

# Evaluating Data Quality in the Consumer Expenditure Survey's Test of a Fully Self-Administered Diary in an Online Panel

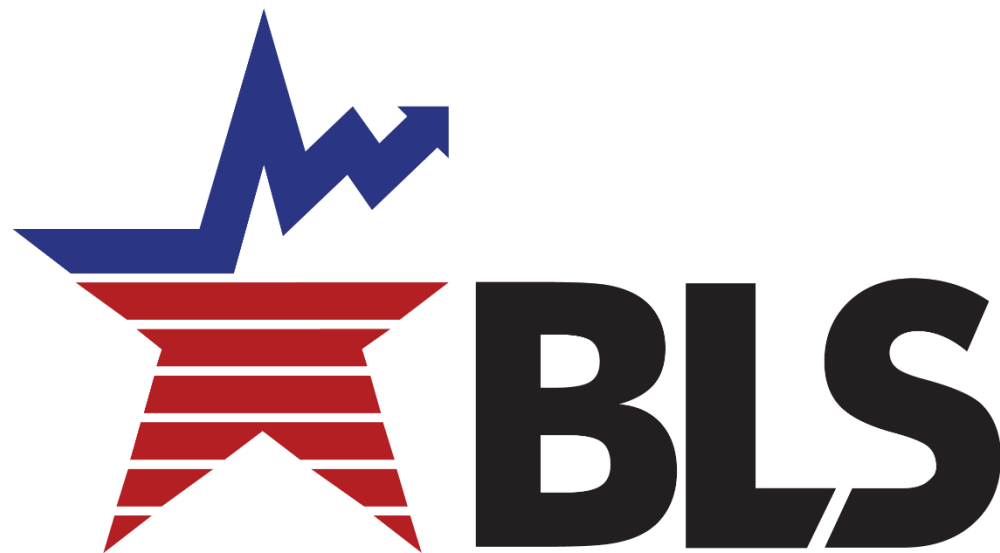
Nikki L. Graf, Graham Jones, and Tucker Miller  
U.S. Bureau of Labor Statistics

AAPOR Annual Conference  
May 15, 2024



# Outline

- Background – Consumer Expenditure (CE) Diary Survey
- Online Panel Self-Administered Diary (OPSAD) test design
- Data quality analysis – Comparison between OPSAD and CE Online Diary
- Examining entry count differences
- Summary and next steps



# Consumer Expenditure (CE) Diary Survey

- Expenditure data are entered by the respondent into a paper or online diary form
- Role of interviewers
  - ▶ Place the diaries
  - ▶ Train respondents how to record the household's daily expenditures
  - ▶ Contact them during the week to encourage diary keeping
  - ▶ Collect any forgotten expenditures at the end of the diary period
- Online Panel Self-Administered Diary (OPSAD) test – Purposes:
  - ▶ Assess the impact of shifting from current diary placement protocols to a fully self-administered mode on data quality
  - ▶ Investigate whether a self-administered online diary in an online panel could be an avenue for testing changes for CE's production diary



# Online Panel Self-Administered Diary (OPSAD) Test Design

- Conducted on the Ipsos KnowledgePanel<sup>®</sup>, a probability-based web panel designed to be representative of the U.S.
- Field period: November 2021 – January 2022
- Independent samples drawn for each of the 3 months
- Total starting sample: 12,700 adults representing their households
- Stages:
  - ▶ Initial Household Characteristics Survey (developed based on CE CAPI instrument)
  - ▶ Two-week Consumer Spending Diary
- Incentives:
  - ▶ \$2 cash equivalent for completing Household Characteristics Survey
  - ▶ \$50 cash equivalent for qualifying as a complete for the Spending Diary

# Diary Survey

- One household respondent reports on *all* household expenditures
- Expenditures grouped into four categories:
  - ▶ Food and Drinks For Home Consumption
  - ▶ Meals, Snacks, and Drinks Away from Home
  - ▶ Clothing, Shoes, Jewelry, and Accessories
  - ▶ All Other Products, Services, and Expenses



# Diary Instructional Resources for Respondents

	CE Online Diary	OPSAD
Interviewer to guide respondents on diary reporting, encourage diary keeping, and review diaries at the end	✓	✗
<i>In the online diary instrument:</i>		
User Guide pdf	✓	✓
Instructions, FAQs, Definitions, Examples	✓	✓
Video tutorial links	✓	✓
<i>Help Desk:</i>		
Diary access and navigation issues	✓	✓
Survey concepts (e.g., where to report items, level of detail to include)	✗	✓
<i>Email communications:</i>		
Invitation/Welcome email	✓*	✓
3 reminder emails and a closing email	✗	✓

