

Overcoming Paradata Mining Challenges: Survey of Income and Program Participation (SIPP) Audit Trails

Patrick C. Campanello
U.S. Census Bureau

Presented at
FedCASIC
April 2017

This work is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views or opinions expressed in the paper are the authors' own and do not necessarily reflect the views or opinions of the U.S. Census Bureau.

Outline

- Reengineered SIPP Instrument
- Paradata Resources
 - Audit Trail File
- Ongoing Paradata Efforts
- Challenges and Issues Encountered

The Survey of Income and Program Participation (SIPP) Mission

- The survey's mission is to provide a nationally representative sample for evaluating:
 - Annual and monthly income dynamics
 - Movements into and out of government transfer programs
 - Family and social context of individuals and households
 - Interactions among these items

2014 SIPP Reengineering Goals

- Reduce respondent burden
- Reduce costs
- Increase accuracy and timeliness of data
- Increase accessibility of data
- Increase use of administrative records
- Move from DOS to Windows-based system

Some SIPP Evaluation Resources

- Audit trail files
- Field monitoring data
- Certification test instrument
- Debriefing instrument
- Contact History Instrument (CHI)

Audit Trails

- Timestamped logs that depict a Field Representative's (FR) interaction with the survey instrument
- Some elements logged:
 - Keystrokes
 - Mouse clicks
 - Function key usage
 - Answer data storage
 - Checks or validations encountered

Audit Trails - Example File

****LINES REMOVED****

```
00168|6/5/2014 1:23:06 PM Enter Field:BDemographics.BAge[1].DOB_BYEAR  
|Status:Normal|Value:
```

```
00169|6/5/2014 1:23:06 PM (KEY:)1832[ENTR]
```

```
00170|6/5/2014 1:23:07 PM Action:Store Field Data Field:BDemographics.BAge[1].DOB_BYEAR
```

```
00171|6/5/2014 1:23:08 PM Leave Field:BDemographics.BAge[1].DOB_BYEAR Cause:Next  
Field|Status:Normal|Value:1832
```

```
00172|6/5/2014 1:23:11 PM (KEY:)s
```

```
00173|6/5/2014 1:23:11 PM Errordlg action:Suppress|Text:@FThat would make you 182 years  
old. Is that correct?@F @/ @/@Zs@Z @LIf this is correct, supress and  
continue.@L @/@Zs@Z @LIf this is not correct, go back to DOB_BMONTH, DOB_BDAY,  
or DOB_BYEAR and correct.@L  
|Involved:BDemographics.BAge[1].DOB_BDAY;31;BDemographics.BAge[1].DOB_BMONTH;De  
c;BDemographics.BAge[1].DOB_BYEAR;1832
```

```
00174|6/5/2014 1:23:11 PM Action:Error Suppress|Field:BDemographics.BAge[1].DOB_BYEAR
```

```
00175|6/5/2014 1:23:11 PM|Enter
```

```
Field:BDemographics.BHISPANIC_ORIGIN[1].ORIGIN|Status:Normal|Value:
```

****LINES REMOVED****

**Example above contains fictitious data*

Audit Trails –Uses

- Gathering paradata
- Debugging
- Retrieving missing answer data
- Helping detect falsification

Audit Trails - Tool Architecture

- A collection of modules written in Python
- Code designed to iteratively process audit trail files through a series of independent code files
- Interview Time
- Function Key and Don't Know/Refusals
- Calculates data at Household, Person, and Section levels
- Missing Answer Extraction

Problems Encountered

- SIPP instrument complexity
- Original audit trail file format
- Non-displayed events not logged
- Duplication of event logged
- Human (FR) and computer impact

Tool Results

Function Key and Definition		Total Count	Completed Interviews
[F1]	Open help screen	12,329	11,641
[F7]	Display item level notes	28,960	27,237
[SHFT][F2]	Display FAQs	64	49
[SHFT][F4]	Display holiday list	88	76
[CTRL-D]	Don't Know answer	379,855	363,568
[CTRL-R]	Refusal answer	130,821	125,246

