Introduction

The linked birth/infant death data set (linked file) is released in two formats - period data and birth cohort data. This documentation is for the 2004 period linked file. Beginning with 1995 data, the period linked files have formed the basis for all official NCHS linked file statistics. Differences between period and birth cohort data are outlined below.

Period data - The numerator for the 2004 period linked file consists of all infant deaths occurring in 2004 linked to their corresponding birth certificates, whether the birth occurred in 2003 or 2004. The denominator file for this data set is the 2004 natality file, that is, all births occurring in 2004. In addition, NCHS accepted a small number of late-filed birth certificates needed to link to infant deaths. This reduced the number of unlinked records and slightly increased the number of births in the denominator file.

Birth cohort data - The numerator for the 2004 birth cohort linked file consists of deaths to infants born in 2004 whether the death occurred in 2004 or 2005. The denominator file is the 2004 natality file, that is, all births occurring in 2004.

The 2004 period linked birth/infant death data set includes several data files. The first file includes all US infant deaths which occurred in the 2004 data year linked to their corresponding birth certificates, whether the birth occurred in 2003 or in 2004 - referred to as the numerator file. The second file contains information from the death certificate for all US infant death records which could not be linked to their corresponding birth certificates - referred to as the unlinked death file. The third file is the 2004 NCHS natality file for the US (plus late-filed records mentioned above), which is used to provide denominators for rate computations. These same three data files are also available for Puerto Rico, the Virgin Islands, and Guam.

1989 and 2003 Revisions of the U.S. Standard Certificate of Live Birth and Certificate of Death

This data file includes data based on both the 1989 Revision of the U.S. Standard Certificate of Live Birth (unrevised) and the 2003 revision of the U.S. Standard Certificate of Live Birth (revised). The 2003 revision is described in detail elsewhere. (See the 2003 Revision website at: http://www.cdc.gov/nchs/vital_certs_rev.htm). Pennsylvania and Washington implemented the revised certificate in 2003; five additional states (Idaho, Kentucky, New York (excluding New York City), South Carolina, and Tennessee) implemented as of January 1, 2004. Two additional states, Florida and New Hampshire, implemented the revised birth certificate in 2004, but after January 1. Where comparable, revised data are combined with data from the remaining 41 states and the District of Columbia. (Revised data are denoted by "A;" unrevised data are denoted by "S" in the "Rev" column of the documentation.) Where data for the 1989 and 2003 certificate revisions are not comparable (e.g., educational attainment of the mother), unrevised and revised data are given in separate fields in the data file.

This file includes data for ten states (California, Idaho, Michigan, Montana, New Jersey, New York, Oklahoma, South Dakota, Washington, and Wyoming), which implemented the 2003 revision of the U.S. Standard Certificate of Death as of January 1, 2004 or in 2003. Two additional

states, New Hampshire and Connecticut, implemented the revised death certificate in 2004, but after January 1. Data from all other areas are based on the 1989 revision. Most of the variables from the death certificate in this file are comparable despite changes to item wording and format in the 2003 revision. The 2003 revision is described in detail elsewhere. (See the 2003 Revision website at: http://www.cdc.gov/nchs/vital_certs_rev.htm).

Incomplete National Reporting - Using Reporting Flags

As a result of the delayed, phased transition to the 2003 U.S. Standard Certificate of Live Birth, the 2004 linked file includes data for reporting areas that use the 2003 revision of the U.S. Standard Certificate of Live Birth (revised) and data for reporting areas that use the 1989 Standard Certificate of Live Birth (unrevised). Although many data items are comparable across certificate revisions and are available for the entire United States, many items have more limited reporting areas. In addition, the 2004 linked file contains information on infants born in both 2003 and 2004 (see description of the period linked file above). As reporting areas changed between 2003 and 2004, this complicates the analysis. Also, birth data are collected by place of occurrence but are typically tabulated by mother's place of residence. Reporting flags were developed to help the user more readily identify reporting areas for items with less than national reporting for the entirety of 2003 and 2004. Reporting flags are included in the file to assist in accurately excluding records from non-reporting areas when tabulating data by mother's place of residence. Reporting flags are available for most items on the file.

