STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

Working Paper No. 10
English only

CONFERENCE OF EUROPEAN STATISTICIANS

UN/ECE Work Session on Statistical Data Editing
(Cardiff, United Kingdom, 18-20 October 2000)

Topic I: Management and evaluation of editing and imputation procedures

THE ROLE OF DATA EDITING IN THE QUALITY CONTROL PROGRAM FOR THE NATIONAL CENTER FOR HEALTH STATISTICS

Submitted by the National Center For Health Statistics, United States [1]

Contributed paper

I. INTRODUCTION

1. The National Center for Health Statistics (NCHS) is the Federal agency responsible for the collection and dissemination of the nation's vital and health statistics. To carry out its mission, NCHS conducts a wide range of annual, periodic, and longitudinal sample surveys and administers the national vital statistics registration systems. These sample surveys and registration systems form four families of data systems: vital event registration systems, population based surveys, provider based surveys, and followup/followback surveys.

2. A primary objective of the Center's Quality Control Program is to assure that the products of its sample surveys and registration systems meet standards of reliability and validity. The NCHS Quality Control Program is defined as the totality of functions for measuring and evaluating the quality of statistics produced by the Center. Thus, quality control is interpreted broadly to include both statistical surveillance on a continuing basis and data evaluation on a special study basis [1].

3. The Center's Quality Control Program has undergone a number of revisions over the years. While its primary objective remains the same, changes in available resources, standards, data requirements etc., have necessitated periodic adjustments in the program.

II. DEVELOPMENT OF STATISTICAL STANDARDS

4. Statistical surveillance and data evaluation must include a standard by which the particular process or operation is to be measured. Statistical standards are usually represented by the error levels that are tolerated (i.e., acceptable) in carrying out the various operations of a data system.

5. Standards for non-response in data collection and for coding errors in data processing are examples. On the other hand, statistical standards also specify how to perform operations of a data system. For example, specifications for the presentation of errors in statistical reports [2] and procedures for protecting the confidentiality of the data are also referred to as statistical standards. In both cases, the standards serve primarily to assure the quality and integrity of the compiled statistics.

6. Responsibilities of statistical standards programs are not limited to the development of the appropriate statistical standards. Typically, statistical standards require estimates of error rates that are not produced as by-products of the data systems themselves, but instead must be derived from quality

[1] Prepared by Kenneth W. Harris.

control systems. The design of the quality control systems, then, is a part of the overall statistical standards program.

III. FUNCTIONS OF THE QUALITY CONTROL PROGRAM

7. When NCHS first developed a policy statement on its Quality Control Program (1971), it identified six functions that the program should satisfy. Despite the many procedural changes that have occurred since then, these functions, described below, are still viable.

(i) **Design of the surveillance system** - A surveillance system for monitoring the data operations is an integral and essential part of the initial design. There are usually sub-systems to evaluate each operation of the measurement process, such as sample selection, data collection, data processing, data analysis, etc.

(ii) **Evaluation of the survey** - Basically, the statistics produced by the continuing surveillance system are reviewed from three viewpoints: (a) the initial test of assuring an acceptable quality level, (b) an ongoing review to assure that the measurement equipment and processes are under continuous statistical control and meet the established quality standards, including a provision for courses of action when the review indicates one or more processes are no longer in control, (c) an intensive periodic review to observe trends in the quality of statistics, to assure that the surveillance system itself is operating adequately, and to decide on possible changes in that system. In addition, the surveillance system's by-product statistics are often helpful in estimating components of non-sampling errors, thus providing a valuable service in interpreting the statistics derived from the survey operations.

(iii) **Review statistical reports** - The objective is to assure the statistical adequacy of published reports. This function includes: (a) verifying the statistical statements and conclusions presented in the text, (b) checking the analytical methods to assure that they are appropriate and (c) reviewing the statistical appendixes to assure that they contain the estimates of sampling and non-sampling errors required to verify the statistical conclusions presented in the text and to interpret the statistics presented in the summary tables.

(iv) **Conduct experimental studies** - Experimental studies are conducted to obtain estimates of the components of measurement error and of costs of the operations that cannot be derived as by-products of the regular surveillance system. These studies are conducted either independently of, or as adjuncts to, the regular data collection system.

(v) **Conduct research on quality control systems** - Research is conducted to improve the efficiency and effectiveness of quality control systems. This includes both applied and theoretical research.

(vi) **Train statisticians** - An important, though often overlooked, function of any quality control program is the training of analytical statisticians in the principles of statistical inference and the use and understanding of quality control methodology.

