Journal of Official Statistics
Vol. 10. No. 3, 1994, pp. 307–326

Continuous Quality Improvement for Survey Operations: Some General Principles and Applications

Paul Biemer and Rachel Caspar[1]

Abstract: Traditionally, survey organizations have relied on inspection methodologies to ensure an acceptable level of quality in their survey operations. Yet inspection methods suffer from a number of serious drawbacks. In this paper we investigate continuous quality improvement (CQI) as an alternative method for ensuring quality in survey operations. In CQI, the aim is to achieve the smallest error rate possible by continually improving the quality of the product for the duration of the operation. We discuss the fundamental concepts and tools that define our specific CQI approach and provide a framework for implementing CQI strategies within a survey organization. In addition, we describe a study in which we incorporated our approach into an industry and occupation coding operation.

Key words: Total Quality Management; industry and occupation coding; quality control.

1. Introduction

A typical sample survey consists of a number of separate, but interrelated operations that may either change the form or modify the content of the original responses. These operations include data collection (or interviewing), data transmission and receipt, data editing and cleaning, response encoding, and data entry. Depending on the scale of the survey, these operations may involve only a few operators or, as in the case of a census, hundreds. It is also common that the operators are inexperienced, lowly-paid, and minimally trained workers. The operations in which these operators are engaged may be complex and error prone. They may consist of repetitive and monotonous activities or, as in the case of interviewing, may require complex thinking and quick judgements. Thus, almost all survey organizations employ some type of quality control for survey operations in order to ensure final results of acceptable quality.

In most survey organizations, quality control is usually based upon some form of inspection. For example, all or a portion of an interviewer's, editor's, or coder's work may be examined to judge whether the work is acceptable. Sample inspection (also

[1] Research Triangle Institute, P.O. Box 12194, Research Triangle Park, N.C. 27709, U.S.A.
Acknowledgements: We wish to thank the Chief Editor, and the referees for their excellent comments and suggestions for improving the paper. We also express appreciation to Elinor Cifton, RTI computer programmer, for her help in preparing the data sets for analysis. Finally, we wish to thank the RTI I&O coding staff who contributed substantially to the success of this study.

known as acceptance sampling) involves selecting a sample of items from a work unit, inspecting the sample to determine the number of items which are "in error" or which deviate from specified procedures, and then either rejecting the work unit if the number of errors (or defects) exceeds some threshold value or otherwise accepting the work unit. For most operations, rejected units are usually reworked to remove the errors.

There are at least three potential objectives of inspection methods. One is to ensure that errors in the output do not exceed some specified level. This is possible through the use of probability methods for determining the number of units to select for inspection, the number of items within each unit to inspect, and the threshold value for rejecting a unit. This is the simplest form of acceptance sampling, and more complex sampling schemes are often encountered in practice (for a description of alternative acceptance sampling schemes, see, for example, Wadsworth, Stephens, and Godfrey 1986). For some implementations of acceptance sampling, meeting this first objective may be the sole purpose of the activity. However, for other implementations, a second objective may be to improve the skills of the operators. This is usually carried out by providing feedback to the operator responsible for a rejected work unit; that is, the operator is given information on the number and types of errors found in a rejected unit. In this way, it is hoped that the operator can take whatever actions are necessary to avoid these types of errors in the future. A third objective may be simply to "keep the operators on their toes". That is, the threat of inspection dissuades the operators from intentionally deviating from procedures or taking short-cuts that may damage data quality in order to meet production goals. This use of inspection may be criticized on a number of points, as follows:

- Inspection adds significantly to the costs of a product. Inspection is necessary only because the operation is prone to error. If the process could be redesigned so that the error rate were extremely small, inspection would be unnecessary and the cost of inspection would be saved.
- Unless the inspection process has a negligible error rate, inspection may not achieve the desired quality of output. However, often the inspection error rate may be substantial and thus the error in the final product may be unacceptable even with a high rate of inspection (see Minton 1969, 1972 for a discussion of the consequences of an error-prone inspection scheme). Thus, achieving very small levels of error – say, 1% or less – may require 100% inspection, even under perfect inspection; yet, this level of inspection may be unaffordable.
- The quality control literature (see, for example, Ishikawa 1990) provides strong evidence that when inspection methods are used, operators take less responsibility for the quality of their work since quality is perceived as being the job of the inspectors. Therefore, the operators lack the motivation to improve quality.
- When feedback to the operators is based upon rejected units, the implication is that they are solely responsible for the errors in their work unit. However, the quality control literature suggests that for most operations, the operator may only be responsible for 20%–30% of the errors. Thus, placing full responsibility for the errors in a process on the operators risks demora-

lizing them. This demoralizing effect may be exacerbated by an imperfect inspection process that may fail to identify true errors and erroneously classifies correct items as errors.

- The feedback from inspection to the operators is both time consuming and, in many cases, ineffective. Part of the reason for this is the lack of information on the root causes of the errors. The operator may be told that he/she is responsible for various errors, but is given little if any useful information on how to eliminate them in the future. Indeed, as we have stated, many errors may be beyond the control of the operators.

These limitations of the traditional inspection methods of quality control motivated our investigation of alternative methodologies for ensuring high quality in survey operations. A key distinction between continuous quality improvement (CQI) or "kaizen" (Imai 1986) methods and traditional quality control methods is that the former aims, not at simply achieving a specified "average outgoing quality limit" (AOQL), but at achieving the smallest error rate possible by continually improving the quality of the product for the duration of the operation. CQI is the core component of a more comprehensive, organizational management strategy which is referred to by a variety of names, including: Total Quality Management, Strategic Quality Management, Total Quality Control, and Fourth Generation Management, to name a few (see Deming 1986; Juran 1964, 1988, 1989; Ishikawa 1985; Crosby 1979; and Joiner 1994 for a thorough discussion of the TQM philosophy). Joiner (1994) lists three fundamental principles of TQM:

- Quality – understanding that quality is defined by the customer and this is shared with and further developed by every employee.
- Scientific Approach – developing system management and process thinking, basing decisions on data, and understanding variation.
- All In One Team – believing in people and working toward win-win instead of win-lose for all stakeholders.

Unfortunately, the literature on applying CQI methods to survey operations is quite scant. Morganstein and Hansen (1990) review a number of techniques which can result in improved data quality, including process control charts and standardization of survey processes. Colledge and March (1993) describe a number of quality improvement activities that are underway at Statistics Canada as well as other statistical agencies in the U.S., Australia, and New Zealand. However, apparently lacking in the literature are the details of how CQI can be successfully implemented for survey operations. In this regard, the current paper is unique. Furthermore, our implementation design for CQI provides an opportunity to compare the costs and data quality of CQI with a traditional inspection method.

