SAFER • HEALTHIER • PEOPLE™

Monitoring the Nation's Health

Linkage of 1986–2009 National Health Interview Survey With 1981–2010 Florida Cancer Data System

Vital and Health Statistics
Series 2, Number 167
September 2014

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Copyright information

All material appearing in this report is in the public domain and may be reproduced or copied without permission; citation as to source, however, is appreciated.

Suggested citation

Miller EA, Miller DM, Judson DH, et al. Linkage of 1986–2009 National Health Interview Survey with 1981–2010 Florida Cancer Data System. National Center for Health Statistics. Vital Health Stat 2(167). 2014.

Library of Congress Cataloging-in-Publication Data

Linkage of 1986–2009 National Health Interview Survey with 1981–2010 Florida Cancer Data System.
p. ; cm.— (Vital and health statistics. Series 2, Data evaluation and methods research ; number 167) (DHHS publication ; (PHS) 2014–1367)
"September 2014."
Includes bibliographical references and index.
ISBN 978–0-8406–0667-9 (alk. paper)—ISBN 0–8406-0667–2 (alk. paper)
I. National Center for Health Statistics (U.S.), issuing body. II. Series: Vital and health statistics. Series 2, Data evaluation and methods research ; no. 167. III. Series: DHHS publication ; no. 2014–1367. 0276–4733
[DNLM: 1. National Health Interview Survey (U.S.) 2. Florida Cancer Data System 3. Public Health—statistics & numerical data—United States. 4. Health Surveys—United States. W2 A N148vb no.167 2014]
RA407.3
614.4'273—dc23 2014031050

For sale by the U.S. Government Printing Office
Superintendent of Documents
Mail Stop: SSOP
Washington, DC 20402–9328
Printed on acid-free paper.

Vital and Health Statistics

Series 2, Number 167

Linkage of 1986–2009 National Health Interview Survey With 1981–2010 Florida Cancer Data System

Data Evaluation and Methods Research

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Hyattsville, Maryland
September 2014
DHHS Publication No. 2014–1367

National Center for Health Statistics

Charles J. Rothwell, M.S., M.B.A., *Director*

Jennifer H. Madans, Ph.D., *Associate Director for Science*

Office of Analysis and Epidemiology

Irma E. Arispe, Ph.D., *Director*

Contents

Abstract 1

Introduction 1

Methods 2
- Data Files 2
- Linkage Approval and Methods 2
- Evaluation of Linked File 4

Results 5
- Linked NHIS–FCDS Data Files 5

Discussion 7

Conclusion 9

References 9

Figures

1. Number of female participants available for breast cancer risk analysis using survey tobacco questions: 1986–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System 6
2. Percentage of diagnosed late-stage cancers, by education level and sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System after interview 7
3. Percentage of diagnosed late-stage cancers among those above and below federal poverty level, by sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System after interview 7
4. Percentage of diagnosed late-stage cancers, by marital status and sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System after interview 8
5. Percentage of self-rated fair and poor health status, by cancer diagnosis and sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System before interview (cancer survivor) and not linked (not diagnosed with cancer) 8
6. Percentage having any activity limitation, by sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System before interview (cancer survivor) and not linked (not diagnosed with cancer) 9
7. Percentage of current smokers, by sex: 1997–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System before interview (cancer survivor) and not linked (not diagnosed with cancer) 9

Detailed Tables

1. Availability of selected questionnaire topics, by survey year: National Health Interview Survey, 1986–2009 11
2. Number and percentage of linked participants, by demographic and cancer characteristics: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System 12
3. Number of linked participants and weighted percent distribution of survey and cancer time-period characteristics, by sequence of survey and cancer diagnosis: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System 13
4. Number and percentage of cancer types and demographic characteristics: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with Florida Cancer Data System (cancers diagnosed during 2006–2010) compared with overall Florida Cancer Data System, 2006–2010 14

5. Cancer stage at diagnosis, by education level, poverty level, marital status, and sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System after interview 15
6. Percentage of smoking status, self-rated health, and any activity limitation, by sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System before interview (cancer survivor) and not linked (not diagnosed with cancer) 16

Abstract

Background

National survey data linked with state cancer registry data has the potential to create a valuable tool for cancer prevention and control research. A pilot project—developed in a collaboration of the Centers for Disease Control and Prevention's National Center for Health Statistics (NCHS) and the Florida Cancer Data System (FCDS) at the University of Miami—links the records of the 1986–2009 National Health Interview Survey (NHIS) and the 1981–2010 FCDS. The project assesses the feasibility of performing a record linkage between NCHS survey data and a state-based cancer registry, as well as the value of the data produced. The linked NHIS–FCDS data allow researchers to follow NHIS survey participants longitudinally to examine factors associated with future cancer diagnosis, and to assess the characteristics and quality of life among cancer survivors.

Methods

This report provides a preliminary evaluation of the linked national and state cancer data and examines both analytic issues and complications presented by the linkage.

Conclusions

Residential mobility and the number of years of data linked in this project create some analytic challenges and limitations for the types of analyses that can be conducted. However, the linked data set offers the ability to conduct analyses not possible with either data set alone.

Keywords: linkage • cancer registry • data evaluation

Linkage of 1986–2009 National Health Interview Survey With 1981–2010 Florida Cancer Data System

by Eric A. Miller, Ph.D., Donna M. Miller, M.S., Dean H. Judson, Ph.D., Yulei He, Ph.D., Hannah R. Day, Ph.D., Keith Zevallos, and Jennifer D. Parker, Ph.D., National Center for Health Statistics; Jill A. MacKinnon, Ph.D., Monique N. Hernandez, Ph.D., Brad Wohler, M.S., and Recinda Sherman, Ph.D., Florida Cancer Data System, University of Miami Miller School of Medicine; Christina A. Fernandez, M.S.Ed., Laura A. McClure, M.S.P.H, William G. LeBlanc, Ph.D., Stacey L. Tannenbaum, Ph.D., Diane D. Zheng, M.S., and David J. Lee, Ph.D., University of Miami Miller School of Medicine; and Sharon L. Christ, Ph.D., Purdue University

Introduction

This report describes a pilot linkage between the 1986–2009 National Health Interview Survey (NHIS) and the 1981–2010 Florida Cancer Data System (FCDS), the cancer registry for Florida. The report includes a preliminary evaluation of the linked cancer data and describes some of the analytic issues and complications presented by the linkage. This project is a collaboration between FCDS at the University of Miami and the Centers for Disease Control and Prevention's (CDC) National Center for Health Statistics (NCHS). Its goals are to demonstrate and evaluate the feasibility of performing a record linkage between NCHS national sample survey data and a state-based cancer registry, and to assess the value and utility of the data it produces. In addition to anticipated cancer studies that would use the linked data, information from the linkage process can help inform future linkages between sample survey data and cancer registry data, and may be able to be extrapolated to inform other linkages between national sample surveys and state data.

Linkage of NHIS with cancer registry data can potentially produce a valuable tool for cancer prevention and control research. NHIS data contain detailed demographic and health information that is not available from cancer registries. For example, NHIS collects information on income, education level, occupation and industry, health insurance, and self-reported health conditions. In addition, linking multiple years of cancer registry data to multiple survey years adds a longitudinal component to the cross-sectional survey data. For survey participants interviewed before their cancer diagnosis, researchers are able to examine characteristics and risk factors associated with future cancer diagnoses, cancer stage at diagnosis, and survival time. For survey participants interviewed after a cancer diagnosis, researchers are able to examine issues related to cancer survival. Although NHIS asks participants if they have ever been

diagnosed with cancer, NHIS–FCDS linked data allow researchers to examine the health characteristics and quality of life of cancer survivors in much greater detail. For example, FCDS data provide information about cancer stage at diagnosis, which has a substantial impact on cancer survival and future quality of life. As a result, linkage between these data sources provides researchers with an opportunity to conduct a wide array of cancer studies and examine relationships not possible with either data set alone.

Preliminary evaluation of the data describes the results of the linkage and compares the cancer types and demographic characteristics of participants in the linked file to cancer patients represented in FCDS data alone. In addition, some of the unique analytic issues presented by the linkage of multiple years of national health survey data to multiple years of state-specific cancer registry data are discussed. Specifically, the creation of Florida-specific survey weights, the potential impact of residential mobility on the linked file, and NHIS questionnaire and sample design changes over time are detailed. Finally, preliminary tabulations of cancer variables by factors available through NHIS are shown to illustrate use of the linked file.

