Data on causes of death are released by NCHS in a variety of ways including published reports, special tabulations to answer data requests, and public-use data sets. Since the inception of the multiple cause-of-death program in 1968, a public-use data set has been released for each data year. Each set contains a data record for all deaths processed by NCHS, including the underlying and multiple causes of death. For deaths occurring from 1979 to 1998, each record which NDI Plus links with the annual NCHS cause of death data set contains an underlying cause of death field and two sets of multiple cause of death fields which have been coded using the Ninth Revision of the International Classification of Disease, ICD-9. Beginning in 1999, ICD-10 codes are used. The two sets of multiple cause of death fields can contain up 20 cause of death codes. These two sets of multiple cause of death fields are referred to as "entity-axis codes" and "record-axis codes" (which are described below). The NDI Plus records contain the complete level of detail coded by NCHS except where precluded by NCHS and state confidentiality restrictions, lack of data reliability, or when data are not available.

Lists of the ICD-9 and ICD-10 codes can be found in two separate files in the CD containing your NDI results. The complete, detailed ICD-9 and ICD-10 coding manuals can be found at the following URL's:


Underlying Cause of Death Classification

Since 1949 cause-of-death statistics have been based on the underlying cause of death, which is defined as "(a) the disease or injury which initiated the train of events leading directly to death, or (b) the circumstances of the accident or violence which produced the fatal injury".¹

For each death the underlying cause is selected from an array of conditions reported in the medical certification section on the death certificate. This section provides a format for entering the cause of death sequentially. The conditions are translated into medical codes through the use of the classification structure and the selection and modification rules contained in the applicable revision of the International Classification of Diseases (ICD), published by the World Health Organization (WHO). Selection rules provide guidance for systematically identifying the underlying cause of death. Modification rules are intended to improve the usefulness of mortality statistics by giving preference to certain classification categories over others and/or to consolidate two or more conditions on the certificate into one classification category.

As a statistical datum, underlying cause of death is a simple, one-dimensional statistic; it is conceptually easy to understand and a well-accepted measure of mortality. It identifies the initiating cause of death and is therefore most useful to public health officials in developing measures to prevent the onset of the chain of events leading to death. The rules for selecting the underlying cause of death are included in ICD as a means of standardizing classification, which contributes toward comparability and uniformity in mortality medical statistics among countries.

Automated selection of underlying cause of death

Before data year 1968, mortality medical data were based on manual coding of an underlying cause of death for each certificate in accordance with WHO rules. Effective with data year 1968, NCHS converted to computerized coding of the underlying cause and manual coding of all causes (multiple causes) on the death certificate. This system is called Automated Classification of Medical Entities (ACME). The multiple cause codes serve as inputs to the ACME software that employs WHO rules to select the underlying cause. Most states also have implemented ACME and provide multiple cause and underlying cause data to NCHS in electronic form.

The ACME system applies the same rules for selecting the underlying cause as would be applied manually by a nosologist; however, under this system, the computer consistently applies the same criteria, thus eliminating intercoder variation in this step of the process. The ACME computer program requires the coding of all conditions shown on the medical certification. These codes are matched automatically against decision tables that consistently select the underlying cause of death for each record according to the international rules. The decision tables provide the comprehensive relationships among the conditions classified by ICD when applying the rules of selection and modification. The decision tables were developed by NCHS staff on the basis of their experience in coding underlying causes of death under the earlier manual coding system and as a result of periodic independent validations. These tables periodically are updated to reflect additional new information on the relationship among medical conditions. For data year 1988, these tables were amended to incorporate minor changes to the previously mentioned classification for HIV infection (ICD-9 codes *042-*044) that originally had been implemented with data year 1987.

Beginning with data year 1990, another computer system was implemented. This system, called “Mortality Medical Indexing, Classification, and Retrieval” (MICAR), automates the coding of the multiple causes of death. The MICAR system is a major and logical step forward in the evolution of processing mortality data. MICAR takes advantage of the increasing capabilities of electronic data processing to produce information that is more consistently handled than manually processed information. In addition, MICAR provides more detailed information on the conditions reported on death certificates than is available in the ICD classification. Beginning with data year 1993, another computer system was implemented for automating cause-of-death coding. This system, called SuperMICAR, is an enhancement of the MICAR system, which allows for total literal entry of the multiple cause-of-death text as reported by the certifier. This information is automatically coded by the MICAR and ACME computer systems. Records that SuperMICAR is unable to process are coded manually and then processed using ACME.

Multiple Cause Data

The original scheme for coding conditions contained on the death certificate was designed with two objectives in mind. First, to facilitate etiological studies of the relationships among conditions, it was necessary to reflect accurately in coded form each condition and its location on the certification in the exact manner given by the certifier. Secondly, the codification needed to be carried out in a manner by which the underlying cause-of-death could be assigned through computer applications. The approach was to suspend the linkage provisions of the ICD for the purpose of condition coding and code each entity with minimum regard to other conditions present on the certification. This general approach is hereafter called entity coding.

Unfortunately, the set of multiple cause codes produced by entity coding is not conducive to a third objective—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 certificate and modified or linked to such conditions as provided by ICD-9. By definition, the entity data cannot meet this requirement since the linkage provisions distort the character and placement of the information originally recorded by the certifying physician.
Since the two objectives are incompatible, the Division of Vital Statistics creates from the original set of entity codes a new code set called record axis multiple cause data. Essentially, the axis of classification is 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) through selective use of 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 which are free of contradictions and are the most precise within the constraints of ICD-9 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-9 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 5715 (cirrhosis of liver without mention of alcohol) and 303 (alcohol dependence syndrome). Tabulation of records with 5715 would on the surface falsely imply that such records had no mention of alcohol. A preferable codification would be 5712 (alcoholic cirrhosis of liver) in lieu of both 5715 and 303.