In the next section, we provide a framework for implementing CQI strategies to a wide variety of survey operations. Also discussed in this section are the fundamental concepts and tools that define our specific CQI approach. In Section 3, we describe a study to test our approach for an industry and occupation coding operation. In presenting the results, we compare the error rates and costs of the CQI approach with that of the traditional inspection method. Finally, in Section 4, we summarize the lessons learned in our study and discuss a number of issues related to the

Table 2.1 The stages of four typical survey operations

Operation	Input	Action	Output
interviewing	procedures or questions	execution of procedures or question delivery	respondent reaction or response
editing	completed questionnaire and procedures	execution of editing procedures	edit marks, new entries, etc.
data entry	responses and data to be keyed	key strokes	keyed responses and data
coding	responses to be coded and the coding procedures	execution of the coding procedures and code assignment	assigned codes

implementation of the CQI methodology to other operations.

2. Basic CQI Principles

2.1. A conceptual framework for survey operations

In this section, we present a general strategy for quality improvement which is applicable to a variety of survey operations. In describing this strategy, it is useful to provide a conceptual framework consisting of the fundamental components of the typical survey operation. Using this model of the survey operation, we will describe in general terms the objective of CQI and how these objectives can be realized.

Most survey operations consist of three major components or stages: the input (or stimulus), the action (or task), and the output (or result). The input stage, which may be the output of some previous operation, may consist of data, forms, or other information requiring some action by an operator. These input items may be assigned to an operator for processing in work units of some homogeneous size. In the action stage, the operator performs the tasks associated with the operation on the input items. The results of these actions constitute the output for the operation. Table 2.1 presents a brief description of these components for four survey operations: interviewing, editing, data entry, and coding.

Associated with each of three components of a survey operation are the actual and the preferred inputs, actions, and outputs. As an example, the actual inputs for interviewing are the questions and procedures as they are currently defined. The preferred inputs are those questions and procedures that encourage preferred actions from the operators. Likewise, the actual actions for interviewing are the actions taken by an interviewer during an interview whereas the preferred actions are those actions which would have elicited the best response (or the preferred output). Finally, observations or interview results constitute the actual output while the preferred output are results which are completely accurate and free of nonsampling error. In brief, the actual component is what exists in the current operation and the preferred component is the ideal input, action, or output.

We shall assume that the actual survey operation component can be observed and that the preferred survey operation component can be uniquely and unambiguously defined so that actual and preferred components can be compared. The difference between actual and preferred performance for a particular item will be referred to as a nonconformity. Thus, the ultimate goal of CQI is to change the actual performance of an operation to agree perfectly with the preferred performance so that the number of nonconformities in the operation is reduced over time to zero. Progress toward this goal is achieved if, at each implementation of the operation, the number of nonconformities is reduced from the previous implementation. Note that, unlike inspection methods which tend to focus only on the actions of the operator, CQI addresses all three components of the operation.

2.2. *A general strategy for CQI*

Using this conceptual framework for a survey operation, in this section we propose a general strategy for implementing CQI. Our approach may be viewed as an integration of the three fundamental principles of TQM. First, a critical ingredient in our approach is the use of teams to identify problems, to determine their solutions, and to implement corrective measures. Secondly, the actual components of a survey operation are evaluated quantitatively using quality indicators which are functions of the number of nonconformities in the operation. Finally, priority is given to identifying and addressing the root causes of the nonconformities without regard to where they are in the system or the organization.

Our CQI plan is a four-step approach and an adaption of Deming's (1986) Plan-Do-Check-Act (PDCA) cycle. However, our approach is especially adapted for survey operations and is more specific regarding the activities to be performed under each step of the cycle, particularly the planning and checking steps. The four steps are as follows:

Step 1. Perform the operation and observe the nonconformities.

"Observe the nonconformities" implies that there is a comparison of the actual performance and the preferred performance for all three components of the operation. As an example, for telephone interviewing, the comparison may be made by a call monitor who is proficient in survey procedures and who, while listening to the interview, determines whether the observed behavior agrees with the preferred behavior. For editing, data entry, and coding, this step may entail reworking a sample of the items by an expert or by using some other process which produces the preferred output. Inspection methods such as independently reworking a sample of items, comparing the outputs, and then adjudicating the differences to obtain a final adjudicated output may be used to produce the preferred output.

Step 2. Classify the nonconformities as to their type and perform a Pareto analysis.

Step 1 may identify many different types of nonconformities in the operation – too many to address simultaneously. Step 2 sorts the nonconformities by type and performs a Pareto analysis. A Pareto analysis, as it is used in this application, is essentially a histogram showing the most frequently encountered nonconformity, followed by the next most frequent, and so on. We found that showing only the top five most frequently occurring nonconformities is sufficient for directing quality improvements (for more information on Pareto analysis, see, for example, Wadsworth et al. 1986). This analysis allows us to focus on a few, more important types of nonconformities in Step 3. As an example, for interviewing, the system of monitoring proposed by Couper, Holland, and Groves (1991) can be used to classify the types of nonconformities observed during interviewing. These type classes correspond to nonconformities in the delivery of the question (wording changes or skipped questions), probing for an adequate response (probe neutrality, completeness, or failure to probe), interviewer feedback to the respondent (neutrality or appropriateness), respondent behaviors (requests for clarification or to repeat the question) and so on. With this system, a sample of questions is monitored and the type of nonconformity observed is coded for each question and post-question interaction. Thus the type classes for the Pareto analysis may be based upon these interviewer or respondent behavior codes for all questions on the questionnaire combined, for particular sections of the questionnaire, or for individual questions.

Likewise, type classes may be defined for data entry or data editing. This may require constructing a list of the various types of nonconformities observed in the operation and developing a classification system on the basis of the most frequently observed errors. The nonconformities may be further stratified by input, action, or output. For example, nonconformities in the input affect the appearance of the data as they are presented to the editors, keyers, coders, etc., and should be reported to the previous operation. Nonconformities observed in the output may be reported by the subsequent operations in the sequence of survey operations, but may also be observed in the current operation.

Step 3. Meet in teams to identify the root causes of the most important types of nonconformities

A key feature of our approach is the use of teams to fully investigate the nonconformities until their root causes are well-understood and agreed upon by the group. Then collectively and individually the team members can set out to address the causes, thereby reducing the number of resulting nonconformities. The team structure and composition is critical to its success. At a minimum, the team should include the operators, the adjudicators or inspectors, the supervisor of the operation, and a quality advisor. For large operations, multiple teams can be formed to reduce the team size. The quality advisor's role is to keep the team on track, advise on survey methodology as well as CQI, assist in the preparation of summary reports and data analyses, and act as a liaison between the team and higher management, if necessary.

The so-called CQI team for the operation may meet frequently (weekly or biweekly) when the operation is active to review the results from the period since the last meeting. The primary objective of the meetings is to consider the most prevalent types of nonconformities as identified by the Pareto analysis and, using whatever data are available, discuss the possible causes and remedies. These discussions may lead to changes in the procedures, feedback to operations upstream regarding the quality of their outputs, retraining of the operators, changes in the work environment, and so on. In some cases, it may be determined that the process by which the preferred performance of the operation is determined is faulty. For example, there may be misconceptions among the adjudicators which lead to inaccuracies in the results of the inspections and false reports of nonconformities. These problems can be discovered in the CQI meetings and, in this way, both the original operation and the adjudication or inspection tasks can be improved.

