THE SOI QUALITY CONTROL PROGRAM

Thomas M. Durkin and Otto Schwartz, Internal Revenue Service

This paper describes the Quality Control System applied in the Statistics of Income (SOI) Program for Individual returns (Forms 1040 and 1040A). Twelve processing locations are involved. The major SOI processing phases are data abstraction from Federal tax returns, key-entry, consistency testing including error resolution, and tabulation. Different quality techniques are used at various points in processing. Flexibility is built into the System in order to accommodate diverse taxpayer reporting variables and resource fluctuations.[1]

This paper may prove useful to others who work with multi-phase and multi-location processing. It may also assist users of the end product by giving them a better understanding of the kinds and sources of error that are reflected in statements of the limitations of the data. Data are given which shows the product improving over time and also the effectiveness of certain techniques. Some reasons for errors are given, as well.

BACKGROUND

As already noted at this session, Federal tax returns are filed at ten Service Centers serving specific geographical areas. Data from the returns are transcribed onto transaction tapes. For revenue processing purposes, tax payments are accounted for, records of filing are made, and return data are recorded for use in selecting returns for audit, etc. SOI sampling is done by matching sampling criteria data in the computer against the transaction tapes, either at the National Computer Center in Martinsburg or at a Service Center.

The SOI Quality Control Program was initiated in 1962 at the time revenue processing was being converted from manual processing at District Offices to automated data processing at regional Service Centers. At that time the position of Service Center Statistician was established at each center to monitor the decentralized operation. Prior to that time, the manual editing portion of statistical processing for SOI was done in the Statistics Division. The program called for a sample of completed work from the Service Centers to be sent to the Statistics Division where experienced clerks performed two independent verifications on selected items. These were compared with the original edit sheet from the Service Center. If two or more edit sheets agreed it was assumed that they were correct; if the third one was different, that edit sheet was assumed to be in error. Feedback consisted of a statistical report, usually issued well after the end of the program. The reports had limited item detail.

In 1967 the program changed so that only one independent verfication was performed, but all codes and items (data elements) were verified. These results were sent to the Service Centers on a continuous (generally weekly) basis. If there was any difference which the Service Center did not understand, a problem referral slip was submitted for resolution by the Statistics Division. This program change gave the supervisors and procedure writers something definite to take corrective action on. A negative factor was lack of timeliness of feedback. The processing was well underway before the initial feedback was transmitted and much of the feedback arrived after processing was completed.

Edit verification was also introduced in 1967. A review of work is made by a peer reviewer and, if necessary, corrections are made by the original editor. In 1974, after a pilot study at the Memphis Service Center, we began the practice of printing some transaction tape data on the SOI edit sheet.[2] The SOI editor merely verified the data and corrected it if necessary. This resulted in a lower error rate and better control of the sampled returns. In 1978, an error register was added as an enhancement to the system. This enabled the Service Center to correct errors, detected in the consistency tests, by referring to the tax return itself.

The procedural changes are, of course, not the only factors which affect the quality of SOI products. The training given at the National Office and the Service Centers, the continuity of assigned personnel, and the relative complexity of the SOI program in any given year, all affect

Table 1.--Editing Quality Form 1040 Edit Sheets as Shown in the Quality Measurement Program

Tax Year	Percent of Edit Sheets Defective	Defects per 100 Edit Sheets
1966	26.0	65.1
1967	7.7	18.8
1968	11.9	23.3
1969	10.3	18.5
1970	10.5	16.9
1971	5.6	10.6
1972	5.8	12.0
1973	7.8	14.4
1974	3.9	6.8
1975	2.8	5.3
1976	3.7	5.7
1977	5.4	9.1
1978	3.4	5.3
1979	5.9	9.3

the error rate. Quality Measurement figures show total defects to be generally declining, although the introduction of new items to the SOI program mitigates against the trend in particular years. The data are shown in Table 1.

PREPRODUCTION ACTIONS

Quality control is an important consideration during planning for our user meetings.[3] In some cases, items from the tax forms are planned for inclusion in the program prior to new tax legislation; in other cases, items are added after the user meetings, particularly if the subject becomes of sufficient interest. It is important to recognize that new items are more prone to diverse taxpayer reporting. Consequently, we may have to make substantial edit instruction changes for new items after the SOI program has begun, as new problems are revealed.

Instructions--The instructions for Statistics of Income processing of returns are published in the Internal Revenue Manual (IRM). These detailed instructions cover all operations at the field processing locations (Service Centers and Data Center), including shipping between locations. Explanations include special instructions for accounting terms which may be applicable only to certain areas of the country, as well as the handling of disparate returns. The instructions are being designed to break out the "core" section of each program which does not change from year to year, in order to minimize the annual costs of reproducing the instructions.

