# MATHEMATICA
## Policy Research

# Commercial Locating Database Efficacy for Telephone Surveys of Low-Income Populations

**Presentation at the American Association for Public Opinion Research Annual Conference**
**New Orleans, Louisiana**

**May 2017**

Kim Mook • Sarah Forrestal

# Background

# Background

- **Program evaluation often requires contact with specific sample members**

- **Sample frames vary in quality and volume**

- **Locating low-income populations can be difficult:**
  - **High mobility rates**
  - **Variable employment**
  - **Phones that cycle in and out of service**

- **Locating databases are often used to find contact information**

MATHEMATICA
Policy Research

# Background

- **Commercial locating databases are paid services that provide additional contact information**

- **Databases aggregate from both public and private records**
  - Public records such as USPS, voter registration, and motor vehicle registration
  - Private records obtained from proprietary sources

- **More research is needed to understand how low-income, hard-to-reach populations are represented in these databases**

MATHEMATICA
Policy Research

# Background

- **More research is needed to determine how to most effectively utilize results from locating searches**

- **Contact information from multiple sources is more likely to be accurate**

  - **Balance against the costs of using locating services**

**MATHEMATICA**
Policy Research

# Research Questions and Methods

# Sample

- **Evaluation of federal demonstration projects targeting low-income households with children**

- **Two grantees provided contact information from program administrative records (n = 7,246) or consent forms (n = 4,750)**

- **Before data collection, project submitted contact information from the two grantees (n = 11,996) to commercial locating databases**
  - **LexisNexis Accurint**
  - **TransUnion TLO**

**MATHEMATICA**
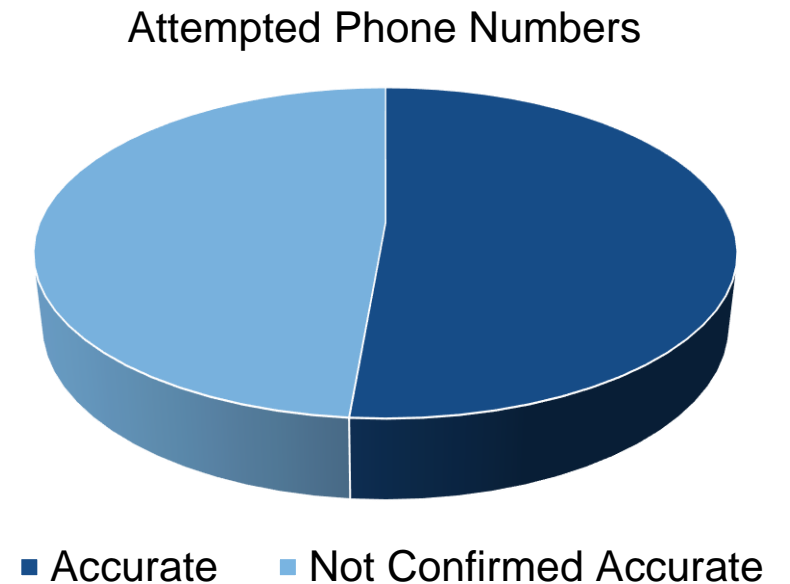Policy Research

# Research Questions

- **How representative are two different locating databases compared with the sample frame?**

- **How accurate are the telephone numbers from each source compared with the sample frame?**

- **Is telephone number accuracy associated with the combination of sources that provided the number?**

# Methods: Coverage

- **Determining sample representativeness:**
  - Number of returned records (hit rate) from each locating database
  - Compared the age, gender, and race/ethnicity of the returned records with the frame

**MATHEMATICA**
Policy Research

# Methods: Phone Number Accuracy

- **Confirmed the accuracy of 9,585 of the attempted 18,659 phone numbers obtained from original sources**

  - "Accurate" means we confirmed the respondent was reachable at that number (via direct contact, voicemail, etc.)

