cancer deaths among the black population in the United States for the upcoming year. *Cancer Facts & Figures for African Americans* contains prevalence estimates of tobacco use, diet, obesity, and physical activity among blacks. It also contains screening prevalence for mammography, Pap smear, proctoscopy, FOBT, and digital rectal exam (DRE) for blacks.

With regard to surveillance research, the ACS plans to expand its presentation of special population data in 2000. Preliminary ideas for the expansion include developing *Cancer Facts & Figures* for Asians and Pacific Islanders, Hispanics, and American Indian/Alaska Native populations. These publications meet several service needs within the larger ACS organization, including program planning for cancer control and patient services, tracking progress against this disease by monitoring trends in cancer incidence and mortality, and assessing trends in cancer risk factors. The nature of these trends are used to assess progress toward Healthy People 2000, Healthy People 2010, and ACS 2015 goals.

Dr. LaMar McGinnis indicated that the ACS does not collect surveillance data for the major cancer sites; however, it does consume such data for analysis and interpretation. The ACS then widely disseminates this analyzed data for public information, for use in cancer control, and for advocacy. Because the ACS has renewed its commitment to be data driven, it has become more dependent on

quality cancer data that are widely reflective of all of the Nation's communities. These data will enable the ACS to develop more cancer activities that have a greater impact. Such data are used by the ACS for surveillance, epidemiologic, behavioral, and cancer control research and program development.

The ACS is a nationwide organization with a presence in more than 3,000 communities, served by staff in 17 divisions and a National Home Office. Approximately 2 million volunteers are assisted by a large national and divisional staff utilizing data to affect local cancer control activities. In 1994, the ACS, in concert with the Commission on Cancer and the National Cancer Registrars Association (NCRA), formed a TRIAD program. This program involves representatives from the three organizations to develop local cancer control programs in each state. This activity was enhanced significantly in 1998, when the TRIAD program was converted from a state to a division activity, thereby bringing a larger critical mass of representatives together to study and develop programs. The TRIAD program—which now includes the Mid-South, Mid-Atlantic, and Heartland Divisions and will extend to involve all ACS divisions—utilizes local surveillance data to impact on local and state cancer control and advocacy programs. The Mid-South TRIAD was the first pilot to be implemented. The Mid-South TRIAD decided to target breast cancer, cervical cancer, and tobacco control. They established partnerships and defined the activities and measures. As the TRIAD program moves forward, different cancer sites will be selected. Each TRIAD will determine, from available data—National Cancer Data Base, state registries, NCHS, SEER, BRFSS, and YRBS—its cancer control priorities, and then develop cancer control and advocacy programs based on those priorities.

The ACS operates a National Call Center at 1-800-ACS-2345 and a Web site (**http://www.Cancer.org**). Data to assist patients in decisions related to prevention, detection, and treatment for cancer are available through these two venues. In closing, Dr. McGinnis pointed out that the data presently collected have minimal influence on local cancer control. These data are difficult to obtain and interpret. They also are difficult to apply because of a clinical disconnect and a lack of relationship to local public health activities and to clinical practice. It is anticipated that the data needs will include stage-related incidence, stage-related outcome (mortality), and therapy-related outcome to include short- and long-term complications, the development of secondary malignancies, and quality of life issues including return to productivity. This local activity, when analyzed on a national basis, may depict regional variations that could again focus activities. Participation in clinical trials also would be an important aspect for analysis. Dr. McGinnis noted that all of this ultimately will impact upon an increased emphasis for evidence-based medical practice.

American College of Surgeons/Commission on Cancer (CoC)

Ms. Lynda Douglas described the community-based activities of the CoC. She reported that the CoC is working with the ACS on the TRIAD program to develop cancer control initiatives at the regional, state, and local levels. Local initiatives will be carried out by ACS staff, cancer liaison physicians, and targeted organizations at the community level based on specific needs of that community related to prevention, early detection, and treatment burden identified (through the NCDB, state registry data, and the BRFSS) in their area. Major cancer sites will be reviewed first, looking at stage, treatment, ethnicity, socioeconomic levels, and other items. Work has been initiated in the Mid-South and Heartland Divisions of the ACS and it will be expanding over the next few years throughout the country. Facilities are required to show an effort to coordinate programs/services with relevant community agencies, resources, cooperative groups, and other health care institutions.

Ms. Douglas provided an overview of the collection of CoC's race, ethnicity, and sociodemographic/ socioeconomic data. The *National Cancer Data Base Report on Malignant Epithelial Ovarian Carcinoma in African American Women* used multivariate statistics to control simultaneously for region of the country, patient race, and residential area income. The variables used were nationality, place of

birth, and race. The *National Cancer Data Base Report on the Relationship of Race and National Origin to the Histology of Nasopharyngeal Carcinoma* addressed one of the few circumstances where race and nationality are known to play a role in etiology. This was part of a series looking at rates of African American women with reproductive malignancies.

The NCDB has a program that generates "residential area income" from zip codes. Ms. Douglas suggested that residential area income may be a more appropriate measure than trying to determine a way to measure personal or household income, especially in areas where most patients currently are retired and many of the females had no personal incomes even in their younger years. She noted that "surrounding income culture" may have a more direct effect on patients and provider choices for medical care.

The CoC will be complying with the new standards for race coding set forth by the OMB. Ms. Douglas mentioned that race is unknown for less than 2 percent of NCDB data. She noted that NAACCR, SEER, and American College of Surgeons (ACoS) have expended considerable time and resources to develop a complete data set for collection of race codes. Because there is minimal race information available to the registrar who looks at the medical record (primarily the face sheet), the registrar must interpret the race information in the medical record into the most appropriate code available. This interpretation is probably the cause of a great deal of race misclassification (numerator). She suggested that the most efficient way to handle this disconnect would be to invite the Joint Commission on Accreditation of Health Organizations (JCAHO) to future ICC meetings with the intent of developing a uniform data set for collection of information on intake data for the medical record. This also could include other possible SES variables (e.g., level of education, occupation) for collection on the hospital intake form. The ACoS currently has a contractual agreement with JCAHO, and JoAnne Sylvester has offered to facilitate the invitation of JCAHO to future ICC meetings for this purpose. Ms. Douglas also suggested that future meetings include individuals with direct data collection experience.

Race and Hispanic Origin Reporting in Census 2000: The Implementation of the Revised OMB Directive on Race and Ethnicity in Census 2000

Dr. Manuel de la Puente provided a chronology of OMB's review of standards for collecting and reporting data on race and ethnicity. The OMB review began in 1993 and Congressional hearings were held from 1993 to 1994. Public hearings were held and the Federal Interagency Committee were appointed in 1994. Research was conducted by the Census Bureau and other agencies from 1994 through 1997. Public comments were received in response to *Federal Register* notices in 1994 and 1996. In July 1997, the report and recommendations of the Interagency Committee were published in the *Federal Register* for public comment. In October 1997, the OMB issued "Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity." The new OMB race categories include: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and white. The Census Bureau allowed the addition of "some other race" for Census 2000 and its dress rehearsal.

The following two changes will substantially affect data products: (1) respondents may mark (select) one or more races, and (2) the "Asian and Pacific Islander" category was split into "Asian" and "Native Hawaiian and Other Pacific Islander." Question 5 on the Census 2000 dress rehearsal form refers to Hispanic origin and race. It asks the respondent to identify whether he/she is Mexican/Mexican American/Chicano, Puerto Rican, Cuban, or Other Spanish/Hispanic/Latino. Dr. de la Puenta pointed out that the Hispanic question precedes the race question because research has indicated that this order will improve reporting. The revised OMB Directive specifies that when there is self reporting as in Census 2000, a two-question format should be used and the Hispanic origin question should precede the race question. The categories for the Hispanic origin question are the same as those used in 1990 Census.

How will the Census 2000 data be tabulated? The Census Bureau and the OMB have grappled with this issue and have established a working group to examine it. The OMB provided some preliminary guidance for tabulations when the revised Directive was issued in October 1997. This guidance included:

- Show number of respondents who marked only one category, separately for each of the five racial categories.
- "Consistent with criteria for confidentiality and data quality, show as much detailed information on race as possible."
- The "overriding objective is to provide the most accurate and informative body of data."

The Census Bureau is trying to determine how much detail regarding race/ethnicity should be shown in data tabulations and data products from Census 2000. All of the detail will be available in the data tapes, but not in the data products. Basically, there will be seven mutually exclusive categories of race: white alone, black or African American alone, American Indian or Alaska Native alone, Asian alone, Native Hawaiian and Other Pacific Islander alone, Some Other Race alone, and two or more races. These categories are consistent with the preliminary guidance issued by the OMB. Data on Hispanic origin also will be available. Dr. de la Puenta pointed out that there are 63 possible combinations of racial categories.

Dr. de la Puenta emphasized the importance of the Census and noted that the Census Bureau is required to provide redistricting data (Public Law 94-171) to the President and Congress, which includes:

- Total and voting-age population
- Race and Hispanic or Latino origin
- Lowest geography: Census block level
- 63 race categories by Hispanic or Latino.

Official state apportionment counts must be provided on December 31, 2000, and redistricting counts on April 1, 2001. He acknowledged that reporting of multiple races may make the determination of denominators quite complex.

