

Non-Consent Error, Nonresponse Error, and  
Measurement Error:

*Total Survey Error in Linked Survey and  
Administrative Data*

Joe Sakshaug

Program in Survey Methodology

Institute for Social Research

University of Michigan

June 22, 2011

# Acknowledgements

- Joint work with Frauke Kreuter, JPSM/IAB
- Data obtained from Institute for Employment Research (IAB)
- IAB visit July, 2010; Nuremberg, Germany
- IAB/Ann Arbor Research Data Center
  - Mark Trappmann
  - Stefan Bender
  - Joerg Heining

# Outline

- Advantages of administrative data linkage
- Selectivity of consent
- Research questions
- Data sources
- Results
- Conclusions
- Planned extensions

# Advantages of Data Linkage

## Substantive research

- Obtain population-based inferences for key survey and administrative variables of interest
- Address complex policy-oriented research questions
  - e.g. health reform, federal assistance programs

## Survey research

- Reduction in respondent burden
- Reduction in data collection costs
- Assessment of data quality

# Informed Consent

- Many surveys require respondent consent to link survey and administrative records
  - HRS, PSID, BHPS, MCS, ELSA, NIS, NHIS, etc...
- Informed consent often requires a signature and unique ID number (Medicare, social security)
- Respondent consent is not universal
  - Range: 19.0% - 96.5% (McCarthy et al., 1999; Rhoades and Fung, 2004)
- Concern: non-consent bias

