A Methodological Study of Quality Control Procedures for Mortality Medical Coding

This report describes the procedures and results of a study designed to investigate certain difficulties in coding mortality medical data and to evaluate the verification procedures used in the medical coding operation.

DHEW Publication No. (PHS) 80-1355

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Office of Health Research, Statistics, and Technology
National Center for Health Statistics
Hyattsville, Md. March 1980

Library of Congress Cataloging in Publication Data

Harris, Kenneth Weldon
A methodological study of quality control procedures for mortality medical coding.

(Vital and health statistics: Series 2, Data evaluation and methods research; no. 81) (DHEW publication; no. (PHS) 79-1355)
Bibliography: p. 30
1. Death–Causes–Classification. 2. Death–Proof and certification. 3. Nosology. 4. Errors, Theory of. I. French, Dwight K., joint author. II. Title. III. Series: United States. National Center for Health Statistics. Vital and health statistics: Series 2, Data evaluation and methods research; no. 81. IV. Series: United States. Dept. of Health, Education, and Welfare. DHEW publication; no. (PHS) 79-1355. [DNLM: 1. Death certificates. 2. Mortality–United States. W2 AN148vb no. 81]
RA409.U45 no. 81 [RA405.A1] 312'.07'23s 79-9173
ISBN 0-8406-0164-6 [616.07]

Vital and Health Statistics—Series 2—No. 81

DHEW Publication No. (PHS) 80-1355
Library of Congress Catalog Card Number 00-000000

PREFACE

Mortality records for the National Center for Health Statistics are processed by their Medical Coding Unit, Data Preparation Branch, at Research Triangle Park, North Carolina. The cooperation of its staff was instrumental in accomplishing much of the work necessary to carry out the study. Ms. Joyce Scott, then Chief of the Vital Records Section, provided valuable assistance in establishing definitional and procedural guidelines. Dependent verification was performed by Ms. Bobbe Doby, Ms. Carolyn Fowler, Ms. Nalda Garner, Ms. Lee Long, Ms. Judy Myers, Ms. Mary Patterson, Ms. Katie Phelps, Ms. Lois Poole, Ms. Barbara Porterfield, and Ms. Mary Weaver. "Expert" coding was provided by Ms. Vicki Long, Ms. Katheryn Lyndon, Ms. June Pierce, Ms. Tanya Pitts, Ms. Julia Raynor, and Ms. Carolyn Watkins. Mr. Jerry Barber of the Computer Center Branch, Division of Operations, furnished the necessary programming; Mr. Guadalupe Gallegos of the Statistical Methods Staff helped to analyze the results.

In addition to internal review, NCHS policy stipulates that methodological reports are to be given a peer review for technical merit and readability by one or more persons who are familiar with the subject matter of the report but who are not involved in its production. Dr. Tore Dalenius, Department of Statistics, University of Stockholm, and Division of Applied Mathematics, Brown University, and Mr. George Minton, retired Mathematical Statistician, U.S. Bureau of the Census, reviewed this report and made constructive suggestions.

CONTENTS

Preface iii

Introduction 1
  Processing of Death Certificates 1
  Verifying Medical Coding 2
  Purpose of the Study 2

Sources and Limitations of the Data 3

Summary of Findings 4

Design of the Study 5
  Determination of Sampling Procedure 5
  Population Coverage 6
  Coding Assignments 7
  Record Match and Assignment of Errors 8

Analysis of Condition Coding 9
  Expert Agreement 9
  Determination of the "True" Production Coding Error Rate 11
  Production Coding Error Rate Measured by Alternative Verification Systems 12
  Agreement in the Three-Way Verification System 15
  Agreement by Quality of Coder 16
  Multiple Sets Coded by the Experts 18
  Characteristics of Multiple Sets of Codes 20
  Index of Confusion 23

Analysis of Underlying Cause-of-Death Coding 25
  Expert Agreement 26
  Underlying Cause-of-Death Coding Error Rate for Production Coding 26
  Underlying Cause-of-Death Coding Agreement in the Three-Way Verification System 27

References 30

Appendixes
  I. ICDA Cause-of-Death Codes Appearing in Three or More Ambiguous Coding Situations and Their Indexes of Confusion 32
  II. Technical Notes 35
  III. U.S. Standard Certificate of Death 40

TEXT FIGURE

1. Number of occurrences of ambiguous cases and descriptions of these occurrences by type of ambiguity 22

LIST OF TEXT TABLES

A. Number of lots in the population and number of sample lots, by production coding error rate 6

B. Percent distribution of sample records in the study and of all 1974 records, by the 17 major underlying cause-of-death groups 7

C. Distribution of coder assignments by coder identification and work lot 8

D. Number of sample records, number of code comparisons, agreement rate between coders 6 and 7, and allocation of their difference rate, by number of sets entered by coders 6 and 7 11

E. Number of records coded, number of condition codes, number of errors, and error rate, by coder identification and assignment 12

F. True production coding error rates and estimated production coding error rates, by verification system 12

G. Number of cases done by coder 7 by agreement classification among the three original coders and percent distribution of coder 7's condition codes by agreement with the three original coders, according to agreement classification among the three original coders 15

H. Number of majority rule cases among the three original coders by number of sets designated by the experts, and percent distribution of these cases by coder 7's agreement with the three original coders, according to the number of sets designated by the experts 16

J. Number of cases coded by the three original coders by number of sets designated by the experts and percent distribution of condition codes by agreement status of the three original coders, according to number of sets designated by the experts 16

K. Number of sample lots, number of 3-way agreements, number of majority rule cases, and number and percent of coder 7 agreements with minority coders, by composition of the three original coders 17

L. Number of errors incorrectly charged, number of errors not charged, and difference between errors not charged and errors incorrectly charged, by composition of groups of good and poor coders 17

M. Number of records with multiple sets, number of differences between the production coder and the sample coder, and number and percent of code differences for which the expert agreed with both coders, by experts entering multiple sets 19

N. Number of records coded and number and percent of records with multiple sets, by coder identification and assignment 19

O. Number of ambiguous cases identified, number and percent of cases for which experts agreed on the condition causing ambiguity, and number and percent of cases for which they agreed on both the condition and the alternative ways of coding it, by expert coder identification 20

P. Number of records with multiple sets by reasons for coding them and expert coder identification 20

Q. Number of records with multiple sets and expected number of multiple set records for specified reasons, by expert coder identification 21

R. Number and percent distribution of ambiguous codes by reason, according to type of ambiguity 22

S. Distribution of the 157 condition codes appearing in three or more multiple sets by frequency of occurrence in all sets and index of confusion 24

T. Number of sample records, number of code comparisons and percent of condition agreements between the experts, and number and percent of underlying cause agreements, by number of sets designated by the experts 27

U. Production coder's condition coding error rate and underlying cause coding error rate for selected cause groupings, by coder 6, coder 7, and the three-way independent verification system 27

W. Number of records and percent distribution of coder 7's underlying cause-of-death codes by agreement with original coders at the each cause level, according to agreement classification among the three original coders 28

Y. Number of records and percent distribution of coder 7's underlying cause-of-death codes by agreement with original coders at the 281-cause level, according to agreement classification among the three original coders 28

Z. Number of records for which multiple sets of underlying cause-of-death codes were entered at the each cause level, numbers of records for which the production coder and sample coder disagreed, and number and percent of code differences for which the expert agreed with both coders, by coder identification 29

AA. Number of records for which multiple sets of underlying cause-of-death codes were entered at the 281-cause level, number of records for which the production coder and sample coder disagreed, and number and percent of code differences for which the expert agreed with both coders, by coder identification 29

SYMBOLS

Data not available----------	- - -
Category not applicable----------	. . .
Quantity zero----------	-
Quantity more than 0 but less than 0.05-----	0.0
Figure does not meet standards of reliability or precision----------	*

A METHODOLOGICAL STUDY OF QUALITY CONTROL PROCEDURES FOR MORTALITY MEDICAL CODING

Kenneth W. Harris, Statistical Methods Staff, and Dwight K. French, formerly with Office of Statistical Research

INTRODUCTION

The National Center for Health Statistics fulfills its mission of providing data on the health of the U.S. population by collecting and publishing information from inventories, registration systems, and *ad hoc* and continuing probability sample surveys. The information from each data system goes through several stages of data processing, finally becoming coded information on computer storage hardware (tapes, disks, or cards) that can be readily recalled. Some data are processed by other organizations under contract. The remainder is processed by the National Center for Health Statistics (NCHS) computer and data preparation facility located in Research Triangle Park (RTP), North Carolina.

Quality control procedures of various types have been instituted for each data system at each step from data collection through dissemination. This report describes a methodological study of one of the quality control systems used at Research Triangle Park, that is, verification of the medical codes assigned to the conditions reported on death certificates.

Processing of Death Certificates

One of the most difficult data processing jobs handled at the data preparation facility is the coding of medical conditions listed as causes of death in the annual file of almost 2 million death records. The standard death certificate, shown in an appendix, contains a large number of demographic items, most of which are easy to code. The medical portion of the certificate (item 25) consists of three lines on which the attending physician or other official is instructed to enter the sequence of medical conditions that led to death and a fourth line for listing other significant conditions. Completed death certificates are collected by the States, reproduced on microfilm, and sent for processing to the data preparation facility at Research Triangle Park.[a] The records are grouped into work lots of 2,000-4,000 certificates. Each lot is identified by a number and is given to a nosologist (medical coder), who assigns numerical codes to the medical conditions according to the *Eighth Revision International Classification of Diseases, Adapted for Use in the United States* (ICDA).[1] The coder separates the conditions listed on lines *a*, *b*, and *c* of item 25 of the death certificate by a slash (/) to indicate the line change, and by an asterisk (*) to separate the conditions in Part II from the previous entries. Although the instructions above line *a* indicate that no more than one condition should be listed on lines *a*, *b*, and *c*, physicians often deem it

[a] All States process their certificates independently, and as part of its Cooperative Health Statistics System, NCHS purchases and uses the codes entered by seven States. The number of States submitting usable codes is expected to increase in the future; however, it will be many years before State processing is accepted for a large proportion of U.S. deaths.

necessary to enter more than one code on a line to properly describe the sequence of conditions that led to death. An example of the format for the medical codes of a death certificate is:[b]

a / *b* / *c* * Part II
782.9/411.9 410.9/412.3*593.1 441.1

The nosologist types the codes on a special sheet that can be read by an optical character recognition (OCR) scanner, which transfers the codes to magnetic tape for temporary storage and later manipulation. The condition codes serve as input to a computer program that assigns a code for one condition, called the "underlying cause of death," to represent all the conditions on a certificate. The program can assign an underlying cause for about 95 percent of all certificates. The remainder, certificates that involve conditions that occur infrequently and are difficult to code, are rejected by the computer and must be coded by nosologists who have special training in underlying cause coding.

Verifying Medical Coding

The assignment of underlying cause codes is not subject to ongoing verification because the process has been tested and is considered sufficiently accurate; a formal verification system is neither cost nor quality effective. However, the original condition codes, assigned by a staff of about 26 coders with varying degrees of proficiency, are subject to three-way independent sample verification.

After the original production coder completes a work lot, two other coders independently code a 10-percent systematic sample of records from the lot. These two new sets of codes are matched by computer with the production coder's work, line by line and position by position. If any two coders enter the same code in the same position on the same line of a record, that code is placed into a "correct" or "preferred" set of codes for the record. If the third coder enters the code in any position on the same line, it matches; otherwise, the coder is charged with an error. If two coders (no code by third coder) or all three coders enter nonmatching codes in the same position on the same line, an "X" is placed in the preferred set position, indicating that an unknown code belongs in that position. All three coders are charged with an error even if a code is matched by one or both of the other coders elsewhere on the line. A code entered by only one coder is not included in the preferred set, and the coder is charged with error.

After the matching procedure is completed for an entire lot, an error rate is computed for each coder by dividing her number of errors by the sum of the number of preferred codes for the sample records. The error rates are used to evaluate employee performance and to determine whether the work lot, as coded by the production coder, is acceptable for underlying cause processing. If the production coder's error rate is 5 percent or less, the work lot is acceptable as coded; otherwise, the entire lot is recoded by a fourth coder and rematched against the work of the two original sample coders.

Purpose of the Study

Three-way independent verification of mortality medical coding was instituted for the 1970 data year because it was considered a more reliable method of measuring coding error than the two-way dependent system previously used. Studies on other types of data have shown that independent verification yields more accurate estimates of the amount of error in the data than dependent verification does; a dependent verifier tends to be biased toward the work of the original coder.[2-8] However, no thorough study has ever been conducted to test the accuracy of mortality medical coding error rates based on the three-way system. The accuracy of such error rates is questionable for the following reasons:

1. Poor handwriting, incorrect or confusing placement of conditions on the death certificate, or poor quality microfilm may make it impossible to determine a unique correct code for some conditions.

[b]The actual coding is done without using decimals. They are inserted here for consistency with other references to ICDA codes in the text.

2. Although a certificate is legible and filled out properly, the coding instructions may be so vague as to allow two or more acceptable codes for a particular condition. The appropriateness of three-way independent verification is based on the assumption that a medical condition leads to only one code; thus when 2 or 3 out of 3 coders with comparable ability arrive at the same code independently, there is a high probability that the majority code is correct. If this assumption is invalid, a nosologist with an acceptable code can be charged with an error when the other two coders match on a different acceptable code.

3. It is possible that a coder with an acceptable code will occasionally be charged with an error when the other two coders match on an unacceptable code.

4. When two coders (no code by third coder) or all three coders enter nonmatching codes, all three coders are charged with an error, although it is likely that at least one coder entered an acceptable code.

The primary purpose of the study was to measure the accuracy of error rates produced by three-way independent verification and to compare them with error rates produced by two other commonly used methods of verification: two-way dependent verification and two-way independent coding with adjudication of differences. In addition, the study was designed to provide data in other areas related to the quality of medical coding. These areas are discussed in the analysis sections of the report. The estimation procedures, the method used to estimate sampling errors, and the sources of nonsampling error and bias in this study are discussed in appendix II.

SOURCES AND LIMITATIONS OF THE DATA

The data for the study were produced from the 10-percent quality control sample of death records in 30 of the 472 work lots processed through the mortality medical coding unit between July 1974 and March 1975. Each coder who worked in the unit during that 9-month period was represented at least once in the 30 sample lots as a production coder or verifier, and the coders were generally represented in proportion to the amount of work they did on the 472 lots. The 30 sample lots were selected by a single systematic procedure after the 472 lots had been sorted into 10 production-coding error-rate strata and the lots in each stratum had been ordered randomly.

Six individual coders provided condition-code input for each sample record. The first three coders were the production coder and the two sample coders who performed the three-way independent procedure during the original processing. Their codes were used exactly as they had been entered during the data year. Three additional coders were assigned to each lot for the purpose of the study. The work of one coder, who was given access to the work of the production coder, represented a dependent verification assignment. The work of the other two coders, who were selected from a small group of "experts," served as measures of "truth" for the study. One expert was given access to the work of the production coder and one sample coder, thus serving as a kind of dependent adjudicator of a two-way independent verification system, and the other expert coded independently. The instructions for the experts differed from the instructions for the other coders. The expert coders were allowed to enter more than one set of codes if they felt that one or more conditions on the record could be correctly coded in more than one way. When an expert entered more than one set of codes, she was instructed to identify a "set of first choice," that is, the set of codes she preferred, or the set she would have entered had she been forced to choose one.

The three extra coding assignments for the study were subject to two major sources of bias that complicate the interpretation of certain statistics presented here. First, there were procedural differences between the dependent verification procedure used in the study and the comparable procedure used in previous data years. The dependent verifiers for the study were aware that they were working on a special project and that their work would be reviewed. In previous data years, dependent verifiers knew that there would be virtually no review of their work. In the study, the dependent verifier's set of codes was matched against the production coder's work, and the production coder was charged with errors when differences occurred. In previous data years, a dependent verifier could change the codes of the production coder in ambiguous situations, but not charge her with errors for codes that the verifier felt were acceptable, if not preferable, alternatives. This difference was not discovered until after the fieldwork for the study had been completed.

