Cancer Epidemiology Supplement
Notes

Introduction

The 1992 NHIS included two special supplements on cancer: Cancer Epidemiology
and Cancer Control. In each NHIS household, a sample adult respondent was
randomly selected to complete one of these 2 cancer supplements (ie., the sample
was split between the two cancer surveys). These supplements are similar to
1987 NHIS Cancer Epidemiology and Cancer Control surveys; however, there were
changes in each questionnaire for 1992. Due to budgetary constraints in 1992,
these two cancer supplements were taken out of the field after the second week
of quarter 3 and thus, did not run for a full year.

Data Layout

The 1992 Cancer Epidemiology data file contains data items on 8 selected topic
areas; Some of these topics are repeated on the Cancer Epidemiology file. The
areas and the file locations for the Cancer Epidemiology file are listed below:

Section :   File Location:

 B- Acculturation   336 - 362
 C- Food Frequency   365 - 1006
 D- Vitamin and Mineral Intake   1007 - 1032
 E- Height and Weight   1033 - 1049
 F- Food Knowledge   1050 - 1083
 G- Cancer Survivorship   1090 - 1168
 H- Smoking Habits   1175 - 1187
 J- Occupational Exposure   1188 - 1207

The 1992 Cancer Epidemiology file is structured as follows:

     a. The NHIS person record from the Core questionnaire
        (locations 1-200)
     b. Weight fields (locations 201-212)
     c. The Cancer Epidemiology data (locations 336-1207)

Sample features and response rate:

The 1992 survey was designed to oversample the Hispanic population.   This was
accomplished by re-contacting the Hispanic participants in the 1991 NHIS and
asking them to participate in the 1992 NHIS. Oversample households can be
identified by using the "processing year" variable (data locations 3-4).

The response rate for the 1992 Cancer Epidemiology was 86.1% and was calculated as follows:

 Household response rate from Core (95.7 percent) multiplied by
 the Cancer Epidemiology response rate (90.0 percent) = 86.1 percent.

Item Non-response

Item non-response was handled in several ways in the 1992 file:
     (1) "Not ascertained" (codes 8, 98, or 998)- impossible
          codes or blanks when there should have been a
          response.

     (2) "refused" (codes 7 or 97)- respondent refused to answer
          the question.

     (3) "Don't know" (codes 9, 99, or 999) - a response of "don't
          know" was given.

For some items, there was not meaningful difference between "not ascertained" and "don't know" or the codes for "not ascertained" were used to indicate an actual response. In these cases, codes of 9 (or 99 or 999) were used. Occasionally, if codes of 8 were used for actual responses, another code was assigned to represented "not ascertained".

Weights and Variances

Since the NHIS uses a multistage sampling design to represent the civilian non-institutionalized population of the United States, weights must be used to make accurate estimates based on data from the NHIS. Two weights are included on the 1991 Cancer Epidemiology File:

a. The Final Basic Weight (locations 207-212) for the Cancer Epidemiology file is calculated for each sample person included in this file. This weight is the functional equivalent of the Annual Final Basic Weight found on the NHIS Person Record of the Core questionnaire. The Cancer Epidemiology Final Basic Weight differs from the weight for the same person on the Core data file in that the Cancer sample was re-weighted to account for non-response, for the selection of one adult per NHIS family, for the use of a split sample, and for the shorter time period that the survey was in the field.

b. The Interim Basic Weight (before age-sex-race adjustment) is also provided since it is required by some software packages for variance estimation for surveys like the NHIS with complex sample designs (locations 201-206).

Those file sections which are the same on both the Cancer Epidemiology and Cancer Control files can be merged. However, the weights should be halved for estimate production.

Estimating numbers of events or conditions

To reduce respondent error, the recall period for questions about some events in the CORE questionnaire is limited to two weeks. These events are: bed days, other restricted activity days, work loss and school loss days, and doctor visits. The two-week variables are found in locations 98-107 and 120-121 of the Cancer Epidemiology data. Estimates of the total number of occurrences of these events in the population can be derived as follows:

> Number of events x 26 (number of two-week periods in a year)
> x Final Basic Weight
>
> = Total number of events occurring in the population during
>   the data year, i.e. 1992.

Example: Number of bed days (loc. 100-101) x 26 x Final Basic Weight (loc. 207-212) = total number of bed days reported for the population in 1992.

The recall period for acute incidence conditions is also two weeks and an annual estimate of the total number of acute incidence conditions is calculated using the same procedures as for two-week events:

> Number of acute incidence conditions x 26 x Final Basic Weight
>
> = Total number of acute incidence conditions occurring in the
>   population during 1992.
>
> NOTE: An acute incidence condition is an acute condition with
> onset during the two weeks preceding the date of interview.

The recall period for information on hospitalizations is 12 months. However, in calculating number of discharges and number of days in the hospital (loc. 132-141), only discharges occurring in the past 6 months are counted. Therefore, the weighted estimates for these events must be calculated as follows:

> Number of discharges x 2 x Final Basic Weight
>
> = Total number of discharges occurring in the population in
>   1992.

Number of days in hospital associated with discharges
        occurring in the past 6 months x 2 x Final Basic Weight

        = Total number of days of hospitalization occurring in the
           population in 1992.

Calculations of rates for events and conditions

The number of events or conditions estimated for the population, as described
above, can be used as the basis for calculating rates of occurrence of these
events (or condition) per person and per 100 persons for the total U.S.
population and for various population subgroups.

NOTE: Only rates can be estimated from these data.   The percent of the
population experiencing a particular type of event during the data year cannot
be estimated. The percent of the population experiencing the event in the
reporting period (eg., two weeks) can be estimated but is generally not
meaningful.

Data on hospital episodes and days, based on a 12-month recall are in locations
122-131.   The Final Basic Weight is used for calculating estimates of these
events in the same way it is used for all other person-based variables. These
variables do permit estimating the percent of the population experiencing a
hospital episode in the past year and the percent of the population having a
specified number of hospital days.

Guidelines for Citation of Data Source

With the goal of mutual benefit, the National Center for Health Statistics
(NCHS) requests that recipients of data files cooperate in certain actions
related to their use.

Any published material derived from the data should acknowledge NCHS as the
original source. The suggested citation to appear at the bottom of all tables
is as follows:

        Source: National Center for Health Statistics (1992)

When cited in a bibliography, the suggested citation should read:

        National Center for Health Statistics (1992). Data File Documentation,
        National Health Interview Survey of Topics Related to Cancer
        Epidemiology, 1992 (machine readable data file and documentation),
        National Center for Health Statistics, Hyattsville, Maryland

The published material should also include a disclaimer that credits any
analyses, interpretations, or conclusions reached to the author (recipient
of the data file) and not to NCHS, which is responsible only for the initial
data. Consumers who wish to publish a technical description of the data
should make a reasonable effort to insure that the description is not
inconsistent with that published by NCHS.