- **Compared accuracy across sources**
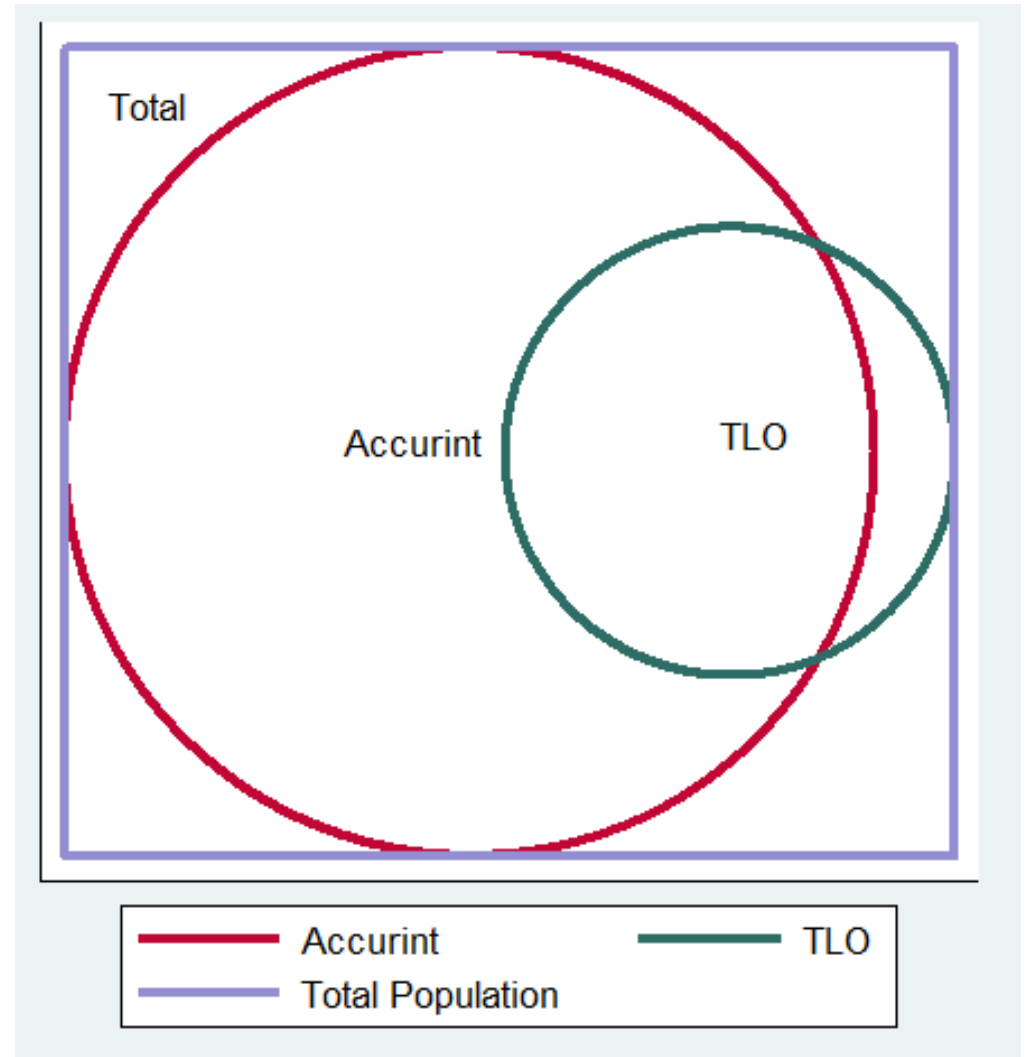
- **Compared accuracy by combination of sources**

