



Examining the relationship between the accuracy of self-reported data and the availability of respondent financial records

Emily Geisen

Charles Strohm

M. Christopher Stringer

Brandon Kopp

Ashley Richards

Disclaimer

The findings and conclusions in this presentation are those of the authors and do not necessarily reflect official views of the U.S. Bureau of Labor Statistics or the U.S. Census Bureau.

Introduction

- **Problem:** self-reported data are subject to recall error due to the following:
 - Event characteristics
 - Question characteristics
 - Respondent characteristics

- **Possible solution:** incorporate the use of respondent records

- **Limitation:** records can be incomplete and not available for all survey items of interest

Research Questions

- What impact does an incomplete set of records have on our ability to measure and reduce recall error?
 - Are respondents more or less likely to have records for events they recall accurately?
 - How does the relationship between records and accuracy affect estimates of recall error?

Survey Data

U.S. Consumer Expenditure Quarterly Interview Survey

- Monthly survey, captures continuous information about buying habits of Americans.
- Data used to calculate the weight of index items in the Consumer Price Index (CPI), a leading economic indicator
- U.S. Census Bureau conducts survey for Bureau of Labor Statistics

Survey Data: Consumer Expenditure Records Study

- Non-probability, feasibility study of 115 participants
- Designed to explore:
 - Availability and feasibility of using financial records
 - Direction and magnitude of measurement error in CE
- Respondents completed 2 interviews in 7-day period
 - 1st interview: abbreviated version of the CEQ interview
 - 2nd interview: participants reviewed receipts and other financial records with interviewer
- Compared self-reports to records to determine accuracy

Statistical Methods

1. Determine if there is potential for bias in the availability of records
2. Determine if there is measurement error bias
3. Use logistic regression to predict probability of having a record and create 5 propensity strata
4. Examine how accuracy of self-reports changes across propensity strata

1a. Demographic factors affecting likelihood of having a record

Respondent Characteristic	Odds R	P-value
Household Size	0.89	0.18
Age	1.01	0.32
Non-Hispanic White (vs other)	1.60	0.04
Owners (vs renters)	1.73	0.06
Lower income (vs higher incomes)	0.27	0.02
Lower education (vs higher education)	0.10	0.02
North Carolina (vs DC)	0.66	0.17
Working (vs not working)	0.86	0.48
Female (vs male)	1.59	0.05
Married (vs not married)	0.70	0.32

