UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS
Expert Meeting on Statistical Data Editing
7-9 October 2024, Vienna

Assessment of Manual vs Automated Survey Editing and Imputation

Sean Rhodes, Megan Lipke, Jennifer Maiwurm, Darcy Miller, Denise A. Abreu, Linda J. Young
National Agricultural Statistics Service, United States Department of Agriculture, USA
Sean.Rhodes@USDA.gov

I. Introduction

1. The United States Department of Agriculture (USDA) National Agricultural Statistics Service's (NASS's) mission is to provide timely, accurate, and useful statistics in service to U.S. agriculture. NASS has two primary programs: the Census of Agriculture (COA) and the agricultural estimates program. The Census is conducted every five years, in years ending in two and seven. Census data inform decisions on community planning, company locations, availability of operational loans, staffing at service centers, and farm programs and policies. The agricultural estimates program produces reports and findings on nearly every characteristic of U.S. agriculture. These reports provide timely market-sensitive information. In conjunction, the COA and agricultural estimates reports provide all market contributors accurate supply/demand information for the agricultural sector, which promotes efficiency and fairness in competitive markets.

2. As part of the agricultural estimates program, NASS conducts hundreds of surveys and publishes more than 400 reports every year. Some examples of areas covered in NASS's surveys are livestock inventories, prices paid and received by farmers, farm labor and wages, chemical usage, and demographics of U.S. producers. A wide variety of topics are covered within these different areas. The subject matter ranges from traditional crops, such as corn and soybeans, to agricultural prices of land in farms. The target population size can vary quite significantly from survey to survey. Some survey's target population is less than 50 producers whereas the COA questionnaire is mailed to about 3 million producers. The sampling design, data collection modes, processing, estimators, and publication timeline also varies from survey to survey.

3. One series of surveys that is conducted within the agricultural estimates program is the quarterly Agricultural Production Survey (APS). The survey series occurs four times a year in the months of March, June, September and December. These quarters are tied closely to how crops develop across the United States. In March, the focus is on planting intentions, the number of acres of each crop the producer plans on planting. By June, most crops have been planted unless weather interfered with the producer's plans. So, the producer is asked how many acres have been planted and are yet to be planted of each crop. Another critical piece of information collected is how many of those acres do the producers plan on harvesting in the fall. Thirdly, the September survey is geared towards the annual final acreage yield and production information for small grain crops, such as oats, wheat, barley, and rye. The timing of the survey is intentional because, in a typical year, harvest has been completed. The producer provides information about their final planted and harvested acres as well as final yield. The fourth and final survey, December, focuses on row crops. Any crops that were not collected in the September survey, including but not limited to corn, soybeans, and cotton are collected in December. Both September and December are post-harvest and become the basis for the official estimates of final acreage planted acreage, harvested acreage, yield, and production for each crop.

4. The results and estimates created by these surveys via producers' responses are provided to all stakeholders at the same time. Producers obtain information that can inform them on how best to market their crops. Agribusinesses, which supply the goods, products, and equipment that the producers need to cultivate the various crops, rely on the statistics to assess where they need to be located and what products are needed in the various locations. The statistics are useful in international trade. Other entities, including but not limited to farm organizations, universities, elected officials, extension agents, and private and governmental economists, also use these results. Partners of USDA need solid data to help administer, support, and implement programs.

5. NASS surveys have a three-step data cycle: data collection, analysis, and publication. First, reported survey data are collected from producers. The data collection methods utilized at NASS fall primarily into two categories: computer-assisted and paper-assisted interviews. In a computer-assisted interview, the producer submits their questionnaire online through the NASS website. For paper-assisted interviews, producers are mailed a questionnaire, and they return the results via mail. Both methods may be assigned for phone follow-up if the producer has not responded. For mail interviews, an in-person follow-up interview may be used, for a small subset of records.

