AN OVERVIEW OF THE QUALITY CONTROL PROGRAM FOR THE NATIONAL CENTER FOR HEALTH STATISTICS

Kenneth W. Harris, National Center for Health Statistics

Introduction

A primary mission of the National Center for Health Statistics (NCHS) is to provide data on the health of the United States population by collecting and publishing information from inventories, registration systems, and ad hoc and continuing probability sample surveys.

A primary objective of the Center's quality control program is to assure that the products of its inventories, registration systems, and sample surveys meet standards of reliability and validity. The NCHS quality control program is defined as the totality of functions for 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].

Development of Statistical Standards

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 in carrying out the various operations of a data system. Standards for nonresponse 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 either event, the standards serve primarily to assure the quality and integrity of the compiled statistics.

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

Functions of the Quality Control Program

When NCHS developed a policy statement on its quality control program,it identified six functions that the program should satisfy. These functions are described below.

1. 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.

2. 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 nonsampling errors, thus providing a valuable service in interpreting the statistics derived from the survey operations.

3. 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 nonsampling errors required to verify the statistical conclusions presented in the text and to interpret the statistics presented in the summary tables.

4. 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.

5. 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.

6. 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.

Measurement Processes

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 know the least about from the viewpoint of measurement error. This situation reflects two things: the basic difficulty in designing quality control procedures for data collection activities, and, until recently, the relatively low priority this work has been assigned. More effort is now being devoted to the development of procedures for adequately measuring this process.

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 a manual [2]

that describes when and how to use existing analytical methods to verify statistical statements and how to assure that statistical appendices contain appropriate estimates of sampling and nonsampling errors required to verify the statistical conclusions presented in Center reports.

Finally, the operation that is the most well developed, in terms of quality control resources, is that of data processing. At NCHS, the most important staff function of data processing is coding, particularly medical coding. The remainder of this paper will focus primarily on the Center's quality control program as it pertains to the measurement of NCHS coding operations.

Types of Quality Control

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 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, it meets a previously established quality standard, or "unacceptable," in which case it is rejected and recoded.

Types of Verification

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 [3-10]. 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. Figure 1 compares the operating characteristic curves for one sample plan with two different values of the Type II error rate, 0 and 0.50, respectively. The latter value means the verifier misses 50 percent of the errors to which he is exposed. There is very little difference between the two curves up to the 0.04 error rate level, but the difference in acceptance levels increases significantly after that. When the error rate is 0.14, for example, the Type II=0 curve shows a probability of acceptance less than 10 percent (.053), - while the Type II=0.50 curve shows a probability of acceptance greater than 70 percent (.706).

The studies cited above 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. Until recently, the Center's medical coding operations were verified under a three way independent system.

Three Way Independent Sample Verification

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.

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.

Modification of the Three Way System

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. During the past several years, NCHS has investigated a number of ways of reducing verification costs while maintaining the capability of deriving timely, unbiased estimates of coding quality.

1. Sequential Sampling

A few years ago, NCHS tested the feasibility of using a modified sequential sampling scheme. 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) α = Probability of rejection when true error rate is P_1 (Type I error)

(3) P_2 = Unacceptable error rate

(4) β = Probability of acceptance when true error rate is P_2 (Type II error)

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, the ASN, based on P_1, P_2, α, β, and the actual incoming error rate, was 61. Thus the sample size, in theory, could have been reduced by almost 88 percent $(\frac{500-61}{500})$.

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.

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.

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.

2. Two Way Independent Verification

The introduction of two way independent verification in the Center's three major medical coding operations occurred in 1982 [13]. It had become apparent, by then, that the Center had to identify program areas that could absorb staff reductions with minimum impact on their functions. The Quality Control Program appeared to be one such program.

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). This conversion procedure is shown in Figure 2. Three way differences (ABC) are rare, but they, too, become AB cases (i.e., two non-matching codes) in the two way system.

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 9 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).

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.

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.

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.

Characteristics of Two Sample Plans

The switch from three way independent verification to two way independent verification was accompanied by a reduction in the sampling rate for the medical coding operations. It was felt that the high quality demonstrated over long periods of time justified such a move.

Figure 3 shows the Operating Characteristic Curves for two sample plans, the original plan for Mortality Medical Coding and the modified plan for that coding operation. The probability of accepting a work lot, L(p), whose error rate, p, is 4 percent or less, is about the same under each plan, with the smaller sample actually providing a higher probability (.972) when p = .04 than the larger sample (.943). Since the Type I error, probability of rejecting good work, is 1-L(p), it is clear that both plans offer excellent protection when the incoming error rate does not exceed 4 percent. During the 1980-1982 period, the average annual error rate ranged

between 2.8 percent and 3.0 percent, with the lowest rate occurring in 1982.

