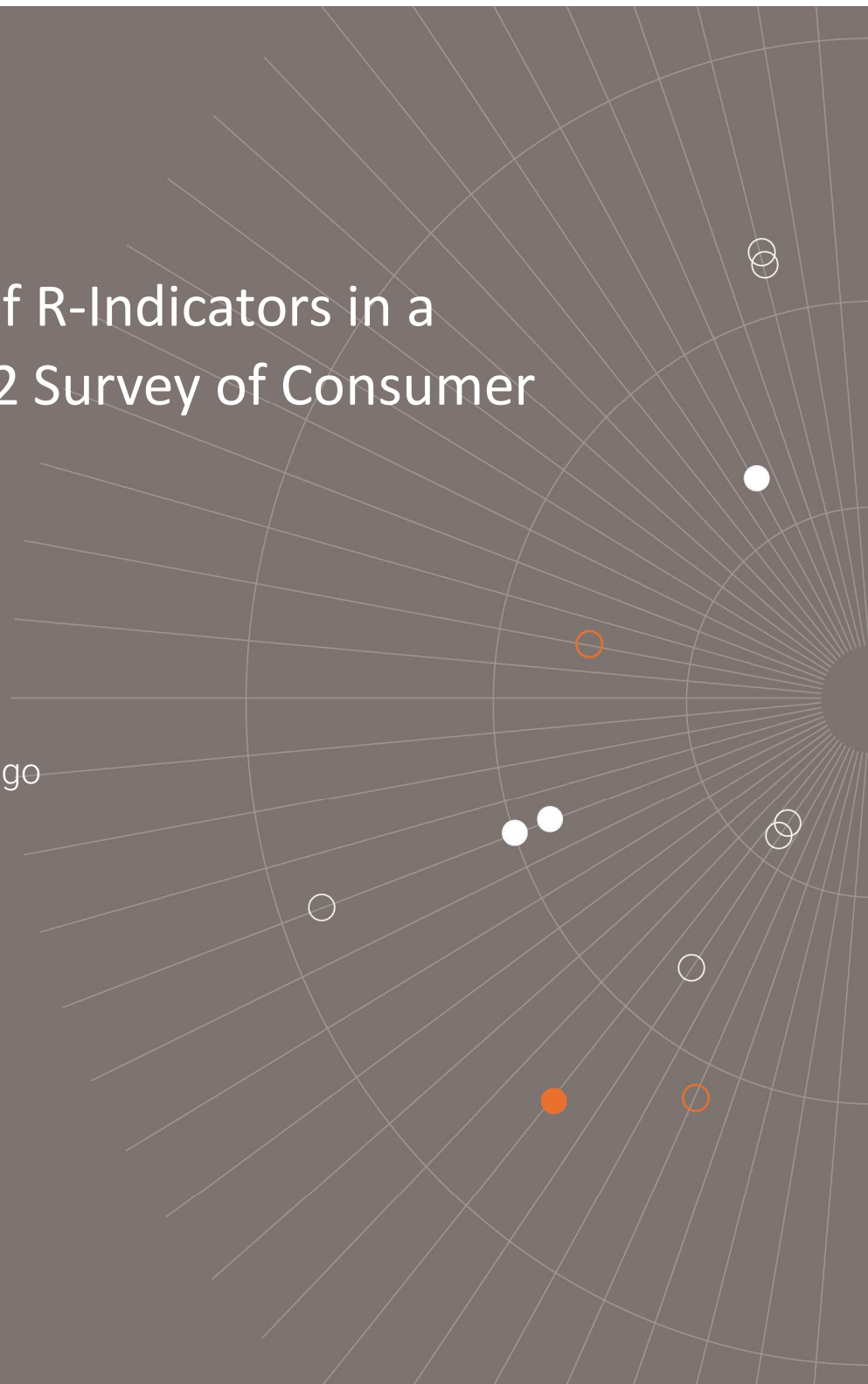


Sharpening our Tools: The Use of R-Indicators in a Challenging Context for the 2022 Survey of Consumer Finances

FedCASIC 2024

April 16, 2024

Katie Archambeau, NORC at the University of Chicago
Steven Pedlow, NORC at the University of Chicago
Kate Bachtell, NORC at the University of Chicago
Cathy Haggerty, NORC at the University of Chicago