6. During the analysis phase of the data cycle, the data are tabulated at the state and national levels. However, prior to tabulation, unit and item nonresponse as well as invalid responses are addressed. That is, the data are edited, imputed, and analyzed prior to tabulation, which produces the estimates. In the final step, the survey estimates are reviewed and then published or released on the NASS website.

7. Editing and imputation is a critical part of the survey cycle. Because this has largely been a manual process, it consumes a large portion of the Regional Field Office (RFO) staff's time (~40%). NASS is continually exploring methods and tools to improve this step from both the data quality and time efficiency perspectives. During the past three years, NASS has taken steps to automate and generalize the current editing and imputation practices. In this paper, a case study comparing the overall operational efficiency and data quality for automatic and manual editing is presented. In Section II, the machine-aided, manual approach to editing and imputation is described, followed in Section III by a description of the automated system being developed. The metrics used to compare the two approaches are presented. The results of the comparison are discussed in Section IV, followed by a discussion of both the successes and challenges with using and implementing a new automated system in the last section.

II. Manual Imputation and Editing Using Blaise

8. The primary system used to conduct editing and imputation for quarterly agricultural surveys is an interactive editing system called Blaise. This system also handles various other tasks throughout the survey process including data collection and system processes. Blaise interactive edit flags errors that violate a survey's edit logic and prompts the analysts who are editing to make manual edits to correct these errors. Blaise's interactive edit provides information on the error yet provides very little information on how to resolve said errors. Editors will use multiple auxiliary resources to help inform how they can correctly resolve the error. All surveys that do not meet the edit logic, which can also include missing or invalid survey responses, follow this process.

9. Blaise's manual interactive editing process can be time consuming and resource intensive for the agency. RFO analysts frequently manage multiple surveys at one time. Consequently, the valuable time spent editing and resolving invalid survey responses can be used for other important survey tasks. The large amount of manual editing required can lead to a variation in editing practices due to an individual's experience, agricultural knowledge, or time constraints.

III. Automated Editing and Imputation

A. Imputation, Deterministic Edits, Automation and Logic (IDEAL)

10. NASS is building a generalized system to help address some of the current challenges that arise from relying so heavily on manual editing and imputation (Lipke 2022). The system being developed is named IDEAL (Imputation, Deterministic Edits, Automation and Logic). IDEAL is a multi-year phased project with the goal of delivering a generalized system to automate and modularize the editing and imputation processes for Blaise surveys. Implementation of this new automated system will decrease RFO analysts' time editing and lead to more time analyzing data and other survey related tasks. Data quality will be enhanced and time burden on the staff will be reduced due to the benefits of the new automated system.

11. One of the initial challenges with developing the IDEAL system was the lack of specific documentation of Blaise code for editing and imputation. To help address this challenge, an automated parser was developed in Python to isolate, record, and classify edit logic. Once the code was documented and reviewed for accuracy the code was then converted to the programing language R. Since this information is now properly documented the edit rules and logic can be more easily standardized, evaluated, and reviewed as the system's development continues to evolve.

12. The IDEAL system has three components working simultaneously. First, the "R Engine" applies edit logic, conducts imputation, and automates corrections. Second, the User Interface (UI) is designed to monitor changes and take any manual intervention still required. Third, the Microsoft Azure Cloud provides the environment in which the R engine and User Interface can function.

B. JIMMY

13. Since IDEAL is a large, long-term project, NASS has set subgoals to implement various components in a phased approach. First, the R engine has been incorporated into the production environment to capture the progress the project has made prior to the full system being prepared. The R engine, which is called JIMMY, applies the post-data collection Blaise edit logic and performs some item level imputations. JIMMY is not an acronym but a way to easily identify the tool when comparing to other systems inside NASS that also use the R programming language.

