Trends in Survey Statistics

Ashley Amaya, Rachel Harter, Katherine Morton, & Stephanie Zimmer

Women in Statistics and Data Science Conference
La Jolla, California
October 19-21, 2017

Introductions

- Ashley Amaya

- Katie Morton

- Rachel Harter

- Stephanie Zimmer

Acknowledgement

- Thank you to our other RTI co-author, Derick Brown.

- This work was partially supported by the Energy Information Administration and used frame and sample data from the 2015 Residential Energy Consumption Survey (RECS).
  - Contract Nos. DE-EI-0000515.

- The views expressed in this presentation do not necessarily reflect the official policies of the EIA, Department of Energy, nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

Today's Goals

- Provide a history of statistics in survey research
- Give an example of a statistical research agenda within survey research
  - Minimizing error (bias) using address-based sampling

Reflections on the History of Survey Statistics: A personal and selective review

Sources for History of Survey Sampling

- Bethlehem, Jelke (2009). *The rise of survey sampling.* Statistics Netherlands.

- Hansen, Morris H. (1987). "Some History and Reminiscences on Survey Sampling," *Statistical Science*, 2(2), pp. 180-190.

- Brewer, Ken (2013). "Three controversies in the history of survey sampling," *Survey Methodology*, 39(2), pp. 249-262.

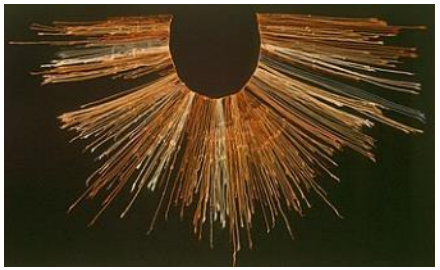
Origins of Official Statistics

- Censuses were recorded in ancient times
  - Babylonian censuses of agriculture
  - Chinese censuses of populations for revenue and military evaluation
  - Egyptian summary overviews
  - Roman censuses of people and property

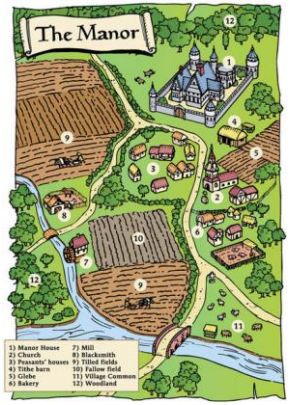
Bible History Online

Official Statistics from the Middle Ages

- 1000-1050 A.D. - Incan records of people, houses, llamas, marriages, and draftable men



- 1086 - William the Conqueror's first *Domesday Book* about each manor and village



- Parish records

National and Colonial Censuses

- 1666 census in New France
- 1746-1769 first censuses in Scandinavia
- 1789 decennial enumeration mandated by U.S. Constitution

“The census in its modern form can be closely associated with the rise of democracy.” Bethlehem 2009

Early 19th Century

- Several statistical principles developed
- Statistical agencies, societies, and journals founded
- ASA founded 1839

Inferences from Partial Data

- "Political Arithmetic" – life tables and vital statistics – from mid-17th century
  - William Petty, Nicolaas Struyck, Willem Kersseboom
- "Social Calculus" – early 19th century
- Late 19th Century
  - "Partial investigations" (close examination of subdomains)
  - "Monograph studies" (surveys)

Representative Method

- Anders Kiaer - Founder and director of Statistics Norway
  - 1897 paper to International Statistical Institute (ISI)
    - Intuitive principles, not theory, of surveys
- 1903 ISI officially supported the “Representative Method”
- Others (March, Bowley) advocated probability sampling
- 1925 ISI accepted the idea of sampling

Probability Sampling

- Jerzey Neyman's 1934 paper to Royal Statistical Society
  - Principles of inference from samples of finite populations based on randomization
  - Theoretical basis for optimal allocation to strata
  - Defined confidence intervals
  - Compared probability sampling with purposive sampling
  - Advocated use of probability sampling *exclusively*
  - Later developed other major sampling concepts for USDA

Other Survey Developments

- Tippet (1927) – tables of random numbers
- Jessen (1942) – concepts that led to rotation sampling
- Cochran (1942) – regression estimation in surveys
- Hansen & Hurwitz (1943) – multi-stage samples
- Hansen & Hurwitz (1943) – probability proportional to size sampling
- Madow & Madow (1944) and Cochran (1946) – systematic sampling
- Horvitz & Thompson (1952) – unbiased estimates

- Most classical sampling theory developed by 1952

- Early classic sampling textbooks:
  - Deming 1950
  - Cochran 1953
  - Hansen, Hurwitz, and Madow 1953

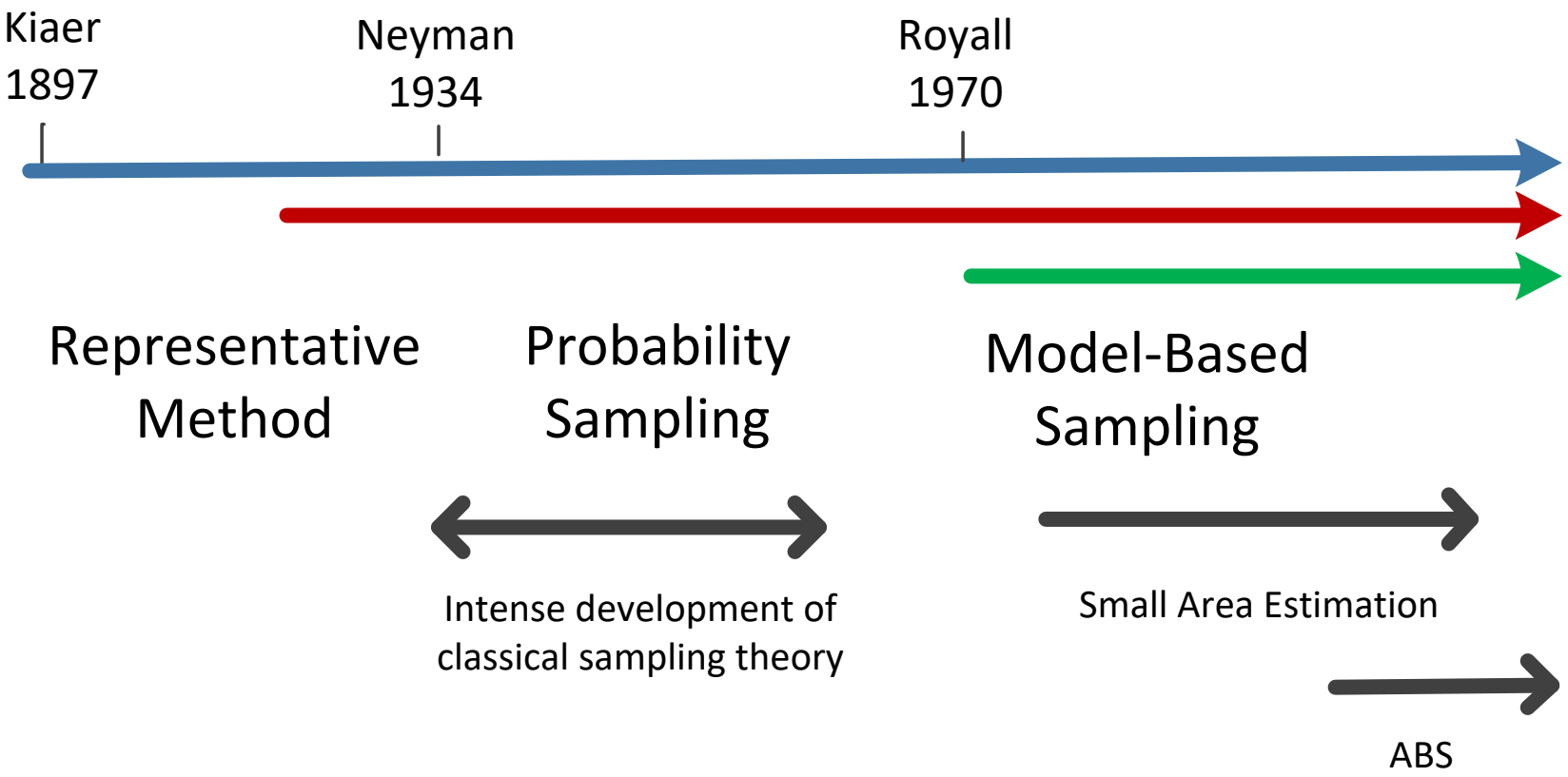
Model-Based Survey Sampling and Estimation

- Richard Royall's 1970 paper
  - Reinstate purposive sampling with prediction-based inference
  - Use best linear unbiased prediction (BLUP)
  - Controversial – depends on the model

Developments in the Last Few Decades

- Sampling for Rare and Clustered or Networked Populations
  - Snowball sampling
  - Network sampling
  - Respondent-driven sampling
  - Adaptive sampling
- Nonprobability sampling
- Small Area Estimation
- Bayesian methods take advantage of modern computing power
- Address-Based Sampling

Development of Modern Survey Sampling



Survey Modes through the Years

- **In-Person** – highest response rates, most expensive, most time-consuming
- **Mail** – lower cost, often a supplemental contact mode, increasing in popularity with availability of postal lists, somewhat time-consuming
- **Telephone** – biased in the early years, primary mode for many years, becoming biased again if cell is not included, becoming more expensive
- **Mobile Device** – relatively new, inexpensive, better at reaching hard-to-find young adults
- **Web** – relatively new, potentially biased, inexpensive, not usually for initial contacts
- **Mixed Mode** – combine the best features of 2 or more methods (e.g., mail contact with web survey)

Changing Technologies

- **PAPI** – paper and pencil

- **CAI** – computer-assisted interviewing – cleaner data
  - CATI – telephone
  - CAPI – personal interviews on laptops
  - CAWI – web interviews
  - CARI – recording of interviews

Changing Technologies (cont.)

- **Computers** - Univac to mainframes to PCs to laptops to tablets and smartphones

- **IVR** – interactive voice response (pre-recorded interviewer)
- **Text-to-speech** – on-the-spot conversion of questionnaire text to synthesized interviewer

Surveys By Industry, and My Connection to Them

- Agricultural Surveys
- Business Surveys
  - Industry
  - Marketing
  - Opinion Polling
- Government Surveys
- Social Science Surveys
  - Independent Not-for-Profit Research

Agricultural Surveys

Agricultural Surveys and Experimental Design

- Rothamsted Experimental Station (1843)
  - R.A. Fisher was hired in 1919 – statistical powerhouse
  - Oscar Irwin, John Wishart, Frank Yates, William Cochran, John Nelder
    - Promoted randomization, replication, and stratification
    - Theory to estimate precision using only the sample
- Indian Council of Agricultural Research (1929)
  - P.C. Mahalanobis
    - Statistical Laboratory (1920s)
    - Indian Statistical Institute (1931)
- Statistical Laboratory at Iowa State University (1934)
  - George Snedecor, William Cochran, T.A. Bancroft, Gertrude Cox, Oscar Kempthorne, and others
    - Department of Statistics (1947)

Iowa State University

- Land Grant University
- Strong Agricultural emphasis
- Connections to Rothamsted

- My Ph.D. Major Professor – Wayne A. Fuller
  - Ag Economics background

- My Support – USDA surveys

My Research

- Small Area Estimation
  - Estimating corn & soybean acreage by county
  - Using LANDSAT imagery as auxiliary data
  - Battese, Harter, and Fuller (1988)
  - Fuller and Harter (1987)

Small Area (or Domain) Estimation

- Desire estimates for domains smaller than those for which the survey was designed.
- Combine sparse survey data with auxiliary data to "borrow strength"
  - Survey data from other small areas
  - Survey data from other time periods
  - Related variables available for the population
- Use model relationships between sample data and population auxiliary data.
- Early work at U.S. Census Bureau in the 1970's.
- Random components/hierarchical models are now common.
- Modern computing power made Bayesian methods feasible.

- Lesson: Users want more than the survey was designed to provide.

