



# SURVEY FUTURES

SURVEY DATA COLLECTION  
METHODS COLLABORATION

## Evaluating the utility of linked administrative data for non-response bias adjustment in a piggyback longitudinal survey

Joseph Sakshaug (IAB & LMU-Munich)

Tobias Büttner (German Federal Employment Agency)



University of Essex



University of  
Southampton



Economic  
and Social  
Research Council

NCRM NATIONAL CENTRE FOR  
RESEARCH METHODS

Office for  
National Statistics

UCL

WARWICK  
THE UNIVERSITY OF WARWICK

MANCHESTER  
The University of Manchester

CITY  
UNIVERSITY OF LONDON  
EST 1894

National Centre  
for Social Research

LSE  
THE LONDON SCHOOL  
OF ECONOMICS AND  
POLITICAL SCIENCE

Ulster  
University

Unil  
UNIL | Université de Lausanne

Ipsos

KANTAR  
PUBLIC

# Introduction

- Unit nonresponse is a key component of TSE framework
- Analytic value of panel surveys grows over time, but cumulative attrition threatens representativeness
- Traditional auxiliary data (e.g. paradata, prior-wave data) have limitations
- Proposal: Use linked administrative data in “piggyback” panel designs
  - Respondents from a preceding cross-sectional survey (which included a data linkage request) are recruited for a subsequent longitudinal survey
- Goal: Utilize linked administrative data to evaluate and adjust for panel nonresponse bias

# Examples of Piggyback Panel Surveys

- US Medical Expenditure Panel Survey-Household Component
  - Subsampled from participants of the US National Health Interview Survey
- English Longitudinal Study of Ageing (ELSA)
  - Sampled from participants of the Health Surveys for England
- GESIS Panel
  - Sampled from respondents of the German General Social Survey (ALLBUS)

# Background

- Cumulative attrition increases significantly in early waves
  - influenced by location, contact, cooperation, and life events
- Limitations of traditional auxiliary data sources
  - Paradata (weak correlations with survey variables)
  - Commercial data (outdated, questionable quality)
  - Previous-wave survey data (misses between-wave events)
    - E.g. losing one's job in-between waves may affect likelihood of subsequent wave participation
- Advantages of linked administrative data
  - Longitudinal, substantive, captures between-wave events, but requires consent
- Few studies investigate effects of between-wave events on attrition
  - e.g. Trappmann et al. 2015