Discussion

Dr. Hiatt asked if the use tapes will allow investigators/users to separate out the races that are involved in the two or more category. Dr. de la Puenta responded that the tapes will allow such separation. He also mentioned that the Census Bureau's American Fact Finder (on the Web) will allow the user to get the detail required. Dr. Hiatt also asked about the targeted testing performed by the Census Bureau. He noted that 1.5 percent of the white population is larger than any other minority population. What white population was selected for the testing? Dr. de la Puenta responded that in the targeted test, the Census Bureau selected specific samples throughout the country, particularly areas concentrated with specific minority groups (e.g., white ethnics, African Americans, Hispanics/Latinos). In the white targeted sample, a total of 1.5 percent residing in the target area (not all individuals were white) selected more than one race. In areas where Asian and Pacific Islanders were targeted, about 10 percent of the individuals residing in those areas reported more than one race. Therefore, the disparity in reporting more than one race was based on the area and not the race. Dr. Suarez asked what proportion of Hispanics selected two or more races. Dr. de la Puenta replied that in the Hispanic targeted sample,

approximately 7 percent of individuals residing in that area selected two or more races. He noted that about 7 percent of the Hispanics surveyed in the national American Community Survey in 1996 reported that they were multiracial.

Dr. Daniel Miller pointed out that the numerator data collected from medical records will not correlate with the denominator data from Census 2000 for those individuals identifying themselves as two or more races. He noted that there is no incentive for a private, for-profit hospital to make changes in its data systems because of the associated cost. Therefore, the data available to registrars may identify only one race. Dr. West noted that younger respondents tended to select two or more races. If that is correct, the impact of the multi-race reporting in Census 2000 may not be immediate because most of the responses are provided by older individuals. Dr. de la Puenta reminded participants that the Census involves proxy reporting—one adult in the household will respond to the race question for all members of the household.

Dr. Edwards asked Dr. de la Puenta to describe any planned studies to address these unknowns and to comment briefly on the American Community Survey. Dr. de la Puenta responded that the American Community Survey was initiated by the Census Bureau in 1996, with hopes that this survey will replace the Census long form in 2010. It collects the same data elements that are on the Census long form, including race and Hispanic origin. The survey is ongoing in 31 counties in the United States. There are plans to continue to expand the survey until 750,000 households nationwide are surveyed. In parts of the country where the population is dense, updated statistics will be available annually. In those parts of the country that are less populated, updated statistics will be available every 3 years (on average). Dr. de la Puenta indicated that plans for expanding this survey will go forward if funding is available from Congress. To address the unknowns that may result from the Census 2000, the Census Bureau is conducting a study that involves sending the 1990 Census question regarding race to 5,000 households and the 2000 Census race question to another 5,000 households. There is another survey planned for late 2001; this survey will include a race question that will allow respondents to select more than one race. The persons who mail back that form will be interviewed by telephone. During this interview the individual will be asked the race question again, and if he/she selects more than one race, the person will be asked to identify the racial group with which he/she most identifies. The purpose of this study is to obtain some insight into how multiple race responses could be allocated into one category if necessary. No other research to address the unknowns is planned. Dr. de la Puenta predicted that there will be only one question regarding race and Hispanic origin on the Census 2010 form.

Dr. Wilson suggested that the Census Bureau include a followup question on the Census form that asks the individual to identify the race with which he/she most identifies. Dr. de la Puenta replied that such a question could not be included on the Census 2000 form because there is no legal requirement to collect such data. Dr. Wilson asked who was involved in determining the new OMB categories. Dr. de la Puenta responded that the determination was made by a committee that included representatives from the Census Bureau, the Bureau of Labor Statistics, CDC, Justice Department, the Internal Revenue Service, and any other federal agency that collects data on race and ethnicity. The OMB made the decision regarding the new categories based on input from the committee, public hearings, and research.

Dr. Hiatt noted that the new categories can be aggregated into the same categories used in the past. Dr. de la Puenta pointed out that, depending on the area, there may be a significant number of individuals who report two or more races. How do these individuals get assigned to the old categories? Dr. Armin Weinberg asked what the Census Bureau plans to do to validate this self-reported data. Dr. de la Puenta replied that the Census Bureau never validates self-reported information. Although this is not validation, the Bureau does reinterview a subset of persons who mailed back the form to determine consistency of reporting. Whites and African Americans usually are very consistent in reporting, but American Indians are not. Dr. de la Puenta noted that consistency of reporting is an indicator of data quality, but not validity.

Dr. Friedell asked Dr. Howe if NAACCR registries are incidence registries. Dr. Howe replied that the registries are focusing on incidence data at this time. Dr. Vivien Chen asked if the data on specific ethnic groups of Asians would be available from the Census 2000 data tapes. Dr. de la Puenta replied that such data would be available to users through the American Fact Finder. One participant asked about how to define denominators. Is it reasonable to look at denominators as those individuals who identified themselves as a single race or those who identified themselves in a category no matter how many races they selected? Dr. de la Puenta responded that there will be many more options for defining denominators. He suggested that research needs to be done to determine the impact of different methods of defining denominators on rates. Dr. Charles Key noted that everyone will have to continue to do the best they can. He stated that "recognizing the limitations of existing enumeration and classification systems for both the incident cancer cases and the corresponding populations at risk, and the lack of consistency of individuals responding to questions about their race and ethnicity, our goal is to produce useful information that is reasonably reproducible and consistent over time, but never perfect." He noted that this will continue to be the mode of operation.

Dr. Hiatt asked Dr. McGinnis to comment on how the interventions are selected in the TRIAD program. Dr. McGinnis replied that the Mid-South TRIAD assembled a group that identified a specific problem in an area and then they worked together to determine what could be done to address that problem. Dr. Hiatt made a plea for those decisions to be evidence based. He mentioned that the CDC is producing a document entitled the "Guide for Community Prevention Services." This is a compendium of evidence of all that is known about cancer control interventions. Dr. Friedell pointed out that the TRIAD program offers an opportunity to look at quality of care within the hospital and how this affects patients. Dr. McGinnis replied that the participants in the TRIAD program are committed to evidence-based decision making.

Conceptual Issues, Measurement and Data on Culture, Acculturation, and Cancer Control

Dr. Marjorie Kagawa-Singer indicated that much of what has been discussed at this meeting centers on the following two questions: What is culture? What are the more proximal variables that are actually at play when you look at different vulnerability and risk factors within each of the various race and ethnic groups? Race distinctions began in the 17th century, when it was assumed that there were subspecies of *Homo sapiens sapiens*. Implicit in this assumption was a hierarchy of evolution, which stated that those of white skin were at a higher stage of evolution than those of darker skin. This is the social and political underpinning to the racial/ethnic categories. Although there now is evidence that race has no scientific basis, the categories have been retained.

Dr. Kagawa-Singer provided several definitions of relevant terms. What is data? Data are measurable indicators of a conceptual construct. What are we trying to capture in our data collection? We are trying to capture information on culture, not race. Culture is dynamic, heterogeneous, continuous (not dichotomous), and multidimensional. It is important to recognize that there are within group as well as between group variations. Dr. Kagawa-Singer showed photographs of several tapestries and asked participants to identify their culture. She noted that within any cultural group there are expressions of belief systems that are inherent in that culture. In the western culture, tapestries usually contain people and they are bordered, indicating that the human form is central as well as a preference for mutually exclusive categories. In terms of health messages, encouraging a western women to get a mammogram for herself is a powerful message. In the American Indian culture, there are no human forms. There also are no borders, indicative of the oneness of world view. A health promotion message designed for the white American culture probably would have little relevance or motivating power. How do we capture belief systems and world views in our programs that take into account culture?

There are four elements that are considered to define culture: religion/world view, language, social structure, and beliefs and values. However, the more anthropological definition of culture takes into

account the ecologic niche (i.e., environment, economy, and technology) in which a population has developed. Practices and beliefs have evolved to promote the survival and well-being of the individuals in that niche. Because cultures that exist now have been around for thousands of years, these practices and beliefs are functional. When individuals migrate to different areas or come into contact with those who are different from themselves is when these individuals become maladaptive, could be more adaptive, or are neutral. Dr. Kagawa-Singer defined race as genotype, not phenotype. Culture is the system of beliefs, values, lifestyles, ecologic, and technical resources and constraints on a group of people. Ethnicity is that subcultural group in a power structure of a multicultural society. Both culture and ethnicity constrain behavior of a self-identified group membership. Race becomes important with regard to the issue of racism, which is the assertion of power, ego fulfillment, and status at the expense of others by skin color. This power often is exerted unconsciously. An article in the February *New England Journal of Medicine* described how medical decision making and treatment differed considerably by gender and skin color and how it has affected health outcomes.

Dr. Kagawa-Singer identified a number of factors that could be used to measure ethnicity, including SES, access to health care, family structure/support system, ethnic group, reasons for immigration, and religion/spirituality. Most of these are measured in many demographic studies, but the difference is in how they are interpreted. This is where the participatory research models become powerful because individuals in the communities with which the researcher is working will inform him/her of how best to interpret the various factors. Dr. Kagawa-Singer then reviewed the U.S. policy for assigning racial status on birth certificates prior to 1989. She pointed out that African Americans status can be passed down from either parent; however, for the Japanese, the ethnicity depends totally on the father. In Brazil, race is defined by the degree of color in the skin and there are eight different categories, which are closely associated with social status.