A second source of bias stems from the fact that expert coders, whose work was used to measure "truth" for the study, could be expected to make some errors, even though they were the best coders available. Although adjustments were made to certain estimates in this report to compensate for errors by the experts, there is no measure of the accuracy of the adjustments, just as there is no totally correct measure of truth for the coding of the study records.

SUMMARY OF FINDINGS

Since most of the analytical data given here are based on coding by the two experts, it was important to know how often the experts agreed with each other. Rules were established for comparing the work of the two experts when both entered one set of codes and when either or both entered multiple sets of codes. The overall rate of agreement between the dependent and independent experts for condition coding was 97.77 percent. Error rates of 0.56 percent and 1.26 percent were charged to the dependent and independent experts, respectively.

The best measure of the production coder's error rate was 4.27 percent, based on the work of the independent verifier and adjusted for her errors. The dependent expert, acting as an adjudicator, estimated the production coder's error rate at 3.05 percent, and the dependent verifier estimated the rate at 3.67 percent; the three-way independent verification system had previously estimated the production coder's error rate as 3.75 percent. The error rate estimates based on dependent verification and three-way independent verification are similar; however, the unexpectedly high rate obtained with dependent verification was a result of the special study conditions already mentioned. It is doubtful that the conditions that made dependent verification competitive with independent verification in the study can be achieved in day-to-day coding without sacrificing its cost benefits.

When a majority of the three original coders agreed on a code (AAA and AAB cases), the independent expert agreed with A, the majority code, in 98 percent of all cases. However, in AAB cases, the independent expert agreed with A in only about 80 percent of the cases and with B in 23 percent of the cases. These percentages include the cases in which the expert agreed with both A and B (about 6 percent of all cases).

Records given multiple sets of codes by one or both experts are apparently more difficult to code than those records not given multiple sets. For records with no multiple sets of codes, AAA cases comprised 93 percent of majority rule cases; for records with multiple sets, the percentage dropped to 75 percent. For multiple set records, the independent expert agreed with code A in AAB cases in 75 percent of the cases, but the percentage of agreement with B increased to 41 percent, including 20 percent for which the expert agreed with both codes. Also, the average number of condition codes in records with multiple sets of codes was 4.4, almost 50 percent higher than the 3.0 codes per single set record.

An estimator of an "index of confusion" was used to investigate the conditions that created ambiguous coding situations which caused the expert coders to enter multiple sets of codes. The index of confusion for ICDA code c was defined as

$$I_c = \frac{M_c}{M_c + U_c}$$

where

M_c = the number of ambiguous coding situations in which code c appeared in one or more of the alternative solutions.

U_c = the number of occurrences of code c as the unique coding solution for a medical condition in either a single or a multiple set (i.e., code c was counted only once if it appeared as a unique solution on each line of a multiple set).

An index was computed for the 157 codes for which M_c was 3 or more. Fifty-three of these codes had indexes of at least .25, of which 32 had 10 or fewer total occurrences in the denominator, and 51 had 25 or fewer total occurrences. A logical explanation for the high indexes associated with these codes is the relative unfamiliarity of the coders with the causes of death they represent. However, the fact that these causes occur infrequently means that a few miscoded cases can cause a large percent change in the stated frequency of occurrence.

Differences between coders on codes for individual conditions on a record may or may not cause differences between them for the summary condition, namely, the underlying cause of death. The two major reasons for the stability of the underlying cause codes are the following:

1. One or more of the conditions on a record may not affect the underlying cause.
2. Underlying cause codes are virtually always collapsed into more general categories than the detailed four-digit classification (the each cause level) for purposes of analysis. A change in a condition code, especially in the third or fourth digit, may affect the underlying cause at the each cause level but does not remove it from a more general category.

The condition code agreement rate of 97.77 percent between the dependent and independent experts increased to 98.85 percent for underlying cause code agreement at the each cause level and to 99.22 percent at the 281-summary-cause level. The independent expert's estimate of the condition code error rate of the production coder was 4.27 percent. The corresponding error rates for underlying cause codes at the each cause and 281-cause levels were 2.21 percent and 1.78 percent, respectively. During condition coding the independent expert agreed with the majority code in 96.78 percent of AAA and AAB cases. The underlying cause code agreement rates for the majority rule cases were 98.31 percent at the each cause level and 98.62 percent at the 281-cause level.

DESIGN OF THE STUDY

Determination of Sampling Procedure

The goals of the study, along with limitations on coding time, programmer time, and available funds, imposed the following constraints on the design:

1. It was necessary to conduct the study using death certificates that had already been processed through the three-way independent verification system.
2. The number of work lots represented in the sample had to be sufficiently large to include both production and sample verification work for most of the coders on the staff.
3. In order to compare the accuracy of error rates produced by different verification systems, some measure of the "true" amount of error in the data was needed. "Truth" would be determined by having a small group of experts code the data.
4. The number of records in the sample had to be small enough so that the experts would not be overburdened. Each expert was limited to a maximum of 2 weeks of coding time to finish the assignment.

The sampling frame for the study consisted of the 472 work lots of 1974 data that passed through the three-way verification system from July 1974 through March 1975. From this frame, a first-stage sample of work lots was selected. The death certificates for the study represented a second stage of sampling within the sample lots. This two-stage procedure avoided a costly, time-consuming search for sample records throughout all 472 lots. An additional consideration was whether to use the entire 10-percent verification sample of records from each sample lot or to take a larger sample of lots and subsample from the original quality control sample. The first alternative was chosen because lot variation in error rates was not very large (97 percent of the 472 lots had estimated production coding error rates of between 1 and 6 percent, and about 64 percent had error rates of between 2 and 4 percent) (table A). Also, there was concern that subsampling from the quality control sample would increase the amount of time needed to find the sample records. In addition, subsampling might cause coders to make more errors in selecting the sample certificates than they would if they followed the usual 10-percent systematic pattern using the terminal digit of the certificate number.

Combined with the original constraints, the decision to use the entire quality control sample from each lot placed rather rigid restrictions on the number of sample lots. In order to include production coding of most of the 26 women who had major responsibility for coding 1974 data, the number of sample lots had to be at least as large as the number of coders. On the other hand, the small number of experts available and the 2-week limitation on their coding time placed a firm ceiling on the number of lots.

The panel of experts consisted of six nosologists: four supervisors and two coders in the unit who had produced exceptionally high quality work for several years. Each sample lot was to be coded by two experts; their instructions are discussed later. Since the average work lot contained about 3,000 records, the average quality control sample contained about 300 records. The expected production coding rate for verifiers in 1974 was 425 records per day, but the experts had more complicated assignments than sample coders had and were urged to be especially careful. Therefore, each expert could only be expected to finish 1 assignment per day, or a total of 10 assignments during the 2 weeks allotted. The total number of expert assignments was therefore limited to 60; thus the sample could consist of no more than 30 lots. This number was set as the first-stage sample size.

Before the sample lots were selected, the 472 work lots were sorted into the 10 production coding error rate strata shown in table A. The lots in each stratum were ordered randomly, the strata were ordered from smallest error rate upward, and the lots were temporarily renumbered from 001 to 472. A sample of 30 lots was selected using a randomly chosen starting number of 008 and a skip interval of 16.

Table A. Number of lots in the population and number of sample lots, by production coding error rate

Production coding error rate	Number of lots in population	Number of sample lots
Total	472	30
0.00-0.99	7	-
1.00-1.99	54	4
2.00-2.49	59	4
2.50-2.99	81	5
3.00-3.49	87	5
3.50-3.99	73	5
4.00-4.49	37	2
4.50-4.99	28	2
5.00-5.99	31	2
6.00 and over	15	1

Population Coverage

The method of selection of sample lots insured that the lots would be representative of the frame with respect to coding difficulty (and, for that matter, representative of all mortality medical coding, since the content of death certificates changes relatively little from year to year). In addition, the sample lots encompassed a good cross-section of the coding staff. Of the 26 nosologists who had major responsibility for coding during the 1974 data year, 17 were represented one or more times as production coders, and all 26 were represented at least once as a production coder or sample coder. In general, the coders who had completed the most assignments during the data year were the ones represented most often in the experiment, and, conversely, the coders whose production work was not represented had done very little production coding during the 9-month period. Finally, the distribution of underlying cause codes generated from the production coders' work for the sample records closely follows the distribution

Table B. Percent distribution of sample records in the study and of all 1974 records, by the 17 major underlying cause-of-death groups

Cause-of-death group and ICDA code[1]			1974 records	
			All	Sample
			Percent distribution	
I.	Infective and parasitic diseases	000-136	0.81	0.75
II.	Neoplasms	140-239	18.90	18.97
III.	Endocrine, nutritional and metabolic diseases	240-279	2.36	2.37
IV.	Diseases of blood and blood-forming organs	280-289	0.28	0.29
V.	Mental disorders	290-315	0.50	0.50
VI.	Diseases of the nervous system and sense organs	320-389	0.90	0.85
VII.	Diseases of the circulatory system	390-458	53.17	53.42
VIII.	Diseases of the respiratory system	460-519	5.65	5.45
IX.	Diseases of the digestive system	520-577	3.78	3.94
X.	Diseases of the genitourinary system	580-629	1.35	1.38
XI.	Complications of pregnancy, childbirth, and the puerperium	631-650	0.02	0.04
XII.	Diseases of the skin and subcutaneous tissue	680-709	0.11	0.06
XIII.	Diseases of the musculoskeletal system and connective tissue	710-734	0.26	0.23
XIV.	Congenital anomalies	741-759	0.70	0.70
XV.	Certain causes of mortality in early infancy	760-779	1.49	1.38
XVI.	Symptoms and ill-defined conditions	780-796	1.61	1.62
XVII.	Accidents, poisonings, and violence	E800-E999	8.11	8.05

[1]Based on the *International Classification of Diseases, Adapted for Use in the United States.*

of underlying cause codes for all 1974 death records (see table B).

Coding Assignments

Within each sample lot, the medical codes of six coders for the 10-percent quality control sample served as input data for the study. The first three coders were the original production coder (coder 1) and the two independent sample coders (coders 2 and 3). Because these numerical designations correspond to the coder numbers on the lot-by-lot quality control reports issued by the Data Preparation Branch (DPB), it was possible to distinguish coders 2 and 3 for each lot. The original production coder was always coder 1, even if her work was rejected and the lot was recoded (the recoder is coder 4).

An additional coder (coder 5) was assigned to code the sample records, with access to the work of the production coder. This was intended to correspond to the dependent verification system that was used by the DPB during the 1973 data year to speed up processing. Because dependent verification assignments for 1973 data were given to the best coder available at the time the work was assigned, most of the verification was handled by coders with relatively low error rates. In order to follow this system as closely as possible, coder-5 assignments for the study were allocated to the 10 coders with the lowest average error rates for the 1974 data year, after excluding the two best coders, who were previously designated as experts. Each of the 10 coders was given three randomly selected lots that she had not worked on as coder 1, 2, 3, or 4. The distribution of coder-5 assignments for the 30 sample lots is shown in table C.

Another coder (coder 6) was assigned to code the sample records, with access to the work of the production coder and one of the sample coders. Her role corresponded to that of a dependent adjudicator in a two-way independent verification system, except that she coded all records regardless of whether the two original coders disagreed. If the work lot number (table C) was odd, coder 6 was given access to the work of coder 2; if the number was even, she had access to the work of coder 3.

The coding instructions for coder 6 were somewhat more complicated than the instructions for the previous coder assignments. Whereas the other coders entered only one set of codes to represent the causes of death listed on a cer-

Table C. Distribution of coder assignments by coder identification and work lot

Work lot number	Identification number of: Coder 5	Coder 6	Coder 7
10	D5	E6	E1
24	D4	E3	E1
32	D9	E2	E1
45	D9	E1	E4
76	D1	E4	E1
81	D1	E6	E4
87	D4	E1	E5
121	D2	E6	E3
149	D1	E5	E6
151	D4	E6	E3
166	D5	E4	E1
191	D2	E5	E2
197	D3	E2	E6
202	D10	E4	E2
215	D5	E2	E6
224	D10	E5	E6
238	D8	E3	E5
275	D6	E1	E3
282	D8	E3	E2
299	D10	E2	E5
336	D9	E5	E2
370	D2	E4	E3
376	D8	E2	E5
611	D6	E6	E4
630	D3	E4	E2
633	D7	E5	E4
640	D6	E1	E5
654	D7	E1	E3
684	D3	E3	E6
706	D7	E3	E4

tificate, coder 6 was instructed to list all sets of codes she considered acceptable. If she thought that each condition on a certificate had a single correct code, she entered one set of codes. However, if she thought that one or more conditions could be coded in more than one way, she entered all possible acceptable sets of codes, changing only the questionable code(s).

Whenever coder 6 entered more than one set of codes, she listed the sets according to the following rules:

1. If two or more sets were acceptable, but one was clearly *preferable*, the preferred set was listed first and "P" was written next to that set.
2. If two or more sets were equally acceptable, first the set the coder would choose if she had to *decide* on one set was listed, and "D" was written next to that set.

The set of codes marked with "P" or "D" is referred to as the "set of first choice."

Coder 6 also entered another letter code to the right of the "P" or "D" for each set of first choice. This letter represented her reason for thinking there was more than one acceptable set of codes. The reasons were coded as follows:

A. Poor handwriting.

B. Formatting problem (difficulty in determining which conditions belong on each line of the medical section of the death certificate).

C. Poor quality microfilm.

D. Vagueness in the 1974 coding instructions, rectified in the 1975 instructions.

E. Vagueness in the 1974 coding instructions, not rectified in the 1975 instructions.

F. Other or multiple reasons.

A final coder (coder 7) was assigned to code the sample records without having access to the work of any of the other coders. Her coding instructions were exactly the same as those for coder 6.

The coder-6 and coder-7 assignments for the 30 sample lots were randomly distributed among the six experts so that each had 10 assignments: 5 as coder 6 and 5 as coder 7 (table C). No expert was given coder-6 and coder-7 assignments for the same lot, and no expert was given an assignment in a lot she had previously coded.

Record Match and Assignment of Errors

After coding assignments 5, 6, and 7 were completed for all sample lots, the codes were run through a computer program along with the work of the three original coders. The program produced a printout of the codes of each of the six coders for every record in the sample. An example of the printout is shown.

Record Number	*State Identification*	*Lot Number*	*Coder Number*	*Codes*
006661	50	045	1	429.9/412.3
			2	429.9/412.3
			3	429.9/412.3*795.2
			5	429.9/412.3*795.2
			6	429.9/412.3
			7	429.9/412.3*795.2 PE
			7	429.9/412.3

The program matched the work of the three original coders (1, 2, 3) with that of each of the coders used in the study (5, 6, 7), assigning errors to the original coders when their entries did not match those of coders 5, 6, or 7. When coder 6 and/or coder 7 entered more than one set of codes, the program observed the following rules in tallying codes and errors:

1. The number of code comparisons between each original coder and the expert (the denominator for computing the original coder's error rate) was taken from the expert's set of first choice for purposes of consistency.

However,

2. The number of errors charged to each original coder was taken from the expert's set that minimized the number of errors. In the example, coder 1 would be charged with no errors by coder 7 because her set matched coder 7's second set, and she would be given credit for three correct comparisons, the number of codes in coder 7's set of first choice.