Reporting areas for the 2004 linked file are different from those for the 2004 birth file, as items had to be reported by a state in both 2003 and 2004 to be able to provide complete data. Thus, data for non-comparable items from states that revised in 2004 are excluded from all tabulations. Positions for reporting flags are noted along with each data item in the file layout. Reporting flags must be invoked to generate accurate numbers by residence for items which are not reported by all states. Where applicable, reporting flags are shown in the column "Reporting Flag Position" in the file documentation. Reporting flag codes are "0" (item reported in neither the current or previous year), 1 (item reported in both current and previous year), 2 (item reported in the previous but not in the current year), and 3 (item reported in the current but not in the previous year). When using these data, select reporting flag=1 to get valid and complete data for an item (see SAS code examples below).

Translating "blanks" - In the 2004 linked file, for data items which are not common or comparable across certificate revisions, events to residents of a revised state occurring in an unrevised state, and events to residents in an unrevised state occurring in a revised state, are represented by "blanks." Blanks should be treated as "unknowns" for tabulations.

The correct use of reporting flags and translation of blanks will result in an accurate tally of births and infant deaths for items with incomplete national reporting. For further information please contact us at births@cdc.gov or (301)458-4111.

Example of SAS code using reporting flags (and translating blanks)

An example of SAS code that may be used to incorporate the correct use of reporting flags and the translation of blanks is shown below. This example is for the revised prenatal care item. Prenatal care data based on the revised certificate are not considered comparable to data based on the unrevised certificate, and are shown separately. Accordingly, use of the reporting flag for this item will produce 2004 data for the month prenatal care began for the two revised states which had implemented the revised certificate by January 1st 2003. Data for states which implemented the revised certificates in 2004 are excluded, as part of their linked file births (those born in 2003) were reported on the unrevised certificate.

Sample SAS program

```
01   DATA work;
02        INFILE 'c:link04us.dat' LRECL=1500;
03        INPUT
04             restatus 138
05             precare 245-246
06             f_mpcb 668;
07
08        /*Exclude foreign residents*/
09        IF restatus NE 4;
10        /*Select reporting area*/
11        IF f_mpcb=1;
12        /*Convert blanks to unknown*/
13        IF precare=. THEN precare=99;
14
15   PROC FREQ;
16        TABLE precare;
17   RUN;
```

In this example, "restatus" is used to exclude births to foreign residents (this is standard practice for all NCHS tabulations). Also in this example, blanks are represented by numeric values SAS code = (.). However, for some items in the file, e.g., obstetric procedures, blanks are represented by character values for which the SAS code is empty quotes (' ').

Alternatives to the use of reporting flags - The use of reporting flags provides a relatively quick, accurate way to select records for all areas reporting comparable data for a given item in a particular year. However, should a limited reporting area be needed, specific state(s) of residence may also be selected, or unselected. This approach may be useful, for example, in trend analysis where reporting areas have changed over time. See Table A in the "Natality Technical Appendix" for state-specific information on reporting areas. This approach may also be used to limit the reporting area to only states reporting multiple or single race data (see multiple race section below).

Example of SAS code using state of residence (and translating blanks)

```
01   DATA work;
02        INFILE 'c:link04us.dat' LRECL=1500;
03        INPUT
04             restatus 138
05             xmrstate $ 107-108
```

```
06              precare 245-246;
07
08         /*Exclude foreign residents*/
09         IF restatus NE 4;
10         /*Select reporting area*/
11         IF xmrstate in ('ID','KY','NY','PA','SC','TN','WA');
12         /*Convert blanks to unknown*/
13         IF precare=. THEN precare=99;
14
15    PROC FREQ;
16         TABLE precare;
17    RUN;
```

Single, Multiple, and Bridged Race

In 1997, the Office of Management and Budget (OMB) issued revised standards requiring Federal collection programs to allow respondents to select *one or more race categories.* Beginning with 2003 data, six areas reported multiple race data for all or part of the year: California, Hawaii, Pennsylvania, Utah, and Washington (full year); and Ohio (partial year). Beginning in 2004, six additional areas reported multiple race data for the full year: Idaho, Kentucky, New York State (excluding New York City), South Carolina, Tennessee, and Minnesota. In addition, Florida, and New Hampshire reported multiple race data for part of the year, and Michigan for births at selected facilities only. For the 2004 period linked file, complete data for both 2003 and 2004 births is available from 5 states: California, Hawaii, Pennsylvania, Utah, and Washington.