# Data Quality – Comparison between OPSAD & CE Online Diary



# CE Online Diary and OPSAD Comparisons

- Focus on comparisons between CE Online Diary and OPSAD to control for mode effects
  - ▶ Past CE research has found meaningful differences between Paper and Online Diary users
- Complete diaries from November 2021 – January 2022
  - ▶ Diaries with no entries (no reported expenditures) are excluded
- CE Online Diary: 660 diaries (348 households)
- OPSAD: 5,856 diaries (3,154 households)
- Results shown here are unweighted



# Item Nonresponse

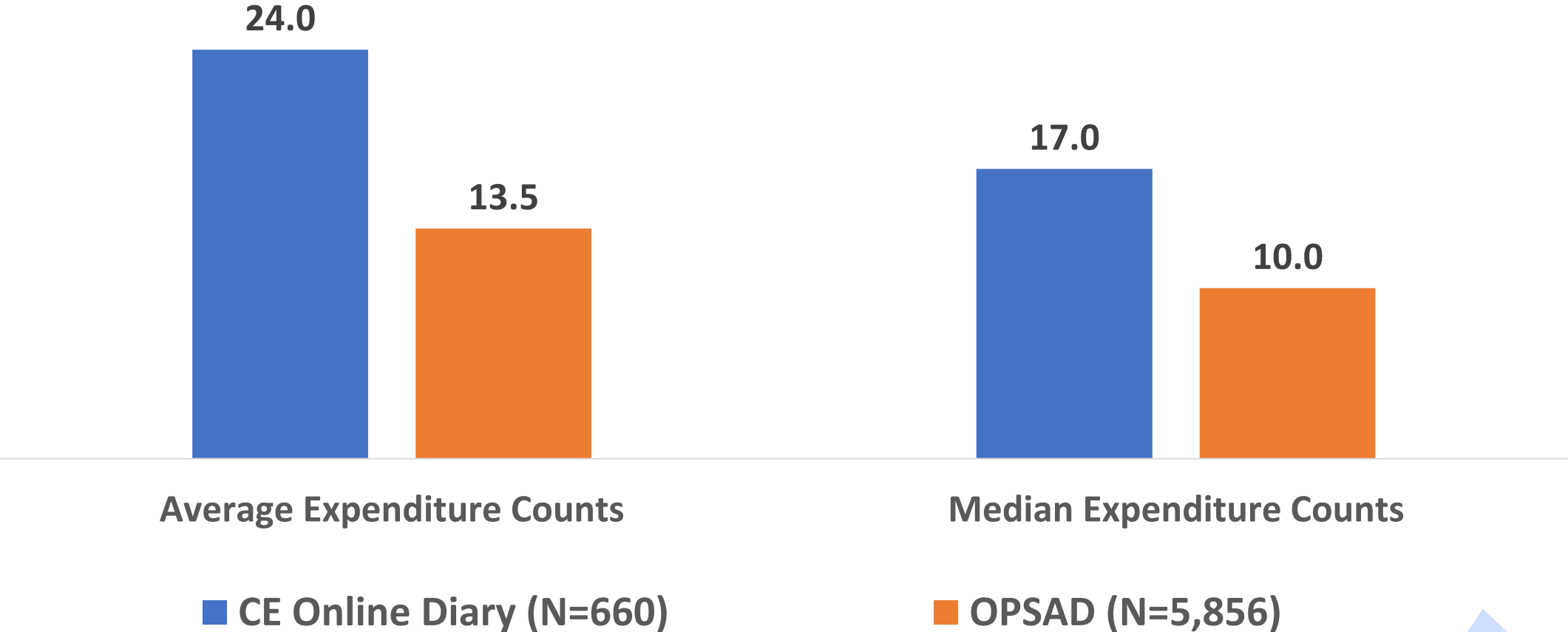
- For each item, respondents were required to provide:
  - ▶ Date of purchase
  - ▶ Business where item was purchased
  - ▶ Expense category (e.g., ‘Food and Drinks for Home Consumption’)
- But other fields could be left blank:
  - ▶ Cost
  - ▶ Item Description

% of Diary Entries with...	CE Online Diary	OPSAD
Cost field blank	0.3%	2.1%
Item description blank*	0.0%	1.1%
Invalid text in item description field* (e.g., Don't know, NA)	0.1%	1.6%
Invalid text in business name field (e.g., Don't know, NA)	0.5%	0.4%

