Evaluating Recall Error in Survey Reports of Move Dates through a Comparison with Records in a Commercial Database

Mary H. Mulry, Elizabeth M. Nichols, and Jennifer Hunter Childs, U.S. Census Bureau

Parvati Krishnamurty, NORC

2013 International Total Survey Error Workshop

June 3, 2013

United States™ Census Bureau

Motivation for study

- Surveys ask respondents to recall events
- U.S. censuses ask respondents to recall where they lived on Census Day, April 1
- Some census evaluation interviews may occur 11 months after April 1
- Assumption: respondents who move around April 1 are able to give correct answers
- Recall errors may cause respondents to be enumerated at the wrong location

Literature on Event Recall

- Much research on respondent recall and several models of survey response process
  Cannell et al. (1981), Sudman et al. (1996)
- Telescoping may be backward or forward
  Rubin & Baddeley (1989), Huttenlocher et al. (1990)
- Wording of questions may aid recall
  Prohaska et al. (1998), Tourangeau et al. (2000)
- Use of move date as anchor may aid recall
  Wagenaar (1986), Belli et al. (2008)
- Could not find research on error in recalling move dates

Survey that asks move months: National Longitudinal Survey of Youth

- NORC conducts NLSY for BLS
  - 85% of interviews in person
- Annual interviews, Oct to May
- Focus on NLSY97 cohort
  - ages 12 to 17 when first interviewed in 1997
- Retention rate high
  - If miss one round, use Accurint database to find respondent for next round

United States Census Bureau

Commercial database

- Accurint
  - Compiles records from over 10,000 sources, some public & some proprietary
    - Ex: utility bills, credit cards, magazine subscriptions, bank accounts, court records
- Maintains records for over 400 million unique identities
- Meets federal confidentiality requirements under CIPSEA

Research strategy

- Use reported moves for NLSY97 cohort
  - Ages 23 to 29 for 3 interview rounds 2006 - 2008
- Match NLSY97 movers to Accurint move records for their address
  - Matching conducted in 2011
- Investigate error in reported move month by using matched results to develop regression model

Define move month variables relative to interview month

NLSY-lag = (interview month - NLSY reported month)

ACC-lag = (interview month - Accurint record month)

True-lag = (interview month - True month) (not observed)

<u>Example for interview in Nov</u>

NLSY report: Sept → NLSY-lag = 2

Accurint record: Dec → ACC-lag = -1

True month: Oct → true-lag = 1

Regression model

$$\text{ACC-lag} = \alpha + \beta * \text{NLSY-lag} + \varepsilon$$

where α is intercept; β is coefficient, and ε is random error

Assumption

True move month and average Accurint start month differ by a constant C

In our notation:

$$\text{ave(ACC-lag)} = \text{True-lag} + C$$

Interpretation of regression parameters

- α = ave(ACC-lag - β*NLSY-lag)

Estimate ave(ACC-lag) by setting NLSY-lag = 0

- β = ave(ΔACC-lag)/ΔNLSY-lag
  = Δtrue-lag/ΔNLSY-lag
  - β = 1 implies no evidence of recall error
  - β < 1 implies backward telescoping in NLSY
  - β > 1 implies forward telescoping in NLSY