Another objective of the CQI team is to assess the success of corrective measures which the team has implemented for the operation. Since the goal of CQI is a steady, continuous reduction in the overall nonconformity rate, the number of nonconformities in the operation should be closely monitored. Over time, there may be considerable variation in the types of nonconformities which are identified as most problematic by the Pareto analyses. Ideally, as the group focuses on and emphasizes improvement for a particular type of nonconformity, the frequency of that nonconformity should be reduced and some other type of nonconformity will rise to the fore. As these nonconformities are reduced, new classes will take their place, and so on. Over time, each type of nonconformity may take its turn in the top position while the overall nonconformity rate is ever decreasing.

Finally, the topics of the CQI meetings need not be limited to a discussion of the causes of the nonconformities or the group's progress toward reducing them. There may be other issues related to the work environment, shift structure, operator's manual, management practices, and so on that the team may discuss. The critical element in the meetings is open, uninhibited communication without fear of retribution. Creating this atmosphere is essential to fully understand the root causes of the nonconformities. It is essential to document the decisions of the group and distribute these to the group and possibly beyond.

Step 4. Implement the corrective measures and return to Step 1.

The measures to be taken to correct problems which give rise to the nonconformities may take a number of forms. For example, the individual operators may need to adhere more closely to procedures, now that these procedures have been clarified. The corrective measures, such as a change in procedure or the work environment, may be the responsibility of the facility manager or operation supervisor. Any changes to procedures, training, etc. should be well-documented.

In the next section, the results of an application of our approach to CQI to industry and occupation coding will be described. This application will illustrate in some detail how the CQI process can be put into practice and the potential benefits that can be derived from this strategy.

3. An Application to Industry and Occupation Coding

In this section we describe a study we conducted using CQI in our industry and occupation (I&O) coding operation. The study took place during the twelve months of 1992. However, due to its success, what began as a test has since become part of the standard operating procedures in RTI's I&O coding division. A general introduction to I&O coding at RTI, the procedures for initiating the CQI process, and the results of the year-long study are presented below.

3.1. The I&O coding quality control process

Questions used to obtain detailed information about industry and occupation are included in many surveys. For the most part the questions are open-ended, requiring the interviewer to record a verbatim response and to probe effectively until a complete answer has been obtained. For the information collected to be useful in statistical analyses, however, these verbatim responses must be coded using a standardized system of industry and occupation codes. In the United States these

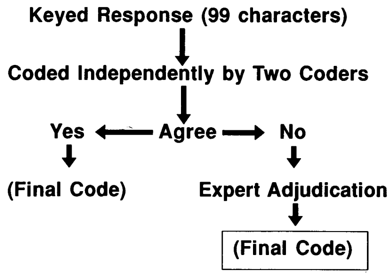

Fig. 3.1. I&O coding quality control system

codes are developed by the U.S. Census Bureau and are updated after each Decennial Census to reflect new industrial and occupational areas. Currently, the system includes more than 30,000 occupation titles classified into approximately 500 occupations, and 20,000 industry titles classified into 230 industries.

At RT1, I&O coding is a manual operation. Specially trained coders are responsible for matching the open-ended survey responses to one three-digit industry code and one three-digit occupation code. In Figure 3.1, the full RTI I&O Coding Quality control system is diagrammed. Upon receipt from the field, the questionnaire responses are keyed. A total of 99 characters are allowed for the industry response and an additional 99 are allowed for the occupation response. I&O coders work at terminals, accessing one data record at a time. Each record is coded independently by two coders. If the two coders are in agreement for both the occupation and the industry codes, the record is finalized with those two codes assigned. However, if one or both of the codes disagree between the two coders, the case is flagged for adjudication. I&O adjudication is handled by more experienced coding personnel. Each record sent to adjudication is reviewed. The codes assigned by each coder are displayed on the adjudicator's screen (though any information indicating which coder assigned the codes is not), and the adjudicator may assign one of those codes or a different code entirely. Regardless of whether the industry or occupation code is in disagreement, both codes will be reviewed during adjudication, as a change in one may result in a necessary change for the other. The codes assigned during adjudication are considered final codes and are written to the permanent data record.

It should be noted that this form of "dependent adjudication" has been criticized in the literature (see Minton 1969) and "independent adjudication" may provide greater accuracy. With dependent coding, there is evidence that the adjudicator may simply pick the better of the two assigned codes instead of arriving at the best code which may differ from the two assigned. In independent adjudication, the adjudicator is not aware of the codes assigned in the previous codings and assigns a code completely independently of the other two coders. Thus the task of the adjudicator is to determine the "best" code after observing three independent codings. However, any additional accuracy provided by independent adjudication comes at a cost since independent adjudication can be substantially slower than dependent adjudication (Fasteau, Ingram, and Minton 1964).

The RTI I&O Quality Control system is a 100% (no sampling) inspection system. Every record in disagreement is referred to adjudication and every code is finalized only when two coders agree on the codes to assign or the adjudicator assigns the code. However, one should not conclude from this discussion that no error exists in these final codes. Two sources of error may still be present: erroneous agreements between coders, and errors made by the adjudicators. Both of these sources of error can be minimized by reducing the overall error in the system. To the extent that all coders are equally well-trained and capable of coding in accordance with the general rules, erroneous agreement between coders can be reduced (and theoretically eliminated entirely). Likewise, the fewer cases sent to adjudication, the fewer chances there are for adjudicator error.

The I&O coding procedure described above has been used at RTI for a number of years. The system has allowed us to fulfill our clients' expectations of obtaining high quality data. Yet, in 1991 we discovered that our quality control system was resulting in especially high costs for the I&O coding operation. These high costs were due to the fact that close to half (46%) of all cases were being sent to adjudication for final code assignment. The added cost was the result of the additional time billed by the adjudicators. With this level of disagreement between coders, it seemed clear that the two sources of error capable of infiltrating our system (erroneous agreement and adjudicator error) were likely to be a nontrivial source of error in the final codes.

In order to document the problem more fully, we developed a new measure of coding accuracy designated as the "coder error rate" (CER). The CER is calculated for each coder individually and is defined simply as the number of disagreements with the *final* code divided by the total number of codes assigned. We prefer to use the CER as our measure of accuracy rather than the disagreement rate because it allows us to classify the coding errors according to the adjudicator's code which we consider to be the most accurate code. In this way, we can identify industries and occupations which are particularly difficult to code. The CER will usually be significantly less than the between-coder disagreement rate – almost half in many cases. Of course, lowering the CER will also result in a decrease in the between-coder disagreement rate.

