DATA EDITING AS A MEASURE OF QUESTIONNAIRE QUALITY

Charles Day and Jaki Stanley McCarthy
Jaki Stanley McCarthy, USDA/National Agricultural Statistics Service, 3251 Old Lee Hwy., Rm. 305, Fairfax, VA 22030 jstanley@nass.usda.gov

Key Words: Data Editing, Total Quality Management, Questionnaire Design

1. Introduction

While editing is done in almost every survey, the results of editing are almost never examined to determine what they can tell us about the survey instrument and its administration. In the language of the quality movement, editing results can be used to measure the quality of questions asked in a periodic survey, and these results can be used for continuous improvement of the questionnaire and data collection. Frequent edits may indicate problems with the wording of a question, the structure of the questionnaire (e.g., skip patterns), understanding of survey concepts, training of interviewers, or the respondent's knowledge of particular information.

While much work has been done examining methods to improve editing and devising ways to adjust for missing data, our objective with this project was to try to identify improvements that would reduce the need for this editing and imputation. Our philosophy is captured well in the following quotation from a 1997 article by Leopold Granquist and John Kovar:

> In the quest to reduce errors in survey data, it is essential to look upstream, rather than attempting to clean up at the end. The adage "do it right the first time" is very appropriate. Editing results can be used to advantage in sharpening survey concepts and definitions and in improving the survey instrument design. More resources should be dedicated to these functions in order to help prevent errors. . . . However, we have as yet to see a report on an editing process where this principle has been applied, and which resulted in changes to, for example, the questionnaire. . . .The role of editing must be re-examined, and more emphasis placed on using editing to learn about the data collection process, in order to concentrate on preventing errors rather than fixing them.

This paper presents the methodology and results of our analysis of editing in the U.S. Department of Agriculture's National Agricultural Statistics Service's (NASS's) Agricultural Resource Management Study (ARMS), a survey of farming practices and farm finances. The authors examined original reported survey data and final edited data, noting where and how edits and imputations had been made. Next, we made a more detailed examination of the questionnaires and final survey data for each question that required frequent editing or imputation. Based on this examination, we were able to offer suggestions for changes in question wording, routing, skip instructions, question content, or interviewer training that may increase the quality of the reported data and reduce future editing and imputation for those items.

2. Methodology

The Agricultural Resource Management Study (ARMS) collects information on production practices, costs, revenues, and assets for a cross section of farm and ranch operations. For 1996, the survey was conducted in three phases. The first phase was a screening process, whereby NASS attempted to determine whether operations had a particular target commodity in 1996 and were currently in business. The second phase concentrated on cropping practices and chemical (fertilizer and pesticide) use. There were separate versions of the phase two questionnaire for a variety of different target commodities. The third phase gathered economic information, such as quantities of agricultural products sold, prices received, costs of inputs, and values of assets for a group of specific target commodities.

The ARMS questionnaire is long and complicated compared to those of most other NASS surveys. The survey is administered in a personal interview using a paper questionnaire. Data collection is coordinated in each of 45 State Statistical Offices (SSOs). Interviews last about one hour on average. The questionnaire is reviewed for completeness and accuracy by the interviewer's supervisor, and by statisticians in the SSO before and after a computer edit. This process of hand editing and imputation by a state office statistician is expensive and time consuming. Concern about the resources being used on manual editing and imputation and the effects of editing and imputation on data quality led us to examine the extent and causes of manual editing and imputation.

The basis of this concern was strikingly illustrated by our research. We examined 8% of the questionnaires completed in Phase 2 and 4% of the completed questionnaires from Phase 3. We found 32 questions that were edited or imputed frequently enough to examine in detail. These 32 questions generated 1,520 edits or imputations on this subset of questionnaires. Therefore the total number of edits and imputations made during the

course of this survey is enormous.

For each item on the questionnaire we counted the number of times there were no data (that is, the question was not applicable to the operation), the value was missing (that is, the question applied, but there was item nonresponse), the value was unchanged, the value was manually imputed using notes made by the interviewer on the questionnaire, the value was manually imputed without interviewer notes, or the value was manually edited. (For the remainder of this paper, unless otherwise noted, all editing and imputation referred to is manual editing and imputation done by a state office statistician.)

