

Leveraging External Data Sources to Improve Federal Government Surveys

Gizem Korkmaz, Ph.D.
Associate Vice President
Statistics and Data Science | Westat
gizemkorkmaz@westat.com

**Joint Statistical Meetings | Portland, Oregon
August 6, 2024**



Introduction

- Statistical agencies are facing challenges of data collection through surveys
- Alternative data to enhance, to be used in conjunction with, or to replace surveys
- The Evidence-Based Policymaking Act of 2018, and the CHIPS and Science Act of 2022
- World of big data and AI-driven technologies

Toward a 21st Century
National Data Infrastructure:
Enhancing Survey Programs by
Using Multiple Data Sources



Stephanie Zimmer: Age-Eligibility Oversampling to Reduce Screening Costs in a Multimode Survey (NSFG)

- Goal: to minimize the number of households with no residents age 15-49 for the National Survey of Family Growth (NSFG)
- Data:
 - Enhanced address-based sampling:
 - USPS Computerized Delivery Sequence (CDS)
 - Consumer marketing data
 - Auxiliary data: Decennial Census, ACS, CDC PLACES, and more
 - Ground-truth: Data from a prior survey
- Methods:
 - Logistic regression and variable selection
 - Propensity scores

Jay Clark: Enhancing Weighting in NHANES with External Data

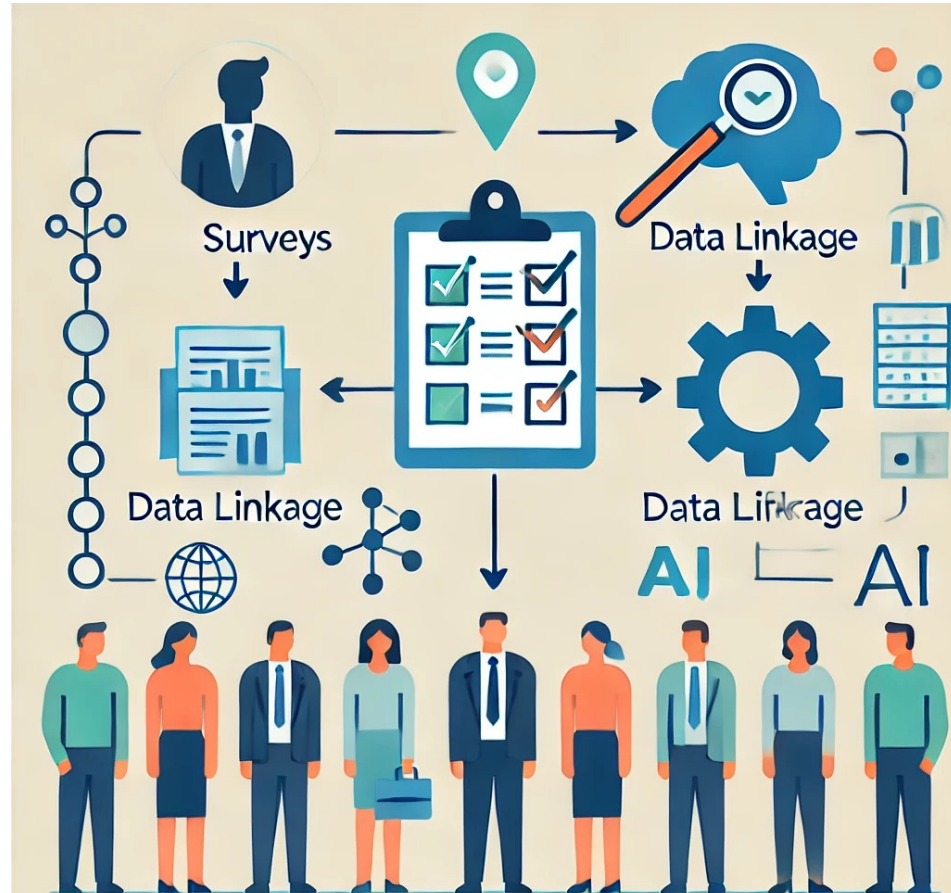
- Goal: to improve the weighting adjustment process to reduce bias from nonrespondents for NHANES
- Data:
 - American Community Survey (ACS)
 - PLACES
 - Interviewer observations
- Methods:
 - Nonresponse adjustments: screener, interview, and exam
 - Calibration: income decile

Wan-Ying Chang: Improving Survey Efficiency with Linked Data: The Survey of Doctorate Recipients Story

- Goal: to assess the feasibility of using linked data to supplement or replace parts of the survey
- Data:
 - Scientific publications
 - Federal awards and other research funding
 - Patents
- Methods:
 - Data linkage
 - Evaluation of the quality of self-reported data on scientific productivity and receipt of federal support
 - Assessment of the feasibility and limitations of using linked data

Questions

- What additional non-survey data could be useful for these studies to enhance and/or replace the surveys?
 - Housing data
 - Resumes, ORCID, etc.
- What is the process to identify and evaluate external data?
- How/whether can you use AI to improve the process? What are concerns?





Thank you!

Gizem Korkmaz, Ph.D.
Associate Vice President
Statistics and Data Science, Westat
gizemkorkmaz@westat.com