14. JIMMY was developed with R packages including validate, error_locate, validatetools, dcmodify, deductive and simputation from Statistics Netherlands (E.de Jonge and M. van der Loo, 2019). These packages provide the ability to utilize the parsed Blaise rules and logic to create and apply data validation rules to survey responses. Once the editing rules and logic have been put into place, error resolution occurs in several sections. First, the various rules are applied to the data to determine which item level edits and imputations are needed. Next, the values that require imputation are imputed via means imputation following a normal distribution. Once imputation is completed any remaining errors are corrected via a Felligi-Holt (1976) methodology. Finally, the data are run through the initial step again to identify any edits that require manual intervention. Data still passes through Blaise since only Blaise performs some types of edits including name, operation, and address changes.

15. A JIMMY prototype was first implemented in December 2023 for the Crop County Estimates Survey (CE). Crops CE and Crops APS having overlap in both scope and questions. Thus, successes and challenges from the initial prototype helped frame the vision and plan for implementation of JIMMY for the quarterly agriculture surveys. The focus of this paper is the testing of JIMMY using the 2023 September APS data conducted in preparation for implementation for the 2024 September Crop APS cycle.

III. Methods Used to Compare Manual vs Automated Editing and Imputation

A. Testing 2023 September Crop APS

16. Having promising results from December 2023, NASS took steps to implement JIMMY in September of 2024 for the quarterly agricultural survey. Improvements were made to the testing procedures based on the successes and challenges faced in the production environment. Categorization and relationships among edits continue to help catalogue the edits and records to quickly prioritize where changes need to be made. However, to address challenges, three major changes were incorporated: 1. using a stratified sample, 2. reading simulated JIMMY data into Blaise Beta and SPS edit summary, and 3. RFO qualitative feedback. These changes were needed to optimize the scope and results given the quick timeline and resources of the testing team.

B. Stratified sample

17. A stratified sample of records was utilized to address scope, pace, and efficiency of review. Previous testing focused on a single region of the United States due to the natural organization of states used within NASS. Testing JIMMY on a handful of states did not sufficiently cover all crops, and therefore did not completely test the entire rule set. Initial solutions to this included running the whole sample through JIMMY. Long durations of time between iteration reviews resulted due to the large number of records within the sample. Therefore, the team worked on creating a stratified sample of around 3000 records where each crop was represented at least once in the sample. There was not a seed set when drawing this sample to incorporate some variability of new records and ensure the same 3000 records were not used for all of testing. Reducing the number of records also assisted in iteration speed of updating the rules. Updates to the rules were then made and the team reviewing the data were able to quickly identify whether the adjustments incorporated were correct.

C. Assessing Quality of Data Processed by JIMMY

18. Various methods have been explored and used to test whether the JIMMY system is providing reasonable edits and imputations. Methods that have been found to be most effective include classifying the edits into three distinct categories: incorrect edits, correct edits, and investigate differences. These categories are based on the changes among the reported data, JIMMY data, and NASS final edited data. The reported data is what the producer provided via data collection. JIMMY data are the values that are produced after the various editing rules and imputation are applied to the reported value. NASS final edited data are the fully edited values used in production; these are taken to be the gold standard or truth. These NASS final edited values had the benefit of both manual editing and analysis. These three datasets (reported, JIMMY, and NASS) of all producers' survey responses were then merged at the item level.

19. Once the three datasets were merged, each item-level value was sorted into one of the three categories: incorrect edits, correct edits and investigate differences. Incorrect edits occur when the reported value and the NASS value are equivalent, but the JIMMY value differs. This could arise because JIMMY changed a reported value, and the NASS analyst changed the JIMMY value back to the reported on. Another example is when JIMMY incorrectly removed the reported value, making it missing, but the analysts editing the survey kept the reported value. These scenarios are inaccurate edits by the JIMMY system.

20. Edits are deemed to be correct edits when the JIMMY value equals the NASS final value. For this category, it did not matter whether the Reported value was the same or different than the JIMMY value.

