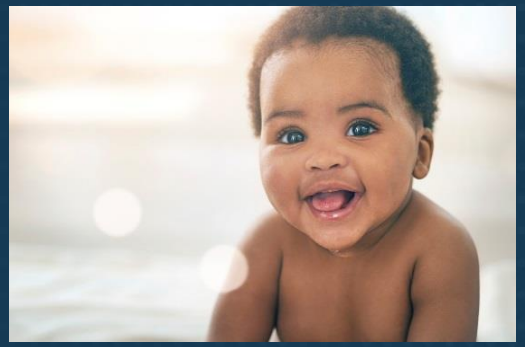
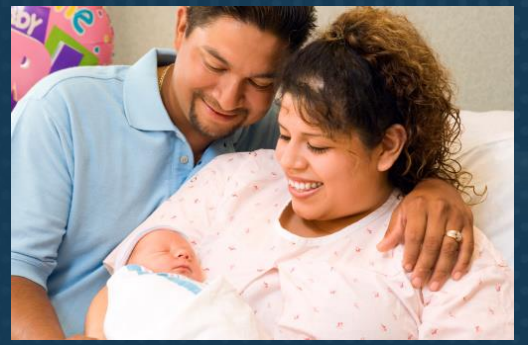
MEASURING NONRESPONSE BIAS IN PRAMS, 2019

HOLLY SHULMAN, MA; PHIL HASTINGS, PHD; JOE PIROZZOLO, PHD

THE FINDINGS AND CONCLUSIONS IN THIS REPORT ARE THOSE OF THE AUTHORS AND DO NOT NECESSARILY REPRESENT THE OFFICIAL POSITION OF THE CENTERS FOR DISEASE CONTROL AND PREVENTION.