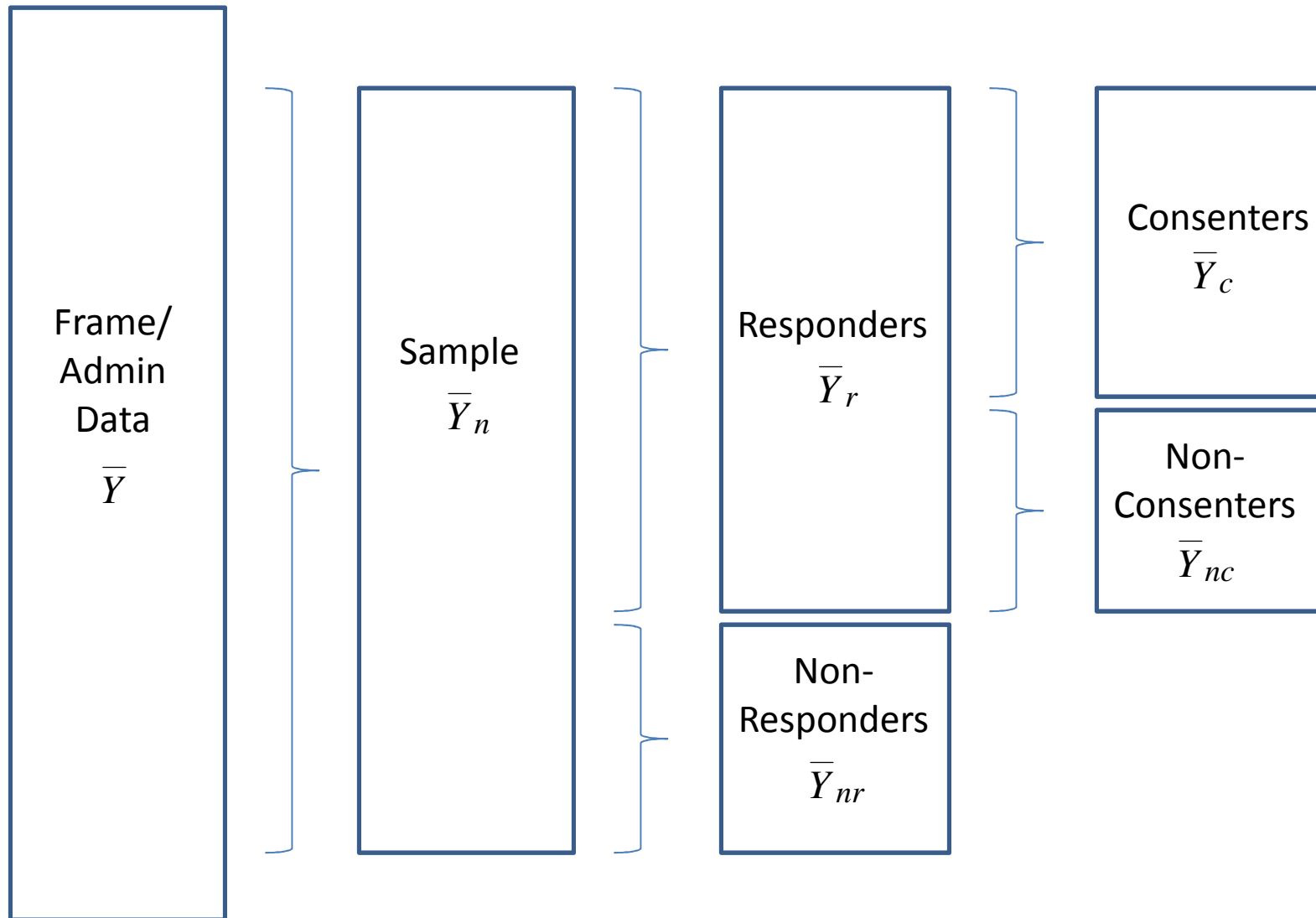
# The Selectivity of Consent

- Common correlates of consent
  - Age, race/ethnicity, gender, education, marital status, health status, employment (Bates and Pascale, 2006; Jenkins et al., 2006; Banks et al., 2005; Dunn et al., 2003; Young et al., 2001; Woolf et al, 2000; Olson, 1999; Pullen et al., 1992)
  - Item missing data, interviewer characteristics, prior-wave outcomes (Sakshaug et al., 2010; Jenkins et al., 2006)
- Unclear whether the selectivity of consent introduces significant biases on **key administrative estimates**

# Relationship between Consent Rates and Bias

- Factors that influence the likelihood of consent
  - Privacy/confidentiality (Bates, 2005)
  - Distrust of survey sponsor/government (Gray, 2009)
  - Interviewer-respondent interaction (Sala, Burton, and Knies, 2010)
  - Motivation, fewer survey questions (anecdotal)
- Are same factors related to non-consent biases?
  - Unclear
  - No study has estimated non-consent biases for administrative outcomes
    - Reason: admin records typically unavailable for non-consenting cases

# Conceptual Pathway to Consent





# Research Questions

1. Do non-consent biases exist?

$$\bar{y}_c - \bar{y}_r \neq 0$$

2. How does non-consent bias compare to other sources of bias?

– nonresponse, measurement error

3. What is the interviewer contribution of consent?

# Survey Data

- German Labour Market and Social Security (PASS)
- Conducted by Institute for Employment Research (IAB)
- Wave 1; 2006/2007
- Response rate: 26.7% (RR1)
- Benefit sample (UB II); head of household
- CATI only (CAPI cases excluded, for now)
- Consent to link employment/benefit records (IEB)
  - Asked early in questionnaire
  - Consent rate about 80%

# Administrative Data

- Integrated Employment Biographies (IEB)
- Longitudinal database
  - employment status, wages, benefit receipt, disability status, and demographic information (age, nationality)
- Variables obtained from official records
- Reliability is high for these variables (Jacobebbinghaus and Seth, 2007)
- Used as sampling frame for benefit sample
- Records available for respondents and non-respondents, consenters and non-consenters

# Bias Estimation Details

- Non-Consent Bias
  - Consent indicator linked to IEB data
  - Biases estimated for IEB variables

$$\bar{y}_{nc\ bias} = \bar{y}_{consenters} - \bar{y}_{resps}$$

- Nonresponse Bias
  - Paradata (e.g., disposition codes) linked to IEB data

$$\bar{y}_{nr\ bias} = \bar{y}_{resps} - \bar{y}_{sample}$$

- Measurement error bias
  - Two versions of same statistic obtained from PASS and IEB

$$\bar{y}_{me\ bias} = \bar{y}_{resps,PASS} - \bar{y}_{resps,IEB}$$

# Estimates based on Administrative data

	Administrative Data			
Variable	Sample (n = 17,167)	Contacts (n = 10,717)	Respondents (n = 4,513)	Consenters (n = 3,538)
Age	39.5 (0.1)	40.3 (0.1)	39.5 (0.2)	39.3 (0.2)
Foreign (%)	16.5 (0.4)	13.6 (0.5)	11.0 (0.8)	10.0 (0.7)
UB II (%)	80.2 (0.3)	80.8 (0.4)	83.4 (0.6)	83.1 (0.6)
Disability (%)	4.9 (1.7)	5.4 (0.2)	5.3 (0.3)	5.3 (0.4)
Employed (%)	29.3 (0.4)	30.4 (0.5)	30.3 (0.8)	30.6 (0.8)
Income (monthly)	799.9 (11.2)	788.8 (14.0)	728.5 (21.3)	730.2 ( 24.8)

*Notes.* Estimates account for clustering, stratification, and unequal probabilities of selection.

# Estimates based on Survey data

Variable	Survey Reports	
	All Respondents (n = 4,513)	Consenters (n = 3,538)
Age	39.5 (0.2)	39.2 (0.2)
Foreign (%)	8.5 (0.6)	7.6 (0.6)
UB II (%)	75.9 (0.7)	76.0 (0.8)
Disability (%)	11.3 (0.5)	11.2 (0.5)
Employed (%)	29.3 (0.8)	30.0 (0.8)
Income (monthly)	1130.9 (29.7)	1124.7 (32.9)

*Notes.* Estimates account for clustering, stratification, and unequal probabilities of selection.

