Sample Loss and Participation Bias in the UASFin Study

Mick P. Couper[1], Marco Angrisani[2] and Arie Kapteyn[2]

[1]ISR, University of Michigan [2]CESR, University of Southern California

MASS Workshop
London, 4-5 June, 2025

We gratefully acknowledge support from the Social Security Administration and the National Institute on Aging under grant U01AG077280

Background

- Surveys are increasingly including a variety of enhancements as discussed at MASS
- These enhancements create the risk of additional sample loss and selection bias
- Adding enhancements to probability-based surveys gives us the opportunity to explore and remediate these potential errors
- We explore one example: the UASFin study

What is UASFin?

- Part of the Understanding American Study (UAS), a probability-based internet panel of adults in the U.S.
- Participants are asked to log into a secure password-protected website (UASFin) where they can share their financial account credentials with a financial aggregator
- This provides real-time access to financial information, including account balances and transactions
- Participants can also see overviews of balances and spending in broad categories

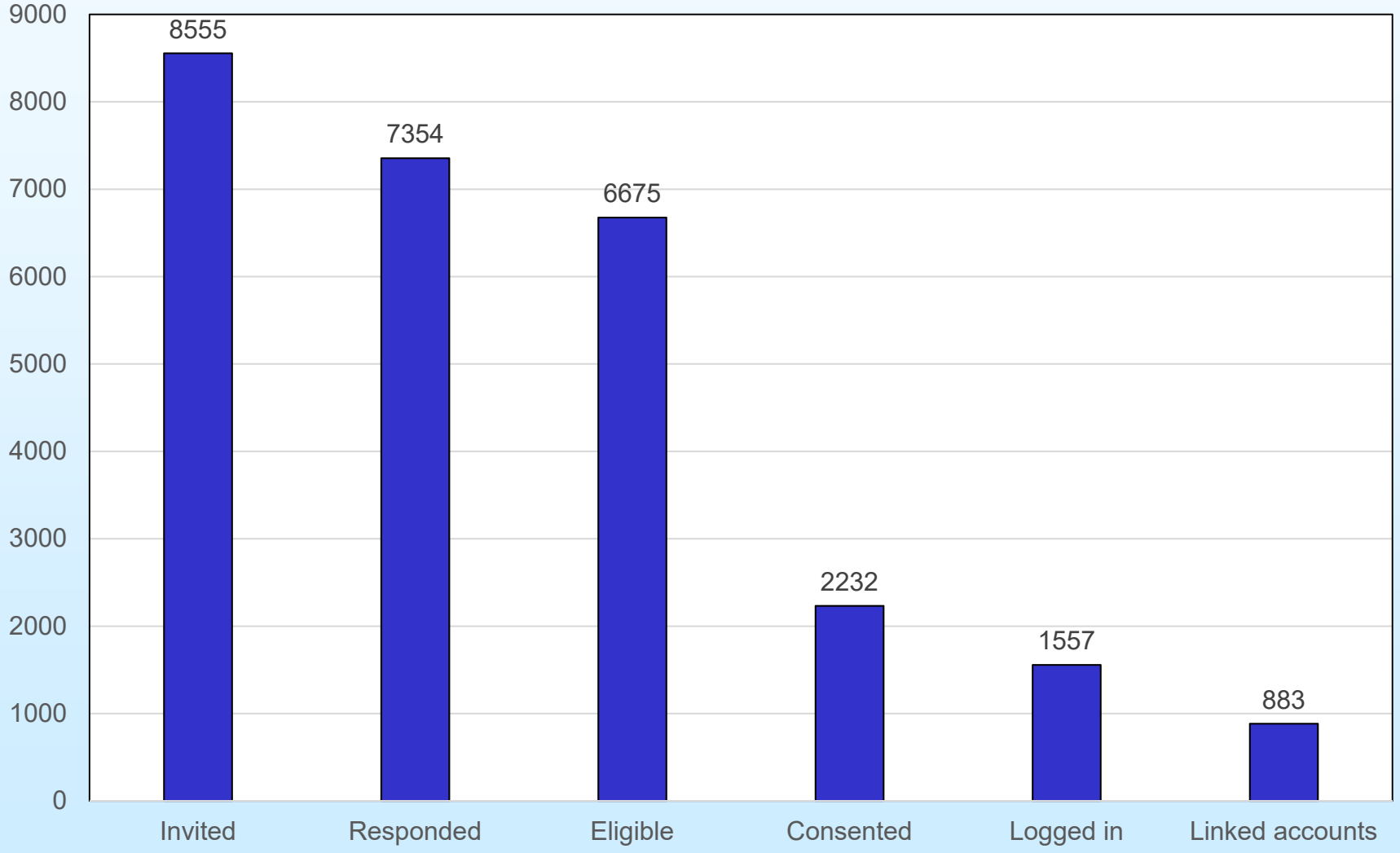
Stages of Selection and Participation in UASFin