Centers for Disease Control and Prevention

National Center for Chronic Disease Prevention and Health Promotion

Division of Reproductive Health

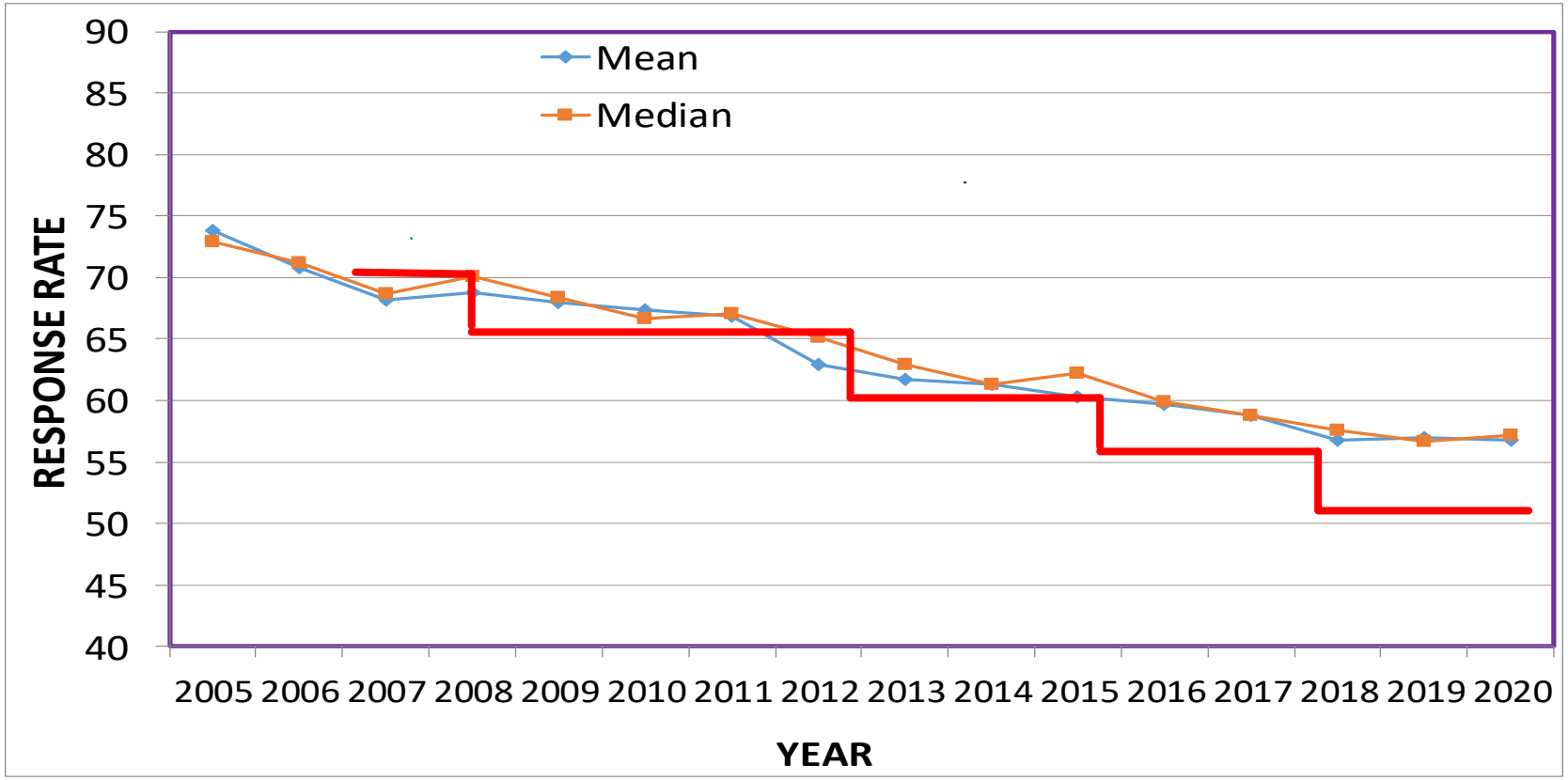
OVERVIEW

- Background
- Justification for Conducting Ongoing Nonresponse Bias (NRB) Analyses
- Previous PRAMS NRB Analyses
- NRB Analyses
  - True Bias
  - Level of Effort and Incremental Bias
  - Model-based Approaches
- Discussion and Next Steps