Rather than using the cultural element approach, the researcher should use the cultural perspective. The cultural element approach is static and stereotypical and does not capture the behaviors pertaining to risk factors that would be of interest. The cultural perspective is a systems approach that looks at variations of principles within the social context of the individual and the group to which he/she belongs. This would be described in an ecologic model showing that one needs to take into consideration the structural and environmental resources, the availability and accessibility of resources, the cultural group or the region of residence, the political class and social constraints of the population of focus, ethnicity, social network and family structure, and the impact on disease-specific issues.

The other major issue related to culture is the concept of acculturation. Dr. Kagawa-Singer pointed out that people are not one thing or another, especially when living in a multicultural community. To determine how acculturation affects a community, she described a study that estimated isoflavone intake in a Japanese population (primarily third generation) that is part of a study on menopause. When comparing Japanese intake to that of the white population, the Japanese intake of isoflavone (primarily from soybean products) was considerably higher for the Japanese. Therefore, isoflavone appears to be a poor indicator of acculturation for Japanese even after years of living in the United States. Referring to another study on food selection in a homogeneous enclave of non-Asians in Korea town, she cautioned that there may be acculturation effects in the referent group from which the control is drawn. She noted that there are considerable changes occurring in the diets of the referent group.

Although Canada does not collect data on race, it does collect population data on ethnicity external to the cancer registry data systems. This occurs within the white as well as ethnic minority populations. How valuable is it to maintain one's cultural identity and characteristics? Is it considered to be of value to maintain relationships with other groups? Is it possible to do so? To measure ethnicity it will be necessary for individuals to self identify ethnicities and heritages (allow selection of more than one). Place of birth also should be included. If an individual was not born in the United States, he/she should be asked for his/her age at the time of immigration. Dr. Kagawa-Singer also suggested that individuals

be asked a question regarding the degree of value/pride in their ethnicity/heritage. For each self-identified ethnicity, an individual should be asked to specify how much he/she identifies with that ethnic group (from not at all to completely). The individuals also should be asked how much they identify with the dominant U.S. modern culture. Dr. Kagawa-Singer concluded her presentation by stating that such questions may provide much needed information on ethnicity for refined analyses.

Dr. Lillian Tom-Orme indicated that many acculturation studies identify indicators to measure acculturation, such as education, insurance, income, occupation, spoken language (fluency, amount, etc.), economic resources, perceptions of control/locus of control (external versus internal), perceived access to care or treatment/services, residency status/how long in the United States, screening behaviors for breast and cervical cancer, cancer attitudes and beliefs (fate, fatalism, fear of screening), foodways (how foods are used), modernity, modesty issues/embarrassment, use of traditional medicine, knowledge (lack) about screening, breast self-exam skills, comfort level in asking for screening tests (mammogram, CBE, and Pap test), specifics (cradle board use, naming ceremonies, use of corn pollen), and thinking/dreaming in English (a measure of fitting into the local culture).

Dr. Tom-Orme described three models—Giger and Davidhizar's transcultural assessment model, Purnell's model for cultural competence, and Leininger's "sunrise" theoretical/conceptual model of transcultural care diversity and universality. She noted that a number of models have come from the nursing field, probably because nurses are very close to the patient population and they understand that models cannot be applied to every person.

Leininger's culturological assessment refers to a systematic appraisal or examination of individuals, groups, and communities as to their cultural beliefs, values, and practices to determine explicit nursing practices within the cultural context of the people being evaluated. The guiding principles for this model are to: (1) maintain objectivity and an open attitude, (2) avoid seeing all people as alike, (3) look for degrees of cultural variations, (4) begin with one culture, and (5) avoid ethnocentrism. Purnell's model is a macromodel that looks at all the various components of culture. The "sunrise" model came out of Leininger's work in the early 1960s. This model identifies parameters that describe people and their world views in relation to health care patterns, needs, and practices.

Discussion

Dr. Hiatt asked Dr. Kagawa-Singer about delivery of cancer control interventions in a multicultural, urban area. He asked if she considered similarities across ethnicities cultural variables. Can culture be used separate from ethnicity or is it a subset of ethnicity? Dr. Kagawa-Singer replied that culture is a subset of ethnicity. Dr. Hiatt asked whether a public health intervention designed on the basis of family is using an ethnic or a cultural concept. Dr. Kagawa-Singer responded that there is a universal element to the concept of family. Infants require nurturing and family fulfills that need. Therefore, family becomes a constant, but the form that it takes varies among ethnic groups. How families infuse a sense of identify, well being, pride, and integrity varies among ethnic groups.

CBOs: Role in Data Collection, Use, Reporting, and Quality

Ms. Susan Shinagawa described the community perspective with regard to cancer data. She pointed out that most CBOs are still struggling to get direct health services to the community because of limitedbudgets; these organizations have not been focused on conducting research or collecting data. Nevertheless, they are a significant source of data and there is a need for data at the community level. Analyses by researchers who are not from the community often result in faulty conclusions and lead to misperceptions. As a result, the reasons for low compliance are often misinterpreted. For example, the

low utilization of health care services among the Asian population has led some researchers to conclude that the Asian population does not need health services. Other researchers have concluded that the disparate health needs among special populations are because the individuals are not compliant (e.g., did not follow up after a screening test). Another example is that physicians have not recommended followup for Asian women who found lumps in their breasts because Asian women "do not get breast cancer." Ms. Shinagawa pointed out that although the incidence of breast cancer in Asian women is lower than in the white population, they do get breast cancer. In fact, after two to three generations in the United States, an Asian woman's risk becomes the same as that for a white woman. She cautioned that researchers often draw conclusions regarding health behaviors and health services utilization from small samples that are applied to the aggregate population. In addition, conclusions about newly immigrated subpopulations get applied to those who have been in the United States for generations.

Many behavioral surveys do not reach certain minority populations because many of these minorities will not talk to anyone on the telephone. In addition, most of these surveys are conducted in English and many minorities in the United States speak little or no English. Institutional racism as well as overt and covert racism have limited the resources available to address the needs of these special populations.

Misclassification is another problem for minority communities. Sustainability also is a problem, primarily because of lack of funding. Disseminating information to the organizations/individuals that need it is another issue. In addition, there is an inadequate number of students from minority/ethnic communities currently in the system who can conduct the needed research and serve as liaisons between the medical community and the population.

Ms. Shinagawa noted that there are some good models available. The best model appears to be the partnering of an academic center with a CBO or a community health coalition that provides direct services to the community. She noted the importance of including the CBO as a full partner in such a model. The academic and community researchers should be co-principal investigators and the CBO should be the fiscal agent. It also is important that the CBO recognize the importance of collecting and analyzing data that can be applied universally as well as those that are culturally specific. Examples of best practices include the Appalachia Leadership Initiative, CDC's Racial and Ethnic Approaches to Community Health (REACH) program, and the California Breast Cancer Research Program. The REACH program is attempting to involve communities in data collection and analysis so that the information can be used by the communities. Ms. Shinagawa noted the following needs: more funding to support fully collaborative research partnerships, an increase in the number of students who pursue community health research, inclusion of medically underserved groups in research efforts, and incorporation of evaluation, replication, and sustainability in research efforts.

Ms. Shinagawa agreed with the OMB Directive that separates Asians and Pacific Islanders. It is very important to separate these two groups because the Pacific Islanders comprise less than 5 percent of the combined group and they often are lost in the data. Another rationale for separating these two populations is that the major cancers in Asians are rare in the white population in the United States; however, the major cancers among Pacific Islanders tend to be the same as the major cancers in the U.S. non-Hispanic white population. She cautioned against assuming that all Pacific Islanders are the same.

Dr. Marjorie Kagawa-Singer presented some data from CBOs. Ethnic misclassification is a major issue in minority communities. She indicated that in an infant mortality misclassification study, there was a 54 percent variance between birth and death certificates among Japanese and a 48 percent variance among Chinese infants. Unless there are accurate data, the predictions and policies based on the data are precarious.

Breast cancer screening data from the 1992 and 1994 National Health Interview Survey (NHIS) indicate that Asians and Pacific Islanders are much more vulnerable to indicators that put them at high risk than are uninsured whites and other ethnic groups. Review of the breast self examination, clinical breast examination, mammogram, and Pap smear screening rates by ethnicity indicates that ethnic matching between physician and patient tends to reduce screening. Therefore, patients who visit an ethnic physician who speaks the same language may not necessarily receive better health care. The NHIS collects data on 10 ethnic groups of Asians/Pacific Islanders. However, most studies do not report data on the Pacific Islanders because the sample sizes are so small. For one study on Pap smear screening of women 18 and older, the sample size was adequate for statistical analysis of Asian Indians. This study indicated that education had less effect for Asians (even at higher levels of education) than for the white population. Therefore, improving education, which is a usual barrier to care, among Asians and Pacific Islanders probably will not improve screening rates.

Dr. Kagawa-Singer identified the following future directions for research and practice:

- Disaggregation of Asians and Pacific Islanders.
- Instrumentation—language and concepts.
    - Validity
    - Reliability
- Sampling bias (potentially missing 37 percent of the population who do not speak English).
- Representativeness.
- Control (e.g., SES).
- Bicultural personnel should conduct interviews.