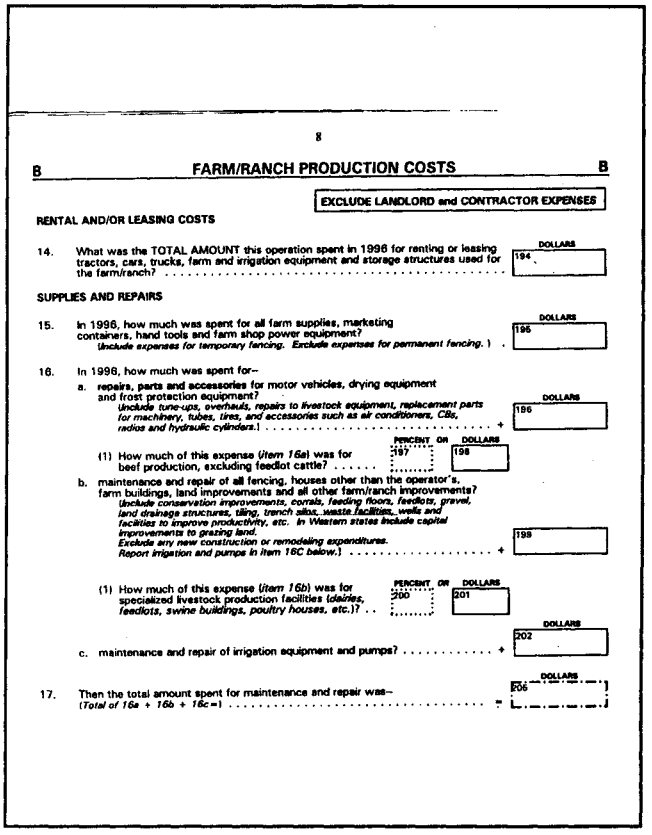
For questions in which imputations or edits were made ten or more times, or imputed or edited values constituted 10% or more of the cases where data were present, a follow up examination of questionnaires and original reported data was done. When we detected a pattern of reasons for such changes, we attempted to recommend improvements to the data collection process that would reduce the need for such changes.

3. Examples

In this paper we present two examples of our results. The first is based on an examination of the editing results from the "Supplies and Repairs" section of the questionnaire reproduced in Figure 1. Sixty-six of the 306 questionnaires with data (21.6 %) in the "total amount spent for maintenance and repair" field were edited. About half (29) of the edits were made because the cost of supplies (Q. 15) was included in the cost of maintenance and repairs (Q. 17). Examination of the questionnaire suggested a solution. The fields for supplies and repairs were grouped together under a single boldface heading "Supplies and Repairs." Question 15 asks how much was spent for supplies in 1996. Question 16 has three parts. Part "a" asks for the costs of repairs to vehicles, drying equipment and frost protection equipment. Part "b" asks for the cost of maintaining and repairing fencing and buildings. Part "c" asks for the cost of maintaining and repairing irrigation equipment. Question 17 then asks for the total of the three parts of question 16. This is indicated by the plus signs and equal sign next to the recording boxes for these items -- a standard NASS questionnaire design format for item sums. It was easy for interviewers to become confused and add supplies (Q. 15) into this total as well. A modification to visually separate the cost of supplies and the cost of maintenance and repair should eliminate these errors. This could be done in several ways. For example, by placing each under its own bold subheading.

The remaining 37 edits were either made because the total field (Q. 17) was left blank, or because arithmetic errors were made in adding the parts of question 16. Between more careful review of questionnaires by

Figure 1. Cost of Supplies and Repairs



8

B FARM/RANCH PRODUCTION COSTS B

EXCLUDE LANDLORD and CONTRACTOR EXPENSES

RENTAL AND/OR LEASING COSTS

14. What was the TOTAL AMOUNT this operation spent in 1996 for renting or leasing tractors, cars, trucks, farm and irrigation equipment and storage structures used for the farm/ranch? ... DOLLARS 194

SUPPLIES AND REPAIRS

15. In 1996, how much was spent for all farm supplies, marketing containers, hand tools and farm shop power equipment? *(Include expenses for temporary fencing. Exclude expenses for permanent fencing.)* ... DOLLARS 195