NHIS–FCDS data are restricted-use data that can be accessed only by approved researchers through the NCHS Research Data Center (RDC). RDC protects the confidentiality of survey respondents by limiting access to approved research proposals, providing a secure work environment, and reviewing researcher output for potential disclosure risk.

Methods

Data Files

National Health Interview Survey

NHIS is an annual, cross-sectional household survey of the U.S. civilian noninstitutionalized population (CNI) conducted by NCHS since 1957. NHIS serves as a principal source of information on the health of the nation. In addition to detailed demographic information, NHIS collects a wide range of health-related information including health status and limitations, health care access and utilization, health insurance, and health behavior information (including specific supplements on cancer health behavior).

Prior to 1997, a core questionnaire, including demographic information and basic health questions, was used for everyone in the household, with supplemental questionnaires used for more specific health topics. In 1997, NHIS underwent a sample redesign and questionnaire revision. While demographic and basic health information were still collected on everyone in the household, one adult and one child within each household were randomly selected to conduct a more detailed survey on more specific health topics (1). These changes are important for the linked data because not all variables have been collected from the entire sample or collected over the entire file time period. This is especially true for many cancer-related variables that are not available from cancer registry data, such as tobacco use and cancer screening history. On the other hand, a number of variables have been collected across the entire survey time period and are not typically available from cancer registry data. These include education level, income, health insurance coverage, occupation and industry, and body mass index (derived from self-reported height and weight). However, data collection in many of these fields has changed slightly over time. Table 1 provides examples of variable availability by survey year for selected topics in NHIS.

Florida Cancer Data System

Funded by Florida and CDC's National Program of Cancer Registries, FCDS is the incidence cancer registry for Florida. Cancer registries are responsible for collecting, managing, and analyzing data on incident cancer cases and cancer deaths, and they are essential in monitoring progress in cancer prevention and control. With more than 110,000 newly diagnosed cancer cases per year, FCDS is the third largest registry in the country and represents approximately 6% of all U.S. cancer cases. FCDS is ''gold certified'' by the North American Association of Central Cancer Registries based on the timeliness and completeness of the data collected. FCDS began collecting data in 1981; however, the data system contains cancer records dating to the 1950s, either from voluntary reporting to the Florida Department of Health prior to FCDS or from previous diagnoses of cancer among those with cancers diagnosed from 1981 forward (2).

Cancer registries collect information on the type of cancer, extent (i.e., stage at diagnosis), initial treatment, and basic demographic characteristics. Cancer registries collect data at the tumor level; therefore, persons diagnosed with more than one incident cancer are included in the registry for each tumor. Approximately one-fifth of persons in FCDS have been diagnosed with multiple incident cancers. Tumors that occur as a result of cancer cells spreading (metastasis) from the original location are not incident cancers. Cancer registries are required to collect direct identifiers in order to eliminate duplicate records, consolidate multiple records, and conduct follow-up on vital status through data linkage with the NCHS National Death Index.

Linkage Approval and Methods

Designated Agent Agreement and project approval

NHIS data are protected by Section 308(d) of the Public Health Service Act and by the Confidential Information Protection and Statistical Efficiency Act (CIPSEA). Under CIPSEA legislation, FCDS was required to sign a Designated Agent Agreement, and all FCDS analysts were required to complete NCHS confidentiality training and paperwork before accessing confidential NCHS data to conduct the linkage. Additionally, all NCHS record linkage activities must be reviewed and

approved by the NCHS Ethics Review Board (ERB). ERB is a formally appointed ethics review committee established to ensure that research involving human participants protects their rights and welfare and conforms to federal regulations. ERB granted approval for the NHIS–FCDS data linkage in May 2008. The project then underwent review and approval by the Florida Department of Health Institutional Review Board, and the linkage was initiated in spring 2009.

Linkage eligibility

To be eligible for FCDS linkage, NHIS participants had to provide sufficient direct identifiers [e.g., name, date of birth, and Social Security number (SSN)] and had not refused to have their data linked. Linkage eligibility requirements varied across NHIS years due to changes in how ERB determined whether a survey participant was deemed eligible or ineligible for linkage. Before 2007, participants who refused to provide direct identifiers were implicitly considered to have refused record linkage. The refusal rate increased between 1997 and 2006, reducing the number of NHIS participants eligible for record linkages. Beginning in 2007, NHIS successfully reduced linkage refusal rates by adding a short introduction prior to the request for SSN, requesting only the last four digits of SSN rather than the full nine digits, and asking participants who did not provide SSN (or Medicare number) for their explicit permission to link to administrative records. Demographic differences have been found between those who refuse to provide an SSN and those who do not (3), which could potentially bias estimates calculated from the linked data. Currently, this issue is treated as a nonresponse bias, and analytic methods are used to adjust sample weights in an attempt to account for differences (4).

NCHS submission file preparation and description

NCHS provided FCDS with electronic data files containing direct identifiers and an NCHS-created control number. The files used for matching did not contain the NCHS survey public-use identifiers available from the public-use files. For the NHIS–FCDS linkage, all NHIS survey participants meeting the eligibility criteria, including non-Florida residents, were submitted for linkage. A match was not requested for participants deemed ineligible for the linkage. The data were securely sent to FCDS on an encrypted CD.

Linkage process

FCDS used LinkPlus software (5) to conduct the linkage. LinkPlus, which CDC developed for cancer registries, uses probabilistic linkage methods. In addition to finding exact matches, probabilistic methods can identify matching records when not all matching variables agree or when variables only partially match. Records are given scores for the number of matching or partially matching fields, with certain fields given more weight than others. For example, matching on an SSN is given more weight than matching on place of residence.

For this linkage, matching scores were based on SSN (nine and four digits), date of birth, last name, first name, middle initial, sex, race (white, black, American Indian or Alaska Native, Asian, other), five-character city name, and five-digit zip code. The city-name field was truncated to five characters to improve matching efficiency and reduce errors due to misspellings or nonstandardized city names. To account for possible differences in first and last names due to misspellings, alternative spellings, and clerical errors, LinkPlus includes the New York State Identification and Intelligence System, or NYSIIS, which is commonly used in data linkage to convert names to phonetic codes.

Matches with scores of 30 and above were considered true matches, and scores below 17 were considered false matches. Matches with scores from 17 through 29 were manually reviewed to determine if they were true matches. To assess the number of false or missed matches, 100 consecutive records just above and below the cutoffs were reviewed. Of those above the upper-score threshold, one false-positive match was identified. Of those below the lower threshold, matches appeared to be missed in six records. The six false-negative matches were all women and fell just below the score cutoff, primarily due to inconsistencies between last names on the two files.

Processing of return file at NCHS

The return file containing the extracted FCDS data for matched NHIS records was remerged with the public-use survey identifiers to enable the addition of survey data, and then intermediate files were destroyed. The return file was processed to create one record per person. Because so few children were linked with FCDS, the file was limited to adults aged 18 and over as of their NHIS interview. Variables were added to provide linkage eligibility status of survey participants and identify survey participants who moved after the survey interview. In addition, sample weights appropriate to the linked data were created and added to the file.

Creation of Florida-specific sample weights

NHIS public-use and restricted-use analytic survey files include annual sample-weight variables, which are inversely proportional to each participant's selection probability and have been adjusted for oversampling of specific subgroups and differential nonresponse. NHIS sample weights are intended to be representative of the CNI population of the United States in each survey year.

Although other approaches are possible, sample weights for the linked NHIS–FCDS data were created to represent the CNI population of Florida. The Florida-specific sample weights were calculated for all linkage-eligible NHIS participants, regardless of whether they linked with FCDS. Because the sample weights created for the linked NHIS–FCDS data were representative of the Florida CNI population in each survey year, survey participants interviewed in other states consequently received a sample weight of zero, even

if they linked with FCDS. As a result, using the Florida CNI weighting strategy, survey respondents interviewed outside of Florida who later moved to Florida where they were diagnosed with cancer drop out of weighted analyses.