Using the same 1991 data, we calculated the CER for every coder – for both industry and occupation. It is typical for occupation error rates to be higher than those for industry as the responses are usually more difficult to code and also because the codes are driven from the industry code which is assigned first. Thus, if the industry code is assigned incorrectly, it is more likely that the occupation code will also be in error. Our error rates reflected this. Overall, the error rate was 17.4% for industry and 21.1% for occupation. However, there was a wide fluctuation of error rates among the coders. Industry error rates ranged from 13.0% to 28.1%. Occupation error rates ranged from 15.8% to 33.2%. While it is difficult to know what an acceptable error rate "should" be, discussions with Quality Assurance Staff at the U.S. Census Bureau indicated the bureau averages error rates of 13.0% for industry and 18.9% for occupation (P. Gbur, personal communication, January 30, 1992). Based on this information we felt certain that our error rates could be decreased and costs reduced. Using the 1991 error rates as a starting point, in January 1992 we began to implement changes in our I&O coding operation.

3.2. The effects of system changes on the error rate: Quarters 1 and 2

During the first quarter of 1992, a few significant changes were made to the I&O coding system. First, I&O coding was restricted to the day shift and the number of coders was reduced from 15 to 5. This change was motivated by several considerations related to work load and the cost

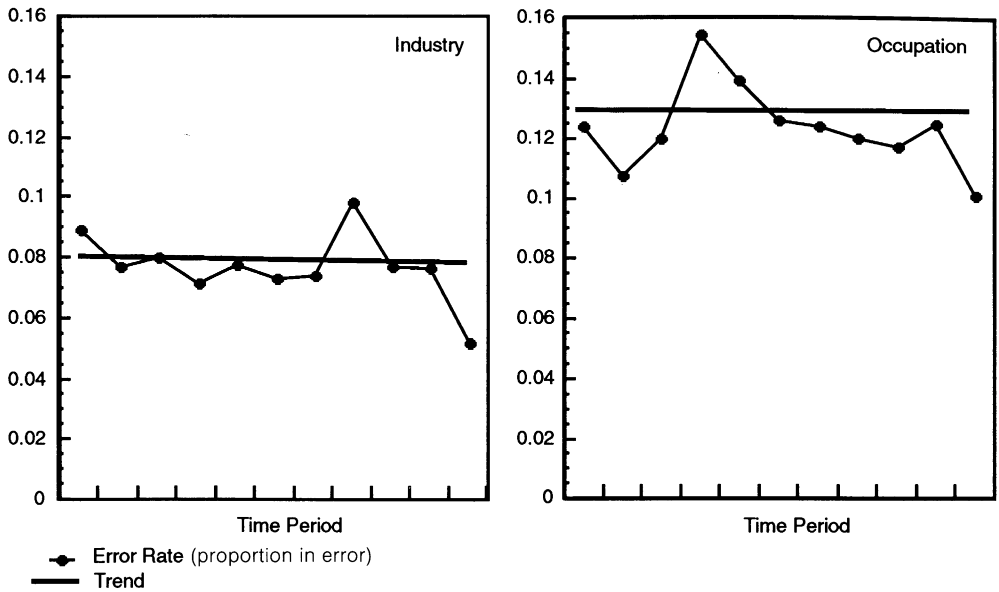

Fig. 3.2. CER for industry and occupation coding: Quarter 1 results

of providing for an additional shift. In addition, the error rates for night shift operators were consistently higher than those of the day shift and it was suspected that the temporary, inexperienced personnel recruited for the late shift were responsible for this. Secondly, based on comments received from the coders, an enhancement was made to the on-line coding system. Rather than simply accepting a three-digit code keyed by the operator, the computerized system was reprogrammed to display a written description of the code after the code was keyed by the operator. After reviewing the description, the operator can either keep the code or reject it and reenter a different code. All operators agreed that this change improved their coding accuracy, particularly for catching typographical errors before they are entered into the system. Finally, all operators were encouraged to focus more on accuracy in completing their assignments.

The results from Quarter 1 are presented in Figure 3.2. The improvement from 1991 to the first quarter of 1992 are clear. The CER for industry fell from 17% during 1991 to 8% by the end of the first quarter. Likewise, the error rate for occupations decreased from 21% to 12%. While these are dramatic improvements, our success was tempered by the fact that after an initial drop in both error rates, there appeared to be no additional improvement throughout the first quarter period. Statistical analyses show that the slopes of the lines displayed in Figure 3.2 do not differ significantly from zero. Thus, while an improvement was achieved, it was clearly not the continuous quality improvement described in Section 2 of this paper.

At the start of the second quarter of 1992, we reviewed the results from Quarter 1 with all coding staff, and continued to stress the importance of quality in the operation. The results for Quarters 1 and 2 are presented in Figure 3.3. The error rate for industry decreased to approximately 7% by the end of Quarter 1. During the same period, the occupation error rate fell to

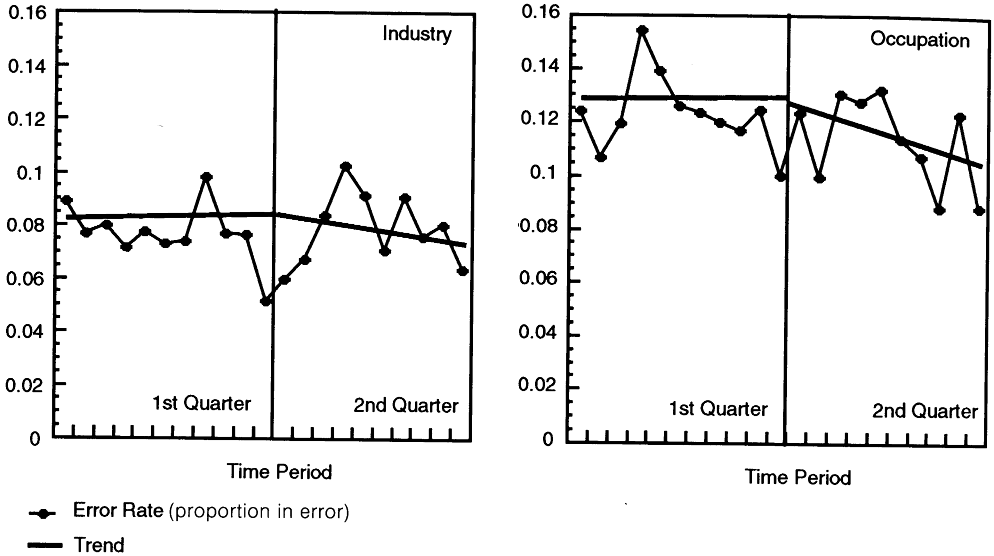

Fig. 3.3. CER for industry and occupation coding: Quarters 1 and 2 results

just over 10%. From the graphics in Figure 3.3, it appears that some improvement may have occurred during this quarter. However, due to the fluctuation in the error rates throughout the quarter, our analyses once again showed no statistically significant change occurring during the quarter. Thus by the end of Quarter 2, while we had dramatically improved our overall error rates compared to 1991 data, we still had not been able to create an environment of continuous quality improvement.

