DEVELOPING PROCESS CONTROLS FOR SURVEYS AND CENSUSES

David Diskin, Peter Karcher, Michael Mersch, Paul Pavnica [1]/
U.S. Bureau of the Census

I. INTRODUCTION

The Census Bureau has always made a concerted effort to improve the quality of its operations and products. Previously, the Census Bureau has attempted to control quality for operations such as data keying and data coding by using traditional inspection and repair. However, Census Bureau studies have shown that reliance on inspection and repair does not assure a quality product. There are several reasons for the ineffectiveness of this approach, principally:

1. This approach does nothing to prevent errors before they occur.

2. Our experience has shown that repair of unacceptable work is ineffective. It is also very costly to inspect and repair outputs of an operation.

In the fall of 1985, many Census Bureau personnel viewed and discussed Dr. W. Edwards Deming's video tapes on "Quality, Productivity, and the Competitive Position". Our former colleague's eminently sensible presentation persuaded us to try to incorporate his approach into census operations. Simply stated, we decided to focus on improving processes to assure quality rather than control it after the fact.

We are attempting to improve our processes in several ways. First, a quality assurance microcomputer software system has been developed and is being constantly updated and improved. This system generates charts, graphs, matrices, and information interpreting statistical tests. It is designed to provide timely feedback to supervisors, managers, and clerks who use this information to identify special and frequent problems for clerical processes. Second, we have tried to simplify clerical tasks, as well as to improve task training. Job procedures and other materials are being written in clear English and are being supplemented when appropriate with better visual aids and straightforward decision tables. As in many areas of education, the Bureau has adopted a practice of "active learning." Trainees are required to participate to a greater extent in classroom training. Throughout the training process, much greater emphasis is being placed on quality than ever before. This paper focuses on the development of the automated quality assurance system, our experiences in applying the system, and our plans for future applications of the methodology.

II. THE AUTOMATED SYSTEM

A. WHY AN AUTOMATED TOOL IS USEFUL/ NEEDED

When we first decided to try to implement Dr. Deming's philosophy to improve the quality of clerical processes, we were convinced that an automated tool was essential. We felt that without computer assistance, we would be overwhelmed by the problem of managing the massive amounts of quality assurance data for the persons and the process that accumulate over time. Recording, retrieving, calculating, analyzing, and plotting data are tedious and error-prone jobs for humans, but an ideal application for a computer. A computer-driven quality assurance (QA) system can provide the capability to quickly and effortlessly produce the various kinds of reports needed to provide information to help in identifying problems and in analyzing the process. A computer system also can automatically analyze each person's data and indicate those points, such as runs and trends, which should be investigated.

B. DESIGN GOALS

We set four primary goals for ourselves at the beginning of the software development:
1. Make the software as intuitive and natural to use as possible.

2. Make the software as robust as possible.

3. Enable the user to easily tailor the software to the application's needs.

4. Minimize input and response time.

C. DESIGN STRATEGY

The first question we addressed was whether there

already existed a process control software package that met our basic functional requirements, as well as the goals already discussed. Although we found several commercially available packages, none sufficiently met our specific functional needs. We then chose to develop our own system.

We decided to develop the software system by iteratively refining prototypes. Building a working model is a good approach to take when designing a new software system for which there are questions about specifications and feasibility. Development of this type of automated QA system had not been attempted at the Census Bureau and this approach seemed appropriate. The first working prototype was developed in about two months for the IBM PC [2]/ environment and it demonstrated the feasibility of the QA system. We used successive prototypes to gain a better understanding of the system requirements, the user interface, and the performance issues. The prototype was then refined into the first version of the complete system.