“Find a good auxiliary variable. After that you are just keeping statisticians busy.” Fuller

Brushes with Greatness

- Dr. Fuller introduced me to other “big-name” survey statisticians
  - J.N.K. Rao
  - Tore Dalenius
  - Leslie Kish
  - Graham Kalton
  - Dan Horvitz
  - Joe Waksberg

Business Surveys

W. Edwards Deming and Industrial Statistics

- Promoted Shewhart's statistical process control
- Helped rebuild Japanese industry after WWII
- Deming's 14 Points
- Awarded National Medal of Technology and Distinguished Career in Science

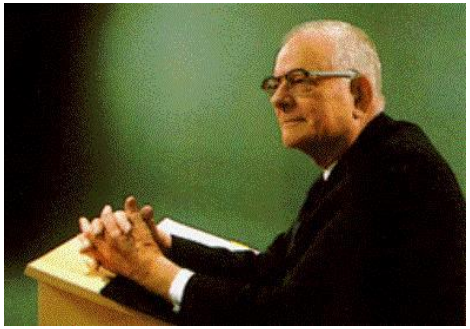
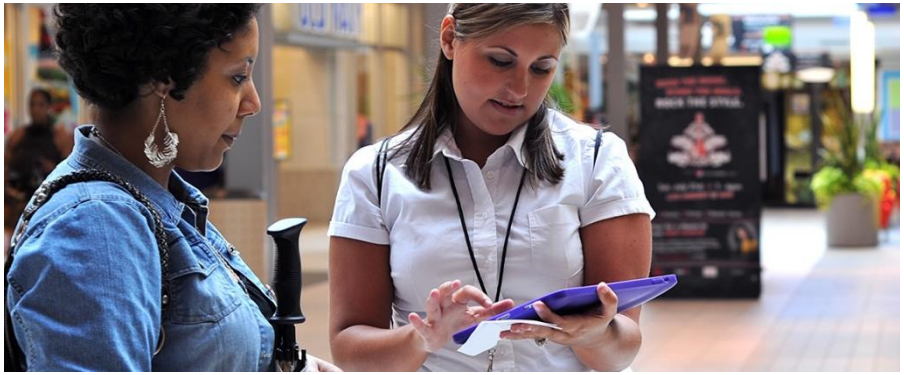


Photo: By FDA - http://www.fda.gov/oc/initiatives/criticalpath/stanski/stanski.html, Public Domain, https://commons.wikimedia.org/w/index.php?curid=3239071

Statistics can have an enormous positive impact on business success, if permitted.

Survey Methods Associated with Marketing



- Intercept sampling
- Quota sampling
- Commercial sample providers
- Marketing databases
- Panel surveys

Commercial Surveys and Market Research

- A.C. Nielsen Company (1923)
  - First market research company
  - Radio (1936) and Television (1950) Ratings
  - Surveys of Supermarkets, Drug Stores, Mass Merchandisers (from 1920s)
  - High tech
  - Helping manufacturers spend their advertising dollars

- A “Big Data” company

One Nielsen Project

Restating historical audit data to match new scanner data

Selected Themes from my Nielsen Experience

- Measuring trends is most important! Don't disrupt trends.

- Customers want more and more data, until they can't handle it. Then they need help seeing what is important.

Public Polling

Opinion Polling for Marketing and Politics

Rulers have an easier job when they have the goodwill of the governed.

- French Revolution
- U.S. Declaration of Independence
- Newspaper straw polls
- FDR enlisted polling organizations for his administration

- George Gallup
  - American Institute of Public Opinion (1935)
  - Gallup Organization and Gallup Poll
- Roper Center (1947)
- Harris Poll (1963)
- American Association of Public Opinion Research (1947)

Lessons from Political Polling

- Public opinion research is critical to maintaining a free society.
- Legitimate polls use sound statistical methods.
- Politics is risky business, and nonpartisan not-for-profits stay away.

Surveys in Statistical Agencies in the Federal Government

Founding of Modern Federal Statistical Agencies

IRS 1862

NCES 1867

BLS 1884

Census 1903

SSA 1935

NSF 1950

NCHS 1960

NASS 1961

And More Federal Statistical Agencies

ERS 1961

BEA 1972

EIA 1977

BJS 1979

BTS 1992

Unlike other countries, the U.S. does not have one central statistical agency, and sharing data is extremely difficult.

U.S. Census Bureau

- 1936 – under Morris Hansen, began research on sampling and applications
- 1937 – Congress authorized voluntary registration of unemployed
  - Converted to mail survey – the 1937 Enumeration Check Census
  - Demonstrated usefulness of probability sampling on national study
- Sampling in the 1940 Census
- Labor Force Survey moved to the Census Bureau in 1942
- 1954 Labor Force Survey redesigned into Current Population Survey
- Burgeoning survey operations

- Small area estimation started at the Census Bureau in 1970's
- The Census Bureau has its own address frame.
- Sampling for the decennial census remained controversial.

- Individuals from the Census Bureau have had a major impact on my career.

Social Science Research

NORC at the University of Chicago

- Founded 1941 at the University of Denver by Harry H. Field
- Social research in the public interest
- Give people a voice in political, social, and economic decisions
- Contractor for Office of War Information, WWII
- Mix of business, academia, and government

I learned so much more about survey statistics!

(and government contracting and corporate management)

NORC at the UNIVERSITY of CHICAGO

Statistical Leaders at NORC

- Roger Tourangeau
- Kirk Wolter
- Colm O'Muircheartaigh
- Fritz Scheuren
- John Thompson
- Ken Copeland
- Dan Kasprzyk
- Avi Singh
- Consultant Brenda Cox

Example Project

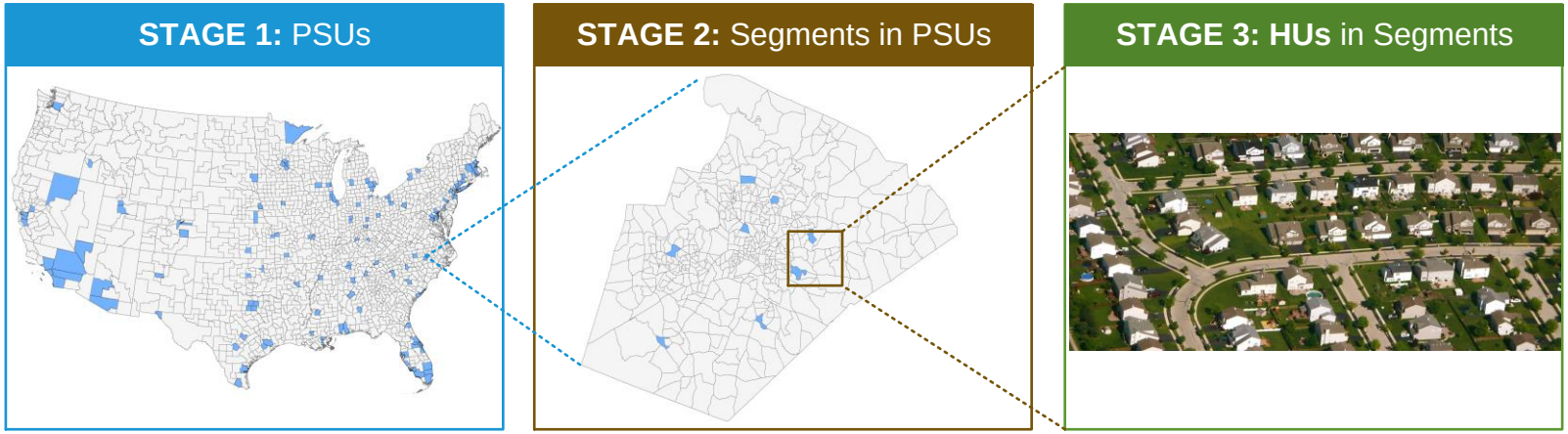
Bureau of Labor Statistics – Illinois Dept of Employment Security

- Current Employment Statistics (CES)
  - Started in the 1920s – Fed-State cooperative program
  - Survey of employers for employment and wages
  - Trend disruption in 1991 – Don't disrupt trends!
  - ASA investigation
  - Advisory panel on redesign

- Illinois' need for county-level estimates
  - Small Area Estimation (Hooray!)
  - Production system

Example Project – NORC National Frame

- Area probability design



- List all housing units within selected segments as a frame for general population surveys.

Frames of Housing Units Revolutionized by ABS

History of Address-Based Sampling (ABS)

- Census Bureau has had a *proprietary* frame of addresses for many years.
- Marketing database companies provide mailing lists to marketers.
- Licensed companies got mailing lists “corrected” by USPS.
- Iannacchione et al. (2003) saw potential in address lists as frames.
- NORC implemented ABS in that decade’s national frame.

Early ABS Research

- How well do the address lists cover the population of housing units?
- Where is coverage not so good, and why?
- How do we operationalize a survey that uses an address frame?
- How can we supplement the address list where needed?
- Can ABS surveys replace telephone surveys?
- How much can we save by using ABS for in-person surveys?
- What other data are available from the USPS and the marketing vendors?
- How useful are the auxiliary variables?

Selected Themes from my NORC Experience

- Response rates are dropping.
- Don’t disrupt trends.
- Be smart about design and operations.
- Learn to use new technologies and methods.

Research Triangle Institute

- Founded 1959 in North Carolina
- Founding Universities
  - University of North Carolina in Chapel Hill
  - North Carolina State University in Raleigh
  - Duke University in Durham
- Cornerstone of new Research Triangle Park
- One of the select few independent not-for-profit research institutes
- Mission: Improve the human condition by turning science into practice.

RTI INTERNATIONAL

Statistical Leaders at RTI

- James Chromy
- Paul Biemer
- Ralph Folsom
- Phil Kott
- Steve Cohen
- Roy Whitmore
- Vince Iannacchione
- Jill Dever
- Alan Karr
- Karol Krotki
- Consultant Gauri Datta

Current Initiatives in RTI Statistical Research

- Data science
- Nonprobability sampling
- Redirected calling
- Drones and geosampling
- Artificial Intelligence
- Predictive Modeling
- Small Area Estimation
- Address-Based Sampling

My RTI Experience with ABS

- AAPOR Task Force (2016)
  - http://www.aapor.org/Education-Resources/Reports/Address-based-Sampling.aspx
- Sampling lead on ABS surveys
- Literature review on ABS for major client
- Internal R&D on ABS
- Multiple presentations and papers on ABS research
- Ongoing research agenda – more on that from my colleagues

Selected Themes from my RTI Experiences

- Don’t disrupt trends!
- Response rates are dropping.
- Being a survey statistician means understanding all the sources of error and controlling or adjusting for them as best you can.
  - That’s what ABS research is about.