*\*Excludes Meals Away from Home category*

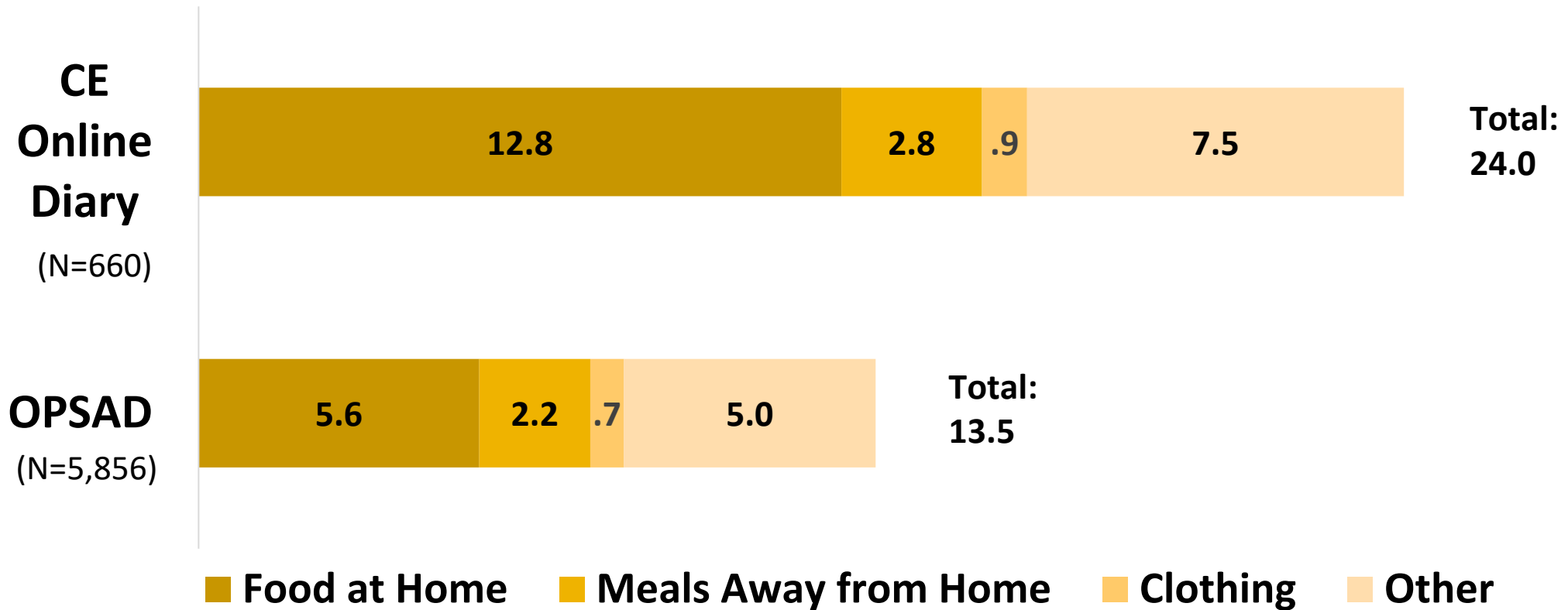
# Expenditure Counts

(among diaries with at least one entry)