D. **CURRENTLY AVAILABLE FEATURES**

1. **Control or Trend Charts**
   Control or trend charts (see figure 1) can be produced for each person or only for those individuals requiring feedback. Feedback may be
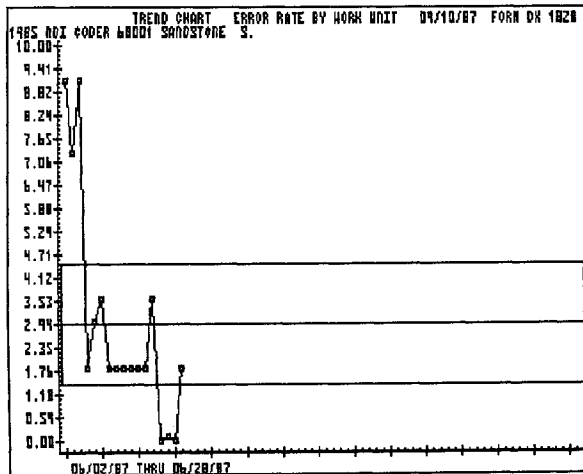

Figure 1

required if the control chart shows a run, a trend, points outside control limits, if the person's average error

rate is significantly higher than the group's average error rate, or if the person's variation is too wide. This last point is an important one because we need to keep in mind that a primary cause of poor quality is uncontrolled and excessive variation.

2. **Reports**
   Several types of reports can be requested. An error matrix of person by type of error or item (question) can be produced (see figure 2), as well as a scatter diagram, which is a plot of error rate against production rate (see figure 3). A planned enhancement will provide an error matrix of error type by item.
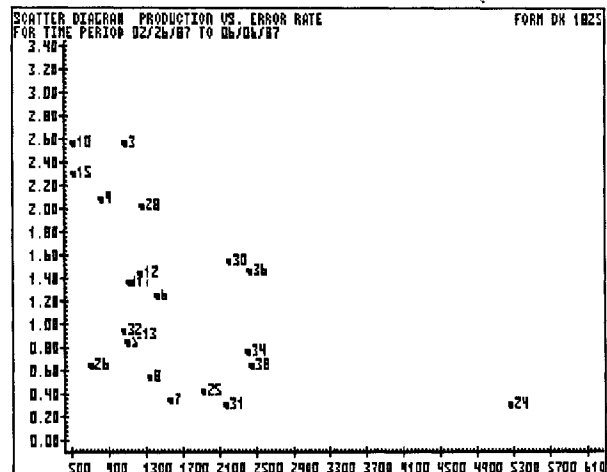

Figure 2

Figure 3

553

The user can also obtain reports which provide information in summary or disaggregated by work unit about fields coded or keyed, fields verified, fields in error, and average error rate. These can be requested for specified individuals or for everyone. Date intervals can be specified if desired.

3. Batch Mode

A batch facility is provided. This option enables the user to create and print all charts and reports for all persons or just for those requiring feedback. This option is provided for users who do not always want to employ the interactive features of the system.

4. Data Capture

Currently, data is keyed into the system via screens designed to look like the usual forms on which the QA data is recorded. Consistency and validity edits are built into the system to detect as many QA data keying errors as possible. Work unit error information and personal data for each individual on the project are captured. The system also has data entry facilities for capturing data on production and type of error.

A future enhancement will allow for direct transfer to the microcomputer of data from processes which already have captured such data in computer-readable form.

5. User Interactive Analysis

The software provides an option to work interactively with the control charts. For example, the user can mark certain points on the control charts for detailed reports or can mark points to be excluded from the process average calculations.

6. User Customization

To support our design goal of flexibility, the software allows the user to specify many output and analysis parameters as indicated in Figure 4. These include:

a. General

- Enabling the user to hear the computer beep or not at the completion of a data entry field.

- Specifying the project start date, which is used to edit work unit error data.

- Specifying the operation and type of work of the person being analyzed, such as "Edit Clerk", or "Laguna Niguel Prelist Keyer". This identifying information is displayed on screens and outputs.

b. Control Chart Display

Names, control limits, the group process average, and the individual's process average can each be displayed or hidden as desired. Additionally, the y-axis error rate scaling can be specified.

c. Feedback Parameters

The user may specify how to define the types of feedback:

- The number of points to include for runs and trends.