Training--Training is conducted by the Statistics Division subject matter statisticians and economists.[4] Two experienced editors come from each Service Center to a central location for about one week of review each year. New data requirements are explained; background is given; and the importance of quality data is emphasized. The participants find these sessions to be very useful. In most cases, the same personnel from the Service Centers process the same tax returns for both revenue and statistical processing purposes and are thus quite knowledgeable concerning taxpayer reporting behavior.

Preproduction Consensus Sample--This year for the first time we prepared a package of 110 returns (Preproduction Consensus Sample) in the 1040 area for each Service Center.[5] These were selected from the prior year New Item Sample [6], the current Preproduction Sample, and the current year Taxpayer Usage Study Sample.[7] Because these preproduction returns had not completed revenue processing, the computer-printed data was entered manually. The package of 110 returns was edited at each of the ten Service Centers by as many as possible of their editors who had completed training. A consensus edit sheet was decided upon for each return at each Service Center by the supervisor and senior editors.

This was sent to the National Office where it was compared with the National Office version. Feedback was sent promptly to the centers and, where necessary, the instructions revised. The results were useful in providing prompt, early feedback. Table 2 shows the reasons for differences in the 1980 Preproduction Consensus Sample.

Table 2.--1980 SOI Preproduction Consensus Sample

Edit sheets with differences	147
Number of differences	249
Error rate	13.4%
Reasons for differences:	
Training problems	4
Unclear instructions	15
Editor errors	230

Analysis of the results showed that 67% of the errors attributable to editors were concentrated at three Service Centers. Editors were making a large number of careless errors of omission (not going into attached schedules.) We were able to bring the detailed error analysis (citing specific items) to the attention of the Service Center Statisticians within five calendar days after receipt of their shipments. Omission errors arise chiefly if the employee is working too hurriedly to satisfy improperly a production goal (self or management initiated) by skipping some required operations. For the Individual program for SOI 1980 the majority of the Service Centers started the processing late. The cutoff date was made earlier to improve our service to our users.

However, the Service Center scheduling had already been done and all Service Centers were not able to adjust. These actions reacted negatively and are reflected in the Preproduction Consensus Sample at some centers. Errors of omission always have been a large source of error. They are difficult to correct because of the psychological and attitude factors involved. Table 3 shows our experiences from 1974 to 1979 for errors attributable to editors.

Table 3.--Percentage of Errors by Type Attributable to Editors by SOI Year

Tax Year	Items with Entry Required		No Entry Required But Entry Made
	Omission (1)	Incorrect Entry (2)	(3)
1979	41	48	11
1978	40	44	16
1977	44	49	7
1976	27	54	19
1975	39	47	14
1974	42	53	5

PRODUCTION AND POSTPRODUCTION ACTIONS

Quality control attention must continue during the production phase. This is where the anomalies of the data appear and have to be acted upon; we must be particularly on the lookout for systemic errors.

Edit Verification Sample--The Edit Verification Sample is drawn by a continuous sampling plan based on the review of an editor's work by another editor, referred to as a verifier.[8]

This work is given 100% inspection until a certain number of consecutive edit sheets (clearance number, i) are found defect free. The verifier then inspects on a sample basis such as a frequency (f) of one out of five (f = 1/5). When a defective edit sheet is found, one hundred percent inspection is then reinstated and the process repeats. The system has been very successful. Reviewing each other's work increases communication between the editors and results in a more uniform product. Table 4 shows the error rates for edit sheets which were subjected to verification compared with those which were not.

Table 4.--Percentage of Defective Edit Sheets Showing Comparison Between Unverified and Verified Edit Sheets

SOI Year	Edit Sheets	
	Unverified	Verified
1979.....	7.30	5.31
1978.....	4.84	1.82
1977.....	7.00	4.56
1976.....	4.59	4.67
1975.....	3.19	2.55
1974.....	5.73	3.34

Errors found during verification are corrected. In every one of the last six years, except one, the percent of defective edit sheets was lower in the verified portion. (The anomaly of the one year could be due to sampling variability.)

Although historically we have been receiving counts of the gross number of errors corrected, we have not tallied what type or what items are involved. After discussion with several Service Center statisticians, we have instituted the following new procedures in the Edit Verification System:

1. A tally of errors by edit sheet field will be made each week at each Service Center editing unit. These data will be entered on a blank edit sheet.