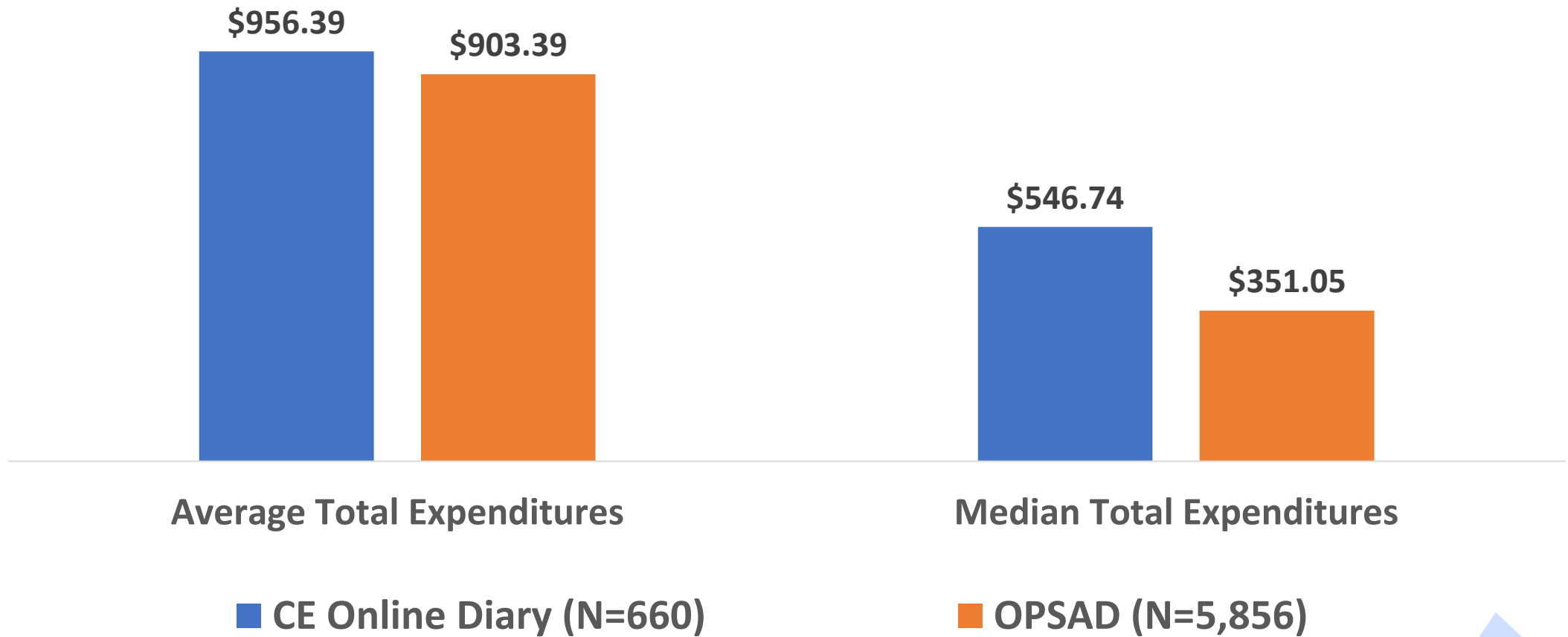


# Counts by Section

Average number of weekly diary entries, by expense category:



# Total Expenditure Amounts



# Drop-Off in Entry Counts

- Respondents may experience fatigue from the reporting process and report fewer expenditure entries over the diary period
- Drop-off measure: Averaged percentage of each household's two-week entries reported in Week 1
- Calculated for households having at least one expenditure in both weeks

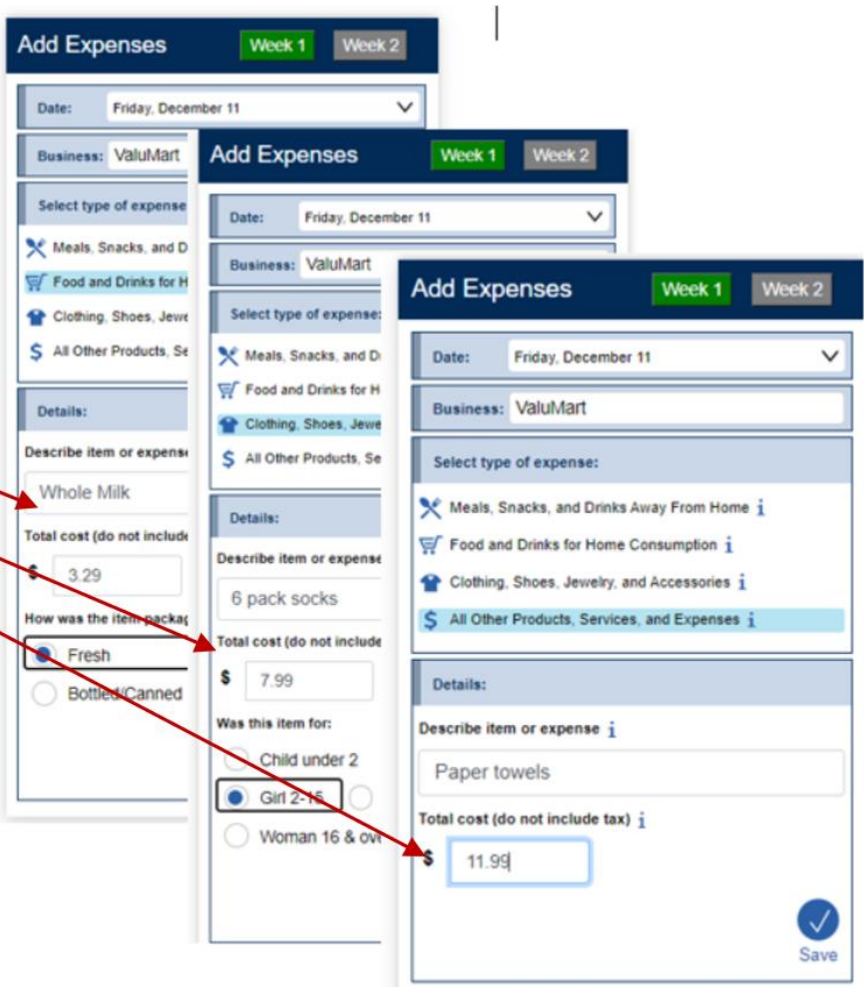
	CE Online Diary (n=312)	OPSAD (n=2,702)
Week 1 mean entries	25.5	15.1
Week 2 mean entries	23.3	13.2
Mean entries (Weeks 1 + 2)	48.8	28.2
<b>Avg. % of two-week entries reported in Week 1</b>	<b>52.6%</b>	<b>54.4%</b>

# Examining Entry Count Differences

- Why does OPSAD have lower weekly entry counts than the CE Online Diary?
  - ▶ Bundling expenses?
  - ▶ Compositional differences in the underlying samples?
  - ▶ Other factors related to study design and lack of interviewer guidance?