- The difference between the group process average and individual process average which is considered "significant".

```
Screen: D1              System Defaults                Page 1

  ┌─────────────────────┐    ┌──────────────────────────────┐
  │       General       │    │         Chart Display        │
  ├─────────────────────┤    ├──────────────────────────────┤
  │Bell On? (Y/N):    F │    │Display names on charts? (Y/N): T │
  │Confirm On? (Y/N):  T│    │Display control limits? (Y/N):  T │
  │Project start date: 05/01/87│ Display group average? (Y/N):  F │
  │Display for person type 1:  │ Display coder average? (Y/N):  T │
  │Jeffersonville MDI Coding   │ Y-axis maximum error rate: 10.00%│
  │(E)xperimental/(C)ontrol: E │ Y-axis minimum error rate:  0.00%│
  └─────────────────────┘    └──────────────────────────────┘

       ┌────────────────────────────────────────────────┐
       │            Target Group Definitions             │
       ├──────────┬──────────┬──────────┬────────────────┤
       │ Outside  │Differs from│ Runs   │    Trends      │
       │Control Limits│Group Average│       │              │
       ├──────────┼──────────┼──────────┼────────────────┤
       │Number of │Tolerance:│Consecutive│Consecutive    │
       │points:  1│   100.00%│points: 12│points:    9    │
       └──────────┴──────────┴──────────┴────────────────┘

Enter or Re-Enter Information, Press F1 for HELP
```

Figure 4.

E. IMPLEMENTATION

We have tried to achieve ease of use in several ways.

- The software uses a familiar and consistent user interface similar to popular spreadsheet programs like Lotus 1-2-3. (Figure 5). Menu selections are made by highlighting the choice with the cursor keys or by selecting the first letter of the menu choice.
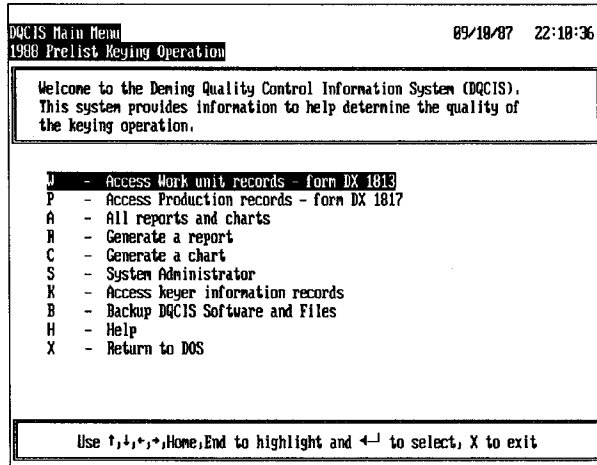
```
DQCIS Main Menu                              09/10/87  22:10:36
1988 Prelist Keying Operation

 Welcome to the Deming Quality Control Information System (DQCIS).
 This system provides information to help determine the quality of
 the keying operation.

   W   -  Access Work unit records - form DX 1813
   P   -  Access Production records - form DX 1817
   A   -  All reports and charts
   R   -  Generate a report
   C   -  Generate a chart
   S   -  System Administrator
   K   -  Access keyer information records
   B   -  Backup DQCIS Software and Files
   H   -  Help
   X   -  Return to DOS


       Use ↑,↓,←,→,Home,End to highlight and ↵ to select, X to exit
```

Figure 5

- On line help (Figure 6) is available to the user from every menu selection screen as well as from other screens.

```
                Deming Quality Control Information System
 [W] - Select "W" to add, modify, or delete Work Unit Records. Work Unit
       Records provide information about the # of codes verified and how many
       errors were detected in a Work Unit. Forms or questionnaires
       are grouped into Batches for coding or keying and verifying.
 [P] - Select "P" to add, modify, or delete Production Records. Production
       Records provide information about production rates for coders.

 [A] - Select "A" to print all DQCIS reports and charts

 [R] - Select "R" to generate printed or on-screen reports only.

 [C] - Select "C" to generate control charts for groups or individuals.

 [S] - Select "S" to set default values used by DQCIS routines.

 [K] - Select "K" to Access Coder Information records.
 [B] - Select "B" to format or backup the DQCIS software and files.
 [H] - Select "H" to obtain this menu.
 [X] - Select "X" to exit DQCIS.
                Press the Enter key (↵) to exit
```
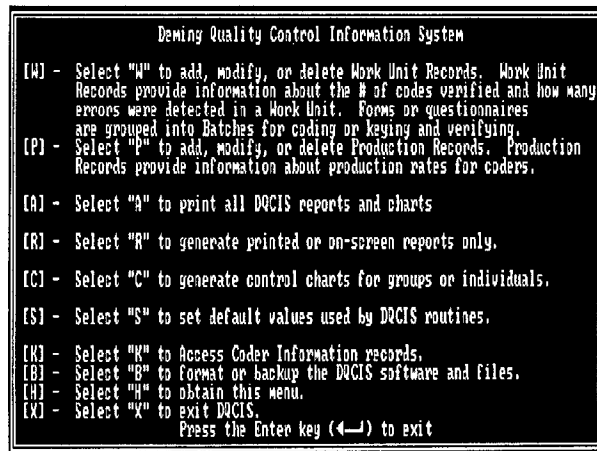
Figure 6