A part of the program that had to be modified was the assignment of errors to the original coders when their work matched that of coders 5 and/or 6 and/or 7 by line but not position. As noted earlier, a code was listed as "preferred" in the three-way verification system when any two coders placed that code in the same position of the same line. The third coder was credited with a correct entry if she placed the code anywhere on the same line. For example:

Coder 1:	203.0	207.5/123.4
2:	203.0	207.5/123.4
3:	207.5	203.0/123.4

Although the work of coder 3 did not match that of coders 1 and 2 by position on line 1, it did match by line. No one was charged with an error.

Originally the program charged errors when codes did not match by position. However, entries by coders 5, 6, and 7 were, by definition, preferred codes; thus no errors should have been charged in cases of line agreement. The program modification substantially reduced the number of errors charged to the original coders and thus reduced the estimated error rates.

ANALYSIS OF CONDITION CODING

Expert Agreement

Almost any sample verification system will provide an estimate of the amount of error in the original process, but the reliability of the estimates may be questionable. In order to compare the merits of the competing verification systems, it was first necessary to determine the "true" value of the statistic being estimated, that is, the error rate of the production coder (coder 1). Since coding assignments 6 and 7 were carried out by the best nosologists available, the work of coders 6 and 7 provided competing "true" error rates for the sample of 30 lots. Obviously, two separate measures of truth would not be exactly the same because of occasional errors, the vagueness of some cases, and the differences between dependent and independent coding. Hopefully, however, the two versions of "truth" would be very close.

The degree of consistency between dependent and independent expert coding was measured as an agreement rate. The numerator and denominator of the rate were determined by comparing the work of the two coders record by record. Each record yielded a number of agreements (again, line agreement on a code was sufficient) and difference cases. When both coders entered only one set of codes, one comparison determined the number of agreements and disagreements. When coder 6 and/or coder 7 entered more than one set of codes, the following rules were established to determine which sets should be compared.

1. The comparison that minimized the number of differences between the two experts was selected. If two or more comparisons yielded the same number of differences,
2. One of them was selected to maximize the number of agreements. If two or more comparisons were still indistinguishable,
3. One of them was selected using the following priority order:
   a. The comparison involving the set of first choice of both coders.
   b. A comparison involving coder 7's set of first choice.
   c. A comparison involving coder 6's set of first choice.
   d. Any other comparison.

Example:

Coder	*Set*	*Entry*	
6	(1)	122.3 234.5/150.0 175.0/200.1*500.2	P
6	(2)	122.3 234.5/150.1 175.0/200.1	
6	(3)	122.3/150.1 175.1/200.1*500.2	
7	(1)	122.3 234.6/150.0 175.0/200.1*500.2	P
7	(2)	122.3 234.5/150.1 175.0/200.2	
7	(3)	122.3 234.5/150.1 175.1/200.1*500.2	

Rule 1

6(1) vs 7(1): 1 difference
6(1) vs 7(2): 3
6(1) vs 7(3): 2
6(2) vs 7(1): 3
6(2) vs 7(2): 1
6(2) vs 7(3): 2
6(3) vs 7(1): 3
6(3) vs 7(2): 4
6(3) vs 7(3): 1

Rule 2

6(1) vs 7(1): 5 agreements
6(2) vs 7(2): 4 agreements
6(3) vs 7(3): 5 agreements

Rule 3

Select comparison 6(1) vs 7(1)

The agreement rate between coders 6 and 7 for record j in lot i is given by

$$R_{ij} = \frac{\text{number of agreements}}{\text{number of agreements + number of disagreements}}$$

$$= \frac{A_{ij}}{A_{ij} + D_{ij}}. \qquad (1)$$

The agreement rate for sample lot i is given by

$$R_i = \frac{\sum_j A_{ij}}{\sum_j A_{ij} + \sum_j D_{ij}} \qquad (2)$$

and the overall agreement rate for the study is

$$R = \frac{\sum_i \sum_j A_{ij}}{\sum_i \sum_j A_{ij} + \sum_i \sum_j D_{ij}}. \qquad (3)$$

Any other agreement rate can be computed by adding agreements and differences over all applicable sample records and calculating a quotient similar to equation (2) or (3).

As can be seen in table D, coders 6 and 7 coded 8,973 records, resulting in 27,752 code comparisons. Under the rules outlined above, the agreement rate for the study was 97.77 percent. Since the production coding error rate for the best medical coders on the regular file is about 2 percent, the 98-percent agreement here (which is lowered by errors from both coders) is probably as high as could be expected.

Whenever a difference between coders 6 and 7 occurred, one expert was credited with a correct response if at least 3 of the other 4 coders (1, 2, 3, and 5) matched her code. The nonmatching expert was charged with an error. If at least three coders agreed on a code different from the nonmatching codes of the experts, both experts were charged with an error. Under any other circumstances, the difference was considered unresolved. This method of assigning errors was unquestionably biased in

Table D. Number of sample records, number of code comparisons, agreement rate between coders 6 and 7, and allocation of their difference rate, by number of sets entered by coders 6 and 7

Sets entered by coders 6 and 7	Sample records	Code comparisons	Agreement rate	Difference rate	Error charged to: Coder 6	Coder 7	Both coders 6 and 7	Unresolved
	Number		Percent					
All records	8,973	27,752	97.77	2.23	0.47	1.17	0.09	0.49
Both coders entered one set	8,289	24,743	98.14	1.86	0.39	1.04	0.04	0.38
Only coder 6 entered multiple sets	280	1,215	95.23	4.77	0.91	1.32	0.66	1.89
Only coder 7 entered multiple sets	290	1,244	93.81	6.19	1.45	2.89	0.48	1.37
Both entered multiple sets	114	550	95.45	4.55	0.91	2.91	0.18	0.55

favor of coder 6 because she had access to the work of 2 of the other 4 coders, and coder 5 had access to the work of one of the same coders as coder 6 (the production coder). However, no other more suitable measuring procedure was available. The breakdown of the 2.23 percent difference rate (table D) attributes more than half to errors by coder 7 and less than one-quarter to errors by coder 6, even though the same six coders were responsible for both assignments.

Table D also shows that the agreement rate decreased somewhat for the presumably more complicated cases in which one or both coders entered multiple sets of codes, but that coder 6 was consistently charged with fewer errors than coder 7.

Table E shows the overall error rate of each of the six experts in their roles as coder 6 and coder 7, respectively. Each expert's error rate as coder 6 was less than her error rate as coder 7. Error rates for the experts in their coder-6 assignments ranged from 0.32 to 0.88 percent, and the error rates in their coder-7 assignments ranged from 0.85 to 1.27 percent, except for expert 3, who had a relatively high error rate of 1.96.

Determination of the "True" Production Coding Error Rate

Although the error rate of coder 6 over the 30 lots (given in the preceding section) was smaller than that of coder 7, production coding error rates based on the work of coder 7 were considered to be the best estimate of the true level of production coding error. Coder 7 independently arrived at her code selections; while coder 6, because of her access to the work of two other coders, was at least minimally subject to their influence. One valid counterargument to such reasoning is that access to the work of other coders gives the expert a broader perspective on different acceptable coding strategies. However, this argument is not supported by the number of multiple sets coded by the experts (394 by coder 6 and 404 by coder 7). A detailed discussion of multiple sets is found later in the report. With the goal of obtaining a measure of truth as free as possible from any source of bias, coder 7's work was chosen as the ultimate determinant of accuracy.

After the obvious typing and punching errors made by coders 6 and 7 were corrected and the problem of code positioning was resolved, the error rate of production coding for the sample was measured as 4.11 percent by coder 6 and 5.37 percent by coder 7. However, these rates include errors made by the two assignment groups. Coder 6 incorrectly charged 140 errors to the production coder, and coder 7 incorrectly charged 303 errors to the production coder. When the incorrectly charged errors were removed, the production error rate was estimated as 3.60 percent by coder 6 and 4.27 percent by coder 7 (see table F). The second estimate is considered to be the better estimate of the true production coding error rate.

Table E. Number of records coded, number of condition codes, number of errors, and error rate, by coder identification and assignment

Coder identification and assignment	Records coded	Condition codes	Errors	Error rate
	Number			Percent
All experts, both assignments	17,946	55,504	507	0.91
As coder 6 (dependent)	8,973	27,752	156	0.56
As coder 7 (independent)	8,973	27,752	351	1.26
Expert 1, both assignments	2,961	9,396	92	0.98
As coder 6	1,348	4,179	26	0.62
As coder 7	1,613	5,217	66	1.27
Expert 2, both assignments	2,905	8,743	57	0.65
As coder 6	1,582	4,701	15	0.32
As coder 7	1,323	4,042	42	1.04
Expert 3, both assignments	2,900	9,157	111	1.21
As coder 6	1,463	4,712	24	0.51
As coder 7	1,437	4,445	87	1.96
Expert 4, both assignments	3,125	9,729	77	0.79
As coder 6	1,510	4,755	17	0.36
As coder 7	1,615	4,974	60	1.21
Expert 5, both assignments	3,067	9,169	71	0.77
As coder 6	1,518	4,605	32	0.69
As coder 7	1,549	4,564	39	0.85
Expert 6, both assignments	2,988	9,310	99	1.06
As coder 6	1,552	4,800	42	0.88
As coder 7	1,436	4,510	57	1.26

Table F. True production coding error rates and estimated production coding error rates, by verification system

Verification system	Production coding error rate	
	True	Estimated
Coder 7	4.27	...
Three-way independent	...	3.75
Two-way independent with dependent adjudication	...	3.05
Dependent	...	3.67

Production Coding Error Rate Measured by Alternative Verification Systems

The primary goal of the study was to determine which of the following verification systems provides the best estimate of the true error rate:

1. Three-way independent coding,
2. Two-way independent coding with dependent adjudication of differences by a highly qualified coder, or
3. Dependent verification.

The adjusted estimates for these three systems are shown in table F.

The ongoing three-way system estimated the production error rate for the 30 sample lots at 3.75 percent. Coder 6 provided the best measure of error that would be obtained under the second system, the use of dependent adjudication. The estimate of 4.11 percent need not be adjusted for the expert's errors because an adjudicator's work is not normally subject to review. However, the 4.11-percent error rate is not an appropriate measure because a dependent adjudicator would review only those situations in which the production coder and the sample coder disagreed. In the study there were 27,952 comparisons between the production coder and the sample coder to whose work coder 6 had access. Of this total, they agreed in 25,897 cases, leaving 2,055 cases to be adjudicated. From the remaining cases, coder 6 charged the production coder with 852 errors, resulting in an estimated error rate of 852/27,952 or 3.05 percent.

Analysis of dependent verification based on the work of coder 5 produced surprising results. Coder 5 initially charged the production coder with an error rate of 4.24 percent, an estimate higher than the one based on the three-way independent system and almost identical to the one based on coder 7's work. Such a result is contrary to the findings of other studies, which show that error rates based on dependent verification of coding operations tend to understate substantially the true error rate. The two reasons most often given for this understatement are:

1. The verifier is biased in favor of the original coder's work, especially in questionable cases.
2. There is more time and work involved in changing a code than in letting it stand, so the verifier tends to gloss over the work.

In fact, the mortality medical coding operation itself provides evidence of the inaccuracy of dependent verification. The estimated production coding error rate for all 1974 medical coding, based on three-way independent verification, was 0.7 percent larger than the corresponding error rate for 1973 data, which was verified dependently (3.2 percent compared to 2.5 percent).[9] This difference occurred even though the coding staff and the coding instructions were virtually identical in 1973 and 1974.

Analysis of the errors charged to the production coder by coder 5 shows that coder 5 did not identify errors accurately, even though her estimate of the production coder's error rate, 4.24 percent, was close to the true error rate of 4.27 percent. The true error rate of 4.27 percent is based on 1,180 errors made by the production coder. Coder 5 correctly identified 811 of the errors (68.73 percent) and missed the remaining 369 (31.27 percent). However, she offset the missed errors by incorrectly identifying 357 correct codes as errors. The net effect, then, was a charge of 1;168 errors to the production coder and a resultant error rate of 4.24 percent. The percent of errors missed by coder 5, 31.27 percent, is consistent with findings of other studies that measured the quality of dependent coding.[6,7] These studies have shown that a dependent verifier's Type II error rate, that is, the percent of errors that will not be detected, ranges from 30 to 70 percent, depending on the level of coding difficulty. On the other hand, the large number of errors inserted by coder 5 represents an unusual occurrence, since it is quite easy for a dependent verifier to simply go along with the work of the production coder. The insertion of errors by a verifier, called a Type I error, has been negligible in previous studies.[4]

There are two possible explanations for the large number of errors inserted by coder 5. First, it seems reasonable to assume that the coders who worked as coder 5, knowing that they were working on a special project, were more concerned about correcting errors than they would have been had they been working on the regular data file. The coders knew that the study data would be analyzed extensively, whereas dependent verification of the basic mortality file (e.g., during the 1973 data year) was given very little review. The effect of a special assignment on the quality of dependent verification is discussed elsewhere.[4]

In addition, during the 1973 data year the coders in the Data Preparation Branch (DPB) were working under a system of performance standards that emphasized productivity. Significant cash awards were given to coders who produced above-average volumes of work. It seems reasonable to assume that a dependent verifier, not subject to verification and competing against the clock, would tend to agree with most of the production coder's entries rather than repeat the coding process for all records, especially since the verifier had to retype only those records for which changes were made. The dependent verifiers for the study were not on performance standards, and so were not under any such time pressure.

The second explanation involves procedural differences between dependent verification of the 1973 file and dependent verification during the study that were not discovered until after the study had been completed. In the coding of 1973 data the dependent verifier could resolve many ambiguous situations by agreeing with the production coder's entries or by changing those entries without charging the production coder with any errors. In the study coder 5 was allowed to enter only one set of codes for each sample record. If, in ambiguous situations, the preferred set of codes entered by coder 5 differed from that of the production coder, the production coder was charged with one or more errors. These procedural differences undoubtedly caused the production coders to be assigned more errors in the study than they would have received if the 1973 verification procedure had been followed.

In order to estimate the effect of the procedural differences, the multiple sets of codes produced by coders 6 and 7 were examined. Multiple sets generally resulted from ambiguity and confusion about the identity and placement of codes for certain medical conditions. It was expected that coder 5, with training and experience comparable to that of coders 6 and 7, would recognize the same kinds of ambiguous coding situations as they would. Given this assumption, coder 5, had she been permitted, would have entered multiple sets of codes for approximately 400 records, 60 percent of which would have been given the preferred set designation "P" following the set of first choice. (Coder 6 designated a preferred set for 59.1 percent of 394 multiple sets, and coder 7 for 59.4 percent of 404 multiple sets.) These sets represent cases in which, using the 1973 verification procedure, coder 5 could have disagreed with the production coder while not charging her with errors. In the case of the multiple sets of codes produced by coder 6, the expert whose coding procedures more closely resembled those of coder 5, the production coder was charged with 156 errors by preferred sets that were eliminated by the accompanying secondary sets. If the number of errors charged to the production coder by coder 5 is subtracted, the revised production coding error rate based on dependent verification is 3.67 percent.

Table F summarizes the estimated production coding error rates for the 30 sample lots using the three verification systems, compared with the best estimate of the true error rate based on the work of coder 7. All three systems underestimate the true error rate. The major shortcoming of the adjudication system, which yields the smallest estimate, is that errors pass through the system undetected when the sample coder agrees with the production coder. The estimated error rates based on the dependent and three-way independent systems are very similar.

The study data suggest that an error rate based on dependent verification could approximate an error rate based on three-way independent verification if the dependent verifiers (1) are among the best coders in the unit, (2) are sufficiently aware that their work will be reviewed, and (3) can do their work free from any time pressure. However, the closeness of the approximation depends on the number of errors the dependent verifier inserts to offset the errors she misses.