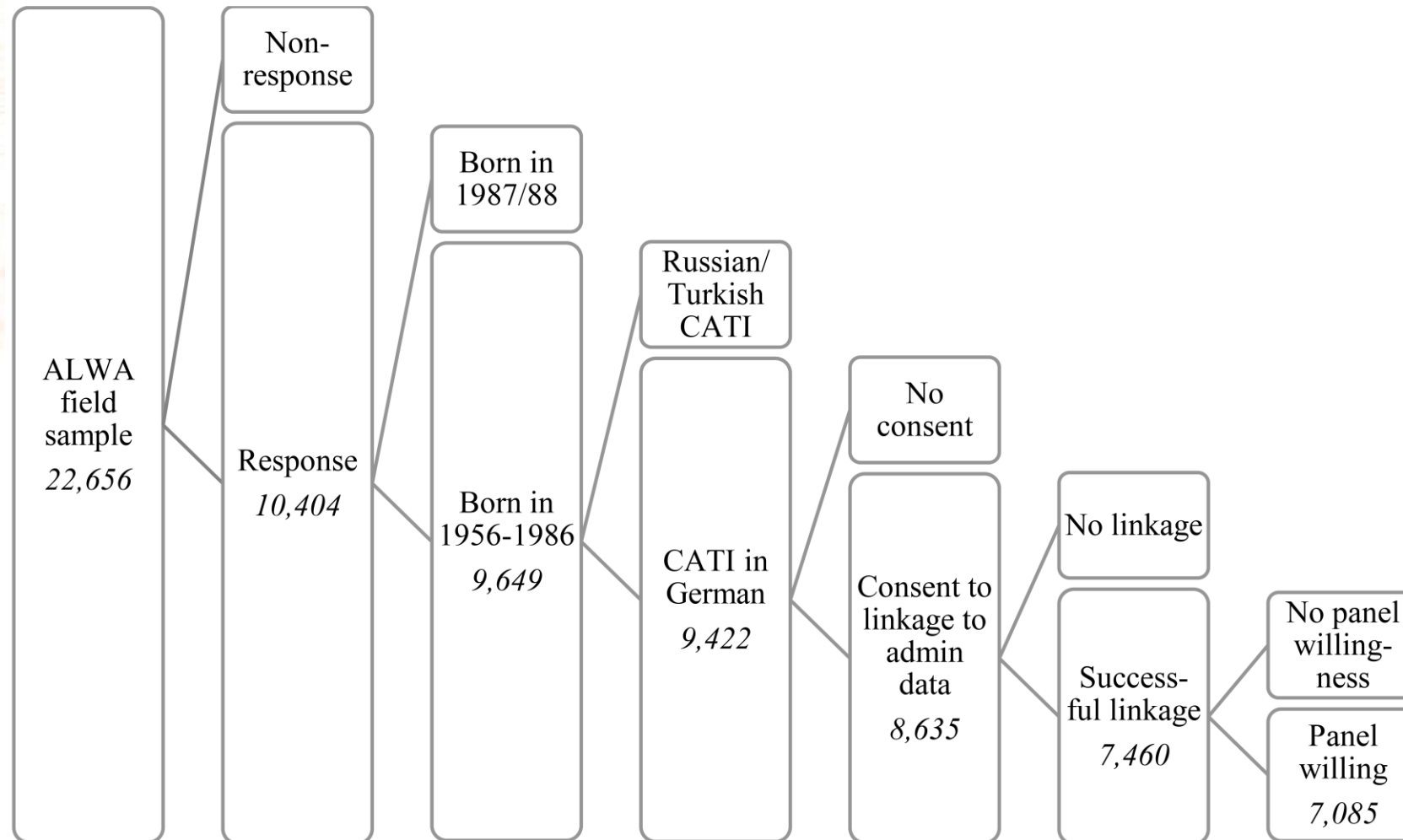
# Research Questions

- RQ1: Do admin variables correlate with response outcomes and substantive survey variables?
- RQ2: Does the inclusion of admin variables in attrition models improve model fit?
- RQ3: Does the inclusion of admin variables in an enhanced weighting scheme reduce attrition bias, compared to a standard weighting scheme?