16. In 1996, how much was spent for--

a. repairs, parts and accessories for motor vehicles, drying equipment and frost protection equipment? *(Include tune-ups, overhauls, repairs to livestock equipment, replacement parts for machinery, tubes, tires, and accessories such as air conditioners, CBs, radios and hydraulic cylinders.)* ... + DOLLARS 196

(1) How much of this expense *(item 16a)* was for beef production, excluding feedlot cattle? ... PERCENT OR 197 DOLLARS 198

b. maintenance and repair of all fencing, houses other than the operator's, farm buildings, land improvements and all other farm/ranch improvements? *(Include conservation improvements, corrals, feeding floors, feedlots, gravel, land drainage structures, tiling, trench silos, waste facilities, wells and facilities to improve productivity, etc. In Western states include capital improvements to grazing land. Exclude any new construction or remodeling expenditures. Report irrigation and pumps in item 16C below.)* ... + 199

(1) How much of this expense *(item 16b)* was for specialized livestock production facilities *(dairies, feedlots, swine buildings, poultry houses, etc.)*? .. PERCENT OR 200 DOLLARS 201

c. maintenance and repair of irrigation equipment and pumps? ... + DOLLARS 202

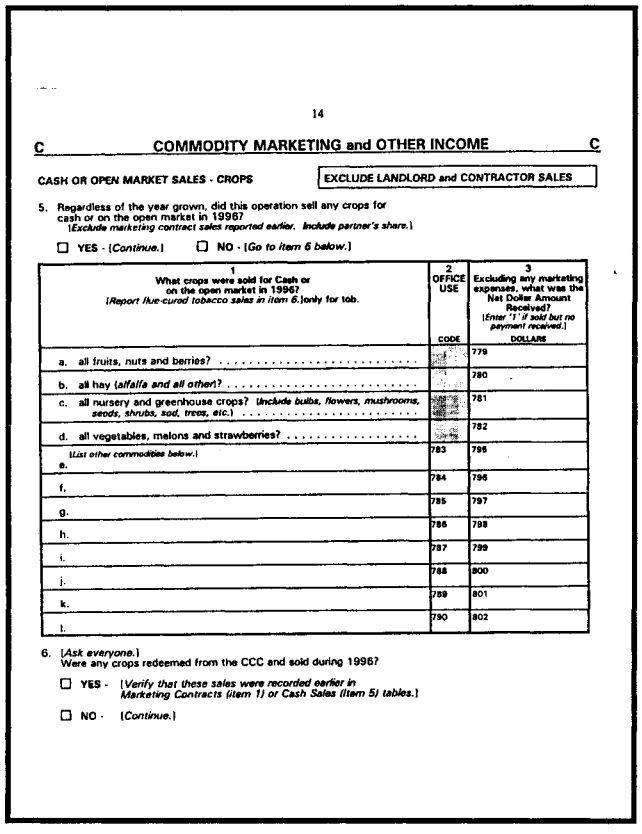
17. Then the total amount spent for maintenance and repair was-- *(Total of 16a + 16b + 16c =)* ... = DOLLARS 506

interviewers and an addition to the computer edit to take care of these errors automatically, the need for so many hand edits should be eliminated. Perhaps a better solution, the total (Q. 17) could be eliminated from the questionnaire since it clearly was not functioning as a check on the sum of items 16a-c. This total can easily be calculated by data users if needed.

Our second example is based on the "Cash or Open Market Sales Table" shown in Figure 2. When recording responses to this question interviewers were to enter data on a line for a specific crop, or enter the data on an open line and supply the name of the crop if it was not printed on the questionnaire. Thirty-eight of the 283 times that data were listed on the "other commodities" lines (13.4 percent of the time) they were changed. These 38 edits appeared on 23 questionnaires. This table often (on 10 of the 23 records) needed to have data moved from one of the free formatted lines to a crop-specific line. For example, on one questionnaire, 'strawberries' and a corresponding dollar amount was written on line e. This had to be crossed out and reentered on line d. so it would be properly summarized as 'all vegetables, melons and strawberries'.

