# Improving Editing Efficiency: How a Comprehensive Program Interface Reduces the Time Cost of the Comment Review Process

**By**

**Richard Windle[1]**

**Federal Reserve Board, 20th and C Streets NW, Washington, DC 20551**

## Abstract

For complex surveys, one of the most effective tools for reducing nonsampling survey error is interviewer comments. In the Survey of Consumer Finances (SCF), these comments have been particularly useful in explaining unusual respondent situations, allowing editors to alter case data after the interview has been completed in order to restore the data to the state they should have been in had they been correctly gathered originally. However, this method of error reduction is also extraordinarily time consuming, requiring months of careful analysis by multiple editors. Furthermore, in order to even become a qualified editor, extensive training is required, necessitating even more time. Given these twin issues, any method for speeding up the training and editing processes while still maintaining the data quality improvements they generate is worth exploring. With this goal in mind, a system was designed to incorporate survey data, interviewer comments, and a series of data checks into a single, easy-to-use program interface. Perhaps most helpfully, the program also generates financial summary sheets—such as a household balance sheet and an income statement—for the quick identification of anomalies. Using this system, data editors can, at a glance, understand the basic fundamentals of a case, identify potential problems, make corrections, and then check to see that these corrections did not create further issues. The program also serves to encapsulate knowledge about the survey that previously had to be memorized during the training process. This program, named the Editor Assistant (EA), was fully employed for the 2013 SCF and was used to swiftly train three new editors and speed up the editing process. The four- to five-month reduction in required editing time—compared with previous years—has been credited in large part to the EA.

---

[1] E-mail: richard.a.windle@frb.gov. The analysis and conclusions set forth are those of the author and do not indicate concurrence by other members of the research staff of the Board of Governors of the Federal Reserve System.

## Introduction

When conducting a nationwide survey, it is unavoidable that a large spectrum of respondents will be interviewed. Some of these people will have extensive financial knowledge, while others will not. Some will be living in complex situations, while others will have rather straightforward lives. The diversity of respondents makes formulating question wording for gathering complicated data challenging. Any given wording will be confusing to a certain segment of the population. One method for combating this issue is to interview more families in the hope that, through averaging, these data collection errors will be minimized. However, for an extensive household finance survey such as the Survey of Consumer Finances (SCF),[2] this method can become quite expensive. There is also the risk that all or most people in certain situations of interest might make the same mistake, and, thus, no matter how many of them are interviewed, the data error will persist.

Ultimately, then, what is sought is a way for people to provide information in their own words. Since the SCF is interested in actual information—such as income, assets, and debts—and, for the most part, not opinions, the use of interviewer and respondent comments can be safely employed as a tool of data error reduction.

For example, if a person with limited financial knowledge is trying to describe his or her defined benefit pension, but is unfamiliar with the terminology, the person can make the comment "I'm going to be paid a set amount each month, based on my pay when I retire" when the interviewer gets to that section of the survey. This comment more accurately reflects the household's situation than a simple "Don't Know" answer, and the data can be evaluated and changed with this additional knowledge.

Another check on data errors is the collection of similar data at different points during the survey. For example, wage data are collected in both the employment section and the income section, and sometimes these data disagree. Perhaps, during the course of the survey, the respondent will recall something that he or she had previously forgotten, or perhaps slightly different wordings will reveal the complexities of a situation. When these values do disagree, there is sometimes enough evidence to justify one value or the other.

Once the sometimes-conflicting data and clarifying comments are sent back to the SCF group, there needs to be a process for altering that data so that they reflect the true situation of the respondent. This process is called data editing. In order to ensure that data quality is, in fact, improved using this method, a great deal of time and training for editors is required. In years prior to the 2013 SCF, data editing made up a significant portion of the post-collection, pre-data-release period, typically about 14 months. In each of the roughly 6,000 cases, each data point had to be reviewed in light of the comments made, possible issues needed to be weighed, values were changed if necessary, and then those changes were evaluated to determine if they required changes to other variables. The programming code needed to make those changes to the data set then had to be written to a master program, the program was run, data errors checks were conducted, and any problems with the edited data then had to be resolved by the original editors, sometimes weeks after the initial case review.

After reviewing the state of the editing process, the SCF group determined that significant increases in efficiency, as well as reductions in time cost and burden on the editors, could be achieved while still maintaining the data quality gains. The method of this improvement

---

[2] For more information on the 2013 SCF, see Bricker, Dettling, Henriques, Hsu, Moore, Sabelhaus, Thompson, and Windle (2014).

would be the creation of several programs, all easily accessible from a single unified dashboard. This would put all of the tools, data, comments, and error checks immediately at the editor's fingertips.