Unfortunately, it would be difficult to satisfy the three conditions in an ongoing dependent verification system. First, if the quality of coding within a coding unit is heterogeneous, the better coders should be assigned to production coding rather than to dependent coding, so that the outgoing product represents the best quality attainable. Second, dependent verification is often instituted because of pressure to complete the work of coding and verifying data in a given time; consequently, the idea of performing dependent verification without time pressure is contradictory. Finally, in order to review the work of a dependent verifier, at least one additional coder must be assigned; thus the manpower saving inherent in dependent verification would be greatly reduced. Also, the additional verification should be done independently, creating other problems, such as scheduling work, selecting samples, comparing codes, and so forth. If the additional verification is dependent, then, according to the criteria, it should also be verified.

The point is that the conditions necessary to make the use of dependent verification worthwhile are very difficult to attain in practice. Thus, for most NCHS coding operations, it would be difficult to institute a dependent verification system that would be cost effective and quality oriented.

Agreement in the Three-Way Verification System

In addition to comparing three-way independent verification with other systems, the study was designed to investigate the following questions about the mechanics of the three-way system (discussed elsewhere[4,10,11]).

1. When two or three original coders agree, how often is the agreed-upon code correct?

2. When the three original coders enter three different codes, how often does at least one coder enter an acceptable code?

Findings concerning these questions are presented in table G. The entries of the three original coders can be classified as AAA, AAB, AAX, ABC, ABX, and XXB, where A, B, and C represent distinct codes and X represents no code.

When two or three coders agreed (AAA, AAB, AAX, XXB), coder 7 agreed with the majority code in 96.78 percent of cases (96.25 + 0.53). After adding coder 7's errors (each such error represented a majority agreement among the three original coders, since at least three of coders 1, 2, 3, and 5 had to agree by position in order to charge coder 7 with an error), the majority code was the correct code in about 98 percent of the cases. However, when two coders agreed (cases AAB, AAX, and XXB, hereafter referred to as "majority rule" or AAB cases) coder 7 agreed with the majority code in only about 80 percent of the cases (74.02 + 5.79) and with the minority code in more than 23 percent of the cases (17.57 + 5.79). Table H shows that the rate of agreement of coder 7 with the minority code was even higher for the most difficult records, namely, those for which one or both experts entered multiple sets of codes. For AAB cases in these records, coder 7 considered the minority code acceptable in more than 40 percent of the cases (21.46 + 19.97), compared with 16-percent agreement when each expert coded one set only. A further indication of the ambiguity and confusion and the resultant

Table G. Number of cases done by coder 7 by agreement classification among the three original coders and percent distribution of coder 7's condition codes by agreement with the three original coders, according to agreement classification among the three original coders

Agreement classification among the three original coders	Number of cases	Codes for which coder 7 agreed with:					
		Total	A only	B only	A and B	A, B, and C	None
		Percent distribution					
Two or three coders agreed (AAA and AAB)[2]	28,077	100.00	96.25	2.61	0.53	...	0.61
Three coders agreed (AAA)	25,521	100.00	98.48	1.11	...	...	[1]0.41
Two coders agreed (AAB)[2]	2,556	100.00	74.02	17.57	5.79	...	2.62
All coders disagreed (ABC, and ABX)	194	100.00	[3]53.09	...	32.99	4.12	9.79
Three coders disagreed (ABC)	128	100.00	[3]34.38	...	49.22	6.25	10.15
Two coders disagreed, no code by third coder (ABX)	66	100.00	[3]89.39	...	1.52	-	9.09

[1]No code by coder 7.
[2]Includes cases where only one coder coded, i.e., XXB, X=A and where one coder did not code, i.e., AAX, X=B.
[3]Coder 7 agreed with one of the coders.

Table H. Number of majority rule cases among the three original coders by number of sets designated by the experts, and percent distribution of these cases by coder 7's agreement with the three original coders, according to the number of sets designated by the experts

Set designation	Majority rule cases	Cases in which coder 7 agreed				
		Total	A only	B only	A and B	None
	Number	Percent distribution				
Total	2,556	100.00	74.02	17.57	5.79	2.62
Coder 6 and coder 7 each coded one set	1,815	100.00	81.32	15.98	. . .	2.70
One or both experts coded more than one set	741	100.00	56.14	21.46	19.97	2.43

NOTE: Majority rule means 2 of the 3 original coders agreed.

Table J. Number of cases coded by the three original coders by number of sets designated by the experts and percent distribution of condition codes by agreement status of the three original coders, according to number of sets designated by the experts

Set designation	Number of cases	Agreement status		
		AAA	AAB	ABC
		Percent distribution		
Total cases	28,271	90.27	9.04	0.69
Coder 6 and coder 7 each coded one set	25,109	92.22	7.23	0.55
One or both experts coded more than one set	3,162	74.79	23.43	1.77

NOTE: AAA = all coders agreed; AAB = 2 of 3 coders agreed; ABC = all coders disagreed.

coding difficulties associated with multiple sets can be seen in table J. When both experts coded only one set of codes, the original three coders agreed (AAA) on more than 92 percent of the codes. When one or both experts coded multiple sets of codes, the percent of AAA cases was below 75 percent. Although the majority code, code A in AAA and AAB cases, was correct about 98 percent of the time for the study as a whole, the percent decreased for the records identified as most difficult to code. These results support the validity of a three-way verification system when experts agree that only one course of action is valid. If one course is not clearly indicated, the outcome of the three-way system is not as reliable.

Agreement by Quality of Coder

Some members of the DPB staff at Research Triangle Park have expressed concern that "good" coders tend to have higher error rates when they are matched with less qualified coders in the three-way verification system. It was contended that the good coders are overruled in situations in which they have in fact entered the correct code. In order to investigate this hypothesis, each coder in the unit was assigned to one of two groups, based on the quality of her coding. The first group of about 21 coders consistently produced coding assignments at the acceptable error level (5 percent or less as measured by the three-way verification system). This group was defined as "good" coders. The second group of 5 coders produced a significant proportion of unacceptable coding assignments. This group, defined as "poor" coders, accounted for about one-quarter of the original coding assignments in the frame (356/472 X 3) but more than two-thirds of the unacceptable coding assignments (104 out of 152). Almost 30 percent of this group's assignments were unacceptable, compared with a 4.5-percent rate of unacceptable assignments for good coders. Table K shows the distribution of majority rule cases (AAB, AAX, and XXB) by coder ability, according to

Table K. Number of sample lots, number of 3-way agreements, number of majority rule cases, and number and percent of coder 7 agreements with minority coders, by composition of the three original coders

Composition of original coders	Sample lots	AAA agreements	Majority rule cases: Total	Majority rule cases: Good coder in minority	Majority rule cases: Poor coder in minority	Coder 7 agreed with minority coder only: Total	Coder 7 agreed with minority coder only: Good coder in minority	Coder 7 agreed with minority coder only: Poor coder in minority	Coder 7 agreed with minority coder only: Total	Coder 7 agreed with minority coder only: Good coder in minority	Coder 7 agreed with minority coder only: Poor coder in minority
	Number								Percent		
0 poor, 3 good	10	8,823	782	782	...	123	123	...	15.7	15.7	...
1 poor, 2 good	13	10,628	1,045	585	460	172	129	43	16.5	22.1	9.3
2 poor, 1 good	7	6,070	729	223	506	154	61	93	21.1	27.4	18.4

the number of good and poor coders among coders 1, 2, and 3.

As the number of poor coders increased, so did the ratio of majority rule cases to three-way agreements (AAA cases). With 0 poor coders, the ratio was 782/8,823 = .089. With 1 poor coder, it was 1,045/10,628 = .098, and with 2 poor coders, the ratio was 729/6,070 = .120. There was not a great deal of difference in the percent of total cases for which coder 7 agreed with the minority coder among the three compositions. However, within each composition group there were significant differences between the percent of cases in which coder 7 agreed with a good coder in the minority and in which she agreed with a poor coder in the minority. These differences diminished somewhat when the percent agreement with the minority coder was adjusted for the quality of the coders in the majority. When a good coder was overruled by two other good coders, the probability that she was correct was 15.7 percent, compared with a 9.3-percent probability for a poor coder overruled by two good coders. When a good coder was overruled by a good coder and a poor coder, the expert agreed with the overruled good coder in 22.1 percent of cases, slightly more than the 18.4 percent of cases in which the expert agreed with a poor coder overruled by the good coder–poor coder combination. These figures suggest that, although it is more probable that good coders are overruled incorrectly when they are in the minority than are poor coders, the difference is not as great as might be expected. Also, the error rates of both the poor and good coders are biased downward because of these cases, since in each case one error was charged to the minority coder when two errors should have been charged, one to each of the majority coders (table L). The average undercharge for good coders was 3.8 errors per assignment (241 errors undercharged ÷ 63 assignments); for poor coders

Table L. Number of errors incorrectly charged, number of errors not charged, and difference between errors not charged and errors incorrectly charged, by composition of groups of good and poor coders

Composition of original coders	Number of errors: Incorrectly charged: Poor coders	Number of errors: Incorrectly charged: Good coders	Number of errors: Not charged: Poor coders	Number of errors: Not charged: Good coders	Difference between errors not charged and errors incorrectly charged: Poor coders	Difference between errors not charged and errors incorrectly charged: Good coders
All lots	136	313	344	554	208	241
0 poor, 3 good	-	123	-	246	-	123
1 poor, 2 good	43	129	129	215	86	86
2 poor, 1 good	93	61	215	93	122	32

the average undercharge was 7.7 errors per assignment (208 errors undercharged ÷ 27 assignments) (tables K and L). Since the average number of codes assigned to preferred sets per sample lot was approximately 920, an estimate of the average downward bias in a coder's error rate is approximately 3.8 ÷ 920 = 0.41 percent for good coders and 7.7 ÷ 920 = 0.84 percent for poor coders. Again, good coders did not fare as well as poor coders, but the average error rate was biased downward for both, accounting for part of the difference between the true production coding error rate and the error rate as estimated in the three-way verification system.

Table G shows that there were only 194 cases of three-way disagreement in the study, representing less than 1 percent of the total number of agreement cases for the three original coders. In these cases coder 7 agreed with at least one coder 90.2 percent of the time and with two or three coders 37.1 percent of the time. Each agreement represents an excess error charged in the three-way system. The total number of excess errors charged in the three-way system in the study is given by 194 (.5309) (1) + 194 (.3299)(2) + 194 (.0412)(3) = 255 (table G), or an average of 255 ÷ 90 = 2.83 excess errors per coder assignment. This overcharge partially compensates for the undercharge in majority rule cases. The remainder of the difference between the true production coding error rate and the estimate based on three-way verification is accounted for by the cases in which coder 7 agreed with none of the original coders.

Multiple Sets Coded by the Experts

As shown in table D, coders 6 and 7 coded 8,973 records, with 27,752 code comparisons. For 8,289 of the records (92.4 percent), containing 24,743 comparisons (89.2 percent), the two experts agreed that the conditions listed on the record generated only one acceptable set of codes. For the remaining 684 records (7.6 percent), at least one expert entered multiple sets of codes. However, coders 6 and 7 jointly agreed that multiple sets of codes were necessary for only 114 of the 684 records (16.7 percent). There was apparently considerable disagreement among the six expert coders as to what constituted a vague or ambiguous coding situation. The average number of comparisons for the 8,289 records with no multiple sets was 3.0, and the comparable average for the 684 records involving multiple sets was 4.4. For the 114 records for which both expert coders entered multiple sets of codes, the average increased to 4.8, leading to the obvious conclusion that multiple sets occur for the more complicated death certificates.

Because coder 6 had access to the work of two coders who sometimes disagreed with each other, she was expected to code more multiple sets than coder 7, even though the same set of six coders was responsible for all coder-6 and coder-7 assignments. It is surprising that coder 7 had more multiple sets, 404 to 394 (table M). However, more of coder 6's than of 7's multiple sets occurred for records for which the two coders to whose work she had access were in disagreement (table M). Furthermore, coder 6 agreed with both coders in a much larger proportion of code difference cases than did coder 7.

It appears that coder 6 was influenced to code multiple sets, at least to some extent, by the alternative opinions she was able to review.

As shown in table N, 4 of the 6 expert coders entered about the same number of multiple sets in their coder-6 and coder-7 assignments. Expert 3 entered more multiple sets as coder 6 than she did as coder 7, and expert 6 entered more as coder 7 than as coder 6. The total number of multiple sets coded by each expert were similar, except for expert 4, who coded multiple sets almost twice as often as any of the other coders.

An analysis of the 798 multiple sets shows that the average number of sets coded was 2.24 (2.22 for coder 6 and 2.26 for coder 7). Table O shows that the total number of vague or ambiguous conditions that caused the experts to enter multiple sets was 585 for coder 6 and 599 for coder 7, or an average of 585/394 = 599/404 = 1.48 ambiguous conditions per multiple set for coders 6 and 7. Since coders 6 and 7 entered 27,622 and 27,630 codes, respectively, for the 8,973 sample records, they both considered between 2.1 and 2.2 percent of the conditions to be ambiguous.

However, these similarities do not mean that coders 6 and 7 agreed on what constituted an ambiguous coding situation. As mentioned

Table M. Number of records with multiple sets, number of differences between the production coder and the sample coder, and number and percent of code differences for which the expert agreed with both coders, by experts entering multiple sets

Coders who entered multiple sets	Number of records with multiple sets	Differences between production coder and sample coder[1]		Expert agreed with both coders	
		Number of records	Number of condition codes	Number of condition code differences	Percent of condition code differences
Coders 6 and 7	114	71	131	...	...
Coder 6	114	71	131	75	57.25
Coder 7	114	71	131	52	39.69
Coder 6 only	280	148	283	179	63.25
Coder 7 only	290	94	213	58	27.23

[1]The production coder (coder 1) and the sample coder (coder 2 or 3) whose work was available to coder 6.

Table N. Number of records coded and number and percent of records with multiple sets, by coder identification and assignment

Coder identification and assignment	Records coded	Records with multiple sets	
	Number		Percent
All experts, both assignments	17,946	[1]798	4.45
As coder 6	8,973	394	4.39
As coder 7	8,973	404	4.50
Expert 1, both assignments	2,961	132	4.46
As coder 6	1,348	67	4.97
As coder 7	1,613	65	4.03
Expert 2, both assignments	2,905	106	3.65
As coder 6	1,582	52	3.29
As coder 7	1,323	54	4.08
Expert 3, both assignments	2,900	83	2.86
As coder 6	1,463	52	3.55
As coder 7	1,437	31	2.16
Expert 4, both assignments	3,125	243	7.78
As coder 6	1,510	120	7.95
As coder 7	1,615	123	7.62
Expert 5, both assignments	3,067	97	3.16
As coder 6	1,518	54	3.56
As coder 7	1,549	43	2.78
Expert 6, both assignments	2,988	137	4.59
As coder 6	1,552	49	3.16
As coder 7	1,436	88	6.13

[1]Includes 114 records with multiple sets coded by both coders 6 and 7.

Table O. Number of ambiguous cases identified, number and percent of cases for which experts agreed on the condition causing ambiguity, and number and percent of cases for which they agreed on both the condition and the alternative ways of coding it, by expert coder identification

Coder	Number of ambiguous cases identified			Cases for which experts agreed on condition causing ambiguity			Cases for which experts agreed on the condition causing ambiguity and the alternative ways of coding the condition		
	Records with multiple sets (1)	Records for which one expert coded multiple sets (2)	Records for which both experts coded multiple sets (3)	Number (4)	Percent of column 3 (5)	Percent of column 1 (6)	Number (7)	Percent of column 3 (8)	Percent of column 1 (9)
6	585	415	170	121	71.2	20.7	90	52.9	15.4
7	599	411	188	121	64.4	20.2	90	47.9	15.0

earlier, both coders jointly entered multiple sets for only 114 records. In these 114 records coder 6 identified 170 ambiguous situations, and coder 7 identified 188. In 121 cases the two coders agreed on the condition that caused ambiguity, but for only 90 of these cases did they completely agree on the alternative ways of coding the ambiguous condition (about 15 percent of each coder's total number of ambiguous conditions).