8. Each of these six functions is served by the Center's Quality Control Program, though the degree of service varies between functions and, sometimes, varies for the same function between data systems. In general, the allocation of Center resources to these functions is roughly in the order in which they are listed. Functions (i) and (ii) are at the highest level of involvement; functions (iii), (iv) and (v) are at a lower level, and function (vi), though improving, is at the bottom. With regard to NCHS areas of responsibility, functions (i), (ii), and (v) tend to be centralized in the Center's major methodological office, while the remaining functions are much more decentralized and, thus, more subject to fluctuations in levels of effort devoted to them.

IV. MEASUREMENT PROCESSES

9. As stated earlier, a primary objective of the Center's Quality Control Program is to assure that the statistical products of its data systems meet standards of reliability and validity. Although the operations or measurement processes of these data systems can be described in varying levels of detail, we generally identify three major measurement processes: data collection, data processing, and data analysis. Of the three, data collection is the one we have historically known the least about from the viewpoint of measurement error. Because of the many possible sources of error, there is an inherent difficulty in designing quality control procedures for data collection. Tremendous strides have been made in recent years, however, particularly in the area of data editing. Within the NCHS Quality Control Program, data editing is defined as:[3]

> any procedure designed and used for detecting erroneous and/or questionable survey data (survey response data or identification type data) with the goal of correcting (manually and/or via electronic means) as much of the erroneous data as possible, usually prior to data imputation and summary procedures.

11. The increasing use of Computer Assisted Personal Interviewing (CAPI) and Computer Assisted Telephone Interviewing (CATI), for example, have done much to reduce data collection error, as well as improve our capacity to measure it. CAPI and CATI have introduced a number of technical enhancements that provide data editing functions at the very beginning of the data collection process. Some examples are automated skip patterns, signals to indicate when invalid or inconsistent codes/items are entered, etc.

12. Concurrent with the increasing uses of CAPI and CATI is the Center's greater reliance on "Source Point Data Editing" (SPDE). This refers to editing survey data by any means of access to either the interviewer (or other data collector), the respondent, or records within a limited time following the original interview or data collection. The time limit reflects the period within which the person(s) involved can reasonably be expected to remember details of the specific interview or, in the case of data collected from records, a time within which there is a reasonable expectation that there has been no change to the records which would affect the data collected. Thus, data completion and accuracy are much more likely to result when source point data editing is used.

13. With regard to data analysis, the basic difficulty is the lack of uniform proficiency in the use of analytical methods. This difficulty has been lessened by the development of protocols, one of which is a manual [2] that describes when and how to use existing analytical methods to verify statistical statements and how to assure that statistical appendixes contain appropriate estimates of sampling and non-sampling errors required to verify the statistical conclusions presented in Center reports. Of more importance, perhaps, is the development of classes taught by NCHS methodologists and selected contractors on the uses of computer software packages such as SUDAAN (SUrvey DAta ANalysis). This package provides for analyses of complex sample survey designs and includes modeling procedures such as regression, logistic regression and survival analysis. One of its values as a data editing tool is that it allows analysts to specify how data are correlated and weighted. This is a very important feature in analyzing NCHS surveys, most of which are multi-stage sample surveys and, thus, frequently produce correlated data.

14. A data editing procedure that is used as a preliminary analytic tool in the analysis of mortality and natality summary statistics is the use of forecast tables. For all demographic variables, forecast tables provide standards of the acceptable percentage of item incompleteness. For example, the percentage of deaths where the marital status of the decedent is unstated or unknown. These standards are based on an upper band (usually two or three times the median of the previous year=s data); any reporting area that falls above the standard is automatically flagged for review.

Similarly, both medical conditions and causes of death that are classified as Arare@ are also automatically flagged for review.

15. Finally, the operation that is the most well developed, in terms of quality control resources, is that of data processing. Historically, the most important staff function of data processing has been coding, particularly medical coding. Although NCHS now uses an automated multiple cause of death coding system, medical coding continues to be a very important function because of the need to maintain in-house underlying cause of death coding expertise.

V. TYPES OF QUALITY CONTROL

16. There are two types of quality control: process control and acceptance sampling. Process control is designed to measure the quality of a process and to determine when that process is “out of control” and needs changing in order to get back in control. Defective work resulting from the “out of control” process is seldom corrected. The quality control procedures for the data collection of two NCHS surveys, the National Health Interview Survey (NHIS) and the National Hospital Discharge Survey (NHDS), are examples of process control. In the NHIS, poor quality work by an interviewer is not corrected, but the interviewer, who represents the process, is brought back in control through retraining. In the NHDS, if the quality control procedure determines that the work of the hospital abstractor is unacceptable, that work is not corrected, but the abstractor is retrained so that future work will be acceptable.