Attempted Phone Numbers



■ Accurate   ■ Not Confirmed Accurate

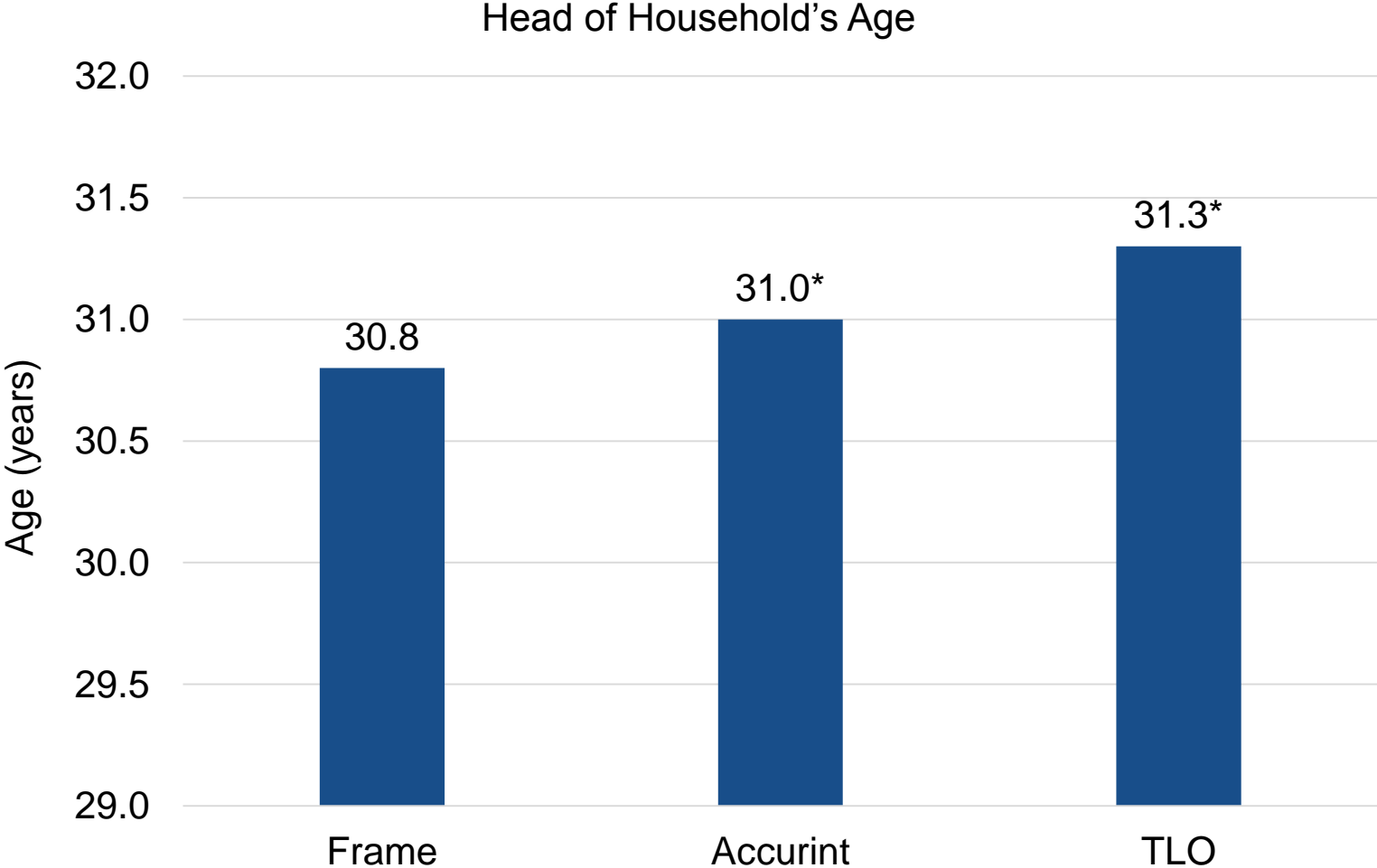MATHEMATICA
Policy Research

# Findings

# How Representative Are Two Different Locating Databases Compared with the Sample Frame?

# What Is the Coverage of Each Database?

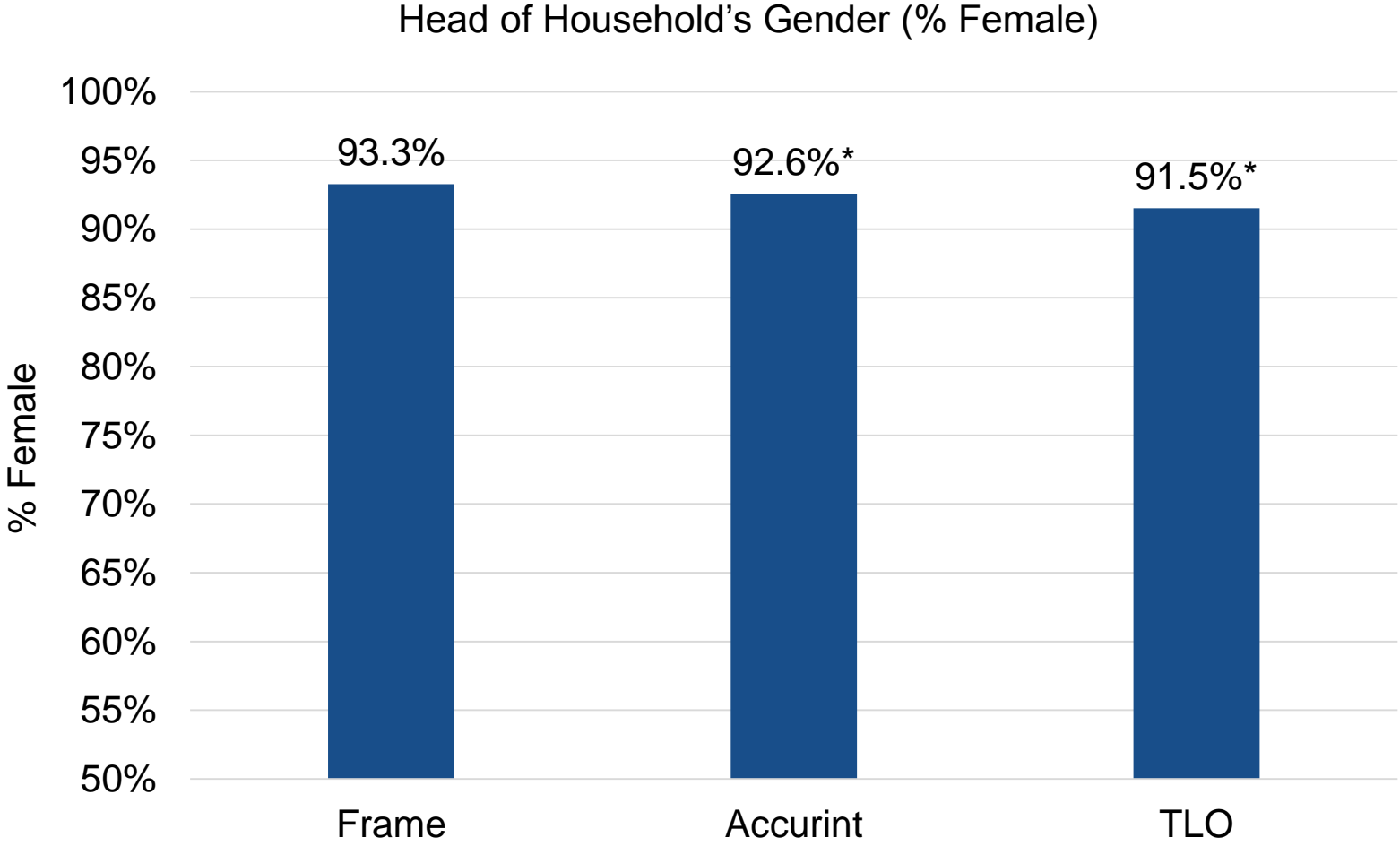- **Submitted 7,246 households**

