Using Audit Trail Files for Missing Data Extraction in Wave 2 of the 2014 Survey of Income and Program Participation (SIPP)

Patrick C. Campanello and Shelley K. Irving

U.S. Census Bureau

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

SIPP's 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

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Outline

- Reengineered SIPP Instrument
- Addition of Event History Calendar (EHC) Tool
- Wave 1 Missing Answer Data Retrieval
- Wave 2 and Wave 1 Differences
- EHC Processing Logic Solutions
- Wave 1 and Wave 2 Results
- Future Application of Lessons Learned

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

2014 SIPP Survey 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

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Event History Calendar (EHC) Tool

*Note: This graphic contains fake data.

United States **Census** Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Event History Calendar (EHC) Tool

* Note: This graphic contains fake data.

Missing Answer Data Retrieval

CaseID (Record Number)	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Variable 6
00001231	9	2	Missing Answer	3	1	1
00001232	Don't Know	Refusal	6	4	Missing Answer	2
00001233	6	1	Don't Know	3	1	2
00001234	2	Refusal	3	1	2	1
00001235	Missing Answer	1	5	Missing Answer	2	1
00001236	6	1	2	2	5	2
00001237	Missing Answer	2	4	1	Missing Answer	2

Note: This graphic contains fake data.

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

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
**LINES REMOVED**
```
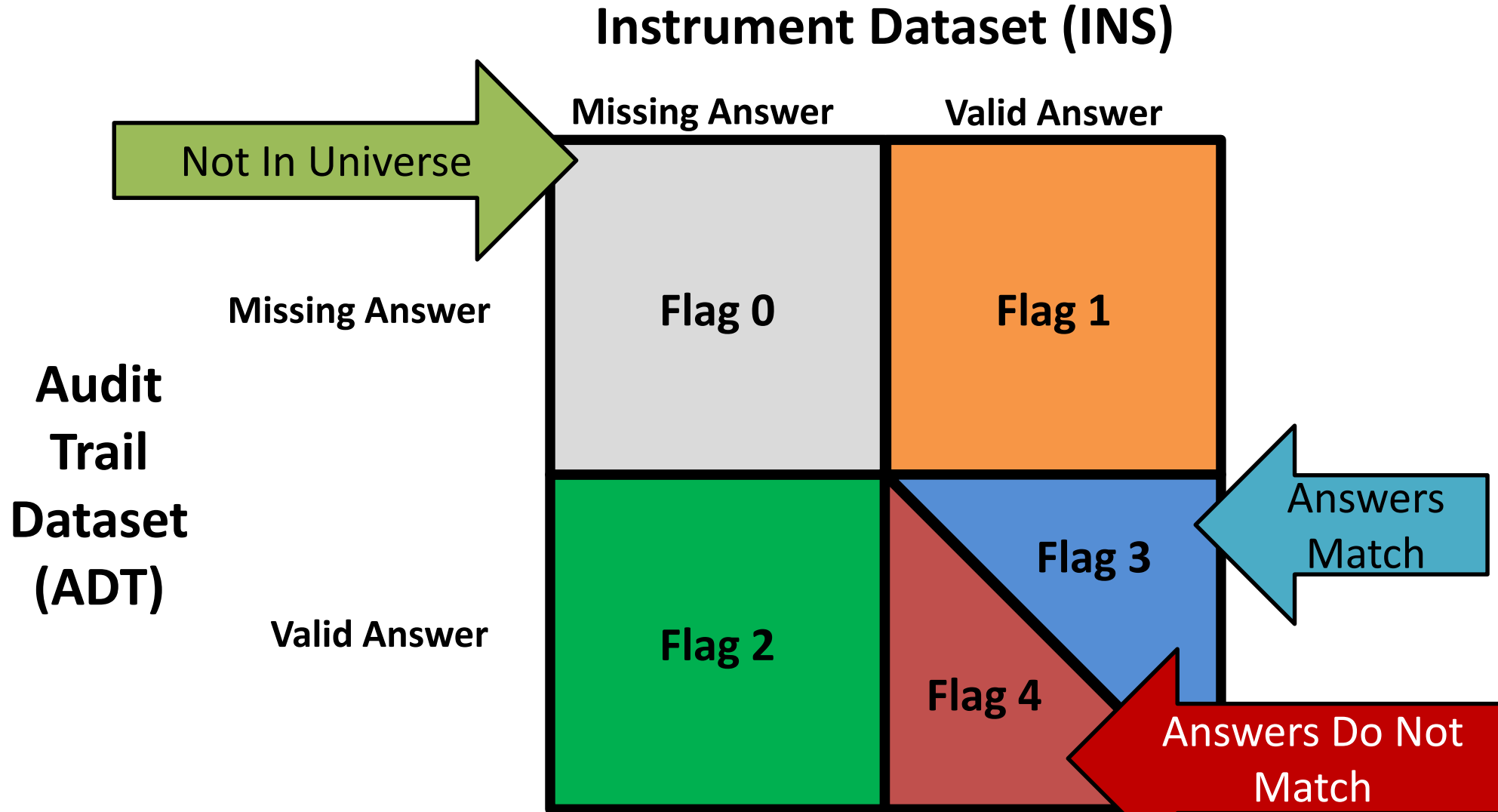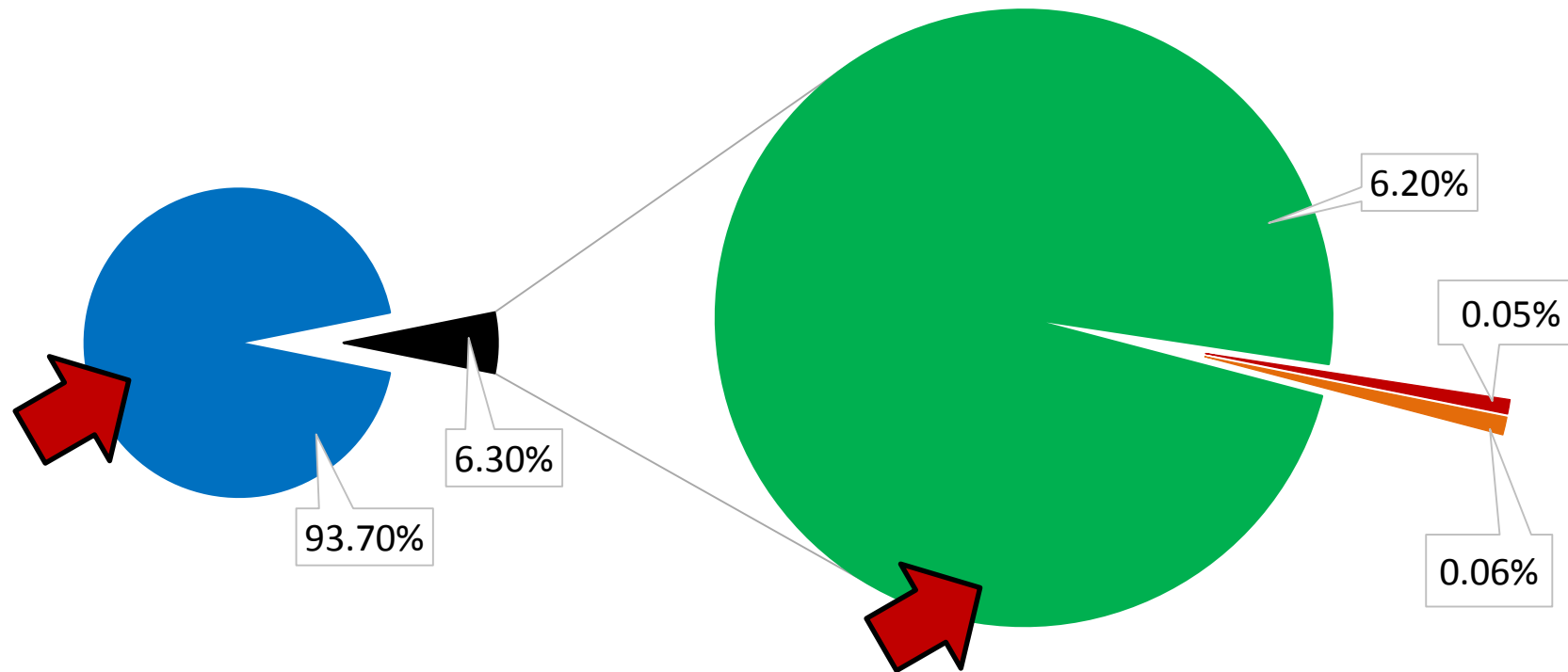
Note: This graphic contains fake data.

Multipoint Answer Comparison Flag System

Results of Wave 1 Data Retrieval - EHC Dataset

6.20%

0.05%

6.30%

93.70%

0.06%

■ Instrument dataset has an answer, but the audit trail is missing

■ Instrument dataset is missing and the audit trail has an answer

■ Both instrument dataset and the audit trail have answers and match

■ Both instrument dataset and the audit trail have answers and don't match

Results of Wave 2 Data Retrieval - EHC Dataset

33.12%

66.88%

31.30%

1.46%

0.36%

■ Instrument dataset has an answer, but the audit trail is missing

■ Instrument dataset is missing and the audit trail has an answer

■ Both instrument dataset and the audit trail have answers and match