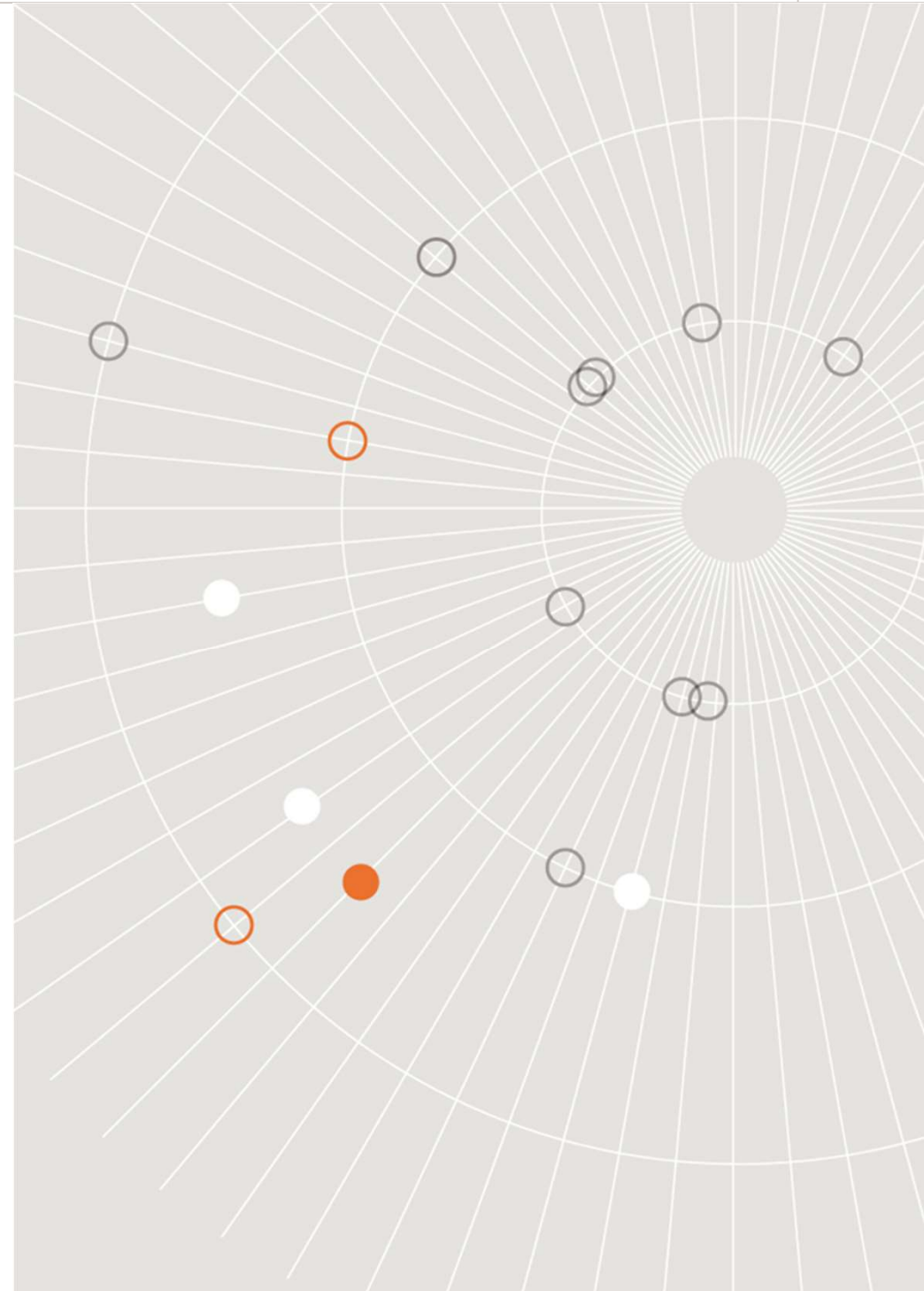


R-Indicator Background

What are R-Indicators?

Representativity Indicators

- Integral part of Adaptive Survey Design (ASD)
- Output from a Logistic Regression Equation
 - Independent variables predict which cases are completed
 - Dependent variable is whether or not interview is completed



3 Levels of R-Indicators

Overall Sample R-Indicator

- Value between 0 and 1
 - Overall indicator of sample's representativeness
 - 1 indicates perfect representation

Unconditional Variable R-Indicator (ex: Age)

- Value between 0.0 and 0.5
 - Larger values indicate more differences between sample and completes
 - 0 indicating the sample with respect to the associated variable has perfect representation

Unconditional Category R-Indicator (ex: Age over 65 years)

- Value between 0.0 and ± 0.5
 - Tells us whether the category is over- or under-represented in the completed interviews
 - 0 indicating the category of the associated variable has perfect representation