21. Finally, any edits that did not fall into either the correct edits or the incorrect edits categories were named "investigate differences". An example is when the reported value and the NASS final value are not equivalent, and JIMMY has a third and separate number. These edits are generally not as straight forward to categorize as correct or incorrect.

22. For the unit-level analysis, aggregation of the count of the number of edits within a given category are calculated. Hence, records that have a high portion of incorrect edits are easily identifiable. Once identified, the edit rules and logic as documented earlier are revisited and adjusted to reflect what changes need to be made to yield a correct edit. For records with large number of correct edits, rules were reviewed to ensure these records stay constant once edit rules have been adjusted. Finally, records that had a high number of investigate difference edits were not able to be resolved as easily by updating the edit logic and rules. Additional analysis and methodology need to be considered when reviewing these edits.

D. Assessing Workload and Data Quality (from current Blaise & SPS Edit Systems)

23. To capture a larger aggregate picture of how JIMMY was performing during testing, internal systems were utilized. First, data were read back to the Blaise beta testing system. Blaise beta is a non-production sandbox environment that provides a more comparable experience to what analysts see in the production environment with JIMMY data. The tool also provides the ability to create some preliminary analysis on the change in the number of records that meet and do not meet the edit logic. Information created by this tool was used to provide a sense of how much manual editing was introduced or reduced. Next, data were moved to the Survey Processing System (SPS) edit. The SPS edit is a supplemental edit, used for Crops APS, which provides descriptive information about the item level data and its relationship to other items. Output created by these systems enabled patterns of errors or discrepancies in the JIMMY data to be identified.

E. Stakeholder (RFO) Analyst Review

24. One of the biggest challenges faced during the previous iteration of JIMMY testing was records with large amounts of edits that fell into the investigate differences category. One strategy used to help address this challenge was shifting from a quantitative to a qualitative approach. RFO analysts who had experience editing the quarterly agricultural surveys were recruited to help test JIMMY. Analysts were provided a few records weekly and were asked a series of questions. These questions included the following:

Individual Record Review Questions
1. What is JIMMY doing well for this record?
2. Are the JIMMY values reasonable?
3. What is JIMMY not doing well with for this record?
4. For the items JIMMY is not doing well, do you have any suggestions on how JIMMY should be editing this differently?

Information-rich responses yielded a better understanding of not only the record but also how JIMMY should have handled these edits.

25. With the addition of these new strategies and incorporating best practices from previous iterations of testing JIMMY on historical data, the testing strategy was completed. Once completed the testing strategy was applied to September 2023 Crops APS data. Results of this testing are critical to implementing the new automatic editing system in the production environment for September Crops APS 2024.

IV. Results

A. Assessing Quality of Data Processed by JIMMY

26. JIMMY testing was completed on the whole US 2023 September Crops APS sample, which totaled over 60,000 records. For this study, to measure the impact of JIMMY on efficiency and data quality, a subsample of 11,085 records was drawn from 2023 September Crops APS. Only the corn belt states of Illinois, Iowa, Minnesota, Missouri and Wisconsin were included. Records within the subsample have a strong representation of size of operation, various crops, and completeness of report.

27. Several evaluation criteria were developed for measuring the impact of implementing JIMMY in the production environment. First, the differences in the JIMMY and NASS final edited data are compared to the differences in the reported survey and the NASS final edited data at the record level. For these results to be considered reasonable, large differences at the respondent level must be minimized. Differences do not necessarily mean JIMMY edits are incorrect, but many large differences may indicate an issue with data quality.

28. Results of this phase of testing are based on a single variable from the 2023 September Crops APS– specifically the acres of total land operated. Total land operated for a producer is calculated using the following variables: land owned (A_o), land rented from others (A_{rf}), and land rented to others (A_{rt}). Relationships that exist among these variables can be expressed by the following formula for total land operated by the producer (A_t).

$$A_t \geq A_o + A_{rf} - A_{rt}$$

Note that the sum of land owned, and land rented from minus land rented to must be less than or equal to the total land operated. Therefore, if total land operated, land owned, and land rented to, were reported, a reasonable edit for land rented from can be obtained by subtracting total land operated by land owned and adding land rented to.