- **Accurint hit rate: 71%**
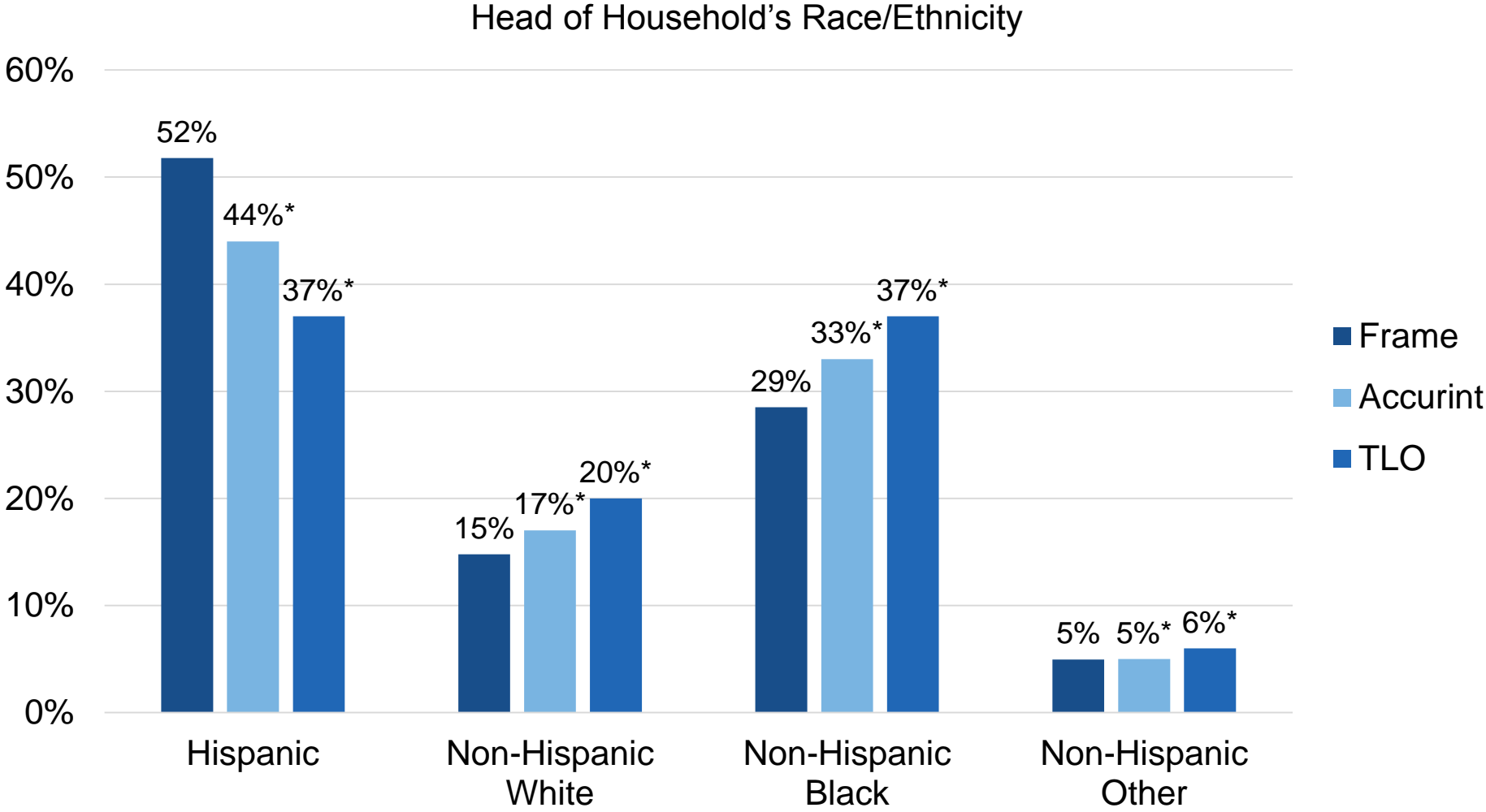  - **5,170 households**

- **TLO hit rate: 22%**
  - **1,592 households**

MATHEMATICA
Policy Research

# What Is the Coverage of Each Database?

Head of Household's Age

# What Is the Coverage of Each Database?

Head of Household's Gender (% Female)