2. The Service Center statistician or unit supervisor will phone the Statistics Division each week with the top five errors in terms of frequency at their center. The edit sheet with error tallies will be transmitted to the Statistics Division with the Weekly Verification Summary Report.

3. The Statistics Division project manager, in conjunction with Quality Control Section, will review the error tallies and take steps to correct the apparent deficiencies by further training, new instruction, and/or systemic adjustments.

These changes should increase the usefulness of the Edit Verification process by enabling us to obtain data about the major errors (particularly the ones which are connected with the new program) at an early time in processing. We can use this information to take more timely corrective action as necessary.

Key-Entry Checkpoint--Over a period of years, we have had difficulty implementing the Service Center key-entry instructions, and we have had no formal feedback process. Our instructions are written to be compatible with processing at the IRS Data Center which uses different key-entry machines than the Service Centers. This has caused a need for changes to our instructions midstream and also caused considerable confusion at the beginning of the key-entry process.

To correct the situation, we are consulting with the Service Center statistician to meet with the key-entry supervisor at least weekly at the beginning of processing.[9] If any difficulties are being encountered, it will be the responsibility of the Service Center statistician to contact the computer specialist involved, so that any difficulties can be rectified quickly.

Service Center Consistency Testing And Error Resolution--At the present time, the data in the Form 1040 program are subjected to an initial error resolution which is performed at the Service Centers with the returns and edit sheets present.[10] Only limited analysis is now being made of the results of error resolution.

Depending on the volume, we are having the Service Center statistician review the error

Table 5.--SOI 1980 Individual Program as of August 1981

Service Center	Total Volume Input to DDES (1)	To Error Register	
		Number (2)	Percent (3)
I.....	12,027	2,373	19.73
II.....	8,062	749	9.29
III.....	4,392	572	13.02
IV.....	9,740	1,936	19.88
V.....	5,730	728	12.71
VI.....	14,583	2,481	17.00
VII.....	10,636	1,668	15.68
VIII.....	7,722	968	12.54
IX.....	14,817	4,070	27.47
X.....	5,351	878	16.41
Total	93,070	16,423	17.65

NOTE: DDES See Footnote [11]

printouts on either a 100% or a sample basis in an effort to discover patterns of errors. This is then fed to the editing or key entry supervisor and brought to the attention of the Statistics Division project manager to determine if there are implications for other processing centers. Table 5 shows the total volume of edit sheets which printed out on the error register.

The high incidence of errors shown for Service Center IX was primarily due to a computer operator error. As soon as the Service Center statistician received these data, he took appropriate action.

Concurrent Consistency Testing--In addition to the Service Center consistency testing for Forms 1040 and 1040A there is extensive consistency testing at the Data Center without the benefit of the tax return. It is not feasible to have the 1040 Forms available at the Data Center in volume. For returns other than 1040's, several changes will alleviate some of our problems in the Data Center error resolution procedures. The first involves a change in internal controls at the IRS Data Center. If the return is sent there for editing, the work flow is being changed to make the return available for error resolution. Where there is microfilm of the return available, this, of course, can be utilized. Also, the consistency testing will take place concurrently with editing so that the returns will be available for error resolution.

Information Listings--The present practice in the Individual area (and, indeed, generally in SOI) is to identify returns with unusual conditions which appear on information or consistency test listings or which surface as a result of table review.[12] Since obtaining these returns is a time-consuming and expensive operation, we will also utilize them for other purposes such as in the Preproduction Consensus Sample as well as in the design of the following year's edit and consistency test instructions.

Error Measurement Approach --The accumulation of Error Measurement data in our work operation enables us to appraise whether we are moving in the right direction, whether or not our procedural changes are improving the quality of the statistics. Our task is to estimate statistically how a large number of respondents react to an administrative document which is not designed with statistics in mind. The users of our statistics, who are using them as source material, have a right to know the extent of nonsampling error so that they can take this into account. We plan to maintain an historical series, at the National level, of error data accumulated chiefly through the Quality Measurement program.

We have accumulated a significiant amount of data over the years through our Quality Measurement program. We know what items continue to generate errors. Table 6 for example shows what items on the edit sheets were more frequently found to be in error from 1974 through 1979.