In 2004, multiple race data was reported on the death certificates of 15 states: California, Hawaii, Idaho, Maine, Michigan, Minnesota, Montana, New Hampshire, New Jersey, New York, Oklahoma, South Dakota, Washington, Wisconsin, and Wyoming. In order to provide uniformity and comparability of the data before all or most of the data are available in the new multiple-race format, it was necessary to ‘‘bridge’’ the responses of those for whom more than one race was reported (multiple race) to one, single race. See the “Technical Notes” of “Births: Final Data for 2004” and “Deaths: Final Data for 2004” which are included on this CD ROM for more information.

Weighting

Beginning with the 1995 linked file, a weight was added to the linked numerator file to correct in part for biases in percent of records linked by major characteristics (see section below on Percent of records linked). The number of infant deaths in the linked file are weighted to equal the sum of the linked plus unlinked infant deaths by age at death and state. The formula for computing the weights is as follows:

$$\frac{\text{number of linked infant deaths + number of unlinked infant deaths}}{\text{number of linked infant deaths}}$$

A separate weight is computed for each state of residence of birth and each age at death category (<7 days, 7-27 days, 28 days-1 year). Thus, weights are 1.0 for states which link all of their infant deaths. The denominator file is not weighted. Weights are not computed for the Puerto Rico, Virgin Islands, and Guam file.

Birthweight

Beginning with the 1995 linked file, an imputation for not-stated birthweight was added to the data set, to reduce potential bias in the computation of birthweight-specific infant mortality rates. Basically, if birthweight is not stated and the period of gestation is known, birthweight is assigned the value from the previous record with the same period of gestation, race, sex, and plurality. Imputed values are flagged. The addition of this imputation has reduced the percent of not-stated responses for birthweight from 4.10 to 0.44 in the numerator file, and from 0.09 to 0.01 in the denominator file, thus greatly reducing (but not eliminating) the potential for underestimation when computing birthweight-specific infant mortality rates.

Comparisons of infant mortality data from the linked file with infant mortality data from the vital statistics mortality file

Although the time periods are the same, numbers of infant deaths and infant mortality rates by characteristics are not always identical between the period linked file and the vital statistics mortality file. Differences in numbers of infant deaths between the two data sources are primarily due to geographic coverage differences. For the vital statistics mortality file, all deaths occurring in the 50 states and the District of Columbia are included regardless of the place of birth of the infant. In contrast, to be included in the linked file, both the birth and death must occur in the 50 states and the District of Columbia. Also, although every effort has been made to design weights that will accurately reflect the distribution of deaths by characteristics, weighting may contribute to small differences in numbers and rates by specific variables between these two data sets. In most cases, differences between numbers of infant deaths and infant mortality rates between the linked file and those computed from the vital statistics mortality file are negligible.

Methodology

The methodology used to create the national file of linked birth and infant death records takes advantage of two existing data sources:

1. State linked files for the identification of linked birth and infant death certificates; and

2. NCHS natality and mortality computerized statistical files, the source of computer records for the two linked certificates.

Virtually all states routinely link infant death certificates to their corresponding birth certificates for legal and statistical purposes. When the birth and death of an infant occur in different states, copies of the records are exchanged by the state of death and state of birth in order to effect a link. In addition, if a third state is identified as the state of residence at the time of birth or death, that state is also sent a copy of the appropriate certificate by the state where the birth or death occurred.

The NCHS natality and mortality files, produced annually, include statistical data from birth and death certificates that are provided to NCHS by states under the Vital Statistics Cooperative Program (VSCP). The data have been coded according to uniform coding specifications, have passed rigid quality control standards, have been edited and reviewed, and are the basis for official U.S. birth and death statistics.

To initiate processing, NCHS obtained matching birth certificate numbers from states for all infant deaths that occurred in their jurisdiction. We used this information to extract final, edited mortality and natality data from the NCHS natality and mortality statistical files. Individual birth and death records were selected from their respective files and linked into a single statistical record, thereby establishing a national linked record file.

After the initial linkage, NCHS returned to the states where the death occurred computer lists of unlinked infant death certificates for follow up linking. If the birth occurred in a state different from the state of death, the state of birth identified on the death certificate was contacted to obtain the linking birth certificate. State additions and corrections were incorporated, and a final, national linked file was produced. Characteristics of the natality and mortality data from which the linked file is constructed are described in detail in the Technical Appendix and Final Reports included in this document.