- Respond to the invitation survey
- Be eligible for UASFin (use online banking)
- Consent to participate in UASFin
- Log in to the UASFin website
- Link one or more financial institutions
- Link one or more accounts (opt out of linking selected accounts)

Research Questions

- RQ1: What are the sample losses at each stage of the process?
- RQ2: What are the reasons for non-consent?
- RQ3: What are the predictors of participation at each stage of the process?
  - Causes and correlates of sample loss
  - Are these the same or different across the stages?
- RQ4: What are the participation biases at each stage of the study?
  - Consequences of cumulative sample loss
  - Are the biases compounding or offsetting across stages?

RQ1: Participation Counts and Rates



Conditional losses	86.0%	→	90.8%	→	33.4%	→	69.8%	→	56.7%
Cumulative losses	86.0%	→	78.0%	→	26.1%	→	18.2%	→	10.3%

RQ1: Description of Sample Loss

- Biggest loss at consent: 33.4% of eligible persons consented
- Conditional on consent, 69.8% logged in and 39.6% linked 1 or more accounts
- Cumulative sample loss was 89.7%
  - 10.3% of invited sample complied with all steps
- For subsequent analyses, we focus on 3 steps
  - Responded and eligible
  - Consented
  - Logged in and linked 1+ accounts

RQ2: Reasons for Non-Consent

- Check-all-that-apply question with an “other, specify” option
  - Multiple mentions possible
- See next slide

