# Proxy Survey Cost Indicators: Are they Actually Correlated with Costs?

James Wagner, *University of Michigan's Survey Research Center*Lena Centeno, *Westat*Rick Dulaney, *Westat*Brad Edwards, *Westat*Tuba Suzer-Gurtekin, *UM SRC*Stephanie Coffey, *US Census Bureau* 

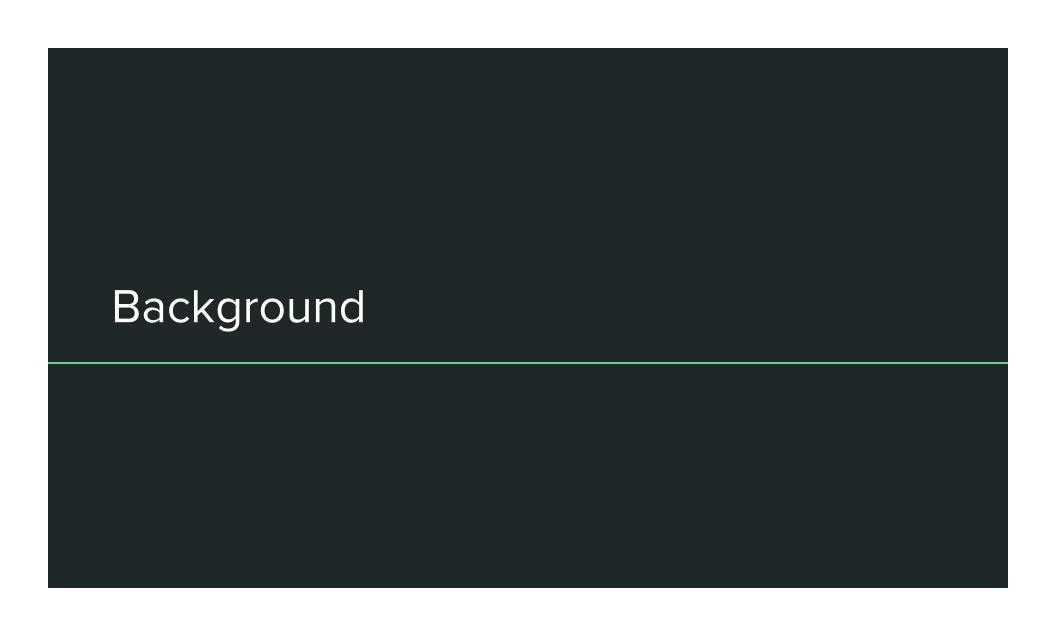
#### Acknowledgement/Disclaimer

The National Survey of Family Growth (NSFG) is conducted by the Centers for Disease Control and Prevention's (CDC's) National Center for Health Statistics (NCHS). The data collection included in this paper was performed under contract #200-2010-33976 with University of Michigan's Institute for Social Research with funding from several agencies of the U.S. Department of Health and Human Services, including CDC/NCHS, the National Institute of Child Health and Human Development (NICHD), the Office of Population Affairs (OPA), and others listed on the NSFG webpage (see <a href="http://www.cdc.gov/nchs/nsfg/">http://www.cdc.gov/nchs/nsfg/</a>). The views expressed here do not represent those of NCHS or the other funding agencies.

Any opinions and conclusions expressed herein are those of the author and do not necessarily reflect the views of the U.S. Census Bureau. The Census Bureau has reviewed this data product (for SIPP data) for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release (Approval ID: CBDRB-FY22-CES014-026)