# Estimates of NR Bias vs. Non-Consent Bias

Variable	Administrative Data			
	Noncontact	Refusal	Total NR	Non-consent
Age	0.8 (0.1)***	-0.7 (0.1)***	4.6 (0.2)***	-0.3 (0.1)*
Foreign (%)	-3.0 (0.2)***	-2.6 (0.3)***	-5.6 (0.4)***	-0.9 (0.2)***
UB II (%)	0.6 (0.2)*	2.6 (0.5)***	3.2 (0.5)***	-0.3 (0.3)
Disability (%)	0.5 (0.1)***	-0.1 (0.3)	0.4 (0.3)	0.01 (0.2)
Employed (%)	1.0 (0.3)***	-0.1 (0.5)	1.0 (0.5)	0.3 (0.4)
Income (30 days)	-11.1 (8.6)	-60.3 (15.5)***	-71.4 (15.6)***	1.7 (9.5)

Notes. \* < 0.05; \*\* 0.001<p<0.01; \*\*\* p < 0.001

Standard errors estimated using random group method (Wolter, 2009)

# Estimates of ME Bias vs. Non-Consent Bias

Variable	Non-consent	Measurement (All Rs)	Measurement (Consenters)
Age	-0.3 (0.1)*	-0.4 (0.01)	0.03 (0.02)
Foreign (%)	-0.9 (0.2)***	-2.5 (0.3)***	-2.5 (0.3)***
UB II (%)	-0.3 (0.3)	-7.5 (0.4)***	-7.1 (0.1)***
Disability (%)	0.01 (0.2)	6.1 (0.4)***	6.0 (0.5)***
Employed (%)	0.3 (0.4)	-1.0 (0.6)	-0.6 (0.6)
Income (30 days)	1.7 (9.5)	402.4 (28.4)***	394.5 (31.4)***

Notes. \* < 0.05; \*\* 0.001<p<0.01; \*\*\* p < 0.001

Standard errors estimated using random group method (Wolter, 2009)



# Consent Model

- Random-effects logistic regression
  - Respondents nested (non-randomly) within interviewers
- Outcome: linkage consent
- Covariates: survey variables
  - socio-demographics
  - paradata (call attempts, panel cooperation)
  - interviewer characteristics (age, education, gender)

# Consent Model Summary

- Sociodemographics
  - Age (-), Employed (+)
- Paradata
  - Panel cooperation (+)
- Interviewer characteristics
  - Gender (+), Education (-)
  - Interviewer variance component ( $p < 0.05$ )
- Model Diagnostics
  - Pseudo  $R^2 = 0.05$
  - Adj. Pseudo  $R^2 = 0.03$

# Conclusions

- Non-consent bias present for some variables
- Overall non-consent biases are small
- NR/ME biases tend to be larger than non-consent biases
  - data linkage makes sense from TSE perspective
- Significant interviewer variation in consent rates

# Limitations

- PASS response rate is low (26.7%)
- Special population (benefit recipients)
  - Correlates of consent similar in general population
- Quality of administrative data is unknown
- Admin data come from different sources

# Future Work

- Extensions
  - Explaining interviewer variance
  - Non-consent bias trends over time
  - Other surveys/populations
  - Exact linkage vs. statistical matching/imputation
- New data collection possibilities
  - Mechanisms of linkage consent
  - Gain/loss framing (Tourangeau and Ye, 2009)
  - Confidentiality assurance (Singer, Hippler, and Schwarz, 1992)
  - Effect of consent on response accuracy
  - Placement of consent request (anecdotal evidence)

# Thank you

- Slides/manuscript available upon request
- [joesaks@umich.edu](mailto:joesaks@umich.edu)

# Extra Slides

# Consent Propensity Strata

- Question: Is the propensity of consent related to non-consent bias?
- Estimated propensity scores grouped into quintiles
  - Sorted from low-to-high propensity of consent
- Non-consent bias estimated within quintiles



# Non-Consent Bias within Quintiles

Consent propensity stratum	Age	Foreign (%)	UB II (%)	Disability (%)	Employed (%)	Income
Q1 (Low)	-0.2	-1.8	-0.3	0.3	-0.2	31.9
Q2-Q5 (High)	-0.3	-0.4	-0.2	0.01	0.4	1.3
P-value	0.85	0.37	0.96	0.85	0.71	0.75

# % Item Missing Data within Quintiles

- Question: Is the likelihood of consent related to item nonresponse?

Consent propensity stratum	UB II	Foreign	Income (ref/dk)	Income (ref)	Income (dk)
Q1 (Low)	3.2	20.9	4.2	2.4	1.8
Q2	1.3	8.2	2.9	1.1	1.8
Q3	1.0	6.2	3.2	1.4	1.8
Q4	0.8	4.9	2.7	1.8	0.9
Q5 (High)	1.1	2.7	4.7	1.5	3.2
P-value (trend)	0.16	0.05	0.84	0.84	0.84