Case 2: If "gastric ulcer" and "bleeding gastric ulcer" are reported on a record they are coded to 5319 (gastric ulcer, unspecified as acute or chronic, without mention of hemorrhage or perforation) and 5314 (gastric ulcer, chronic or unspecified, with hemorrhage). A more concise codification would be to code 5314 only since the 5314 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 releasing the data to the public. The following paragraphs describe the format and application of entity axis data.

Entity Axis Code Format

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

- **Line indicator:** The first byte represents the line of the 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 certificate of death. Line "6" represents Part II of the certificate.

- **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.

- **Cause category:** The next four bytes represent the ICD-9 or ICD-10 cause code.

- **Nature of injury flag:** ICD-9 uses the same series of numbers (800-999) to indicate nature of injury (N codes) and external cause codes (E codes). This flag distinguishes between the two with a one (1) representing nature of injury codes and a zero (0) representing all other cause codes. (NOTE: ICD-10 nature of injury codes are S000 –
T983 and do not have a flag of one (1) at the seventh position. The seventh position 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, line 1 third code, etc.
- line 2 first code, line 2 second code, line 2 third code, etc.
- line 3 first code, line 3 second code, line 3 third code, etc.
- line 4 first code, line 4 second code, line 4 third code, etc.
- line 5 first code, line 5 second code, line 5 third code, etc.
- line 6 first code, line 6 second code, line 6 third code, etc.

Fields without codes are left blank. The specifics of locations are contained in the record layout given in Sections II and III of this document.

**Entity Axis Applications**

The entity axis multiple cause data set is appropriate to analyses which 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 certificate. Within this framework, the entity data are appropriate to the examination of etiological relationships among conditions, accuracy of certification reporting, and the validity of traditional assumptions in underlying cause selection. Additionally, the entity data provide in certain categories a more detailed code assignment which is linked out in the creation of record axis data. Where such detail is needed for a study, the user should selectively employ entity data. Finally, the researcher may not wish to be bound by the assumptions used in the axis translation process preferring rather to investigate hypotheses of his/her predilection.

By definition, the main limitation of entity axis data is that an entity code 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, ICD-9 code 5750 is titled "Acute cholecystitis without mention of calculus". Within the framework of entity codes this is interpreted to mean that the codable entity itself contained no mention of calculus rather than that calculus was not mentioned anywhere on the record. Tabulation of records with a "5750" as a count of persons having acute cholecystitis without mention of calculus would therefore be erroneous. This illustrates the fact that under entity coding the ICD-9 (or ICD-10) titles cannot be taken literally. The user must study the rules for entity coding as they relate to his/her research prior to utilization of entity data. The user is further cautioned that the inclusion notes in ICD-9 (or ICD-10) which relate to modifying and combining categories are seldom applicable to entity coding.

In tabulating the entity axis data, one may count codes with the resultant tabulation of an individual code representing the number of times the disease(s) represented by the code appears in the data set. In this kind of tabulation of morbid condition prevalence, 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 record axis multiple cause data are the basis for NCHS core multiple cause tabulations. The following paragraphs describe the format and application of record axis data.

Record Axis Code Format

Each record (or person) axis code is displayed in five (5) bytes. Location information is not relevant. The Code consists of the following components:

- **Cause category**: The first four bytes represent the ICD-9 or ICD-10 cause code.
- **Nature of injury flag**: For ICD-9 codes, the last byte contains a 0 or 1 with the 1 indicating that the cause is a nature of injury category. (NOTE: ICD-10 nature of injury codes are S000 – T983 and do not have a flag in the last byte. 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. 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.

Record Axis Applications

The record axis multiple cause data set is the basis for NCHS core multiple cause tabulations. Location of codes is not relevant to this data set 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-9 or ICD-10 category. This is in contrast to the entity code which is assigned each time such a disease is reported on two different lines of the certification. Secondly, the linkage implies that within the constraints of ICD the most meaningful code has been assigned. The translation process creates for the user a data set which 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. Likewise, they are comparable to general morbidity coding where the linkage provisions of ICD are usually utilized. 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 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 are utilized.

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 which combine individual cause categories must take into account the possible interaction of up to 20 codes on a single certificate.
For purposes of publishing data on causes of death over the period 1979-1998, NCHS groups the data for selected ICD cause of death codes. Consequently each cause of death code is provided with a "recode" to indicate which group or data category a particular cause of death is associated with. There are separate lists of recodes for ICD-9 and ICD-10. These lists appear on two separate files on the CD that contains your NDI results.

Listed below are 3 sets of ICD-9 recodes. These recodes are included in your NDI results in data fields next to the ICD-9 underlying cause of death code.

- **282 recodes** of selected causes of death
- **72 recodes** of selected causes of death
- **61 recodes** of selected causes of infant deaths

IMPORTANT: ICD-10 recodes are used for deaths occurring after 1998 and appear in a separate file entitled ICD-10 Recodes on your NDI results CD. The ICD-10 recodes are positioned in the same data fields used for the ICD 9 recodes; however, the ICD-9 recodes change as follows:

- **282** (ICD-9) recodes become the **358** (ICD-10) recodes.
- **72** (ICD-9) recodes become the **113** (ICD-10) recodes.
- **61** (ICD-9) recodes become the **130** (CD-10) recodes.