



# **Improving data quality in the Survey of Consumer Finances: What do we learn from interviewer-respondent interactions?**

Jesse Bricker, Kevin Moore, and Richard Windle  
Federal Reserve Board

The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors of the Federal Reserve System.

# Conducting a Survey

- Ideal:
  - Survey questions written so that all parties perfectly understand what is being asked
  - No misclassification issues occur
- Reality:
  - There is a distribution of understanding of questions
    - Any question wording will cause confusion for some set of Respondents
  - The Interviewer must facilitate the interview, while still maintaining consistency
  - Data most vulnerable and fixable during the interview itself

# Collecting Interviewer and Respondent Comments

- Reality in the Survey of Consumer Finances (SCF):
  - Respondents are asked to provide complex data
  - Mistakes can result, due to:
    - Misunderstanding of terminology used in the SCF
    - Incomplete understanding of concepts or circumstances
- Solution:
  - Allow a space for both the Respondent and Interviewer to explain and clarify the provided data in their own words
- How useful is this space?

# Investigating the Value of Comment Collection

- Collecting and analyzing Interviewer and Respondent comments is costly
  - Requires a significant staffing investment
  - Can delay data releases
- Therefore, it's important to ask if such an investment is worth the cost
  - How often do we repair data due to these comments?
  - How useful are these comments in assisting with data repair?

# Analyzing the Frequency of Comments and Data Edits

- 10 percent of data are repaired
- Interviewer comments drive 1/3 to 3/4 of these repairs
- Comments less helpful in some parts of the survey, which may indicate that:
  - Interviewers require more training
  - Question needs to be re-written

# Survey of Consumer Finances (SCF)

- Triennial XS 1983-2013
  - Sponsored by the FRB, conducted by NORC (since 1992)
- Household finances
  - Assets (houses, vehicles, checking, stocks, IRAs, 401(k)s...)
  - Debts (mortgages, consumer loans...)
- Sensitive topics
  - Account balances
  - Income
  - Birthdates
- Confusing topics
  - Mutual funds directly-held or held within my IRA
  - Determining type of pension plan
  - Financial events from decades ago

# Data Collection

- Interview is conducted without any immediate outside assistance
- Interviewer required to do more than simply read words, must *facilitate* the interview
  - Resources
    - Extensive pre-field training (led by NORC)
      - Classes
      - One-on-one training
    - Glossary (terms by FRB)
    - Weekly memos during the field (by NORC)

# Differences Between Data Collection Methods

- Old (PAPI) days:
  - Interviewer saw the data and corrected it
  - Left marginal notes to explain complicated data
  - FRB reviewed notes during the editing process
- CAPI:
  - Data cannot be seen during the interview
  - Interviewer can leave notes during and after the interview
    - Pushing F2 during the interview brings up a comment box
    - Debriefing comments are required at the end of every interview



# Motivation for Edits to Data

- Comments left by interviewer *during* the interview
  - “Checking account is really a savings account”
- Comments left by interviewer *after* the interview in the debriefing
  - “My earlier statement meant R has no checking account”
- Inconsistent Questionnaire data
- Warning flags
  - “R is 25 but lives with parent – check ownership”
- Record of calls, used to verify correct respondent

# J Variables

- A set of shadow variables
  - = 0 if question was originally answered
  - = 1 if originally skipped
  - = **5** if **repaired based on comment** (minimal judgment)
  - = **13** if **repaired based on comment** (using some judgment)  
***or*** based on other data structure
  - = 14 if skipped based on previous edit
  - = **2098** if **no useful edit** and unable to repair – **set to impute**

# Editing Example

- Original data

CaseID	X804	J804	X805	J805
0001	750,000	0	600,000	0
0002	250,000	0	200,000	0
0003	400,000	0	50,000	0

- X804: How much was originally borrowed?
- X805: How much do you currently owe?

# Editing Example

- Example: CaseID=0002
  - In the original data, we see that the Respondent has borrowed \$250,000 for the purchase of their home (mortgage)
  - In the case notes, though, the interviewer tells us that the Respondent was confused, and actually borrowed \$200,000

# Editing Example

- Final data

CaseID	X804	J804	X805	J805
0001	750,000	0	600,000	0
0002	200,000	5	200,000	0
0003	400,000	0	50,000	0

## Editing Example 2

- Example: CaseID=0003
  - In original data, the respondent has borrowed \$400,000
  - In case notes, though, the interviewer tells us that the respondent was confused.

The interviewer could not determine what the amount borrowed was
  - Impute the amount borrowed

