STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

Working Paper No. 11
English only

CONFERENCE OF EUROPEAN STATISTICIANS

UN/ECE Work Session on Statistical Data Editing
(Rome, Italy, 2-4 June 1999)

Topic (i): Measuring the impact of editing in various phases of statistical survey processing

A COMPARISON OF EDIT FLAGS FOR COMPUTER ASSISTED TELEPHONE INTERVIEW VS. MAIL

Submitted by the U.S. Department of Energy[1]

Contributed paper

I. INTRODUCTION

1. A census survey of businesses retailing or reselling major petroleum products was historically conducted every three years. Due to major budget reductions, this mail survey of approximately 22,000 companies was changed in frequency to every four years with even further reductions in funds, starting with the 1998 reference year. Two major cost areas of the survey were targeted to potentially reduce the cost of the survey. One of these areas was the editing and response correction portion of the survey processing. Approximately 40% to 50% of the companies were thought to have failed either: (i) the control data quality (CDQ) edit, which identifies the inconsistencies in the company's name, identification number, address, phone, and operating status, or (ii) the volume data quality (VDQ) edit, which validates the reported state level sales volumes of retail and wholesale petroleum products with respect to the previously reported (historical) volumes.

2. Survey forms that failed the control edits were flagged with CDQs and were called to obtain the information necessary to complete or correct the record. Survey forms that failed the volume edits were flagged with VDQs and intended to be called, but were prioritized below CDQs. While the survey aimed to resolve all VDQs, as well as contact all non-respondents, in reality, this was never accomplished in the past, due to the large number of failures, and other priorities.

3. Table 1 below illustrates the type of control edit flags and the number of failures at a certain point in the processing cycle for the 1994 survey. Some CDQ failures are non-verifiable and represent definite errors that require a change to one or more of the items. When the edit flag is resolved by changing the response, the CDQ goes away, so, a cumulative count throughout processing is not provided. The other CDQs are query edits and do not necessarily require a change to the response. These flags may be verified if the response is confirmed as correct.

[1] Prepared by Paula Weir.

Table 1. Control Data Editing Flags, 1994

CDQs	Verifiable?	CDQ Flags to be be reviewed (as of 6/2/95)	CDQ Flags verified (as of 6/2/95)
2,13, 22, 24, 66, 67	Verifiable CDQs	74	812
	Non-verifiable CDQs		
1, 3, 4, 6, 8, 9, 23 29, 30, 41, 45, 46, 82, 83	Incomplete/missing status (respondent error)	84	n.a.
7, 28, 43	Processing conflicts	3	n.a.
5, 12, 40, 42, 43, 80, 81	Manual coding needed	74	
	Total of flags	235	812
# CDQs/#Forms	Rate of flag occurrence	.98%	3.40%

4. Table 2 illustrates the type of volume edit flags and the number of failures for the same survey.

Table 2. Volume Data Edit Flags, 1994

VDQ	Description	# Verifiable to be reviewed	# Non-verifiable to be reviewed
1	Only product totals reported	n.a.	80
2	Sum of details not equal to reported total	n.a.	1358
3	Prior, no current	3440	n.a.
4	Low compared to prior	557	n.a.
5	High compared to prior	505	n.a.
6	Current, no prior	2845	n.a.
7	New state compared to prior	328	n.a.
8	Missing state volumes compared to prior	420	n.a.
9	Expected propane volumes	472	n.a.
A	Oxygenated should be zero	n.a.	509
B	Oxygenated too high	414	n.a.
C	Oxygenated too low	221	n.a..
Total		9202	1947
Rate	Number of VDQs/Number of forms processed	38.5%	8.1%

5. Again it should be noted that these counts represent work outstanding as of a selected point in the processing cycle. Counts of flags over the entire processing cycle would be higher. Despite the large number of failures, which in turn resulted in such a large number of telephone follow-ups, it seemed reasonable to consider conducting the survey as a CATI survey. Phone numbers were available for approximately 93% of the companies. Companies in the previous survey were able to correct their data during telephone follow-up. It followed that the majority of companies should be able to respond initially by telephone. Most large companies were exempt from the survey because of their filing status on an annual survey from which their data could be taken. As a result, most respondents were expected to report data for only a couple cells. The frequency of companies by the number of states and cells reported in the previous survey validated this notion. In 1994, 91%

of the companies filing the survey reported in only one state. Another 7% reported in two states. Furthermore, 96% of the filing companies reported five or fewer cells for all states combined. As a result, it was decided that a pilot survey be conducted that compared the efficiency of CATI in collecting and editing data simultaneously versus a mail survey for this particular data collection effort. This pilot study would not only measure the overall effectiveness of CATI, but would also serve to pre-test a CATI survey instrument.