# What Is the Coverage of Each Database?

## Head of Household's Race/Ethnicity



Legend: Frame, Accurint, TLO

- Hispanic: Frame 52%, Accurint 44%*, TLO 37%*
- Non-Hispanic White: Frame 15%, Accurint 17%*, TLO 20%*
- Non-Hispanic Black: Frame 29%, Accurint 33%*, TLO 37%*
- Non-Hispanic Other: Frame 5%, Accurint 5%*, TLO 6%*

# How Accurate Are the Telephone Numbers from Each Source Compared with the Sample Frame?

# Phone Number Accuracy by Source

- **Accurint returned 7,323 phone numbers that were attempted**

- **TLO returned 1,320 phone numbers that were attempted**

| Source | Total attempted |
|---|---|
| Frame | **13,585** |
| Accurint | **7,323** |
| TLO | **1,320** |

MATHEMATICA
Policy Research

# Phone Number Accuracy by Source

- **Numbers from the frame were accurate 62% of the time**

- **Numbers from Accurint were accurate 39% of the time**

- **Numbers from TLO were accurate 36% of the time**

| Source | Confirmed accurate | | Not confirmed | | Total |
|--------|--------|--------|--------|--------|--------|
| Frame | 8,457 | 62.3%* | 5,128 | 37.8%* | **13,585** |
| Accurint | 2,834 | 38.7%* | 4,489 | 61.3%* | **7,323** |
| TLO | 476 | 36.1%* | 844 | 64.0%* | **1,320** |