# Adherence to Itemization Instructions

- For all expense categories except Meals Away from Home, respondents are instructed to itemize – enter each item purchased as a separate entry



# Bundled Expenses

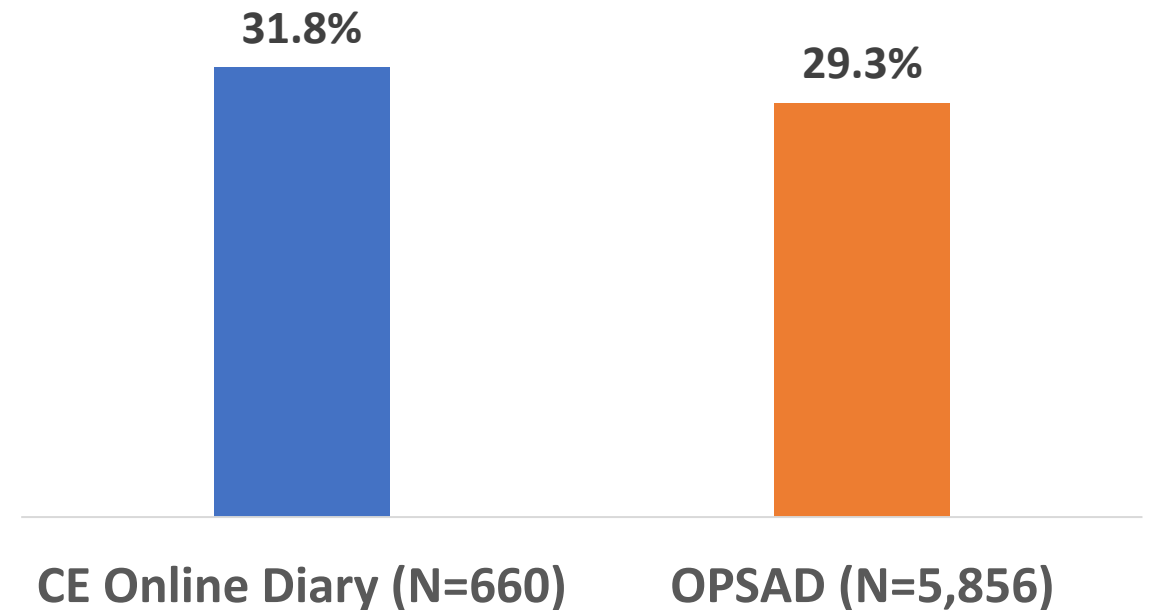
- Item descriptions containing the following strings were flagged as “bundled” entries:
  - ▶ GROCERIES
  - ▶ FOOD
  - ▶ CLOTHES
  - ▶ CLOTHING
  - ▶ HOME IMPROVEMENT
  - ▶ HOUSEHOLD
  - ▶ MISCELLANEOUS
  - ▶ MISC



# Expenditure Entry Bundling

- Similar shares of diaries contained bundled expenses
- Bundled expenses are most common in the Food at Home category