Table 6.--Average Error Rate for Items on Edit Sheets Expressed as a Percentage of Frequency of Usage, 1974-1979

Item	Average Error Rate
Unemployment compensation	0.31
Other income	0.29
Moving expenses	0.26
Taxable portion of unemployment compensation	0.23
Occupation of self	0.22

The Quality Measurement Sample is randomly selected after edit verification, or after Service Center error resolution, for programs which have on-line consistency testing such as in the 1040 program. It is a multi-purpose sample which not only measures the quality of the editing process, but also provides a vehicle for feedback to the submitting centers, thereby helping to insure uniformity in the interpretation of instructions among the various processing sites. Since we are putting more resources into the feedback from the new Preproduction Consensus Sample, which specifically addresses uniformity among processing locations, we are designing the Quality Measurement sample to measure quality at the National level. We are at the same time reducing Service Center samples since the resources are going into the Preproduction Consensus Sample. Some feedback will be provided to the Centers from the Quality Measurement Sample; however, our principal goal is to develop more information in the SOI texts on data limitations.

ADDITIONAL POST PRODUCTION ACTIONS

Edit Verification and Operating Characteristic (O.C.) Curves--Because of the uneven flow of work to the unit doing the Quality Measurement review, the results of the prior-year program may not be available at the time the Internal Revenue Manual procedures for the next year go to printing. Until these results become available, we generally keep the prior-year edit verification scheme. However, as soon as the results do become available, we analyze them and make our judgment based on the use of Operating Characteristic (O.C.) Curves. These give the probability of accepting the product on a sampling basis and the average outgoing quality. They are plotted as functions of incoming quality, prior to verification. Naturally, other factors enter into these decisions, such as complexity of program, use of experienced personnel, initial reports of verification, etc. We are able to notify the Service Centers as soon as our decision is made and then the Statistician at each Service Center implements the change.

Tying Together Cost Data--While our work is in process, we receive volume and staff-hour data on a weekly basis. These reports are monitored and

management is kept informed. Since Edit Verification accounts for the biggest share of the Quality Control costs, particular attention is paid to see that the proper sampling plan is being utilized. Sometimes, editors are kept on 100% review when they should not be doing 100% review.

The other main element of costs are for the Quality Measurement program. The errors are tallied by item. If the number of errors detected is lower than what we think is acceptable, we are able to reduce the size of the Quality Measurement Sample and review fewer returns.

FUTURE PLANS

We have reviewed our present practices in our Quality Control System and find that we have complied with recommendations given by authors writing about a Total Quality Control Plan.[13] Although they concern themselves with manufacturing operations, the principles are equally applicable in our paper-work operation and its later conversion to magnetic tape. Planning must be done. The employee must know what to do. There must be a corrective mechanism built in to correct the design errors as well as the employee errors. The System must be alert to systemic errors. Last year's errors must be this year's points of corrective action.

Nevertheless, there are shortcomings in our System. The major deficiency is the lack of use of the data accumulated in the Quality Measurement program. Prior to this time, the data has been used primarily for quality improvement. They were not really used to describe the limitations of the various SOI statistical series. (This was pointed out by Harry Grubert in writing for the Report of the President's Commission on "How Much Do Agencies Know About Error Structures?"[14]. Since such data have now been accumulated for several previous years, we plan to tabulate, analyze, and evaluate them and make the results available to our users. In future years we will make this operation part of our basic program. In this manner we will be able to help our users make more meaningful use of our data.

In the analysis of quality costs, it is evident that a philosophy of "Do it right the first time" is generally more cost effective than a rework operation. Reworking is often very time-consuming and plays havoc with the schedule of later production steps. The changes we have made this year (Preproduction Consensus Sample and revised Edit Verification procedure) are steps in the right direction. We will continue to explore other procedural changes which may be able to give added emphasis to this approach.

ACKNOWLEDGEMENTS

Because this paper covers the entire gamut of SOI processing, space limitations prevent acknowledging all personnel who helped the authors in the endeavor. Particular recognition must go to Ralph Bristol of the Office of Tax Analysis of the Treasury Department. He provided helpful comments as well as leading a thought provoking discussion at the A.S.A. meeting in Detroit. Several typists were involved but Joyce Coleman did the Final Copy. All members of the Quality Control Section were involved and were the source of the statistical data. Many other personnel of the Statistics Division were very helpful as were the Service Center Statisticians and the Data Center Statistician. Their prompt response to numerous questions about their areas was very helpful.

NOTES AND REFERENCES

[1] With the large population which we sample from (over ninety-million individual returns) and the variations in reporting, the process is sometimes described as "chameleonic." In fact the changing nature of the environment originally led us to title the present paper "Total Quality Control For a Chameleonic Input."

[2] Figure 1, items 3 and 4.

[3] Figure 1, item 1.

[4] Figure 1, item 2.

[5] Figure 1, item 2a.