3.3. The effect of CQI on the error rates: Quarters 3 and 4

Prior to the start of Quarter 3, we implemented the full four-step CQI process outlined in Section 2 of this paper. Our first step was to make one simple modification to the RTI I&O coding quality control system (see Figure 3.4). The addition of the feedback loop allowed coders to receive information about cases they had coded incorrectly and to use this information to improve their future performance. More detailed information regarding the type of feedback given to coders and the way in which this took place is provided below.

To begin, the entire coding staff received about two hours of training which explained the CQI process as it would be applied to I&O coding. Then, weekly quality circle meetings were organized. All five coders, two adjudicators, two supervisors, and a quality advisor (or facilitator) took part in these meetings. During these meetings the coding staff was encouraged to share any problems they were encounter-

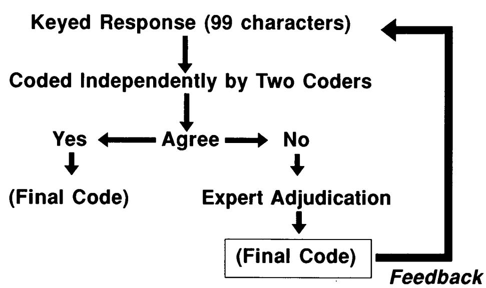

Fig. 3.4. I&O coding quality control system with modification for a feedback loop

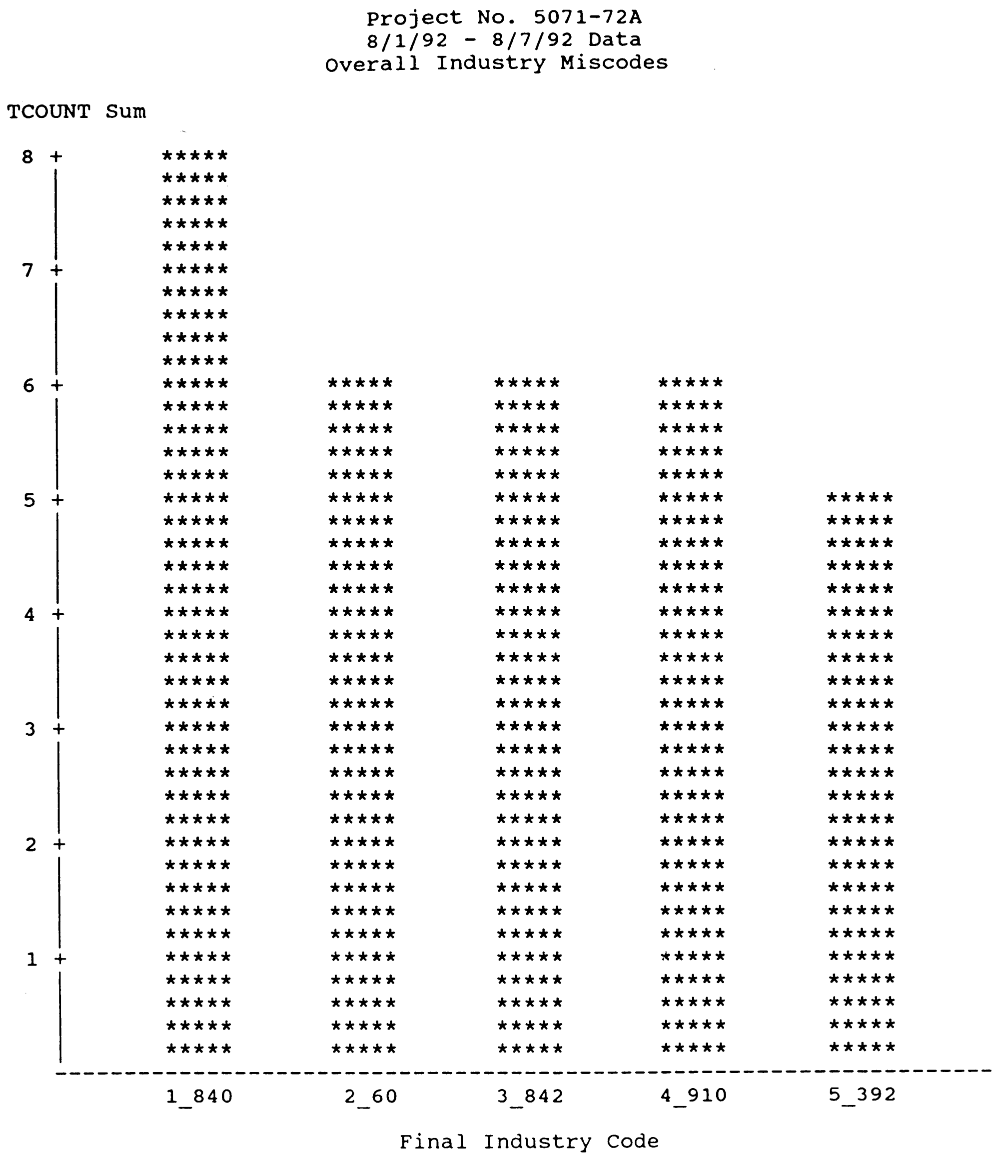

Fig. 3.5. Example of the Pareto charts used with I&O coding staff

ing as they completed their work and to discuss possible solutions. Pareto charts were provided to the coders the day before the meeting. These charts (for an example see Figure 3.5) documented the most often misassigned codes for the group as a whole, based on all cases coded during the previous work week. This allowed the coders to see exactly which codes were causing difficulty for the group. Five industry codes and five occupation codes were documented in this way. In Figure 3.5, which shows the results for industry coding during the week of August 1, 1992, the five most problematic codes were 840, 060, 842, 910, and 392. In addition to the overall charts, each coder also received an individual listing which showed the ranking of

```
                         Project No. 5071-72A
                         8/1/92 - 8/7/92 Data

FIVE OF OCCUPATION MISCODES FOR ADJUDICATED CODE: 19  FOR CODER:  BH
--------------------------------------------------------------------

Incorrect Code:   8           Other Code:  19
 Occ:MANAGING STAFF, OVERSEEING COMPUTERS
 Dty:INFORMATION SYSTEMS MANAGER
 Bus:OFFICE AUTOMATION
 Des:U S GOV'T
 Type of Business: Other
 Adj:IND CODED TO IND NOT REPORTED


Incorrect Code: 174           Other Code:  19
 Occ:COOKING, HOUSE CLEANING, TAKING CLIENTS OUT
 Dty:COUNSELOR FOR MENTALLY HANDICAPPED
 Bus:BEHAVIORAL RESEARCH ASSOCIATION
 Des:BEHAVIORAL RESEARCH
 Type of Business: Other
 Adj:


Incorrect Code: 254           Other Code:  19
 Occ:HELPING PEOPLE WITH RELOCATION TRANSITION
 Dty:RELOCATION COUNSELLOR
 Bus:REAL ESTATE RELOCATION SERVICE
 Des:SERVICE
 Type of Business: Other
 Adj:


Incorrect Code: 653           Other Code: 706
 Occ:IRON FABRICATION
 Dty:SELF EMPLOYED
 Bus:MAKE PLAY GROUNDS DYDROLIC GATES, POOL FENCING OUT
 Des:
 Type of Business: Retail
 Adj:OCC CODED TO OWNER
```

Fig. 3.6. Example of the personal errors listing sheets used with I&O coding staff

these same problematic codes for his/her work alone.

