

# Data quality in the Consumer Expenditure Survey's Large Scale Online Diary Feasibility Test

Laura Erhard  
Joint Statistical Meetings  
August 8-12, 2021

Gray Jones, Parvati Krishnamurty, and Brett McBride



# Outline

- Brief Background
- Test Design
- Results
- Next steps



# Consumer Expenditure Survey

- Collects spending data on the U.S. Population
- Sponsored by the Bureau of Labor Statistics, collected by the Census Bureau
- Survey participants report dollar amounts for all non-investment purchases. Business expenses and reimbursements are excluded.
- Provide expenditure weights for the U.S. Consumer Price Index (CPI)
- Quarterly Interview and Diary Survey
- Diary Survey: two, consecutive one-week paper diaries\* designed to collect small, frequently purchased items



# Large Scale Feasibility Test (LSF) of the Online Diary Design

- Starting sample size: 2,500 addresses
- Field period: October 2019-April 2020
- Two visits, one for diary placement and one at the end of the diary keeping period (“pickup”)
- Push to online diary if the household respondent is eligible, if they refuse or are ineligible-- they get paper diaries
- Summary reports for interviewers to monitor the progress of online diary participants
- Recall tab in the CAPI instrument for interviewer to record additional expenses reported by respondent at pickup interview (normally transcribed in the paper diary)
- Embedded incentives/postcard experiment



# Response Rates and Mode



# Estimation Response Rates: LSF Test vs. CE Production

Disposition	CE Production (2019q4-2020q1)		LSF Test (2019q4-2020q1)	
	n	%	n	%
<b>Eligible Sample</b>	12,690	100	4,268	100
<b>Complete Cases</b>	<b>5,838</b>	<b>46.0</b>	<b>1,782</b>	<b>41.8</b>
<b>Overall Nonresponse</b>	6,852	54.0	2,486	58.3
<i>Noncontact</i>	939	7.4	284	6.7
<i>Refusal</i>	3240	25.5	1108	26.0
<i>Diary placed late</i>	939	6.5	408	9.6
<i>Minimal expenditure</i>	481	3.8	253	5.9
<i>Other</i>	1,253	9.9	433	10.1



# Mode at Completion

LSF	n	%
Online	663	37.2
Paper	1074	60.3
Other	45	2.5





# Demographics of Mode

- Similar to other mode studies
- LSF online households: Younger, with higher household income, higher education
- Less likely to be Hispanic or Latino origin