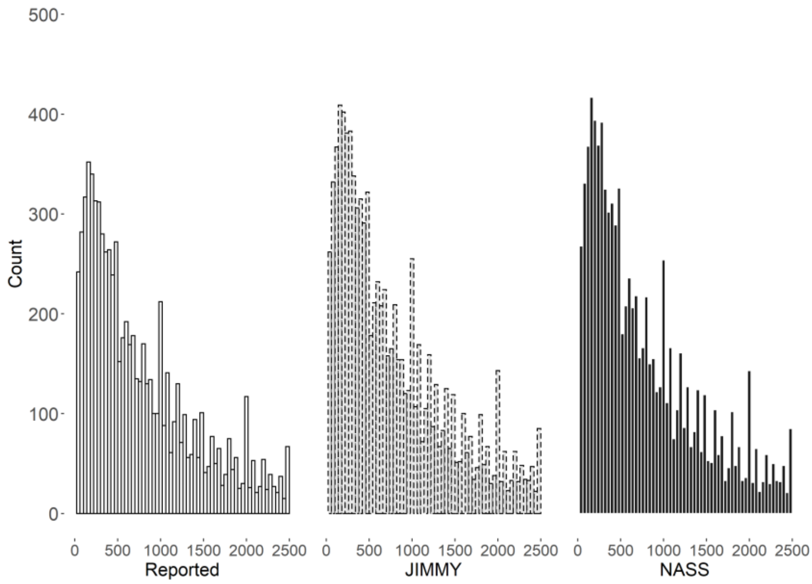

Figure 1.1: Distribution of total land operated by data source.

Number of observations are represented on the y axis and number of acres operated is on the x axis.

29. Visually, the distribution of the total land operated based on the JIMMY data is more similar to the distribution for the NASS final data than that for the reported data (see Figure 1.1). It is important to note that the JIMMY data were produced by applying the automated edits and imputation to the reported data, and the NASS final edited data are the values produced from the reported data in September 2023 after Blaise manual interactive editing. The distributions do not follow a normal distribution. The median number of acres operated was calculated for each distribution as a measure of central tendency, or the middle, of the distribution.

Table 1: Average Acers Operated			
	N	Median	ΔNASS
Reported	9,258	700	38.5
JIMMY	10,930	736	2.5
NASS	10,968	738.5	--

30. Comparison of medians reveals that the difference between JIMMY (Md = 736) and NASS (Md = 738.5) is 2.5 acres whereas the difference between reported (Md = 700) and NASS (Md = 738.5) is 38.5 acres. These median differences are represented by ΔNASS in the table above. This implies that the measure of central tendency for JIMMY values is closer to the NASS true values than the reported values alone.

31. Once exploration of the distributions was completed, test observations were assigned to the three distinct categories that measure differences among the reported, JIMMY, and NASS final data for total land operated. Of the 11,085 subsample records, only 10,968 records had non-missing total land operated data. Of the non-missing data, 59 records fell into the Incorrect edits group, 509 records in the investigate differences and 10,400 records in the correct edit group (see Figure 1.2). That is, over 93% of the JIMMY values for total land operated fall into the correct edit category and less than 1% of the values fall into the incorrect edit category. Approximately 4% of the edits fell into the investigate differences category. These are the edits that are not as apparent to categorize as correct or incorrect and were previously difficult for the review process. To address this, RFO analysts joined the testing team to review some of these records.