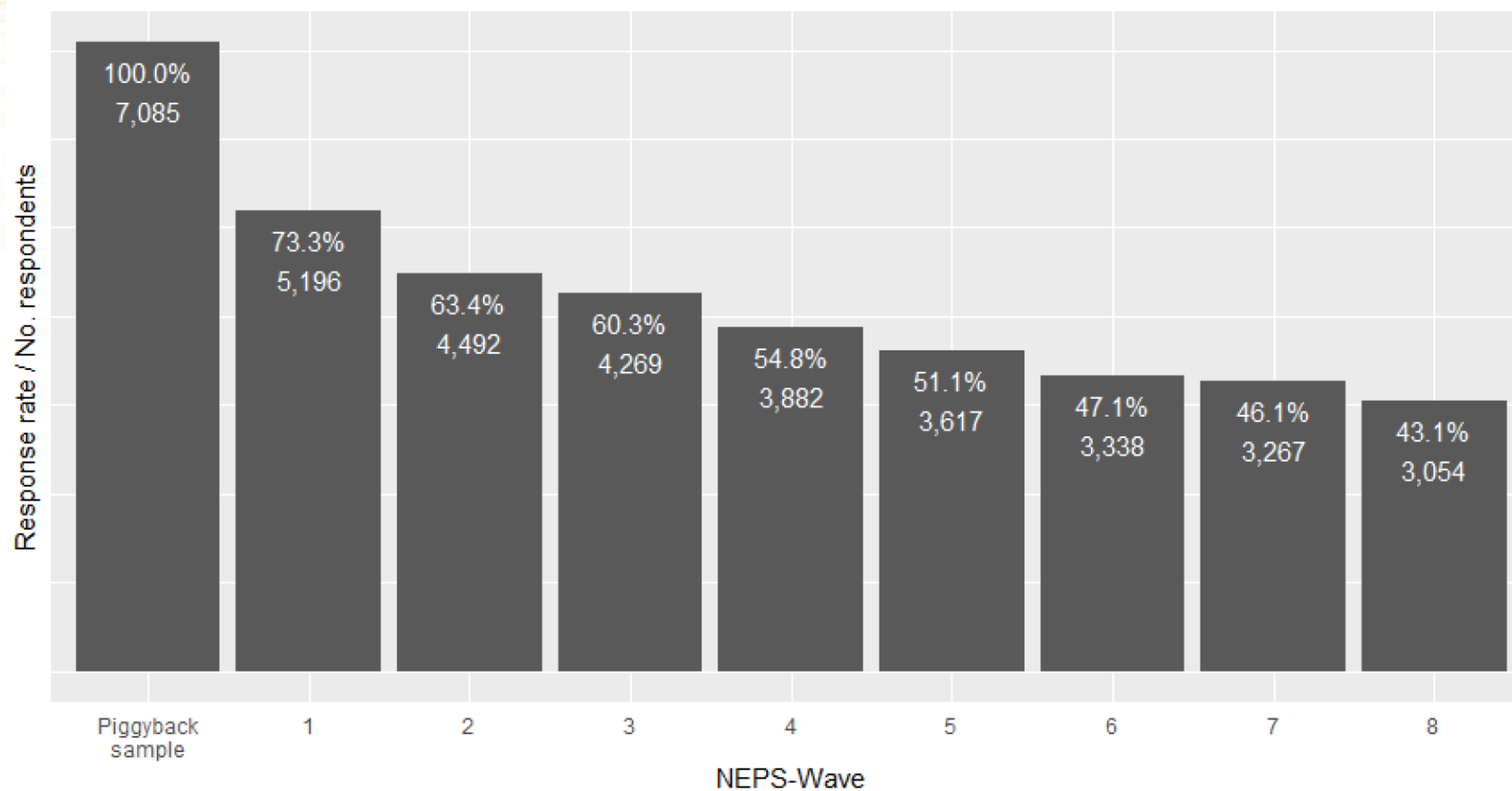
# Survey Data

- National Educational Panel Study (NEPS) – Adult Cohort (SC6)
  - Multicohort panel study on education/employment in Germany; 2009-2017
  - Mixed-mode data collection (telephone and face-to-face)
- NEPS subsample recruited from a prior cross-sectional survey
  - “Working and Learning in a Changing World” (ALWA); 2007-2008
  - Probability-based sample drawn from selected municipality records
- Respondents in cross-sectional survey (ALWA) asked for consent to follow-up (i.e. panel willingness) and data linkage consent