Further information about these problematic codes was provided to the coders in the form of a Personal Errors Listing Sheet (see Figure 3.6). For each of the codes identified in the Pareto analyses, coders received a listing of up to five cases which they had not coded correctly and thus had been sent to adjudication. This listing displays the entire text of the response as the coder originally viewed it. The listing also shows the incorrect code assigned by the coder, the code assigned by the other coder, and any comments made by the adjudicator. Thus, in Figure 3.6 we see a list of four cases for coder "BH". Each case was adjudicated to an occupation code 019 after the first and second coder disagreed. For the first three cases, the other coder had assigned the correct code (i.e., agreeing with the adjudicator's code), but in the last example, both coders incorrectly coded the case. During the meetings coders were able to look at these examples and discuss how they had arrived at the incorrect code. Supervisors could then provide explanations and retraining to reduce misunderstandings about the codes and to increase the likelihood that these codes would be

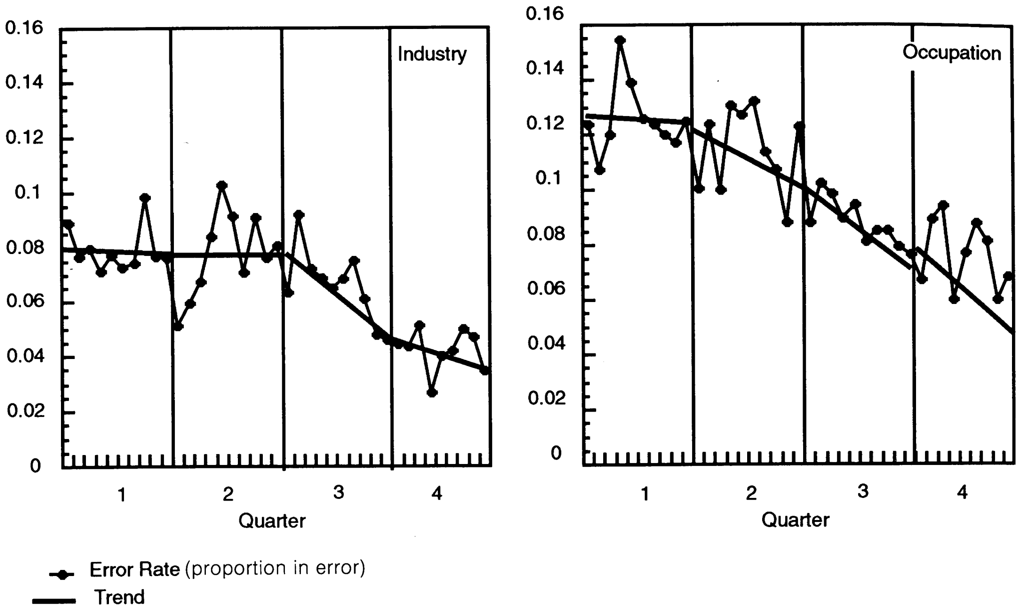

Fig 3.7. CER for industry and occupation coding: Quarters 1, 2, 3, and 4 results

used correctly in the future. The adjudicators could also provide their rationale for assigning a particular code to a case.

The critical component of these weekly meetings was the use of a team approach. The Pareto analyses identified the most problematic codes for the coders as a group. The goal was not to identify poor coders and replace them, but to have all the coders focus on the most important issues for the group. By having each member of the coding staff strive to improve his/her work on the problematic codes for the week, the overall error rates decreased and individual performance improved as well.

The weekly meetings were not restricted solely to discussion of the Pareto analyses and personal listings, however. Coders and adjudicators were encouraged to bring up problematic issues related to their work environment, the quality of the data they worked with, and other demands on their time which impinged on their ability to work efficiently. The quality advisor, who attended each meeting would, when necessary, act as a liaison to upper management and staff in other divisions of RTI whose decisions affected the coding operation. In this way, the coding operation could be improved both by increasing the skill-level of the coders and by improving the external environment in which the coding operation occurs. The quality advisor kept the team focused on the objectives of CQI and served as a resource to the team, providing additional tabulations of the data upon request and responding to questions regarding the effect of certain types of errors on overall coding quality and other technical questions.

Results in Figure 3.7 show the error rates across all four quarters of 1992. By the end of Quarter 4 the error rate for industry had fallen to 4% and the rate for occupation had decreased to just under 5%. Not only is the overall decrease from the end of Quarter 2 to the end of Quarter 4 dramatic, but the slope of the lines for both industry and occupation during Quarters 3 and 4 show

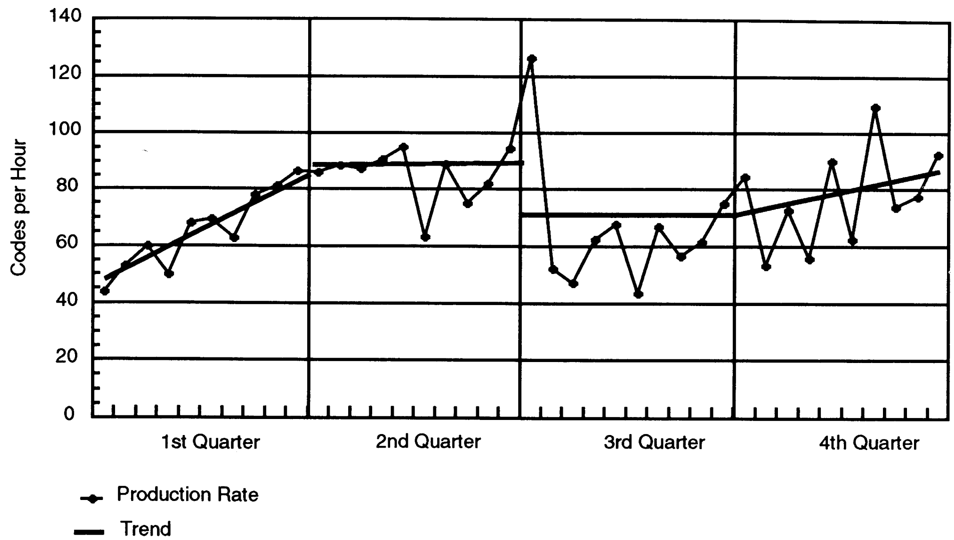

Fig. 3.8. Production rate in codes per hour for I&O coding: Quarters 1, 2, 3, and 4 results

a statistically significant downward trend. Our goal of continuous quality improvement has clearly been realized.