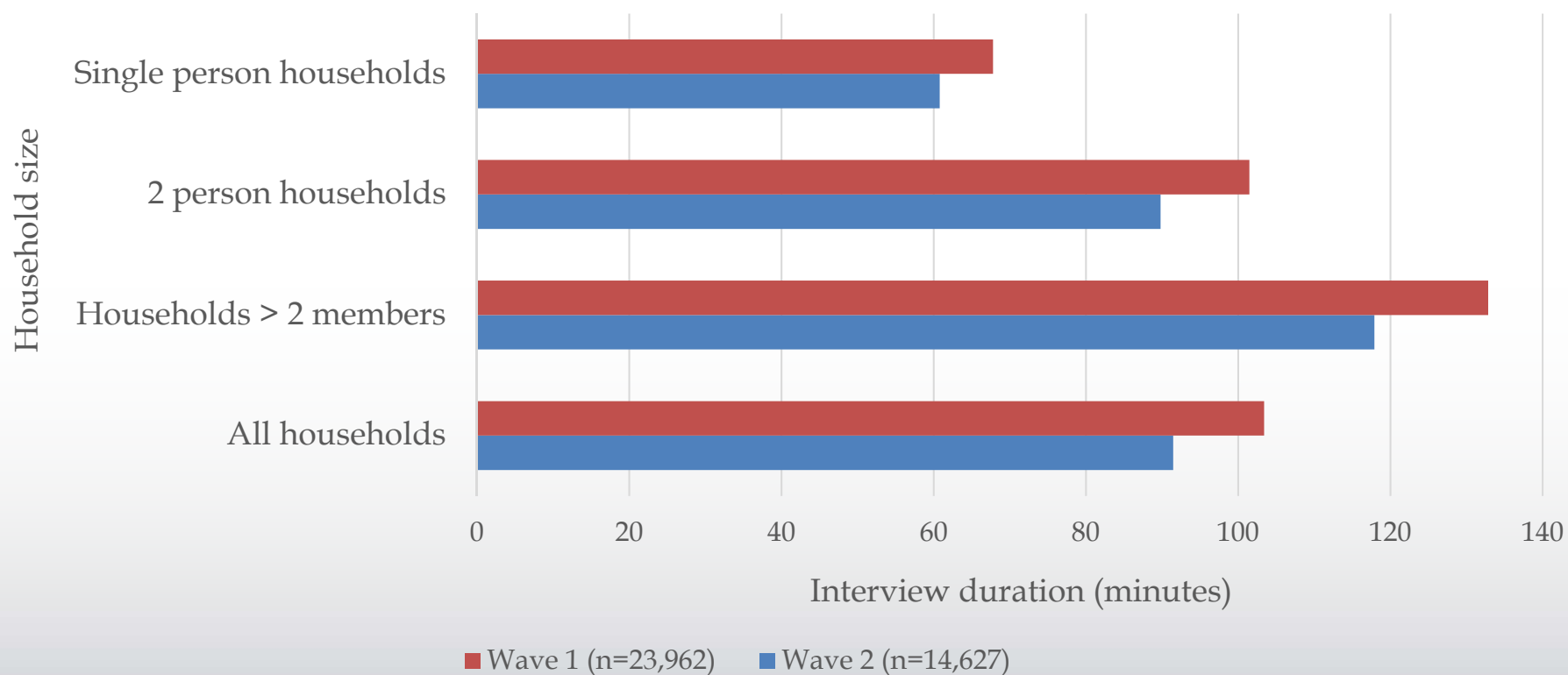
Tool Results

- How many cases did we retrieve missing data for?

	Household Level Dataset	Person Level Dataset
Answer values updated	1,976	17,446
Variables updated over total possible variables	80 / 250	71 / 109
Records with answers updated over Records with data in a missing answer field	893 / 29,419	6,251 / 30,434

Tool Results

Average Interview Duration by Household Size for Complete, In-person Interviews for 2014 SIPP Panel



**Data includes complete and sufficient-partial interviews, excluding outlier households that exceeded a threshold based on household size and total interview time.*

Conclusion

- Despite the hurdles, audit trail files are valuable resource
- Steadily making progress toward reengineering goals
- Help other survey teams realize similar goals
- Code written to be applicable outside SIPP