Characteristics of Unlinked File

For the 2004 linked file 1.1% of all infant death records could not be linked to their corresponding birth certificates. Unlinked records are included in a separate data file in this data set. The unlinked record file uses the same record layout as the numerator file of linked birth and infant death records. However, except as noted below, locations reserved for information from the matching birth certificate are blank since no matching birth certificate could be found for these records. The sex field contains the sex of infant as reported on the death certificate, rather than the sex of infant from the birth certificate, which is not available. The race field contains the race of the decedent as reported on the death certificate rather than the race of mother as reported on the birth certificate as is the case with the linked record file. The race of mother on the birth certificate is generally considered to be more accurate than the race information from the death certificate (see section on Race and Hispanic origin in the Mortality Technical Notes included in this documentation). Also, date of birth as reported on the death certificate is used to generate age at death. This information is used in place of date of birth from the birth certificate, which is not available. Documentation table 6 shows counts of unlinked records by race and age at death for each state of residence. The user is cautioned in using table 6 that the race and residence items are based on information reported on the death certificate, whereas tables 1-5 present data from the linked file in which the race and residence items are based on information reported on the birth certificate.

Percent of Records Linked

The 2004 linked file for the 50 States and D.C. includes 27,612 linked infant death records and 308 unlinked infant death records. The linked file is weighted to the sum of linked plus unlinked

records, thus the total number of weighted infant deaths by place of occurrence is 27,920. While the overall percent linked for infant deaths in the 2004 file is 98.9%, there are differences in percent linked by certain variables. These differences have important implications for how the data is analyzed.

Table 1 below shows the percent of infant deaths linked by state of occurrence of death. While many states link all of their infant deaths, linkage rates for some states are below the national average. Note in particular the percent linked for California (96.9), Massachusetts (97.0), New Jersey (97.3), and Texas (96.7). When a high percentage of deaths remain unlinked, unweighted infant mortality rates computed for these states are underestimated. It is for this reason that weights were added to the file to correct for biases in the data due to poor data linkage for particular states.

Table 1. Percent of infant deaths linked by state of occurrence of death: United States, 2004 linked file

United States	98.9	Nebraska	99.5
Alabama	100.0	Nevada	99.5
Alaska	100.0	New Hampshire	100.0
Arizona	98.7	New Jersey	97.3
Arkansas	99.7	New Mexico	100.0
California	96.9	New York State	97.8
Colorado	100.0	New York City	99.6
Connecticut	100.0	North Carolina	100.0
Delaware	100.0	North Dakota	100.0
District of Columbia	100.0	Ohio	98.5
Florida	99.8	Oklahoma	99.0
Georgia	100.0	Oregon	99.6
Hawaii	100.0	Pennsylvania	99.6
Idaho	99.2	Rhode Island	100.0
Illinois	97.6	South Carolina	100.0
Indiana	99.4	South Dakota	100.0
Iowa	100.0	Tennessee	99.9
Kansas	100.0	Texas	96.7
Kentucky	99.4	Utah	100.0
Louisiana	98.6	Vermont	100.0
Maine	100.0	Virginia	100.0
Maryland	100.0	Washington	99.8
Massachusetts	97.0	West Virginia	100.0
Michigan	100.0	Wisconsin	100.0
Minnesota	100.0	Wyoming	100.0
Mississippi	99.2	Puerto Rico	99.5
Missouri	99.9	Virgin Islands	100.0
Montana	100.0	Guam	100.0

In general, a slightly higher percentage of postneonatal (28 days to under 1 year) than neonatal (less than 28 days) deaths were linked (99.1 and 98.8, respectively.) While the weighting protocol has been designed to correct for possible bias due to variations in match rates by characteristics, no statistical method can correct perfectly for data limitations. Therefore, variations in the percent of records linked should be taken into consideration when comparing infant mortality rates by detailed characteristics.

Confidentiality

To minimize the risk of disclosure of individual or institutional information NCHS public-use data files do not contain the actual day of the birth or the dates of birth of the mother or father. Also, for the linked files, only counties and cities of a population size of 250,000 or more are separately identified.

Geographic classification

Geographic codes in this data set are based on the results of the 2000 census, and only identify areas with a population size of 250,000 or more. Users should refer to the geographic code outline in this document for the list of available areas and codes.

For events to be included in the linked file, both the birth and death must occur inside the 50 states and D.C. in the case of the 50 states and D.C. file; or in Puerto Rico, the Virgin Islands or Guam in the case of the Puerto Rico, Virgin Islands and Guam file. In tabulations of linked data and denominator data events occurring in each of the respective areas to nonresidents are included in tabulations that are by place of occurrence, and excluded from tabulations by place of residence. These exclusions are based on the usual place of residence of the mother. This item is contained in both the denominator file and the birth section of the numerator (linked) file. Nonresidents are identified by a code 4 in location 138 of these files.