II. THE PILOT STUDY

6. The pilot study used the previous 1994 survey as the starting point. Companies who reported an active status at that time and were not reporting on the annual survey were considered to be in scope. This resulted in 20,419 companies that could be sampled for the pilot.

7. Two variables from the 1994 survey were then used for stratification: (i) number of states reported and (ii) number of cells reported. The pilot survey used a matched pair design with each sampling unit matching on the two stratification variables. These variables were considered to be the determining factors in the respondent's decision to complete the survey by phone or by mail. The previous survey's respondents, as described above, were then allocated to each stratum. A random sample of 500 each for CATI and for mail permitted 95% confidence intervals on estimates with plus or minus 5%. The majority of companies were one-state companies. Two and three state companies were also expected to behave differently than those in only one state. Half the sample was therefore allocated to the single state status companies and half to the two and three state status companies. A fixed number was then allocated to each number of cells reported stratum within the number of state groupings, except in the three state status companies where the population was not large enough to achieve the allotment designated. This yielded a total sample size of 568 for each survey instrument mode, as shown in Table 3, sufficient for the confidence intervals stated above.

Table 3. Pilot Sample Allocations by Stratum

# Cells/States Reported	One State	Two States	Three States	**Total**
0	50	25	5	80
1	50	25	10	85
2	50	25	35	110
3	50	25	21	96
4	50	25	22	97
5 or more	50	25	25	100
Total	300	150	118	568

8. The design was intended to allow t-tests for significant differences between the two instruments for matched pairs, as well as yield overall and survey process cost estimates, and some quality estimates for the two instrument populations. However, it was also intended that the pilot survey reflect the upcoming full survey and survey respondents' behaviors. In that survey, it was decided that respondents could not be forced into one or the other instrument modes. While a respondent would be designated as CATI or mail, the respondent could choose to report either way. A CATI designated respondent would be allowed to report by mail if he so preferred. Similarly, mail non-respondents would be followed up via CATI after the form due date. Total costs and average costs, exclusive of programming costs, were estimated for the two modes for the main processes of the surveys. Quality indicators such as response rates, and edit flag counts were also tracked for the two surveys.

9. While the original CATI designated respondents had a higher response of 520 completed survey forms, a rate of 91.5%, within 15 weeks, and the mail profile respondents had 473 forms, an 83.3% response rate, within 23 weeks, only 65% of the CATI designated respondents chose CATI as the reporting instrument. This resulted in the rate of survey forms processed as mail being 122%. Even though t-tests on differences for matched pairs have not yet been performed, comparisons of the control data quality (CDQ) flags can be made between the two reporting modes by the type of edit check failed across all strata. Any one company could fail more than one check, and some failures are correlated. The percentages that each CDQ is of all CDQs flagged are shown in Table 4. Even though the number of respondents allocated to each instrument was equal, the number of survey forms processed CATI and mail was not equal, so only the percentage comparisons are appropriate.