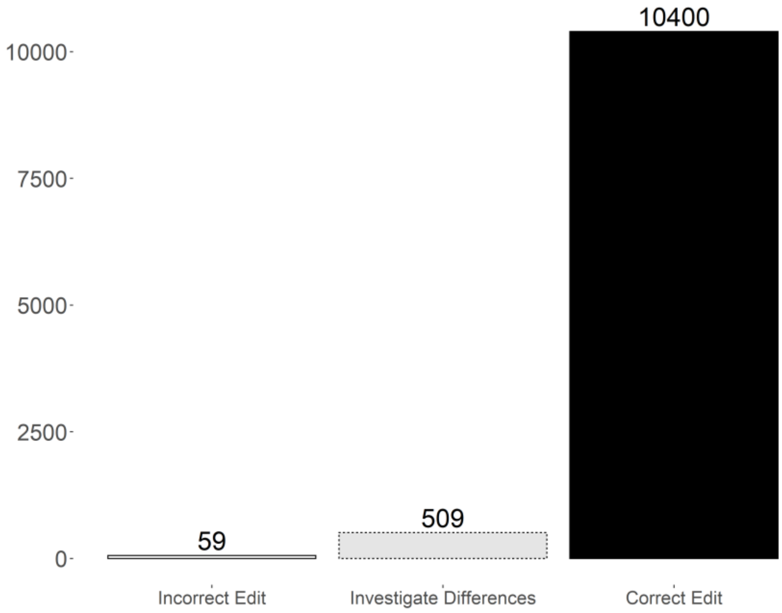

Figure 1.2: Number of observations by edit group.

B. Assessing Workload and Data Quality (from current Blaise & SPS Edit Systems)

33. For the second component of testing, reported data from September 2023 were read into the Blaise beta environment and SPS edit. Outcomes from this phase of testing provided a baseline of the number of records that do not meet the survey's edit logic. For records that do not meet the edit logic, a count was created to determine how many edits will be required per record. Once these baseline metrics were captured, the JIMMY data were read in to see how many unit-level edits and item-level changes were made. Below are the results for the 11,085 survey respondents in the subsample.

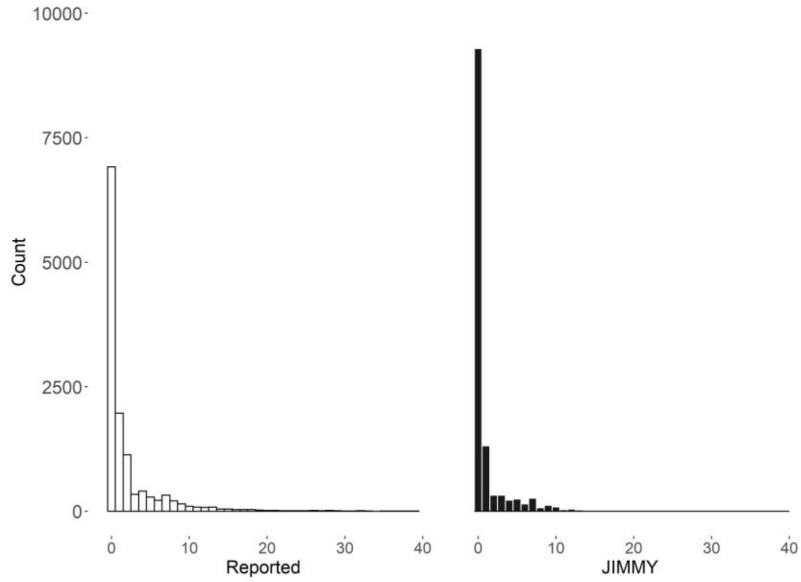

Figure 1.3: Distribution of the number of errors per record pre and post JIMMY automatic edits.
Number of observations are represented on the y axis and number of edits required per record is on the x axis.

The distributions of the number of errors per record for the reported data and for the JIMMY data are positively skewed (see Figure 1.3). When the number of errors is close to zero, the corresponding records require little to no manual editing. When comparing the reported distribution to the JIMMY distribution, a larger number of records are closer to zero due the automatic editing.