## The Survey of Consumer Finances

As already briefly mentioned, the SCF is a nationwide household finance survey. It is conducted in the United States every three years and gathers data on household balance sheets, income, and demographics. Interviews are conducted by NORC, an independent research institution from the University of Chicago, usually from April to December of the survey year.

The sample is created using two methods: A list sample is constructed using tax data to cover all income groups, and an area probability sample is used to cover all geographic regions.

Interviews can be extensive. Generally, they take about an hour and a half, but, for particularly wealthy households or households in unusual circumstances, the survey can take up to four hours. All interviews are conducted by an interviewer, either in person or over the phone.

## The Data Editing Process

In total, about 10 percent of the data have been edited in previous surveys.[3] Since edits are made only when they improve data quality, this statistic means that 10 percent of the data in the past have been corrected to more accurately reflect the household's financial situation. As mentioned already, data editing reduces measurement error and makes it cost effective to represent the entire population with a relatively small sample, but these benefits do come with challenges.

The first is a high time cost. The Federal Reserve Board and users want the public data released as soon as possible, but the time required to perform editing is significant. Great care must be taken to ensure that data quality is actually being improved. An editor must review all data, all of the comments made on the case, and the records of contact between the interviewer and respondent. All of this caution and thoroughness takes a significant amount of time.

The second challenge is accounting for the effect that multiple editors can have on data editing consistency. Even with multiple editors, the amount of time required for the data editing process is substantial, but using multiple people means that they all must be trained extensively to ensure that everyone is editing the same, correct way. This training, again, takes a great deal of time, as well as considerable effort on the part of the editors, who must develop an extremely intimate knowledge of the survey so that they can make edits properly. Unless some more advanced system is used, it falls on them to memorize the details of survey data editing.

Both of these data editing challenges can be better met by using a comprehensive program interface—that is, a program that brings together the survey data, comments, summary statistics, and edit-code-generating software into one quick and easy-to-use dashboard.

---

[3] For more information about editing rates, see Bricker, Moore, and Windle (2014).

Such a system saves time by implementing dozens of small, time-saving measures, which, when added up over thousands of cases, create a significant effect. Also, by making editing easier and clearer, it reduces the amount of time required for effective training. It also increases consistency by encapsulating accumulated knowledge, taking some of the burden of memorization off of editors, and making the whole process less taxing.

It was decided, then, to design such a system and implement it for the 2013 data editing process.

**The Editor Assistant**
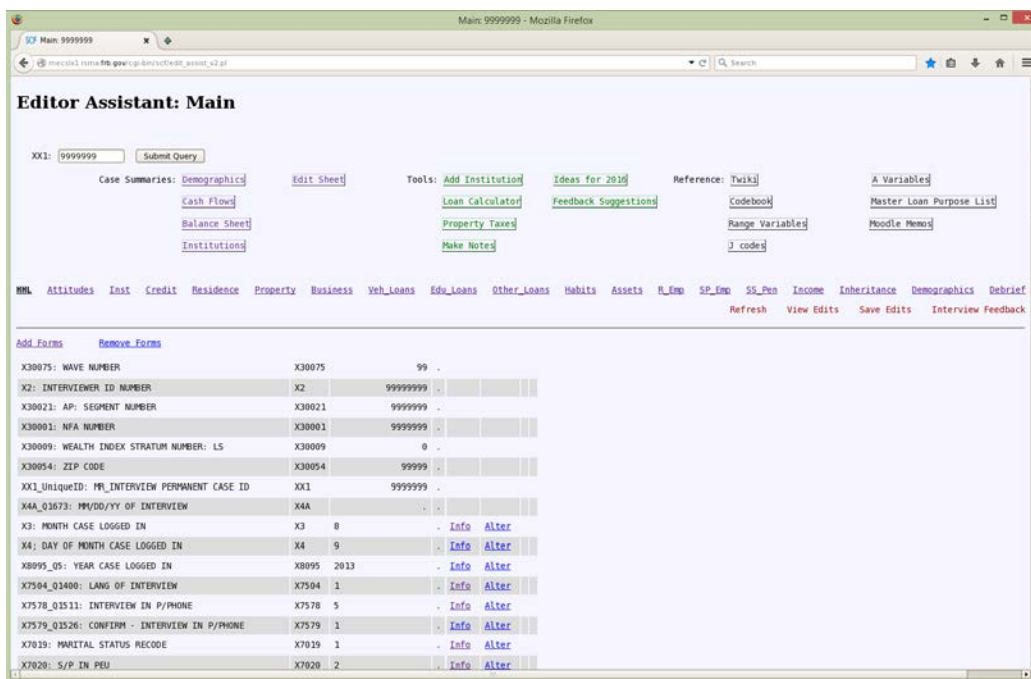


The image above shows the final design of this interface: a comprehensive web form accessed through an internal browser and named the Editor Assistant (EA). Along the top are links to summary information, tools, and reference materials. Beneath the line are the actual data themselves, along with variable summaries and links to edit the data.