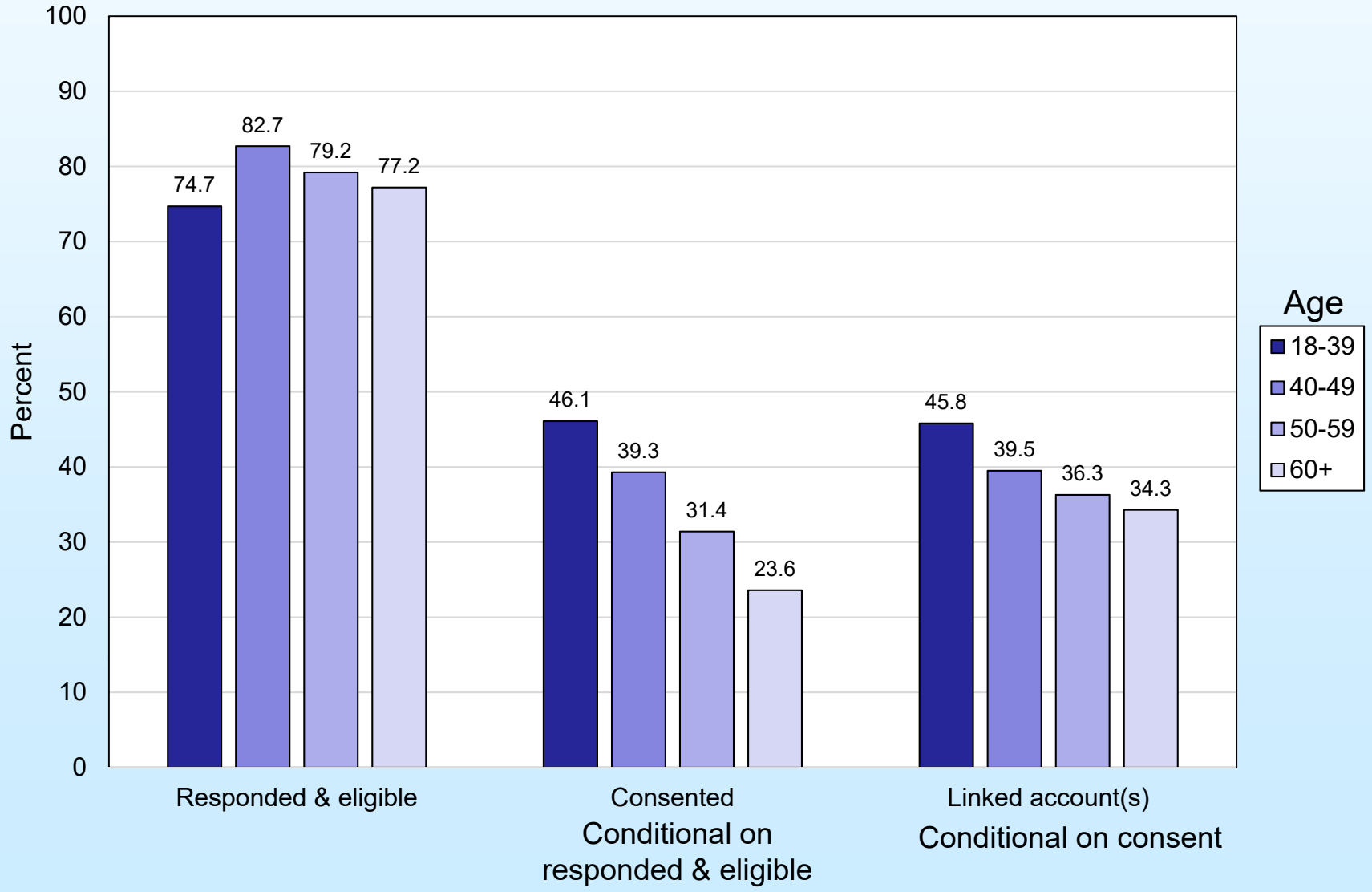
RQ2: Reasons for Non-Consent (%)*

* Multiple mentions possible

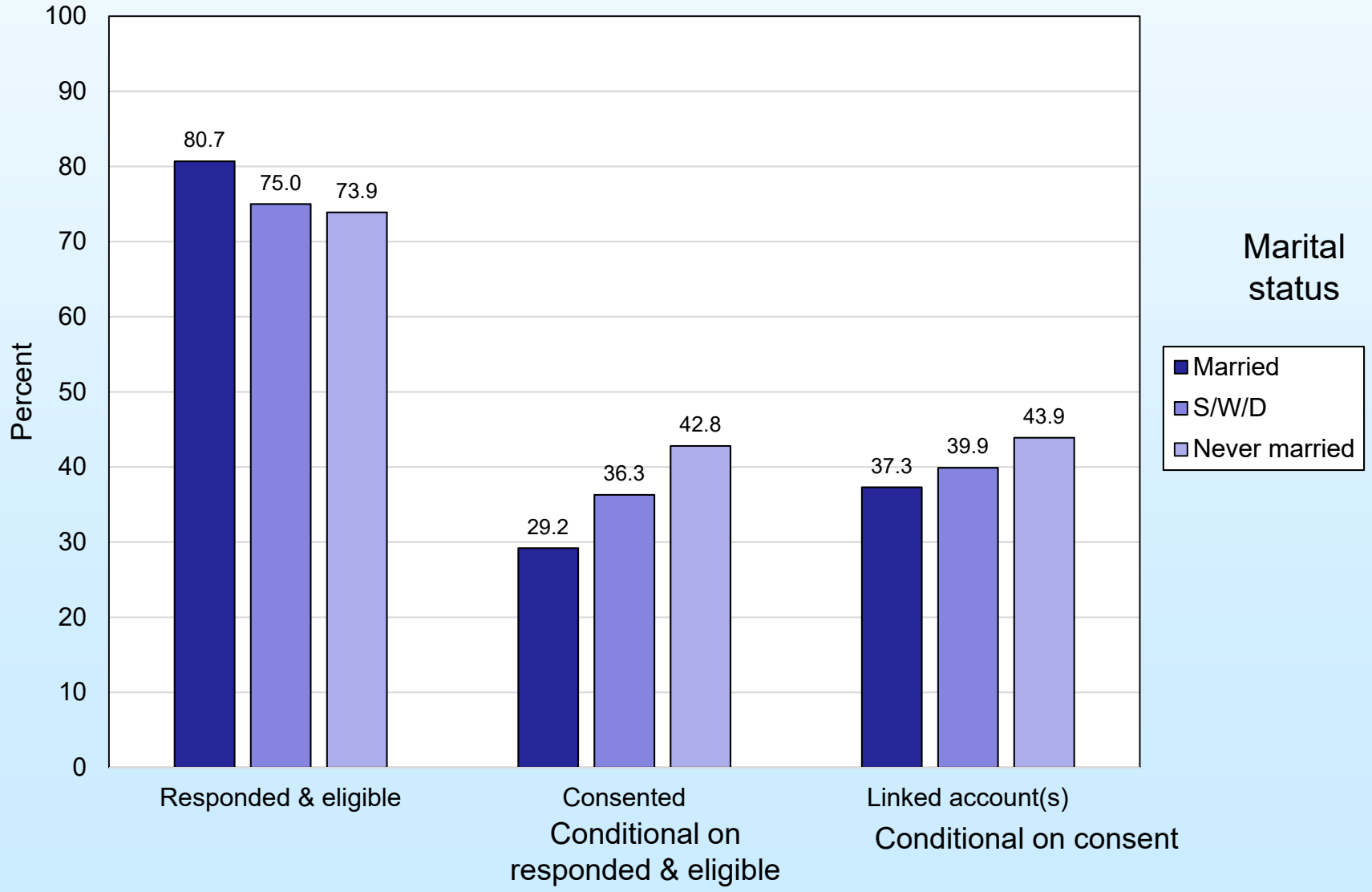
RQ3: Predictors of Participation

- We examined both bivariate and multivariate relationships
- Examined socio-demographic, survey engagement and substantive variables
- Selected bivariate examples follow