Current R-Indicator Work

SCF 2022

Survey of Consumer Finances (SCF) Overview

- Sponsored by the Board of Governors of the Federal Reserve System
- Premier source of data on U.S. household finances
- Dual Frame Sample
 - Nationally representative Area Probability (AP) Sample
 - List Sample (oversamples wealthy households)
- Fielded every 3 years



<http://scf.norc.org>

Building the Model

Identify key statistics for measuring representativeness

1. Must use covariables available prior to the start of data collection
2. Keep variables consistent throughout the field period

Block Group Level Census Variables

Examples:

- Owner Occupied Rate
- Public Assistance Rate
- Age Group Percentages
- Median House Value
- College Graduate Rate
- Other Language Rate

SCF Paradata

- Oversampling Variables
 - Likelihood of Asian Resident
 - Likelihood of Black Resident
 - Likelihood of Hispanic Resident
- Multi-Unit Building

Strategy for 2022 R-Indicators

Work collaboratively with field staff

+

Use propensity scores to identify most under-represented cases

=

Identify 10-15 percent of non-completed cases with lowest propensity scores and work with interviewers to increase effort for those flagged cases!

How did we increase effort for flagged cases?

- Identified a list of 'higher priority' AP cases based on R-indicators results and asked Field Interviewers to prioritize these cases in front of 'normal priority' AP cases
- Sent additional mailings to priority groups
- Escalated post incentives for priority groups
- Increased use of maps showing pending cases by priority to inform FI travel plans

Two Interventions occurred during data collection

October 24

- Flagged 641 cases as 'higher priority'
 - Propensity scores between 0.15–0.2
 - Bottom 9.9% of propensity scores

January 9

- REPLACED October 24 cases with a new set of cases
- Flagged 809 cases as 'higher priority'
 - Propensity scores between 0.26-0.27
 - Bottom 13.7% of propensity scores
- Less than 50% overlap with original 'higher priority' cases

SCF Completion of October 24 Pending Cases

Category	Cases	Completed 1/9	Rate 1/9	FINAL Completed	FINAL Rate
Higher Priority	641	81	12.64%	130	20.28%
Normal Priority	5,851	496	8.48%	968	16.54%
TOTAL	6,492	577	8.89%	1,098	16.91%

SCF Completion of January 9 Pending Cases

Category	Cases	FINAL Completed	FINAL Rate
Higher Priority	809	115	14.22%
Normal Priority	5,118	420	8.21%
TOTAL	5,927	535	9.03%

Model Before/After October Intervention

Data Set	Best October Model	Best October Model on 1/9	Best October Model at end of data collection
Concordance score	56.2%	55.1% ↓	54.7% ↓
Overall R-score	0.9186	0.9256 ↑	0.9246 ↑
Max-Scaled R-Square	0.0116	0.0080 ↓	0.0074 ↓
PARTIAL R-SCORES			
Owner Occupied Rate	0.0279	0.0249 ↓	0.0136 ↓
Urban-Suburban-Rural	0.0207	0.0186 ↓	0.0257 ↑
Public Assistance Rate	0.0205	0.0221 ↑	0.0128 ↓
Under 18 Pop. Percentage	0.0154	0.0116 ↓	0.0186 ↑
Other Lang. Spoken Rate	0.0058	0.0007 ↓	0.0032 ↓

Model Before/After January Intervention

Data Set	Best January Model	Best January Model at end of data collection	Final Best Model at end of data collection
Concordance score	55.3%	53.6% ↓	54.8% ↓
Overall R-score	0.9190	0.9327 ↑	0.9179
Max-Scaled R-Square	0.0093	0.0057 ↓	0.0086 ↓
PARTIAL R-SCORES			
Oversample (Black)	0.0292	0.0308 ↑	0.0308 ↑
Owner Occupied Rate	0.0249	0.0136 ↓	Not Significant
Public Assistance Rate	0.0221	0.0128 ↓	Not Significant
Uninsured Rate	0.0167	0.0069 ↓	Not Significant

Final Results of SCF 2022 R-Indicator Interventions

- **Interventions improved representativeness**
 - Perfect representativeness is a 50 percent Concordance Score
 - Best October model could predict completion status for 56.2% of cases
 - Final best model could predict completion status for 54.8% of cases
 - Identifying 'higher priority' AP cases removed $1.4/6.2 = 22.6$ percent of biases
- **Likely only one intervention was needed**
 - Less than 50 percent overlap between October cases and January cases
 - Challenge of explaining shift in cases to the field
 - The Final Best model included two variables in October model, but not January model
 - When we replaced our set of 'higher priority' cases, variables from earlier model tried to return
 - Possibly add additional 'higher priority' cases instead of fully replacing them