Two sets of Florida CNI sample weights were created. The first were created for all survey participants aged 18 and over in all NHIS years. An additional sample-weight variable was created to account for the sample redesign and Sample Adult file beginning in 1997. Because not all NHIS participants were eligible for the linkage, NHIS sample weights for Florida residents at the time of survey were first adjusted for linkage ineligibility using PROC WTADJUST in SUDAAN software (4). The adjusted weights were then poststratified to the estimated annual CNI population of Florida for the corresponding survey year. Two different sex- and age-specific (18–39, 40–64, and 65 and over) estimates of the annual Florida CNI population were used to create the poststratified weights. One method used population estimates directly from NHIS sample weights for each survey year among Florida residents. Because of substantial year-to-year variability among population subgroups in earlier years of NHIS, a second method used total Florida population data from the Florida Department of Health's Office of Health Statistics and Assessment (available from: http://www.floridacharts.com/charts/default.aspx). This method estimated the CNI proportion of the total population using NHIS data from 1997–2009, when CNI population estimates were more stable. The first step compared the Florida CNI population data from NHIS to the total Florida population for these 13 years to calculate the average percent CNI population over the time period. The second step applied the average percent CNI to the total Florida population data over the entire time period covered by the linkage, 1986–2009, to estimate the CNI population for each survey year. The two sample weights were highly correlated (r = 0.99).

Evaluation of Linked File

To conduct the evaluation, Integrated Health Interview Series (IHIS) data, which harmonizes NHIS data across survey years (6), were merged with the linked file to obtain several NHIS variables. For overall evaluations, the following NHIS variables collected at interview were used: sex, age (18–39, 40–64, and 65 and over), race and ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, and all other races and ethnicities), educational level [less than high school degree, high school graduate/General Educational Development (GED) certificate/some college, post-high school degree], and self-rated health status (excellent/very good/good, fair/poor). IHIS documentation contains further information on how the collection and subsequent harmonized coding of race and ethnicity and education information have changed over time (6). Except for the presentation of unweighted numbers, (see Results and Detailed Tables), analysis of the file accounted for the complex survey design, and estimates were weighted using the Florida-specific sample weights created for this file. Comparisons of unweighted percentages were not statistically tested.

Evaluation of NCHS–FCDS data first consisted of an overall description of NHIS participants who were linked with FCDS, including the unweighted and weighted distributions of demographic and cancer characteristics in the linked population. Because linked NHIS participants who were not residents of Florida at the time of survey receive a zero weight and drop out of weighted analyses, an assessment was made of how their characteristics differed from the linked NHIS participants who were Florida residents at the time of survey. Additional weighted examinations of cancer types and characteristics related to the linkage time period were made by sequence of cancer diagnosis and survey participation.

To assess the representativeness of the linked cancer cases in NHIS–FCDS data relative to all Florida cancer cases, the demographic characteristics and cancer types were examined for Florida linked survey participants and the unlinked FCDS data for 2006–2010, the most current 5 years of data available for analysis. The evaluation was limited to survey participants who had their first cancers diagnosed during 2006–2010 and those diagnosed after survey participation, to examine incident cancers only. Because the two groups are not statistically independent, comparisons were not statistically tested.

Due to the potential loss of sample size from linkage ineligibility, weights that were specific to Florida residents at the time of survey, and the unavailability of certain variables during the linkage time period, an examination was made of how these factors could affect an analysis of the linked NHIS–FCDS data. Specifically, the reduction in sample size after accounting for each of these factors was documented, using a hypothetical analysis of female breast cancer and tobacco use as an example.

As an initial analysis of the linked data, basic statistics available from the linked NHIS–FCDS data but not available from either data source alone were calculated to examine preliminary relationships between demographic and health factors and cancer. For this analysis, an examination was made of the relationship between cancers diagnosed at a late stage (i.e., with tumor having spread regionally or to a distant site in the body) and demographic characteristics of those who were linked with FCDS. In addition, selected health characteristics were compared between cancer survivors (survey participants linked with FCDS prior to their survey interview) and survey participants who had not been diagnosed with cancer (participants not linked with FCDS). All estimates were calculated overall and by sex. Other information from IHIS used for this evaluation included marital status (married, not married/widowed/separated), smoking status (never, current, former), and poverty status based on income above or below the federal poverty level. Poverty

status used the POORYN variable from IHIS, which is based on income and family size and its relation to the poverty threshold set by the U.S. Census Bureau for each survey year (6). Income is not imputed for this variable; as a result, data were missing from 676 Florida NHIS participants who were linked with FCDS (11.4%). The demographic variables and activity limitation variable were available for all years of NHIS. Because smoking status was available only for select years before 1997 and was not always collected on the full sample, smoking estimates reported here were based on sample adults in 1997–2009 NHIS data.

Results

Linked NHIS–FCDS Data Files

Characteristics of linked survey participants

NHIS participants aged 18 and over totaled 1,708,723 during 1986–2009. Approximately one-third of the participants were ineligible for the linkage (n = 575,327). After running LinkPlus and performing manual review, NHIS participants who were linked with FCDS totaled 8,210. For this report, linked records with cancers missing a date of diagnosis (n = 59) or cancers reported prior to 1981 with no additional cancers diagnosed after 1981 were excluded (n = 41).

Of the remaining 8,110 linked participants, 1,443 (17.8% unweighted) were linked with more than one tumor in FCDS (Table 2). Approximately one-half (unweighted) of the linked survey participants were men (50.2%), approximately 90% were aged 40 and over at the time of survey (89.9%), and more than 70% were non-Hispanic white (73.3%). Almost two-thirds (unweighted) of the linked participants (65.0%) were interviewed during 1986–1996; the remaining 35.0% were interviewed during 1997–2009, after the NHIS sample and questionnaire redesign. By time period of cancer diagnosis, 46.0% (unweighted) of linked participants had cancers diagnosed during 1996–2005.

When the Florida-specific weights were applied, limiting the analysis to Florida residents at the time of survey, the percentage of linked participants who were Hispanic decreased and the percentage of participants aged 65 and over increased (Table 2). This difference between the unweighted and weighted distributions is likely a result of the weights accounting for the oversampling of Hispanic participants in the NHIS survey design. In addition, limiting the sample to Florida residents at the time of survey resulted in a larger percentage of linked survey participants from survey years 1997–2009 and with 0–5 years between survey and cancer diagnosis.

The total number of NHIS–FCDS linked participants interviewed outside of Florida was 1,829, 22.6% (unweighted) of all linked participants. Although not statistically tested, the sample of non-Florida linked survey participants had higher percentages (unweighted) of men, non-Hispanic white persons, and post-high school degrees than the Florida linked survey participants. Possibly due to being healthy enough to move, 85.3% (unweighted) of the sampled non-Florida linked survey participants reported excellent/very good/good health status at the time of survey; 76.7% (unweighted) of Florida linked survey participants reported excellent/very good/good health status. The majority of non-Florida linked participants were interviewed during 1986–1996 and close to 80% (unweighted) of their cancers (78.9%) were diagnosed during 1996–2010.

Table 3 presents the number and distribution of cancers among linked Florida participants by sequence of cancer diagnosis related to survey participation. The cancer cases were limited to the first cancer diagnosis among those who were linked to more than one cancer. The majority of linked survey participants were diagnosed with cancer after they had participated in the survey, which allows researchers to examine the characteristics and health behaviors of survey participants before they were diagnosed with cancer. Among five commonly diagnosed cancers (prostate, female breast, lung, colorectal, and bladder), the number of cases ranged from 200 to 724 when the cancer was diagnosed after survey participation. More than 60% of linked participants with cancer diagnosed after the survey (63.6%) were interviewed during 1986–1996, and 84.3% of the cancers were diagnosed during 1996–2010. When examining the length of time between survey and cancer diagnosis, 44.3% of the linked survey participants were diagnosed within 5 years after the interview.

A total of 1,908 linked survey participants had cancer diagnosed before participating in NHIS (Table 3). The number of cases for the specific cancer types ranged from 89 to 403. Consistent with poor survival among patients diagnosed with lung cancer, the proportion of lung cancer survivors was substantially smaller for linked survey participants diagnosed with cancer before rather than after the survey interview. Additionally, the slightly higher percentages of breast and prostate cancers among cancer survivors may be due to the fact that many of these cancers are diagnosed at an early stage and have high survival rates when detected early (7). Approximately two-thirds (weighted) of linked survey participants with a cancer diagnosis prior to the survey (66.8%) were aged 65 and over at the time of survey. By time period, 67.5% (weighted) of linked survey participants with a previous cancer diagnosis were interviewed during 1997–2009, and 79.9% of the cancers were diagnosed during 1986–2005. More than one-half (weighted) were interviewed within 5 years from their cancer diagnosis date in FCDS (53.5%).