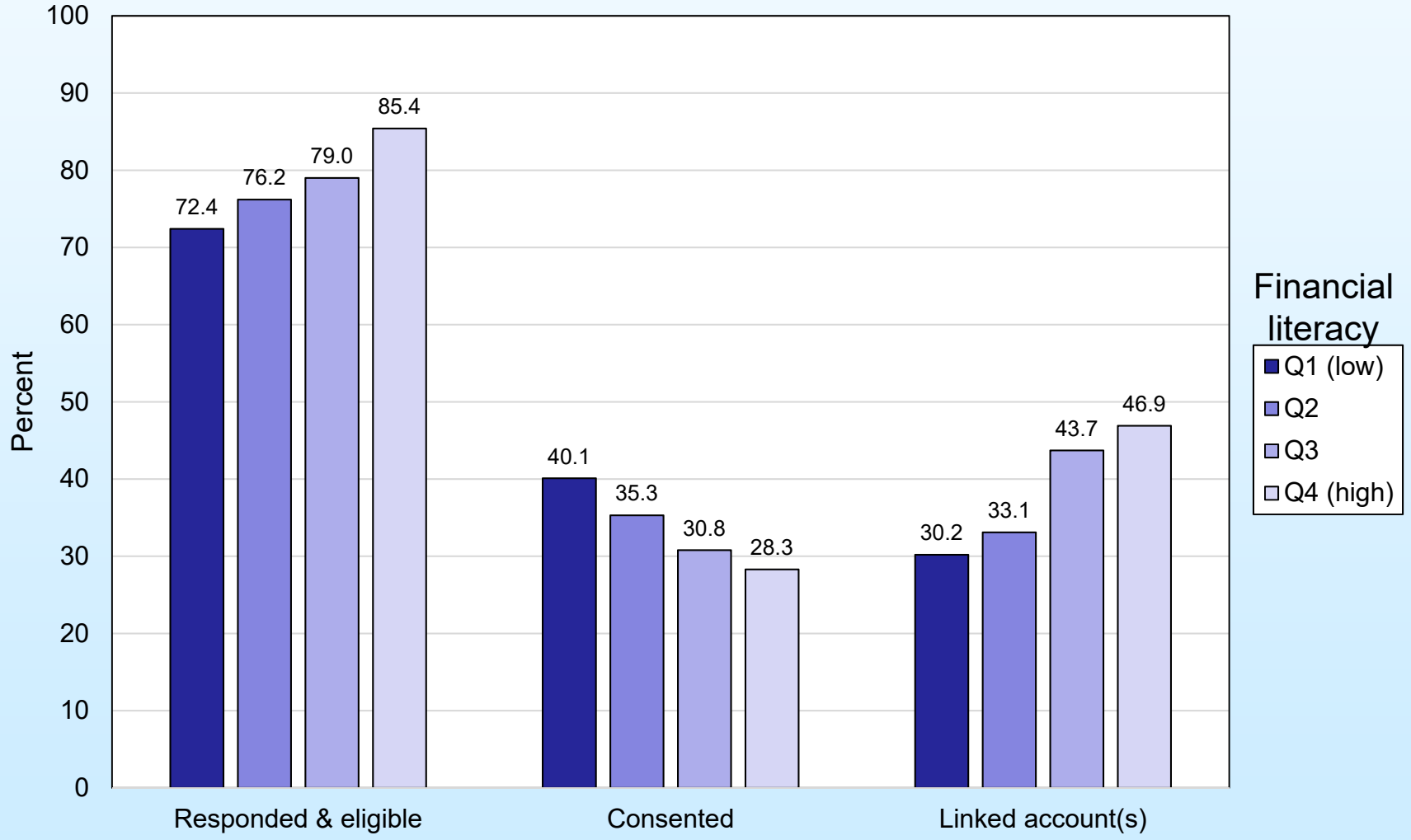
RQ3: Conditional Participation Rates by Age