Characteristics of Multiple Sets of Codes

Each time coder 6 or coder 7 coded a multiple set, she entered one of the six reason codes listed earlier. Reasons A and B (handwriting and formatting) represent problems that occur during the completion of the death certificate, reason C (poor microfilm) is generated at the State office, reasons D and E (vague instructions) are essentially NCHS inhouse problems, and reason F is a catchall category. The six expert coders were cautioned to use reason F as little as possible. Nevertheless, as shown in table P, F was the reason most frequently coded. Although reason F could have been used to denote a reason other than reasons A-E, according to the experts it was used almost exclusively to indicate combinations of reasons A-E.

Of the 232 and 245 records, respectively, coded to reasons A-E by coders 6 and 7, 42.2 percent and 40.4 percent, respectively, were coded to reason A, 16.4 and 20.4 percent to reason B, 9.9 and 7.3 percent to reason D, and 31.0 and 31.4 percent to reason E. These percents can be calculated from the data in table P. Reason C was virtually nonexistent in the study, although poor quality microfilm is encountered from time to time in everyday medical coding. It is possible that reason C generally accom-

Table P. Number of records with multiple sets by reasons for coding them and expert coder identification

Coder	Number of records with multiple sets						
	Total	Reason					
		A	B	C	D	E	F
Coder 6	394	98	38	1	23	72	162
Coder 7	404	99	50	1	18	77	159
Both coders	[1]114	25	5	-	3	1	32

[1]For 48 of the 114 records, coders 6 and 7 did not agree on the reason for multiple sets.

NOTE: A = poor handwriting; B = formatting problem; C = poor quality microfilm; D = vagueness in 1974 coding instructions, rectified in the 1975 instructions; E = vagueness in 1974 coding instructions, not rectified in the 1975 instructions; F = other or multiple reasons.

panied other reasons for ambiguity, and such cases are thus embedded in F codes. On the other hand, if each F code encompasses two of reasons A-E and these reasons are distributed among the F codes according to their distribution among the single reason codes, the expected distribution of records by reason is as shown in table Q. To illustrate the derivation of the data in table Q, the expected number of records coded with multiple sets due to reason A according to coder 6 is given by

$$N_E = N_A + N_{FA}$$

where

N_A = the number of records originally coded to reason A by coder 6

N_{FA} = the expected number of records coded to reason F by coder 6 that have reason A as a contributing cause of ambiguity.

Since coder 6 entered reason F for 162 records,

$$N_{FA} = 162 \{P_r \text{ (1st reason is A)} + P_r \text{ (1st reason is not A and the 2d reason is A)}\}.$$

If reasons A-E are distributed proportionally among the F codes, the first term in braces is .422. Using elementary probability theory, the second term inside the braces can be expressed as

$$\sum_{i=B}^{E} [P_r \text{ (2d reason is A|1st reason is } i)] [P_r \text{ (1st reason is } i)]$$

The probabilities for the second term in brackets can be calculated from the data in table P. They are .164 for reason B, .004 for reason C, .099 for reason D, and .300 for reason E. The conditional probabilities for the first term in brackets are computed by subtracting the occurrences of reason i from 232, then dividing 98, the number of occurrences of reason A, by the revised denominator. These revised probabilities of occurrence for reason A estimate the probability that A is listed as a second reason for ambiguity given that reason i has already been cited. The revised probabilities for reason A are .505 with reason B removed, .424 with C removed, .469 with D removed, and .613 with E removed. Substituting the values in the formula for N_{FA} yields

$$\begin{aligned} N_{FA} &= 162 \{.422 + (.505)(.164) + (.424)(.004) + (.469)(.099) + (.613)(.310)\} \\ &= 162 \{.422 + .321\} \\ &= 162 \{.743\} \\ &= 120 \end{aligned}$$

so that $N_E = 98 + 120 = 218$, as shown in table Q. The remaining entries were produced by similar calculations.

As a result of the assignment procedures, reason A was at least a contributing reason for

Table Q. Number of records with multiple sets and expected number of multiple set records for specified reasons, by expert coder identification

Coder	Number of records with multiple sets	Expected number of records with multiple sets: Reason[1] A	B	C	D	E
6	394	218	99	3	62	174
7	404	214	122	3	46	178

[1]Reason F occurrences were allocated to A-E according to their distribution in single reason codes.

NOTE: A = poor handwriting; B = formatting problem; C = poor quality microfilm; D = vagueness in 1974 coding instructions, rectified in the 1975 instructions; E = vagueness in 1974 coding instructions, not rectified in the 1975 instructions; F = other or multiple reasons.

ambiguity in 55.3 percent and 53.0 percent of the records coded with multiple sets by coders 6 and 7, respectively. The corresponding percents for reasons B, D, and E were 25.1 and 30.2 percent, 15.7 and 11.4 percent, and 44.2 and 44.1 percent, respectively. The contribution of reason C was, of course, still negligible.

As mentioned in table O, coder 6 encountered a total of 585 ambiguous conditions for records for which she coded multiple sets, and coder 7 encountered 599. There were four ways in which an ambiguity could occur (see figure 1). The number of type-3 ambiguous conditions is probably too large because in some cases a condition could be acceptably represented by a single "combination" code or two or more "component" codes. Although these circumstances represent a variation of a type-1 ambiguity, they were counted as a type 1 and one or more type-3 ambiguities, depending on the number of codes that could be combined. Unfortunately, the authors did not have the coding expertise to identify such cases, and the expert coders had already devoted their allotted time to the study.

Type of ambiguity	Number of occurrences	Description
1. Alternate codes	381	A stated condition can be coded as A or B.
2. Order	15	Two stated conditions can be coded on the same line in the order AB or BA.
3. Inclusion-exclusion	603	A condition is either valid and should be coded A, or it is not valid and therefore should not be coded.
4. Format	185	A condition is valid and should be coded A, but it is not clear on which line of the record code A should be entered

Figure 1. Number of occurrences of ambiguous cases and descriptions of these occurrences by type of ambiguity

The 1,184 ambiguous coding situations involved 1,615 condition codes from 495 distinct ICDA code categories. Each occurrence of type-1 and -2 ambiguities involved at least two codes, and types 3 and 4 involved only one code per occurrence. Table R shows the frequency distribution of the 1,615 ambiguous codes by reason and type of ambiguity, and the percent distribution by reason according to type of ambiguity. The most surprising result is that in cell 4B. It was expected that virtually all the format problems (type 4) would have reason B (format)

Table R. Number and percent distribution of ambiguous codes by reason, according to type of ambiguity

Reason	Ambiguous codes				
	Total	Type 1	Type 2	Type 3	Type 4
	Number				
Total	1,615	797	30	603	185
A	350	194	-	138	18
B	161	14	12	42	93
C	2	-	-	1	1
D	95	54	4	34	3
E	308	182	2	101	23
F	699	353	12	287	47
	Percent distribution				
Total	100.0	100.0	100.0	100.0	100.0
A	21.7	24.3	-	22.9	9.7
B	10.0	1.8	40.0	7.0	50.3
C	0.1	-	-	0.2	0.5
D	5.9	6.8	13.3	5.6	1.6
E	19.1	22.8	6.7	16.7	12.4
F	43.3	44.3	40.0	47.6	25.4

NOTE: A = poor handwriting; B = formatting problem; C = poor quality microfilm; D = vagueness in 1974 coding instructions, rectified in the 1975 instructions; E = vagueness in 1974 coding instructions, not rectified in the 1975 instructions; F = other or multiple reasons.

given as the reason for coding a multiple set. However, reason B was given for only 50 percent of the type-4 ambiguities, and, even assuming that when reason F was given it included reason B, the remaining one-quarter of the ambiguous conditions were assigned multiple sets for reasons other than format.

The question arises: Which code categories were the most troublesome for the experts to code? Presumably, the categories that gave expert coders the most trouble should also present problems to medical coders as a whole. One way to measure the degree of ambiguity associated with particular code categories is the "index of confusion" methodology.

Index of Confusion

As part of the analysis of industry and occupation descriptions in the 1960 census of population, a sample of records containing these descriptions was independently coded by three U.S. Bureau of the Census coders. Using the data from this study, the Census Bureau developed an "index of consistency," which was computed for each industry and occupation code.[12] This index was used to answer the question, "How consistently was code X independently applied by three different coders looking at the same description?"

The methodology has been modified somewhat for use in analyzing the ambiguity data from this study. Since the six expert coders were allowed to list more than one set of acceptable codes for a record, a particular code category could appear either as a fixed, unique solution to a given coding situation or as one of two or more alternative solutions. The expression for an index of confusion for code category c is given by

$$I_c = \frac{M_c}{M_c + U_c}$$

where

M_c = the number of ambiguous coding situations in which code c appears in one or more of the alternative solutions.

U_c = the number of occurrences of code c as the unique coding solution for a medical condition in either a single or a multiple set (i.e., code c is counted only once if it appears as a unique solution on each line of a multiple set).

An index of confusion was computed for 157 codes for which $M_c \geqslant 3$ (table S). Although these codes comprise less than one-third of the ICDA categories represented in ambiguous coding situations, they include 1,171 (appendix I) of the 1,615 codes (table R) involved in those situations (72.5 percent). The work of coders 6 and 7 was combined in computing indexes of confusion because their numbers of multiple sets and their distributions by reason and type of ambiguity were quite similar, and because of the relatively small number of ambiguous conditions coded in the study.

Table S shows the distribution of the 157 codes by total occurrences ($M_c + U_c$) and index of confusion. (A description of each code and its index of confusion is found in appendix I.) Since coders 6 and 7 entered a total of 55,504 codes (table D), the average index of confusion for the study is 1,615 ÷ 55,504 = .029. The average index of confusion for the 157 codes also happens to be .029. However, when the 10 codes for which $M_c + U_c > 1{,}000$ are excluded, the remaining 147 codes have an average index of confusion of .049, with 92 having indexes of .050 or more. The average number of total occurrences for the 157 codes is 256; however, the average is heavily weighted by the codes in the <.050 category. Of the 92 codes with indexes of at least .050, the largest number of total occurrences is 261, and the average is 33. Of the 53 codes with indexes of at least .250, 32 have 10 or fewer total occurrences, 51 have 25 or fewer total occurrences, and the average number of occurrences is 10 (table S). A logical explanation for the high indexes associated with these codes is the infrequent occurrence of the causes of death they represent. These causes may be difficult for physicians and other certifiers to describe accurately, and the coders may not be sufficiently familiar with them to code them properly. However, it is especially important that the codes that occur infrequently are

Table S. Distribution of the 157 condition codes appearing in three or more multiple sets by frequency of occurrence in all sets and index of confusion

Index of confusion	Occurrence in all sets											
	Total	10 or less	11-25	26-50	51-100	101-200	201-300	301-400	401-500	501-750	751-1,000	1,000 or more
Total	157	32	27	16	16	23	14	6	5	3	5	10
.050 or less	65	-	-	-	4	20	12	6	5	3	5	10
.050-.099	14	-	-	4	10	-	-	-	-	-	-	-
.100-.149	13	-	1	7	1	2	2	-	-	-	-	-
.150-.199	7	-	3	2	1	1	-	-	-	-	-	-
.200-.249	5	-	4	1	-	-	-	-	-	-	-	-
.250-.299	8	-	8	-	-	-	-	-	-	-	-	-
.300-.349	3	-	3	-	-	-	-	-	-	-	-	-
.350-.399	7	2	4	1	-	-	-	-	-	-	-	-
.400-.449	6	4	1	1	-	-	-	-	-	-	-	-
.450-.499	1	-	1	-	-	-	-	-	-	-	-	-
.500-.599	5	5	-	-	-	-	-	-	-	-	-	-
.600-.699	4	3	1	-	-	-	-	-	-	-	-	-
.700-.799	4	3	1	-	-	-	-	-	-	-	-	-
.800-.899	3	3	-	-	-	-	-	-	-	-	-	-
.900-.999	-	-	-	-	-	-	-	-	-	-	-	-
1.000	12	12	-	-	-	-	-	-	-	-	-	-

handled properly because each misclassification results in a large percent change relative to a code's actual frequency of occurrence. Furthermore, the misclassification of several causes of death into a rare category could have severe ramifications. For example, in 1968 smallpox, which had last occurred in the United States in 1953, was listed as the underlying cause of a death in one of the preliminary tabulations for the annual mortality publication.[13] Investigation revealed that the actual cause of death was cowpox. Failure to correct the misclassification would have, at the least, resulted in an embarrassment to the United States when subsequent investigation uncovered the error.

Appendix I shows that of the 53 codes with indexes of confusion of at least .250, 32 had either or both "other" or "unspecified" in their descriptions. These words are used to indicate ill-defined or catchall categories. An additional five codes belonged to major cause group XVI, "Symptoms and ill-defined conditions" (table B), and three more represented secondary neoplasms. Many of the 104 codes with smaller indexes of confusion also had the words "other" and/or "unspecified" in their descriptions. The high frequency of occurrence of these equivocal descriptions is consistent with the concept of confusion due to vague descriptions of conditions on the death certificate, inadequate coding instructions, and questions of interpretation among the coders.

In order to try to explain the reasons for ambiguity in certain codes, the confused codes were examined by type of ambiguity and the reason given by the expert for coding a multiple set of codes. The following occurrences were the most noteworthy:

1. There appeared to be a problem in distinguishing between ICDA codes 199.0 (multiple malignant neoplasm, site unspecified) and 199.1 (other malignant neoplasm, site unspecified) and between these two codes and other codes for malignant neoplasms. Codes 199.0 and 199.1 were listed as alternative codes five times, all with reasons D, E, or F. The problem is more noticeable for code 199.1 because it occurs less frequently than 199.0.

2. Anemia, unspecified (ICDA code 285.9) and uremia (code 792) were given as alternative codes nine times, with reason A (handwriting) given as the reason eight times. Anuria, not of newborn (code 786.5) was given twice as an alternative for anemia, once for reason A and once for reason F. These two cases were the only ambiguous occurrences of anuria.

3. Codes 410.9 (acute myocardial infarction without mention of hypertensive disease) and 412.3 (chronic ischemic heart disease, without mention of hypertension) were given as alternative codes 10 times, with 5 of the occurrences attributed to reason A and 4 others to reason F. These occurrences and the other ambiguous occurrences of the two codes are relatively unimportant because the codes occur so often.

4. All four ambiguous occurrences of codes 537.9 (other and unspecified diseases of stomach and duodenum) and 560.9 (other and unspecified intestinal obstruction) were cases of inclusion or exclusion.

5. There was some confusion about whether to code certain conditions under general heading XVI, "Symptoms and ill-defined conditions." Most notable were ICDA codes 795.2 (any sudden death, age 2 years and over), 796.0 (other ill-defined conditions), 796.3 (died without sign of disease), and 796.8 (unknown cause of morbidity and mortality). These four codes accounted for 57 cases of ambiguity, of which 45 were cases of inclusion or exclusion.

6. The expert coders were somewhat confused about assigning nature of injury codes for conditions due to adverse effects of chemical substances. ICDA code N977.8 (other specified drugs, not elsewhere classified) occurred eight times in the study, all in ambiguous situations: it was an alternative in two cases each to code N960.2 (adverse effect of the antibiotic chloramphenicol), N961.9 (adverse effect of other anti-infectives), N987.4 (toxic effect of freons), and N989.9 (toxic effect of other substances chiefly nonmedicinal as to source, other). The reasons for the two associations with N960.2 were both A, the reasons for N961.9 were both F, and the four remaining cases were given reason B by the expert. Code N987.4, in addition to its two associations with N977.8, was associated twice with code N989.9 in its four occurrences in the study.