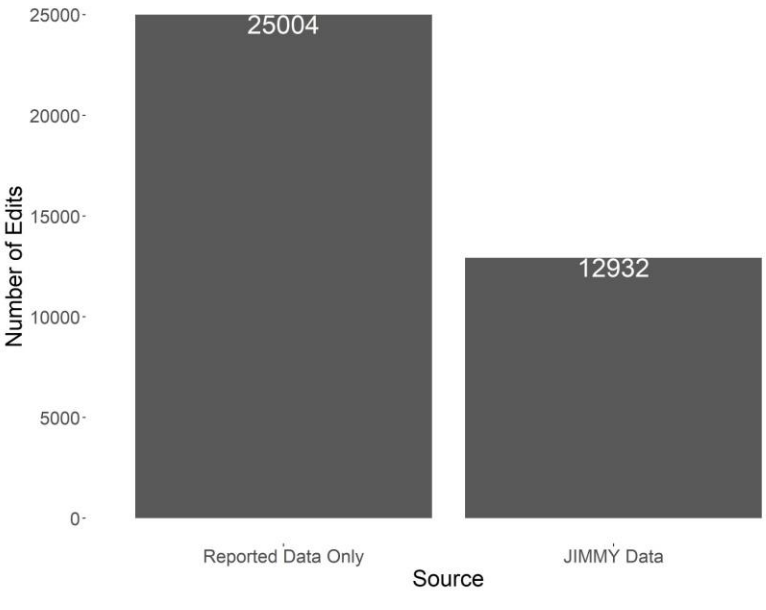

Figure 1.4: Number of edits required by source of data.

34. When exploring the reduction of edits required by an RFO analysts with and without JIMMY automatic editing the total number of edits were aggregated for the sample. In Figure 1.4 the total number of edits for the reported data only was 25,004. By dividing the total number of edits by the sample size of 11,085 an average of about two edits per record was produced. Once JIMMY automatic edits are applied to the reported data, the total number of edits is 12,932. Following a similar procedure to the reported data by dividing the total number of edits after JIMMY automatic edits by the sample size produced an average of about one edits per record. Reduction of over 50% in the number of edits due to JIMMY's automatic editing is promising to see and implies less time would be needed to edit these records.

C. Stakeholder (RFO) Analyst Review

35. RFO analysts and other content matter experts were recruited to assist with testing JIMMY for complicated records and edits. RFO analysts completed reviews of 76 records over a nine-week period. The number of records reviewed per week are presented in table below.

Table 2: RFO Weekly Review

Week	Number of Records
1	15
2	9
3	13
4	9
5	18
6	10
7	6
8	10
9	9

Feedback produced by the four questions produced deeper insight into the complex records. For many of the records, the analysts found the JIMMY data to be reasonable. For edits analysts thought were unreasonable, extensive insight was gained through their discussion of the edit they thought needed to be made. Some examples and justification for edits different than JIMMY included using additional data sources that JIMMY does not have access. Information acquired from these additional data sources can provide insight into why the record is challenging for JIMMY. Also, the analysts supplied the steps and reasoning needed to resolve all required edits on the reviewed records. Perspective provided from their feedback helped refine editing rules and proposed new ideas and vison for future iterations of JIMMY.

V. Discussion

36. NASS has developed an automatic editing and imputation prototype to reduce the number of edits required during a given survey cycle. Goals of this new system include modernizing and standardizing NASS's current editing and imputation methodology and reducing RFO analysts time editing. To incorporate this new system into the production environment, an assessment between automated vs manual editing and imputation on operational efficiency and data quality was completed. Assessment of this system was conducted on historical data from the September 2023 quarterly agricultural survey.