Dr. Armin Weinberg described the role of CBOs in restructuring and using cancer data. He noted the importance of recognizing that the data needs of the community may differ from those at the national or state level. At the community level, there is a desire to ensure that any unique difference—whether in screening rates, lifestyle factors, incidence, prevalence, or outcomes—is understood. Similarities with other communities when comparing one community to another are helpful, but not if the comparison is not valid. Organizations typically mobilize members of their communities by finding commonalities. Participation is predicated on the belief that by working together they can make a difference. How they measure their success may come in a variety of outputs. For example, increased screening may be measured by simple participation or by improved detection rates at early stages of disease when applicable. Some may measure success by the amount of community support gauged by fund-raising efforts or through participation at educational programs.

There are hundreds of organizations interested in combating cancer at the national, regional, and local community levels. Some are broad based in their scope while others are organ specific. Some are interested more in research while others focus on prevention, screening, and survivorship. Some are led by professionals and others are led by lay persons. The bottom line is that each organization believes that it has sufficient evidence to rally financial and other support and to draw attention to its mission. What are some of the roles CBOs can play in improving data collection, surveillance, and analysis? It is critical that CBOs understand their roles in supporting the funding necessary for data collection and analysis. Dr. Weinberg speculated that most CBOs have little idea of the cost involved in achieving the level of detail and completeness of the data on which they often rely. These organizations need to gain an appreciation of this as well as an understanding of the tradeoffs that must take place when funding is limited. It is equally important that CBOs are granted the option to “purchase” specific add-ons, special analysis, over sampling, etc. Likewise, national and state efforts should create opportunities for tailoring

"standard" data collection procedures to accommodate this need as much as possible. At a minimum, efforts to educate CBOs about the strengths and limitations of existing databases must be undertaken. For example, if a local community data collection or survey is conducted it should be clear whether or not it can be correlated with other existing data sets, and if not, why not. Assistance also might be given on design methodologies and questionnaire design and sampling. Partnerships in data efforts should be a cornerstone in mobilizing efforts to bring relevance at all levels.

CBOs should become advocates for local data collection efforts. First, they must be made aware of what is being done and by whom. Dr. Weinberg speculated that most CBOs are unaware of the process, labor, and cost associated with data collection for cancer registry programs. Most hospital-based efforts are increasingly being asked to control costs and physicians are being asked for more with little appreciation of the time and effort required to respond to information requests and requirements. As the shift from inpatient to outpatient care continues, the question arises of whether dependence on hospital registries is justified. Can we find support for community-based approaches rather than those that are traditionally institutionally focused?

While it is fair to assume that health care providers are in tune with the issues of data collection, it would be imprudent to fail to engage them in any discussion of retooling cancer data efforts. At a community level, this is especially true of the physicians. One example where physicians can play a key role relates to the major investment currently underway to capitalize on computerization of medical records. As this is accomplished, there will be an opportunity for standardizing reporting for cancer incidence, treatment, therapy, and outcome as well as a tremendous potential for harnessing population data. Local organizations should become advocates to be certain that proper linkages exist between community-based health care and hospital care, that data can be extracted easily for importing to cancer data efforts, and that other variables of interest (beyond financial) are incorporated as these are developed. Similarly, academic institutions should draw upon cancer data more routinely to focus educational efforts.

Consideration should be given to how national cancer information data efforts can become more meaningful at the local level and of greater benefit to CBOs, recognizing the importance, at the community level, of issues such as cancer clusters, sources of local variance and concerns, and the role that their data might play in evaluating local efforts to change outcomes. It is essential that the time from data collection to data application at community levels be reduced. The delay in translating what is learned from data collection efforts and the impact that this delay may have on the timeliness of community-based efforts often is not considered. In this age of rapid dissemination of knowledge and information, can we afford to have time bias? Can we afford to have the gaps go undetected for long periods of time?

Most CBOs would like to know they are making a difference. For this to become a reality they must be able to see that their efforts have produced or contributed to changes in outcomes and trends. This will require evidence and objectivity. By engaging local organizations in the decision-making process regarding data collection and application, CBOs will be proactive in sharing with their constituencies the progress that is sought and achieved. Likewise, failures will be better understood.

Discussion

Dr. McGinnis asked if data based on small numbers are meaningful. What are NCI's plans for addressing the issue of small data sets? Dr. Edwards asked the data collectors if they find small data sets better than no data. How small is too small? She has been looking for some examples of CBOs that have collected small sets of data that were analyzed and used by the community to improve services/outcomes. Ms. Shinagawa pointed out that CBOs usually do not have the funding or the expertise needed for data collection and analysis. They often do not know why data are needed and how to use the data. We need to do a better job of educating the CBOs about the usefulness of data. Dr. Daniel Miller asked the

questions: What is temporarily small? What will always be small? He indicated that many of the NPCR registries were quite incomplete in 1994, but by 1999, 16 NPCR registries were certified as 95 percent complete and a number of others were 90 percent complete. Obviously, progress is being made and as completeness of reporting is improved for the entire population, more data on racial/ethnic groups will be available. This will provide us with larger numbers and more resources with which to assess the accuracy and misclassification of data on racial/ethnic groups and to analyze what types of interventions work for these special populations. There also are groups for which the numbers will always be small even if reporting across the United States is 95 percent complete. That is where decisions must be made regarding aggregating the data. How many years of data should be aggregated? How will the data be used? Dr. Miller pointed out that although there may not be statistical significance, some data is better than no data.

Dr. Friedell mentioned that the timeliness of the data is important. In Kentucky, timely reporting of incomplete data serves several purposes—it is used by the state legislature to make decisions about funding and it is used by organizations that are serving the communities. He pointed out that it takes time to build community coalitions, because it requires a change in the nature of the academic, regional, and community settings of the community. Dr. Friedell also noted that most of these community health centers will require continual support from an entity that provides both technical support and funding.

Ms. Pat Golden asked for clarification regarding the purpose of this meeting and its expected outcome. Do we need to develop a pool of knowledge about cancer data? Dr. Friedell responded to Ms. Golden's request by indicating that this meeting has been focused on the work done by various agencies and where those agencies are planning to go in the future. The objectives of the meeting are to: identify needs that are not being addressed by the agencies, develop plans to address those needs, and recognize problems that cannot be solved now or in the near future. Dr. Wilson pointed out that there are a number of issues that must be revisited because they have not been solved in the past. Perhaps we need to develop new ways to approach these problems—new paradigms. Dr. Wingo suggested that participants should develop a list of issues and then separate the list into those that can be addressed now and those that cannot be addressed in the near term. Dr. Chen acknowledged that there are gaps, but warned that there are tradeoffs. For some purposes, less complete, lower quality data are adequate. This group needs to determine what is needed and then prioritize those needs. Dr. Hiatt indicated that the group should focus on what it can do to advance the uniform collection of race, ethnicity, and socioeconomic data as well as what can be done to improve the quality and accuracy of the numerators. In addition, we should focus on how to bring together the organizations (i.e., JCAHO, ACoS, and NCQA) that have influence over physicians, hospitals, and organizations that collect data to facilitate the development of a uniform set of measures. Ms. Douglas responded that the ACoS already is engaged in dialogue with the JCAHO. With regard to improving timeliness, one of the strategic directions for the ACoS and the NCDB is to initiate some real-time studies with the PCEs using NCDB data.

Dr. Liu noted that registry data are not perfect, but we cannot ignore the usefulness of population-based registry data. Misclassification is a problem, but we need to develop innovative, creative ways to supplement current registry data. She noted that participants should not let the unavailability of denominator data deter us when proportional incidence rates can be used.

National Center for Health Statistics

Dr. John Horm provided an overview of the NCHS Research Data Center (RDC), which has been in operation for about 1 year. He distributed several handouts to participants including a list of grants funded by NCHS' Minority Health and Statistics Grants Program. More information on these grants is provided on the Web at **http://www.cdc.gov/nchs/otheract/grants/projects**. Most of these grants have been methodologic in nature and few have involved data collection. Dr. Horm also distributed a brochure on the RDC and procedures for using the RDC.

Since 1988, the NHIS has included questions on health insurance, access to care, family resources, physician visits, cancers, smoking, and alcohol use. Therefore, this data set has been a good source of information on behavior, knowledge, risks, and attitudes related to cancer. In 1993, the NHIS expanded its racial categories to include 15 races and 7 categories of Hispanic ethnicity. In 1995, the NHIS combined some of those categories because of confidentiality issues. The NHIS still collects the information for these racial categories, but the information is no longer released to the public. The RDC was developed to allow researchers and data users to access internal data files from its numerous surveys that have not been available to the research community until now. These files have not been made available to researchers outside the NCHS to protect the identify of the respondents, to maintain control over the data, and to prohibit the use of these data for any purpose other than the reason identified to the respondents. These data can be accessed only by collaborating with NCHS staff or working with the RDC. The NCHS does not release the files, but it does collaborate with researchers.

The RDC, located at the NCHS headquarters in Hyattsville, MD, is open from 9 a.m. to 5 p.m. weekdays. An advisory and proposal review committee receives, reviews, and approves researcher proposals. Researchers may not remove any materials from the RDC until they have received a disclosure avoidance examination by RDC staff and may not remove any restricted data files nor linked data files from the RDC. The internal files accessible through the RDC contain lower levels of geography such as state, county, Census tract, block-group, or blocks depending on the survey. The data that are available through the RDC include the NHIS (1987-1996), National Survey of Family Growth (1990, 1993, and 1995), Third National Health and Nutrition Survey (1988-1994), and other NCHS survey data including vital statistics and the Longitudinal Study on Aging. The RDC also allows access to linked data sets that would not otherwise be available to researchers.