17. The Center’s coding operations are evaluated by the second type of quality control, i.e., acceptance sampling. Coding assignments are made on the basis of work lots or batches. Within each data system, the size of the batch is usually within some range consistent with the manner in which the data are received. For example, in the mortality medical coding unit, a batch of about 2,500-3,000 mortality records is formed from the monthly submissions of one (or more) states. The monthly submissions from larger states may form two (or more) batches. Each batch is then assigned to a production coder, who codes all of the records in the batch. The batch is then verified to determine whether it is “acceptable”, that is, if it meets a previously established quality standard, or "unacceptable”, in which case it is rejected and recoded.

VI. TYPES OF VERIFICATION

18. Virtually all of the Center’s coding operations are verified under an independent verification system, as opposed to a dependent verification system. The difference in the two systems is that the verifier in an independent system codes a batch (usually a sample) without having access to the codes of the production coder; a dependent verifier knows what the production coder has entered. Although the independent system is usually more expensive, the additional cost is often justified by its greater accuracy in measuring quality. A number of studies at the Bureau of the Census and at NCHS, to name two places, have shown that dependent verifiers tend to be biased toward the work of the production coder [4-10]. In some cases, the dependent verifier’s type II error rate, i.e., agreeing with an incorrect code, was as high as 70 percent. On the other hand, the type I error rate, i.e., changing a correct code to an incorrect code, was almost non-existent. These studies have shown that dependent verification is more effective in non-judgmental or check coding. For example, verifying sex, age, place of birth, etc. Medical coding, which requires extensive training for proficiency, is verified much more accurately under an independent system. Initially, the Center’s medical coding operations were verified under a three way independent system.

VII. THREE WAY INDEPENDENT SAMPLE VERIFICATION

19. Under this procedure, a batch of records is coded by three coders. A production coder codes all of the records and two sample coders independently code the same sample of records from the batch.

The two sets of sample records and the corresponding records from the production coder are computer matched and the majority code (i.e., coded by at least two of the coders) is selected as the correct code. Error rates for each coder are computed on this basis. If the production coder's error rate exceeds the acceptable level, the entire production batch (100 percent) is recoded and re-verified.

20. Three way independent verification has long been the preferred method by many organizations that employ difficult or extensive coding schemes in their data reduction operations. This verification system assumes that a condition or description leads to only one valid code; thus, when two or three out of three coders with comparable coding skills independently arrive at the same code, there is a high probability that the selected code is correct.

VIII. MODIFICATION OF THE THREE WAY SYSTEM

21. Although three way independent verification is considered a highly efficient system, some have felt that its cost, in light of the recognized less precise knowledge regarding measurement errors in data collection processes, is too high. As pointed out earlier, one of the functions of the Center's Quality Control Program is to conduct research to improve the efficiency and effectiveness of its quality control systems. In carrying out that function, NCHS has investigated a number of ways of reducing verification costs while maintaining the capability of deriving timely, unbiased estimates of coding quality.

A. Sequential Sampling

22. One of the systems considered early on by NCHS involved sequential sampling. Under sequential sampling, a decision to accept or reject a batch can usually be made with a sample substantially reduced from that required by regular acceptance sampling. Given four parameters,

(1) P_1 = Acceptable error rate

(2) a = Probability of rejection when true error rate is P_1 (Type I error): a=alpha

(3) P_2 = Unacceptable error rate

(4) b = Probability of acceptance when true error rate is P_2 (Type II error): b=beta

the Average Sample Number (ASN) can be computed [11]. For example, batch records from the National Hospital Discharge Survey (NHDS) were then being verified under the three-way system at a rate of 10 percent, resulting in sample batches of 250 codes (for each of two sample coders). However, based on P_1, P_2, a,b, and the actual incoming error rate, the ASN was 61. Thus the sample size, in theory, could have been reduced by almost 88 percent (500-61)/500.

23. However, there were several problems associated with the use of sequential sampling. In a production setting, where items are inspected (verified) to determine whether they meet or do not meet specifications, sequential sampling involves inspecting items in the order in which they are produced. Item number 1 is inspected first, item number 2 is inspected second, etc., until enough items have been inspected to make a decision to accept or reject the batch. In a coding operation, however, where errors are generally assumed to be non-randomly distributed throughout the batch, the use of sequential sampling would severely bias the validity of the decision-making process. Initial good quality that deteriorates significantly after a period of time would lead to an accept decision, while initial poor quality that improves significantly after a period of time would lead to a reject decision. In each case, the decision might be wrong.