The ARMS survey is one of NASS's longest, and there is considerable pressure on interviewers to complete the

Figure 2. Cash or Open Market Sales Table



14

C COMMODITY MARKETING and OTHER INCOME C

CASH OR OPEN MARKET SALES - CROPS | EXCLUDE LANDLORD and CONTRACTOR SALES

5. Regardless of the year grown, did this operation sell any crops for cash or on the open market in 1996?
(Exclude marketing contract sales reported earlier. Include partner's share.)

☐ YES - (Continue.) ☐ NO - (Go to item 6 below.)

1 What crops were sold for Cash or on the open market in 1996? (Report flue-cured tobacco sales in item 6.) only for tob.	2 OFFICE USE CODE	3 Excluding any marketing expenses, what was the Net Dollar Amount Received? (Enter "1" if sold but no payment received.) DOLLARS
a. all fruits, nuts and berries?		779
b. all hay (alfalfa and all other)?		780
c. all nursery and greenhouse crops? (Include bulbs, flowers, mushrooms, seeds, shrubs, sod, trees, etc.)		781
d. all vegetables, melons and strawberries?		782
(List other commodities below.) e.	783	795
f.	784	796
g.	785	797
h.	786	798
i.	787	799
j.	788	800
k.	789	801
l.	790	802

6. (Ask everyone.)
Were any crops redeemed from the CCC and sold during 1996?

☐ YES - (Verify that these sales were recorded earlier in Marketing Contracts (item 1) or Cash Sales (item 5) tables.)

☐ NO - (Continue.)

interview as rapidly as possible. Interviewers may feel the need to minimize burden on operators to reduce the likelihood of incomplete interviews and future refusals. Sometimes, under this pressure, interviewers get the information down on the questionnaire as best they can. They do not always check if a commodity belongs on a crop-specific line or if it belongs on one of the "other commodities" lines during the interview. These types of edits can clearly be reduced by stressing to interviewers in training that these things should be reviewed and corrected before questionnaires are submitted to the office.

Another problem that arose on five records for this table, but more frequently for other questions, was recording of information in marginal notes, rather than in the boxes provided. This occurred especially frequently when the information was given in a different format or in different units than were required by the questionnaire. These problems can be addressed by reminding interviewers that they are to review the questionnaire carefully after leaving the respondent. During this review, they should move responses to the proper places, instead of leaving this job for statisticians.

4. Recommendations and Conclusions

Using editing information, we have been able to identify areas where changes in questionnaire design and enumerator training could be made to increase the quality of reported data and thus reduce statistician editing and imputation. We examined all the items on our questionnaires, not just particular items, to show how resources could be saved by reducing editing for items throughout the questionnaires.

It was clear to us that for a survey of this complexity, the involvement of the statisticians involved in the design and development of the survey and subject matter experts was critical to the analysis. Often, we relied on the expertise of this ARMS team to explain why a question was being asked in a particular way, or to suggest better ways to ask a question we had identified as a possible problem. We also relied on them to help identify critical interrelationships between the data items within the questionnaire. They are truly the ones who have the knowledge to interpret the results of this analysis. We also believe that it is most appropriate to embed this kind of process evaluation in the work process itself.

While we were able to do this, it was by no means a trivial task. NASS does not currently capture and retain original questionnaire data. This project required re-keying of the original responses and manual review of questionnaires, following the survey. However, in the future, such an analysis could be incorporated into the operational program.

If such an analysis is to be done on a continuing basis in any survey, we suggest that information such as an identifier for imputed values, or original values for edited fields, be captured as part of the data file. Capturing the needed information as part of the regular data collection program would allow analysis of the full data set, rather than a sample, as we did, and these metadata may prove useful for other purposes.

While data editing will likely never be eliminated, we have clearly shown the potential for reductions in the extreme amount of resources currently dedicated to this task. By reducing manual editing and imputation, several things are accomplished. With fewer edits, statisticians and subject matter experts can more easily isolate and focus on true problems in their data. In addition, less time spent editing frees more statistician resources for data analysis or other tasks. Finally, with more accurate data collected and reported initially, the quality of the survey results is ultimately improved.

5. References

Granquist, Leopold and Kovar, John, "Editing of Survey Data: How Much Is Enough?" in Lyberg, Beimer, Collins, de Leeuw, Dippo, Schwarz and Trewin, eds., *Survey Measurement and Process Quality,* p. 430, John Wiley and Sons, Inc., 1997.