An important question to raise at this point is: How much did the reduction in error rates cost to the coding operation? To address this issue, we first looked at the effect of the reduced error rate on coder production rates. In Figure 3.8 the production rates by quarter are detailed. During Quarter 1, production rates increased markedly from approximately 40 codes assigned per hour to just over 80 codes per hour. In Quarter 2 there is virtually no additional improvement, though the rate of 80 codes per hour is clearly sustained. Beginning with Quarter 3, when our efforts at adapting the full CQI model began, production rates dropped to about 65 codes per hour and remained fairly stable at this rate throughout the quarter. In Quarter 4, however, the production rate began to increase such that by the end of 1992 our coders were assigning nearly 80 codes per hour. At this rate, our coding operation has rebounded to nearly the same rate that was achieved prior to implementing CQI but with half the level of error present in the operation. We believe that the initial decline in production during Quarter 3 is entirely attributable to our CQI efforts. Coders began to concentrate more fully on the task and to take more time in looking up and assigning codes. But, as they received feedback during weekly meetings and began to understand where and why errors were occurring they were able to combine speed with accuracy more effectively.

It is clear from Figure 3.8 that production rates returned to the levels attained prior to implementing CQI. However, this is not to say the costs to the operation were exactly the same as those incurred prior to adopting CQI. The lower error rates mean that fewer cases are in disagreement between the two coders. Thus, fewer cases must be sent to adjudication which reduces the cost of coding a case. However, the CQI methodology is not without added costs. There is the cost of the weekly meetings which involve the entire coding staff as well as the quality advisor. There are also costs associated with producing the Pareto

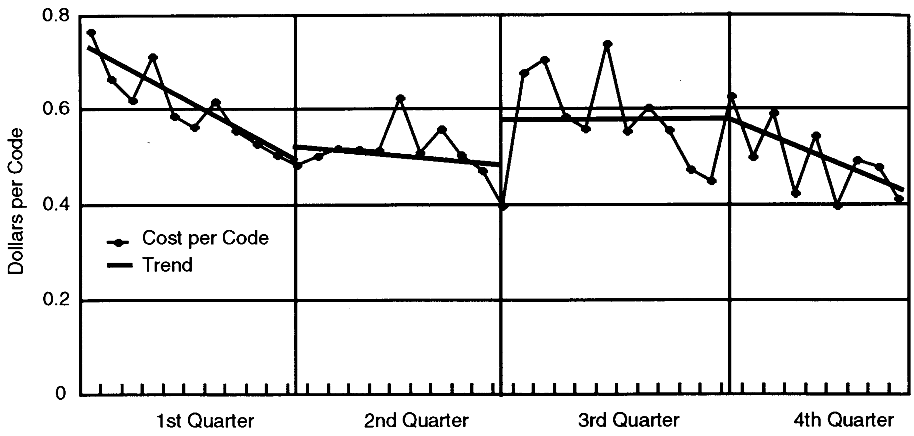

Fig. 3.9. Cost per code for I&O coding: Quarters 1, 2, 3, and 4 results

analyses and costs for the quality advisor to handle issues that involve staff outside the coding unit.

In order to compare the costs associated with the traditional and CQI approaches we developed a simple cost model which takes into account two types of variable costs: the cost of the coding and the cost of adjudication. The model for the average cost per code (CPC) is

$$CPC = \frac{CHC}{CPH} + p\frac{AHC}{APH}$$

where CHC is the average coding hourly cost, CPH is the average number of codes assigned per hour, AHC is the average adjudicating hourly cost, APH is the average number of adjudications completed per hour, and p is the estimated probability that a code is sent to adjudication. Thus, CPC decreases as coder productivity (CPH) and adjudicator productivity (APH) increase and as the proportion of cases being sent to adjudication, p, decreases. Figure 3.9 is a graph of CPC for the four quarters of the study. Despite the fact that the average coder production rate under CQI in Quarters 3 and 4 was lower, the average cost per code for Quarters 3 and 4 is almost identical to the average for Quarters 1 and 2 under the traditional inspection approach. Furthermore, a clear downward trend is exhibited in Quarter 4 which inspires hope that even greater cost efficiency may be realized under CQI. This dramatic reduction in cost is somewhat unexpected since virtually no emphasis was placed on coder productivity in this experiment. With appropriate control and feedback of coder productivity in future implementations of the CQI coding operation, we expect the downward trend in CPC to continue with no increase in coder error rates. These cost savings are substantial considering the thousands of cases which are coded in the operation each quarter and more than offset the additional costs of implementing CQI that were noted above.

Finally, at the end of Quarter 4 we asked the coding staff to critique the CQI process. Opinions were unanimously positive. The staff felt the weekly meetings provided a nonthreatening environment in which they could raise questions and concerns. The meetings also allowed the coders to receive additional training in how to assign some of the more problematic codes. The meetings were viewed as an excellent forum for this type of retraining since all coders and adjudicators were present and thus there was no danger of some staff members failing to obtain the information. The Pareto charts and error listings were viewed as an important component of the weekly meetings. Coders felt the retraining was more practical because there were "true life" examples from which to work. All staff members felt the focus on group improvement was especially useful. The coders reported that they felt increasingly comfortable reporting problems to their supervisors or to other coders because the goals of CQI were based on the group working together rather than each coder working individually.

The benefits of having a quality advisor involved in the CQI process were also noted. The quality advisor was able to effect change in other areas of the survey process which were beyond the jurisdiction of the coding staff. Examples of the tasks undertaken by the quality advisor include: (1) development of an improved interviewer training module on how to collect sufficiently detailed I&O data, (2) enhancements to the on-line coding software, and (3) development of a mechanism to notify interviewers who are not collecting sufficiently detailed I&O data. It has not yet been determined how these changes may have contributed to coding quality; nevertheless, they serve to illustrate and emphasize our view of the coding operation as an integral part of a larger survey system which is affected by multiple inputs (or suppliers) and which affects multiple customers of the operation. Thus, to achieve the goals of CQI, communication with suppliers and customers is essential since continual improvement of the inputs to operations is a necessary ingredient for continually improving output quality.

4. Other Areas of Application and Conclusions

The coding example provides a useful model for implementing CQI in other survey operations. Applications of the four-step approach to data entry, editing, and other coding operations are readily apparent. One factor that these operations have in common which we feel is key to the success of CQI is the ability of the operators to perform a step-by-step review of their actions in creating the output of the operation and to discuss, in a group setting, how their actions deviated from the preferred action. As mentioned in Section 3, the personal error listings that were provided to the coders prior to each CQI meeting were critical to the success observed for that operation.