ABS Research – Understanding the sources of error in ABS surveys and how to manage/control them

RTI's ABS Team

- 12 experts across Statistics and Survey Methods groups

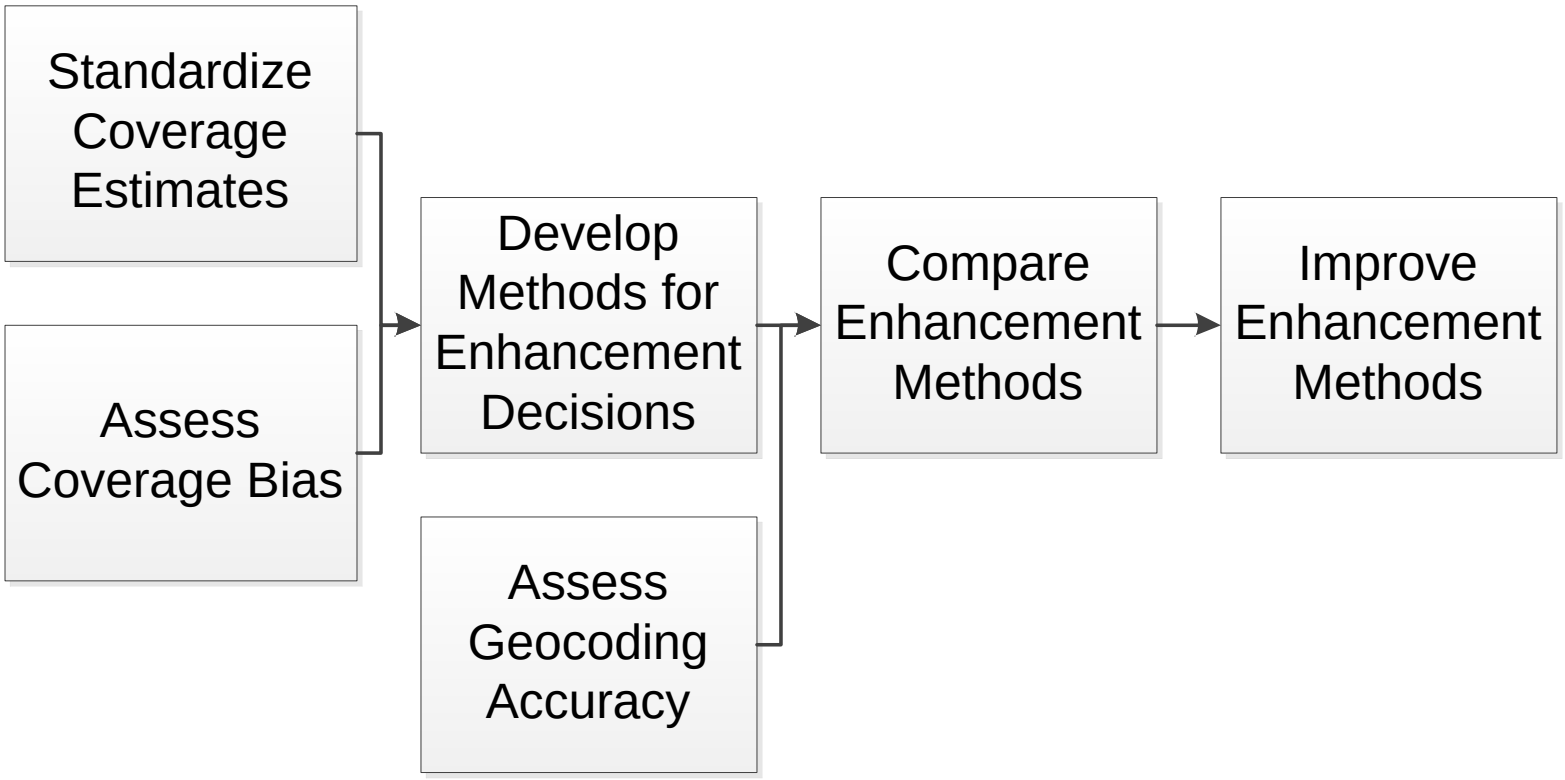
- Our roles are to:
  - Write proposals that call for ABS
  - Estimate productivity prior to data collection
  - Design and draw sample
  - Monitor data collection
  - Develop weights
  - Collaborate with other groups to optimize design
  - **Advance ABS best practices**

Address-Based Sampling (ABS)

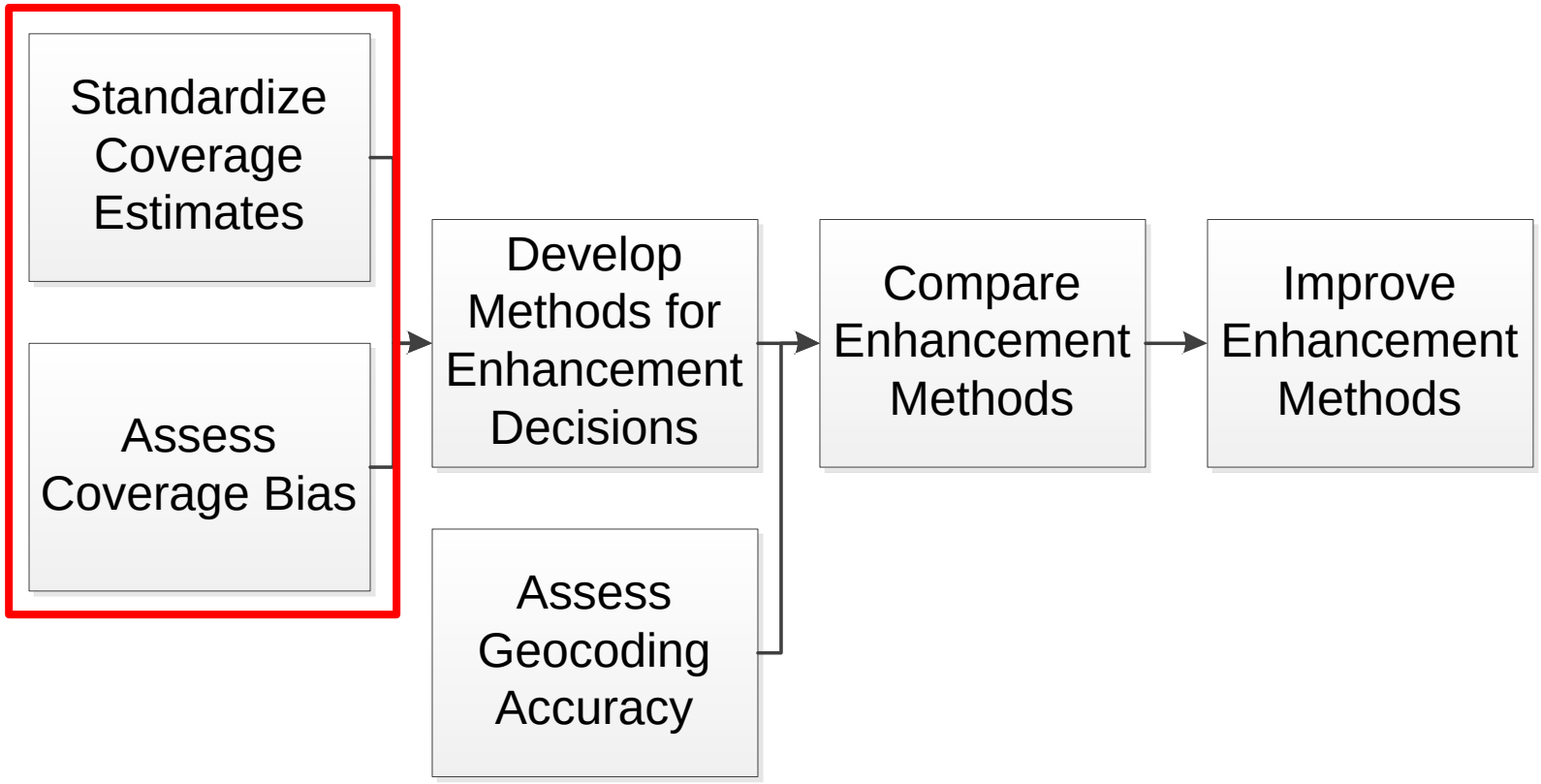
- US Postal Service's Computerized Delivery Sequence File (CDS)
  - Contains all addresses for which USPS delivers mail
    - 90–98% estimated coverage of residential housing units (AAPOR 2016)
    - Names are not included