# Editing Example

- Original data

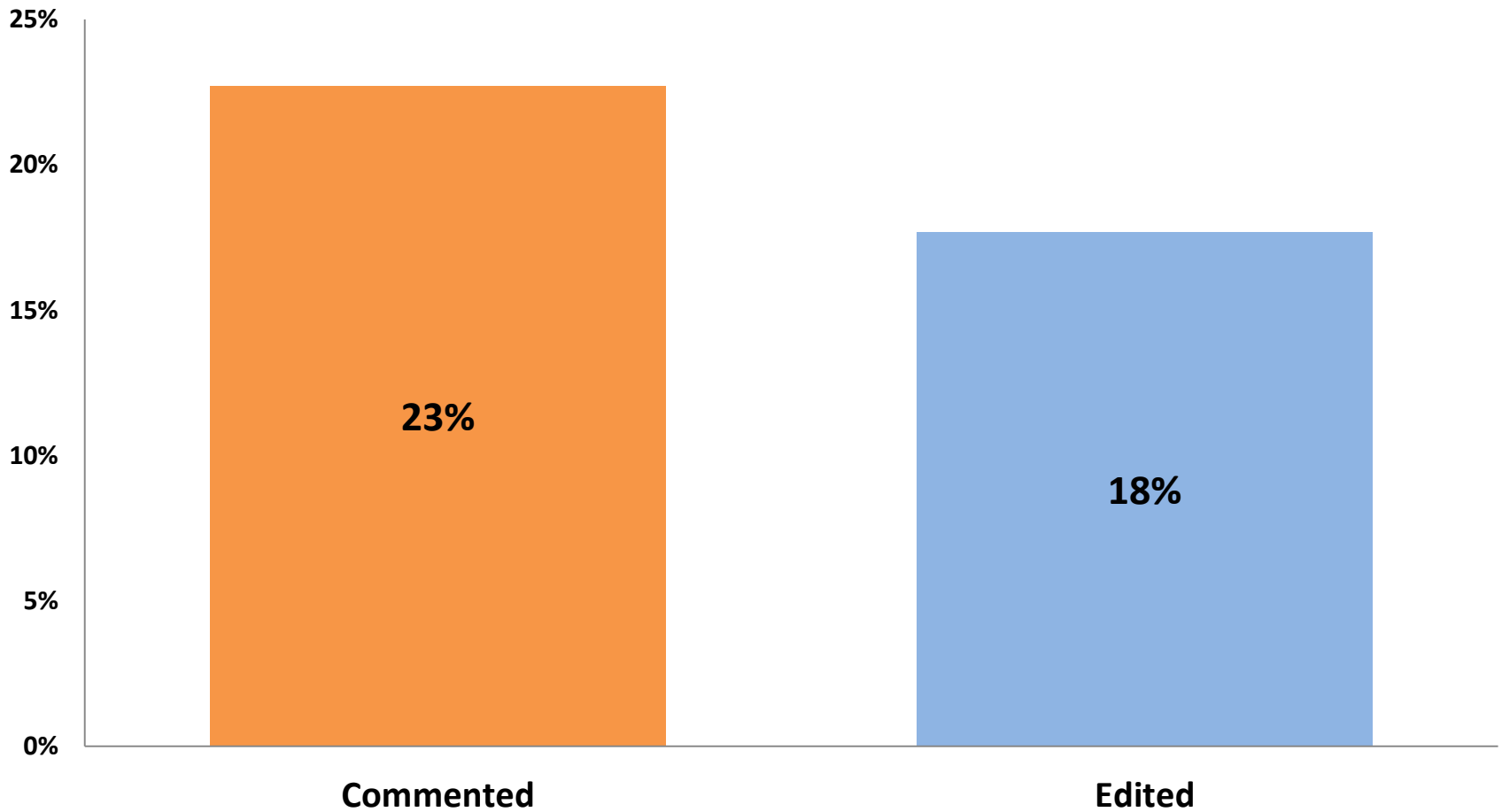
CaseID	X804	J804	X805	J805
0001	750,000	0	600,000	0
0002	200,000	5	200,000	0
0003	.N	2098	50,000	0