The smaller sample is not as effective as the larger sample in protecting against the Type II error, the probability of accepting poor work. For example, a work lot with an error rate of 6 percent will be accepted 11 percent of the time under the larger sample plan, but has an 83 percent probability of acceptance under the smaller sample plan. A 10 percent error rate has virtually no chance of acceptance under the larger, but has a 30 percent chance under the smaller. These differences are not significant for infrequent random occurrences of high error rate batches in an operation under statistical control. However, in instances where the quality of the coding operation itself has deteriorated, the smaller sample plan will not detect the change as rapidly as the larger plan will.

Figure 4 presents another graphic comparison of the two sample plans. The Average Fraction Inspected (AFI) is the average proportion of a work batch that is verified in order to make a decision to accept that batch. If a batch is rejected on the basis of the sample, the entire batch must then be recoded. The AFI is determined by the sample rate and L(p), the probability of acceptance. The minimum AFI equals the sample rate; the maximum AFI equals 100 percent. For the two plans shown in Figure 4, the AFI when the error rate, p, is 4 percent is .037 and .245 for the smaller and larger sample plans, respectively.

Conclusion

The use of a two way independent verification system, coupled with lower sample rates, has reduced the annual resource expenditures for measuring the quality of three medical coding operations by 75 percent, from 8 person years to 2 person years. This represents a much needed savings during the current climate of fiscal restraint. Equally important, though, is the continuing ability of the Center to provide unbiased estimates of the quality of these coding operations.

Responsibility for the NCHS quality control program is shared by all offices that collect, process, or analyze data. As this paper has indicated, much is known and has been done with regard to data processing operations. In order to increase its efficiency in managing its statistical programs, the Center needs to generate more information on costs and non-sampling errors associated with data collection and data analysis operations.

References

[1] National Center for Health Statistics, "Policy Statement on the NCHS Quality Control Program," August 1971, internal memorandum.

[2] National Center for Health Statistics, "Manual on Standards and Procedures for Reviewing Statistical Reports," Washington, D.C., March 1976.

[3] 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, pp. 197-206.

[4] Gilford, L., "Estimating the Error of the Verifying Clerk When His Work is Not Independent of the Original Clerk's," unpublished document.

[5] Minton, G., "Inspection and Correction Error in Data Processing," Journal of the American Statistical Association, 64, pp. 1256-1275, December 1969.

[6] Hansen, M.H., H.H. Fasteau, J.J. Ingram, 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, WI, 1962, pp. 311-339.

[7] Minton, G., "Comments on 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, VA, March 1970, pp. 3-G(1) - 3-G(23).

[8] Jaehn, A.H. and R.S. Bingham, "Decision Rules for Inefficient Sorting, in American Society for Quality Control," Annual Technical Conference Transactions, 26th Annual Technical Conference, Washington, D.C., May 8-10, 1972, pp. 311-318.

[9] U.S. Bureau of the Census, "United States Censuses of Population and Housing 1960," Quality Control of Preparatory Operations, Microfilming and Coding, Washington, D.C., 1965.

[10] Harris, K.W. and D.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] Wald, A., Sequential Analysis, John Wiley and Sons, Inc., New York, NY, 1947

[12] Harris, K.W., W.M. Spillane, et. al., "Use of Sequential Analysis in Quality Control of NCHS Coding Operations," Proceedings of the American Statistical Association, Social Statistics Section, Washington, D.C., 1978, pp. 748-753.

[13] Harris, K.W., "Quality Control Procedures for Medical Coding: National Center for Health Statistics," Proceedings of the American Statistical Association, Social Statistics Section, Washington, D.C., 1983, pp. 480-484.