- Became available to survey researchers to replace in-person frames

ABS Frame Coverage Research

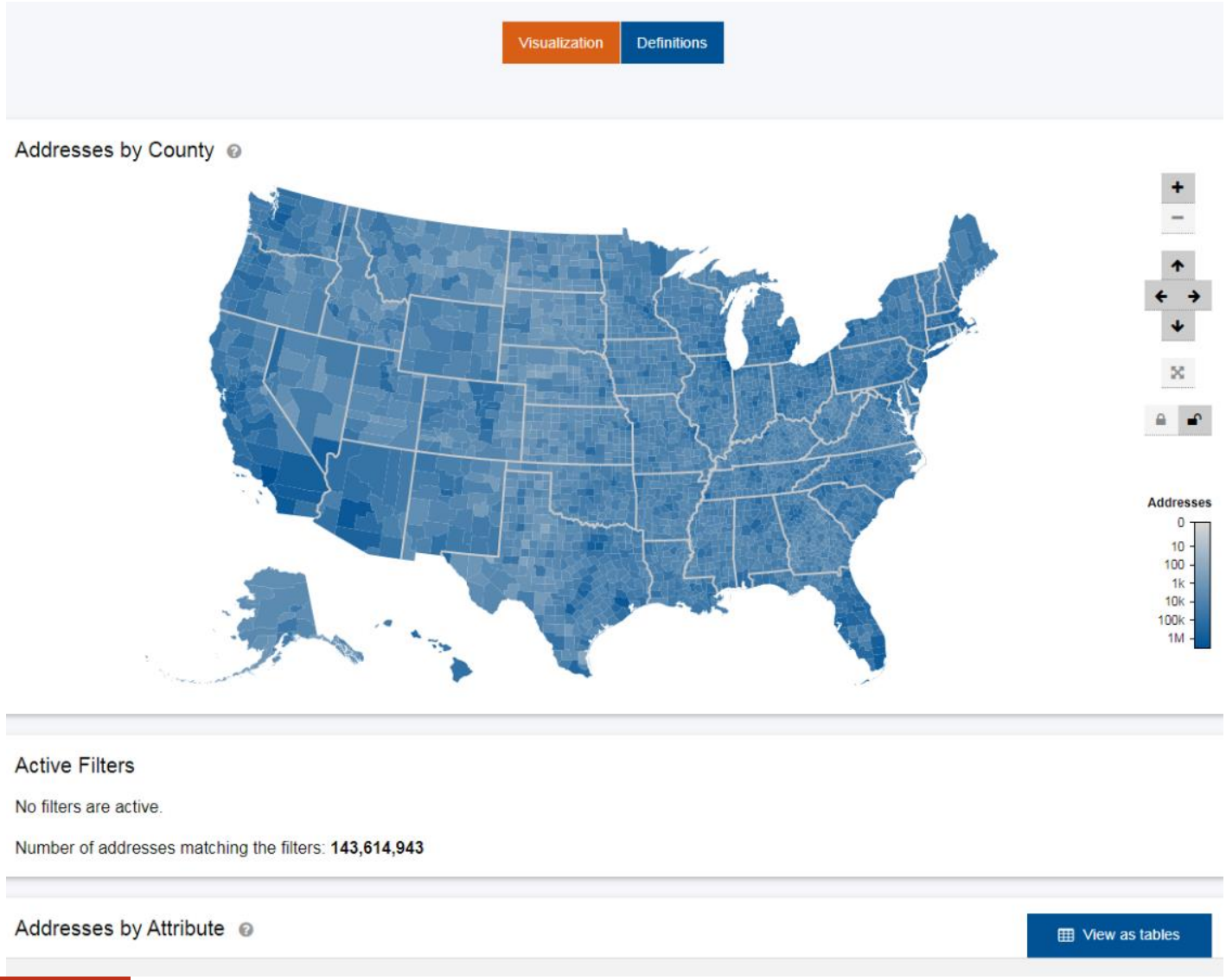


ABS Frame Coverage Research



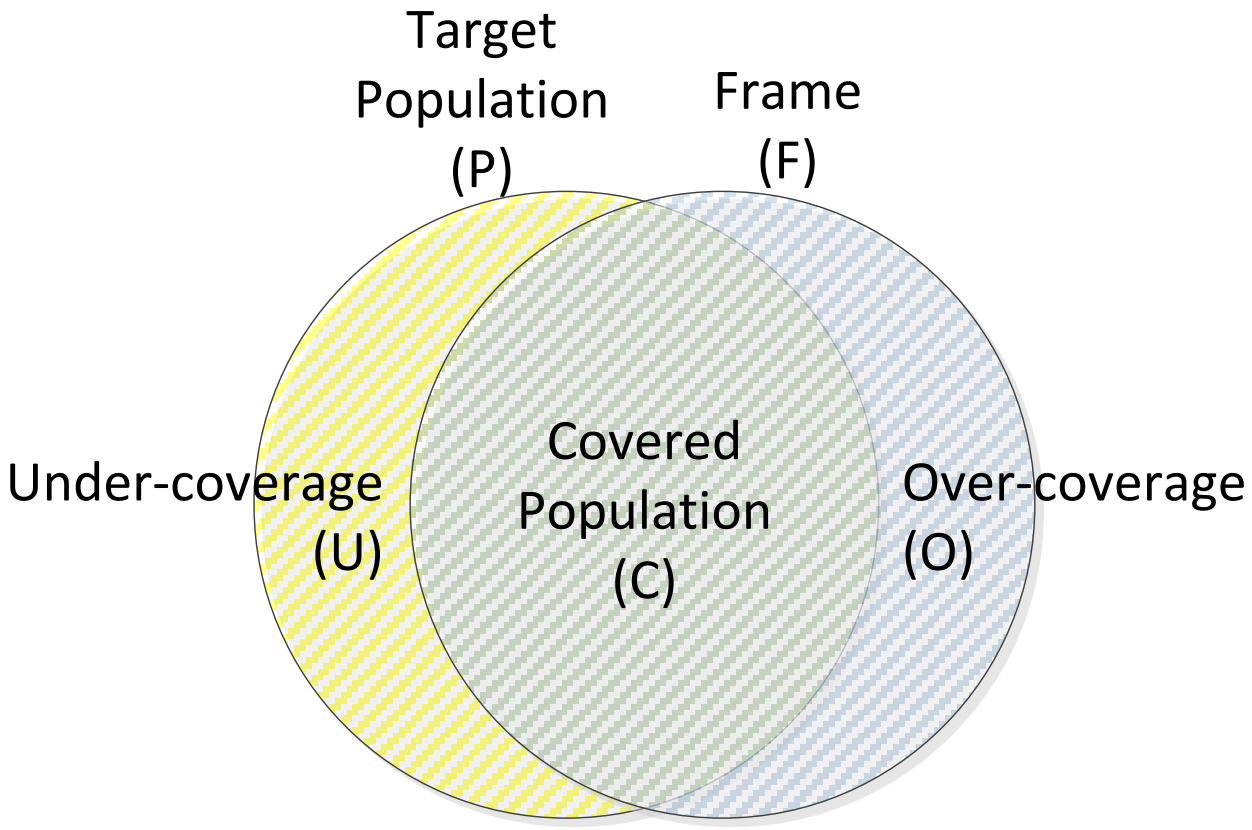
Dissemination

http://abs.rti.org/



What Method Should We Use to Estimate Net Coverage of the ABS Frame?

What is coverage?



Coverage is the extent to which the frame covers the target population.

Known Coverage Problems with ABS Frames

- **Unlocatable addresses** (e.g. PO boxes)
- Group quarters (e.g. college dormitories, shelters)
- Mobile homes
- Simplified addresses
- AIAN lands
- Units within drop points
- Irregular homes
- Vacant and newly constructed homes

- Geocoding error

In general, coverage is more likely to be a problem in rural areas.

General Approach to ABS Frame Construction

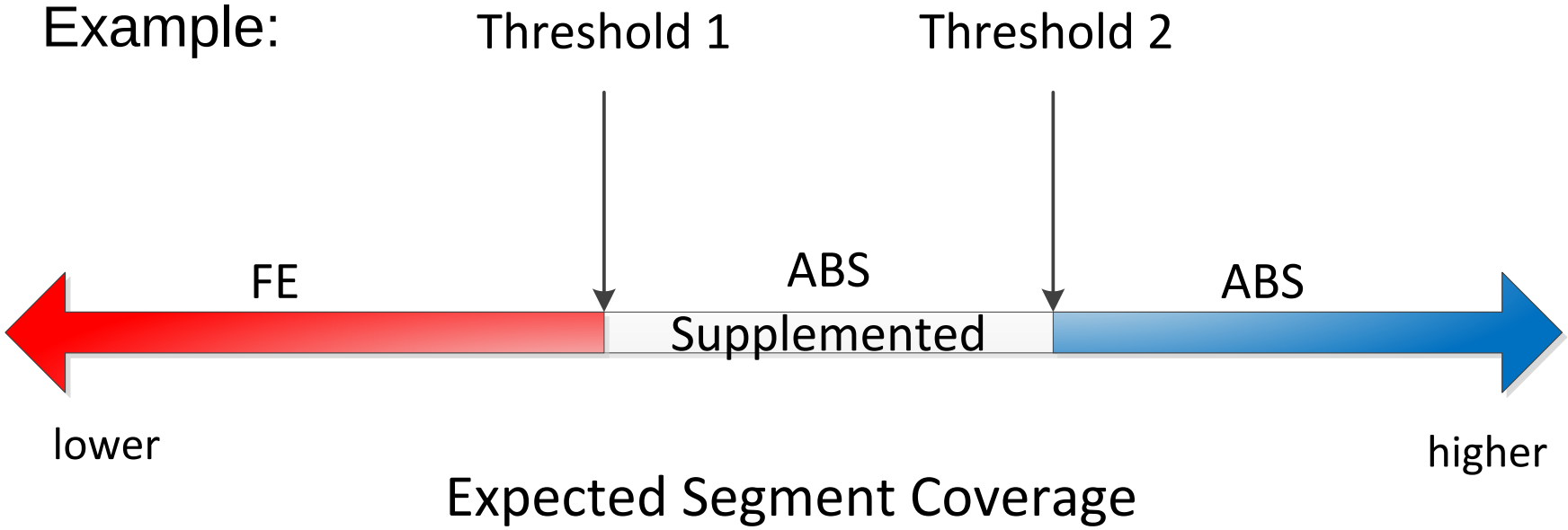
- Goal: create a frame which maximizes coverage and minimizes cost.

- General approach:
  1. Select area probability sample.
  2. Predict ABS coverage for sampled areas (segments).
  3. Select coverage threshold(s).
  4. Determine frame construction method by comparing expected coverage to threshold(s).
  5. Construct frame for each sampled area according to the method determined in Step 4.

Coverage Thresholds

Predict if coverage is above or below a specified threshold value.

Example:



A study typically has 1 or 2 thresholds, depending on the number of frame construction methods.

Predicting Coverage: Models

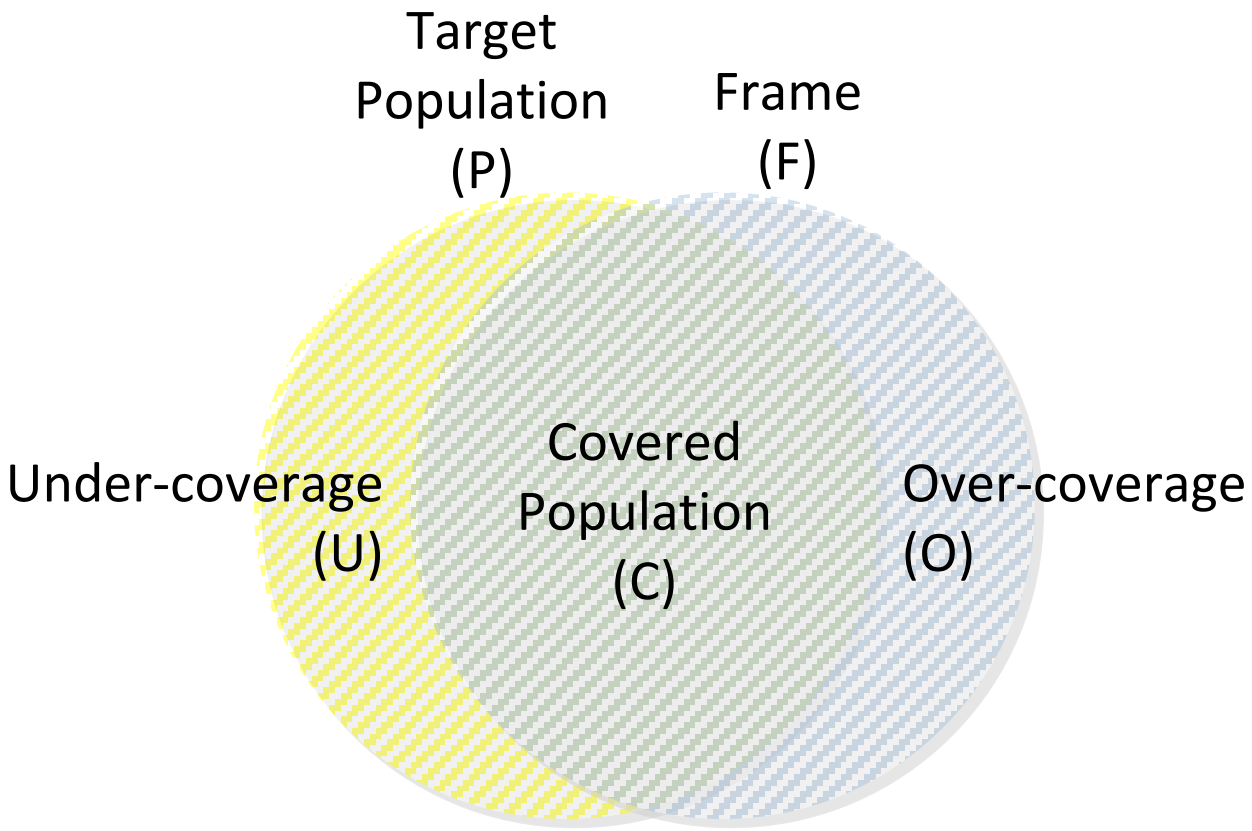
Models for estimating coverage or classifying segments relative to thresholds:

- O'Muircheartaigh et al. (2007, 2009): regression trees.
- Hsu et al. (2010) and Montaquila et al. (2011): linear regression model.

Must have prior FE listings (estimates of truth) to fit models.

Must have auxiliary variables at the segment level.

Predicting Coverage: Net Coverage Ratio



Coverage rate = $\frac{C}{P}$

Net coverage rate = $\frac{F}{P}$

(Kish 1965)

Options for Net Coverage Ratios

- Numerator - ABS frame inclusions/exclusions
- Denominator - Choice of “truth” for comparison

ABS Frame Counts for Net Coverage Numerators

Remove PO boxes that are not an HU's only way to get mail.

Optionally add active addresses from No-Stat file to increase coverage. (Shook-Sa et al. 2013)

Face-to-Face Surveys

- Remove unlocatable addresses (all PO boxes, etc.)

Mail Surveys

- Remove drop points and their units

“True” Counts for Net Coverage Denominators

Source

- Decennial Census (Census)
- American Community Survey (ACS)
- American Housing Survey (AHS)
- Claritas (CL)

Measure

- Total HUs (Tot)
- Occupied HUs (Occ)

2015 Residential Energy Consumption Survey (RECS)

- Purpose - estimate energy costs and usage
- Target Population – Occupied Primary Housing Units (HUs) in U.S.
- Stratified 3-Stage Design
- Within 2nd stage units (segments), create HU frames

3 Types of Segments in RECS

- ABS frame (547 **ABS** segments)

  Threshold 2 = 90%

- ABS frame supplemented with field searches (213 **SUP** segments)

  Threshold 1 = 56%

- Field Enumerated frame (40 **FE** segments)

ABS frames available for all 3 types of segments, even if used only to determine how to classify each segment.

Analysis 1: How well did our net coverage estimates categorize segments for RECS?

- For FE segments (initially estimated to have 56% net coverage or less), match addresses found in the field to the ABS frames.
  - Assumes FE listings are “truth”, which is not necessarily a good assumption. (Eckman and Kreuter 2013, Cunningham et al. 2006)

- Summarize match rates, undercoverage, and overcoverage.

RECS Address Match Rates, 40 FE Segments

Coverage and Match Rates for 40 FE Segments Combined		
Class of HUs	**Count**	**Percent**
Matched Addresses (coverage lower bound)	2,839	27%
FE Matched outside segment (geocoding error)	1,762	17%
FE Only (other undercoverage and match error)	5,844	56%
Total FE HUs	10,445	100%
ABS only (overcoverage upper bound)	2,219	21%
Total ABS (net coverage, coverage upper bound)	5,058	48%

Analysis 2: Does choice of net coverage ratio matter?