*$p < 0.001$

MATHEMATICA
Policy Research

# Is telephone number accuracy associated with the combination of sources that provided the number?

MATHEMATICA
Policy Research

# Telephone Accuracy by Source Overlap

| Source | Total | |
|--------|------:|------:|
| Frame only | 10,848 | 58.1% |
| Accurint only | 4,068 | 21.8% |
| TLO only | 409 | 2.2% |
| Frame & Accurint | 2,423 | 13.0% |
| Frame & TLO | 79 | 0.4% |
| Accurint & TLO | 597 | 3.2% |
| All sources | 235 | 1.3% |
| **Total** | **18,659** | **100.00%** |
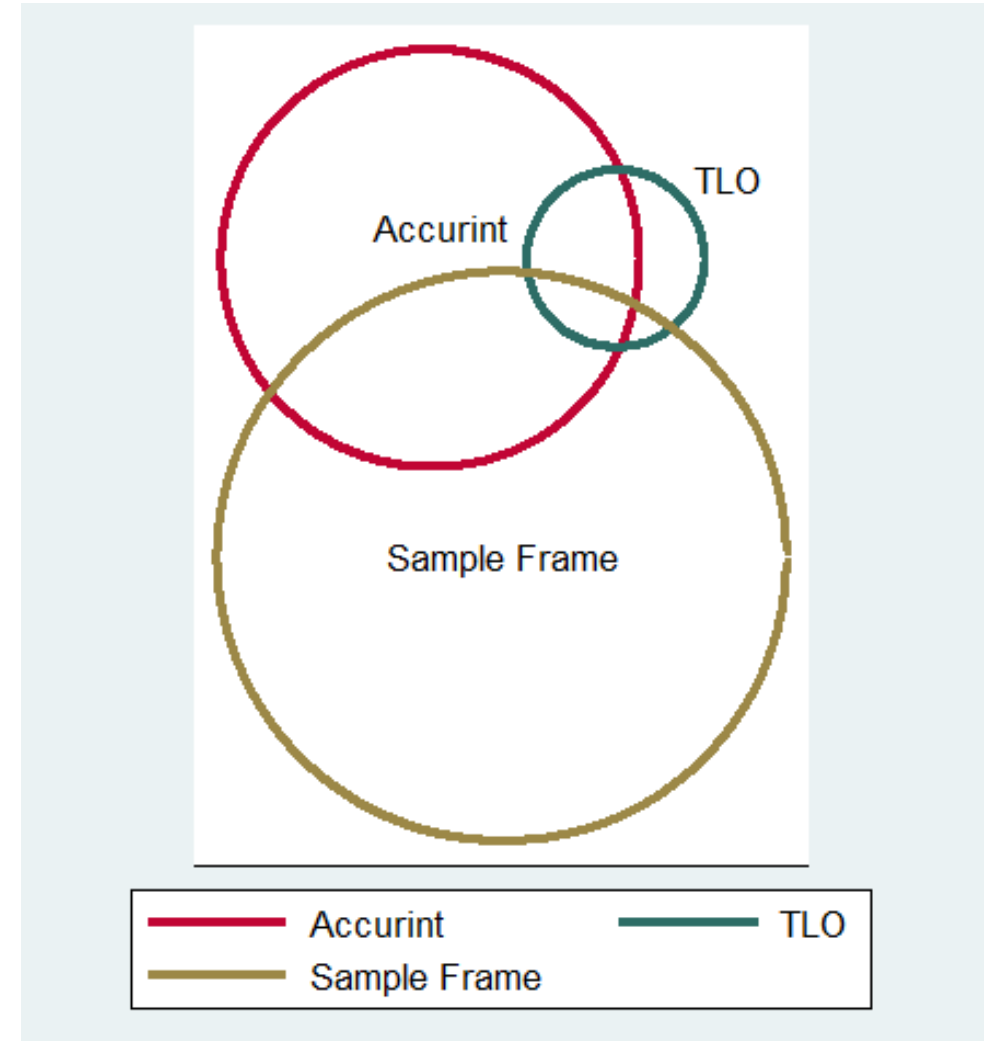
Telephone Accuracy by Source Overlap
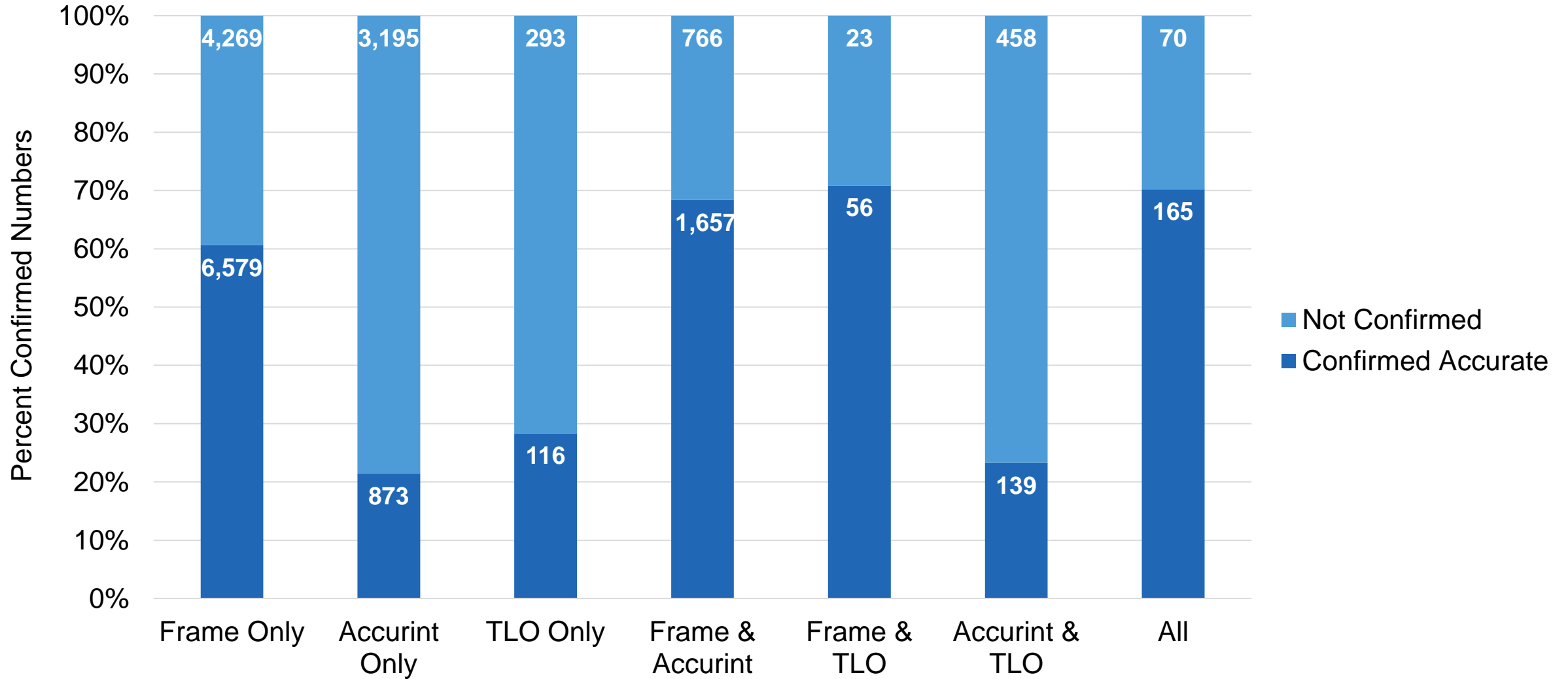
# Telephone Accuracy by Source Overlap

- **Odds ratio for probability of a accurate number, compared with the frame alone**

  – **Databases alone are less likely to be accurate than the frame alone**

  – **Overlap between the frame and a database is more likely to be accurate than frame alone**