7. The two codes that were involved in the largest variety of ambiguous situations were nature of injury code N998.9 (other complications of surgical procedures) and external cause of injury code E930.0 (complications and misadventures in surgical treatment, excluding effects of anesthetic management). The 26 ambiguous occurrences of code N998.9 included 6 format problems, 10 questions of inclusion or exclusion, and 10 associations with 8 different conditions, ranging from congestive heart failure (code 427.0) to gangrene (code 445.9) to perforated ulcer of the stomach (code 531.1). The 32 ambiguous occurrences of code E930.0 included 5 format problems, 16 questions of inclusion or exclusion, 5 associations with code E947 (late effects of surgical operation), 2 associations with code 686.9 (other local infections of skin and subcutaneous tissue), and 4 other single associations with conditions ranging from other and unspecified infective and parasitic diseases (code 136) to other and unspecified stomach ulcer (code 531.9).

8. Five of the six ambiguous occurrences of code E947 (late effects of surgical procedures) were in association with code E930.0. The reason codes for these associations were two E's and three F's. One possible explanation is that the coding instructions did not clearly define what constitutes late, as opposed to immediate, effects of a surgical misadventure.

ANALYSIS OF UNDERLYING CAUSE-OF-DEATH CODING

The preceding analysis was generated from the ICDA codes that represent individual causes on the death certificate. As mentioned briefly in the introduction, condition codes serve as input to a computer program entitled Automated Coding of Medical Entities (ACME). The ACME program is supplied with extremely complex decision tables that take into account the identity and format of the individual ICDA codes and assign one condition, called the "underlying cause of death," to represent all conditions on a certificate. One commonly used definition of the underlying cause of death is "the condition that initiated the sequence of events that resulted in death." However, the underlying cause selection and modification rules, drafted by the World Health Organization (WHO) and built into ACME, sometimes deviate from the sequence initiation concept. These same rules are used by the nosologists to handle the certificates that are rejected by ACME. The underlying cause as defined by these rules is used to produce the statistics shown in the annual mortality volumes of *Vital Statistics of the United States* and in newspapers and other public information sources.

There are two major reasons why variation in the coding of detailed causes of death may not produce corresponding variation in the underlying cause of death for a particular record.

1. One or more of the conditions on a record may not have any bearing on the underlying cause; thus, errors in coding the condition cannot affect the underlying cause code. One large group of ICDA codes with this property is the set of approximately 500 four-digit nature of injury codes, which by convention cannot be used as underlying causes of death.
2. Except in one detailed table in *Vital Statistics of the United States*, the 2,000 underlying cause codes are virtually always collapsed into more general cause groups for analysis. Thus an error in an individual condition code, especially in the third or fourth digit, may change the underlying cause yet not remove it from the more general category.

In order to produce underlying cause coding error rates and agreement rates comparable to the condition coding rates analyzed earlier, the condition codes of the six coders for each sample record were processed by ACME, and an underlying cause of death was assigned for each coder for each record. When an expert coder entered multiple sets of codes, a separate underlying cause was assigned for each set. The computation process for underlying cause coding rates is much simpler than the process for condition coding rates because there is only one cause per record. An underlying cause code agreement, disagreement, or error rate at a specified level of cause grouping is given by

$$R = \frac{\text{number of agreements, disagreements, or errors}}{\text{number of sample records matched}}$$

The cause groupings used for analysis of underlying cause code data in this report are the total detailed cause list, excluding nature of injury codes (commonly referred to as the "each cause list"), the 281-cause list, which is the most detailed collapsed cause list used in *Vital Statistics of the United States*, and the ICDA list of 17 major cause headings, shown earlier in table B.

Expert Agreement

Table T compares the condition coding agreement rates between coders 6 and 7 (given in table D) with underlying cause coding agreement rates at the each cause and 281-cause levels. Underlying cause code agreement was computed in the same way as condition code agreement; that is, coders 6 and 7 were counted as agreeing on the underlying cause code for a record if the underlying cause code generated from any of coder 6's sets matched the underlying cause code from any of coder 7's sets. The data in table T show that the underlying cause code agreement between coders 6 and 7 was greater than their condition code agreement. As expected, the agreement rate increased when the causes were combined into 281-cause groups.

Underlying Cause-of-Death Coding Error Rate for Production Coding

The procedures for calculating estimates of the "true" underlying cause coding error rate based on the production coder's work were similar to the procedures used to estimate the

Table T. Number of sample records, number of code comparisons and percent of condition agreements between the experts, and number and percent of underlying cause agreements, by number of sets designated by the experts

Sets entered by coders 6 and 7	Number of sample records	Condition agreement		Underlying cause agreement			
				Each cause level		281-cause level	
		Number of code comparisons	Agreement rate (percent)	Number of agreements	Agreement rate (percent)	Number of agreements	Agreement rate (percent)
All records	8,973	27,752	97.77	8,870	98.85	8,895	99.13
Both coders entered one set	8,289	24,743	98.14	8,202	98.95	8,224	99.22
Only coder 6 entered multiple sets	280	1,215	95.23	273	97.50	275	98.21
Only coder 7 entered multiple sets	290	1,244	93.81	282	97.24	283	97.59
Both coders entered multiple sets	114	550	95.45	113	99.12	113	99.12

"true" condition coding error rate. As before, the differences between the underlying cause codes generated from the work of coders 6 and 7 at each level of specificity were adjudicated. If the underlying cause code generated from the work of at least three of coders 1, 2, 3, and 5 matched the underlying cause code of coder 6 or coder 7, the other expert was charged with an error. If the underlying cause code generated from the work of at least 3 of the 4 other coders agreed, but did not match any of the codes derived from either expert's work, both experts were charged with an error. In all other cases the difference was considered to be unresolved. All charged errors were subtracted from the production coder's total number of errors as measured by coder 6 and coder 7, except for cases in which the production coder's underlying cause code agreed with that of the expert charged with the error.

Table U shows the production coder's condition code error rate and her underlying cause code error rate at the three levels of specificity as measured by coder 6, coder 7, and the three-way independent verification procedure. For all three measures the estimated underlying cause code error rate at the each cause level is slightly more than one-half the estimated condition code error rate. The rates for the less specific cause groupings are correspondingly lower. It is interesting that the error rate generated from coder 7's work is the highest of the three rates in each column, but the difference between the condition code error rates of coder 6 and three-way independent verification virtually disappears when underlying cause coding is measured.

Underlying Cause-of-Death Coding Agreement in the Three-Way Verification System

Another area of investigation that can be carried over from condition coding to underlying cause coding concerns the mechanics of

Table U. Production coder's condition coding error rate and underlying cause coding error rate for selected cause groupings, by coder 6, coder 7, and the three-way independent verification system

Measurement standard	Condition coding error rate	Underlying cause error rate		
		Each cause	281-cause	17 major categories
	Percent			
Coder 6	3.60	1.87	1.48	0.79
Coder 7	4.27	2.21	1.78	0.94
Three-way independent verification	3.75	1.89	1.48	0.77

the three-way verification system. Two questions asked in a previous section were rephrased as follows in terms of underlying cause:

1. When the work of two or three original coders leads to the same underlying cause code (even if the condition codes do not match exactly), how often is the agreed-upon code correct?
2. When the work of the three original coders leads to three different underlying cause codes, how often is one or more of these codes correct?

Findings on these questions are summarized at two levels of specificity, the each cause level (table W) and the 281-cause level (table Y). The 10 to 1 ratio of AAA to AAB cases for condition coding (table G) increased for underlying cause coding to about 19 to 1 at the each cause level and 25 to 1 at the 281-cause level, reinforcing the conclusion that many condition-code differences do not affect the underlying cause code. As might be expected, the percent distributions in tables W and Y closely resemble those in table G.

A related issue that was discussed for condition coding was the inclination of coder 6 to use multiple sets of codes to account for differences between two coders to whose work she had access (see table M). Similar findings for underlying cause coding are summarized at the each cause level (table Z) and the 281-cause level (table AA). At both levels of specificity, coder 6

Table W. Number of records and percent distribution of coder 7's underlying cause-of-death codes by agreement with original coders at the each cause level, according to agreement classification among the three original coders

Agreement classification	Records	Codes on which coder 7 agreed					
		Total	A only	B only	A and B	A, B, and C	None
	Number	Percent distribution					
All coders agreed (AAA)	8,488	100.0	99.21	...	...	...	[1]0.79
Two coders agreed (AAB)[2]	455	100.0	74.51	16.48	7.03	...	1.98
Two or three coders agreed (AAA and AAB)[2]	8,943	100.0	97.95	1.59	0.36	...	0.10
All coders disagreed (ABC and ABX)	30	100.0	[3]83.33	...	-	-	16.67

[1]No code by coder 7.
[2]Includes cases where only one coder coded, i.e., XXB, X = A, and where one coder did not code, i.e., AAX, X = B.
[3]Coder 7 agreed with one of the coders.

Table Y. Number of records and percent distribution of coder 7's underlying cause-of-death codes by agreement with original coders at the 281-cause level, according to agreement classification among the three original coders

Agreement classification	Records	Codes on which coder 7 agreed					
		Total	A only	B only	A and B	A, B, and C	None
	Number	Percent distribution					
All coders agreed (AAA)	8,609	100.0	99.42	...	...	...	[1]0.58
Two coders agreed (AAB)[2]	346	100.0	70.81	19.65	8.38	...	1.16
Two or three coders agreed (AAA and AAB)[2]	8,955	100.0	98.31	1.32	0.32	...	0.05
All coders disagreed (ABC)	18	100.0	[3]83.33	...	5.56	-	11.11

[1]No code by coder 7.
[2]Includes cases where only one coder coded; i.e., XXB, X = A, and where one coder did not code, i.e., AAX, X = B.
[3]Coder 7 agreed with one of the coders.

Table Z. Number of records for which multiple sets of underlying cause-of-death codes were entered at the each-cause level, number of records for which the production coder and sample coder disagreed, and number and percent of code differences for which the expert agreed with both coders, by coder identification

Coders who entered multiple sets	Number of records		Expert agreed with both codes	
	With multiple sets	Production coder and sample coder[1] disagreed	Number of differences	Percent of differences
Coder 6 only	280	32	16	50.00
Coder 7 only	290	27	8	29.63
Coders 6 and 7	114	19	{12 {11	{63.16 {57.89

[1]The production coder and the sample coder whose work was available to coder 6.

Table AA. Number of records for which multiple sets of underlying cause-of-death codes were entered at the 281-cause level, number of records for which the production coder and sample coder disagreed, and number and percent of code differences for which the expert agreed with both coders, by coder identification

Coders who entered multiple sets	Number of records		Expert agreed with both codes	
	With multiple sets	Production coder and sample coder[1] disagreed	Number of differences	Percent of differences
Coder 6 only	280	25	13	52.00
Coder 7 only	290	21	5	23.81
Coders 6 and 7	114	17	{10 {12	{58.82 {70.59

[1]The production coder and the sample coder whose work was available to coder 6.

resolved a larger proportion of differences between the production coder and the sample coder to whose work she had access than did coder 7, the independent expert. However, these results hold only for cases in which only coder 6 or coder 7 entered multiple sets of codes. The proportions are similar for the 114 records for which both coders entered multiple sets, although the number of observations is too small to permit any definitive conclusions.

——— ooo ———

REFERENCES

[1]National Center for Health Statistics: *Eighth Revision International Classification of Diseases, Adapted for Use in the United States.* PHS Pub. No. 1693. Public Health Service. Washington. U.S. Government Printing Office, 1967.

[2]McKeon, A. J.: Estimating the true process average by means of controlled errors. *Science in Management,* Proceedings of 1959 Conference of the Administrative Applications Division. Washington. 1959. pp. 197-206.

[3]Gilford, L.:Estimating the Error of the Verifying Clerk When His Work Is Not Independent of the Original Clerk's. Unpublished document.

[4]Minton, G.: Inspection and correction error in data processing. *J. Am. Statistical Assoc.* 64: 1256-1275, Dec. 1969.

[5]Hansen, M. H., Fasteau, H. H., Ingram, J. J., et al.: Quality control in the 1960 Census, in American Society for Quality Control, *New Frontiers in Administrative and Engineering Quality Control,* Proceedings of the 1962 Middle Atlantic Conference. Milwaukee, 1962. pp. 311-339.

[6]Minton, G.: Comments in quality control and research in data processing programs, in American Society for Quality Control, *Statistical Science in Management,* Proceedings of Administrative Applications Division Conference. Arlington, March 1970. pp. 3-G(1)-3-G(23).

[7]Jaehn, A. H., and Bingham, R. S.: Decision rules for inefficient sorting, in American Society for Quality Control, *Annual Technical Conference Transactions,* 26th Annual Technical Conference. Washington, May 8-10, 1972. pp. 311-318.

[8]U.S. Bureau of the Census: *United States Censuses of Population and Housing 1960: Quality Control of Preparatory Operations, Microfilming and Coding.* Washington, 1965.

[9]Harris, K. W.: Analysis of the independent three-way verification system in mortality medical coding, Oct. 1974. Memorandum.

[10]Fasteau, H. H., Ingram, J. J., and Minton, G.: Control of quality of coding in the 1960 Censuses. *J. Am. Statistical Assoc.* 59(305): 120-132, March 1964.

[11]U.S. Bureau of the Census: *1970 Census of Population and Housing Evaluation and Research Program*; Coding Performance in the 1970 Census. Washington, April 1974.

[12]Fasteau, H. H., Ingram, J. J., and Mills, R. H.: Study of the reliability of the coding of census returns. *Proceedings of the Social Statistics Section,* American Statistical Association. Minneapolis. 1962. pp. 104-115.

[13]National Center for Health Statistics: *Vital Statistics of the United States, 1968,* Vol. II, DHEW Pub. No. (HSM) 72-1101. Health Services and Mental Health Administration, Washington. U.S. Government Printing Office, 1972.

[14]Hansen, M., Hurwitz, W., and Madow, W.: *Sample Survey Methods and Theory,* vol. I. New York. John Wiley & Sons, 1953.