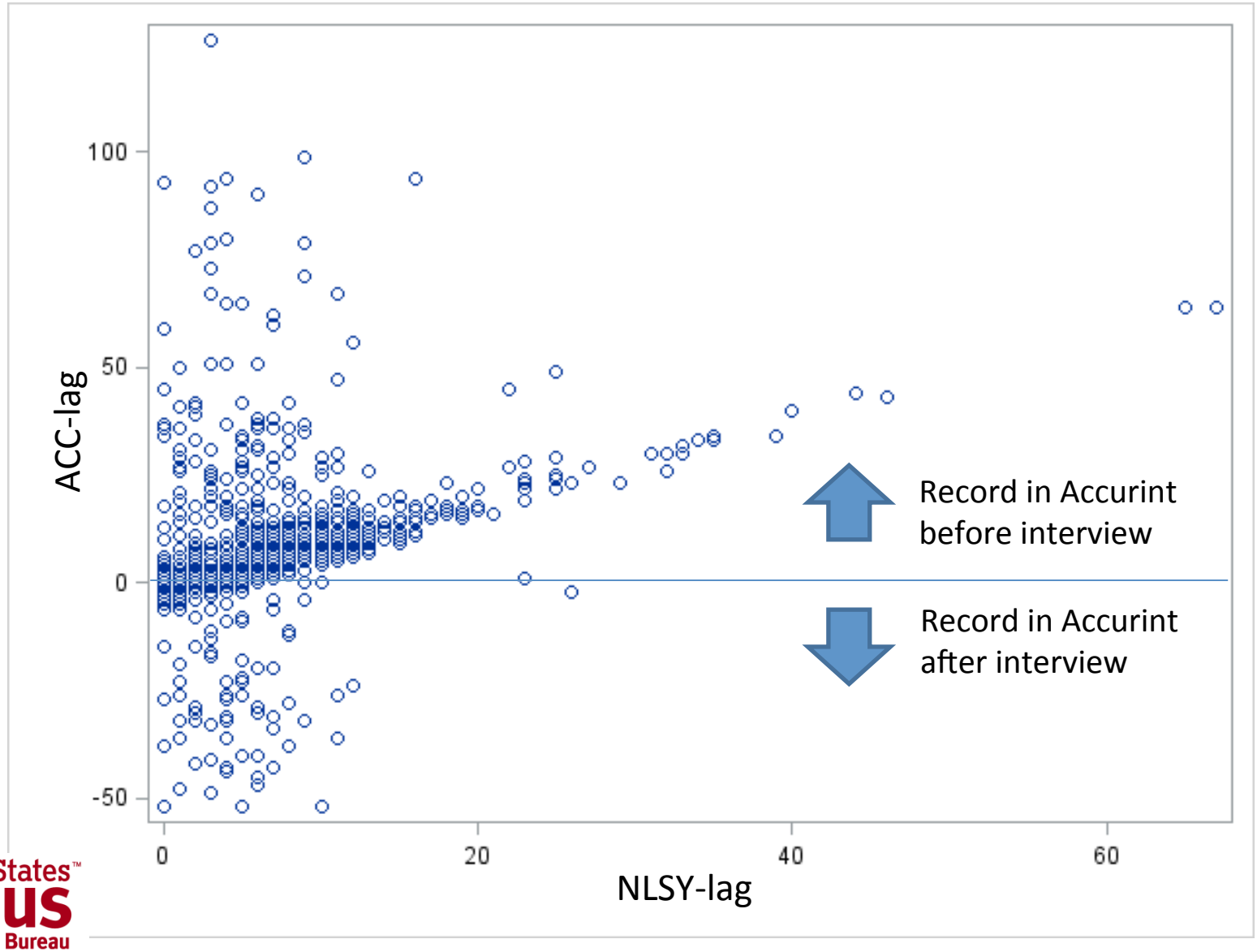
Matching limitations using NLSY data on moves

- Move to address at time of interview
  - Complete street address, city, state, zip code
- Moves between cities since last interview
  - Only city, state, zip code; no street address
- Moves within cities since last interview
  - Only number of moves; no address information
- Move month may be missing in NLSY or start month may be missing in Accurint