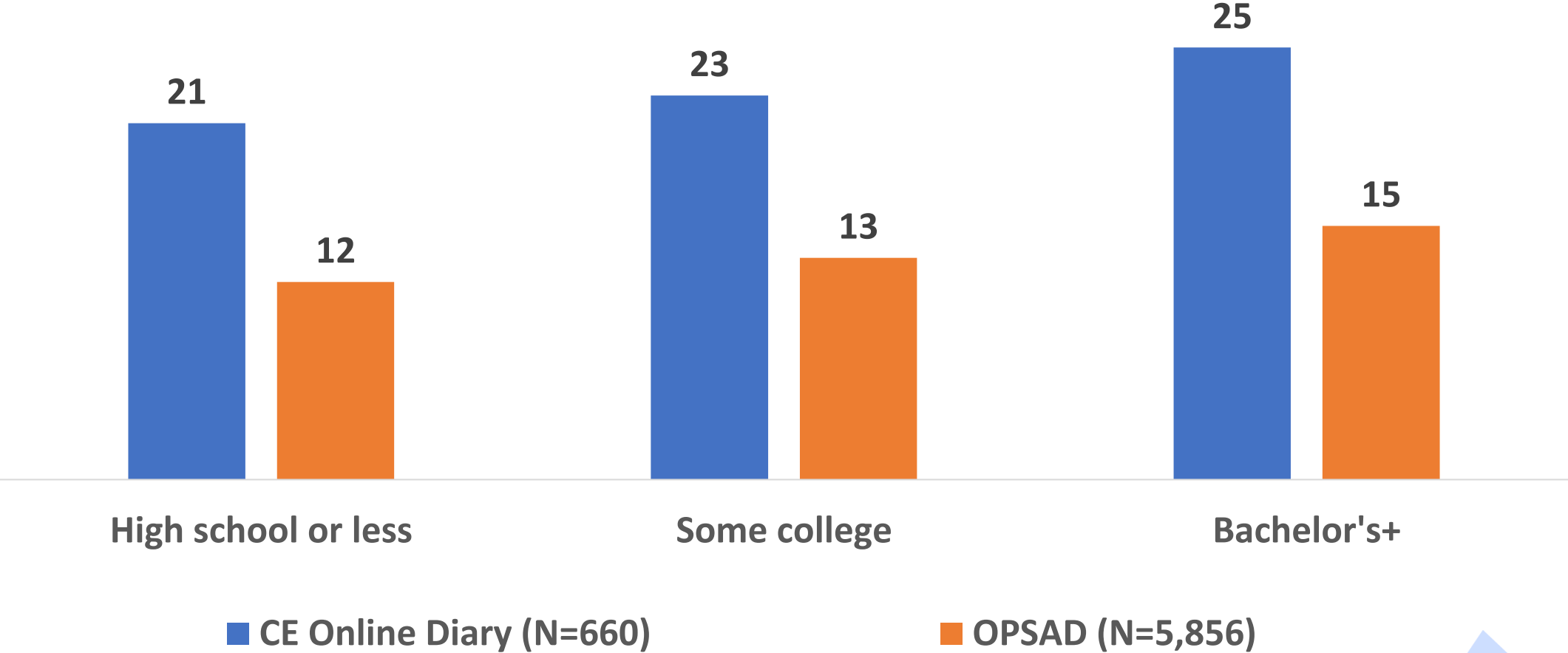
% of Diary Cases with Any Bundled Expenses



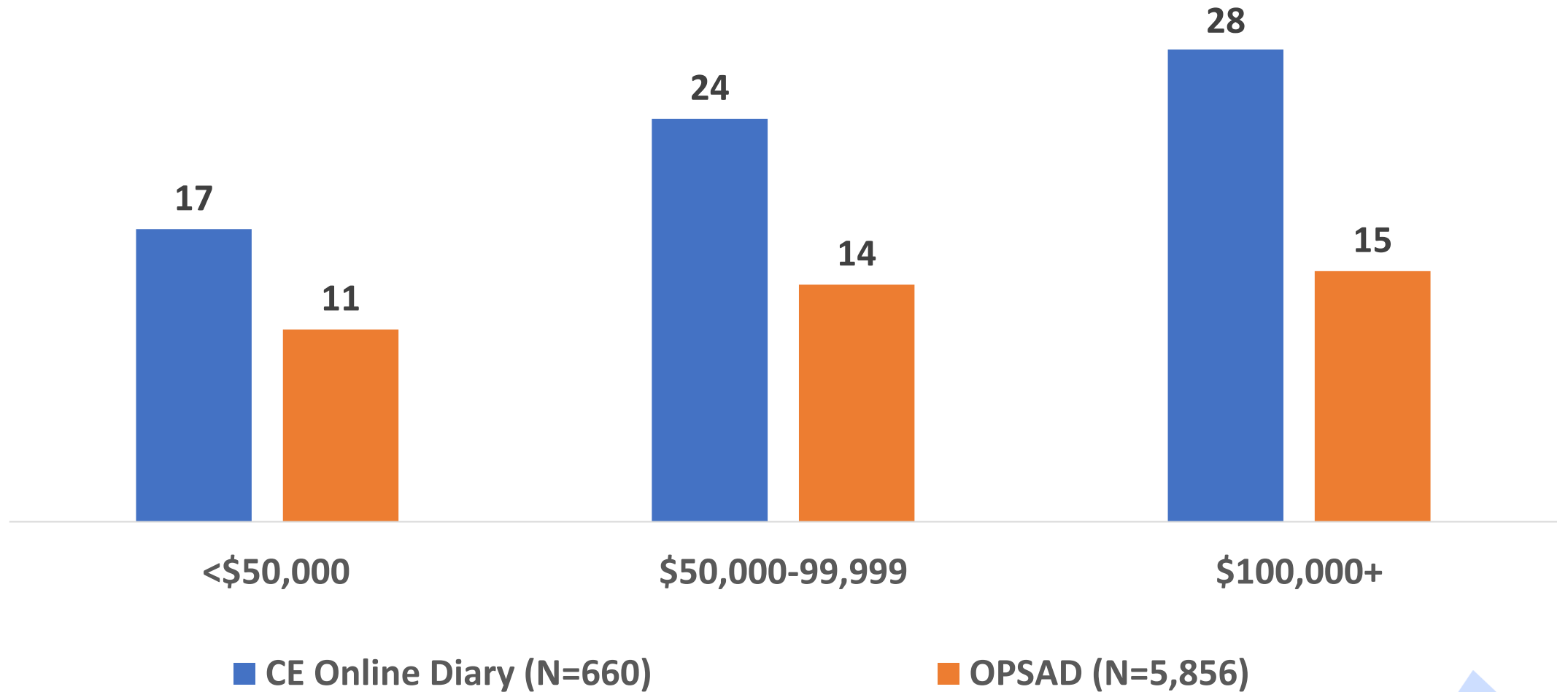
# Demographics and Entry Count Differences