Metropolitan statistical areas - Metropolitan statistical areas in this file are based the 1994 Office of Management and Budget (OMB) definition effective July 1, 1994. This definition has been used to define metropolitan statistical areas for natality files since 1994. A listing of the Metropolitan Statistical Areas (MSA's), Primary Metropolitan Statistical Areas (PMSA's), and New England County Metropolitan Areas (NECMA's) is included in this documentation. The 18 Consolidated Metropolitan Statistical Areas (CMSA's) are also included. In June of 2003 the OMB substantially revised the methodology for classifying and coding metropolitan areas in the United States. NCHS plans to convert to the new classification scheme with the release of 2005 natality data.

Demographic and Medical Classification

The documents listed below describe in detail the procedures employed for demographic classification on both the birth and death records and medical classification on death records. These documents, while not absolutely essential to the proper interpretation of the data for a number of general applications, should nevertheless be studied carefully prior to any detailed analysis of demographic or medical data variables. In particular, there are a number of exceptions to the ICD rules in multiple cause-of-death coding which, if not treated properly, may result in

faulty analysis of the data. Volumes 1, 2 and 3 of the ICD-10 may be purchased from the World Health Organization (WHO) Publication Center USA, 49 Sheridan Avenue, Albany, New York, 12210 (http://www.who.int/whosis/icd10/index.html). Many of the instruction manuals listed below are available electronically on the NCHS website at: http://www.cdc.gov/nchs/about/major/dvs/im.htm. In addition, users who do not already have access to these documents may request them from the Chief, Mortality Medical Classification Branch, Division of Vital Statistics, National Center for Health Statistics, P.O. Box 12214, Research Triangle Park, North Carolina 27709. The technical appendices for natality and mortality included in this document also provide information on the source of data, coding procedures, quality of the data, etc.

A. National Center for Health Statistics. Vital statistics, Instructions for Classifying the Underlying Cause-of-Death, 2007. NCHS Instruction Manual, Part 2a. Hyattsville, Maryland: Public Health Service.

B. National Center for Health Statistics. Vital statistics, Instructions for Classifying Multiple Cause-of-Death, 2007. NCHS Instruction Manual, Part 2b. Hyattsville, Maryland: Public Health Service.

C. National Center for Health Statistics. Vital statistics, ICD-10 ACME Decision Tables for Classifying Underlying Causes-of-Death, 2007. NCHS Instruction Manual, Part 2c. Hyattsville, Maryland: Public Health Service.

D. National Center for Health Statistics. Specifications for U.S. Standard Certificate of Birth – 2003 Revision. (replaces NCHS Instruction Manual, Part 3a). Available at: http://www.cdc.gov/nchs/about/major/dvs/im.htm.

E. National Center for Health Statistics. Specifications for U.S. Standard Certificate of Death – 2003 Revision. (replaces NCHS Instruction Manual, Part 4). Available at: http://www.cdc.gov/nchs/about/major/dvs/im.htm.

F. National Center for Health Statistics. Vital statistics, Computer Edits for Natality Data, Effective 1993. NCHS Instruction Manual Part 12. Hyattsville, Maryland: Public Health Service.

G. National Center for Health Statistics. Vital statistics, Computer Edits for Mortality Data, Effective 2007. NCHS Instruction Manual Part 11. Hyattsville, Maryland: Public Health Service.

Also see: http://www.cdc.gov/nchs/vital_certs_rev.htm for the most recent information about revised certificates.

Underlying Cause of Death Data

Mortality statistics by cause of death are compiled from entries on the medical certification portion of the death certificate. The U.S. Standard Certificate of Death is shown in the Mortality Technical