# Comment Usage During the Editing Process

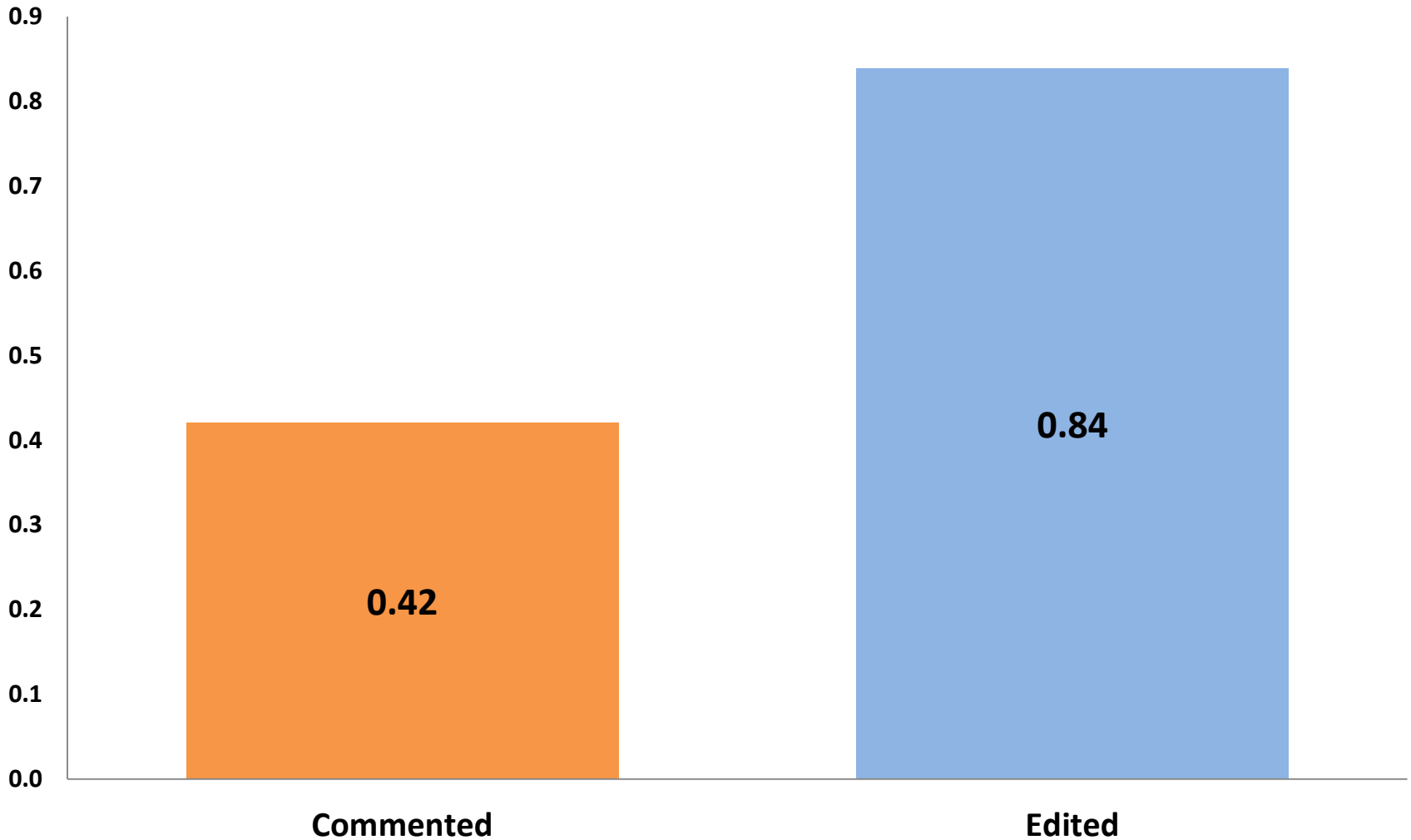
- Overall, about 10 percent of the data are edited
- All cases have comments
- Comments improve the quality of the edits
  - Mortgages
    - Comments influence 3/4 of edits, and directly influence 1/2 of edits
  - Pensions
    - Comments influence 3/4 of edits, and directly influence 1/3 of edits
  - Income
    - Comments influence almost all edits, and directly influence 1/3 of edits



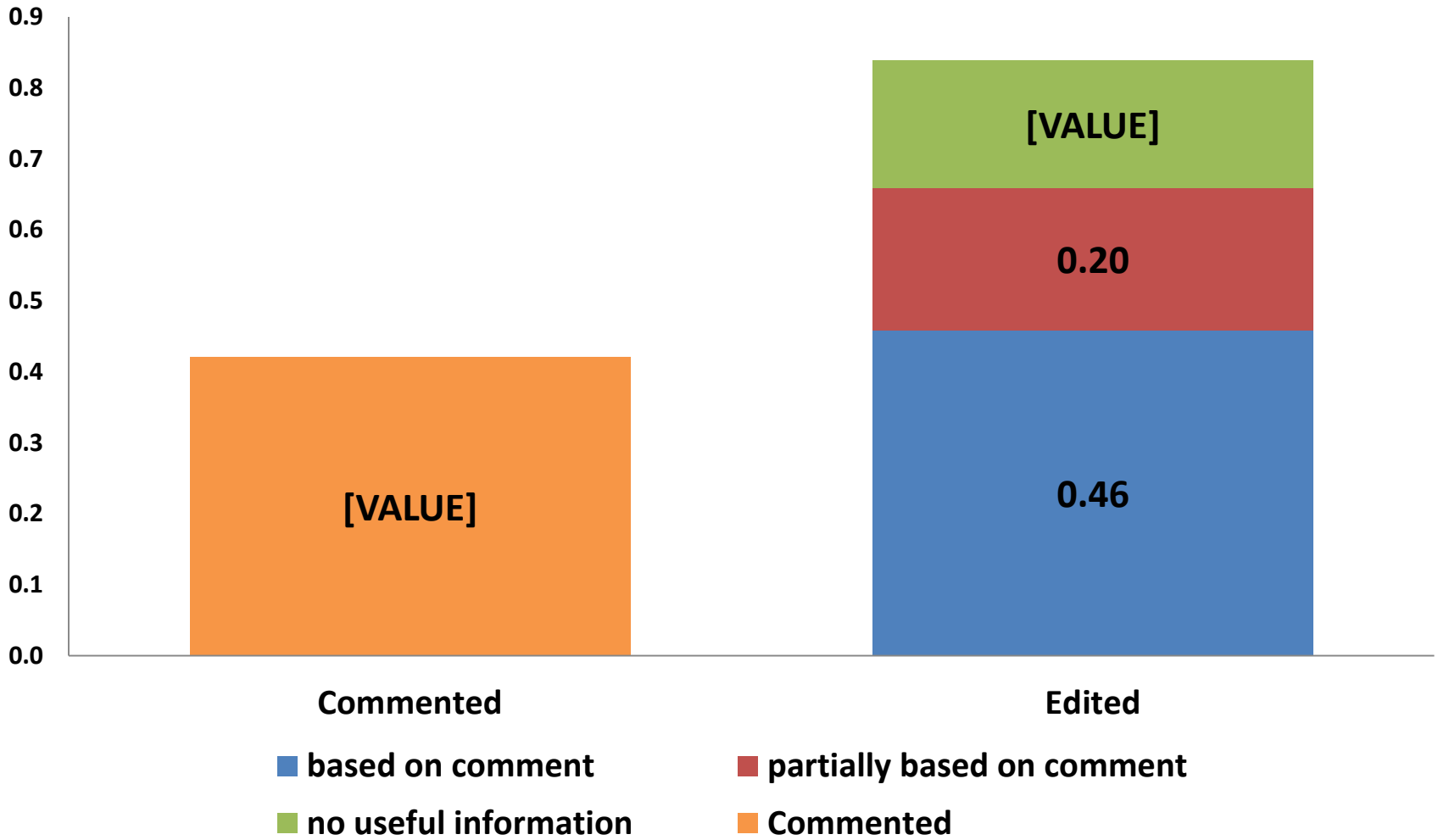
# Mortgages: Percent of Cases with Comments, Edits