- Whenever possible, user interaction with the software is accomplished through a point and select method (Figure 7) rather than requiring the user to remember and enter information.

```
                     All Keyers (Charts)

  Budzinski    ( #60002 )        Meisner     ( #60005 )
  Foster       ( #60007 )        Monahan     ( #60003 )
  Iskow        ( #60008 )        Nunn        ( #60006 )
  Johnson      ( #60004 )        Sandstone   ( #60001 )








       Use ↑,↓,←,→,Home,End to highlight and ↵ to select, X to exit
```
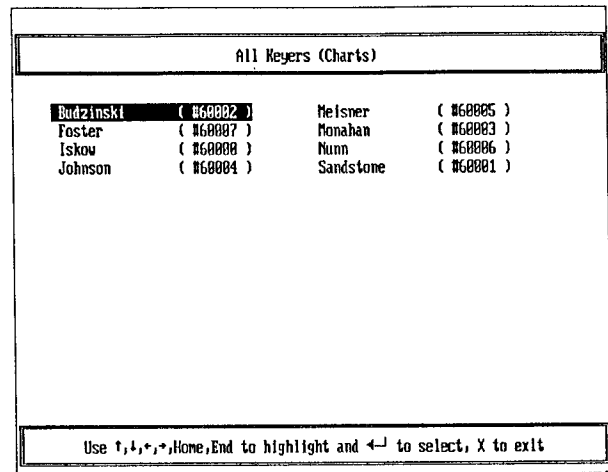
Figure 7

We also have addressed other design goals:

- As already discussed, the user is able to customize the software to some extent through the use of specifiable parameters.

- Response time has been minimized by using a compiled version of the dBase III+ code and through careful algorithm analysis.

Two other things we did proved quite helpful. First, we attempted to employ reusable code modules whenever possible in order to reduce the time needed for development and changes to the system. For example, all menu screens are produced by one parameter driven generalized module. Second, we monitored the use of the system in California and North Dakota, sites for decennial census pretests, from our Washington, D.C. headquarters through the use of modems and communications software running at the remote site as well as at headquarters.

III. EXPERIENCES

Our experiences with the current automated record-keeping system only covers two decennial census operations--the 1987 Test Census clerical edit and the 1988 Dress Rehearsal prelist keying. But from these two operations, a few things are already clear:

(1) The error matrix of clerk (keyer) by item or type of error is very valuable.

555

(2) Computer interpretation of data is very useful.

(3) Further simplification of the data entry portion should be investigated.

The Error Matrix - This report has turned out to be the supervisors' favorite output to identify personnel with potential problems. This is not all that surprising when one considers that at a glance the supervisor can look at all personnel across all types of errors or items. It facilitates the identification of potential systems problems by highlighting items or types of errors which are disproportionately contributing to total errors. It also highlights special problems individual clerks are having.

Control Charts - We first believed that these charts would be the most relied on by the supervisors and other personnel to help control and analyze the process. However, the actual chart was not relied on during these two operations. We suspect the primary reason for their lack of use was that the supervisors were not given adequate training in the interpretation of the charts. Also, due to their unfamiliarity with using control charts, the supervisors felt that they did not have the time to use them. This result indicates that the supervisor's job responsibilities should increasingly stress not only production but also quality.

Computer Interpretation - This refers to outputs from the system that interpret the control chart data for the supervisors (e.g., determining runs, trends, and points outside control limits). This computer-assisted interpretation has been very useful for the supervisors because the outputs indicate specific personnel who are having problems and who are in need of retraining or even reassignment from the operation.