RQ3: Conditional Participation Rates by Marital Status



RQ3: Conditional Participation Rates by Financial Literacy Score*



*Similar pattern for income quartiles

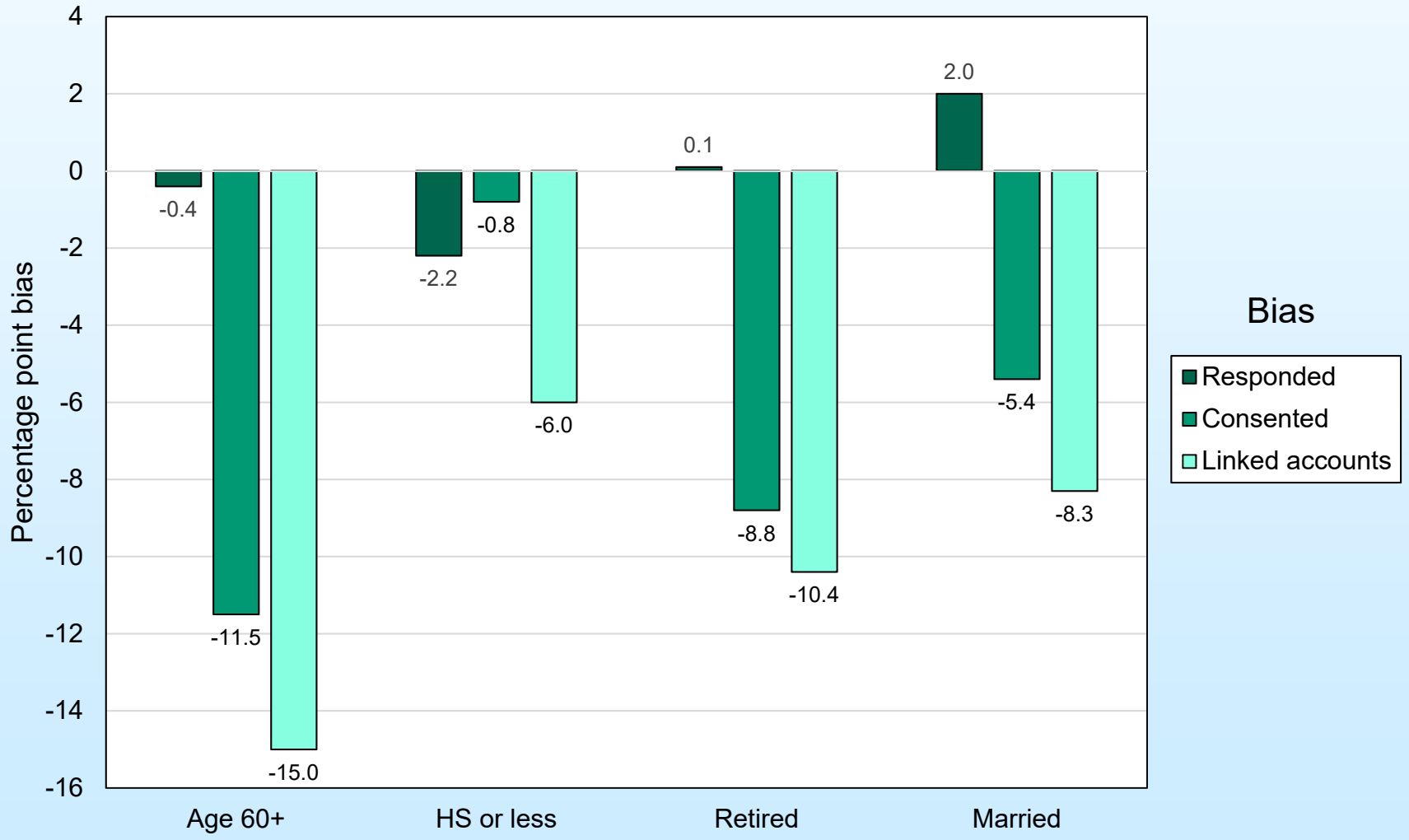
RQ3: Predictors of Participation

- Some variables significant predictors of participation in all stages
  - Age, survey activity ratio, risk tolerance, conscientiousness
- Others are significant in some stages but not others
  - E.g., gender, marital status, education, income, basic internet skills, numeracy, extroversion
- Others are not significantly associated with any stage
  - Region, race/ethnicity
- Some effects are consistent across all stages, but others are not
  - E.g., financial literacy (see earlier slide), numeracy