# Mortgages: Mean number of Comments, Edits per Case



# Mortgages: Mean number of Comments, Edits per Case



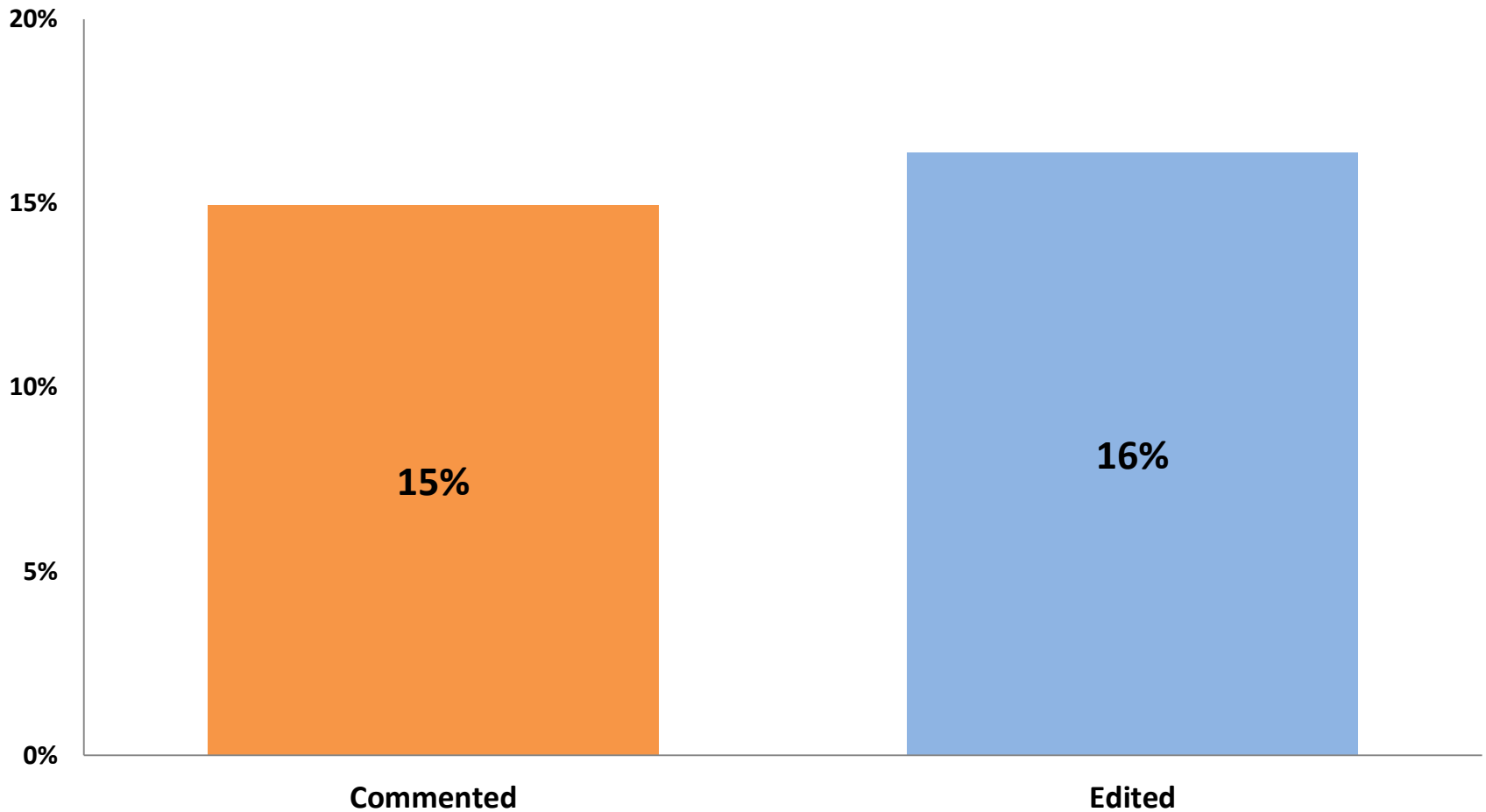
# Mortgages: Most Edited Variables

**Table. Mortgages - most common edits and reason for edits**

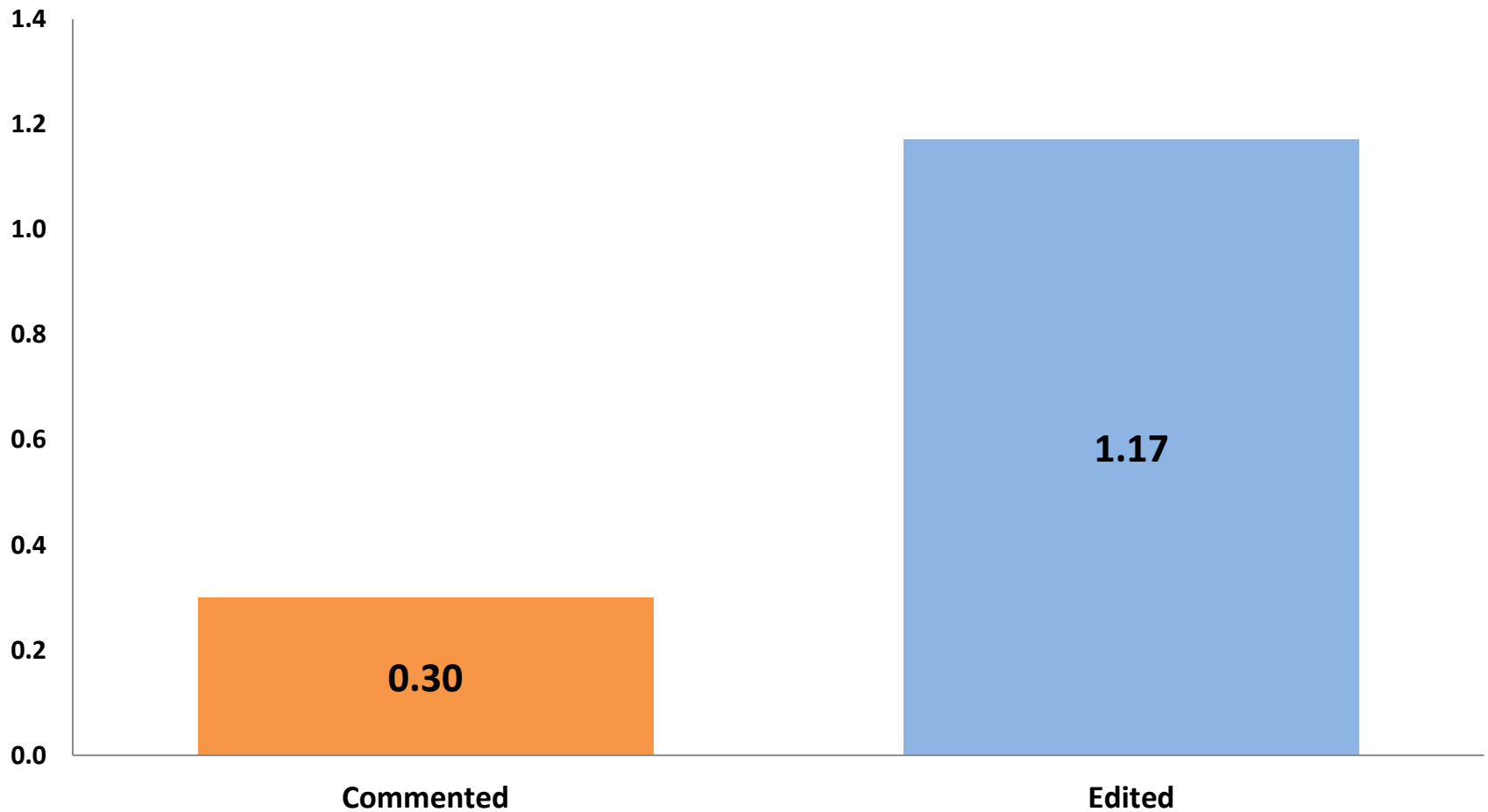
	Mortgage amount	Mortgage balance
Percent edited	6	5
<hr/>		
<u>Percent by edit reason</u>		
Comment	2	1
Partial comment	3	2
No useful information	1	1
Skip based on earlier	1	1

*Note:* Reason edited columns sum to total percent edited, though imperfectly due to rounding. Percent edited based on ratio of edits to original non-inaps.