Data sets

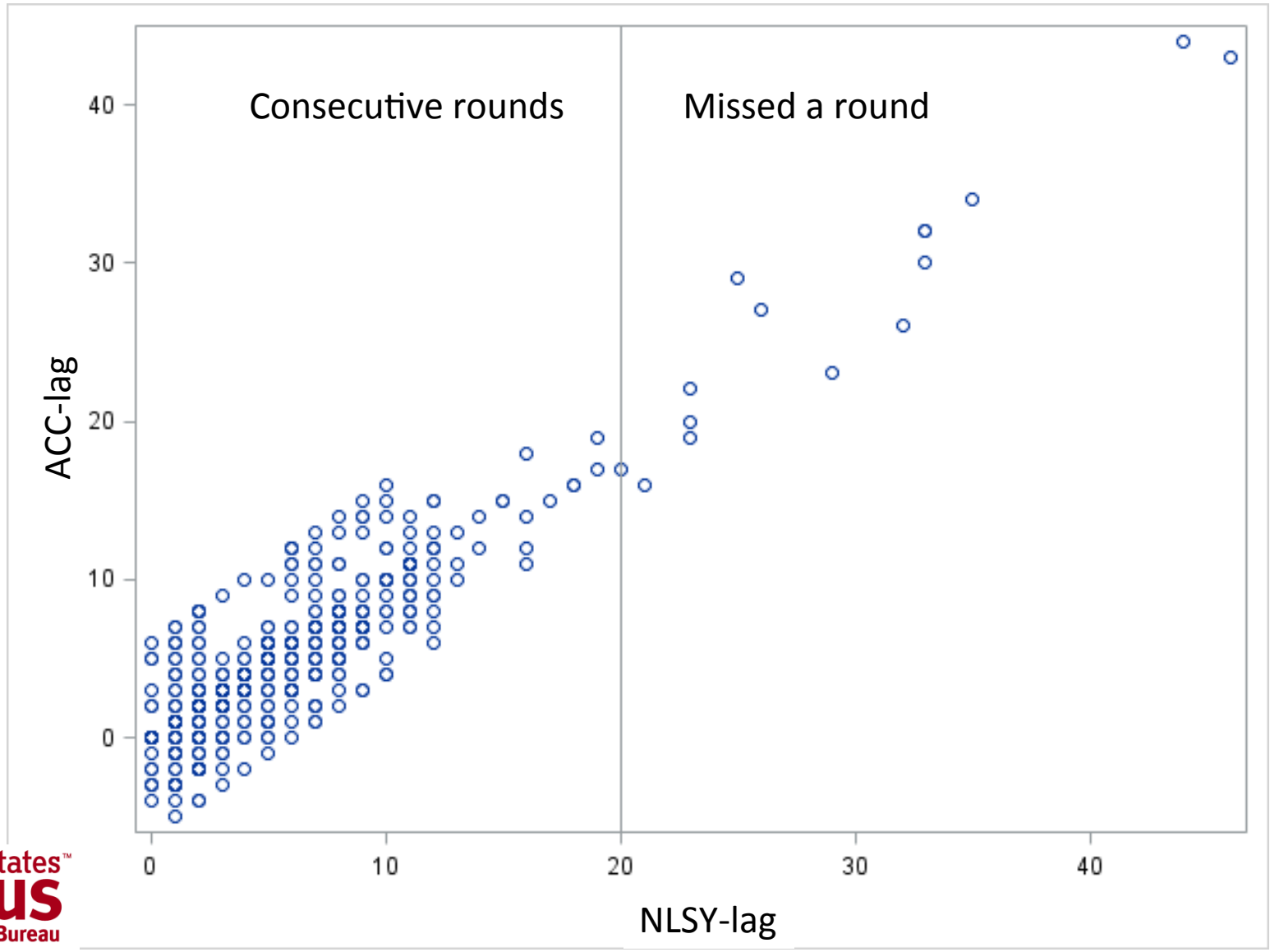
- Geographic match (2600 moves)
  - Match on city-state or zip code
  - Have NLSY move date & Accurint start date
  - Includes inter-jurisdictional moves
- Address match (1159 moves)
  - Match on complete address
  - Have NLSY move date & Accurint start date
  - Only moves to interview address

ACC-lag vs. NLSY-lag
Address match (zoom in)



United States Census Bureau

ACC-lag vs. NLSY-lag
Address match & difference < 6 months



United States Census Bureau

Restrictions on datasets

- NLSY-lag < 20 months
  - Only those interviewed in consecutive rounds
- $|$ NLSY-lag – ACC-lag $|$ ≤ 6 months
  - Attempt to assure both sources refer to same event

Address match data model

- n=410 moves; stronger matching criteria

	estimate	p-value
α	-0.171	0.460
β	0.919	<.0001

- $\alpha < 0$, but not significant
  - no evidence of nonzero average lag in Accurint
- $\beta < 1$ indicates backward telescoping
  - Is β different from 1?
  - t test indicates Yes, with p-value = 0.009

Geographic match data model

- n=1,094 moves; weaker match criteria

	estimate	p-value
α	-0.704	<.0001
β	0.954	<.0001

- $\alpha < 0$, & significant
  - indicates constant average delay in Accurint
- $\beta < 1$ indicates backward telescoping
  - Is β different from 1?
  - t test indicates Yes, with p-value = 0.007

Summary

- Both models indicate backward telescoping, but differ on when it starts affecting responses for ages 23 to 29
  - Address match data: 6 months after move
  - Geographic match data: 10 months after move
- Models differ on existence of nonzero average lag in Accurint records
  - Address match data: none
  - Geographic match data: 0.7 months after move

Future research

- More research needed to identify point when error in recall of move month becomes problematic
- Complete address and strong matching criteria necessary for definitive results
- Database search requires special instructions to identify the last start date when a person lives at an address, moves out, and then moves back in
- Suggest collecting source of database records to determine the best sources

Contact

mary.h.mulry@census.gov