- Compute various net coverage ratios for all 800 RECS segments.
  - Numerators:

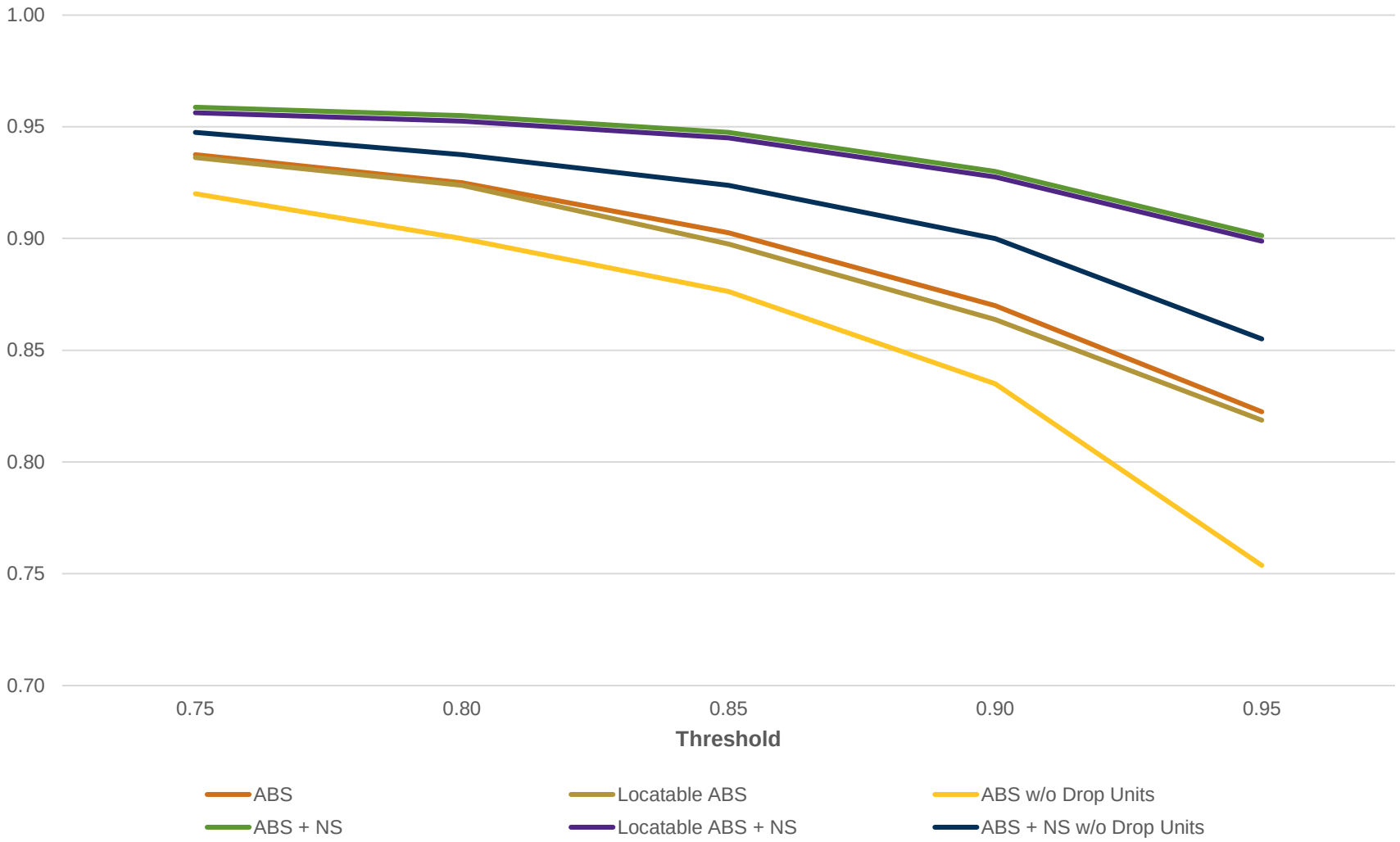
ABS	Locatable ABS	ABS - Drop Units
ABS + NS	Locatable ABS + NS	ABS + NS - Drop Units

  - Denominators:

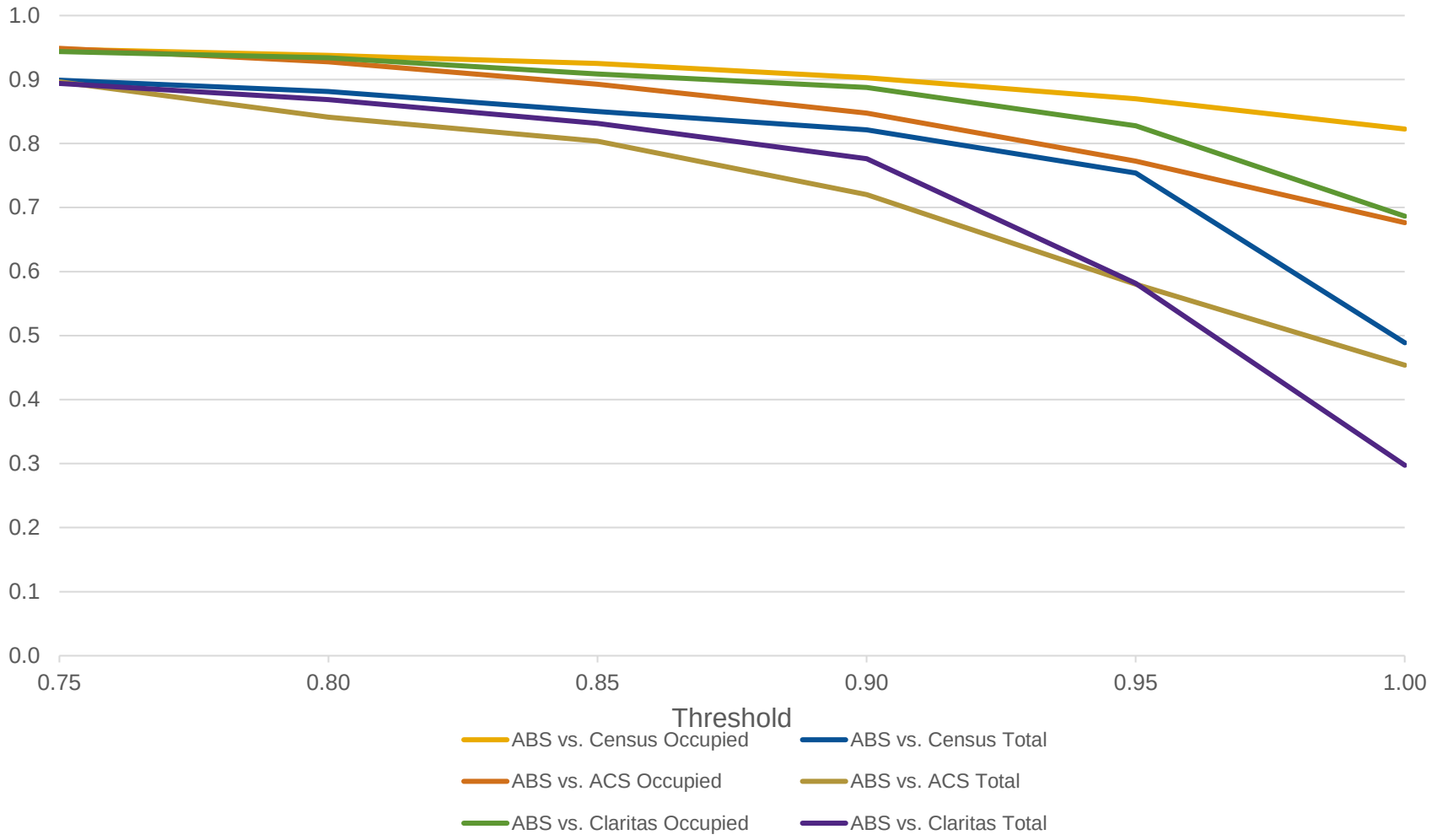
Census Occupied	ACS Occupied	Claritas Occupied
Census Total	ACS Total	Claritas Total

- Compare classifications for different threshold values.

Proportions of Segments Exceeding Threshold: Ratio = Various Numerators / Census Occupied



Proportions of Segments Exceeding Threshold: Ratio = ABS / Various Denominators



Summary: Steps in Estimating Coverage or Net Coverage of ABS Frames

1. Determine ABS inclusions/exclusions.
2. Determine whether occupied HUs or total HUs is more appropriate for estimating “truth.”
3. Select best source of “true” HU counts based on level of geography, time period, and cost.
4. Compute net coverage ratios (or use a coverage prediction model to estimate coverage).

Does Undercoverage translate to bias?

Our Question and Goal

- Research Questions
  - Does undercoverage lead to bias?
  - What is the risk of coverage bias when using the USPS CDS in a face-to-face survey?
- Goal
  - Inform decisions on whether to
    - Use the ABS frame for a given survey, and/or
    - Enhance the ABS frame (e.g., listing some segments, using supplementation in some segments)

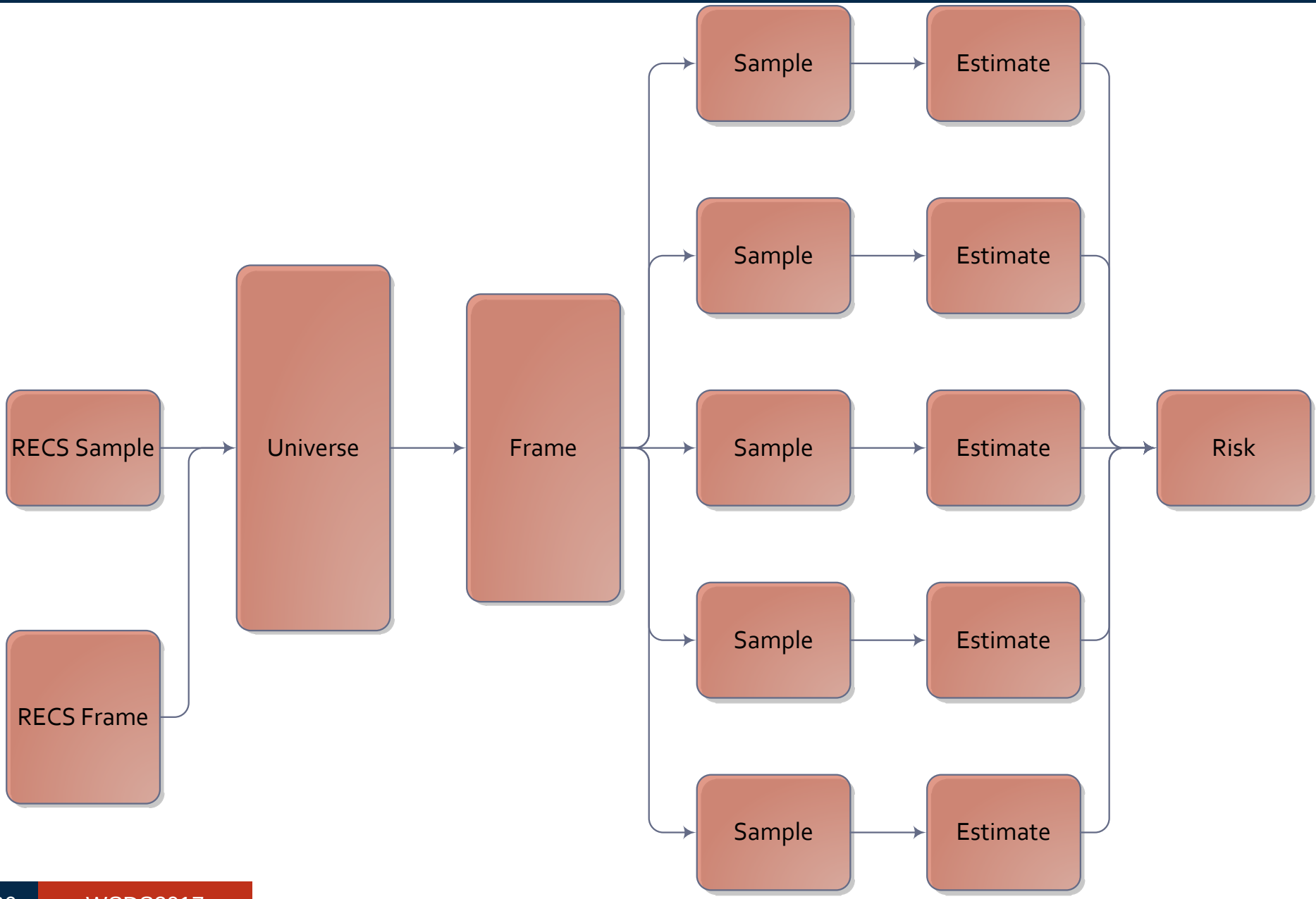
Coverage Bias

- Three studies have assessed the impact of undercoverage on bias
  - English et al (2011)
    - Fertility in Cumberland, Maine
  - Morton et al (2010)
    - Substance abuse with small uncovered counts
  - Eckman & Kreuter (2013)
    - Fertility, health, sexuality, and demographics of two list frames (not the CDS)