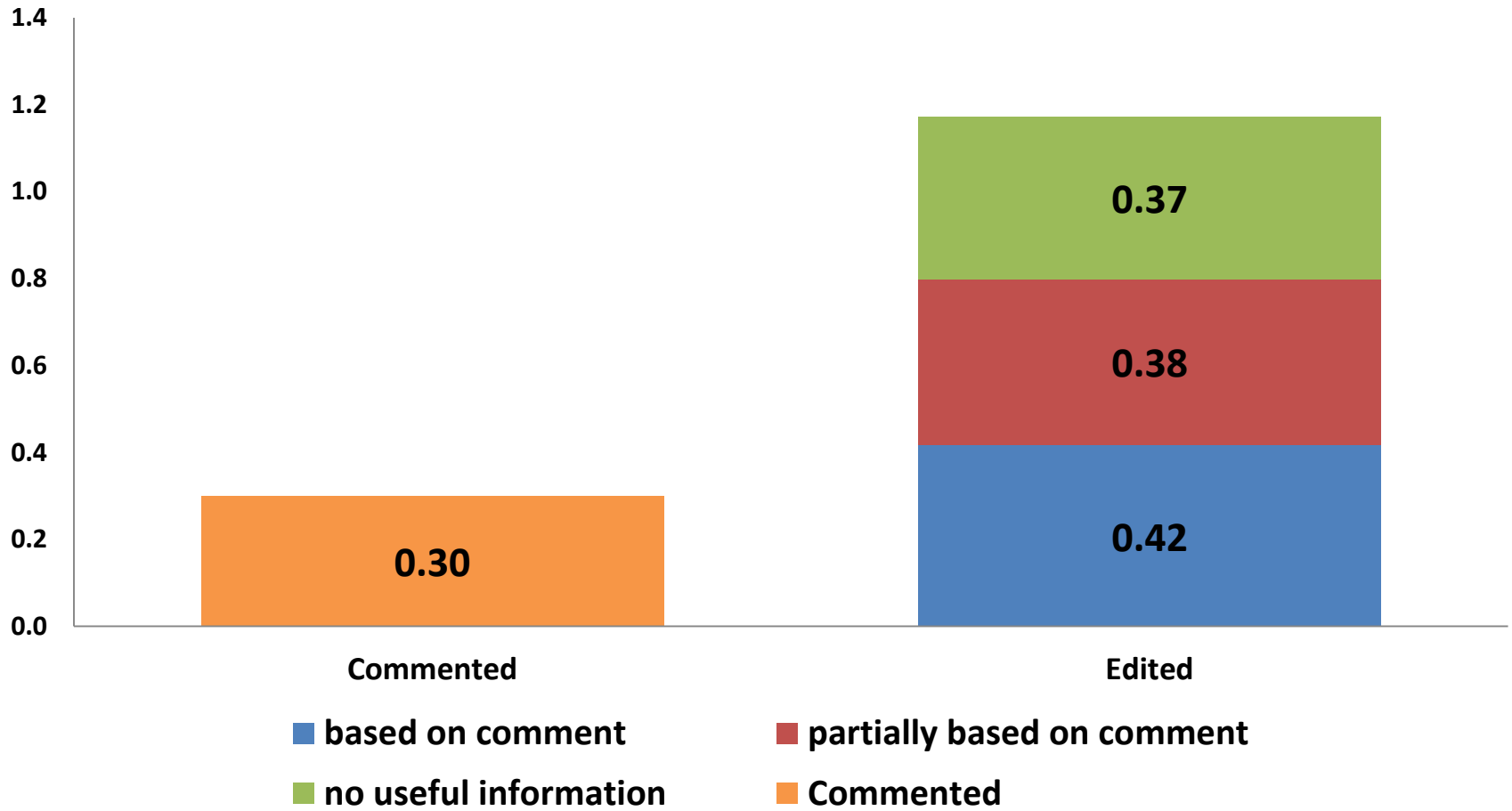
# Pensions: Percent of Cases with Comments, Edits



# Pensions: Mean Number of Comments, Edits per Case



# Pensions: Mean Number of Comments, Edits per Case



# Pensions: Most Edited Variables

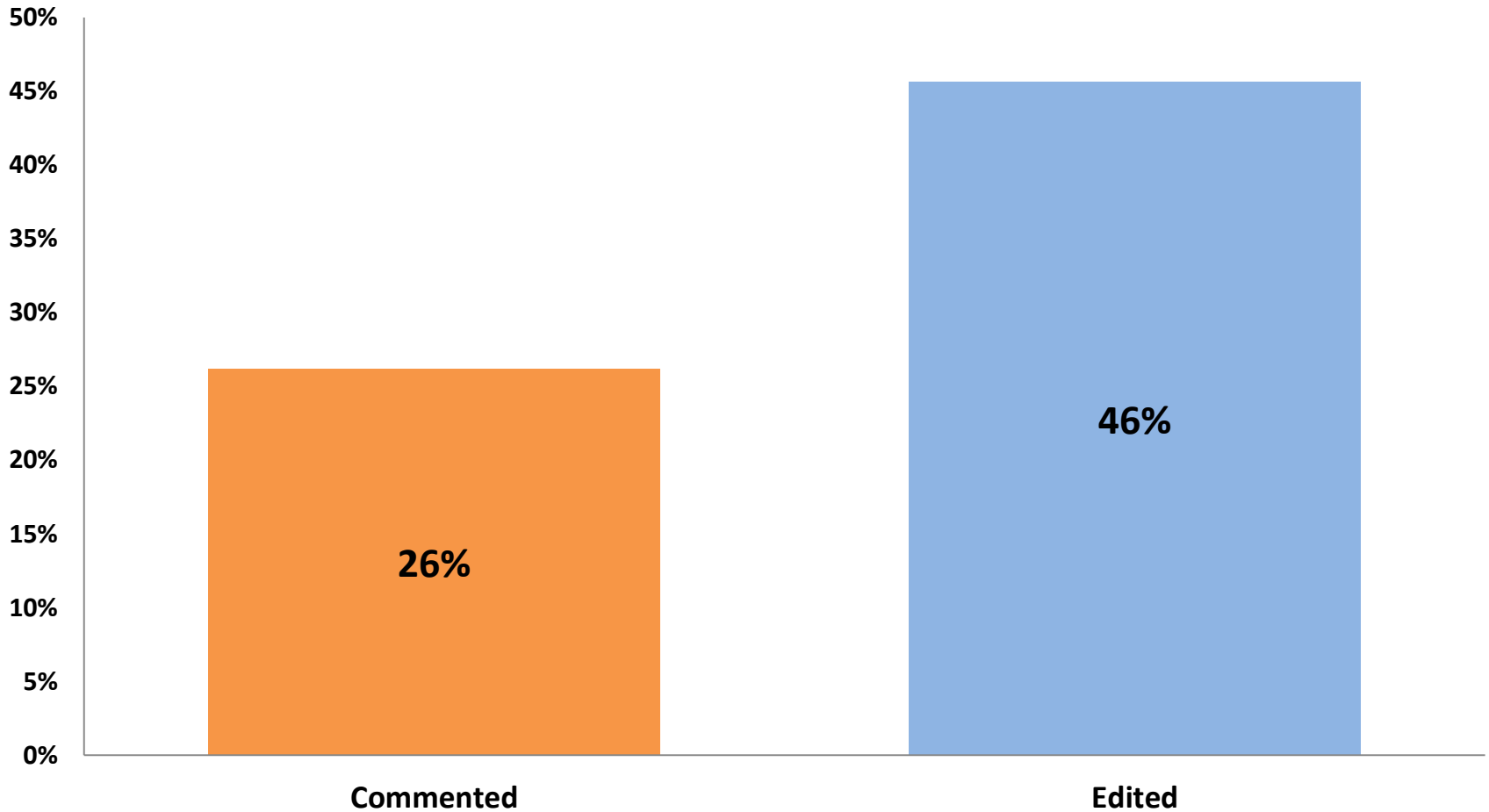
**Table. Pensions - most common edits and reason for edits**