- Compositional differences in the underlying samples
  - ▶ CE Online Diary users have higher levels of education, higher incomes, and are younger
  - ▶ Higher education and income are typically associated with higher entry counts
  - ▶ Are these compositional differences driving differences in entry counts in CE Online Diary vs. OPSAD?
- How do entry counts vary by demographics in CE Online Diary vs. OPSAD?

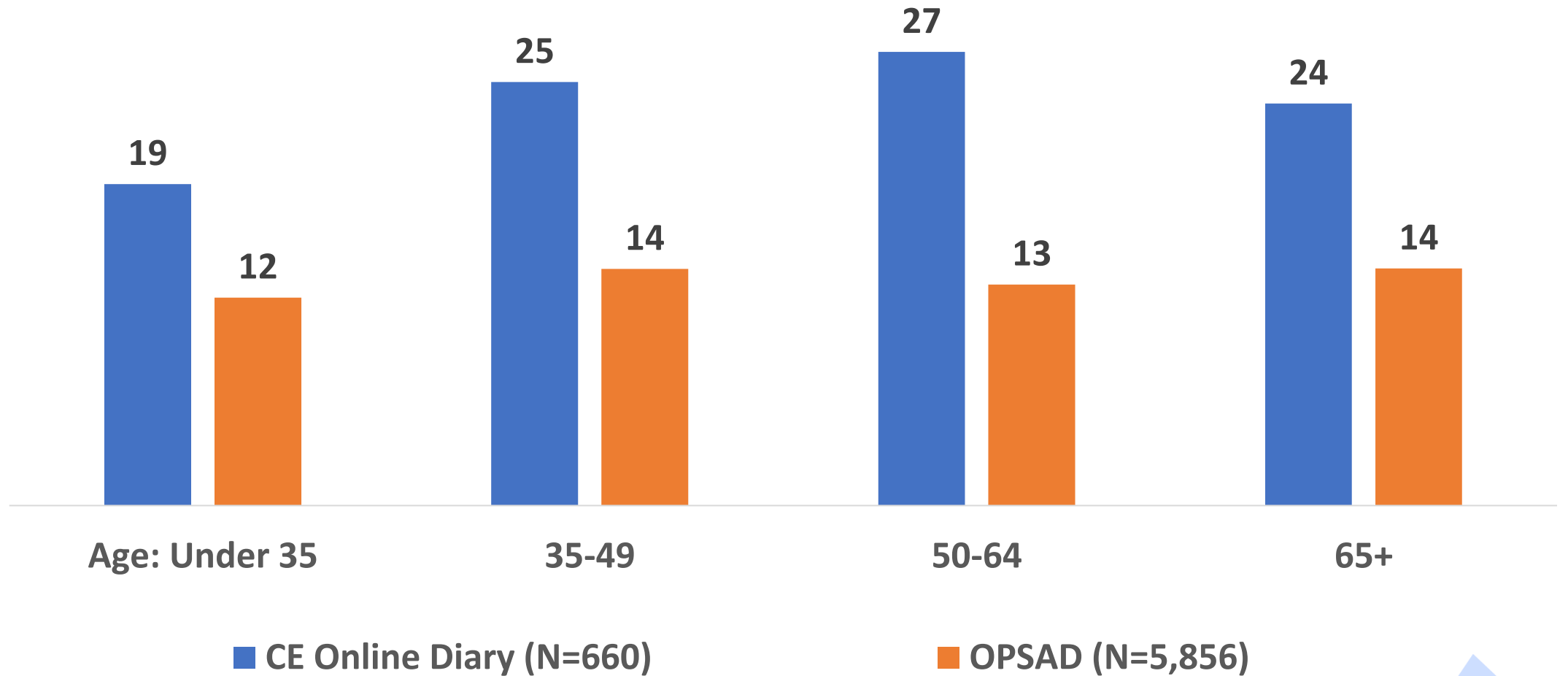
# Average Expenditure Counts – by Education