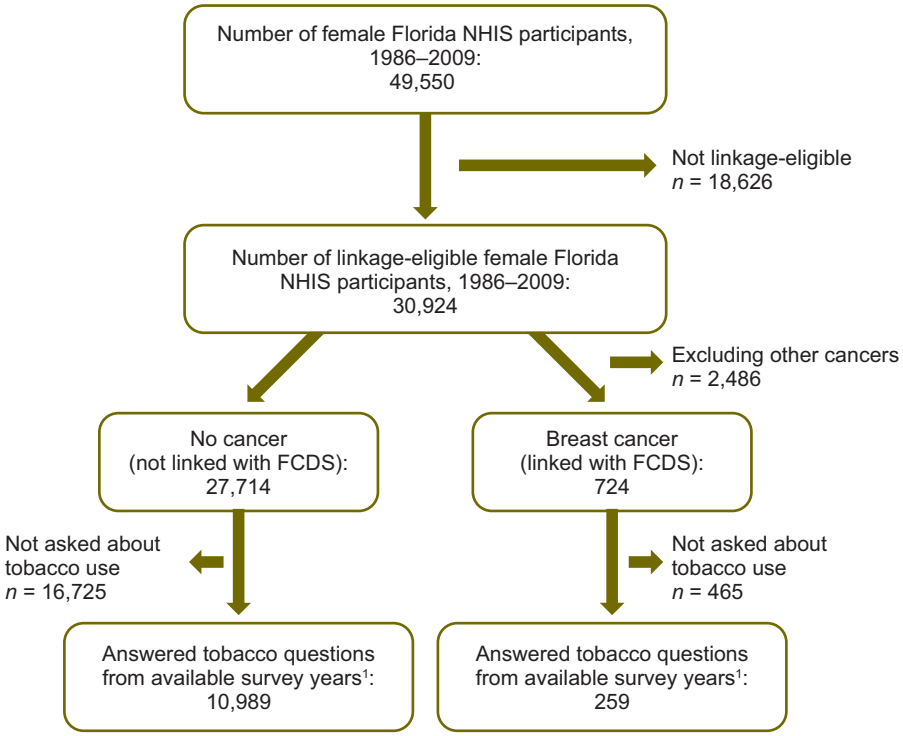
Comparison with unlinked FCDS

By inspection, the five commonly diagnosed cancers among Florida linked survey participants diagnosed during 2006–2010 were consistent with unlinked FCDS data, with some exceptions. For example, female breast cancer was the most commonly diagnosed cancer among Florida survey

participants, and lung cancer was the most commonly diagnosed cancer in the unlinked FCDS data (Table 4). By sex, a similar percentage of men and women are represented in the linked data, but the percentage of men in the unlinked FCDS data is 53.2%. The racial and ethnic distribution of Florida linked survey participants was generally similar to those with cancer diagnosis in the unlinked FCDS, except that the estimated percentage of non-Hispanic black linked survey participants was higher than the corresponding percentage in the unlinked FCDS.

Potential loss of analytic sample size in linked NCHS–FCDS data analysis

Although the large annual sample populations of NHIS make this linkage possible, not all participants can be included in analyses of the linked data for various reasons. This loss of analytic sample size can substantially reduce the power of certain analyses or the ability to examine differences between subgroups, and can introduce bias if differences exist between those who are included in the sample and those who are not. For example, consider tobacco use and breast cancer risk (Figure 1): Of the 911,262 women aged 18 and over in NHIS during 1986–2009, 49,550 were interviewed in Florida. Of those, 18,626 were ineligible for linkage. After excluding women diagnosed with other cancers in FCDS, 27,714 women without cancer and 724 women diagnosed with breast cancer were available for analysis. However, to conduct an analysis based on tobacco use, which was not collected every year or from the full sample, data are available only for 10,989 women without cancer and 259 with breast cancer. Similarly, questions regarding cancer screening and family history of cancer were included only in select years with NHIS cancer supplemental questionnaires. Of the same 724 breast cancer cases, information on previous mammography use is available only for 112 women.



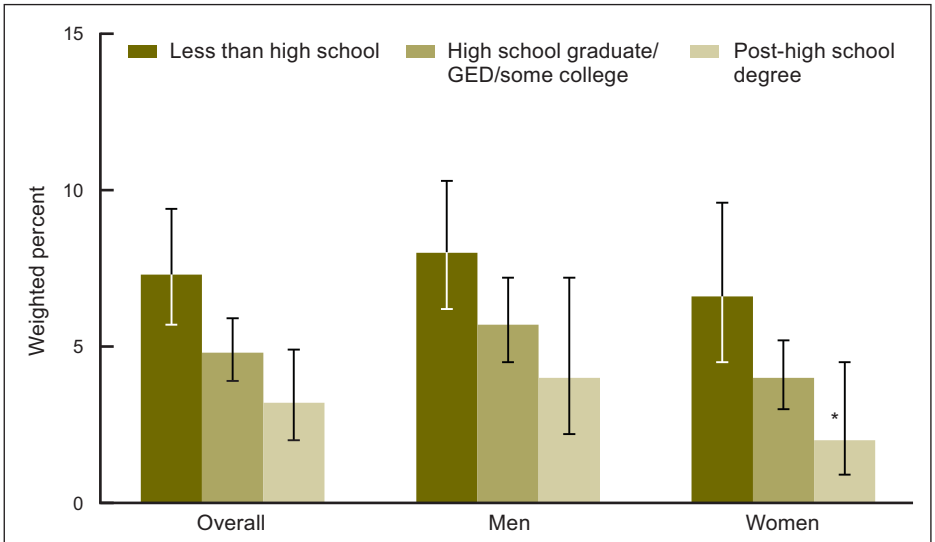
[1]See Table 1 of this report for details on the availability of tobacco questions across NHIS years.
NOTE: NHIS is National Health Interview Survey; FCDS is Florida Cancer Data System.
SOURCE: CDC/NCHS, 1986–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System.

Figure 1. Number of female participants available for breast cancer risk analysis using survey tobacco questions: 1986–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System

Initial analysis of linked NCHS–FCDS data

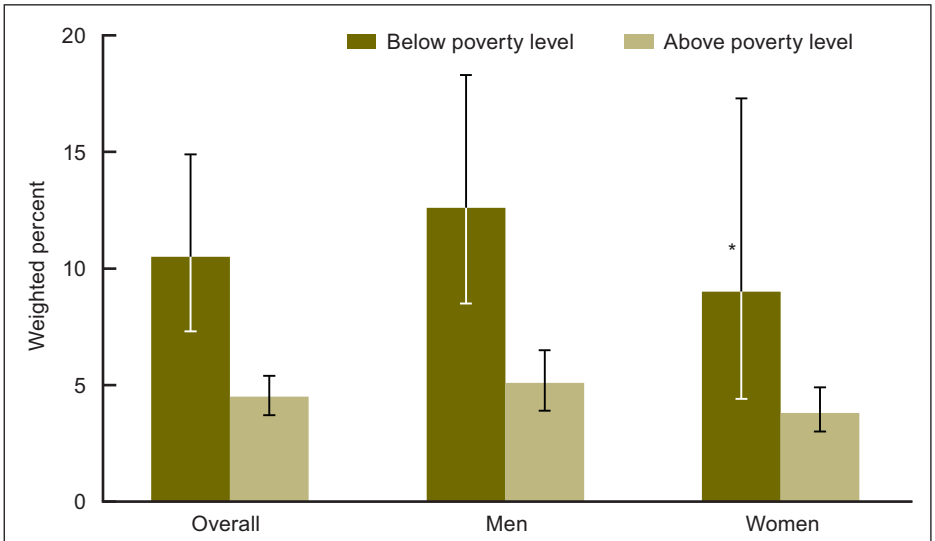
The weighted percentage of participants diagnosed with late-stage cancer in the linked NHIS–FCDS data was greater with decreasing education level for both men and women (Figure 2, Table 5). Men (4.0%) and women (2.0%) with a post-high school degree were less likely to be diagnosed with late-stage cancer compared with those having less than a high school degree (8.0% men and 6.6% women). Similarly, men (5.1%) and women (3.8%) living above the federal poverty level were less likely to be diagnosed with a late-stage cancer compared with those below the federal poverty level (12.6% men and 9.0% women), but this difference was statistically significant only among men (Figure 3, Table 5). By marital status, only married women were significantly less likely to be diagnosed with a late-stage cancer (2.9%) compared with those who were not married/divorced/separated (6.2%) (Figure 4, Table 5); however, married men were significantly more likely to be diagnosed with late-stage cancer than married women.