Common Practice

- Areas are ranked by coverage
- Areas with lower coverage use frame enhancement or listing instead of ABS frame only
- Typically budget dictates the number of segments that can be enhanced and/or listed rather than optimization
- Only coverage rate is used to prioritize which segments to choose – not other characteristics of segment

Overview of simulation



Data sources for basis of simulation

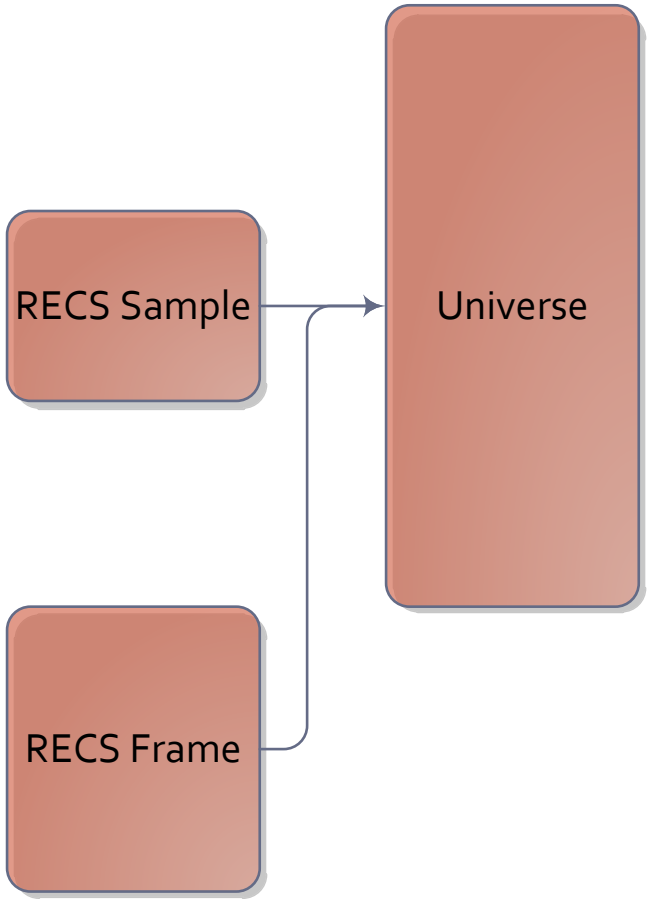
RECS Sample

- 12 demographic and building characteristics

RECS Frame

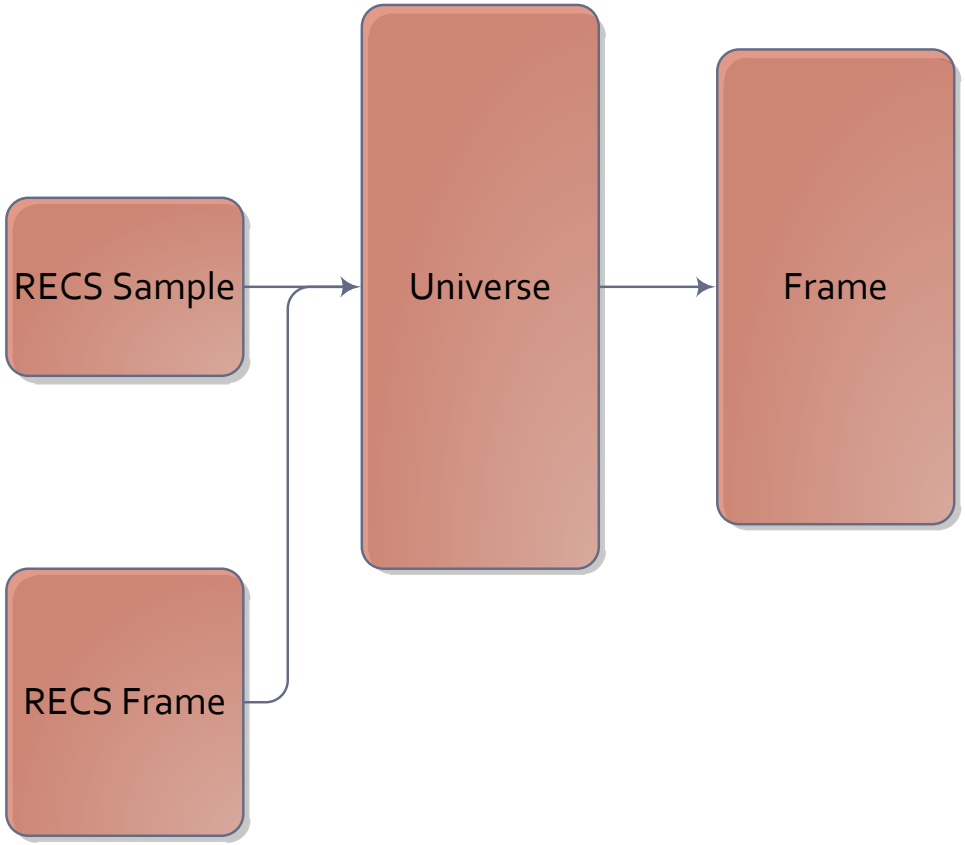
- 800 Census block groups across the US
- 579,459 CDS addresses
- 6,841 enumerated addresses

Construct a universe



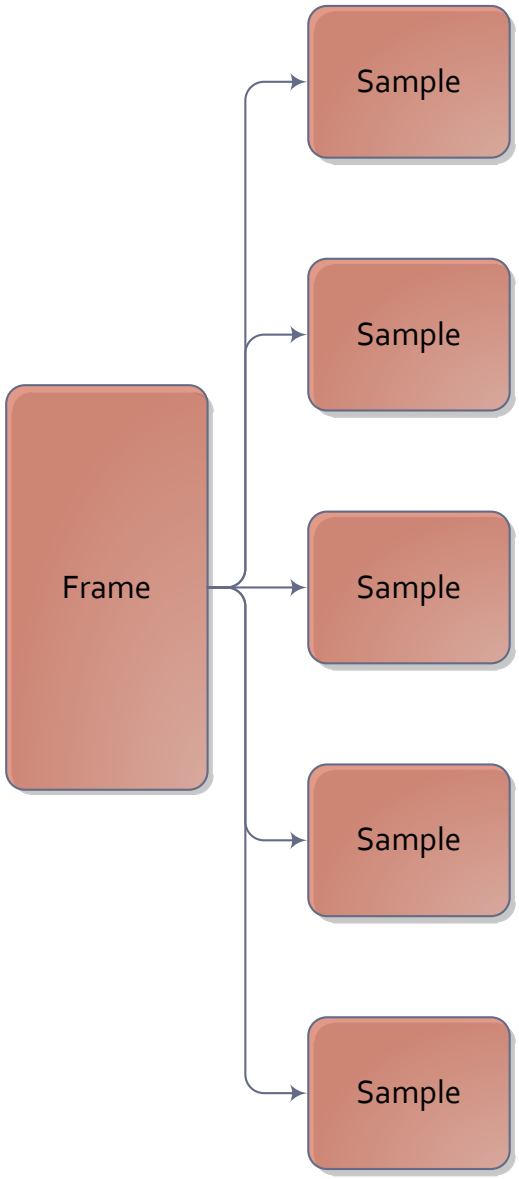
- Created one universe
- Replicated cases from the RECS survey by their final weights
- Used frame information and appended ACS data to assign coverage propensities from model built using RECS data

Made multiple frames



- Created frames ranging in coverage 1-100% (n=100)
- Frame constructed at the x% coverage rate consisted of units with modeled propensity in the x percentile and higher

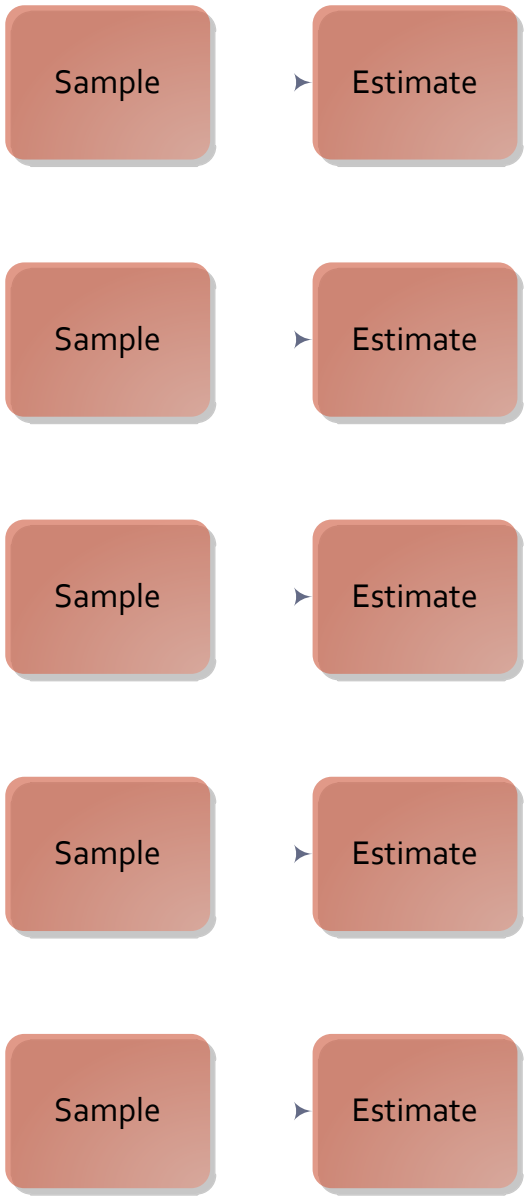
Generate multiple samples from frames



- For each frame
  - Drew 1,000 samples of 1,000 addresses per sample
- 2-stage design
  - 200 PSUs
  - Proportional allocation