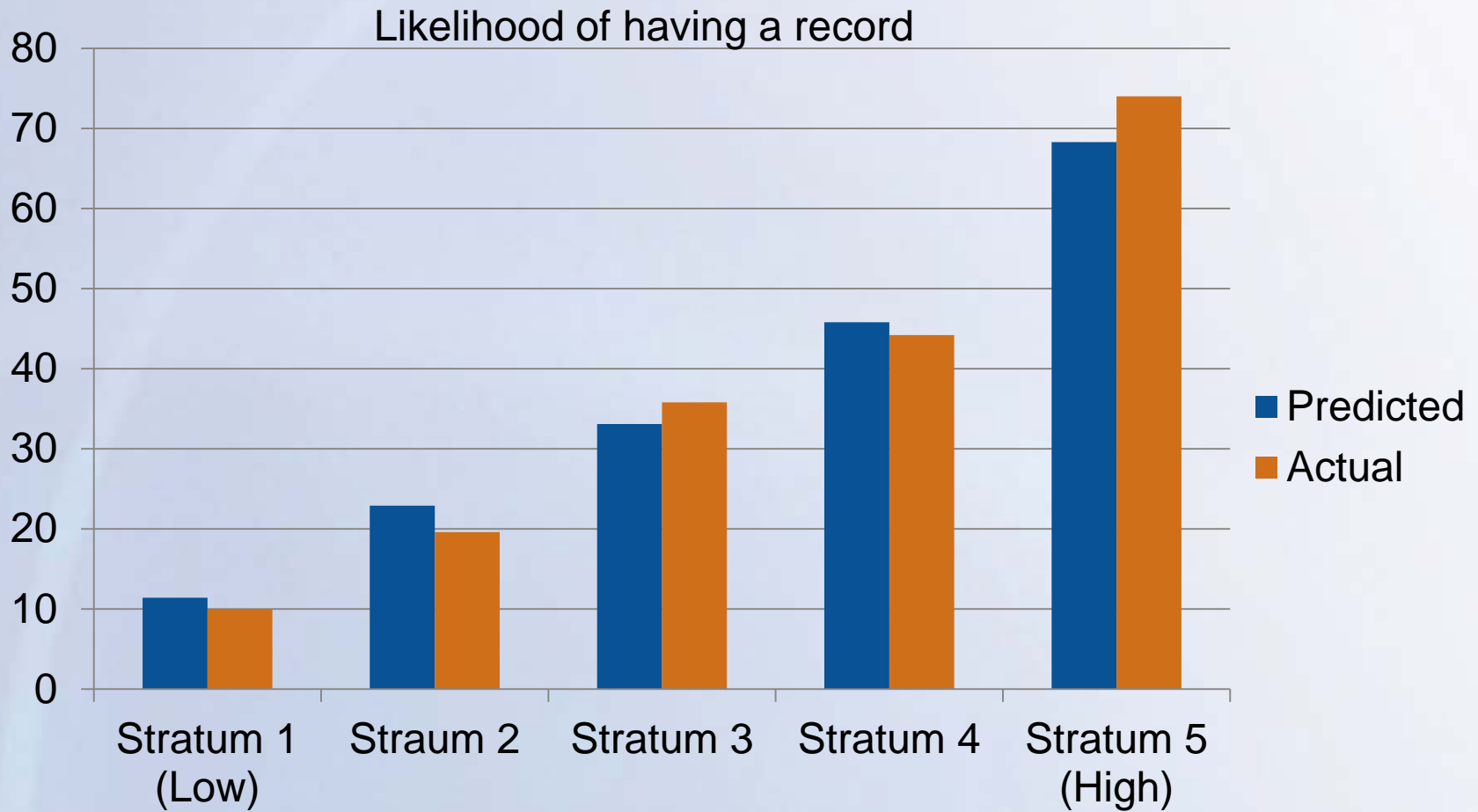
1b. Item characteristics affecting likelihood of having a record

Item Characteristic	Odds R	P-value
Order reported	0.96	<0.01
Recent	0.74	<0.01
Amount of item (\$)	1.00	<0.01
Housing expenditures (vs. other)	7.84	< 0.0001
Phone expenditures (vs. other)	8.78	< 0.0001
Utility expenditures (vs. other)	7.47	< 0.0001
Appliance or furniture exp (vs. other)	10.9	< 0.0001
Clothing expenditures (vs. other)	1.19	0.46