# ALWA Sample → Piggyback Sample



# Cum. Response Rates, NEPS Waves 1-8





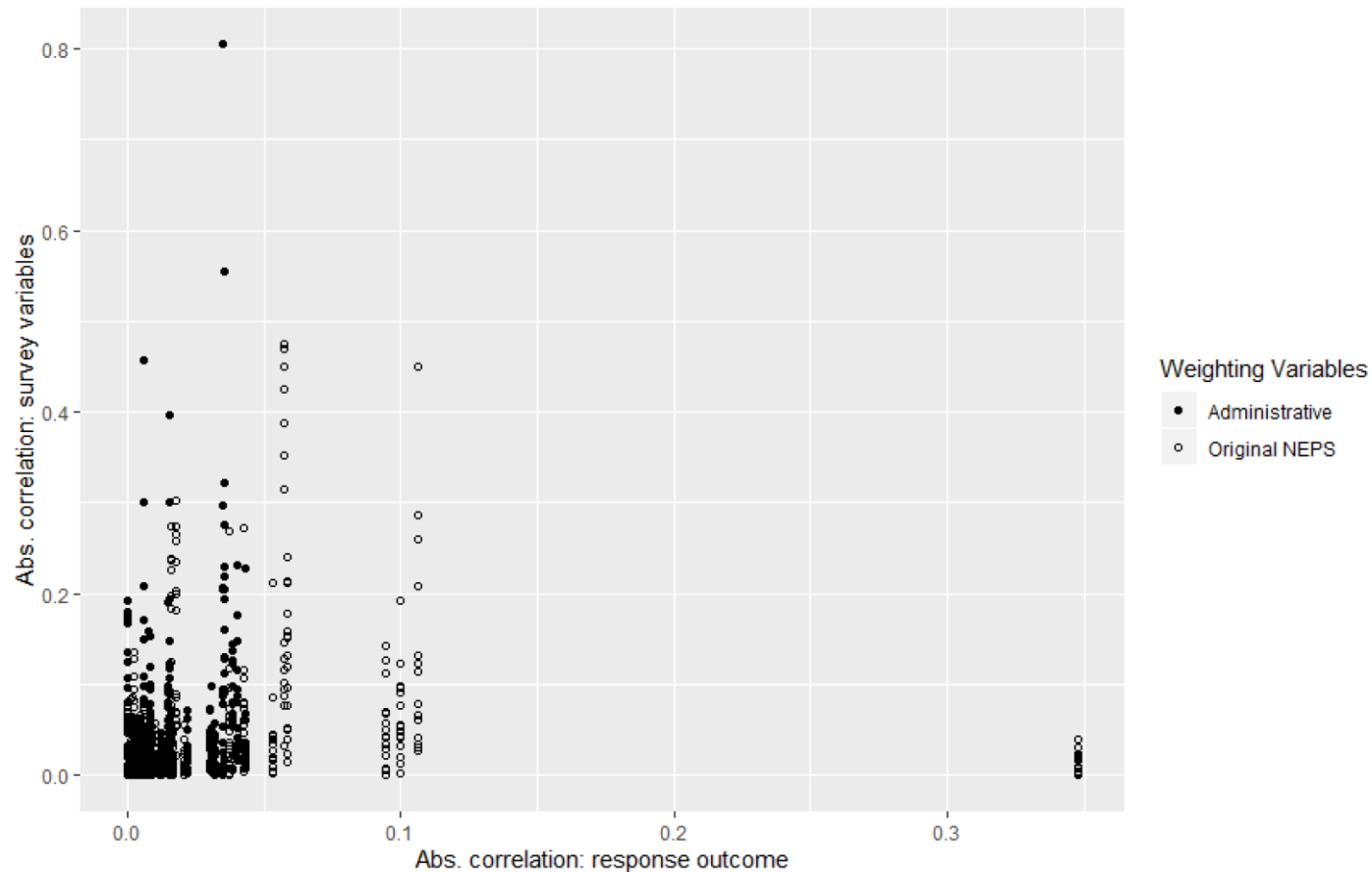
# Administrative Data

- Integrated Employment Biographies (IEB) database
  - Compiled from compulsory notifications by employers to the German Federal Employment Agency
- Contains information on employment spells, wages, working time, benefit receipt, participation in active labor market programs, job search, and personal characteristics
- Respondent consent required for survey data linkage

# Methods

- Original NEPS weighting scheme
  - 8 variables; **paradata** (# contact attempts) and **socio-economic variables** (e.g. federal state, municipality codes, sex, marital status, mother tongue, household size, income, educational attainment)
- Enhanced weighting scheme
  - ...+ 16 **IEB admin variables** (e.g. current employment status, type of employment contract, average daily wages, geodata, welfare benefit receipt, and total number of employment and benefit spells in last 5 years, etc.), including **indicators of status changes in-between waves**
- Fit logistic regression models for 1) linkage consent, 2) successful linkage, 3) panel willingness, and 4) wave-specific response
  - Inverse-Propensity Score Weighting
- Evaluation: Correlations, model fit, absolute bias, differences in survey estimates