# Expenditure Estimates and Quality

## *preliminary*



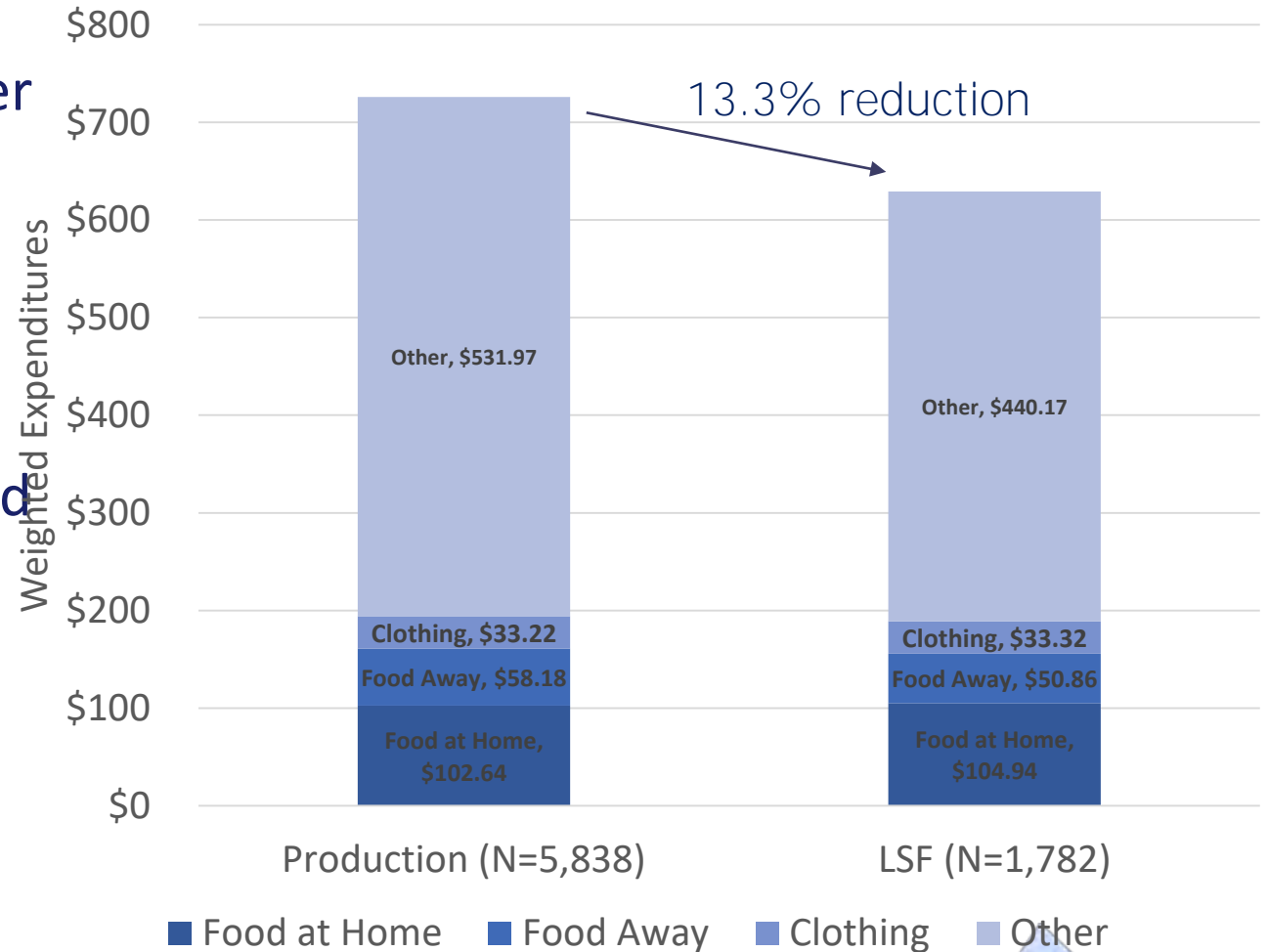
# Earlier analysis quality takeaways

- Item Non-response:
  - ▶ Lower non-response for business of expenditure in online diaries compared to production diaries (by design, since required field)
  - ▶ Other item non-response rates similar across LSF modes and production
- Similar rates of Diary Fatigue (reporting more in the first week than the second week)
- Recalled expenses at pick-up handled differently for online diaries than in the paper diaries → potential missed recall
- Lower counts of expenditures in the LSF



# Expenditure Totals By Section

- Average weighted expenditures higher in Production than LSF.
- Most of difference attributable to reduction in “Other” category. Significant by minor difference in Food away from home
- No differences in Food at home and Clothing.



# Expenditure Regression Results

	<u>Coefficient</u>	<u>SE</u>	
Intercept	-234.008	(239.25)	
LSF online	-114.704	(46.33)	*
LSF paper	-69.693	(67.04)	
Log(pre-tax income)	106.645	(22.63)	***
Non-homeowner CU	-155.335	(40.31)	***
Highest: HS grad or less	-245.555	(75.33)	**
Highest: Some college	-205.940	(76.18)	*
Highest: Assoc. degree	-171.684	(77.13)	*
Highest: Professional degree	108.471	(99.10)	

\* p < 0.05

\*\* p < 0.01

\*\*\* p < 0.0001

The full model did not find the CU's size, respondent's age, sex, race or ethnicity, or the RO in which the interview occurred to be significant factors in reported expenditures.



# Editing Rates Production vs. LSF

	CE Production		LSF Test	
	n	%	n	%
<b>Minimal expenditure</b>	481	3.79	253	5.93
<b>Allocation rate</b>	17,096	9.33	3,424	11.71



# Summary

- Preliminary research showed little to no impact on item nonresponse, diary fatigue
- Response rate slightly lower, but partly attributable to test effect
- Take up rate much lower than expected for the online diary
- Quality indicators show online diary *may* not provide as high quality data
  - ▶ Higher allocation rates
  - ▶ Higher number of removed diaries
  - ▶ Lower expenditures after controls

## BUT....

- Earlier research identified concerns on how recalled expenses were captured during LSF test, particularly for online diaries → since mitigated with further interviewer training and revised procedures
- Online diary helped us “pandemic proof” our survey
- Some respondents *expect* an online diary, Interviewers are happy to have the option.



# Online Diary next steps

- Continued use due to the pandemic
- Go/No-Go decision on planned **official** implementation in July 2022
- Recommendations for changes to full implementation protocols
- Continuous improvement
  - ▶ autosuggest feature, prompts, trip based diary
  - ▶ more use of receipts
  - ▶ self-administration test





# Contact Information

**Laura Erhard**

Branch of Research and Program Development  
Division of Consumer Expenditure Surveys

[Erhard.Laura@bls.gov](mailto:Erhard.Laura@bls.gov)