[6] At the begining of edit processing the Individual returns for Statistics of Income, the Service Centers are instructed to select a New Item sample by screening for returns containing an entry for designated items which are new for that tax year. These returns and edit sheets are photocopied and sent to the Statistics Division which uses them in revising procedures, including consistency tests, particularly for the following year.

[7] The Preproduction Sample consists of 110 unedited returns (11 from each Service Center) which are sent to the Statistics Division for use in the Preproduction Consensus sample. The Taxpayer Usage Study (TPUS) is based upon a systematic sample of 1040 returns taken at the Service Centers as soon as the returns are available for initial processing. The sample is also used for a variety of special studies. The volume was such that we had enough returns to yield a selection for the Preproduction Consensus Sample.

[8] Figure 1, item 4a.

[9] Figure 1, item 5a.

[10] Figure 1, item 5a.

[11] Direct Data Entry System (DDES) is key-entry system used at the Service Centers.

[12] Figure 1, item 7a.

[13] Adams, Clifford C., What is Total Quality Control? Industrial Quality Control, 1966, 22, 341.
Adams breaks down Total Quality Control into eleven sub-functions applied primarily to manufacturing operations.
Cue, Dale E., Some Frustrations and Difficulties in Applying the Total Quality Control Concept, Industrial Quality Control, 1962, 18, 13.

Cue analyzes the functions which should fall under the quality control umbrella for a manufacturing operation. He stresses that each activity should be meaningful and productive. He considers each activity in terms of "Does it add value?"

DiPaolo, E. John, Quality Attitudes-Turn Concepts into Benefits, Industrial Quality Control, 1962, 18, 49-51.

DiPolo gives an automobile manufacturer's approach which stresses cost of quality. Six examples are given. These show practical applications of statistical quality control techniques.

Feigenbaum, A.V., Total Quality Control, Annual Technical Conference Transactions, American Society for Quality Control, Inc., 1959, 311-315.

Feigenbaum brings out that the twin objectives of product quality and lower quality cost can be achieved only by giving attention to all stages of the production cycle.

An annotated bibliography on five subject matter areas of Quality Control: Long Range Planning, Statistical Methodology, Systems Analysis, Total Quality Control, and Motivation Including Quality Control Circles, is available by contacting the authors.

[14] Grubert, Harry, "How Much do Agencies Know about Error Structures?" Chapter 4, Volume II, Federal Statistics: Report of the President's Commission. 1971

Figure 1. Flow of SOI with Related Quality Checks

SOI Processing Operations		Quality Control Operations	
Operation	Description	Operation	Description
1. User Meetings	Division Management meets with users to determine program content	1a. Give data on taxpayer reporting	Supply information on the quality of particular items if available.
2. Processing Instructions and Training	Instructions are prepared for I.R. Manual. Division conducts training for Service Centers.	2a. Preproduction Consensus Sample	A package of 110 returns is sent to each Service Center where they are edited. These are reviewed in the Division and Feedback given.
3. Computer Printed Edit Sheets	Data used in Revenue Processing, which is also in SOI, are printed.	3a. Master File data are analyzed	Extensive Systems Acceptability Testing is made, using live data.
4. Editing	Editor reviews computer entries; adds data items as required.	4a. Edit verification New Item Sample	Work is reviewed by a peer editor on a sample basis. Errors are corrected. New Item Sample. (See Footnote [6]).
5. Transcription Key-Entry	New data items and corrections to computer printed items are Key-entered.	5a. Key-entry checkpoint. 100% review of transcription. S.C. consistency tests checking editing and transcription. Error Resolution.	Service Center Statistician checks Key entry procedure. Error register printed, corrected and reinput. Quality Measurement Sample selected.
6. Ship tapes to Detroit Data Center.	Combine tapes at Data Center	6a. Validate tapes	Read tapes and reconcile counts with Service Center Counts.
7. Create Accepted File	Accepted File is SOI Data Base. File builds as returns clear Consistency tests.	7a. Consistency Testing Error Resolution Information Listings.	Full scale consistency testing is performed. Some error conditions are computer corrected; others are corrected manually. Returns with unusual characteristics are printed to be reviewed by professional staff. (as "Information conditions.")
8. Weighting Program	Data are weighted based on population counts from Martinsburg.	8a. Sample counts are verified	Comparison and analysis is made with Service Center counts and Number of Returns to be Filed (IRS DOC 6186A)
9. SOI Tables	Tables are produced at Detroit Data Center and sent to the Statistics Division electronically or by other expeditious means.	9a. Systems Acceptability Testing and Table Review	Tables are tested for conformity with specifications and reviewed for subject matter content.