■ Both instrument dataset and the audit trail have answers and don't match
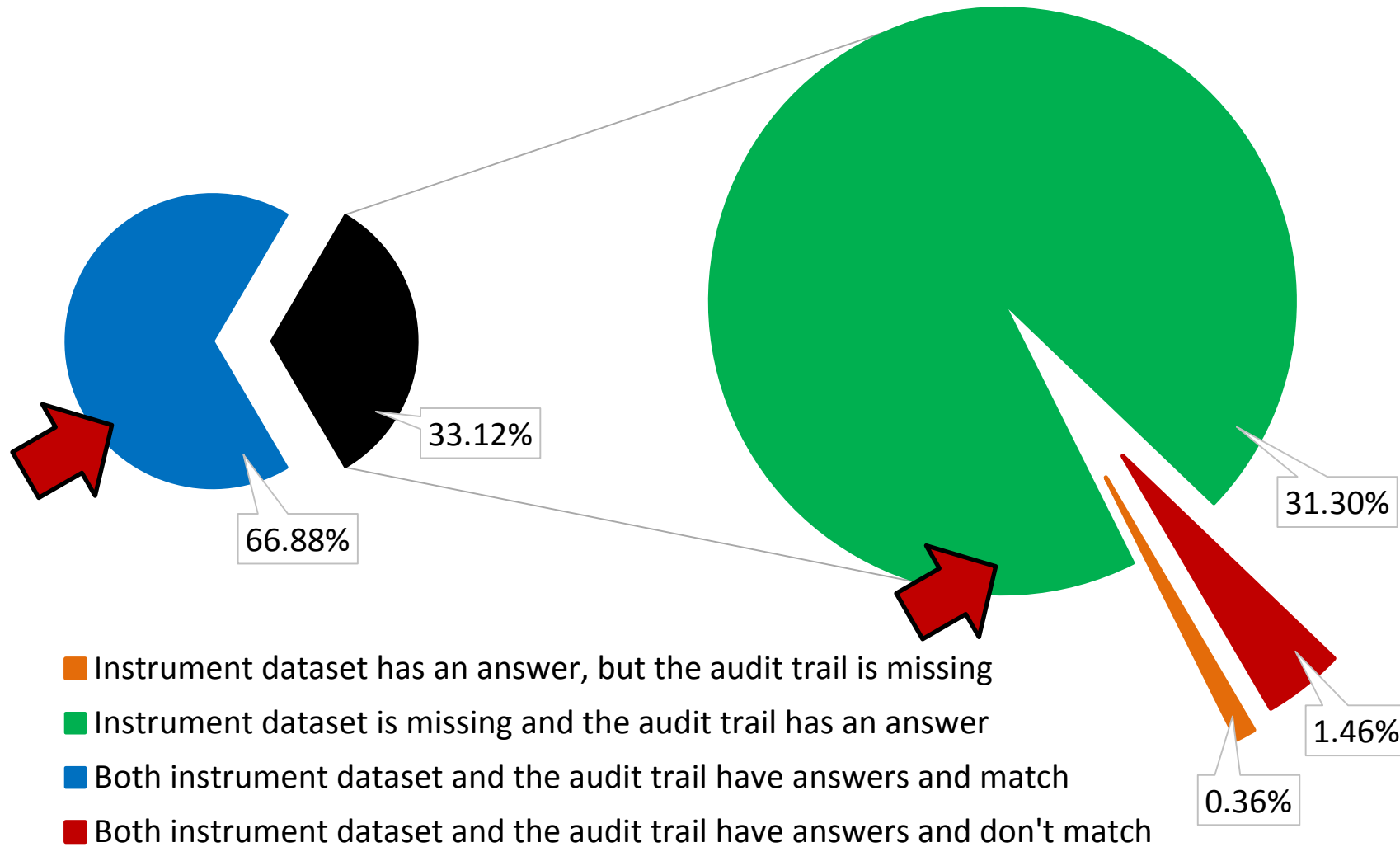
Dataset Levels – Processing Logic

- Household and Person
  - Locate the answer field
  - Retrieve the answer field
  - Assign the answer field to the correct person or household record
- Event History Calendar (EHC)
  - More complicated

EHC Processing Logic - Complications

- C# programming language
- FR actions logged
- Answers copied to other household members
- Provisional spells (Wave 2+ issue)
- NoJob spells

EHC Processing Logic – Wave 2 and Future Waves

- Wave 2:
  - Descriptive data
  - Program Structure
- Future Waves:
  - Expand respondent identification and tracking systems
  - Move to retrieving all answer values from audit trail files
  - EHC topic's spell coverage verification system
    - Identify incomplete/unverified provisional spells

Conclusion

- Retrieving missing answer data is possible

- Beneficial, despite the difficulties

- Data retrieval code needs to adapt to file contents

- Formatted to integrate updates easier

Thank You

Patrick C. Campanello

Survey of Income and Program Participation

Patrick.Campanello@census.gov
www.census.gov
www.census.gov/sipp