Future R-Indicator Plans

Plans for R-Indicators in future studies



Maximize collaboration with field staff

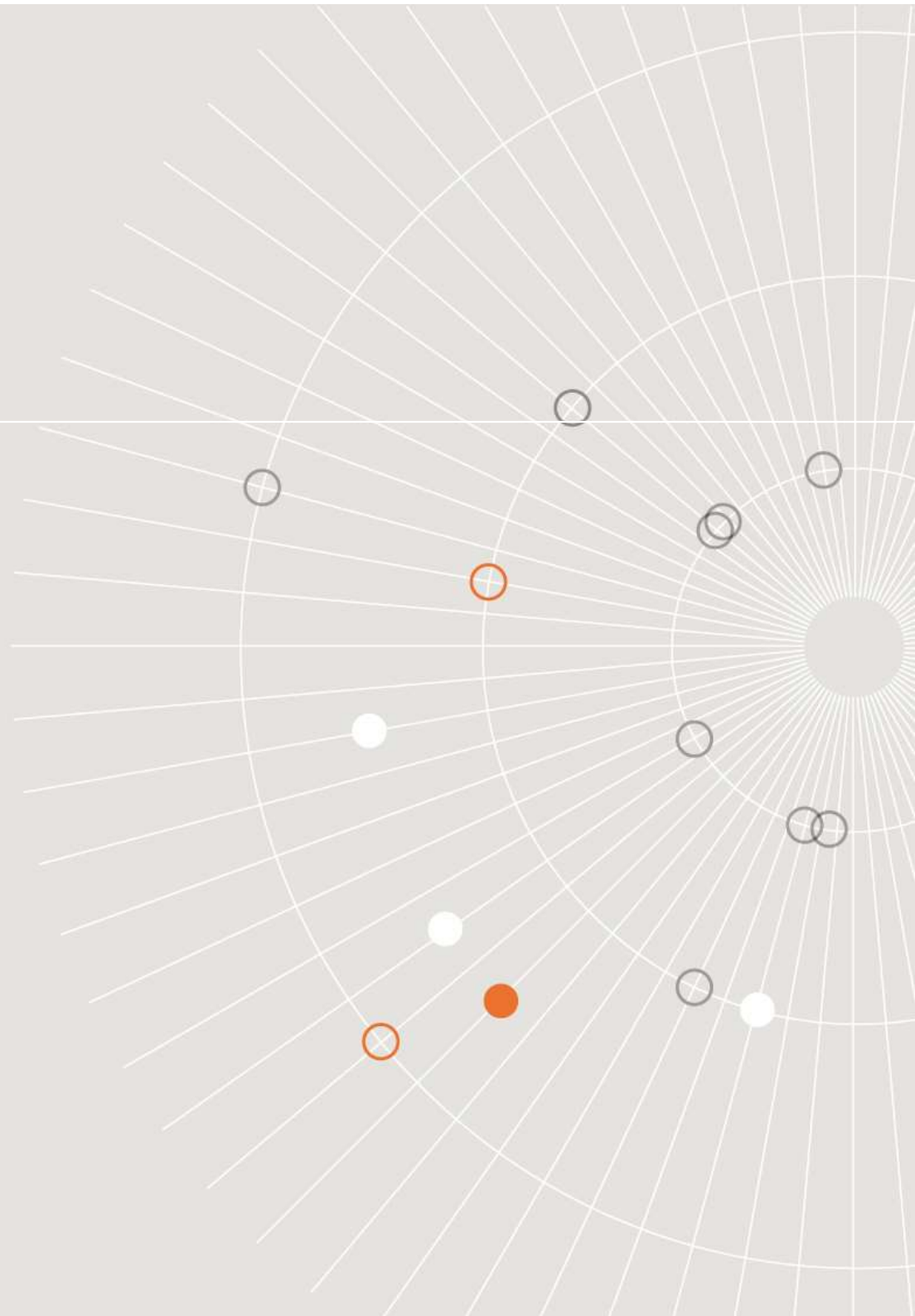


Extend application of R-Indicators to list sample



Identify additional variables prior to data collection that can be used in R-indicator models

Questions?



Thank you.

Katie Archambeau
Senior Data Scientist
archambeau@norc.org

 Research You Can Trust™

 **NORC** at the
University of
Chicago