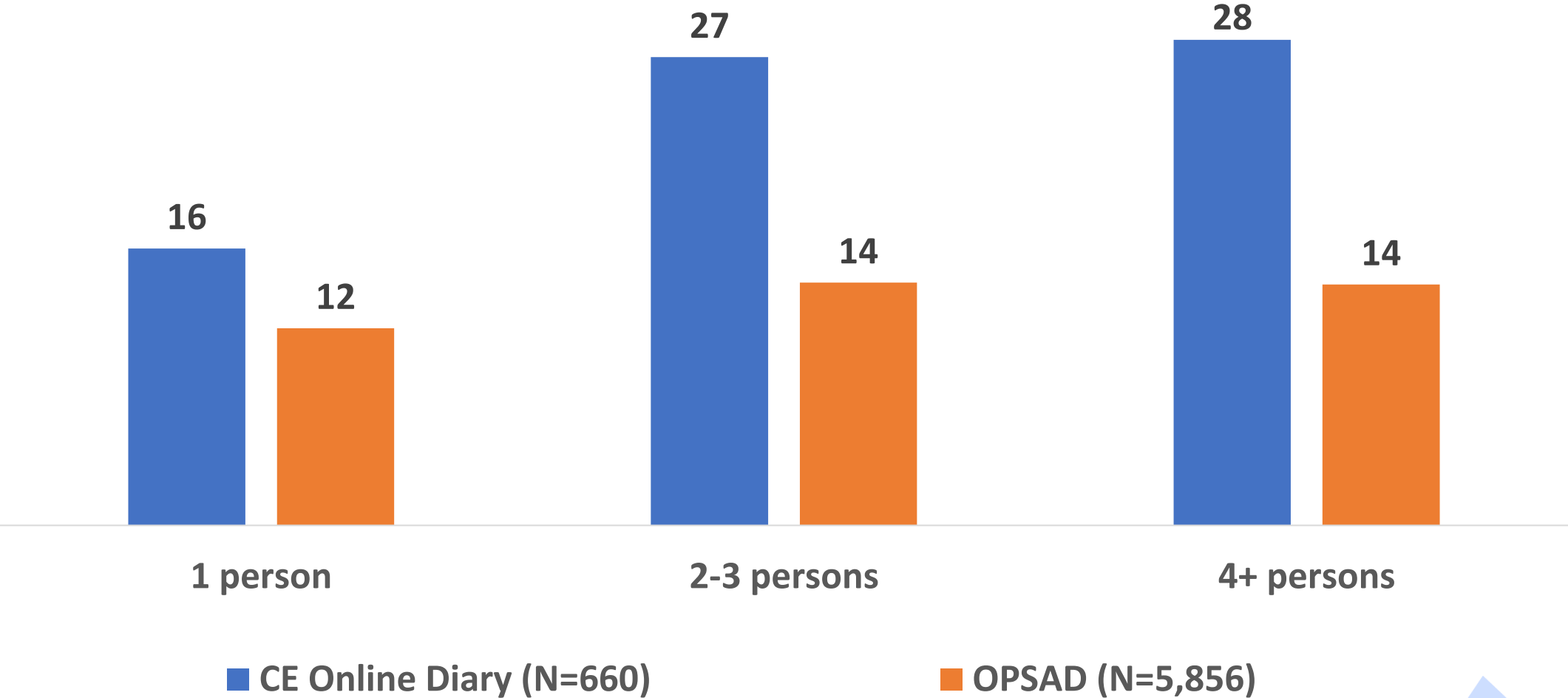
# Average Expenditure Counts – by Household Income



# Average Expenditure Counts – by Age



# Average Expenditure Counts – by Household Size



# Summary

- Lower entry counts in OPSAD are a data quality concern
  - ▶ Gap in counts persists across demographic groups
  - ▶ Attributable to other factors related to study design and lack of interviewer guidance
- In OPSAD, possibly less variation in entry counts by household size
  - ▶ Role of interviewer and respondent communications in emphasizing instructions to record ALL household expenses
- Similar rates of bundling expenses in OPSAD and CE Online Diary
- Lower average expenditure totals in OPSAD
- Higher rates of item nonresponse in cost and item description fields in OPSAD
- Similar rates of drop-off in entries across weeks (among households with entries in both weeks)



# Next Steps for Research

- Expenditure amount totals and predictors
- Further explore relationship between bundling and entry counts
- Paradata analysis – How do user device type, time spent, and number of logins relate to data quality measures?
- Comparisons to full CE production sample
- Is this a platform that we could leverage for future use?



# Contact Information

**Nikki L. Graf**

Branch of Research and Program Development  
Division of Consumer Expenditure Surveys

[Graf.Nikki@bls.gov](mailto:Graf.Nikki@bls.gov)