Using the EA, editors were able to achieve a decrease in time cost while maintaining or even improving consistency and data quality at each of the three steps of the editing process: (I) pre-editing analysis, (II) editing survey data, and (III) analyzing the effects.

*I) Pre-editing Analysis*

After an editor has been trained, he or she can begin to edit cases. The first step is called pre-editing analysis, which is now conducted by reviewing summary sheets generated by the EA.

It is difficult to understand cases quickly by looking at a column of data. Data are diffuse and disaggregated at this point, but summary sheet programs grab the relevant data from multiple places in the survey and calculate statistics about the case. Before the EA, editors

created and ran their own summary programs, but this work interrupted the flow of editing. An editor would have to leave the data, alter the program to run on the case in question, open up the results, and then finally return to the data. Furthermore, these programs were often designed by the individual editor, for their own purposes, and run as needed.

With the EA, however, standardized programs are run by clicking a single button. The EA alters and runs the program automatically, and the results pop up in another tab in the browser. This technology increases efficiency and consistency, since all editors are now using the same programs.

```
X2201_Q577: OWN_VEH_1: CARS, TRUCKS, VANS, JEEPS        X2201   1      .  Info  Alter
X6691_Q578: OWN_VEH_1: HOW MANY VEHS OWN? (RAW)         X6691   1      .  Info  Alter
X2203_Q580A1: OWN_VEH_1: TYPE VEH                       X2203   1      .  Info  Alter
X7027_Q583A1: OWN_VEH_1: MAKE OF VEH                    X7027   93     .  Info  Alter
X2205_Q582A1: OWN_VEH_1: WHAT MODEL YR                  X2205   2006   .  Info  Alter
X8166: OWN_VEH_1:VALUE                                  X8166   .      .  Info  Alter
X7543_Q585A1: OWN_VEH_1: BOUGHT NEW/USE?                X7543   2      .  Info  Alter
X7540_Q586A1: OWN_VEH_1: WHAT YR BOUGHT                 X7540   2012   .  Info  Alter
X2206_Q587A1: OWN_VEH_1: MONEY STILL OWED               X2206   1      .  Info  Alter
X2207_Q588A1: OWN_VEH_1: MO LN TAKEN OUT                X2207   1      .  Info  Alter
X2208_Q589A1: OWN_VEH_1: YR LN TAKEN OUT                X2208   2012   .  Info  Alter
X2209_Q590A1: OWN_VEH_1: HOW MUCH BORROWED              X2209   5000   .  Info  Alter
X2210_Q592A1: OWN_VEH_1: REG INSTALL/OTH TYPE PMTS      X2210   1      .  Info  Alter
X2212A_Q593A1: OWN_VEH_1: AGREED TO YRS OR PMTS         X2212A  1      .  Info  Alter
X2212_Q594A1: OWN_VEH_1: # YRS                          X2212   5      .  Info  Alter
X2211_Q595A1: OWN_VEH_1: # PMTS                         X2211   .      .  Info  Alter
X2213_Q596A1: OWN_VEH_1: AMT REG PMT                    X2213   100    .  Info  Alter
X7537_Q598A1: OWN_VEH_1: FREQ REG PMT                   X7537   4      .  Info  Alter
X2214_Q600A1: OWN_VEH_1: AMT TYP PMT                    X2214   .      .  Info  Alter
X2215_Q602A1: OWN_VEH_1: FREQ TYP PMT                   X2215   .      .  Info  Alter
X7534_Q604A1: OWN_VEH_1: ON/AHEAD/BEHIND SCHED?         X7534   1      .  Info  Alter
X2216_Q605A1: OWN_VEH_1: MO LN TO BE REPAID             X2216   .      .  Info  Alter
```

As an example, the above screenshot of the EA contains a column of data representing a car loan. Some of the data are easy to understand, but much depends on codes and understanding the variable summaries in the leftmost column. But after a summary sheet program has been run, the data can be summarized.

The balance sheet program takes the car loan information, as well as other asset and debt information from the entire survey, and condenses it to a single page (see the figure above), allowing an editor to clearly understand the state of this household with a single look. There are other summary sheets for income and for demographics, as well as a sheet containing all of the interviewer comments. Used together, the summary sheets enable an editor to quickly come to a good understanding of the household.

Most important, though, these summary sheets give editors quick insight into potential problems in a case. For example, say a respondent lists an Actively-Managed Business Asset in the Balance Sheet Summary and makes the following comment on the Edit Sheet: "My construction business did really well last year, so our income is unusually high." But the income sheet has no business income for that year, only wage income. This inconsistency means that it is likely that income was misclassified and should probably be moved from wages to business income.