## Overview

- Background
- Methods and Data
- Results
- Conclusion

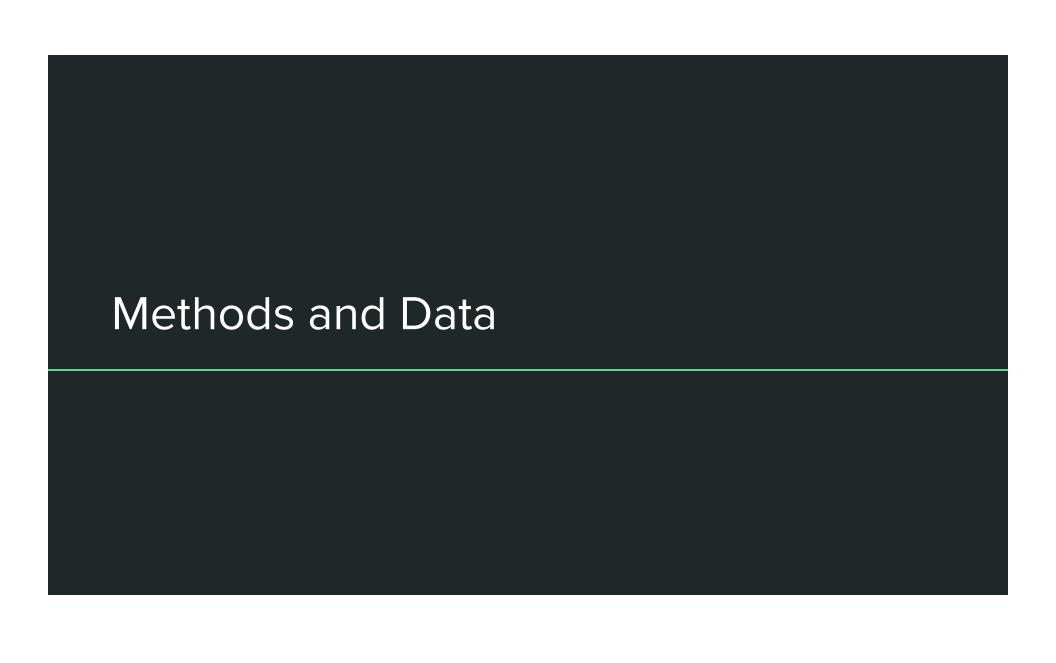


#### Survey Costs are Understudied

- Total Survey Error balances tradeoffs between errors and costs
- Survey methodologists have studied errors
  - Nonresponse, measurement, sampling error, etc.
- Costs are much less studied
  - An important study: Groves, <u>Survey Errors and Survey Costs</u> (1989)
  - A recent review: Olson, Wagner, and Anderson (2021)

#### Survey Cost Measurement

- Cost monitoring can lag behind costs
  - Timesheets may be completed bi-weekly or monthly
  - Expense reports may take time to be processed
  - Other issues...
- Proxy cost indicators can be more timely
  - For example, attempts per interview
  - Drawn from paradata, relatively easy to calculate
- But are these indicators accurate?



#### Correlation of Proxy Indicators with Survey Costs

- Examine 10 different proxy indicators of costs
- Use two different cost measures:
  - 1. Total costs
  - 2. Interviewer Hours
- Each project defines "total costs"
- Interviewer hours are a major component of cost
  - Equivalent across studies

## **Proxy Indicators**

INDICATOR	DESCRIPTION
Total Attempts	A count of the total number of attempts. This includes mail, email, SMS text, telephone and face-to-face attempts.
Attempts Per Interview	Total Attempts divided by Total Completes.
Attempts per Sampled Unit	Total Attempts divided by the Sample Size.
Total Contacts	The count of attempts that have contact with a household member.
Total Completes	Total count of completed interviews across all modes. "Complete" is defined separately for each survey.
Sample Size	The total sample size. This could be a sample of person or households, depending upon the study.
Total Miles	The total number of miles driven by interviewers in their personal vehicles. These miles exclude mileage in rental vehicles. Interviewers using public transportation also do not report mileage.
Mean Number of Attempts to First Contact	For each sampled unit. The attempt at which the first contact was made was recorded. Among those cases with a first contact, what was the mean number of attempts including the attempt when first contact was made.
Minutes of Interviewing	The total number of minutes spent completing a survey, including all modes except paper/mail surveys.
Hours Per Interview	The total hours interviewers have worked divided by Total Completes. Hours spent in training are not included. Used as a proxy for total costs only since it includes total interviewer hours in its calculation.

#### National Survey of Family Growth 2011-2019

- National Survey of Family Growth (NSFG) is study of family formation and fertility
  - Data are collected via face-to-face interviewing
  - New samples each quarter
- Data from Q1 to Q27 included
- Total cost definition:
  - Does not include central office staff
  - Includes interviewers and field supervisors
  - Travel expenses, token of appreciation
  - Excludes interviewer time spent on sampling-related activities