	Account-type (Y/N)	Account balance (\$)
Percent edited	11	15
<u>Percent by edit reason</u>		
Comment	5	5
Partial comment	5	4
No useful information	0	2
Skip based on earlier	0	5

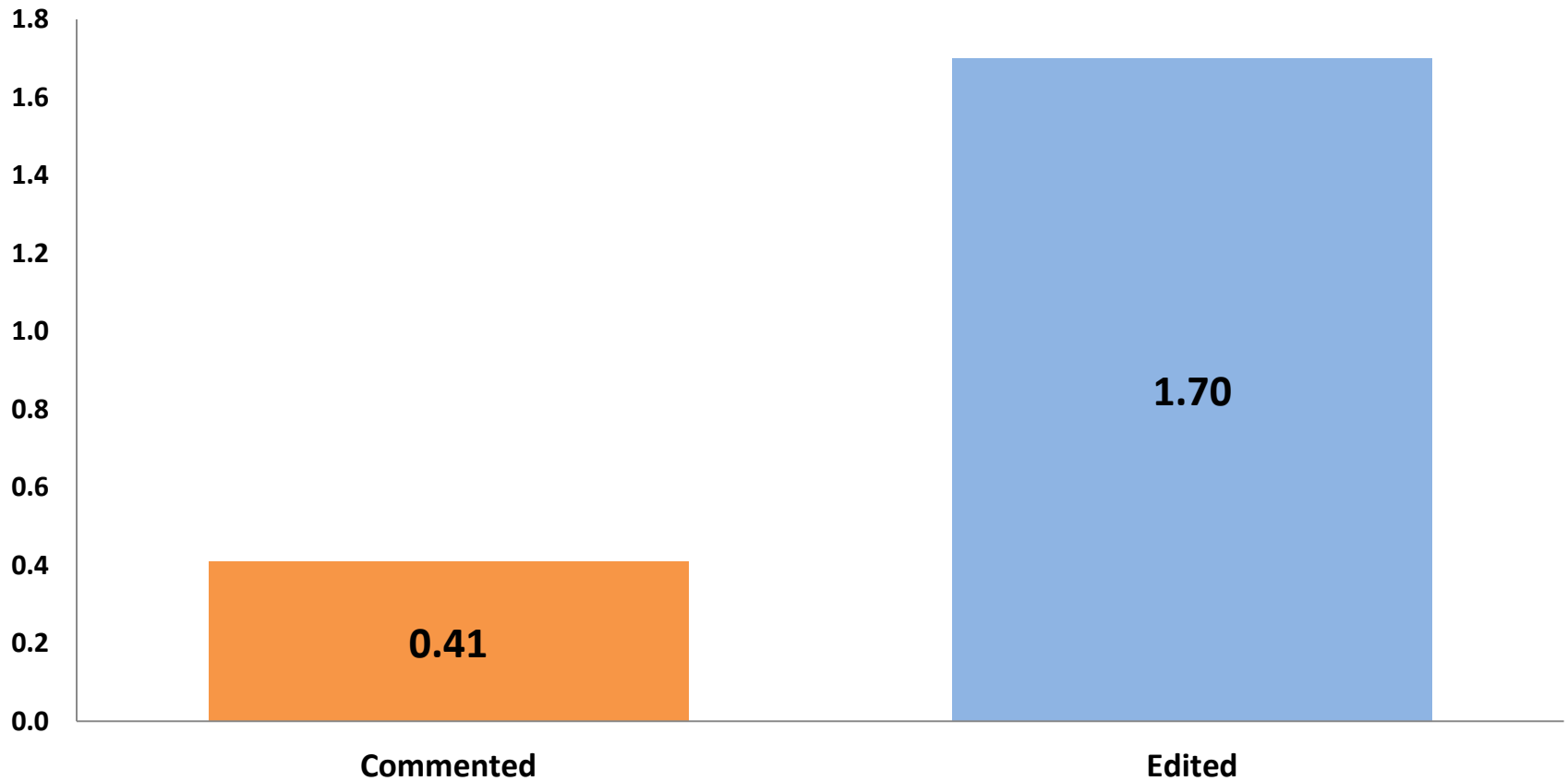
*Note:* Reason edited columns sum to total percent edited, though imperfectly due to rounding. Pensions refer to the respondents current job pension.