For another example, what if a respondent makes this comment on the Edit Sheet: "We're really sad that our daughter is leaving for college soon." On the Demographics page, no daughter is listed. This inconsistency indicates that the household listing is potentially incomplete. It might be required to add the daughter to the listing, since she has not yet left for college.

The summary sheets give editors a quick idea of what needs to be corrected and indicate which sections require special attention as they actually begin going through the data.

*II) Editing Survey Data*

Once an editor has a good idea of a household's situation, as well as an awareness of the case's potential problems, he or she can then begin going through the data and making edits.

Edits can rarely be made automatically, but households are similar enough that combining a smart program interface and editor input can safely reduce time cost without risking data quality. This work is done using tools such as customized forms with instructions and text boxes. For example, adding a checking account requires adding a balance amount and financial institution information, which necessitates entering data into multiple places in the survey. Before the EA, an editor would simply need to know this information or be forced to look up an example of this edit already made to another case, requiring either a significant mental burden on the editor or a considerable time cost. With the EA and one of these forms, however, every variable that needs to be adjusted appears on the screen, along with clear instructions explaining which ones to fill out in different circumstances.



The image above shows an example of such a form—specifically, one for adding or removing a person from the household listing. The data are inputted via textboxes and drop-down menus, allowing the editor to enter new values easily. Instructions for which variables to fill out under slightly different circumstances appear in bold.

Once the editor submits these edits, computer code for making them is automatically generated and saved to a file to be reviewed during the final step of the editing process.

*III) Analyzing the Effects*

After an editor has gone through all of the data and made his or her edits, the final step is to analyze the effect of those edits. How do changes interact with other data collected on the household? Doing this analysis requires entering the saving and error check sequence.

In the EA, the editor clicks a button to run a final review, which runs error checking code that warns the editor if the edits have created some additional problem. In previous years, this error checking code had been run only on large batches of edits, sometimes weeks after they were originally made. So if an error was found, an editor would have to spend extra

time becoming reacquainted with the details of that case in order to fix the problem. With the EA, however, this error checking code is run immediately after each set of edits is made so that an editor can fix problems without delay. Also at this stage, a table of differences between the pre-edit and post-edit data sets created by the edit code is put up onto the screen.

The editor uses all of this immediately generated information to verify that the desired changes were accomplished and no unexpected effects resulted. If there is some discrepancy or error in the code the EA wrote, the editor returns to edit mode, fixes the edits, and then runs the final review again. Once satisfied with the results, the final edits are saved by clicking the "Save Edits" button. This operation writes out the computer code for making those edits to a file and saves it to a special directory, to be applied to the official dataset at a later point. .

At every step in the process, using the EA reduces the amount of time required, increases consistency, and removes burdens from the editor.

**Conclusion**

In summary, then, how much is the EA helping editors? The best measure of the EA's usefulness is the amount of time it takes to complete the post-collection processes. As can be seen in the table below, the EA helped reduce post-survey processing time by about five months relative to the typical amount of time taken previously.

| | | |
|---|---|---|
| 2013 SCF | 860 cases/editor | 9 months |
| 2010 SCF | 930 cases/editor | 18 months (due to 2009 panel) |
| 2007 SCF | 900 cases/editor | 14 months |
| 2004 SCF | 900 cases/editor | 14 months |

Given the slightly reduced case burden per editor in 2013, a few weeks of time reduction could be expected, but not a full five months. Thus, it is reasonable to attribute about four to five months of time saved to the EA.

It should be particularly noted that these time savings were accomplished despite the fact that three new editors joined the SCF group for 2013 and required training. Ultimately, the new editors were able to complete as many cases as the old editors in the same time frame. With the EA, training became much simpler and easier, and mistakes were more swiftly corrected, with minimal time cost for the new editors.

Using smart programming, it was possible to make survey data editing much more manageable for the SCF. The time required and the knowledge burden were dramatically reduced, allowing the data quality gains to be achieved at a much lower cost. Other surveys have resisted engaging in data editing, but, with tools such as a comprehensive programming interface dramatically altering the cost–benefit balance, another look at this practice might be warranted.

## Bibliography

Bricker, Jesse, Lisa Dettling, Alice Henriques, Joanne Hsu, Kevin Moore, John Sabelhaus, Jeffrey Thompson, and Richard Windle (2014), "Changes in U.S. Family Finances from 2010 to 2013: Evidence from the Survey of Consumer Finances," *Federal Reserve Bulletin,* Vol. 100, No. 2.

Bricker, Jesse, Kevin Moore, and Richard Windle (2014), "Examining Interviewer-Respondent Interactions in the Survey of Consumer Finances (SCF)," in *Proceedings of the Survey Research Methods Section, American Statistical Association, 2014.*