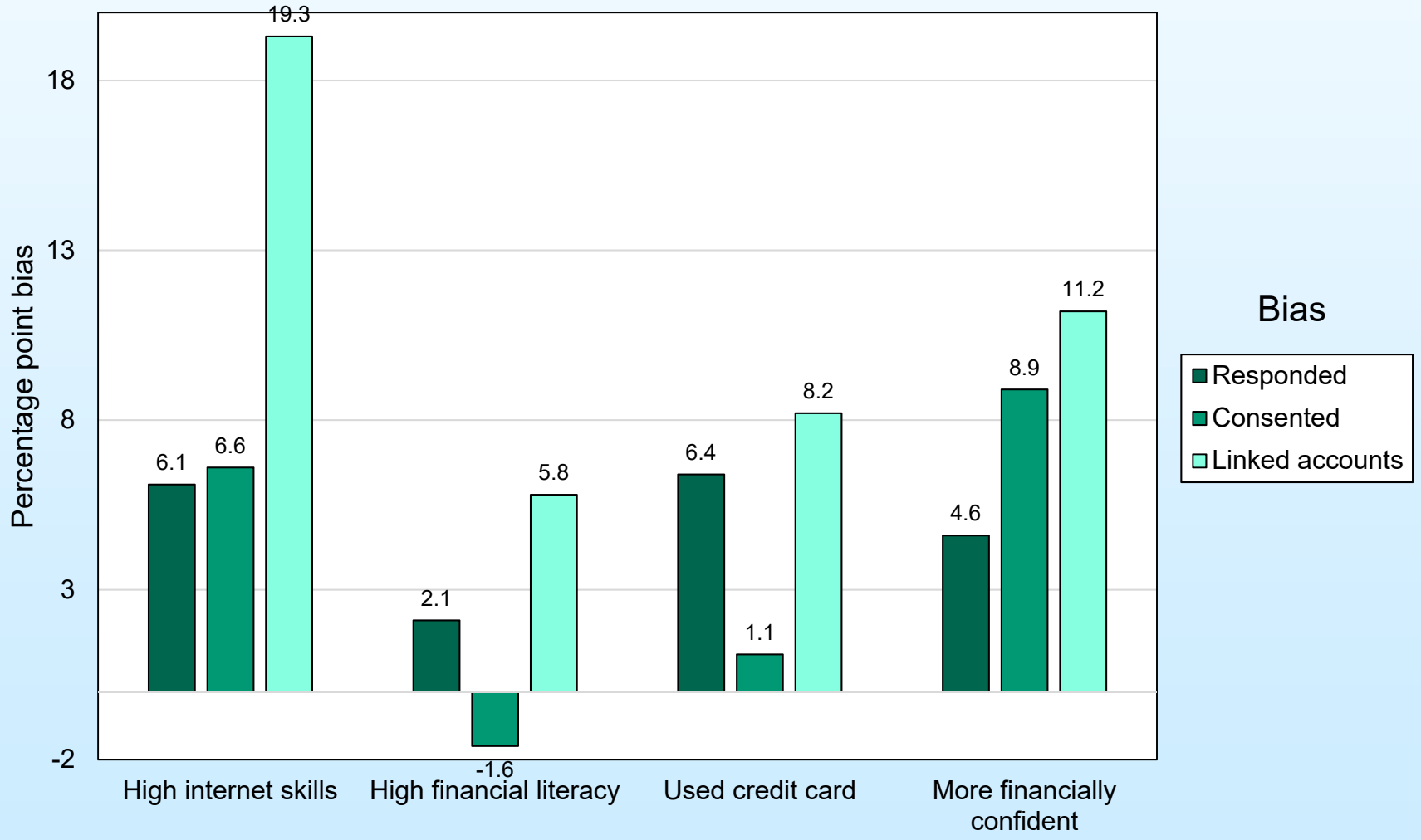
RQ4: What are the Participation Biases at Each Stage of the Study?

- Here we compare each subset to the original sample of invitees
- Do biases compound or are they offsetting?
  - Biases tend to increase over successive stages – sample is getting smaller and is increasingly different from the original set of panelists
  - There are some relatively large biases (age, education, income, employment status, marital status)
- Selected examples follow

RQ4: Participation Biases: Selected Demographic Variables



RQ4: Participation Biases: Selected Substantive Variables



RQ4: Participation Biases

- Biases vary across stages of participation
- Largest observed socio-demographic biases for older, retired, and married
  - All under-represented in the final dataset
- Substantive biases also observed
  - Those with high basic internet skills, high financial literacy scores, used a credit card in past 3 years, and more financially confident than 5 years ago all over-represented
- Patterns of bias do not always suggest compounding (getting larger over stages)

Summary

- The patterns of associations are not always consistent across stages of participation, suggesting that these are qualitatively different
  - Only focusing on the last stage of participation (comparing UASFin participants to all invitees) misses nuances across stages
- Understanding the reasons behind sample loss – and particularly differential sample loss – across stages is important for developing mitigation and correction strategies
- Selection mechanisms can be studied when transaction data are obtained within a probability panel, but remain largely unknown when transaction data are obtained directly from financial institutions

Implications for Practice

- Including measures related to the selection process and key outcomes helps understanding of participation biases
  - Go beyond socio-demographics
  - Include privacy/confidentiality measures, comfort/familiarity with technology, etc.
  - Make this part of the design/planning stage
- Identifying correlates/predictors of participation at each stage can guide fieldwork strategies to minimize differential loss
  - E.g., responsive/adaptive designs, weighting

Thank You!