Figure 1
Operating characteristic curves for one sample plan with different Type II error levels

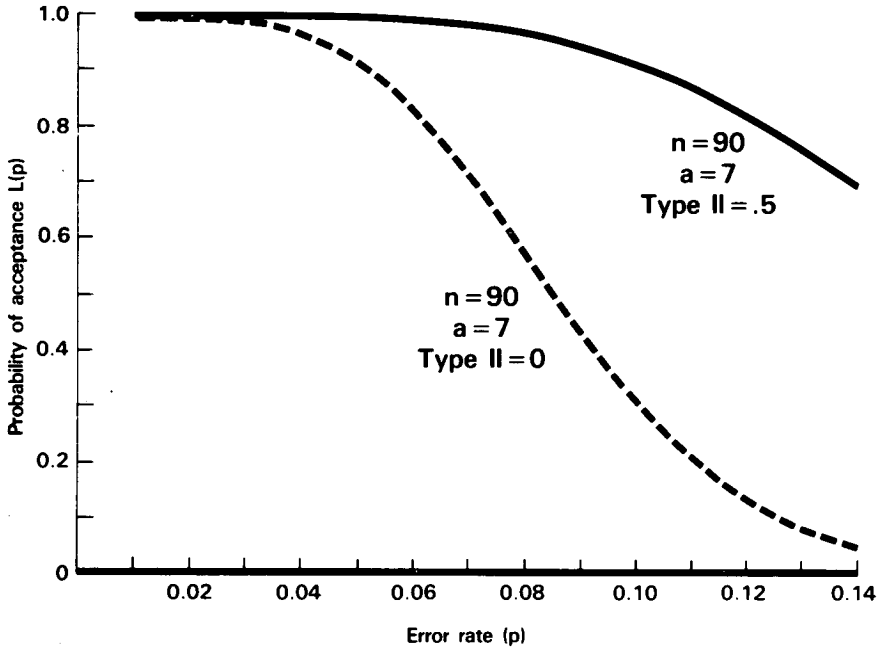

Figure 2
Conversion of three way verification system to two way verification system: effect on majority rule cases[1]

	Three way system			*Two way system*		
	Production coder	*First verifier*	*Second verifier*	*Production coder*	*Verifier*	
	A	A	A	A	A	
$\overline{AAB}$ =	A	A	B	A	A	
				A	B	(adjudicate)
	A	B	A	A	B	(adjudicate)
				A	A	
	B	A	A	B	A	(adjudicate)
				B	A	(adjudicate)

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

Figure 3
Operating characteristic (oc) curves for two sample plans

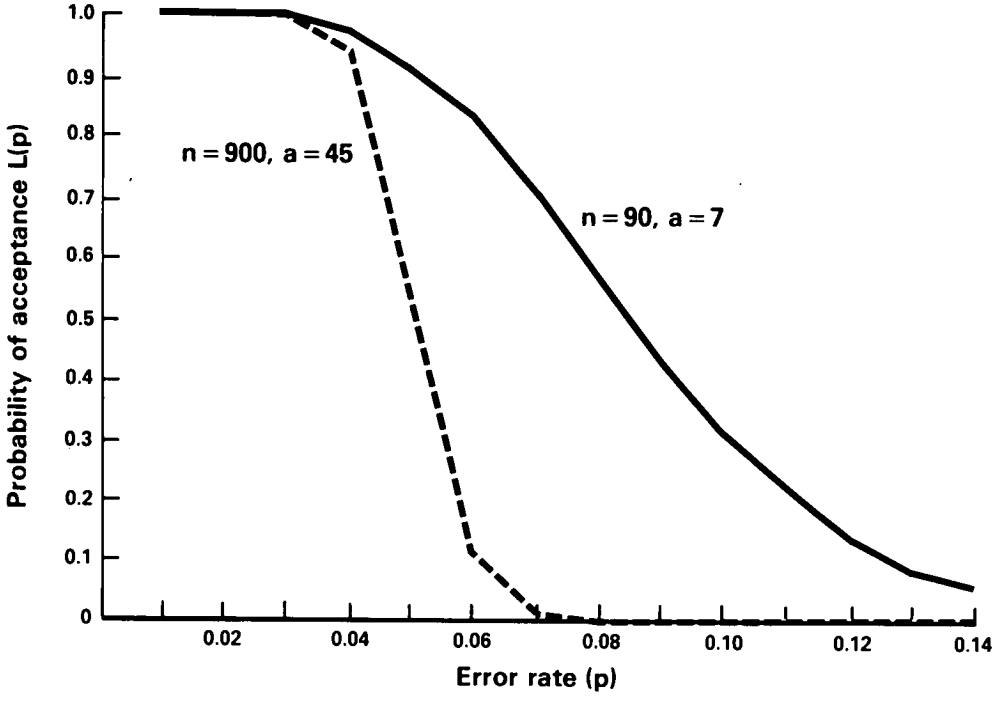

Figure 4
Average fraction inspected (afi) for two sample plans