#### Medical Expenditure Panel Survey 2019

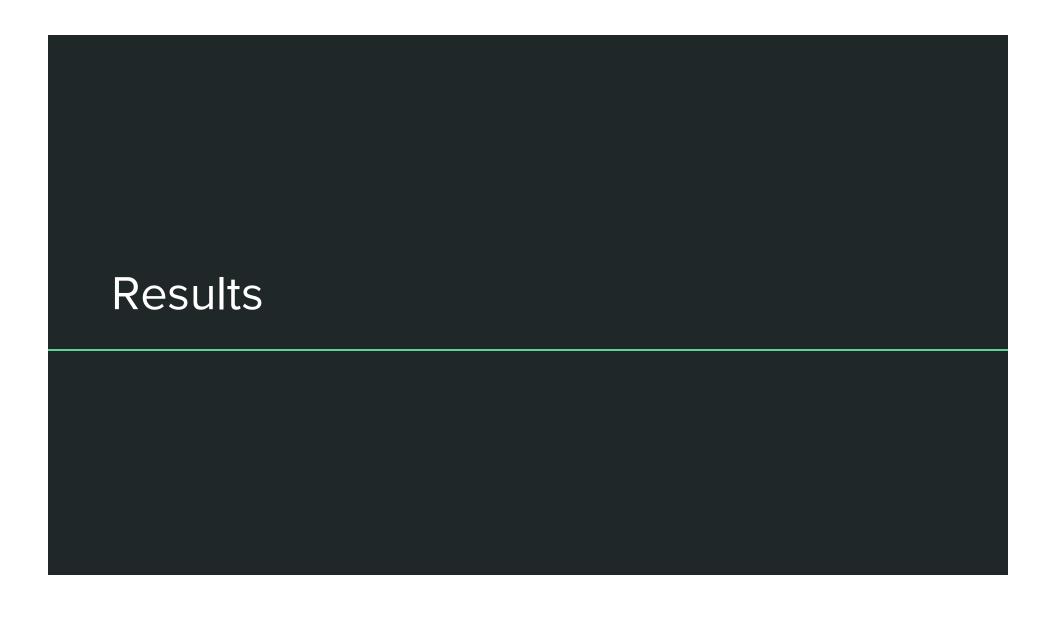
- Medical Expenditure Panel Survey (MEPS) is a primary source of annual data on health care utilization and costs
  - Data are collected via face-to-face interviewing, in 5 interviews over 2.5 years
  - New samples each January; overlapping panel design, so 3 panels in play for all interviewers in first half of every year
- Data from January to July 2019 included
- Total cost definition:
  - Does <u>not</u> include central office and field management staff
  - Includes interviewers only
  - Wages and expenses only, and only for work on newest panel
  - Includes all interviewer time post-initial training

### Surveys of Consumers (SCA)

- Surveys of Consumers (SCA) collect data on the consumer expectations that are strong predictors of business cycle in the U.S.
  - Monthly cell phone RDD survey
  - Rotating panel design
    - Each month's data are from 50% fresh sample and 50% reinterviews
  - No incentives offered
  - A typical survey period is 25-27 days, while there are three longer periods (33-34 days) during the year.
  - Noncontacts followed up to three and ten times in the fresh and reinterview samples
  - Average interview length is 29 minutes per complete across all samples.
  - Data from January 2015 to December 2020 included
- Total cost definition:
  - Includes supervision and interviewers

### Survey of Income and Program Participation (SIPP)

- SIPP is a primary source for information on eligibility for and participation in government assistance programs along with workforce participation, demographic, income information
  - Annual data collection operations, primarily via face-to-face
  - Data collection period lasts four to five months, February May or June
  - Panel design with total annual samples of approximately 53,000
    - 2014 Panel: Single sample selected in 2014, interviewed annually through 2017
    - 2018 Panel: Initial sample selected in 2018, refreshment samples annually thereafter to maintain sample
  - Data from 2014 2020 included
    - 2014 2017 complete panel
    - 2018 2020 ongoing panel
- Total cost definition:
  - Interviewer Total Costs
    - Includes mileage, contact attempts, interviewing, attending training, etc.
    - Excludes Regional Office activities, Census Bureau headquarters costs, overhead costs



## Correlations: Total Costs at the Survey Level

		Proxy Indicator												
	Total	Attempts	Attempts	Total	Total	Sample	Total	Mean # of	Minutes	Hours Per				
	Attem-	Per	per	Con-	Com-	Size	Miles	Attempts to	of	Interview				
	pts	Interview	Sampled	tacts	pletes			First	Interview					
Survey			Line					Contact	-ing					
NSFG	0.496	0.266	0.409	0.370	0.309	0.319	0.203	0.156	0.218	0.509				
(n=27														
quarters														
SIPP-A	0.668	0.403	0.041	0.757	0.509	0.741	0.774	0.206	0.838	0.684				
(n=7 years)	0.000	0.403	0.041	0.757	0.509	0.741	0.774	0.200	0.030	0.004				
SIPP-B	0.570	-0.154	0.544	0.610	0.835	0.340	0.386	0.871	0.608	-0.468				
(n=7 years)	0.570	-0.134	0.544	0.010	0.000	0.540	0.500	0.07 1	0.000	-0.400				
SCA (n=72	0.529	0.341	-0.361	0.497	0.817	0.826	NA		0.222	0.735				
months)														