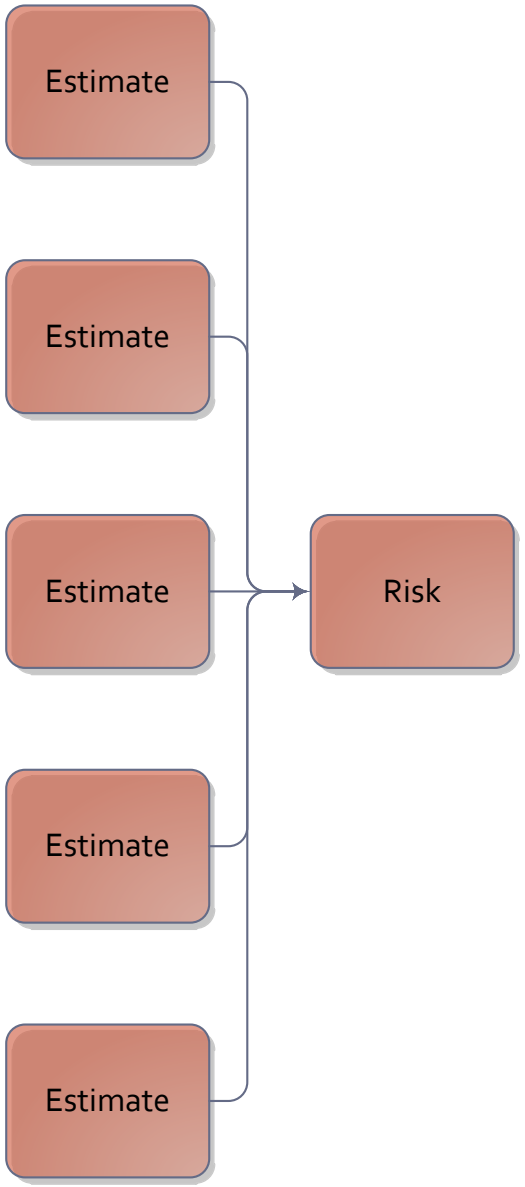
Calculate estimates on samples

- For each sample
  - Calculated the proportion/mean of each variable
  - Calculated bias compared to the universe
    - Bias ($\hat{\theta} - \theta$) and relative bias ($\frac{\hat{\theta}-\theta}{\theta}$)
    - Z-test for significance (using appropriate variance estimate)

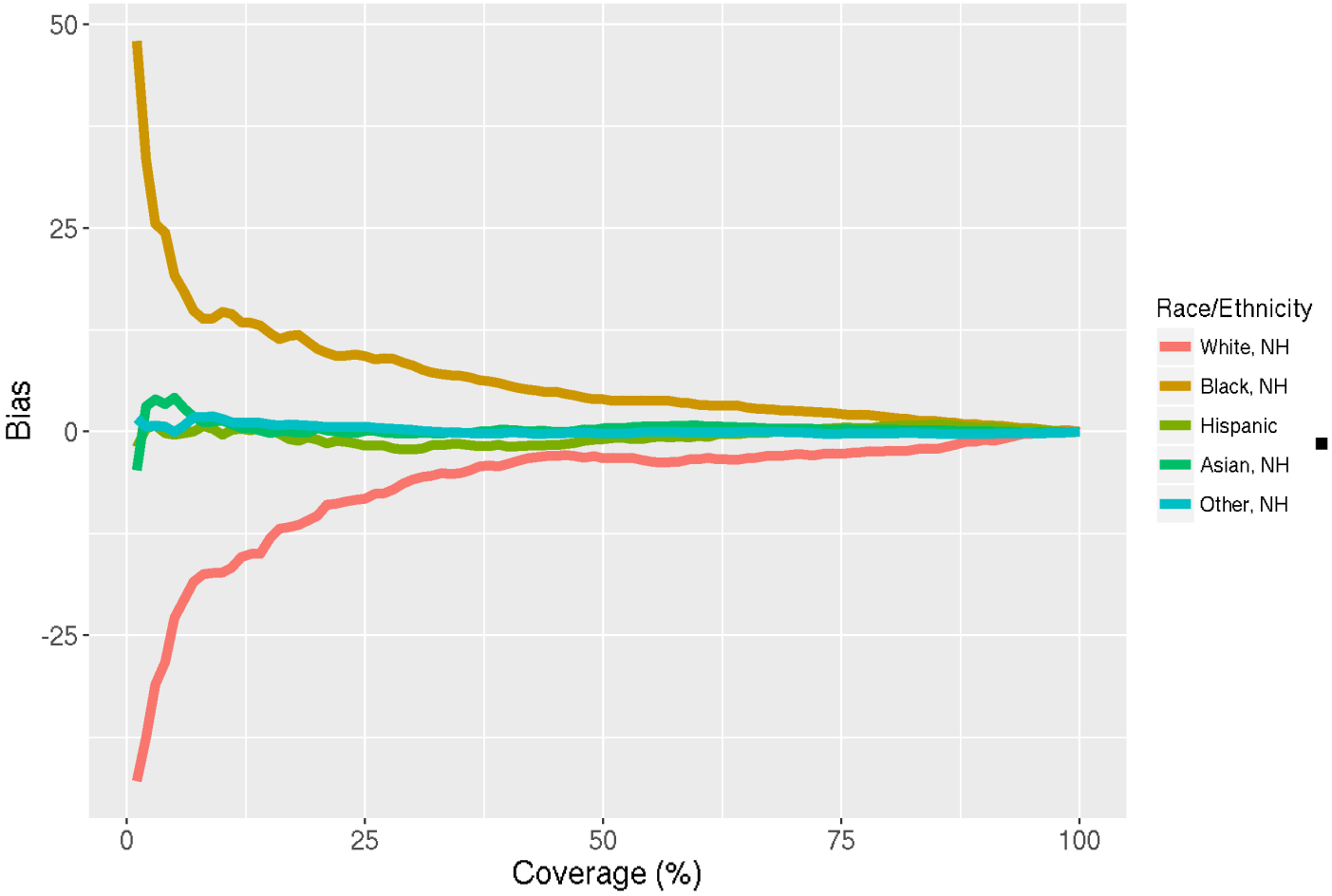


Calculate risk from samples

- For each level of coverage,
  - Risk is the proportion of samples for which the statistic was significantly different from the parameter
  - This is expected to be 5% at 100% coverage

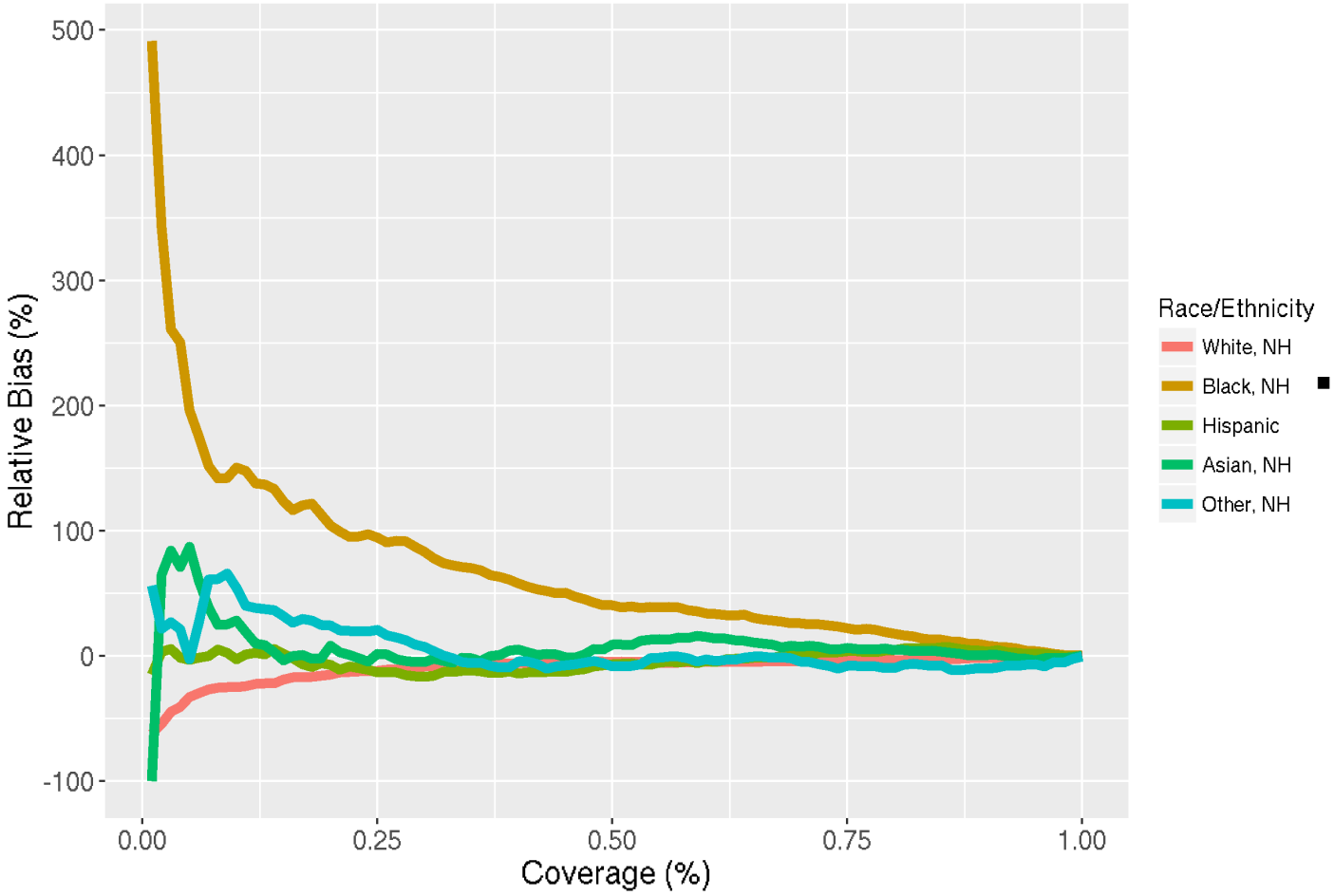


Race/ethnicity bias



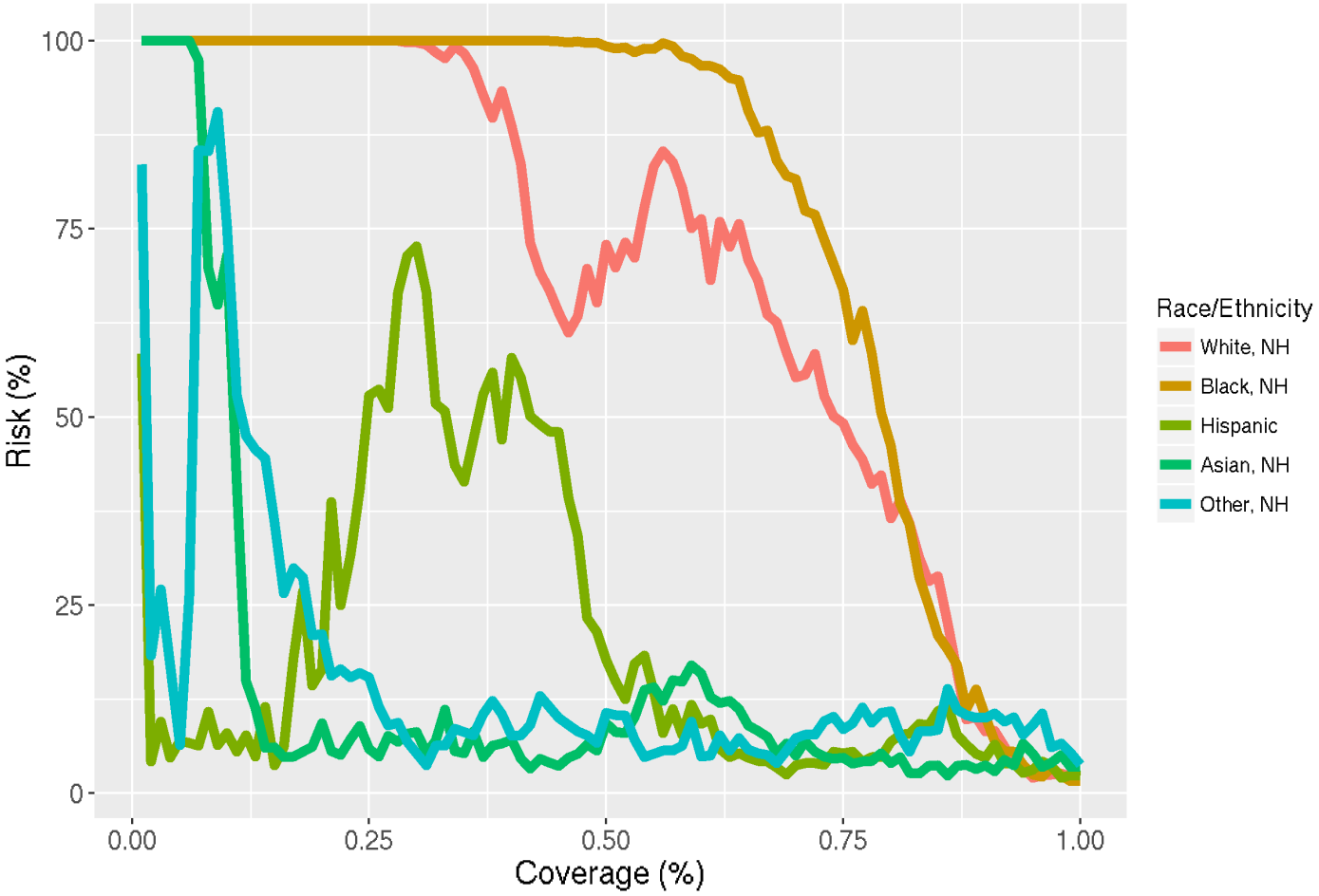
- As coverage decreases, over estimate black, NH and underestimate White, NH
- Other race/ethnicities aren't as affected until coverage gets low