2. Measurement Bias

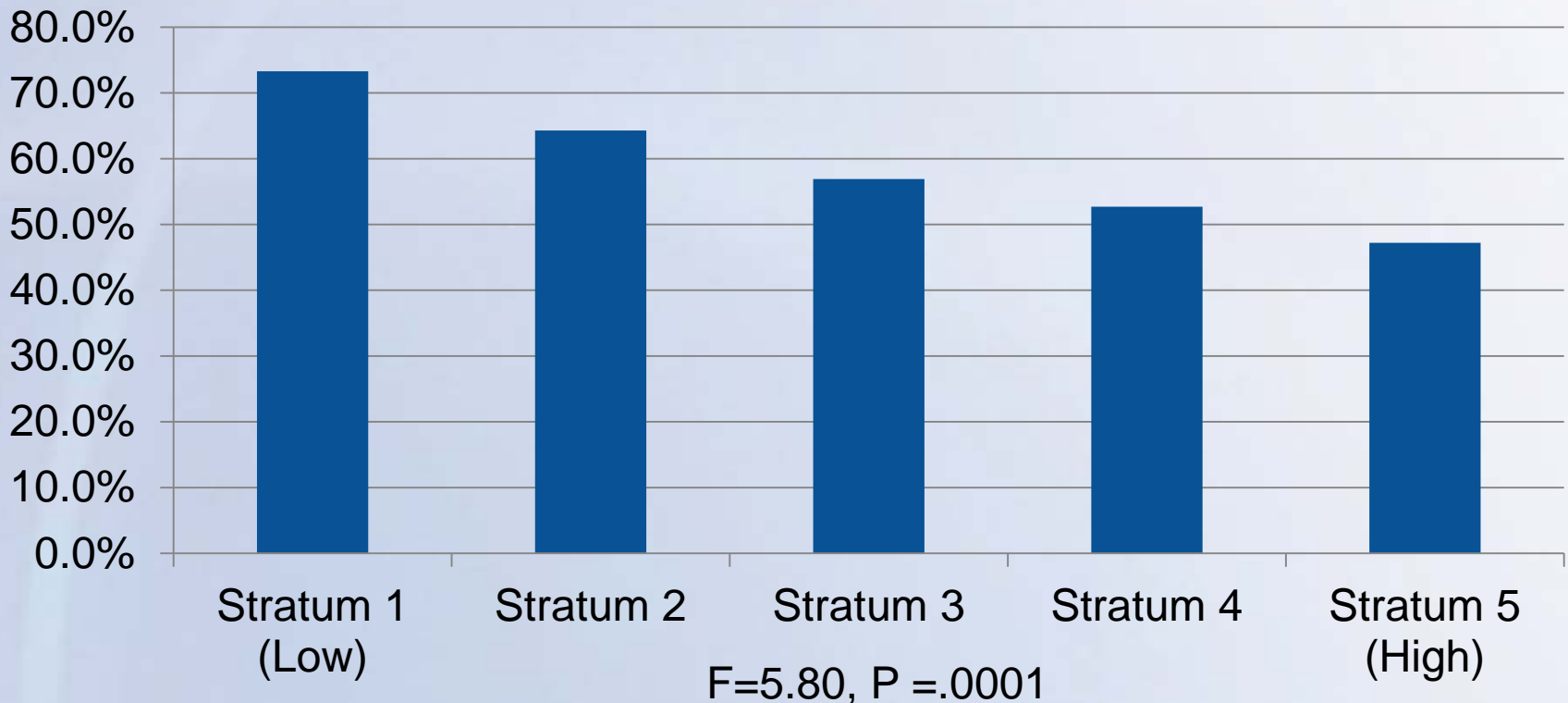
- Respondents reported 3,039 expenditures (Interview 1)
- Records available for 36% of expenditures (Interview 2)
- Of expenditures with records: 46% of self-reports were inaccurate (>10% difference)
- Magnitude of misreporting was 29.9% on average

3. Propensity Strata



4. Relationship between likelihood of having a record and accuracy of self-report

Percent of self-reports matching records



Implications

- When using an incomplete set of records, it is important to investigate the impact on measurement error
- In the CE records study, even an incomplete set of records can reduce error associated with recall failure
- Although respondents only had records for roughly 1/3 of reported expenditures, they were more likely to have records for expenditures they were WORST at recalling

Limitations

- Non-probability, feasibility study limited to 115 people
- Study data were about financial expenditures. It is unclear how well this would apply to other types data where records can be used
- We do not know the accuracy for the self-reports without records.
- There may be other factors affecting availability of records and accuracy that we did not examine.

Next Steps

- Continue using records to supplement recall and reduce respondent burden
- Consider having respondents collect records proactively (instead of retroactively)
- Explore the use of records as a primary source of data collection

Contact Information

Corresponding author:

Emily Geisen

RTI International

919-541-6566

egeisen@rti.org