MATHEMATICA
Policy Research
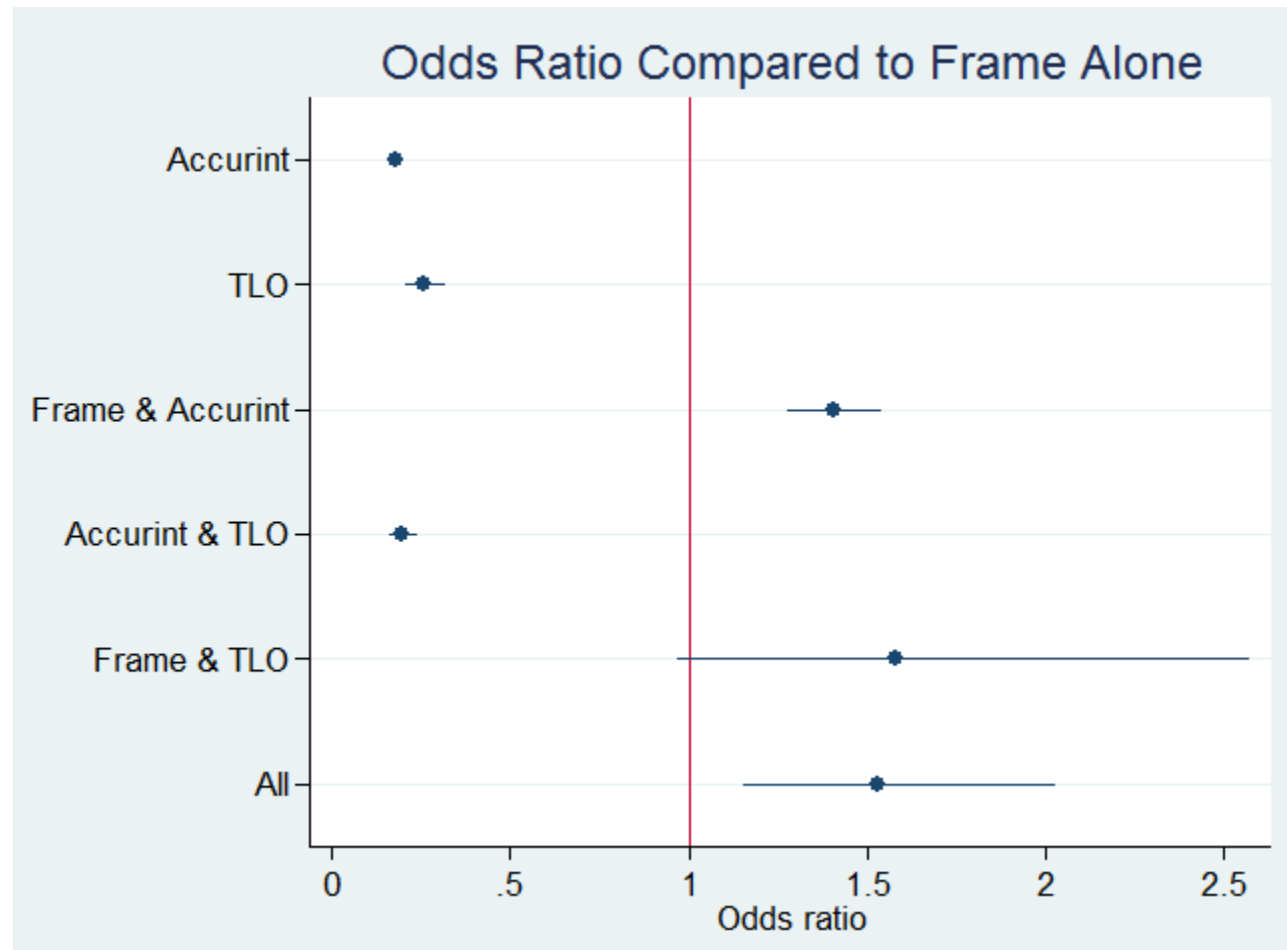
# Conclusions

# Database Coverage

- **Coverage of low-income households with children varied across sources**

- **Age and gender do not show substantive differences**
  - **Future research with less homogeneous sample may reveal larger biases in age and gender**

- **We found substantive differences by race and ethnicity**
  - **Hispanics are under-represented in locating databases**

# Phone Number Accuracy by Source

- **In this study, the sample frame of low-income adults was the best source for accurate phone numbers (62% accuracy rate)**

- **Overlap between sources (including sample frame) increased the accuracy rate by 40% to 60%**