Appendix which is included in this documentation. Causes of death include "all those diseases, morbid conditions or injuries which either resulted in or contributed to death and the circumstances of the accident or violence which produced these injuries". The medical certification of death is divided into two sections. In Part I, the physician is asked to provide the causal chain of morbid conditions that led to death, beginning with the condition most proximate to death on line (a) and working backwards to the initiating condition. The lines (a) through (d) in Part I are connected by the phrase "due to, or as a consequence of." They were designed to encourage the physician to provide the causally related sequence of medical conditions that resulted in death. Thus, the condition on line (a) should be due to the condition on line (b), and the condition on line (b) should be a consequence of the condition on line (c), etc., until the full sequence is described back to the originating or initiating condition. If only one step in the chain of morbid events is recorded, a single entry on line (a) is adequate. Part I of the medical certification is designed to facilitate the selection of the underlying cause of death when two or more causes are recorded on the certificate. The underlying cause of death is defined by the WHO in the ICD-10 as "(a) the disease or injury which initiated the chain of morbid events leading directly to death, or (b) the circumstances of the accident or violence that produced the fatal injury" and is generally considered the most useful cause from a public health standpoint. Part II of the cause-of-death section of the death certificate solicits other conditions that the certifier believed contributed to death, but were not in the causal chain. While some details of the death certificate vary by state, all states use the same general format for medical certification outlined in the U.S. Standard Certificate. The U.S. Standard Certificate, in turn, closely follows the format recommended by the WHO.

If the death certificate is properly completed, the disease or condition listed on the lowest used line in Part I is usually accepted as the underlying cause of death. This is an application of "The General Principle." The General Principle is applied unless it is highly improbable that the condition on the lowest line used could have given rise to all of the diseases or conditions listed above it. In some cases, the sequence of morbid events entered on the death certificate is not specified correctly. A variety of errors may occur in completing the medical certification of death. Common problems include the following: The causal chain may be listed in reverse order; the distinction between Part I and Part II may have been ignored so that the causal sequence in Part I is simply extended unbroken into Part II; or the reported underlying cause is unlikely, in an etiological sense, to have caused the condition listed above it. In addition, sometimes the certifier attributes the death to uninformative causes such as cardiac arrest or pulmonary arrest.

To resolve the problems of incorrect or implausible cause-of-death statements, the WHO designed standardized rules to select an underlying cause of death from the information available on the death certificate that is most informative from a public health perspective. The rules for the Tenth Revision as updated by WHO since the publication of ICD-10 are described in NCHS instruction manual Part 2A. Coding rules beyond the General Principle are invoked if the cause-of-death section is completed incorrectly or if their application can improve the specificity and characterization of the cause of death in a manner consistent with the ICD. The rules are applied in two steps: selection of a tentative underlying cause of death, and modification of the tentative underlying cause in view of the other conditions reported on the certificate in either Part I or Part II. Modification involves several considerations by the medical coder: determining whether conditions in Part II could have given rise to the underlying cause, giving preference to specific terms over

generalized terms, and creating linkages of conditions that are consistent with the terminology of the ICD.

For a given death, the underlying cause is selected from the condition or conditions recorded by the certifier in the cause-of-death section of the death certificate. NCHS is bound by international agreement to make the selection of the underlying cause through the use of the ICD-10 classification structure, and the selection and modification rules contained in this revision of the ICD. These rules are contained in a computer software program called ACME (Automated Classification of Medical Entities). ACME does exactly what a coder would do to select the underlying cause of death. The ACME program has been used for final mortality data since 1968.

The WHO selection rules take into account the certifier's ordering of conditions and their causal relationships to systematically identify the underlying cause of death. The intent of these rules is to improve the usefulness of mortality statistics by giving preference to certain classification categories over others and consolidating two or more conditions on the certificate into a single classification category.

In addition to changes due to the implementation of a new ICD revision, rules for coding a cause of death may occasionally require modification at other times, when evidence suggests that such modifications will improve the quality of cause-of-death data. These changes may affect comparability of data between years for select causes of death. For example, a change was made in a coding rule in 2004 to not code conditions classified to P703-P720, P722-P749, Transitory endocrine and metabolic disorders specified to the fetus and newborn, as the underlying cause of death (there were 20 deaths coded to these categories in the 2003 mortality data). Thus, if this was the only cause listed, the record would be coded to P969, Condition originating in the perinatal period, unspecified; if another cause was also listed, the other cause was preferred.

Multiple Cause of Death Data

The limitations of the underlying cause concept and the need for more comprehensive data suggested the need for coding and tabulating all conditions listed on the death certificate. Coding all listed conditions on the death certificate was designed with two objectives in mind. First, to facilitate studies of the relationships among conditions reported on the death certificate, which require presenting each condition and its location on the death certificate in the exact manner given by the certifier. Secondly, the coding needed to be carried out in a manner by which the underlying cause-of-death could be assigned using the WHO coding rules. Thus, the approach in developing multiple cause data was to provide two fields: 1) entity axis and 2) record axis. For entity axis, NCHS suspends the provisions of the ICD that create linkages between conditions for the purpose of coding each individual condition, or entity, with minimum regard to other conditions present on the death certificate.