# RQ1 Correlations with Response/Survey Variables



- Correlations are generally low for both admin & NEPS weighting variables and response outcome (below ~0.1)
- Most correlations with the survey variables are generally small, with some exceptions.
  - NEPS weighting variables (e.g. education, demographics) are correlated with competency measures
  - IEB admin variables (e.g. current wages, employment status/changes) have moderately high correlations with NEPS employment outcomes

# RQ2: Model Fit Statistics

	Model fit statistics	Panel wil-lingness	NEPS-Wave							
			1	2	3	4	5	6	7	8
Original NEPS	Pseudo R <sup>2</sup>	0.03	0.07	0.14	0.12	0.12	0.11	0.14	0.12	0.10
	AIC	2974.4	7714.5	4349.9	3132.7	3282.3	3760.9	3612.3	2940.1	2541.6
Original NEPS + administrative data	Pseudo R <sup>2</sup>	0.03	0.08	0.14	0.14	0.13	0.12	0.15	0.13	0.11
	AIC	2984.2	7718.4	4362.70	3119.3	3292.7	3756.6	3627.7	2943.4	2558.6

- Adding the administrative variables only slightly increases explanatory power to the already low fit of the original panel willingness and NEPS response models

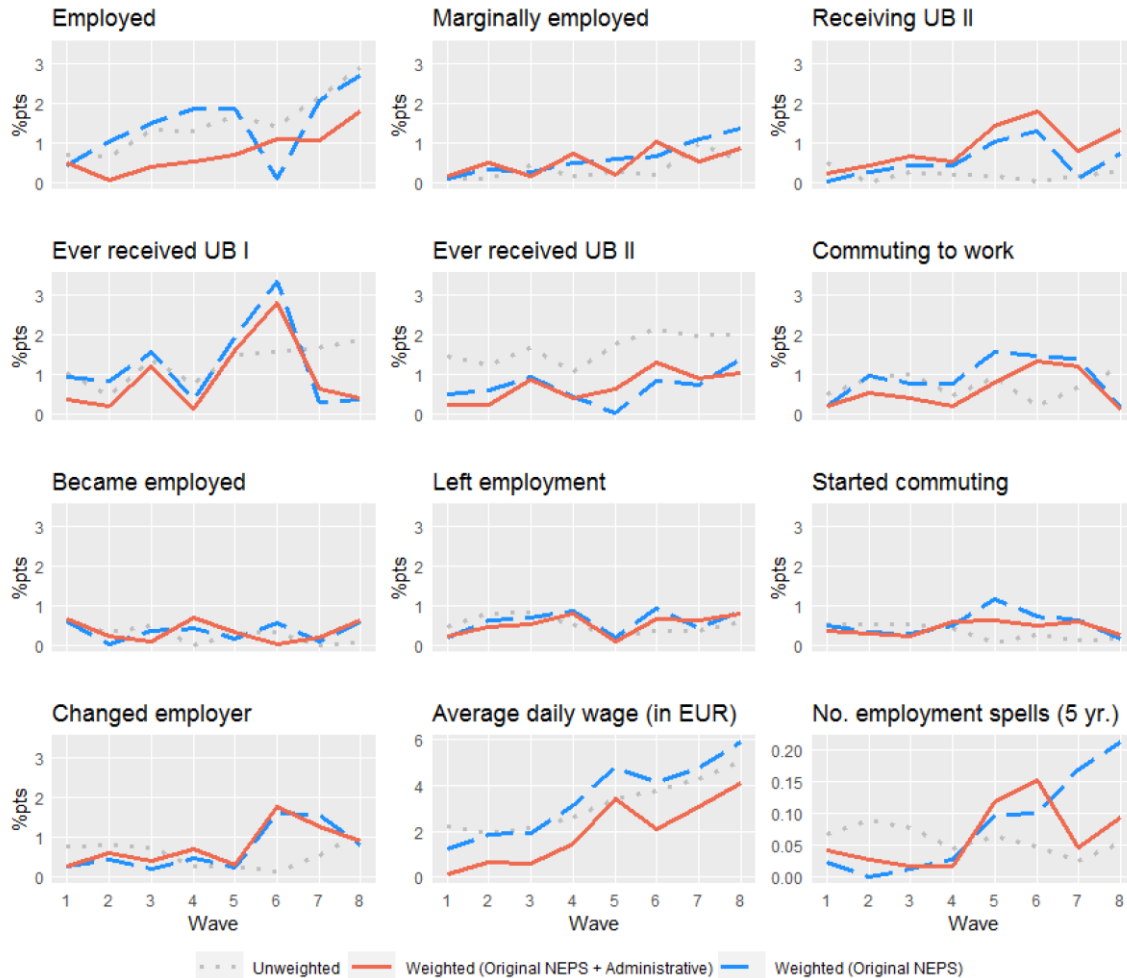


# Significant Predictors of Response in Waves 1-8

Administrative variable	NEPS-Wave							
	1	2	3	4	5	6	7	8
Average daily wage (in Euros)	+				-			
Working hours								
Not employed	REF	REF	REF	REF	REF	REF	REF	REF
Part-time					+		+	
Full-time			+		+			
Receiving Unemployment Benefit II					+			
Commuting to work (5-digit municipality code)		+						
Number of employment spells last 5 years				+				
Number of UBI spells last 5 years			+	-				
Ever received UBII in lifetime					-			
<i>Variables of change since t-1</i>								
Became UBI recipient							+	
Became employed			-					
Left employment			-	-	-			
Stopped receiving UBII			-					
Different employer compared to last wave					+			-

- Some admin variables are significant predictors in multiple waves
  - Being employed at the time of interview attempt is positively related to response
  - Becoming unemployed since the previous interview negatively related to response
- Sometimes direction of effect reverses