37. Results of testing on historical data are promising and have shown that incorporation of additional testing techniques, including stratified sample, Blaise and SPS Edit System, and RFO qualitative feedback, have been advantageous. Results included, over ninety percent of total land operated edits were correct edits and implies JIMMY is performing properly for many records. Over 13,000 edits were resolved automatically that previously needed to be edited manually. In addition, the average number of edits per record has been reduced, indicating that less time was allotted to editing records. Finally, numerous records and edit relationships that the JIMMY systems struggled with in previous iterations of testing were either resolved or improved based on experts' feedback on how to handle said record.

38. Challenges faced during the assessment of the new automatic system include the number of edits to evaluate, relationships between edits, and defining edits to be reasonable. Reviewing of records in more detail is time consuming especially for records with several values. Variables within a given questionnaire have numerous relationships, therefore if an edit is made to one it can directly affect another variable. Finally defining an edit to be reasonable can be hard to quantify; large differences do not necessarily mean the value is incorrect. Research is continuing to address these challenges to ensure measurements of success for this new automatic editing and imputation system are properly documented.

39. Findings determined by these results have been essential in helping NASS leadership determine that JIMMY will be used in the production environment for the quarterly agricultural survey in September 2024. Preparation for implementation and additional testing are currently being conducted. Preliminary results of using JIMMY in the production environment for September 2024 Crops APS will be presented at UNECE Expert Meeting on Statistical Data Editing, 7-9 October 2024, in Vienna Austria.

A. Acknowledgments

40. The findings and conclusions in this paper are those of the authors and should not be construed to represent any official USDA, or US Government determination or policy. The authors also want to acknowledge Summit Consulting, Vikas Agnihotri, Karl Brown, Monic Denwiddie, Peter Kilpatrick, Vito Wagner, and the NASS JIMMY Testing Team for making major contributions to the IDEAL project.

VI. References

Lipke, M., Miller, D., Wagner, V., Brown, K. and Agnihotri, V. (2022). “Growing a Modern Edit and Imputation System”. United Nations Statistical Commission and Economic Commission for Europe, Conference for European Statisticians, Work Session on Statistical Data Editing. Virtual, 3-5 October 2022

Laird, M., Maiwurm, J., Lipke, M., Miller, D., Denwiddie, M., Wagner, V., Brown, K. and Agnihotri, V. (2023) “Growing and Testing a Modern Edit and Imputation System". International Conference on Agricultural Statistics, Washington DC, United States, 17-19 May 2023

Miller, D. (2021). “Growing a Modern Editing and Imputation System”. 2021 Federal Committee on Statistical Methodology Conference.

Miller, D. and Young, Linda (2015). “Imputation at the National Agricultural Statistics Service”. United Nations Statistical Commission and Economic Commission for Europe, Conference for European Statisticians, Work Session on Statistical Data Editing. Budapest, Hungary, 14-16, September 2015.

Dau, A. and Miller, D. (2018). “Dancing with the Software”. 2018 Joint Statistical Meetings. Vancouver, BC, Canada, 28 July – 2 August, 2018.

E.de Jonge and M. van der Loo, "errorlocate: Locate Errors with Validation Rules," 2018.

E. de Jonge and M. van der Loo, "validatetools: Checking and Simplifying Validation Rule Sets," 2019.

Miller, D. (2021). “Growing a Modern Editing and Imputation System”. 2021 Federal Committee on Statistical Methodology Conference.

M. van der Loo and E. de Jonge, "dcmodify: Modify Data Using Externally Defined Modification Rules," 2018.

M. van der Loo and E. de Jonge, "deductive: Data Correction and Imputation Using Deductive Methods," 2017.

M. van der Loo, "simputation: Simple Imputation," 2017.

M. van der Loo and E. de Jonge, "validate: Data Validation Infrastructure," 2018.

Fellegi, I. P. and D. Holt (1976). A Systematic Approach to Automatic Edit and Imputation. Journal of the American Statistical Association 71, pp. 17–35.

Manning, A. and Atkinson, D. (2009). “Toward a Comprehensive Editing and Imputation Structure for NASS – Integrating the Parts”. USDA NASS RDD. United Nations Statistical