Record axis is designed for the generation of person-based multiple cause statistics. Person-based analysis requires that each condition be coded within the context of every other condition on the same death certificate and modified or linked to such conditions as provided by ICD-10. By definition, the entity data cannot meet this requirement since the linkage provisions modify the character and placement of the information originally recorded by the certifier. Essentially, the axis

of the classification has been converted from an entity basis to a record (or person) basis. The record axis codes are assigned in terms of the set of codes that best describe the overall medical certification portion of the death certificate.

This translation is accomplished by a computer system called TRANSAX (Translation of Axis). TRANSAX selectively uses the traditional linkage and modification rules for mortality coding. Underlying cause linkages which simply prefer one code over another for purposes of underlying cause selection are not included. Each entity code on the record is examined and modified or deleted as necessary to create a set of codes that are free of contradictions and are the most precise within the constraints of ICD-10 and medical information on the record. Repetitive codes are deleted. The process may 1) combine two entity axis categories together to a new category thereby eliminating a contradiction or standardizing the data; or 2) eliminate one category in favor of another to promote specificity of the data or resolve contradictions. The following examples from ICD-10 illustrate the effect of this translation:

Case 1: When reported on the same record as separate entities, cirrhosis of liver and alcoholism are coded to K74.6 (Other and unspecified cirrhosis of liver) and F10.2 (Mental and behavioral disorders due to use of alcohol; dependence syndrome), respectively. Tabulation of records with K74.6 would imply that such records had no mention of alcohol. A preferable code would be K70.3 (Alcoholic cirrhosis of liver) in lieu of both K74.6 and F10.2.

Case 2: If “gastric ulcer” and “bleeding gastric ulcer” are reported on a record they are coded to K25.9 (Gastric ulcer, unspecified as acute or chronic, without mention of hemorrhage or perforation) and K25.4 (Gastric ulcer, chronic or unspecified with hemorrhage), respectively. A more concise code is K25.4 which shows both the gastric ulcer and the bleeding.

Entity Axis Codes

The original conditions coded for selection of the underlying cause-of-death are reformatted and edited prior to creating the public-use data file. The following paragraphs describe the format and application of entity axis data.

1. Format. Each entity-axis code is displayed as an overall seven byte code with subcomponents as follows:

1. Line indicator: The first byte represents the line of the death certificate on which the code appears. Six lines (1-6) are allowable with the fourth and fifth denoting one or two written in “due to”s beyond the three lines provided in Part I of the U.S. standard death certificate. Line “6" represents Part II of the death certificate.

2. Position indicator: The next byte indicates the position of the code on the line, i.e., it is the first (1), second (2), third (3) .... eighth (8) code on the line.

3. Cause category: The next four bytes represent the ICD-10 cause code.

4. The last byte is blank.

A maximum of 20 of these seven byte codes are captured on a record for multiple cause purposes. This may consist of a maximum of 8 codes on any given line with up to 20 codes distributed across three or more lines depending on where the subject conditions are located on the certificate. Codes may be omitted from one or more lines, e.g., line 1 with one or more codes, line 2 with no codes, line 3 with one or more codes.

In writing out these codes, they are ordered as follows: line 1 first code, line 1 second code, etc. ----- line 2 first code, line 2 second code, etc. ----- line 3 ---- line 4 ----- line 5 ----- line 6. Any space remaining in the field is left blank. The specifics of locations are contained in the record layout given later in this document.

2. Edit. The original conditions are edited to remove invalid codes, reverify the coding of certain rare causes of death, and assure age/cause and sex/cause compatibility. Detailed information relating to the edit criteria and the sets of cause codes which are valid to underlying cause coding and multiple cause coding are provided in NCHS Instruction Manual Part 11.

3. Entity Axis Applications. The entity axis multiple cause data file is appropriate for analyses that require that each condition be coded as a stand alone entity without linkage to other conditions and/or require information on the placement of such conditions in the death certificate. Within this framework, the entity data are appropriate to examine relationships among conditions and the validity of traditional assumptions in underlying cause selection. Additionally, the entity data provide in certain categories a more detailed code assignment that could be excluded in creating record axis data. Where such detail is needed for a study, the user should use entity data. Finally, the researcher may not wish to be bound by the assumptions used in the axis translation process.