- **Locating sources can be used to prioritize numbers from the frame**



**MATHEMATICA**
Policy Research
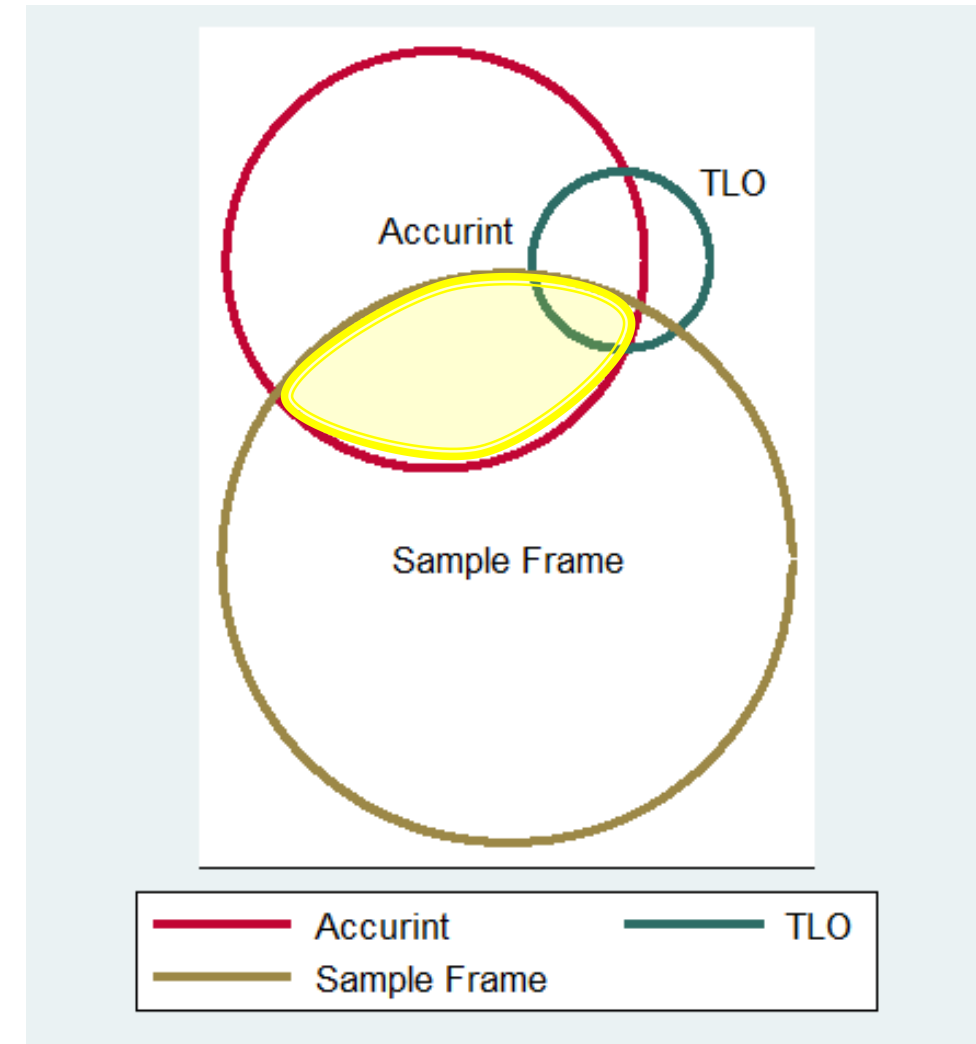
# Takeaway Messages

- **Locating low-income populations can be challenging**

- **Commercial locating databases can vary in coverage**

- **Identifying numbers provided by the frame and by a locating database can increase dialing efficiency**

- **More research is needed to investigate locating database efficacy among:**

  - **Other populations**
  - **Different sample frames of varying quality**

**MATHEMATICA**
Policy Research

# Future Research

- **Investigate representativeness by income, education, home ownership, and number of children**

- **Investigate phone number source and quality by respondent characteristics**

- **Investigate findings with different populations**

**MATHEMATICA**
Policy Research

# For More Information

- **Kim Mook**
  - **KMook@mathematica-mpr.com**

- **Sarah Forrestal**
  - **SForrestal@mathematica-mpr.com**

**MATHEMATICA**
Policy Research

# Phone Number Accuracy by Source Overlap

| Source | Confirmed accurate | | Not confirmed | | Total |
|---|---|---|---|---|---|
| Frame only | 6,579 | 60.6% | 4,269 | 39.4% | **10,848** |
| Accurint only | 873 | 21.5% | 3,195 | 78.5% | **4,068** |
| TLO only | 116 | 28.4% | 293 | 71.6% | **409** |
| Frame & Accurint | 1,657 | 68.4% | 766 | 31.6% | **2,423** |
| Frame & TLO | 56 | 70.9% | 23 | 29.1% | **79** |
| Accurint & TLO | 139 | 23.3% | 458 | 76.7% | **597** |
| All | 165 | 70.2% | 70 | 29.8% | **235** |
| **Total** | **9,585** | **51.4%** | **9,074** | **48.6%** | **18,659** |

**MATHEMATICA**
Policy Research

# Odds Ratio for Source, Compared with Frame Alone

| Source | Odds ratio | Confidence Interval (CI) |
|---|---|---|
| Frame only | 1 | -- |
| Accurint only | 0.18* | 0.16–0.19 |
| TLO only | 0.26* | 0.20–0.32 |
| Frame & Accurint | 1.40* | 1.28–1.54 |
| Accurint & TLO | 0.20* | 0.16–0.24 |
| Frame & TLO | 1.58 | 0.97–2.56 |
| All | 1.53* | 1.15–2.03 |

**MATHEMATICA**
Policy Research