## Correlations: Total Costs at the Survey Level

	Ļ									
					Prox	y Indicator				
	Total	Attempts	Attempts	Total	Total	Sample	otal	Mean # of	Minutes	Hours Per
	Attem-	Per	per	Con-	Com-	Size	l /liles	Attempts to	of	Interview
	pts	Interview	Sampled	tacts	pletes			First	Interview	
Survey			Line					Contact	-ing	
NSFG	0.496	0.266	0.409	0.370	0.309	0.319	.203	0.156	0.218	0.509
(n=27										
quarters										
SIPP-A	0.668	0.403	0.041	0.757	0.509	0.741	).774	0.206	0.838	0.684
(n=7 years)	0.000	0.403	0.041	0.757	0.509	0.741	0.774	0.200	0.030	0.004
SIPP-B	0.570	-0.154	0.544	0.610	0.835	0.340	0.386	0.871	0.608	-0.468
(n=7 years)	0.570	-0.154	0.544	0.010	0.000	0.340	0.300	0.67 1	0.000	-0.400
SCA (n=72	0.529	0.341	-0.361	0.497	0.817	0.826	I <mark>I</mark> A		0.222	0.735
months)										

## Correlations: Total Interviewer Hours at the Survey Level

	Proxy Indicator											
	Total	Attempts	Attempts	Total	Total	Sample	Total	Mean # of	Minutes of			
Survey	Attempts	Per Interview	per Sampled Line	Contacts	Com- pletes	Size	Miles	Attempts to First Contact	Interviewing			
NSFG	0.857	0.183	0.566	0.861	0.808	0.699	0.713	0.141	0.705			
SIPP	0.895	0.433	0.314	0.960	0.779	0.867	0.727	0.448	0.967			
SCA	0.815	0.666	-0.071	0.719	0.786	0.929	NA		0.232			

## Correlations: Total Interviewer Hours at the Survey Level

					Proxy Indica	tor			
		ttempts Fer Interview	Attempts per Sampled Line	Total Contacts	Total Com- pletes	Sample Size	Total Miles	Mean # of Attempts to First Contact	Minutes of Interviewing
Survey									
NSFG	0.857	.183	0.566	0.861	0.808	0.699	0.713	0.141	0.705
SIPP	0.895	0.433	0.314	0.960	0.779	0.867	0.727	0.448	0.967
SCA	0.815	.666	-0.071	0.719	0.786	0.929	NA		0.232

## Correlations: Total Costs at the Component Level

	Proxy Indicator											
	Total	Attemp	Attempt	Total	Total	Sample	Total	Mean # of	Minutes of	Hours	Refusals	
	Attem	ts Per	s per	Con-	Com-	Size	Miles	Attempts	Interview-	Per Inter-		
	pts	Inter-	Sample	tacts	pletes			to First	ing	view		
		view	d Line					Contact				
Survey												
MEPSP24R1 (Component=Iwers)		-0.271	-0.055	0.0498	0.229		0.751	0.053	0.469	0.387	-0.217	

## Correlations: Total Interviewer Hours at the Component Level

	Proxy Indicator											
	Total	Attemp	Attempt	Total	Total	Sample	Total	Mean # of	Minutes of	Refusals		
	Attem	ts Per	s per	Con-	Com-	Size	Miles	Attempts	Interview-			
	pts	Inter-	Sample	tacts	pletes			to First	ing			
		view	d Line					Contact				
Survey												
MEPSP24R1 (Component=Iwers)		-0.305	-0.055	0.033	0.247		0.782	0.069	0.504	-0.265		

#### Conclusion

- Proxy cost indicators provide useful summaries
- Often highly correlated with actual cost measures (but not always!)
- Understanding these correlations should help with interpreting proxy indicators
- Post hoc analyses of costs are helpful and fun!

# Thanks!

jameswag@umich.edu



#### References

- Groves, R. M. (1989). <u>Survey Errors and Survey Costs</u>. New York, N.Y., Wiley.
- Olson, K., J. Wagner and R. Anderson (2021). "Survey Costs: Where Are We and What Is the Way Forward?" <u>Journal of Survey Statistics and Methodology</u> **9(5): 921-942.**