APPENDIXES

CONTENTS

I. ICDA Cause-of-Death Codes Appearing in Three or More Ambiguous Coding Situations and Their Indexes of Confusion 32

II. Technical Notes 35
- Estimation 35
- Sampling Errors 35
- Nonsampling Errors and Sources of Bias 37

III. U.S. Standard Certificate of Death 40

LIST OF APPENDIX TABLES

I. Sampling errors and design effects for condition code agreement rates and underlying cause-of-death code agreement rates for coders 6 and 7, by multiple set status 36

II. Sampling errors and design effects for the production coder's condition code error rate and underlying cause-of-death code error rate, by expert coders 36

III. Sampling errors and design effects of condition code agreement rates and underlying cause-of-death code agreement rates for coder 7, by the three original coders for selected agreement status categories 37

IV. Quality of dependent verification in selected data processing activities 38

APPENDIX I

ICDA CAUSE-OF-DEATH CODES APPEARING IN THREE OR MORE AMBIGUOUS CODING SITUATIONS AND THEIR INDEXES OF CONFUSION

Description and code		Number of occurrences in multiple sets	Number of occurrences in all sets	Index of confusion[1]
Pulmonary tuberculosis NOS	011.9	6	43	0.1395
Pulmonary diseases due to mycobacteria	031.0	3	3	1.0000
Other bacterial septicemia	038.8	4	15	0.2667
Unspecified septicemia	038.9	12	301	0.0399
Viral infection, unspecified	079.9	4	8	0.5000
Moniliasis	112	4	6	0.6667
Other and unspecified infective and parasitic diseases	136	9	12	0.7500
Malignant neoplasm of stomach, part unspecified	151.9	4	143	0.0280
Malignant neoplasm of large intestine (including colon), part unspecified	153.8	9	290	0.0310
Malignant neoplasm of rectum	154.1	3	68	0.0441
Malignant neoplasm of pancreas, unspecified, NOS	157.9	3	135	0.0222
Malignant neoplasm of bronchus and lung	162.1	13	859	0.0151
Malignant neoplasm of bones of skull and face	170.0	3	6	0.5000
Malignant neoplasm of connective and other soft tissue, trunk	171.1	4	4	1.0000
Malignant melanoma of skin, site unspecified	172.9	3	39	0.0769
Malignant neoplasm of breast	174	5	409	0.0122
Malignant neoplasm of ovary	183.0	5	98	0.0510
Malignant neoplasm of prostate	185	8	239	0.0335
Malignant neoplasm of kidney, except pelvis	189.0	3	28	0.1071
Malignant neoplasm of abdomen, intra-abdominal cancer	195.0	3	21	0.1429
Malignant neoplasm of ill-defined sites, other	195.9	6	15	0.4000
Secondary malignant neoplasm of lung	197.0	4	111	0.0360
Secondary malignant neoplasm of large intestine and rectum	197.5	6	21	0.2857
Secondary malignant neoplasm of peritoneum	197.6	3	18	0.1667
Secondary malignant neoplasm of liver	197.7	4	15	0.2667
Secondary malignant neoplasm of liver, unspecified	197.8	3	40	0.0750
Secondary malignant neoplasm of other digestive organs	197.9	7	19	0.3684
Other secondary malignant neoplasm, brain	198.3	4	116	0.0345
Other secondary malignant neoplasm, bone	198.5	7	87	0.0805
Other secondary malignant neoplasm, other sites	198.9	21	121	0.1736
Malignant neoplasm without specification of site, multiple	199.0	30	1,075	0.0279
Malignant neoplasm without specification of site, other	199.1	15	39	0.3846
Other benign neoplasm of muscular and connective tissue	215	3	6	0.5000
Neoplasm of unspecified nature of other and unspecified sites	239.9	4	4	1.0000
Diabetes mellitus, without mention of acidosis or coma	250.9	13	1,313	0.0099
Other nutritional deficiency, other and unspecified	269.9	6	190	0.0316
Anemia, unspecified	285.9	21	166	0.1265
Unspecified psychosis	299	5	55	0.0909
Unspecified neurosis	300.9	4	4	1.0000
Cerebral spastic infantile paralysis, other and unspecified	343.9	4	6	0.6667
Other cerebral paralysis, hemiplegia	344.1	8	194	0.0412
Other diseases of brain, other and unspecified	347.9	10	226	0.0442
Blindness, both eyes, not specifically defined	379.1	3	12	0.2500
Essential benign hypertension	401	13	1,082	0.0120

See footnote at end of table.

Description and code	Number of occurrences in multiple sets	Number of occurrences in all sets	Index of confusion[1]
Hypertensive renal disease ... 403	4	77	0.0519
Acute myocardial infarction, without mention of hypertensive disease ... 410.9	33	3,508	0.0094
Other acute and subacute forms of ischemic heart disease, without mention of hypertensive disease ... 411.9	4	160	0.0250
Cardiovascular heart disease with hypertension ... 412.2	3	169	0.0178
Chronic ischemic heart disease without mention of hypertension ... 412.3	29	4,598	0.0063
Cardiovascular heart disease without mention of hypertension ... 412.4	17	1,389	0.0122
Chronic disease, of other endocardial structures ... 424.9	6	9	0.6667
Pulmonary heart disease ... 426	5	108	0.0463
Symptomatic heart disease, congestive heart failure ... 427.0	10	1,467	0.0068
Symptomatic heart disease, cardiac arrest NOS ... 427.2	3	1,241	0.0024
Symptomatic heart disease, other heart block ... 427.3	4	85	0.0471
Symptomatic heart disease, atrial fibrillation or flutter ... 427.4	3	144	0.0208
Symptomatic heart disease, other and unspecified disorders of heart rhythm ... 427.9	6	282	0.0213
Other myocardial insufficiency ... 428	5	204	0.0245
Ill-defined heart disease, other and unspecified ... 429.9	21	446	0.0471
Subarachnoid hemorrhage, without mention of hypertension ... 430.9	5	104	0.0481
Cerebral hemorrhage, without mention of hypertension ... 431.9	7	423	0.0165
Occlusion of precerebral arteries, without mention of hypertension ... 432.9	4	9	0.4444
Cerebral thrombosis, without mention of hypertension ... 433.9	9	678	0.0133
Acute but ill-defined cerebrovascular disease, without mention of hypertension ... 436.9	15	1,344	0.0112
Generalized ischemic cerebrovascular disease, without mention of hypertension ... 437.9	22	896	0.0246
Other and ill-defined cerebrovascular disease, without mention of hypertension ... 438.9	8	229	0.0349
Arteriosclerosis, of arteries of the extremities ... 440.2	3	3	1.0000
Arteriosclerosis, generalized and unspecified ... 440.9	46	3,448	0.0133
Other aneurysm ... 442	5	13	0.3846
Other peripheral vascular disease, unspecified ... 443.9	4	16	0.2500
Arterial embolism and thrombosis, unspecified arteries ... 444.9	5	16	0.3125
Gangrene not elsewhere classified ... 445.9	6	72	0.0833
Other diseases of arteries and arterioles ... 447	13	31	0.4194
Pulmonary embolism and infarction ... 450	7	464	0.0151
Other venous embolism and thrombosis ... 453	5	36	0.1389
Other and unspecified circulatory diseases ... 458.9	15	86	0.1744
Pneumococcal pneumonia ... 481	5	125	0.0400
Bronchopneumonia, unspecified ... 485	4	832	0.0048
Pneumonia, unspecified ... 486	14	877	0.0160
Bronchitis, unqualified ... 490	6	30	0.2000
Emphysema ... 492	6	504	0.0119
Pulmonary congestion and hypostasis ... 514	5	376	0.0133
Other diseases of lung ... 519.2	16	146	0.1096
Chronic obstructive lung disease without mention of asthma, bronchitis, or emphysema ... 519.3	7	331	0.0211
Other diseases of respiratory system, other ... 519.9	3	16	0.1875
Ulcer of stomach, with perforation only ... 531.1	4	12	0.3333
Ulcer of stomach, other and unspecified ... 531.9	5	45	0.1111
Other and unspecified diseases of stomach and duodenum ... 537.9	4	4	1.0000
Other hernia of abdominal cavity without mention of obstruction, of other specified site ... 551.8	4	4	1.0000
Other hernia of abdominal cavity without mention of obstruction, of unspecified site ... 551.9	3	4	0.7500
Intestinal obstruction, other and unspecified ... 560.9	4	30	0.1333
Diverticula of intestine, colon ... 562.1	4	14	0.2857
Perforation or ulceration of intestine NOS ... 569.4	3	12	0.2500
Other diseases of intestines and peritoneum ... 569.9	11	297	0.0370
Cirrhosis of liver, other specified ... 571.8	3	117	0.0256
Cirrhosis of liver, unspecified ... 571.9	5	305	0.0163
Other diseases of liver, unspecified ... 573.9	10	296	0.0338
Other diseases of gallbladder and biliary ducts, unspecified ... 576.9	5	17	0.2941
Other pyelonephritis, pyelitis, and pyelocystitis ... 590.1	3	47	0.0638
Other renal disease ... 593.2	14	360	0.0389
Other cellulitis and abscess, of other, multiple, and unspecified sites ... 682.9	4	5	0.8000

See footnote at end of table.

Description and code	Number of occurrences in multiple sets	Number of occurrences in all sets	Index of confusion[1]
Other local infections of skin and subcutaneous tissue....686.9	8	12	0.6667
Osteoarthritis....713.0	4	10	0.4000
Other diseases of bone, spontaneous fracture....723.2	5	6	0.8333
Other diseases of bone....723.9	3	4	0.7500
Unspecified anomalies of brain, spinal cord, and nervous system....743.9	4	5	0.8000
Unspecified anomalies of heart....746.9	4	20	0.2000
Other specified anomalies of circulatory system....747.8	3	6	0.5000
Birth injury without mention of cause, to brain....772.0	3	16	0.1875
Other conditions of fetus or newborn, other....778.9	3	77	0.0390
Convulsions....780.2	4	10	0.4000
Encephalopathy....781.7	4	33	0.1212
Acute heart failure, undefined....782.4	6	734	0.0082
Syncope or collapse....782.5	4	9	0.4444
Shock without mention of trauma....782.9	5	261	0.0192
Dyspnea....783.2	3	8	0.3750
Nausea and vomiting....784.1	3	14	0.2143
Hematemesis....784.5	6	38	0.1579
Electrolyte disorders....788.0	3	273	0.0293
Debility and undue fatigue....790.1	8	92	0.0870
Depression....790.2	4	6	0.6667
Uremia....792	20	460	0.0435
Senility without mention of psychosis....794	9	387	0.0233
Any sudden death, age 2 years and over....795.2	6	23	0.2609
Other ill-defined conditions....796.0	35	889	0.0394
Died without sign of disease....796.3	6	148	0.0405
Unknown cause of morbidity and mortality....796.8	10	206	0.0485
Other and unqualified skull fractures....N803	3	110	0.0273
Fracture of rib(s), sternum, and larynx....N807	6	52	0.1154
Fracture of neck of femur....N820	5	246	0.0203
Fracture of unspecified bones....N829	5	29	0.1724
Cerebral laceration and contusion....N851	4	78	0.0513
Subarachnoid, subdural, and extradural hemorrhage, following injury (without mention of cerebral laceration or contusion)....N852	3	50	0.0600
Injury to other and unspecified intrathoracic organs....N862	3	84	0.0357
Injury to other and unspecified intra-abdominal organs....N868	5	56	0.0893
Multiple open wounds of other and unspecified location....N907	6	174	0.0345
Contusion of other, multiple, and unspecified sites....N929	7	21	0.3333
Foreign body in pharynx and larynx....N933	5	105	0.0476
Adverse effect of agents primarily affecting blood constituents, anticoagulants....N964.2	3	4	0.7500
Adverse effect of other central nervous system depressants, anesthetic gases....N968.1	3	8	0.3750
Other specified drugs, not elsewhere classified....N977.8	8	8	1.0000
Toxic effect of freons....N987.4	4	4	1.0000
Toxic effect of other substances chiefly nonmedical as to source, other....N989.9	5	13	0.3846
Asphyxiation and strangulation....N994.7	4	108	0.0370
Injury, other and unspecified, knee, leg, ankle, and foot....N996.7	4	4	1.0000
Injury, other and unspecified, unspecified site....N996.9	6	73	0.0822
Postoperative hemorrhage or hematoma....N998.1	8	22	0.3636
Other complications of surgical procedures....N998.9	26	206	0.1262
Accidental poisoning by other gases and vapors....E876	4	4	1.0000
Other and unspecified fall....E887	6	142	0.0423
Inhalation and ingestion of food causing obstruction or suffocation....E911	5	56	0.0893
Inhalation and ingestion of other object causing obstruction or suffocation....E912	5	38	0.1316
Complications and misadventures in surgical treatment, excluding effects of anesthetic management....E930.0	32	261	0.1226
Complications and misadventures in administration of drugs and biologicals....E931.1	3	15	0.2000
Late effect of surgical operation....E947	6	13	0.4615
Fight, brawl, rape....E960	5	5	1.0000
Assault by other and unspecified means....E968	5	24	0.2083

[1]Index = $\frac{\text{number of occurrences in multiple sets}}{\text{number of occurrences in all sets}}$

NOTE: NOS = not otherwise specified.

APPENDIX II

TECHNICAL NOTES

Estimation

The medical coding study was based on a two-stage sample of 8,973 death certificates processed during data year 1974. These certificates comprised the 10-percent quality control sample from 30 of the 472 work lots processed through the three-way verification system from July 1974 through March 1975.

As mentioned in the text, the certificates chosen for the study are reasonably representative of the content of U.S. death certificates in general, since disease patterns tend to change gradually over time. However, the statistics in this report also reflect the prevailing conditions in the medical coding unit at the time of the study. The magnitude of the statistics can be expected to change over time for many reasons, including increased experience of veteran coders, introduction of new coders into the unit, changes in coding instructions, and, in every 10th year, the revision of the International Classification of Diseases. However, the primary purpose of the study was to investigate certain relationships between these statistics—relationships that will persist even though the magnitude of the statistics may change.

All aggregates shown in this report are observed totals for the 8,973 sample records and are discussed in descriptive rather than inferential terms. Error rates, agreement rates, and percents are presented as estimates for the underlying population. Since the sample was self-weighting, these ratio statistics were computed directly from the raw totals for the numerator and denominator.

Sampling Errors

Variance estimator.—For the purpose of computing standard errors of estimates, the 30 sample lots were treated as though they were selected by simple random sampling, although the frame was stratified by production coding error rate prior to selection, and finite population correction factors were ignored. These simplifications tend to cause overestimates of sampling variance.

An ultimate cluster estimator was used to compute standard errors for error rates, agreement rates, and percents. The ultimate cluster estimate of the variance of a ratio estimate $r = x'/y'$, where x' and y' are aggregate estimates, is given by[14]

$$s_r^2 \doteq r^2 \left(v_{x'}^2 + v_{y'}^2 - 2v_{x'y'}\right)$$

where

$$v_{x'}^2 = \frac{\sum_{i=1}^{m} (x_i' - \bar{x})^2}{m\,(m-1)\,\bar{x}^2} = \frac{\sum_{i=1}^{m} (x_i')^2 - m\bar{x}^2}{m\,(m-1)\,\bar{x}^2}$$

$$v_{y'}^2 = \frac{\sum_{i=1}^{m} (y_i' - \bar{y})^2}{m\,(m-1)\,\bar{y}^2} = \frac{\sum_{i=1}^{m} (y_i')^2 - m\bar{y}^2}{m\,(m-1)\,\bar{y}^2}$$

$$v_{x'y'} = \frac{\sum_{i=1}^{m} (x_i' - \bar{x})(y_i' - \bar{y})}{m\,(m-1)\,\bar{x}\bar{y}} = \frac{\sum_{i=1}^{m} x_i'y_i' - m\bar{x}\bar{y}}{m\,(m-1)\,\bar{x}\bar{y}}$$

$v_{x'}^2$ = rel-variance of x'.

$v_{y'}^2$ = rel-variance of y'.

$v_{x'y'}$ = rel-covariance of x' and y'.

x_i' = estimate of x for sample lot i.

NOTE: A list of references follows the text.

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i'}{m}$$, the mean of the sample lot totals of x.

y_i' = estimate of y for sample lot i.

$$\bar{y} = \frac{\sum_{i=1}^{m} y_i'}{m}$$, the mean of the sample lot totals of y.

m = 30, the number of sample lots.

Presentation of standard errors.—Because of resource limitations, it was not feasible to compute variances for all estimates given in the report. Instead, variances were computed for some key statistics shown in tables I-III. Sampling errors for other proportions, error rates, and agreement rates can be approximated by using the design effects (DEFF's) shown in the tables, using the formula

$$\sigma_r^2 = \frac{[\text{DEFF}(\bar{r})]^2\, r\,(1-r)}{n}$$

where

r = the statistic of interest.

n = the number of sample codes in the denominator of r.

DEFF = the design effect for a statistic $\bar{r}$ in tables that is the same type of statistic as r (error rate, agreement rate, or proportion), of approximately the same magnitude as r, with a similar sample size in the denominator. DEFF ($\bar{r}$) is defined as the ratio of the estimated standard error of $\bar{r}$ based on the study data to the expected value of the standard error of $\bar{r}$ under the assumptions of simple random sampling from a binomial population.