24. Another problem associated with the use of sequential sampling involves estimating the quality of the batch. Since inspection begins with the first item in the batch and continues sequentially until a

decision is made to accept or reject, items coded beyond a certain point have no chance of being selected for sample inspection, thus an unbiased estimate of the error rate cannot be made.

25. In our study, these problems were resolved by modifying the sequential sampling plan to a less economical multiple sampling plan that selected the samples in m stages of size n and made the decision to accept, reject, or continue inspection only at the end of each stage. In order to obtain an unbiased estimate on the basis of each stage or group of stages, the records verified in each stage were systematically selected from the entire batch [12]. Ultimately, though, the fact that the sequential system required dependent verification in order to make an efficient decision on accepting or rejecting a batch led us to explore other possibilities.

B. Two Way Independent Verification

26. The introduction of two way independent verification in the Center's three major medical coding operations occurred in 1982 [13]. The appeal of the two-way independent system is that it also uses the majority rule concept, which is the key feature of the three way system. When at least two coders in a three way system agree on a code (AAB or AAA), then A, the majority code, is considered the correct code. In a two way system, AAA cases become AA cases and AAB cases become AA (one-third) and AB (two-thirds). The conversion procedure is shown in Figure 1. Three-way differences (ABC) are rare, but they, too, become AB cases (i.e., two non-matching codes) in the two way system.

27. In an earlier study of the quality of mortality medical coding [10], Harris and French found that AAA cases comprised 90.3 percent of all cases and AAB cases accounted for just over nine percent of all cases. At these levels, the use of a two way system would provide the same measurement precision (majority rule) for more than 93 percent of the codes [90.3 + 1/3(9)] while reducing the number of coders used from three to two. The third coder would be needed as an adjudicator only for the AB cases (6.7 percent).

28. These findings indicated that substantial resource savings could be realized, at no diminution of quality measuring capabilities, when

(1) The quality of coding within a coding unit is, in general, homogeneous.
(2) Correct coding solutions are unique.

These conditions enhance the concept of the majority rule as a valid identifier of the correct code and a reliable measure for establishing coding quality. As one might expect, however, the validity and reliability are much greater when the majority code is arrived at unanimously.

29. In the mortality study cited above, it was found that Code A was correct 98.5 percent of the time that each of three coders independently selected Code A (AAA cases) but was correct slightly less than 80 percent of the time when only two of the three coders selected it (*AAB* cases). Overall, the selection of Code A was correct about 97 percent of the time. In addition, that study found a medical description led to a unique coding solution 98 percent of the time.

30. These additional findings further supported the plan to convert from three way independent verification to two way independent verification. Since AAA cases comprised more than 91 percent of the majority rule cases (AAA + *AAB*), the AA cases in the two way system are much more likely to represent what would have been AAA cases in a three way system (only a third of the *AAB* cases will become AA cases), thus, the correctness of code A is strongly indicated. If the adjudication of the remaining AB cases yields the correct code 80 percent of the time, then the two-way system will also produce the correct code about 97 percent of the time, thereby providing a measure of coding quality very similar to that of the three way system. Several followup studies (the most recent in 1997) conducted since the introduction of two way independent verification have yielded similar results.

IX. AUTOMATED CODING SYSTEMS

31. The utility of two- way independent verification has been particularly noteworthy in one area. With the advent of MICAR and Super MICAR [14,15], the Center's multiple cause of death coding systems, the two-way system provides an additional mechanism for assessing the quality of those systems.

32. Briefly, MICAR, Mortality Medical Indexing, Classification and Retrieval, allows data entry operators to enter full text, abbreviations, or reference numbers for cause of death terms. Through a series of steps, and the use of a MICAR dictionary containing more than 100,000 entries, this information can then be automatically converted to an ICD (International Classification of Diseases) code that represents the cause of death. A data entry operator can become proficient in one month, compared to the 18 months required to become a proficient nosologist (medical coder). Despite the extensive changes introduced in 1999 when ICD-10 was implemented (ICD-9 had been in effect since 1979), MICAR processes more than 90 percent of all mortality records with a system error rate less than one percent.

33. SuperMICAR, which accepts almost everything in literal text, can be mastered even more quickly. The data entry operator enters condition descriptions exactly as they are shown on the death certificate, which are automatically converted to ICD-10 codes. SuperMICAR can process about 80-85 percent of all records with a system error rate close to two percent. Both systems feature extensive automated data editing procedures covering a broad range of variables.