When comparing health characteristics of cancer survivors (Florida NHIS participants who were linked to FCDS prior to survey) with those who had not been diagnosed with cancer (Florida participants not linked with FCDS before or after interview), the weighted percentage of cancer survivors reporting they were in fair/poor health (33.2%) was more than double that of participants who had not been diagnosed with cancer (13.0%) (Figure 5, Table 6). This association was true for both men and women. Cancer survivors were also more likely to report having some type of activity limitation compared with those not diagnosed with cancer (Figure 6, Table 6). Both men (40.1%) and women (36.1%) who were cancer survivors were more than twice as likely to report having an activity limitation compared with those who had



* Figure does not meet standards of reliability or precision; relative standard error is greater than 30%.
NOTES: Late-stage cancers include tumors that have spread regionally or to distant sites of the body at diagnosis. GED is General Educational Development high school equivalency diploma. Confidence interval is 95%.
SOURCE: CDC/NCHS, 1986–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System.

Figure 2. Percentage of diagnosed late-stage cancers, by education level and sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System after interview



* Figure does not meet standards of reliability or precision; relative standard error is greater than 30%.
NOTES: Late-stage cancers include tumors that have spread regionally or to distant sites of the body at diagnosis. Confidence interval is 95%.
SOURCE: CDC/NCHS, 1986–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System.

Figure 3. Percentage of diagnosed late-stage cancers among those above and below federal poverty level, by sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System after interview

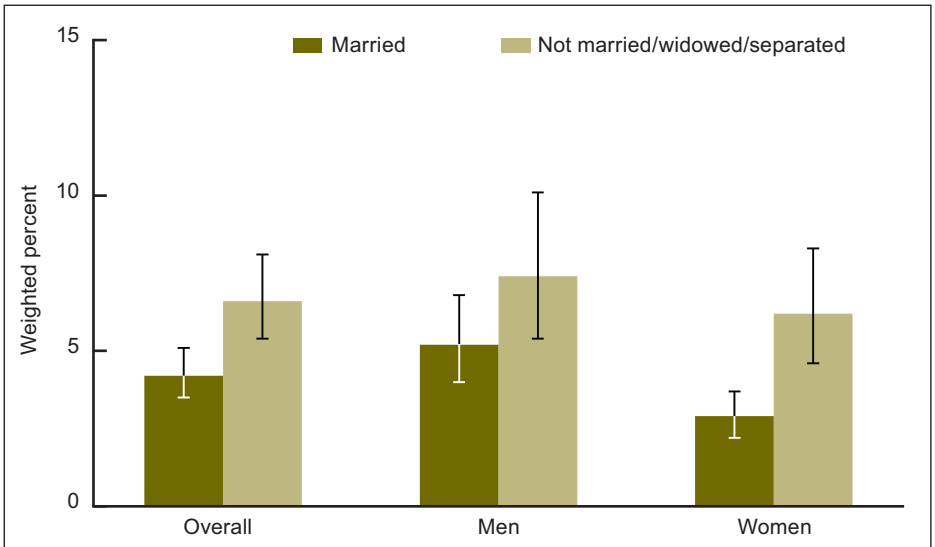
not been diagnosed with cancer (14.4% men and 16.4% women). Among those interviewed during 1997–2009, the percentage of cancer survivors who were current smokers at the time of survey was less than 11% for both men and women (Figure 7, Table 6), approximately one-half the percentage among women (20.2%) and less than one-half the percentage among men (25.0%) who had not been diagnosed with cancer.

Discussion

The linked NHIS–FCDS data allow researchers to follow NHIS survey participants longitudinally to examine factors associated with future diagnosis of cancer and to assess the characteristics and quality of life among cancer survivors. FCDS data provide detailed information on the type of cancer and stage at diagnosis, while NHIS provides detailed sociodemographic information and extensive self-reported data on health characteristics, including health conditions, access to health care, and health care utilization that are not available from cancer registry data. For example, using education and poverty level, the linked data allow researchers to examine socioeconomic characteristics at the individual level rather than relying on census tract-level estimates, which is the common approach. Using self-rated health and the presence of an activity limitation as another example demonstrates the data set's ability to assess morbidity among cancer survivors.

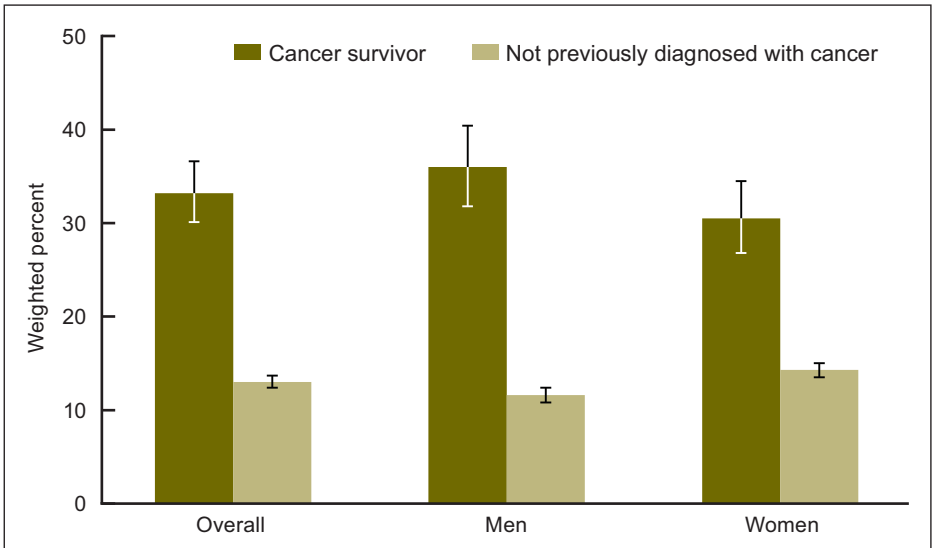
Overall, the linked NHIS–FCDS data had a similar distribution of major cancer types compared with unlinked FCDS data, with a slight over-representation of breast cancer, which may be explained by a higher percentage of women in the linked data. The linked NHIS–FCDS data also had a slightly higher percentage of non-Hispanic black cancer cases than the unlinked FCDS data. This could result from oversampling in NHIS that was not fully adjusted for in the new weights.

Residential mobility proved to be one of the major challenges with the linked NCHS–FCDS data. While the current weighting strategy (creating weights to the year of the survey) is valid, it results in a nontrivial reduction in sample size by giving a zero weight to linked NHIS participants who were interviewed in other states but were diagnosed with cancer in Florida. Alternative weighting strategies are being investigated to reduce the loss of sample size. However, noticeable differences were observed between the



NOTES: Late-stage cancers include tumors that have spread regionally or to distant sites of the body at diagnosis. Confidence interval is 95%.
SOURCE: CDC/NCHS, 1986–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System.

Figure 4. Percentage of diagnosed late-stage cancers, by marital status and sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System after interview



NOTES: Cancer survivors are survey participants whose data were linked with the Florida Cancer Data System before their survey interview; those not diagnosed with cancer were not linked with the data system.Confidence interval is 95%.
SOURCE: CDC/NCHS, 1986–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System.

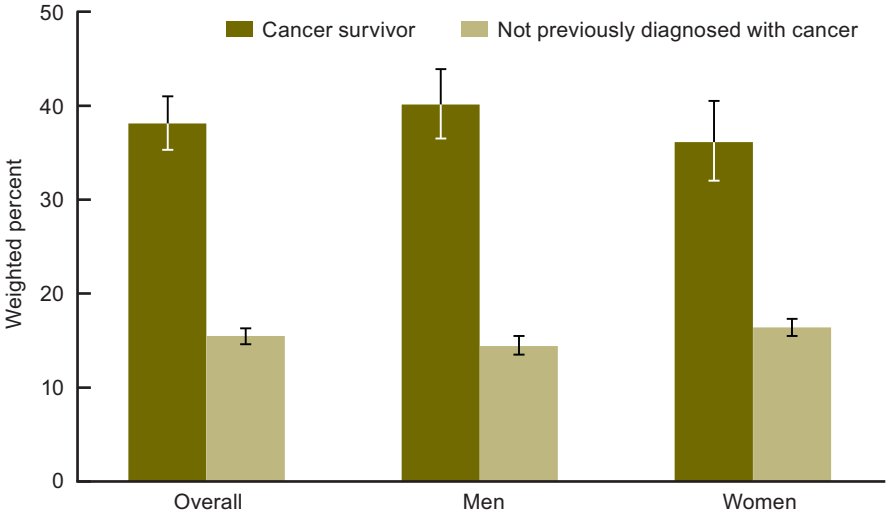
Figure 5. Percentage of self-rated fair and poor health status, by cancer diagnosis and sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System before interview (cancer survivor) and not linked (not diagnosed with cancer)

demographic characteristics of non-Florida residents who were linked with FCDS and the demographic characteristics of linked Florida participants, and caution is needed to not bias inference by including non-Florida residents without appropriate methods.