Table 4. Pilot Control Data Quality Flag Distribution

	% Processed Mail	% Processed CATI	Difference (Mail-CATI)
Verifiable CDQs			
CDQ 2	4.80%	3.60%	1.20%
CDQ 13	0.60%	0.00%	0.60%
CDQ 22	0.60%	0.00%	0.60%
CDQ 24	6.60%	15.90%	-9.20%
CDQ 66	35.90%	51.20%	-15.10%
Total of Verifiable CDQs	**48.50%**	**70.70%**	**-22.20%**
Non-verifiableCDQs			
Incomplete/missing status			
CDQ 1	0.60%	9.80%	-9.80%
CDQ 3	4.20%	0.00%	4.20%
CDQ 4	2.40%	0.00%	2.40%
CDQ 6	1.20%	2.40%	-1.20%
CDQ 8	10.80%	0.00%	10.80%
CDQ 9	0.60%	0.00%	0.60%
CDQ 29	4.80%	3.70%	1.20%
CDQ 30	0.00%	1.20%	-1.20%
CDQ 41	1.20%	0.00%	1.20%
CDQ 82	0.60%	0.00%	0.60%
Total incomplete or missing status	**26.40%**	**17.10%**	**9.30%**
Processing conflicts			
CDQ 7	0.60%	0.00%	0.60%
CDQ 28	0.60%	0.00%	0.60%
CDQ 43	3.60%	0.00%	3.60%
Manual coding needed			
CDQ 5	19.10%	12.20%	6.90%
CDQ 12	1.20%	0.00%	1.20%
Total processing conflicts	**25.10%**	**12.20%**	**12.90%**
#CDQs/#Forms	**28.90%**	**24.10%**	**4.80%**
# Forms with a CDQ/# Forms	**23.70%**	**21.50%**	**2.20%**

10. From the table it can be seen that overall, the number of flags set as a percent of forms processed is 4.8% less for CATI than mail. This translates into 23.7% of the companies processed by mail, vs. 21.5% of the forms processed by CATI, a 2.2% difference, have at least one CDQ failure. Although the total CDQ difference is not dramatic, the most apparent difference is that there are no flags set for twelve of the CDQs for CATI or, conversely, flags were generated for only 8 of 20 types of CDQs possible in the pilot. Six of these CDQs that resulted in no flags for CATI are non-verifiable CDQs which are the result of missing or incomplete status information reported by the respondent. Four more of these are also non-verifiable but are the result of changes in the respondent's status that require manual coding or correction of a new error introduced through manual coding. The other two CDQs are verifiable flags. It appears that these CDQs, which were infrequent for the forms processed by mail, were avoided by the CATI interview. This result may, however, be due to the combined effect of small sample size and the generally low rate of occurrence in the population. On the other hand, flags were set for nineteen CDQs for forms processed as mail, and only one type of CDQ resulted in no flags for forms processed mail.

11. The CATI system flags the responses that fail the CDQ criteria as the responses are entered during the interview and the interviewer requests that the respondent verifies that flagged response. As shown in Table 4, some CDQs are verifiable and others are not. Verifiable CDQs can be resolved by either changing the response, in which case the CDQ goes away, or, verifying the CDQ by marking it with a "V" in the appropriate field. Non-verifiable CDQs require that the response be changed and then the CDQ goes away. CDQ 66 (verifiable) constitutes the largest percent of the CDQ flags (35.9% and 51.2%) for each type of form processing, but its share of all CDQs is 15.1% more for forms processed CATI than for processed mail. This flag, which is set when a company name is entered that is different than the original name, requires the interviewer to verify if just the name changed, or if the original company was sold or merged. If the name change is minor and no sale has occurred, the change is verified. If a sale has taken place, or the original company has a completely different name, the original name is put back, the appropriate status response corrected and a new company identification number must be issued for the new company. The status correction would then result in another CDQ, CDQ5.

12. It was learned from the pilot results that the interviewer screens did not appropriately notify the interviewer to verify name changes, so the scenarios just described did not occur. Because these name changes can easily be verified within the CATI interview, a large number of flags will easily be verified in the full survey, if the flagged rates in the pilot are representative of a full survey. The second largest failure category for processed mail was CDQ 5 (19.1%), but CDQ 24 was the second largest for CATI (15.9%), followed by CDQ 5 (12.2%). CDQ 5 results when there has been a sale and requires that a new company identification number (CIN) be manually assigned to the new or purchasing company. This CDQ is not verifiable but it does not require phoning the company to resolve it. CDQ 24 results if a company reports not currently selling petroleum products but reported volumes in the previous reference period survey. This CDQ is verifiable and, even though the response is not necessarily an incorrect response, it requires validation with the respondent through a phone call. The percent of forms with at least one CDQ flag is also lower for CATI, but only 2.2% lower (23.7% vs. 21.5%). In the case of CDQ 66, while CATI had the larger percent of the share of CDQ flags compared to mail, the rate of occurrence of forms processed that failed this CDQ was 12.6% of forms, a little higher than the rate for mail, 10%. For CDQ 24, the rate of occurrence was 1.9% higher for CATI. On the other hand, CDQ 5 occurred in 2.9% of the forms processed CATI but 5.5% for forms processed mail—2.6% lower for CATI. Because non-verifiable CDQs can not be tracked in the 1994 survey once they were resolved by changing the response (i.e. the CDQ disappears), the rates of occurrence of CDQ flags unfortunately can not be compared between the 1994 survey and the pilot. The same is true of verifiable CDQs that are resolved by changing the data rather than verifying the flag.