Table I. Sampling errors and design effects for condition code agreement rates and underlying cause-of-death code agreement rates for coders 6 and 7, by multiple set status

Coder(s) who entered multiple sets	Condition agreement				Underlying cause agreement (each cause level)			
	Agreement rate (percent)	Denominator (number of codes)	Sampling error (percent)	Design effect	Agreement rate (percent)	Denominator (number of records)	Sampling error (percent)	Design effect
All records	97.77	27,752	0.16	1.77	98.85	8,973	0.13	1.15
Neither coder	98.14	24,743	0.15	1.77	98.95	8,289	0.14	1.21
Coder 6 only	95.23	1,215	0.74	1.20	97.50	280	0.99	1.06
Coder 7 only	93.81	1,244	1.06	1.55	97.24	290	0.91	0.95
Both coders	95.45	550	0.91	1.02	99.12	114	0.88	1.00

Table II. Sampling errors and design effects for the production coder's condition code error rate and underlying cause-of-death code error rate, by expert coders

Expert coder	Condition coding				Underlying cause coding (each cause level)			
	Error rate (percent)	Denominator (number of codes)	Sampling error (percent)	Design effect	Error rate (percent)	Denominator (number of records)	Sampling error (percent)	Design effect
Coder 6	3.60	27,622	0.24	2.15	1.87	8,973	0.18	1.27
Coder 7	4.27	27,630	0.27	2.21	2.21	8,973	0.19	1.22

Table III. Sampling errors and design effects of condition code agreement rates and underlying cause-of-death code agreement rates for coder 7, by the three original coders for selected agreement status categories

Agreement status category	Condition agreement				Underlying cause agreement (each cause level)			
	Agreement rate (percent)	Denominator of rate (number of codes)	Sampling error (percent)	Design effect	Agreement rate (percent)	Denominator of rate (number of records)	Sampling error (percent)	Design effect
AAA cases								
Coder 7 agreed with A only	98.48	25,521	0.14	1.77	99.21	8,488	0.09	0.91
Coder 7 agreed with B only	1.11	25,521	0.09	1.32	0.79	8,488	0.09	0.91
Coder 7 agreed with no one[1]	0.41	25,521	0.06	1.62	...	...	...	...
AAB cases								
Coder 7 agreed with A only	74.02	2,556	1.36	1.56	74.51	455	2.25	1.10
Coder 7 agreed with B only	17.57	2,556	1.13	1.55	16.48	455	2.06	1.19
Coder 7 agreed with A and B	5.79	2,556	0.79	1.71	7.03	455	1.46	1.22
Coder 7 agreed with no one	2.62	2,556	0.30	0.95	*1.98	455	0.57	0.88
ABC cases								
Coder 7 agreed with 1 coder	53.09	194	4.71	1.31	83.33	30	8.15	1.20
Coder 7 agreed with 2 coders	32.99	194	3.14	0.93	-	-	-	-
Coder 7 agreed with all coders	*4.12	194	2.22	1.55	-	-	-	-
Coder 7 agreed with no one	9.79	194	2.22	1.04	*16.67	30	8.15	1.20

[1]No code by coder 7.

Variance of the difference between two rates.—In general, the variance of the difference between two error rates, agreement rates, or proportions is given by

$$\sigma^2_{r_1-r_2} = \sigma^2_{r_1} + \sigma^2_{r_2} - 2\sigma_{r_1r_2}$$

For this study, the covariance term in the equation was considered to be 0. There is probably some correlation between the results for various coding assignments, because all coding assignments were performed on the same work lots, and the same set of coders acted as coders 6 and 7. The magnitude of the effect is unknown, but it seems likely that any correlations would be positive. Therefore $\sigma^2_{r_1-r_2} = \sigma^2_{r_1} + \sigma^2_{r_2}$ will tend to overstate the true variance.

Nonsampling Errors and Sources of Bias

Several precautions were taken in the design and conduct of the study to minimize the effects of nonsampling error and bias on the results. The same group of six coders was assigned to the dependent and independent expert assignments to reduce the effects of variability between coders on the data based on these assignments. After the six coders had completed their work, the differences between the work of coders 6 and 7 were reviewed for all lots. When, in the judgment of the supervisor of the coding unit, a difference was due to keypunch error by either expert, the code was changed so that the work of coders 6 and 7 would be as close to the truth as possible. The remaining differences were examined and errors were charged to coder 6 and/or coder 7 by the adjudication process described in the text. These errors were subtracted in calculating error rates based on the work of coders 6 and 7.

Despite these precautions, some nonsampling error and bias are present in the study. The 2.23-percent overall difference rate between coders 6 and 7 (table D) was probably about as small as could be expected. There is no question, however, that the method of apportioning the difference rate as error rates to coders 6 and 7 was biased in favor of coder 6, since she had access to the work of 2 of the 4 coders whose codes were the basis for assigning errors when coders 6 and 7 did not agree. Because supervisory coding time was not available, the only other method of apportioning the errors would have been to assign them arbitrarily, say, half to coder 6 and half to coder 7. Under these circumstances the true production coding error rate based on the work of coder 7 would have been 4.25 percent (5.37-1.12) and the rate based on adjudication of differences by coder 6 would have been 3.00 percent (4.11-1.11).

Additional problems arose when the expert coders were asked to identify ambiguous coding situations and enter the alternative solutions. Individually and as a group the experts identified about the same number of ambiguities in their coder-6 roles as they did in their coder-7 roles (table O). However, there was considerable variation between the coders in the number of ambiguous situations identified. More important, coders 6 and 7 generally did not agree on which records contained ambiguous situations. Thus it is difficult to make inferences about the true prevalence of ambiguous coding situations and the circumstances that cause ambiguity, since the expert coders themselves do not agree on these issues.

Another source of bias in this study is the method of dependent verification that was used. Although the dependent verification system compared favorably in terms of estimating quality with the independent verification system, numerous studies have shown the dependent system to be inferior. In general, the probability that a dependent verifier will change a correct code to an incorrect code is quite small; however, the probability that she will fail to change an incorrect code can be substantial, depending on the complexity of the operation being verified. Table IV, from a study by Minton,[4] shows the percent of errors detected in a variety of coding operations. The sample verification of mortality medical coding is at least as complex as the sample verification of industry and occupation coding, in which 70 percent (100-30) of the errors were not corrected during dependent verification.

The results of an earlier study by McKeon[2] may shed some light on the favorable showing of dependent verification in this study. It was pointed out in the text that the knowledge that they were working on a special project may have influenced the dependent coders to work more

NOTE: A list of references follows the text.

Table IV. Quality of dependent verification in selected data processing activities

Operation	Unit defective[1]	Percent of defectives detected
Matching of sample cards to questionnaire:		
1960: General coding	Coded item	96
1960: Industry and occupation coding	Coded item	99
Keypunching of data cards:		
1950: 100-percent verification	Punched data card	97
Sample verification	Punched data card	96
1963: 100-percent verification	Punched data card	86
Typing data:		
1963: 100-percent verification	Typed address card	78
1962: 100-percent verification	Survey questionnaire	58
Coding responses:		
1960: General coding (sample verification)	Coded item	50
1960: Industry and occupation coding (sample verification)	Coded item	30
1960: Industry and occupation coding (correction of rejected work)	Coded item	70
Rounding-coding-typing:		
1963: 100-percent and sample verification	Typed questionnaire	10

[1]A defective unit may contain one or more error types.

carefully than they would have under normal circumstances. The McKeon study measured the quality of dependent verification by planting errors in the work to be verified and counting the number of errors remaining after verification. After the verifiers were told that errors were being planted, undetected errors decreased by more than 50 percent, suggesting that the "awareness" of the special procedure was an assignable cause of the improvement in the dependent verification system.

APPENDIX III

U.S. STANDARD CERTIFICATE OF DEATH

(PHYSICIAN, MEDICAL EXAMINER OR CORONER)
U.S. STANDARD
CERTIFICATE OF DEATH

Form Approved
OMB No. 68R 1901

LOCAL FILE NUMBER — STATE FILE NUMBER

TYPE OR PRINT IN PERMANENT INK
FOR INSTRUCTIONS SEE HANDBOOK

DECEDENT

1. DECEDENT-NAME FIRST MIDDLE LAST
2. SEX
3. DATE OF DEATH (Mo. Day, Yr.)
4. RACE-(e.g., White, Black, American Indian, etc.) (Specify)
5a. AGE-Last Birthday (Yrs)
5b. UNDER 1 YEAR MOS. DAYS
5c. UNDER 1 DAY HOURS MINS.
6. DATE OF BIRTH (Mo. Day, Yr.)
7a. COUNTY OF DEATH
7b. CITY, TOWN OR LOCATION OF DEATH
7c. HOSPITAL OR OTHER INSTITUTION-Name (If not in either, give street and number)
7d. IF HOSP OR INST Indicate DOA, OP/Emer Rm, Inpatient (Specify)
8. STATE OF BIRTH (If not in U.S.A. name country)
9. CITIZEN OF WHAT COUNTRY
10. MARRIED, NEVER MARRIED, WIDOWED, DIVORCED (Specify)
11. SURVIVING SPOUSE (If wife give maiden name)
12. WAS DECEDENT EVER IN U.S. ARMED FORCES? (Specify Yes or No)
13. SOCIAL SECURITY NUMBER
14a. USUAL OCCUPATION (Give kind of work done during most of working life, even if retired)
14b. KIND OF BUSINESS OR INDUSTRY

IF DEATH OCCURRED IN INSTITUTION SEE HANDBOOK REGARDING COMPLETION OF RESIDENCE ITEMS

15a. RESIDENCE-STATE
15b. COUNTY
15c. CITY, TOWN OR LOCATION
15d. STREET AND NUMBER
15e. INSIDE CITY LIMITS (Specify Yes or No)

PARENTS

16. FATHER-NAME FIRST MIDDLE LAST
17. MOTHER-MAIDEN NAME FIRST MIDDLE LAST
18a. INFORMANT-NAME (Type or Print)
18b. MAILING ADDRESS STREET OR R.F.D. NO. CITY OR TOWN STATE ZIP

DISPOSITION

19a. BURIAL, CREMATION, REMOVAL, OTHER (Specify)
19b. CEMETERY OR CREMATORY-NAME
19c. LOCATION CITY OR TOWN STATE
20a. FUNERAL SERVICE LICENSEE Or Person Acting As Such (Signature)
20b. NAME OF FACILITY
20c. ADDRESS OF FACILITY

CERTIFIER

To be Completed by CERTIFYING PHYSICIAN Only

21a. To the best of my knowledge, death occurred at the time, date and place and due to the cause(s) stated (Signature and Title)
21b. DATE SIGNED (Mo. Day Yr.)
21c. HOUR OF DEATH M
21d. NAME OF ATTENDING PHYSICIAN IF OTHER THAN CERTIFIER (Type or Print)

To be Completed by MEDICAL EXAMINER or CORONER Only

22a. On the basis of examination and/or investigation, in my opinion death occurred at the time, date and place and due to the cause(s) stated (Signature and Title)
22b. DATE SIGNED (Mo. Day, Yr.)
22c. HOUR OF DEATH M
22d. PRONOUNCED DEAD (Mo., Day, Yr.) ON
22e. PRONOUNCED DEAD (Hour) AT M

23. NAME AND ADDRESS OF CERTIFIER (PHYSICIAN, MEDICAL EXAMINER OR CORONER) (Type or Print)
24a. REGISTRAR (Signature)
24b. DATE RECEIVED BY REGISTRAR (Mo. Day, Yr.)

CAUSE OF DEATH

CONDITIONS IF ANY WHICH GAVE RISE TO IMMEDIATE CAUSE STATING THE UNDERLYING CAUSE LAST

25. IMMEDIATE CAUSE (ENTER ONLY ONE CAUSE PER LINE FOR (a), (b), AND (c))

PART I
(a) — Interval between onset and death
DUE TO, OR AS A CONSEQUENCE OF
(b) — Interval between onset and death
DUE TO, OR AS A CONSEQUENCE OF
(c) — Interval between onset and death

PART II OTHER SIGNIFICANT CONDITIONS-Conditions contributing to death but not related to cause given in PART I (a)

26. AUTOPSY (Specify Yes or No)
27. WAS CASE REFERRED TO MEDICAL EXAMINER OR CORONER (Specify Yes or No)
28a. ACC., SUICIDE, HOM., UNDET. OR PENDING INVEST. (Specify)
28b. DATE OF INJURY (Mo. Day, Yr.)
28c. HOUR OF INJURY M
28d. DESCRIBE HOW INJURY OCCURRED
28e. INJURY AT WORK (Specify Yes or No)
28f. PLACE OF INJURY-At home, farm, street, factory, office building, etc. (Specify)
28g. LOCATION STREET OR R.F.D. No. CITY OR TOWN STATE

HRA-162-1
Rev. 1/78

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE - PUBLIC HEALTH SERVICE - NATIONAL CENTER FOR HEALTH STATISTICS
1978 REVISION

☆U.S. GOVERNMENT PRINTING OFFICE: 1980 311-240/7 1-3

VITAL AND HEALTH STATISTICS Series

Series 1. *Programs and Collection Procedures.*—Reports which describe the general programs of the National Center for Health Statistics and its offices and divisions and data collection methods used and include definitions and other material necessary for understanding the data.

Series 2. *Data Evaluation and Methods Research.*—Studies of new statistical methodology including experimental tests of new survey methods, studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory.

Series 3. *Analytical Studies.*—Reports presenting analytical or interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.

Series 4. *Documents and Committee Reports.*—Final reports of major committees concerned with vital and health statistics and documents such as recommended model vital registration laws and revised birth and death certificates.

Series 10. *Data From the Health Interview Survey.*—Statistics on illness, accidental injuries, disability, use of hospital, medical, dental, and other services, and other health-related topics, all based on data collected in a continuing national household interview survey.

Series 11. *Data From the Health Examination Survey and the Health and Nutrition Examination Survey.*—Data from direct examination, testing, and measurement of national samples of the civilian noninstitutionalized population provide the basis for two types of reports: (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.

Series 12. *Data From the Institutionalized Population Surveys.*—Discontinued effective 1975. Future reports from these surveys will be in Series 13.

Series 13. *Data on Health Resources Utilization.*—Statistics on the utilization of health manpower and facilities providing long-term care, ambulatory care, hospital care, and family planning services.

Series 14. *Data on Health Resources: Manpower and Facilities.*—Statistics on the numbers, geographic distribution, and characteristics of health resources including physicians, dentists, nurses, other health occupations, hospitals, nursing homes, and outpatient facilities.

Series 20. *Data on Mortality.*—Various statistics on mortality other than as included in regular annual or monthly reports. Special analyses by cause of death, age, and other demographic variables; geographic and time series analyses; and statistics on characteristics of deaths not available from the vital records based on sample surveys of those records.

Series 21. *Data on Natality, Marriage, and Divorce.*—Various statistics on natality, marriage, and divorce other than as included in regular annual or monthly reports. Special analyses by demographic variables; geographic and time series analyses; studies of fertility; and statistics on characteristics of births not available from the vital records based on sample surveys of those records.

Series 22. *Data From the National Mortality and Natality Surveys.*—Discontinued effective 1975. Future reports from these sample surveys based on vital records will be included in Series 20 and 21, respectively.

Series 23. *Data From the National Survey of Family Growth.*—Statistics on fertility, family formation and dissolution, family planning, and related maternal and infant health topics derived from a biennial survey of a nationwide probability sample of ever-married women 15-44 years of age.

For a list of titles of reports published in these series, write to:

Scientific and Technical Information Branch
National Center for Health Statistics
Public Health Service
Hyattsville, Md. 20782

DHEW Publication No. (PHS) 80-1355

Series 2-No. 81

NCHS

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Office of Health Research, Statistics, and Technology
National Center for Health Statistics
3700 East-West Highway
Hyattsville, Maryland 20782

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF H.E.W.

HEW 396

THIRD CLASS

U.S. MAIL

For publications in the *Vital and Health Statistics* Series call 301-436-NCHS.