Out-migration is another potential problem in the data set. In this case, all survey participants from Florida are given a Florida sample weight for analysis of the linked NHIS–FCDS data, but it is not possible to identify if some of these participants moved or were diagnosed with cancer in another state. The number of survey participants who were Florida residents at the time of interview but moved out of Florida after the interview is unknown, but the number may not be trivial. Of the 3,526 deaths among Florida survey participants who were linked with FCDS, 7.8% (n = 275) died in another state.

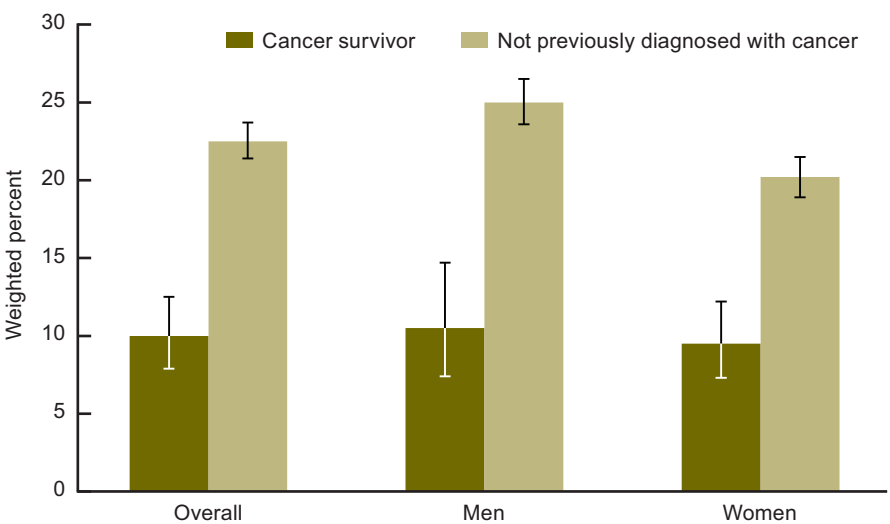
The loss of NHIS participants from the study population, due either to linkage ineligibility or migration, may adversely affect the generalizability of study inferences or the ability to calculate accurate prevalence estimates. Although the sample weights used for the estimates in this report were adjusted for linkage ineligibility and calibrated to the Florida population, it is unclear how differences in characteristics may influence results between those who are linkage eligible and those ineligible, as well as between those who migrated into or out of Florida compared with those who were constant Florida residents.

Using 23 years of NHIS data allowed for a sufficient number of Florida NHIS participants to be linked with FCDS to conduct many types of data analyses. However, changes in the questionnaire during this time period limit some types of analyses as well as the ability to control for potential confounders. While NHIS data from 1997 forward are more consistent, the number of cancers diagnosed from those survey years limits the ability to examine many individual types of cancer; linkage eligibility criteria and the provision of SSN also changed during the time period, further reducing the study population. In addition, survey responses may not be representative for time-variant characteristics because of the potential for a substantial number of years to elapse between the survey and cancer diagnosis. Among participants diagnosed with cancer after their survey participation, nearly 60% had more than 5 years elapse between survey and cancer diagnosis (55.7%). For time-variant characteristics such as



NOTES: Cancer survivors are survey participants whose data were linked with the Florida Cancer Data System before their survey interview; those not diagnosed with cancer were not linked with the data system. Confidence interval is 95%.
SOURCE: CDC/NCHS, 1986–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System.

Figure 6. Percentage having any activity limitation, by sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System before interview (cancer survivor) and not linked (not diagnosed with cancer)



NOTES: Current smokers have smoked at least 100 cigarettes in their lifetime and still currently smoke. Cancer survivors are survey participants whose data were linked with the Florida Cancer Data System before their survey interview; those not diagnosed with cancer were not linked with the data system. Confidence interval is 95%.
SOURCE: CDC/NCHS, 1997–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System.

Figure 7. Percentage of current smokers, by sex: 1997–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System before interview (cancer survivor) and not linked (not diagnosed with cancer)

health insurance coverage or meeting cancer screening guidelines, the survey response may not be representative of the full interval from or until cancer diagnosis.

Conclusion

The pilot linkage of NHIS and FCDS demonstrates the feasibility of linking national-level sample survey data with state-based cancer registry data. The linked data set provides the ability to conduct analyses that are not possible with either data set alone. NHIS adds participant-level characteristics that are highly desired among researchers using cancer registry data, and FCDS provides detailed outcome information for those who are diagnosed with cancer after survey participation and detailed baseline information for cancer survivors who participate in the survey after their cancer diagnosis. The number of years of data linked in this project creates some analytic challenges, and limitations exist in the types of analyses that can be conducted. For example, the loss of sample due to linkage eligibility and mobility in and out of Florida make the data more suited for examining relationships among factors than for other uses, such as prevalence estimates. The methodological issues encountered with these data may be relevant to other national and state data linkages.

References

1. National Center for Health Statistics. 1997 National Health Interview Survey (NHIS) public use data release. 2000. Available from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/1997/srvydesc.pdf.
2. Florida Cancer Data System (FCDS). 2014. Available from: http://fcds.med.miami.edu/inc/welcome.shtml [Accessed February 18, 2014].
3. Miller DM, Gindi RM, Parker JD. Trends in record linkage refusal rates: Characteristics of National Health Interview Survey participants who refuse record linkage. Paper presented at the 2011 meeting of the Joint Statistical Meeting, July 30–August 4, 2011, Miami Beach, FL. 2011.
4. Judson DH, Parker JD, Larsen MD. Adjusting sample weights for linkage-eligibility using SUDAAN. Hyattsville, MD: National Center for Health Statistics. 2013. Available from: http://www.cdc.gov/nchs/data/datalinkage/adjusting_sample_weights_for_linkage_eligibility_using_sudaan.pdf.

5. Centers for Disease Control and Prevention. Registry Plus Link Plus. Available from: http://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm [Accessed July 15, 2014].
6. Minnesota Population Center and State Health Access Data Assistance Center. Integrated Health Interview Series (Version 5.0). Minneapolis, MN: University of Minnesota. 2012. Available from: https://www.ihis.us/ihis/ [Accessed March 12, 2014].
7. American Cancer Society. Cancer Facts & Figures 2013. Atlanta, GA. 2013.

Table 1. Availability of selected questionnaire topics, by survey year: National Health Interview Survey, 1986–2009

Questionnaire topic	Survey year																							
	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Tobacco use	. . .	[1]X	. . .	. . .	X	X	[2]X	[3]X	[1]X	[1]X	. . .	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X
Alcohol use	. . .	X	X	. . .	X	X	X	. . .	. . .	. . .	. . .	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X	[4]X
Family history of cancer	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	[4]X	. . .	. . .	. . .	. . .	[4]X	. . .	. . .	. . .	. . .
Time since last mammogram	. . .	X	. . .	. . .	X	. . .	X	. . .	. . .	. . .	. . .	. . .	. . .	[4]X	[4]X	. . .	. . .	[4]X	. . .	[4]X	. . .	. . .	[4]X	. . .
Time since last colorectal test	. . .	[5]X	. . .	. . .	. . .	. . .	[5]X	. . .	. . .	. . .	. . .	. . .	. . .	. . .	[4]X	. . .	. . .	[4]X	. . .	[4]X	. . .	. . .	[4]X	. . .
Body mass index	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Education level	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Income	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Health insurance coverage	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Occupation or industry	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

. . . Category not applicable.

[1]One-half of sampled adults aged 18 and over.

[2]All sampled adults, with partial-year data collection.

[3]One-half of sampled adults, with partial-year data collection.

[4]Sampled adults only.

[5]One-half of sampled adults aged 40 and over, with partial-year data collection.

SOURCE: CDC/NCHS, National Health Interview Survey, 1986–2009.