The NCHS restricted data can be accessed through the RDC by:

- ✧ Guest Researcher—Researcher submits proposal to the RDC and upon approval, would be able to conduct research onsite at the RDC. NCHS staff would construct the necessary data files and ensure that no restricted data leaves the facility. The researcher is allowed to take the results of the research offsite after disclosure review by NCHS RDC staff.
- ✧ Remote Access—Users can electronically submit analytical computer program using SAS as the programming language. Programs are e-mailed to the RDC, NCHS staff prepare the requested data files, which are submitted for disclosure limitation review. The output files are then e-mailed to the requester.
- ✧ Staff Programming—Users can subcontract with the NCHS RDC to perform the programming tasks necessary for the research project. Analytical programs are written by the NCHS staff with input from the researcher. All tabulations leaving the RDC would be subjected to a disclosure review to ensure that confidentiality of respondents is protected.

Each of these access methods has an associated cost that includes equipment and space rental, staff overhead, and setup. The charge for remote access is $500/month for small files and $1,000/month for large files. Staff programming charges are $500/day. The RDC will obtain any software required by researchers and it is equipped with high-powered state-of-the-art computers that can handle all NCHS data. The NCHS RDC accepts proposals at any time and each proposal needs to explain what files the researcher needs to access, why access to these data are needed, and what the researcher plans to do with the data. Most proposals are reviewed within 2-3 weeks of submission.

Ms. Pat Golden indicated that the NCHS was created to develop data at the national level. The NCHS publishes reports on the data it collects; however, more data are collected than are included in the publications to protect the confidentiality of the respondents. In addition, most of the public use data is

available on the Internet. She stressed the importance of understanding the methodological limitations of data. The NCHS has collected a large amount of data, but each data source is different—the data are collected in a different manner and from different groups. Therefore, comparability is always an issue. In addition, there is a lack of data for subpopulations. Some of these data may be better obtained through small, targeted surveys. Ms. Golden cautioned against adjusting a large, national survey to obtain data for a small group. The NCHS data serve many purposes and meet the needs of many different organizations. The data from all NCHS surveys can be used effectively to generate hypotheses, but researchers may not be able to use these data to answer specific questions about behavior or an intervention. Most NCHS surveys are not designed to provide information at that level; however, NCHS surveys do collect data on emergency room usage and short hospital stays.

Ms. Golden pointed out that terminology is a challenge. Although the OMB directive definitions are limited, they are the only definitions currently available. She noted that the world continues to change and the dynamics that influence health care (e.g., social changes) often are outside the health care system. She identified the following limitations that must be considered by the participants: (1) timeliness and periodicity (it is not possible to get all data systems into sync), (2) core of commonality (common core of needs regarding data collection and interpretation), and (3) points of divergence (differences). She pointed out that no one data system will meet all the data needs. Many researchers have been trying to retrofit data systems to meet needs that they were never designed to meet. Ms. Golden closed her presentation by suggesting that the participants focus on what is doable.

Discussion

Dr. Edwards asked how many of the grantees in the Minority Health Statistics Grants Program use the RDC. Dr. Horm responded that the grantees are not using the RDC because they are working with their own data. Dr. Wilson asked if there is a list of what is in each restricted file so that researchers can determine what they need to access. Dr. Horm replied that most of the data in the restricted files can be released except for the geographic information. He indicated that the restricted files include a number of elements, such as place of birth, that are not available in the public files. However, researchers have to do their own work to find out what is available through the RDC. Ms. Golden suggested that researchers could determine what is available by reviewing the questionnaire on the Web site.

Dr. Lovell Jones suggested that surveys include a question asking respondents to identify the population with which they most identify. Ms. Golden noted that some data sets cannot accommodate the collection of that information. She add that there is a significant amount of cognitive research conducted by research centers and the NCHS. The challenge is to obtain data that are statistically reliable and valid. Dr. Horm noted that the 1997 questionnaire of the NHIS includes a question similar to the one suggested by Dr. Jones.

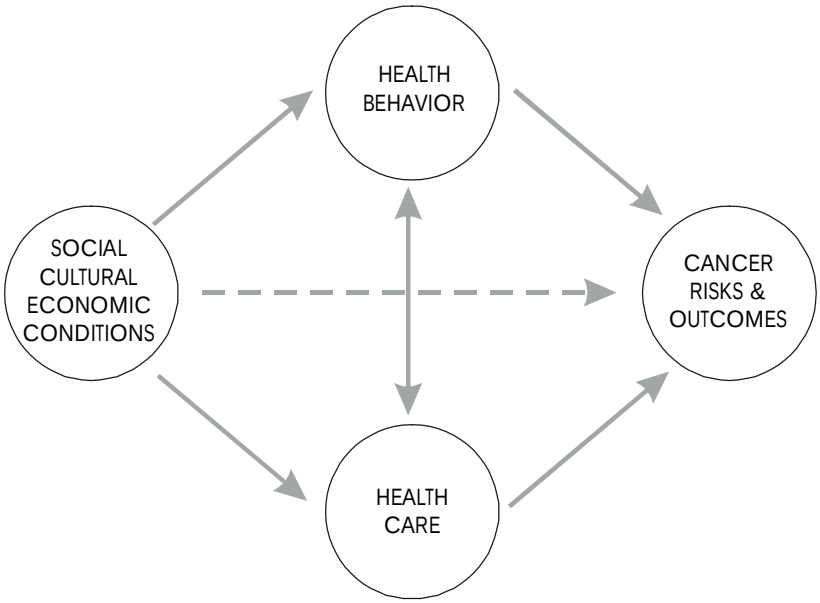
NCI Surveillance Activities Related to Sociodemographic Disparities in Cancer

Mr. Barry Miller presented the diagram in Figure 34 that identifies the relationships between sociodemographic factors (e.g., knowledge, money, power, prestige, and social connections) and cancer risk and outcomes. Because there remains a lot to learn about the influence of social, cultural, and economic conditions on cancer risk and outcomes, the line from that circle in the diagram is dotted.

The goals of NCI's surveillance program with regard to sociodemographic disparities in cancer are to: (1) monitor the differential cancer burden among subgroups of the U.S. population, and (2) support research on the identification of social, psychological, behavioral, and biological mechanisms through which cancer disparities are generated. Some of the ongoing activities undertaken by the NCI to address these goals include:

Figure 34. Relationship Between Sociodemographic Factors and Cancer Risk and Outcomes

36



Source: National Cancer Institute, Division of Cancer Control and Population Sciences, 1999.

- ✧ Supporting the development or refinement of concepts, definitions, indices, and analytical frameworks. Several studies have shown that social characteristics of the environment or the community in which one lives confer health and disease risks independently of individual characteristics. How best to use the sociodemographic information at the individual and community levels requires further investigation. There are a number of studies, some led by SEER registry Principal Investigators, to investigate this issue. Efforts are needed to identify the different roles that socioeconomic factors might play in influencing cancer risk and outcomes within specific ethnic groups.
- ✧ Assessing existing data resources and identifying additional data needs. The SEER program routinely collects detailed information on race, ethnicity, marital status, place of residence, place of birth, tumor characteristics, and first course of therapy for cancer patients. However, there is a need for new data collection efforts, such as SEER expansion, to enable the NCI to capture more information on ethnic populations. In addition, surveillance capabilities are being enhanced through data linkages that take advantage of geographic information systems (GIS).
- ✧ Describing and monitoring sociodemographic patterns in cancer outcomes and cancer control. Studies are focused on monitoring and reporting the extent of social disparities in the cancer burden across time and geographic area. Research is being conducted to develop analytic methods to study potential intervening mechanisms or pathways through which social disparities in cancer occur.

Mr. Miller provided the following examples of projects involving database expansion and linkages:

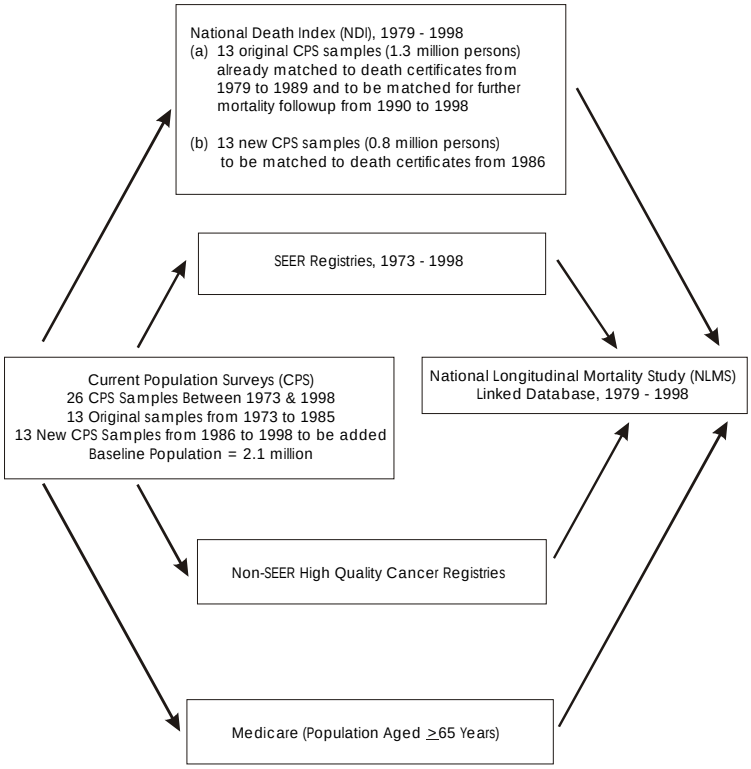
- ✧ Geographic-Level Data
  - ✦ Decennial Census: SES, ethnicity, nativity, acculturation, occupation, urban/rural.
  - ✦ Area Resource File: SES, hospital discharge, medical service.
  - ✦ BRFSS: risk and screening behaviors.