The main limitation of entity axis data is that it does not necessarily reflect the best code for a condition when considered within the context of the medical certification as a whole. As a result, certain entity codes can be misleading or even contradict other codes in the record. For example, category K80.2 is titled “Calculus of gallbladder without cholecystitis.” Within the framework of entity codes this is interpreted to mean that the codable entity itself contained no mention of cholecystitis rather than that cholecystitis was not mentioned anywhere on the record. Tabulation of records with a “K80.2" as a count of persons having Calculus of gallbladder without cholecystitis would therefore be erroneous. This illustrates the fact that under entity coding the ICD-10 titles cannot be taken literally. The user should study the rules for entity coding as they relate to his/her research prior to use of entity data. The user is further cautioned that the inclusion notes in ICD-10 that relate to modifying and combining categories are seldom applicable to entity coding (except where provided NCHS Instruction Manual Part 2b).

In tabulating the entity axis data, one may count codes with an individual code representing the number of times the condition(s) appears in the file. In this kind of tabulation of morbid conditions, the counts among categories may be added together to produce counts for groups of codes. Alternatively, subject to the limitations given above, one may count persons having mention of the

disease represented by a code or codes. In this instance it is not correct to add counts for individual codes to create person counts for groups of codes. Since more than one code in the researcher's interest may appear together on the certificate, totaling must account for higher order interactions among codes. Up to 20 codes may be assigned on a record; therefore, a 20-way interaction is theoretically possible. All totaling must be based on mention of one or more of the categories under investigation.

Record Axis Codes

The following paragraphs describe the format and application of record-axis data. Part 2f of the Instruction Manual Series (ICD-10 TRANSAX Disease Reference Tables for classifying Multiple Causes-of-Death) describes the TRANSAX process for creating record axis data from entity axis data.

1. Format. Each record (or person) axis code is displayed in five bytes. Location information is not relevant. The Code consists of the following components:

1. Cause category: The first four bytes represent the ICD-10 cause code.

2. The last byte is blank.

Again, a maximum of 20 codes are captured on a record for multiple cause purposes. The codes are written in a 100-byte field in ascending code number (5 bytes) order with any unused bytes left blank.

2. Edit. The record axis codes are edited for rare causes and age/cause and sex/cause compatibility. Likewise, individual code validity is checked. The valid code set for record axis coding is the same as that for entity coding.

3. Record Axis Applications. The record axis multiple cause data are the basis for NCHS core multiple cause tabulations. Location of codes is not relevant to this data, and conditions have been linked into the most meaningful categories for the certification. The most immediate consequence for the user is that the codes on the record already represent mention of a disease assignable to that particular ICD-10 category. This is in contrast to the entity code which is assigned each time such a disease is reported on different lines of the certification. Secondly, the linkage implies that within the constraints of ICD-10 the most meaningful code has been assigned. The translation process creates for the user a data file that is edited for contradictions, duplicate codes, and imprecisions. In contrast to entity axis data, record axis data are classified in a manner comparable to underlying cause of death classification thereby facilitating joint analysis of these variables. A potential disadvantage of record axis data is that some detail is sacrificed in a number of the linkages.

The user can take the record axis codes as literally representing the information conveyed in ICD-10 category titles. While knowledge of the rules for combining and linking and coding conditions is useful, it is not a prerequisite to meaningful analysis of the data as long as one is willing to accept the assumptions of the axis translation process. The user is cautioned, however, that due to special

rules in mortality coding, not all linkage notes in ICD-10 are used. (NCHS Instruction Manual Part 2f).

The user should proceed with caution in using record axis data to count conditions as opposed to people with conditions, since linkages have been invoked and duplicate codes have been eliminated. As with entity data, person-based tabulations that combine individual cause categories must take into account the possible interaction of up to 20 codes on a single certificate.

Additional Information

In using the NCHS multiple cause data files, the user is urged to review the information in this document and its references. The instructional material does change from year to year and ICD revision to ICD revision. The user is cautioned that coding of specific ICD-10 categories should be checked in the appropriate instruction manual. What may appear on the surface to be the correct code by ICD-10 may in fact not be correct as given in the instruction manuals.

If on the surface it is not obvious whether entity axis or record axis data should be employed in a given application, detailed examination of NCHS Instruction Manual Part 2f and its attachments will probably provide the necessary information to make a decision. It allows the user to determine the extent of the trade-offs between the two sets of data in terms of specific categories and the assumption of axis translation. In certain situations, a combination of entity and record axis data may be the more appropriate alternative.