This same approach may be used for editing. In editing, the operators (or editors) review the paper questionnaire for nonconformities in the output of the data collection operation that could pose problems for data entry or other subsequent operations. As an example, editors may fail to assign a code for a "refused" or "don't know" or may use an incorrect code. When this occurs the mistake may not be detected until the document is being keyed, and possibly not until the data are analyzed. Such nonconformities in the editing operation can be reviewed by the editors using an approach similar to the personal error listing for coders. That is, the editors can observe the input they originally received and can review the steps they performed in obtaining the output. In this way, CQI teams can constructively discuss the root causes for the editing nonconformities. Likewise, for data entry the input received by the operators can be reviewed, as can the actions taken to enter the data, and the resulting output. Thus, the personal error listings approach is possible since the actions of the operators can be unambiguously reconstructed and examined for possible root causes.

One set of operations that does not neatly conform to the I&O coding model is interviewing. Consider centralized telephone interviewing as an example. Centralized telephone interviewing consists of many interactions and interchanges between the interviewer and the respondent. To obtain a response (output) for a single questionnaire item, the interviewer may deliver the question, clarify the question for the respondent, probe to obtain an acceptable response, provide feedback to the respondent, and enter the response into the data collection system. Thus, to obtain a single output may involve a series of inputs and interviewer actions. For most surveys, none of these inputs and actions is routinely preserved in a form that would allow subsequent objective evaluation in a team setting as described for I&O coding. However, tape recording (video and/or audio) of interviews can provide the basis for constructing a "personal error listing" for interviewing, allowing intensive examinations of the root causes of nonconformities in the interviewing process, particularly with regard to interviewer performance. Since, in the U.S., tape recording requires the respondent to be fully aware that taping is being conducted, the routine tape recording of interviews is seldom done for fear of its damaging effects on respondent cooperation rates. However, there is some evidence that these fears are unfounded (see Moore, Bogen, and Marquis 1992) and thus tape recording may be a viable option for interviewing CQI.

In the absence of the ability to examine, post-hoc, the inputs, actions, and outputs associated with interviewing, the advantages of recreating for the operators the process which lead to a particular nonconformity are lost. Rather the interviewers must rely on a recounting by a trained monitor (inspector) of the series of inputs and actions which produced one or more nonconformities. We find this to be a much less effective device for CQI for several reasons. The process of monitoring can be quite subjective and error prone. Some monitoring systems provide only general, global evaluations of a segment of an interview. Others require that the monitor assess each input and interviewer action associated with a particular response and to record their evaluations of these relative to preferred inputs and actions during a live interview (see, for example, Couper et al. 1991). These assessments, which are usually made at the individual question level, include determining whether deviations from the written question changed the meaning of the question; whether a probe was used appropriately, completely and nondirectively; whether feedback to the respondent was appropriate and neutral; and so on. Our experiences with these systems indicate that monitors may be quite inconsistent in making these assessments in live interviewing situations. (However, monitoring consistency improves somewhat for tape interviews for which monitors are allowed to replay the respondent-interviewer interchanges.) As a result, these data have limited utility for CQI.

For field interviewing, the problem of collecting data on the operation for CQI purposes is even more challenging; although, here too, tape recordings and supervisory live observations may be viable options. Performance measures such as item nonresponse, edit failures, and the results of reinterviews and interview verifications have also been used to monitor field interviewer performance. These would also offer the possibility of post-hoc group and individual assessments of root causes.

Our future research efforts will be directed, in part, at implementing CQI in other survey operations, particularly for centralized telephone interviewing. Additionally, we are exploring extensions of the CQI approach by incorporating the ideas of statistical process control. In particular, we are currently investigating the use of process control charts (see, for example, Wadsworth, et al. 1986) for identifying special causes which result in abnormally high numbers of nonconformities in an operation, relative to historical data. As an example, operators who have nonconformities in their work assignments which tend to be much larger than the group mean may be identified using control charts. Then, corrective measures may be directed toward the root causes which are specific to the operators. However, there are real risks in this approach. In this paper, we have emphasized a team approach to reducing the number of nonconformities. In operations where the team approach has been implemented, the operators have commented that they enjoy the group approach and do not feel threatened or unfairly judged by it. To now focus on the individual operator as an assignable cause could be received quite negatively by the operators and thus adversely affect the morale of the group. Furthermore, it remains to be seen how much additional improvement is realized by adding individual-targeted corrective measures to the team approach relative to the team-only approach. Other uses of control charts in survey operations will also be investigated.

5. References

Colledge, M. and March, M. (1993). Quality Management: Development of a Framework for a Statistical Agency. Journal of Business and Economic Statistics, 11, 157–165.

Couper, M.P., Holland, L., and Groves, R.M. (1991). Developing Systematic Procedures for Monitoring in a Centralized Telephone Facility. Journal of Official Statistics, 8, 63–76.

Crosby, P.B. (1979). Quality is Free. New York: McGraw-Hill.

Deming, W.E. (1986). Out of the Crisis. Cambridge, MA: MIT Press.

Fasteau, H.H., Ingram, J.J., and Minton, G. (1964). Control of Quality of Coding in the 1960 Census. Journal of the American Statistical Association, 59, 120–132.

Imai, M. (1986). Kaizen: The Key to Japan's Competitive Success. New York: McGraw-Hill.

Ishikawa, K. (1985). What is Total Quality Control? The Japanese Way. (Translated by D.J. Lu) Englewood Cliffs, NJ: Prentice Hall, Inc.

Ishikawa, K. (1990). Introduction to Quality Control, 3rd Edition. Tokyo: JOSE Press LTD, (reprinted in English by 3A Corporation).

Joiner, B.L. (1994). Fourth Generation Management: The New Business Consciousness. New York: McGraw-Hill.

Juran, J.M. (1964). Managerial Breakthrough. New York: McGraw-Hill.

Juran, J.M. (1988). Juran on Planning for Quality. New York: McGraw-Hill.

Juran, J.M. (1989). Juran on Leadership for Quality. New York: The Free Press.

Minton, G. (1969). Inspection and Correction Error in Data Processing. Journal of the American Statistical Association, 64, 1256–1275.

Minton, G. (1972). Verification Error in Single Sampling Inspection Plans for Processing Survey Data. Journal of the American Statistical Association, 67, 46–54.

Moore, J.C., Bogen, K.E., and Marquis, K.H. (1992). A Cognitive Redesign of the Survey of Income and Program Participation: Development of Procedures and Initial Test Results. Paper presented to the Census Advisory Committee of the American Statistical Association, April, 9–10, Suitland, MD.

Morganstein, D. and Hansen, M. (1990). Survey Operations Processes: The Key to Quality Improvement. In Data Quality Control: Theory and Pragmatics, G.E. Lepins and V.R.R. Uppuluri (eds.). New York: Marcel Dekker, 91–104.

Wadsworth, H.M., Stephens, K.S., and Godfrey, A.B. (1986). Modern Methods for Quality Control and Improvement. New York: John Wiley & Sons.

Received September 1993
Revised August 1994