- ✧ Individual-Level Data
  - ✦ SEER expansion: rural poor, rural African American, non-Mexican Hispanic, American Indian.

- ✦ Explore possibilities of obtaining new data elements in SEER.
- ✦ NHIS with cancer supplement: screening behavior and knowledge about cancer.
- ✦ SEER special surveillance research studies: GIS, SES.
- ✦ SEER Prostate Cancer Practice Patterns and Quality of Life Study: treatment, quality of life, education, income, marital status.
- ✦ SEER-NLMS-NDI linkage.
- ✦ California Health Interview Survey and Social Security Administration linkage: income and occupation.
- ✦ Behavioral Research Program, NCI (Dr. Robert T. Croyle).
  - – Applied Sociocultural Research Branch
  - – Basic Biobehavioral Research Branch

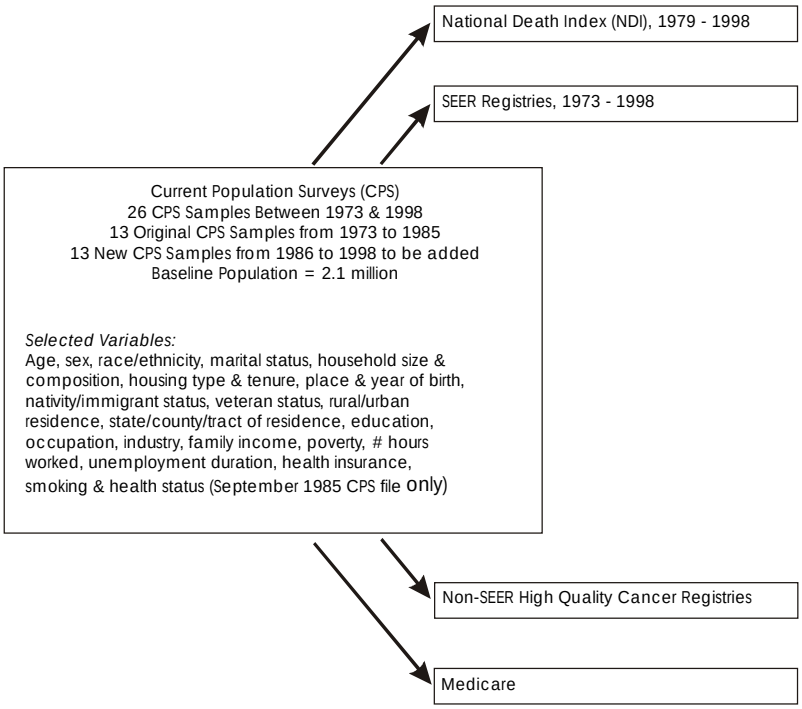
Figures 35 and 36 provide more detail on the SEER-NLMS-NDI Linkage study. These projects link the individual patient records in SEER with a variety of sociodemographic variables collected on individuals participating in the 26 surveys (CPS) conducted annually from 1973 to 1998 by the Census Bureau.

Figure 35. The NLMS/SEER/NDI Linkage Study, 1979-1998



Source: National Cancer Institute, Division of Cancer Control and Population Sciences, 1999.

Figure 36. Current Population Survey Component of the NLMS/SEER/NDI Linkage Study, 1979-1998



Source: National Cancer Institute, Division of Cancer Control and Population Sciences, 1999.

This effort is part of the NLMS. The CPS will provide information on age, sex, race and ethnicity (self reported), martial status, household size and composition, housing type and tenure, place and year of birth, nativity/immigration status, veteran status, rural/urban residence, state/county/tract of residence, education, occupation, industry, family income, poverty, number of hours worked, unemployment duration, health insurance, and smoking and health status. Current efforts are focused on piloting the linkage with data from one of the SEER registries. Once this is completed, data from other SEER registries will be linked. NCI also is working to resolve issues associated with protecting confidentiality.

Discussion

Dr. Weinberg asked if the NCHS would consider bringing in individuals from the community to educate them regarding what data are available. Dr. Horm responded that the RDC has interns from ethnic populations who are working with NCHS staff and learning about these data sets. Dr. Weinberg asked if the NCHS would make a presentation at the next ICC national conference and use this forum to educate the community. Dr. Edwards indicated that the NCI can partner with schools of public health to educate community representatives. However, she noted that the federal government has limited resources available to work directly with communities. Dr. Weinberg commented that this would be a valuable role for the ICC. Ms. Golden added that the NCHS sponsors a data users conference.

Dr. Wilson asked Mr. Miller what he meant by the term “power” in his presentation. Mr. Miller replied that it refers to social connections and cultural influences. Dr. Wilson praised NCI’s efforts to link databases that were not designed for cancer research; he noted that the pharmaceutical industry is linking its files with those of the health insurance industry to examine cost effectiveness. Dr. Friedell asked Mr. Miller how the NCI defined medically underserved. Dr. Edwards responded that it has not been defined,

but the NCI is working to develop a definition. She noted that low income is not synonymous with underserved. Ms. Golden asked what information should be collected to measure the underserved? One recommendation is to collect SES data such as income and education; however, in 5 years these may have changed for an individual. What is the most appropriate vehicle to collect this information? Dr. Friedell suggested that we start by collecting information on education. Dr. Hampton pointed out that most people do not enhance their economic status once they are diagnosed with cancer. Ms. Golden responded that she was referring to the broader population, not just those who have cancer. Dr. Chen commented that SES factors change with time. She also pointed out that different components of SES affect different points in the cancer continuum.

PRIORITIZATION OF THE ICC DATA WORKING GROUP RECOMMENDATIONS

During the September meeting, the ICC Data Working Group developed 13 recommendations to address data gaps for ethnic minorities and the medically underserved (see Figure 37). The participants of the November meeting were asked to prioritize these recommendations in two phases. In the first phase, participants were asked to identify the most important recommendations. Participants were given four orange dots and asked to place one or more of these dots next to those recommendations they considered to be the most important. Through this process, participants rated Recommendations 1, 2, 7, 11, and 12 as the most important of the 13. In the second phase, participants were asked to prioritize these five important recommendations by doability. The participants were given four blue dots and asked to place one or more of the blue dots by the important recommendations that were most doable. This process identified Recommendations 1 and 7 as the most doable of the five most important recommendations. The rankings assigned by the participants are presented in Figure 37.

BREAKOUT SESSION I: ACTION PLANS TO IMPROVE DATA COLLECTED FOR USE IN MEASURING CANCER IN POPULATION SUBGROUPS

Following the prioritization process, the participants were divided into two breakout groups to discuss actions that could be taken to address the most important recommendations resulting from the September ICC Data Working Group meeting. Dr. Edwards instructed the participants to focus their efforts on Recommendations 1, 2, 7, 11, and 12. Because Recommendations 2 and 11 were ranked the lowest in terms of doability, she suggested that the breakout groups focus first on Recommendations 1, 7, and 12, and then on Recommendations 2 and 11. Participants were asked to identify specific actions that could be implemented to address these recommendations.

The results from this breakout session are presented by recommendation in this section of the report. The organizations responsible for implementing these actions were not always identified by the participants. These results do not represent the consensus of the meeting participants, rather they present a summary of the discussions that took place during that session. They form a good foundation on which to base future discussions of this subject.

Actions to Collect SES Information for Cancer Patients (Recommendation 1)

1. More expert input is needed concerning what currently is being collected by hospitals and why, and what can be interpreted/extrapolated from that information. A short-term goal could be to identify hospitals that are already collecting SES data and to discuss this issue with them.
2. Conduct study to determine the feasibility of collecting SES data. Potential data items include education (highest grade completed), income (range), insurance (already on admitting sheet), occupation, and residence (primary payer mailing address already included).

Figure 37. Rankings Assigned to Data Working Group Recommendations

Recommendation From September Report	Priority (number of orange dots assigned)	Doability (number of blue dots assigned)
1. SES information should be collected for cancer patients	14*	22
2. Meaningful numerator/denominator statistics must be developed	15*	16
3. Methods must be developed for overcoming data collection communication barriers	2	
4. Nativity data should be collected	6	
5. Cancer data should be collected in a disaggregated manner	2	
6. The community should be involved in data collection	1	
7. Training of epidemiologists and related professionals from minority and medically underserved populations should be conducted	9*	22
8. A coordinated approach should be developed for collecting and analyzing risk factor information	3	
9. SEER should be expanded to include collection of cancer data for all ethnic minorities and the medically underserved	7	
10. CDC should improve the data collection efforts of state tumor registries	7	
11. Federal agencies collecting cancer data should be mandated to collaborate to provide comprehensive national data	10*	11
12. Persons from the community being studied should provide advice on culturally appropriate cancer control materials and instruments	12*	20
13. Special populations with low incidence of cancer should be studied for "protective factors"	2	

* These recommendations were identified as most important so they were ranked according to doability.