Table 2. Number and percentage of linked participants, by demographic and cancer characteristics: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System

Characteristic	All linked survey participants[1]		Florida residents at time of survey[1]			Non-Florida residents at time of survey[1]	
	Unweighted number	Unweighted percent	Unweighted number	Unweighted percent	Weighted percent	Unweighted number	Unweighted percent
Total	8,110	100.0	6,281	100.0	100.0	1,829	100.0
Number of linked tumors							
One	6,667	82.2	5,147	82.0	81.9	1,520	83.1
More than one	1,443	17.8	1,134	18.0	18.1	309	16.9
Sex							
Male	4,074	50.2	3,110	49.5	50.6	964	52.7
Female	4,036	49.8	3,171	50.5	49.4	865	47.3
Age (years)							
18–39	821	10.1	607	9.7	9.2	214	11.7
40–64	3,784	46.7	2,730	43.5	42.1	1,054	57.6
65 and over	3,505	43.2	2,944	46.9	48.7	561	30.7
Race and ethnicity							
Hispanic	1,066	13.1	971	15.5	10.9	95	5.2
Non-Hispanic white	5,944	73.3	4,330	68.9	76.4	1,614	88.2
Non-Hispanic black	978	12.1	881	14.0	11.1	97	5.3
Other	122	1.5	99	1.6	1.6	23	1.3
Education level							
Less than high school	1,826	22.7	1,523	24.5	21.4	303	16.6
High school graduate, GED, or some college	4,562	56.8	3,498	56.3	57.7	1,064	58.4
Post-high school degree	1,643	20.5	1,188	19.1	20.9	455	25.0
Missing	79	. . .	72	. . .	. . .	7	. . .
Self-rated health							
Excellent, very good, or good	6,356	78.7	4,800	76.7	77.1	1,556	85.3
Fair or poor	1,724	21.3	1,455	23.3	22.9	269	14.7
Missing	30	. . .	27	. . .	. . .	4	. . .
Survey years							
1986–1996	5,271	65.0	3,782	60.2	52.6	1,489	81.4
1997–2009	2,839	35.0	2,499	39.8	47.4	340	18.6
Cancer diagnosis[2]							
Before 1986	443	5.5	354	5.6	5.7	89	4.9
1986–1995	1,991	24.6	1,694	27.0	25.4	297	16.2
1996–2005	3,705	45.7	2,812	44.8	45.0	893	48.8
2006–2010	1,971	24.3	1,421	22.6	23.9	550	30.1
Years between survey and cancer diagnosis[2]							
0–5	3,307	40.8	2,861	45.5	47.6	446	24.4
6–10	2,192	27.0	1,720	27.4	26.7	472	25.8
More than 10	2,611	32.2	1,700	27.1	25.8	911	49.8

. . . Category not applicable.

[1]Cancer identified through linkage with the Florida Cancer Data System could have occurred before or after survey participation.

[2]Estimates limited to first cancer diagnosis among participants with more than one cancer diagnosis in the Florida Cancer Data System.

NOTE: GED is General Educational Development high school equivalency diploma.

SOURCE: CDC/NCHS, 1986–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System.

Table 3. Number of linked participants and weighted percent distribution of survey and cancer time-period characteristics, by sequence of survey and cancer diagnosis: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System

Characteristic	Florida residents at time of survey with interview before cancer diagnosis[1]		Florida residents at time of survey with cancer diagnosis before interview[1]	
	Unweighted number	Weighted percent	Unweighted number	Weighted percent
Total	4,367	100.0	1,908	100.0
Cancer type				
Female breast	724	16.4	401	20.9
Prostate	666	14.8	403	21.2
Lung	613	14.4	89	4.1
Colorectal	446	10.2	233	12.0
Bladder	200	4.4	120	6.4
All others	1,716	39.8	662	35.4
Missing	2	. . .	. . .	. . .
Age at survey (years)				
18–39	521	11.7	86	4.6
40–64	2,165	49.5	561	28.6
65 and over	1,681	38.9	1,261	66.8
Survey years				
1986–1996	3,013	63.6	765	32.5
1997–2009	1,354	36.4	1,143	67.5
Cancer diagnosis				
Before 1986	. . .	. . .	349	16.1
1986–1995	753	15.7	941	43.2
1996–2005	2,235	49.5	576	36.7
2006–2010	1,379	34.8	42	4.0
Years between survey and cancer diagnosis				
0–5	1,778	44.3	1,082	53.5
6–10	1,239	27.2	481	25.8
More than 10	1,350	28.5	345	20.7

. . . Category not applicable.

[1]Estimates limited to first cancer diagnosis among participants with more than one cancer diagnosis in the Florida Cancer Data System.

SOURCE: CDC/NCHS, 1986–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System.

Table 4. Number and percentage of cancer types and demographic characteristics: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with Florida Cancer Data System (cancers diagnosed during 2006–2010) compared with overall Florida Cancer Data System, 2006–2010

Characteristic	Florida residents at time of survey and cancer diagnosed during 2006–2010			Florida Cancer Data System, 2006–2010[1]	
	Unweighted number	Weighted percent	Confidence interval[2]	Unweighted number	Unweighted percent
Cancer type					
All cancers	1379	100.0	. . .	527,730	100.0
Female breast	242	17.3	14.9–20.0	68,618	13.0
Prostate	187	13.0	11.2–15.0	73,746	14.0
Lung	172	13.0	11.3–15.0	81,023	15.4
Colorectal	131	9.7	7.6–12.4	49,592	9.4
Bladder	54	4.2	3.3–5.2	24,978	4.7
All others	593	42.8	39.9–45.8	229,773	43.5
Sex					
Male	634	48.4	45.4–51.3	280,767	53.2
Female	745	51.6	48.7–54.6	246,963	46.8
Race and ethnicity					
Hispanic	268	12.4	10.1–15.2	68,008	12.9
Non-Hispanic white	850	71.3	67.5–74.9	399,854	75.8
Non-Hispanic black	234	14.2	11.9–16.8	49,712	9.4
Other	27	2.1	1.4–3.1	10,156	1.9

. . . Category not applicable.

[1]Available from: http://www.cancer-rates.info/naaccr/ [Accessed December 26, 2013].

[2]Equals 95%.

SOURCE: CDC/NCHS, 1986–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System.

Table 5. Cancer stage at diagnosis, by education level, poverty level, marital status, and sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System after interview

Characteristic	Early stage[1]			Late stage[2]		
	Number	Weighted percent	Confidence interval[3]	Number	Weighted percent	Confidence interval[3]
Overall						
Education level:						
Less than high school	975	92.7	90.6–94.3	77	7.3	5.7–9.4
High school graduate, GED, or some college	2,318	95.2	94.1–96.1	108	4.8	3.9–5.9
Post-high school degree	802	96.8	95.1–98.0	23	3.2	2.0–4.9
Poverty level:						
Below threshold	288	89.5	85.1–92.7	29	10.5	7.3–14.9
Above threshold	3,302	95.5	94.6–96.4	150	4.5	3.7–5.4
Marital status:						
Married	2,800	95.8	94.9–96.6	117	4.2	3.5–5.1
Not married, widowed, or separated	1,311	93.4	91.9–94.6	92	6.6	5.4–8.1
Men						
Education level:						
Less than high school	502	92.0	89.7–93.9	41	8.0	6.2–10.3
High school graduate, GED, or some college	1,061	94.3	92.8–95.5	61	5.7	4.5–7.2
Post-high school degree	450	96.0	92.9–97.8	15	4.0	2.2–7.2
Poverty level:						
Below threshold	115	87.4	81.7–91.5	15	12.6	8.5–18.3
Above threshold	1,667	94.9	93.5–96.1	83	5.1	3.9–6.5
Marital status:						
Married	1,565	94.8	93.2–96.0	79	5.2	4.0–6.8
Not married, widowed, or separated	458	92.6	89.9–94.7	37	7.4	5.4–10.1
Women						
Education level:						
Less than high school	473	93.4	90.4–95.5	36	6.6	4.5–9.6
High school graduate, GED, or some college	1,257	96.0	94.8–97.0	47	4.0	3.0–5.2
Post-high school degree	352	98.0	95.5–99.1	8	*2.0	0.9–4.5
Poverty level:						
Below threshold	173	91.0	82.7–95.6	14	*9.0	4.4–17.3
Above threshold	1,635	96.2	95.1–97.0	67	3.8	3.0–4.9
Marital status:						
Married	1,235	97.2	96.3–97.8	38	2.9	2.2–3.7
Not married, widowed, or separated	853	93.8	91.8–95.4	55	6.2	4.6–8.3

* Figure does not meet standards of reliability or precision; relative standard error greater than 30%.