Race/ethnicity relative bias



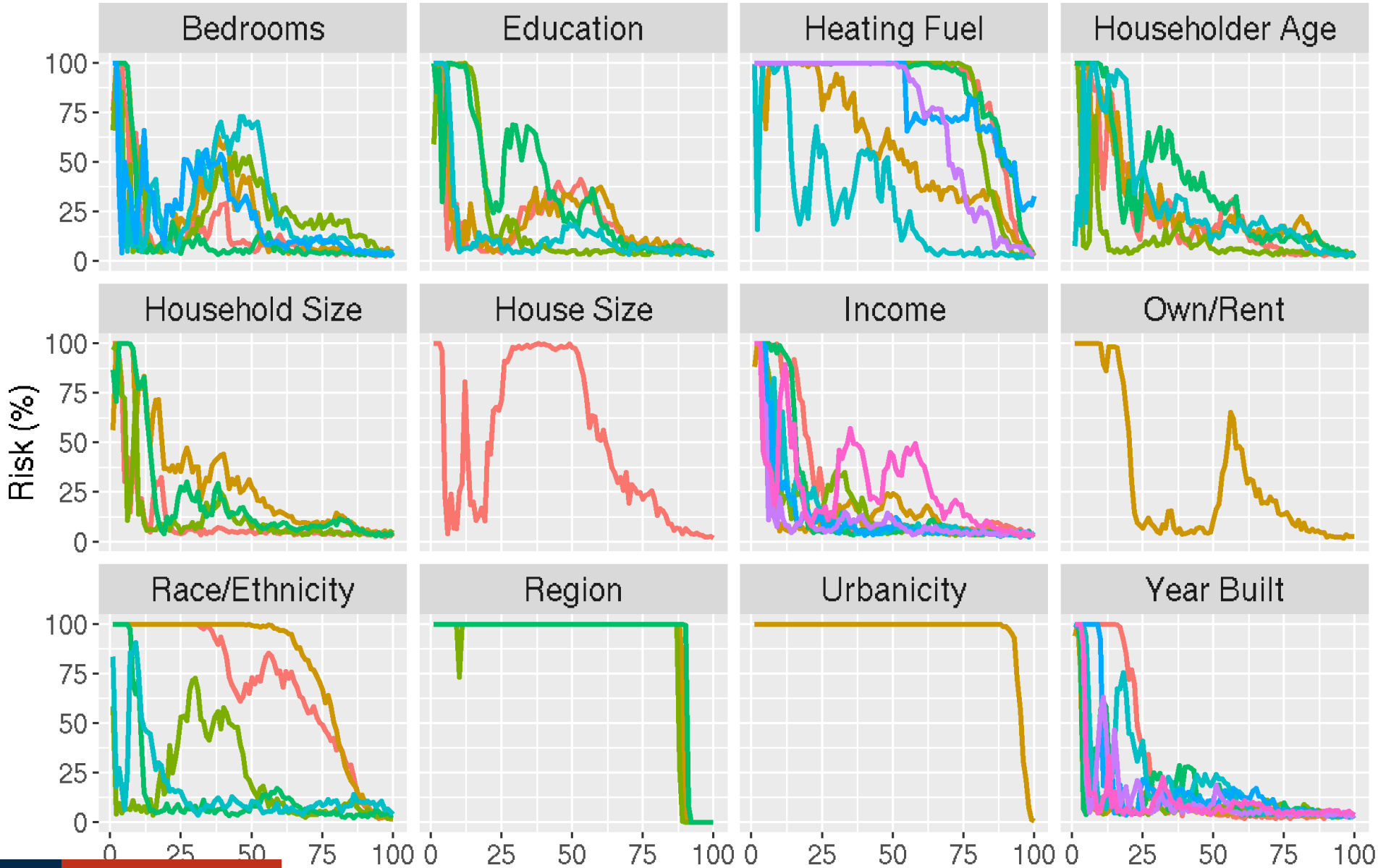
- Relative bias increases quickly as coverage declines
- Race is somewhat geographic and associated with urbanicity which is associated with coverage

Race/ethnicity risk



- Risk increases quickly for black and white

Risk for all variables



Summary

- What is the risk of coverage bias when using the USPS CDS in a face to face survey? It depends!
  - Variable of interest
  - Level of coverage
  - Area of interest (shown in paper, not this presentation)

Next steps

- Replicate
  - RECS frame isn't true universe. What about areas where RECS frame wasn't enhanced?
  - RECS is an energy survey – not applicable to most social science research, how do those variables behave?
- Can post-stratification or other weighting adjustments correct the risk?

More Information

Ashley Amaya
aamaya@rti.org

Rachel Harter
rharter@rti.org

Katie Morton
bowman@rti.org

Stephanie Zimmer
sazimmer@rti.org

http://abs.rti.org/

See us at the RTI Booth

Additional References

- American Association for Public Opinion Research (2016). Address-based sampling. Report prepared for AAPOR Council by the Task Force on Address-based Sampling; R. Harter, Chair. Oakbrook Terrace, IL: AAPOR.
- Battese, G. E., Harter, R. M., & Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association,* **83**(401), 28–36.
- Cochran, W.G. (1942) Sampling theory when the sampling-units are of unequal sizes. *Journal of the American Statistical Association*, **37**, 199-212.
- Cochran, W.G. (1946) Relative accuracy of systematic and stratified random samples for a certain class of populations. *Annals of Mathematical Statistics*, **17**, 164-177.
- Cochran, W.G. (1953). *Sampling Techniques*. John Wiley & Sons, New York.
- Deming, W.E. (1950). *Some Theory of Sampling*. John Wiley & Sons, New York.
- Fuller, W. A., & Harter, R. M. (1987). The multivariate components of variance model for small area estimation. In R. Platek, J. N. K. Rao, E. E. Särndal, & M. P. Singh (Eds.), *Small area statistics: An international symposium*. New York: John Wiley & Sons.
- Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Survey Sampling Methods and Theory*. John Wiley & Sons, New York.

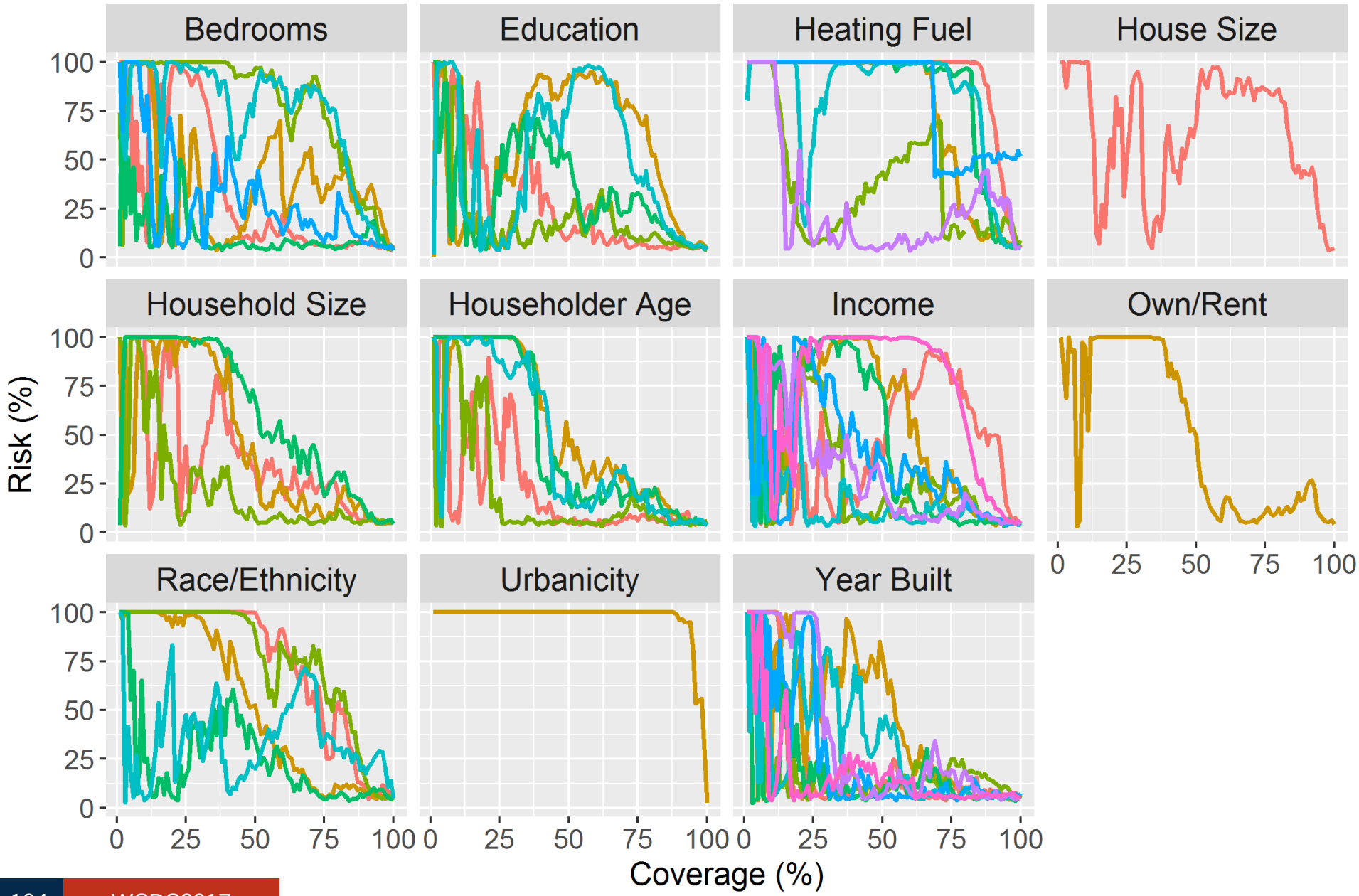
References (cont.)

- Hansen, M.H. and Hurwitz, W.N. (1943) On the theory of sampling from finite populations. *Annals of Mathematical Statistics,* **14**, 333-362.
- Horvitz, D.G. and Thompson, D.J. (1952) A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association,* **47**, 663-685.
- Iannacchione V., Staab, J., & Redden, D. (2003). Evaluating the use of residential mailing lists in a metropolitan household survey. *Public Opinion Quarterly*, **67**(2), 202-210.
- Jessen, R.J. (1942). *Statistical Investigation of a Sample Survey for Obtaining Farm Facts.* Agricultural Experiment Station, Iowa State College of Agriculture and Mechanic Arts, 304, Ames, IA.
- Madow, W.G. and Madow, L.H. (1944) On the theory of systematic sampling. *International Annals of Mathematical Statistics*, **15**, 1-24.
- Neyman, J.(1934) On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97 (4), 557–625.
- Royall, R.M. (1970). On finite population sampling theory under certain regression models. *Biometrika*, **57**, 377-387.
- Tippet, L.C. (1927). Random sampling numbers. *Tracts for computers*, **XV**, Harvard University Press, Cambridge, MA.

Extra slides – time allowing

What if we were sampling only in the west?

Risk of coverage bias in West



What if we were sampling only in the rural areas?

Risk of coverage bias in Rural Areas