3. Work with JCAHO to establish standards for collecting SES data (e.g., information on the face sheet of the medical record), and to include such data in accreditation.
4. Conduct training for physicians and others who will be collecting SES data.
5. NAACCR Uniform Data Standards Committee should consider including SES as a required standard data element for registries, pending results of feasibility studies.
6. Standard SES indicators that should be considered for collection include: education (not currently collected) and insurance (this information is currently collected on the face sheet).
7. Record linkages with Census information at the Census tract or smaller level based on collection of the 9-digit zip code or residential street address and 5-digit zip code.
8. ACS/CoC needs to define residential address variable on medical record face sheet.
9. NAACCR Uniform Data Standards Committee should consider residential street address (not just mailing address) as a recommended standard variable.

Actions to Train Epidemiologists and Related Professionals From Minority and Medically Underserved Populations (Recommendation 7)

1. There are a number of training programs (e.g., University of New Mexico, Northern California Cancer Center, University of Kentucky) for epidemiologists. These programs should be linked together and then linked to communities.
2. The ICC should identify training opportunities and summarize them in an easy and accessible format on the ICC Web Site (i.e., a clearinghouse).
3. Broaden epidemiological training to include genetics, social, and behavioral epidemiology. Also consider distance learning and other innovative approaches for training.
4. Incorporate epidemiological training at the elementary school level.
5. There is a need to designate an agency to focus on career development.
6. ICC should identify available opportunities and then promote those opportunities and coordinate training.
7. Federal agencies should document their actions to meet cancer training needs of minorities and underserved.

Actions to Promote Community Involvement for Culturally Appropriate Cancer Control Materials and Instruments (Recommendation 12)

1. Create a clearinghouse of culturally appropriate materials so communities can determine if appropriate materials exist for their use (e.g., mammography brochure).
2. Require researchers to identify in their proposals how they plan to obtain community input. Develop a checklist to include in review criteria to ensure community involvement.
3. The NCI is committed to ensuring that there is more community involvement (R25 mechanism supports resource center based in major institution, e.g., Native C.I.R.C.L.E.). The culture of the Cancer Centers needs to change to include diversity and community involvement (go beyond cancer care).
4. The Cancer Information Service (CIS) needs to expand its materials to include more ethnic populations. A partnership is being formed between the CIS and the ICC, and CIS outreach staff will attend the ICC meeting in February 2000 so that the two organizations can exchange information.
5. The ICC should identify opportunities to expand community involvement in cancer control activities. The ICC also should identify situations in which such involvement is not taking place.
6. The ICC should identify opportunities to provide input for special data collection efforts (e.g., surveys, studies).

Actions to Develop Meaningful Numerator/Denominator Statistics (Recommendation 2)

1. Registries should collect numerator data that match the denominator data collected by Census 2000. This could be accomplished by encouraging medical records administrative systems to be compatible with the Census form.
2. Registrars should receive training on abstracting race/ethnic data to improve consistency with denominator data.

3. A pilot or feasibility study could be conducted to investigate comparability of numerator and denominator data as well as the cost effectiveness of collecting data such as place of birth.
4. ACS and CoC should work with JCAHO to include birthplace and self-designated ethnicity (following OMB Directive 15 definitions) on the face sheet of the medical record.
5. NAACCR Uniform Data Standards Committee will include birthplace and self-designated ethnicity as required standard data elements for registries.
6. The Census Bureau should produce intercensal estimates for ethnic-specific groups (e.g., Chinese, Mexican Hispanic, specific American Indian tribes). This effort should begin at the national level and then be taken to the state and county level.

Actions to Promote Collaboration Among Federal Agencies Collecting Cancer Data to Provide Comprehensive National Data (Recommendation 11)

1. NAACCR will produce a public aggregated report (i.e., report card) on the quality of the registries (an annual report would be preferable). Also, the NPCR Web Site should continue to identify the states that meet the program standards and receive certification.
2. CDC/NPCR and NCI/SEER should develop accountability indicators based on objective outcome measures for cancer registries.

BREAKOUT SESSION II: HOW CAN WE WORK TOGETHER TO LEVERAGE OUR ACTIONS?

Dr. Lovell Jones suggested that a subcommittee be formed to work with JCAHO to bring about the changes needed to collect SES information. Dr. Friedell thought it might be helpful to ask those who have been working with SES data what indicators are most useful in prognosis and selection of therapeutic measures. Dr. Weinberg suggested that the report from this meeting be disseminated to ICC members and partners. Dr. Edwards mentioned that Dr. Sherry Mills (NCI) convened a small group to look at obtaining SES information from minority populations. This group also is investigating how to use SES information, and the NCI may sponsor a conference on SES and cultural measures. Dr. Jones asked if that conference will involve other institutes/agencies and Dr. Edwards replied that no decision has been made yet regarding who will be invited.

Dr. Key expressed his concern about how hospitals are going to react to additional data requirements—they will be viewed as unfunded mandates. Dr. Friedell indicated that it would be useful to build partnerships with clinical organizations (e.g., ASCO) to facilitate collection of SES data. We need to determine if SES affects quality of care. Dr. Weinberg suggested that we need to stress the importance of data collection to hospitals and physicians. Can we develop leading indicators for cancer? Dr. Edwards responded that a number of groups have debated the issue of leading indicators, trying to reach consensus on what should be collected and how often it should be collected. One participant asked if SES data should be collected on every individual in a registry. What will we do with the data when we get it? Ms. Shinagawa suggested that pilots be conducted to answer that question. Dr. Edwards indicated that SEER could pilot the collection of information on insurance and education. The pilot effort will determine if the information is available and what it will cost to include it in the registry. Dr. Key mentioned that there was an attempt to collect education information under the Breast Cancer Screening Consortium (BCSC).

Ms. Shinagawa identified the following suggestions made during the discussion:

- ✧ Determine what information currently is collected and which organizations collect it.
- ✧ Partner with the National Dialogue on Cancer and provide input into the new National Cancer Act.
- ✧ Examine what information is being collected for other diseases (e.g., heart disease).
- ✧ Partner with Schools of Public Health and other organizations (e.g., survivor groups, nursing organizations, Association of Conference of Cancer Centers, ASCO).
- ✧ Examine the SES data collected under the BCSC.
- ✧ Work with JCAHO to change data collection requirements when collection of SES data is proven feasible.
- ✧ Identify leading indicators for cancer. Could this be accomplished in conjunction with the national report card?

Ms. Shinagawa noted that survivor organizations, which are good at getting information out to their constituents, are an untapped source. Dr. Key stressed the importance of common electronic record systems. If hospitals manage records in a uniform manner, those records could easily feed into the cancer registry. Dr. Edwards mentioned the Cancer Informatics Infrastructure, which is concerned with building compatible information systems for clinical trials. Ms. Shinagawa asked if an Executive Committee should be created to follow up on the suggestions made with regard to Recommendation 1. Dr. Friedell agreed to form an Executive Committee to approach JCAHO.

Dr. Edwards mentioned that the NCI is looking at SES and how it relates to outcomes and interventions. The progress and results of these efforts will be reported to NAACCR. She agreed to keep the ICC apprised of these efforts. Ms. Shinagawa suggested that the ICC interface with the ongoing activities of the NCI and the CDC. She asked if Dr. Key could provide information on how many women in the BCSC have reported their level of education. Dr. Key agreed to provide that information. Ms. Golden suggested that the ICC conduct a search to determine what SES data are available.

With regard to Recommendation 7 on training epidemiologists, it was suggested that the ICC prepare an inventory of training resources and opportunities. Dr. Edwards indicated that Dr. Sanya Springfield (NCI) would be a source of information regarding minority supplements for grants. Dr. Friedell suggested contacting the Office of Research on Minority Health and the Office of Special Populations Research. Ms. Golden commented that CDC's Office of Minority Health and Epidemiology Program Office should be contacted. Dr. Edwards agreed to provide Ms. Shinagawa a list of NCI's R25 training grants for inclusion on ICC's Web site. Other organizations that should be contacted include the Society for Epidemiologic Research, the American Society for Preventive Oncology, the ACS, and the Department of Defense.

For Recommendation 12 on community involvement, most participants agreed that this was occurring. One participant suggested that community involvement should be added to review criteria. Dr. Edwards mentioned that there may be no review criteria published for an R01.

Referring to the collection of meaningful numerators/denominators (Recommendation 2), Dr. Key asked how present data will be incorporated with the new data—the numerator will be the same and there will be a choice of denominators. How will we explain why the rates are different? He suggested that we need to develop consistent language to accompany these data. Ms. Golden has been looking into this issue and she agreed to provide Dr. Key with several contacts with whom he can pursue this question. Dr. de la Puenta added that the Census Bureau also is examining this issue, but it has decided to wait

until the Census 2000 closes before issuing any guidance. He agreed to keep the ICC informed of any developments. Dr. West mentioned that the NCI will be coordinating a workshop at the NAACCR meeting in April 2000, which will focus on the changes in the Census, how this will affect reporting of federal data, and the implications for cancer reporting. Dr. Chen added that ICD-10 also will be discussed at that meeting. Ms. Golden agreed to forward any relevant information regarding this issue to the ICC. Dr. Friedell closed the meeting by thanking the participants for their input.