[1]Includes in situ and localized tumors at diagnosis.

[2]Includes tumors that have spread regionally or to distant sites of the body at diagnosis.

[3]Equals 95%.

NOTES: Education level was missing for 52 NHIS participants, income data were missing for 676, marital status was missing for 31, and cancer stage was missing for 19. GED is General Educational Development high school equivalency diploma.

SOURCE: CDC/NCHS, 1986–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System.

Table 6. Percentage of smoking status, self-rated health, and any activity limitation, by sex: 1986–2009 Florida National Health Interview Survey participants aged 18 and over linked with 1981–2010 Florida Cancer Data System before interview (cancer survivor) and not linked (not diagnosed with cancer)

Characteristic	Cancer survivor			Not previously diagnosed with cancer		
	Number	Weighted percent	Confidence interval[1]	Number	Weighted percent	Confidence interval[1]
Overall						
Self-rated health:						
Fair or poor	669	33.2	30.1–36.6	6,756	13.0	12.4–13.7
Excellent, very good, or good	1,261	66.8	63.4–70.0	44,020	87.0	86.3–87.6
Any limitation:						
Yes	726	38.1	35.3–41.0	7,863	15.5	14.6–16.3
No	1,213	61.9	59.0–64.7	43,074	84.5	83.7–85.4
Smoking status[2]:						
Never	305	43.8	38.5–49.4	7,883	55.7	54.1–57.2
Current	77	10.0	7.9–12.5	3,060	22.5	21.4–23.7
Former	312	46.2	41.0–51.4	2,922	21.8	20.8–22.9
Men						
Self-rated health:						
Fair or poor	355	36.0	31.8–40.4	2,713	11.6	10.8–12.4
Excellent, very good, or good	596	64.0	59.6–68.2	20,445	88.4	87.6–89.2
Any limitation:						
Yes	377	40.1	36.4–43.9	3,383	14.4	13.5–15.5
No	577	59.9	56.1–63.6	19,840	85.6	84.5–86.5
Smoking status[2]:						
Never	108	32.1	24.2–41.2	3,050	49.7	47.9–51.6
Current	39	10.5	7.4–14.7	1,537	25.0	23.6–26.5
Former	199	57.4	49.0–65.5	1,578	25.2	23.8–26.8
Women						
Self-rated health:						
Fair or poor	314	30.5	26.8–34.5	4,043	14.3	13.5–15.0
Excellent, very good, or good	665	69.5	65.6–73.2	23,575	85.8	85.0–86.5
Any limitation:						
Yes	349	36.1	32.0–40.5	4,480	16.4	15.5–17.3
No	636	63.9	59.5–68.0	23,234	83.6	82.7–84.5
Smoking status[2]:						
Never	197	56.2	50.3–61.9	4,833	61.0	59.0–63.0
Current	38	9.5	7.3–12.2	1,523	20.2	18.9–21.5
Former	113	34.4	28.7–40.5	1,344	18.8	17.4–20.2

[1]Equals 95%.

[2]Limited to survey years 1997–2009.

SOURCE: CDC/NCHS, 1986–2009 National Health Interview Survey linked with 1981–2010 Florida Cancer Data System.

Vital and Health Statistics Series Descriptions

ACTIVE SERIES

Series 1. **Programs and Collection Procedures**—This type of report describes the data collection programs of the National Center for Health Statistics. Series 1 includes descriptions of the methods used to collect and process the data, definitions, and other material necessary for understanding the data.

Series 2. **Data Evaluation and Methods Research**—This type of report concerns statistical methods and includes analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. Also included are experimental tests of new survey methods, comparisons of U.S. methodologies with those of other countries, and as of 2009, studies of cognition and survey measurement, and final reports of major committees concerning vital and health statistics measurement and methods.

Series 3. **Analytical and Epidemiological Studies**—This type of report presents analytical or interpretive studies based on vital and health statistics. As of 2009, Series 3 also includes studies based on surveys that are not part of continuing data systems of the National Center for Health Statistics and international vital and health statistics reports.

Series 10. **Data From the National Health Interview Survey**—This type of report contains statistics on illness; unintentional injuries; disability; use of hospital, medical, and other health services; and a wide range of special current health topics covering many aspects of health behaviors, health status, and health care utilization. Series 10 is based on data collected in this continuing national household interview survey.

Series 11. **Data From the National Health Examination Survey, the National Health and Nutrition Examination Surveys, and the Hispanic Health and Nutrition Examination Survey**—In this type of report, data from direct examination, testing, and measurement on representative samples of the civilian noninstitutionalized population provide the basis for (1) medically defined total prevalence of specific diseases or conditions in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics, and (2) analyses of trends and relationships among various measurements and between survey periods.

Series 13. **Data From the National Health Care Survey**—This type of report contains statistics on health resources and the public's use of health care resources including ambulatory, hospital, and long-term care services based on data collected directly from health care providers and provider records.

Series 20. **Data on Mortality**—This type of report contains statistics on mortality that are not included in regular, annual, or monthly reports. Special analyses by cause of death, age, other demographic variables, and geographic and trend analyses are included.

Series 21. **Data on Natality, Marriage, and Divorce**—This type of report contains statistics on natality, marriage, and divorce that are not included in regular, annual, or monthly reports. Special analyses by health and demographic variables and geographic and trend analyses are included.

Series 23. **Data From the National Survey of Family Growth**—These reports contain statistics on factors that affect birth rates, including contraception and infertility; factors affecting the formation and dissolution of families, including cohabitation, marriage, divorce, and remarriage; and behavior related to the risk of HIV and other sexually transmitted diseases. These statistics are based on national surveys of women and men of childbearing age.

DISCONTINUED SERIES

Series 4. **Documents and Committee Reports**—These are final reports of major committees concerned with vital and health statistics and documents. The last Series 4 report was published in 2002. As of 2009, this type of report is included in Series 2 or another appropriate series, depending on the report topic.

Series 5. **International Vital and Health Statistics Reports**—This type of report compares U.S. vital and health statistics with those of other countries or presents other international data of relevance to the health statistics system of the United States. The last Series 5 report was published in 2003. As of 2009, this type of report is included in Series 3 or another series, depending on the report topic.

Series 6. **Cognition and Survey Measurement**—This type of report uses methods of cognitive science to design, evaluate, and test survey instruments. The last Series 6 report was published in 1999. As of 2009, this type of report is included in Series 2.

Series 12. **Data From the Institutionalized Population Surveys**—The last Series 12 report was published in 1974. Reports from these surveys are included in Series 13.

Series 14. **Data on Health Resources: Manpower and Facilities**—The last Series 14 report was published in 1989. Reports on health resources are included in Series 13.

Series 15. **Data From Special Surveys**—This type of report contains statistics on health and health-related topics collected in special surveys that are not part of the continuing data systems of the National Center for Health Statistics. The last Series 15 report was published in 2002. As of 2009, reports based on these surveys are included in Series 3.

Series 16. **Compilations of Advance Data From Vital and Health Statistics**—The last Series 16 report was published in 1996. All reports are available online, and so compilations of Advance Data reports are no longer needed.

Series 22. **Data From the National Mortality and Natality Surveys**—The last Series 22 report was published in 1973. Reports from these sample surveys, based on vital records, are published in Series 20 or 21.

Series 24. **Compilations of Data on Natality, Mortality, Marriage, and Divorce**—The last Series 24 report was published in 1996. All reports are available online, and so compilations of reports are no longer needed.

For answers to questions about this report or for a list of reports published in these series, contact:

Information Dissemination Staff
National Center for Health Statistics
Centers for Disease Control and Prevention
3311 Toledo Road, Room 5419
Hyattsville, MD 20782

Tel: 1–800–CDC–INFO (1–800–232–4636)
TTY: 1–888–232–6348
Internet: http://www.cdc.gov/nchs
Online request form: http://www.cdc.gov/cdc-info/requestform.html
For e-mail updates on NCHS publication releases, subscribe online at: http://www.cdc.gov/nchs/govdelivery.htm.

U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES

Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Road, Room 5419
Hyattsville, MD 20782

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

FIRST CLASS MAIL
POSTAGE & FEES PAID
CDC/NCHS
PERMIT NO. G-284

For more NCHS Series reports, visit:
http://www.cdc.gov/nchs/products/series.htm.

DHHS Publication No. 2014–1367, Series 2, No. 167
CS249357