13. A comparison of the volume data quality (VDQ) between the pilot survey forms processed CATI and mail is presented in Table 5.

Table 5. Volume Data Quality Flags

	Mail				CATI				% Difference
VDQ	# V	% V	# NV	% NV	# V	% V	# NV	% NV	
VDQ 1	N.A.	N.A.	3	0.5%	N.A.	N.A.	0	0.0%	0.5%
VDQ 2	N.A.	N.A.	82	14.2%	N.A.	N.A.	0	0.0%	14.2%
VDQ 3	333	57.7%	N.A.	N.A.	168	49.4%	N.A.	N.A.	8.3%
VDQ 4	40	6.9%	N.A.	N.A.	27	7.9%	N.A.	N.A.	-1.0%
VDQ 5	41	7.1%	N.A.	N.A.	41	12.1%	N.A.	N.A.	-5.0%
VDQ 6	248	43.0%	N.A.	N.A.	149	43.8%	N.A.	N.A.	-0.8%
VDQ 7	23	4.0%	N.A.	N.A.	21	6.2%	N.A.	N.A.	-2.2%
Total	685	**118.7%**	85	**14.7%**	406	**119.4%**	0	0.0%	**14.0%**

14. Two of the VDQs represent definite errors, non-verifiable (NV), while the other five are possible errors, verifiable (V), also known as query edits. The percents in the table are calculated as percent of survey forms processed, not total VDQs. The CATI mode eliminated the definite errors that occurred on 14.7% of the mail forms. The verifiable VDQs which occurred the most frequently of the VDQs for both modes were VDQ 3, prior no current (57.7% and 49.4% mail and CATI, respectively), and VDQ 6, current no prior (43.0% and 43.8% mail and CATI, respectively). Considering verifiable and non-verifiable VDQs, the total percent flagged for mail, 133.4% (118.7% plus 14.7%), represents 1.334 volume data quality flags per form processed and for CATI, 119.4% represents 1.194 volume data quality flags per form, with the difference in the two modes of 14% or .14 flags per form. Similar to the situation for CDQs, direct comparison to the 1994 survey is not possible. While the pilot survey indicates the number of times a response was flagged, regardless of whether the response was then verified or changed, the 1994 survey only tracks errors at a point in processing. So, a response flagged but then changed in the 1994 survey would result in the VDQ going away and not be shown in the counts displayed in Table 2.

15. The costs attributable to each of the main processes in each mode were tracked separately during the pilot. The costs relating to collecting, capturing, and editing the data are show in Table 6.

Table 6. Comparison of Average Cost for Collection and Editing
($ per survey form)

Process	CATI	Mail
Form and instructions mail-out costs	$.33	$ .33
Form return costs		.33
Interviewer costs	3.96	
Telephone costs	1.18	.36
Data entry costs		.93
Data technician pre-screening costs		.61
Data technician editing cost		3.24
Total	$5.47	$5.80

The costs appear to only be different by the amount of the return postage for the mail respondents. These cost comparisons are rough, however, because of the difficulty in tracking costs by the category of how a form was processed. It is not clear that costs were captured correctly in all cases.

III. CONCLUSION

16. The pilot shows that overall response rates are higher in a shorter period of time for the CATI respondents. Most of the control data edits, and all of the volume data edits, were resolved during the phone interview as compared to the telephone follow-up required to resolve the edit flags for the mail survey. The CATI also appears to have the effect of preventing control data quality errors 2.2% and volume errors 14%. These small improvements in data quality are accompanied by small decreases in cost of approximately 6%.

17. It is expected that the cost savings will be larger if implemented for the full survey, because CATI is more efficient for larger sample sizes. Additional analysis by sampling strata will be performed in the future to focus on whether certain respondent groupings are different in CATI vs. mail. This may help to better identify which respondents to designate CATI and further encourage that group to report CATI, and which respondents to designate mail and encourage mail response.