List of Participants

Beverly Campbell
President
The Scientific Consulting Group, Inc.
656 Quince Orchard Road, Suite 210
Gaithersburg, MD 20878
Tel: 301-670-4990
Fax: 301-670-3815
E-mail: bcampbel@scgcorp.com

Vivien W. Chen, Ph.D.
Director
Louisiana Tumor Registry
Louisiana State University Medical Center
1600 Canal Street, Room 900A
New Orleans, LA 70112
Tel: 504-568-6047
Fax: 504-568-2493
E-mail: vchen@lsumc.edu

Manuel de la Puente, Ph.D.
Assistant Division Chief for Survey Methodology
Bureau of the Census
Statistical Research Division
4401 Suitland Road, Room 3207, Building 4
Suitland, MD 20749
Tel: 301-457-4997
Fax: 301-457-2299

Lynda Douglas, C.T.R.
Quality Assurance Administrator
Cancer Department
American College of Surgeons
633 North Saint Claire
Chicago, IL 60611
Tel: 312-202-5435
Fax: 312-202-5009
E-mail: ldouglas@facs.org

Brenda K. Edwards, Ph.D.
Associate Director
Surveillance Research Program, DCCPS
National Cancer Institute
6130 Executive Boulevard
Executive Plaza North, Room 343
Bethesda, MD 20892-7350
(For Federal Express use: Rockville, MD 20852)
Tel: 301-496-8506
Fax: 301-496-9949
E-mail: be2w@nih.gov

Gilbert H. Friedell, M.D.
Director Emeritus
Lucille Parker Markey Cancer Center
University of Kentucky
206 Davis-Mills Building, MRISC 0098
800 Rose Street
Lexington, KY 40536-0098
Tel: 606-323-6541 (main office)
606-257-5323 (voice mail)
Fax: 606-323-6990
E-mail: gil@delos.kcr.uky.edu

Patricia Golden, M.P.H.
Liaison for Minority Health Data
National Center for Health Statistics
Centers for Disease Control and Prevention
Room 750
6525 Belcrest Road
Hyattsville, MD 20782
Tel: 301-436-7032 Ext. 149
Fax: 301-436-8459
E-mail: pmg2@cdc.gov

Marc T. Goodman, Ph.D.
Associate Director, Epidemiology Program
Cancer Research Center of Hawaii
Hawaii Tumor Registry
1236 Lauhala Street, Room 402
Honolulu, HI 96813
Tel: 808-586-2987
Fax: 808-586-2982
E-mail: marc@crch.hawaii.edu

James W. Hampton, M.D.
Clinical Professor of Medicine
University of Oklahoma College of Medicine
Medical Director
Troy and Dollie Smith Cancer Center
Integris Baptist Medical Center
11100 Hefner Pointe Drive
Oklahoma City, OK 73120
Tel: 405-749-0415
Fax: 405-749-6843
E-mail: james.hampton@usoncology.com

Robert A. Hiatt, M.D., Ph.D.
Deputy Director
Division of Cancer Control and Population Sciences
National Cancer Institute
6130 Executive Boulevard
Executive Plaza North, Room 243
Bethesda, MD 20892-7339
(For Federal Express use: Rockville, MD 20852)
Tel: 301-435-7206
Fax: 301-496-8675
E-mail: rh170b@nih.gov

John Horm, Ph.D.
Research Data Center
National Center for Health Statistics
Centers for Disease Control and Prevention
Room 700
6525 Belcrest Road
Hyattsville, MD 20782
Tel: 301-436-3915
Fax: 301-436-3745
E-mail: jdh3@cdc.gov

Holly L. Howe, Ph.D.
Executive Director
North American Association of Central Cancer Registries
2121 W. White Oaks Drive, Suite C
Springfield, IL 62704
Tel: 217-698-0800 Ext. 2
Fax: 217-698-0188
E-mail: hhowe@naaccr.org

Pamela Jackson
Director, Outreach Programs
Intercultural Cancer Council
Baylor College of Medicine
6560 Fannin, Suite 924 Scurlock Tower
Houston, TX 77030
Tel: 713-798-4617
E-mail: pjackson@bcm.tmc.edu

Lovell A. Jones, Ph.D.
Professor and Director of Gynecologic Oncology
Department of Gynecologic Oncology
University of Texas
M.D. Anderson Cancer Center
1515 Holcombe Boulevard, Box 304
Houston, TX 77030
Tel: 713-792-3316
Fax: 713-792-3575
E-mail: exgyn@aol.com

Marjorie Kagawa-Singer, Ph.D.
Assistant Professor
UCLA School of Public Health
Community Health Sciences
UCLA Asian American Studies Center
Box 951772
Los Angeles, CA 90095-1772
Tel: 310-825-9481
Fax: 310-794-1805
E-mail: mkagawa@ucla.edu

Charles R. Key, M.D., Ph.D.
Medical Director
New Mexico Tumor Registry
University of New Mexico Cancer Center
2325 Camino de Salud, NE
Albuquerque, NM 87131
Tel: 505-272-8703
Fax: 505-272-8572
E-mail: ckey@nmtr.unm.edu

Lihua Liu, Ph.D.
Research Associate
Cancer Surveillance Program
University of Southern California
1540 Alcazar Street, Room CHP-204
Los Angeles, CA 90033
Tel: 323-442-1636
Fax: 323-442-2301
E-mail: lihualiu@hsc.usc.edu

Lamar McGinnis, M.D.
Medical Consultant
American Cancer Society
1599 Clifton Road, NE
Atlanta, GA 30329
Tel: 404-329-7625
Fax: 404-329-7530
E-mail: lmcginni@cancer.org

Barry Miller, M.S.P.H.
Commander, Public Health Service
Epidemiologist
Cancer Statistics Branch, SRP, DCCPS
National Cancer Institute
6130 Executive Boulevard
Executive Plaza North, Room 343D
Bethesda, MD 20892-7352
(For Federal Express use: Rockville, MD 20852)
Tel: 301-402-4248
Fax: 301-496-9949
E-mail: bm33q@nih.gov

Daniel Miller, M.D., M.P.H.
Chief, Cancer Surveillance Branch
Centers for Disease Control and Prevention
4770 Buford Highway, NE
Mailstop K-53
Atlanta, GA 30341
Tel: 770-488-4783
Fax: 770-488-4759
E-mail: dsm2@cdc.gov

Roberta Paisano, M.H.S.A.
Epidemiology Analyst
Indian Health Service
5300 Homestead Road, NE
Albuquerque, NM 87110
Tel: 505-248-4226
Fax: 505-248-4393
E-mail: roberta.paisano@mail.ihs.gov

Elsie Pamuk, Ph.D.
Acting Director, Division of Epidemiology
Office of Analysis, Epidemiology and Health Promotion
National Center for Health Statistics
Centers for Disease Control and Prevention
6525 Belcrest Road, Room 750
Hyattsville, MD 20782
Tel: 301-436-5979 Ext. 144
Fax: 301-436-8459
E-mail: exp4@cdc.gov

Susan M. Shinigawa
Co-chair, Intercultural Cancer Council (through February 2000)
1720 Dryden, PMB-C
Houston, TX 77030
Tel: 713-798-4617
Fax: 713-798-3990
E-mail: smshinagawa@juno.com

Lucina Suarez, Ph.D.
Epidemiologist
Associateship for Disease Control and Prevention
Texas Department of Health
1100 West 49th Street
Room G-401
Austin, TX 78756
Tel: 512-458-7729
Fax: 512-458-7229
E-mail: lucina.suarez@tdh.state.tx.us

Lillian Tom-Orme, Ph.D.
American Indian and Alaska Native Populations
Division of Public Health Sciences
Department of Oncological Sciences
University of Utah
546 Chipeta Way, Suite 1100
Salt Lake City, UT 84108
Tel: 801-585-5246
Fax: 801-585-5357
E-mail: lillian.tom-orme@hci.utah.edu

Edward J. Trapido, Sc.D.
Professor and Associate Director
Sylvester Comprehensive Cancer Center
Vice Chair, Epidemiology and Public Health
University of Miami School of Medicine
1550 NW 10th Avenue, Room 308
Miami, FL 33136
Tel: 305-243-3356
Fax: 305-243-2997
E-mail: ejt@miami.edu

Thomas C. Tucker, M.P.H.
Associate Director for Cancer Control
Kentucky Cancer Registry
Lucille Parker Markey Cancer Center
University of Kentucky
Room 312, MRISC Building
800 Rose Street
Lexington, KY 40536-0098
Tel: 606-257-4582
Fax: 606-257-5267
E-mail: tct@delos.kcr.uky.edu

Armin D. Weinberg, Ph.D.
Director
Chronic Disease Prevention and Control
Research Center
1 Baylor Plaza, Suite 924
Houston, TX 77030
Tel: 713-798-4614
Fax: 713-798-3990
E-mail: arminw@bcm.tmc.edu

Dee West, Ph.D.
Executive Director
Northern California Cancer Center
32960 Alvarado-Niles Road, Suite 600
Union City, CA 94587-3106
Tel: 510-429-2502
Fax: 510-429-2550
E-mail: dwest@nccc.org

Jerome Wilson, Ph.D.
Director, Global Outcomes Research
Pfizer Pharmaceutical Group
Pfizer, Inc.
235 East 42nd Street
New York, NY 10017-5755
Tel: 212-733-6413
Fax: 212-808-6472
E-mail: jerome.wilson@pfizer.com

Phyllis Wingo, Ph.D.
Director of Surveillance Research
American Cancer Society
1599 Clifton Road, NE
Atlanta, GA 30329
Tel: 404-327-6552
Fax: 404-327-6450
E-mail: pwingo@cancer.org