X. CONCLUSION

34. In each of the three major survey processes evaluated by the NCHS Quality Control Program, i.e., data collection, data processing, and data analysis, data editing procedures provide critical measurement components in the evaluation process. During the data collection stage, extensive use is made of Computer Assisted Personal Interviewing (CAPI), Computer Assisted Telephone Interviewing (CATI), and Source Point Data Editing (SPDE) to eliminate or minimize the introduction of errors at that stage.

35. At the data processing stage, a number of automated data editing procedures are incorporated in MICAR and SuperMICAR, the Center's automated multiple cause of death coding systems. These include procedures that take into account age, sex, site of condition, onset of condition, severity of condition, etc., when coding cause of death.

36. Finally, the use of computer software packages (e.g., SUDAAN--SUrvey Data Analysis) and evaluation programs such as forecast tables, provide tools that can be used to analyze selected data from complex sample surveys. These and other data editing procedures are critical to the success of the Center's Quality Control Program.

Figure 1: Conversion of three-way verification system to two-way verification system: effect on majority rule cases*

THREE-WAY SYSTEM			**Becomes**	TWO-WAY SYSTEM	
Production Coder	First Verifier	Second Verifier	Production Coder	Verifier	
A	A	A	A	A	
A	A	B	A	A	
			Or A		B (adjudicate)
A	B	A	A	B (adjudicate)	
			Or A		A
B	A	A	B	A (adjudicate)	
			Or B		A (adjudicate)

* Majority rule: At least two of three coders agree on the code. The minority coder is charged with an error.

References

1. National Center for Health Statistics. Policy Statement on the NCHS Quality Control Program. Rockville, MD. August 1971.

2. National Center for Health Statistics. Manual on Standards and Procedures for Reviewing Statistical Reports. Reprinted. Hyattsville, MD. June 1992.

3. National Center for Health Statistics. Data Editing at the National Center for Health Statistics. Internal Report, Hyattsville, MD. November 1994.

4. A.J.McKeon, "Estimating the True Process Average by Means of Controlled Errors". Science in Management, Proceedings of Conference of the Administrative Applications Division. Washington, DC. 1959, pp 197-206.

5. Leon Gilford, "Estimating the Error of the Verifying Clerk When His Work Is Not Independent of the Original Clerks". Unpublished document.

6. George Minton, "Inspection and Correction Error in Data Processing". Journal of the American Statistical Association. Washington, DC. December 1969, pp 1256-1275.

7. George Minton, "Comments in Quality Control and Research in Data Processing Programs". American Society for Quality Control, Statistical Science in Management. Arlington, VA. March 1970, pp 3-G(1)-3-G(23).

8. A.H.Jaehn and R.S.Bingham, "Decision Rules for Inefficient Sorting". American Society for Quality Control, Annual Technical Conference Transactions. Washington, DC. May 1972, pp 311-318.

9. Kenneth W. Harris, "Analysis of the Independent Three-Way Verification System in Mortality Medical Coding". Memorandum. Hyattsville, MD. October 1974.

10. Kenneth W. Harris and Dwight K. French, "A Methodological Study of Quality Control Procedures for Mortality Medical Coding", PHS Pub.No.80-1355,Series 2,NO.81,Public Health Service. Hyattsville, MD. March 1980.

11. Abraham Wald, Sequential Analysis: John Wiley and Sons, Inc. London, England. 1947.

12. Kenneth W. Harris, William M.Spillane,et al, "Use of Sequential Analysis in Quality Control of NCHS Operations". Proceedings of American Statistical Association Social Statistics Section. Washington, DC. August 1978.

13. Kenneth W. Harris, "Quality Control Procedures for Medical Coding: National Center for Health Statistics". Proceedings of American Statistical Association Social Statistics Section. Washington, DC. August 1983.

14. Kenneth W. Harris and Harry M. Rosenberg, "Evaluation of an Automated Multiple Cause of Death Coding System". Proceedings of American Statistical Association Social Statistics Section. Washington, DC. August 1993.

15. Kenneth W. Harris, Harry M. Rosenberg, and Donna E. Glenn, "Further Evaluation of an Automated Multiple Cause of Death Coding System". Proceedings of U.N.Statistical Commission and the Economic Commission for Europe (ECE),Conference of European Statisticians, Work Session on Statistical Data Editing. Geneva, Switzerland. October 1997