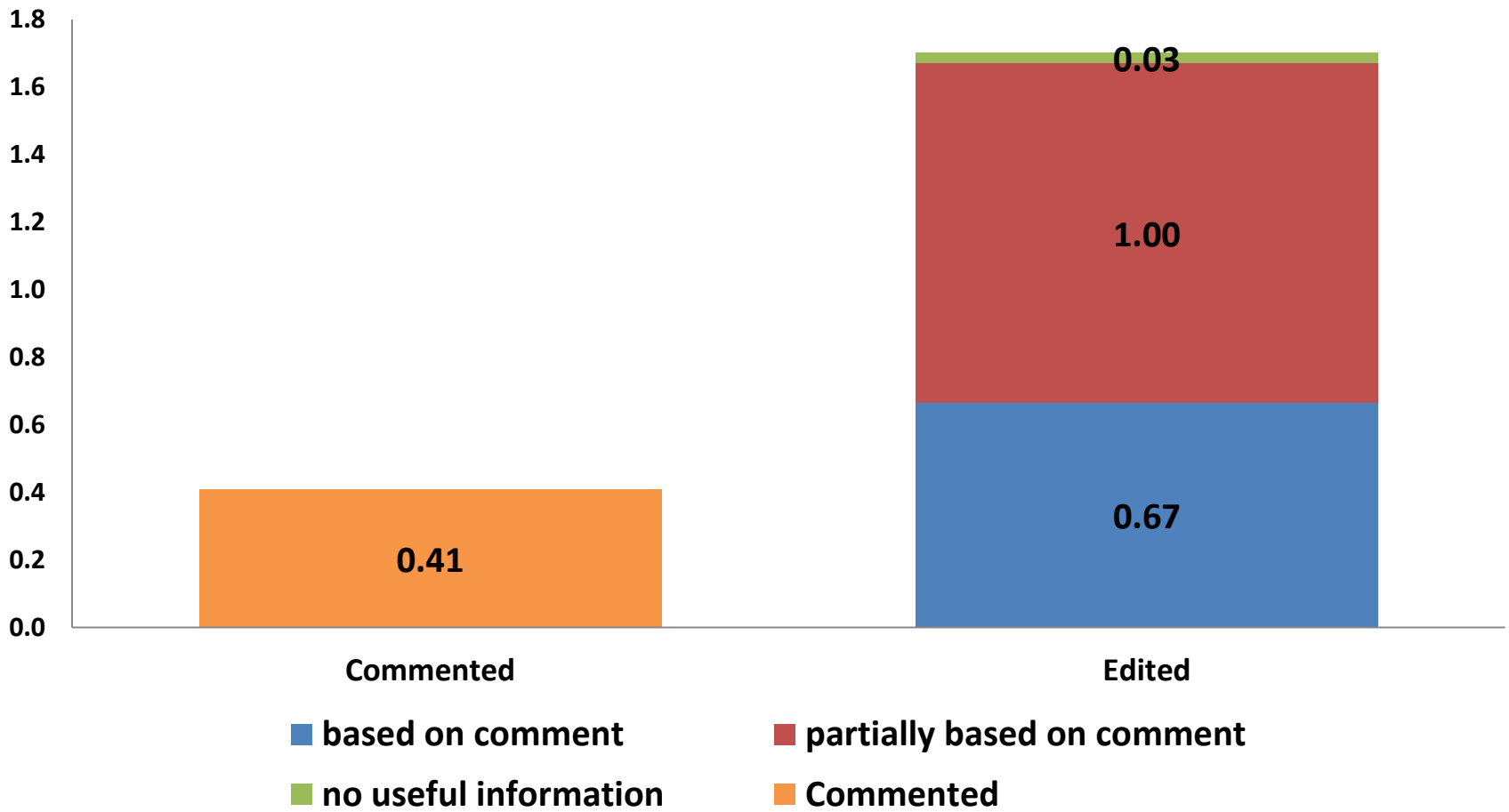
# Income: Percent of cases with Comments, Edits



# Income: Mean Number of Comments, Edits per Case



# Income: Mean Number of Comments, Edits per Case



# Income: Most Edited Variables

**Table. Income - most common edits and reason for edits**

	Have wages (Y/N)	Amount of wages (\$)
Percent edited	6	22
<hr/>		
<u>Percent by edit reason</u>		
Comment	3	6
Partial comment	3	7
No useful information	0	0
Skip based on earlier	0	4
Reconciliation screen	NA	4

*Note:* Reason edited columns sum to total percent edited, though imperfectly due to rounding. Reconciliation screen refers to edit made by R and interviewer during the interview.

# Conclusions

- Providing a space for respondent and interviewer to communicate to the data reviewer significantly improves data quality
  - 10 percent of collected data required editing
  - Most of these edits were motivated by the existence of a comment
- Survey of interviewers
  - 70 percent say pensions are hard, 20 percent say income is hard
  - Difficult because respondents are unfamiliar with pension details or do not do their own taxes

# Thank you

Kevin Moore

kevin.b.moore@frb.gov