Simplification - This is mainly a concern for the data entry part of the system. For the 1990 Decennial Census, data entry will be required in about 350 different locations. As in the past, many of the data entry keyers will have had little, if any, computer experience. For this very reason, data entry and the menu selections corresponding to it must be as uncomplicated as possible. This also has the benefit of making data entry training easier. For example, in the current software, in order to modify a record from the main menu, a keyer has to go through several different submenus. In future systems, some of the submenus will be eliminated. This will greatly simplify the data entry and reduce training time.

In all, the automated system has been an extremely valuable tool. Some modifications are planned, of course, but based on past test census operations, the automation of quality assurance is necessary for 1990 and beyond.

IV. FUTURE PLANS

A. NATIONAL DEATH INDEX PROJECT

We are attempting to compare the traditional method of quality control, based on inspection and repair, with the quality assurance approach which emphasizes constant monitoring and adjustment. A controlled experiment with a split panel design for assignment of death certificates for coding is being used. A unit of coders under the traditional method of control will carry out half of the work. A unit of coders under the new quality assurance procedures will do the other half of the work.

The operation being studied is a transcription and coding operation of items on death certificates from all 50 states for the National Death Index project. It is being carried out by two units of the Bureau's Data Preparation Division in Indiana. Efforts are being made to ensure the statistical integrity of the experiment, consistent with the predicted environment. The two units will be kept as physically separate as possible. The death certificates will be processed by state, with each unit receiving equal numbers of certificates from each state.

The quality assurance software will be used to capture data about coders in both units. However, only one unit will have access to the available charts and reports. The charts and reports will be used to compare units by personnel in Washington. Analysis will be made on such aspects as: learning curves, variability of data, error rate reduction over time, total time to complete, cost, and overall quality of the final data.

B. 1990 DECENNIAL CENSUS

For the 1990 Census, the Census Bureau plans to utilize a family of minicomputers for much of the required data processing. Currently, the automated QA system requires a separate data entry operation to input QA data into a microcomputer. For the 1990 Census, the QA software will be integrated with the Census data capture system on the mini-computers. As processing data is captured, the automated quality assurance software can collect and manipulate the data as desired. Thus, the separate QA data entry operation can be eliminated. No longer will there be a need to overburden a data collection keyer in order to gain more quality assurance data.

However, a few clerical operations will still require manual QA data entry in 1990. There will be an estimated 350 offices where clerical operations will take place in which quality data will have to be manually collected. This means there will be at least 350 keyers to be trained in data entry and 350 supervisors to be trained in data interpretation. Simplifying and standardizing the software and procedures is essential. For the data entry, a tutorial is being developed. This will reduce training time and cost. For the supervisors, an in-depth procedure manual on how to interpret and use the outputs is currently being developed and will be continually refined.

C. PLANS FOR SURVEYS

In addition to the decennial census and National Death Index activities already discussed, the Bureau is applying these quality assurance methods to the keying operation in the Economic Census. We are also considering using the methods and software for several other projects. These include data entry in developing countries under the auspices of our international program, and the Current Population Survey Industry and Occupation Coding operation. Operations for other surveys also will be examined to see how the methodology could be applied.

D. SOFTWARE ENHANCEMENTS

In addition to those enhancements already mentioned, we plan to make the generation of project specific versions easier. We will investigate further parameterization of the system functions so that more of the code changes can be parameter driven.

V. ACKNOWLEDGMENTS

The authors wish to recognize Lawrence T. Love, now on an overseas assignment for the Census Bureau. Mr. Love was instrumental in promoting Dr. Deming's ideas at the Census Bureau.

FOOTNOTE

[1] This paper reports the general results of research undertaken by the Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

[2] HARDWARE REQUIREMENTS FOR PC VERSION OF AUTOMATED SYSTEM

- IBM PC/XT/AT or compatible
- Color Graphics Board and Monitor (CGA or EGA)
- 640 K RAM.
- Local storage such as hard disk or Bernoulli Box or access to network storage.
- IBM or compatible dot matrix printer which supports the IBM extended graphics character

REFERENCES

Deming, W. Edwards _Quality, Productivity, and Competitive Position,_ Massachusetts Institute of Technology Center for Advanced Engineering Study, 1982