  - Avg. daily wage positively related to W1 response, negatively related in W5

# RQ3: Bias Reduction (Wtd. Admin Estimates)



- Results show some large NR biases for later waves, but little difference between both weighting schemes (standard vs. enhanced)
- Few exceptions where enhanced weighting scheme outperforms standard scheme
  - employed
  - no. employment spells
  - average daily wage
- Also identified small reductions in coefficients of variation (CVs) for later waves

# Impact on Weighted Survey Estimates

- Across 16 survey estimates, no significant differences between weighting procedures
- Survey estimate of employment status shifts closer to the corresponding admin estimate, under enhanced weighting scheme
- Reductions in CVs also seen in later waves

# Discussion

- Evaluated whether linked-admin data useful in piggyback designs for reducing attrition bias
- Correlations between admin and survey variables low-to-moderate
  - Higher correlations for construct-similar variables
- Employment administrative variables significant predictors of wave-response, but low model fit
  - e.g. became unemployed since previous interview
- Enhanced weighting scheme improved attrition bias for admin estimates
  - However, only minor effects on point estimates of substantive survey estimates
  - Some evidence of bias reduction for reported employment status
- Practical implications
  - Piggyback design with linked data is worth considering, esp. if data already linked from preceding survey
  - Augmented adjustment weights can be useful for bias adjustment without increasing sampling variance
  - We also envision the strategy to help with monitoring nonresponse over time
- Limitations: Linkage failures (e.g. non-consent), accounting for multiple stages of selection



# Reference

Büttner, T.J.M., Sakshaug, J.W., and Vicari, B. (2021). Evaluating the Utility of Linked Administrative Data for Nonresponse Bias Adjustment in a Piggyback Longitudinal Survey. *Journal of Official Statistics*, 37(4), 837-864. <https://doi.org/10.2478/jos-2021-0037>