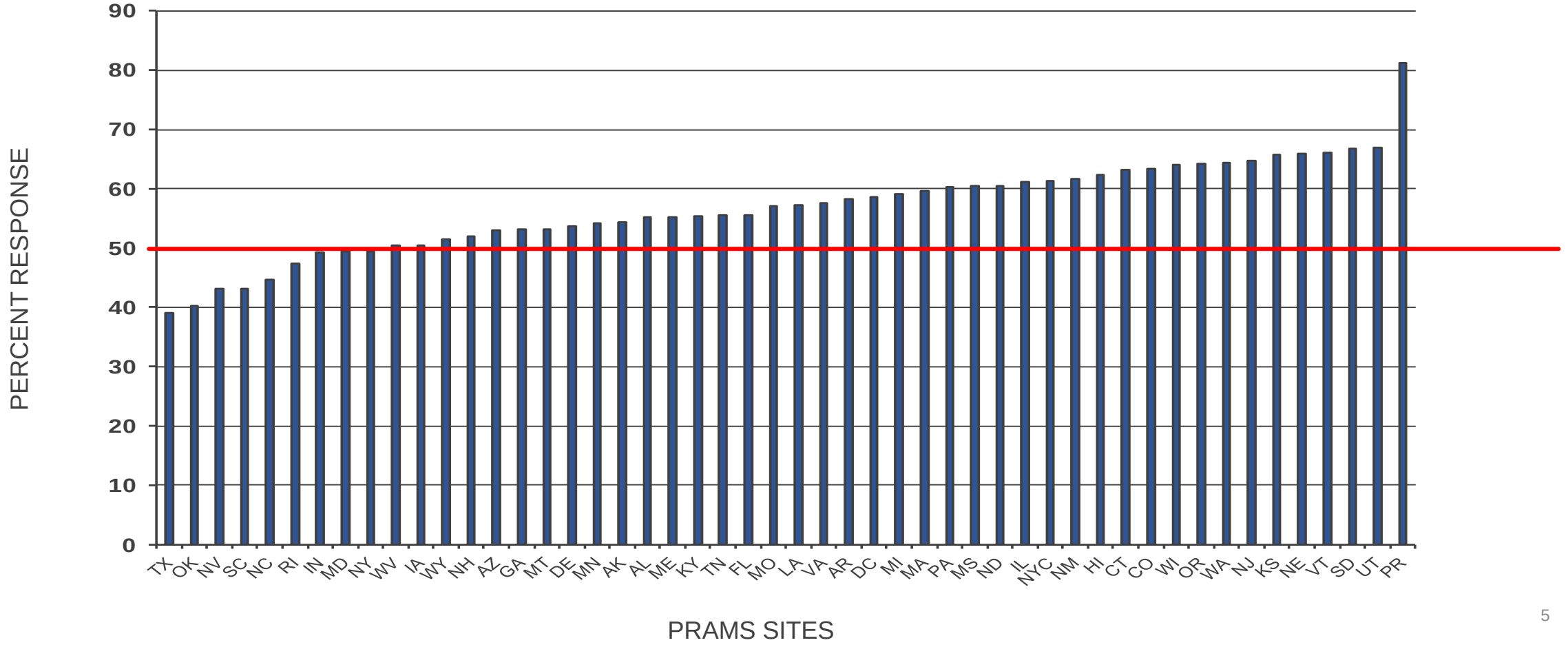
BACKGROUND: RESPONSE RATE THRESHOLDS FOR DATA RELEASE

- PRAMS currently implements response rate thresholds based on what could reasonably be achieved within survey climate using PRAMS protocol
- Data released and included in CDC reports if jurisdiction reaches or exceeds threshold:
  - Before 2007 – 70%
  - 2007 to 2011 – 65%
  - 2012 to 2014 – 60%
  - 2015 to 2017 – 55%
  - 2018 to 2020 – 50%

PRAMS MEAN AND MEDIAN WEIGHTED RESPONSE RATES BY YEAR, 2005 – 2020



PRAMS WEIGHTED RESPONSE RATES BY JURISDICTION, 2020



PURPOSE OF NONRESPONSE BIAS ANALYSES

- Survey response rates steadily falling, with more rapid declines as of late
  - Higher response rate *less likely* to have NRB
  - Lower response rate *not always* indicative of NRB

- 2006 Office of Management and Budget (OMB) Directive
  - Mandated NRB assessment for federal surveys with <80% response rate

- Continually lowering PRAMS response rate threshold unsustainable

- *Systematic Review of Nonresponse Bias Studies in Federally Sponsored Surveys.* FCSM 20-02. Federal Committee on Statistical Methodology. March 2020.
- *Standards and Guidelines for Statistical Surveys.* OMB. 2006

PREVIOUS NONRESPONSE BIAS ANALYSES

Research Questions

- What differences arise in prevalence estimates subject to different response rates?
- Are early respondents different from more difficult-to-reach respondents?

Data Examined

- 2002 data from IL (83% response rate), OK (84% response rate) and VT (86% response rate)
- 8 survey indicators

APPROACH

- Identify states with high response rates
- Based on time of survey completion, order respondents from earliest to latest by batch
- Create response rate groupings of 50%, 60%, 70%, and 80+%
- Weight each response group separately
- Compute and compare weighted estimates for each response group
- Compare demographic distributions of response groups to overall population distributions

RESULTS

- Not much evidence of bias, even at lowest response rate examined at 50% response rate
- From 70% threshold perspective, comparing 70% group to 60% group yielded minimal differences in estimates
- Results varied by site and within site by indicators – not generalizable to other sites or other years
- In other analyses using all sites, we estimated about a 1 percentage point change in estimates per 10% decline in response rates

NEW NON-RESPONSE BIAS EFFORTS IN PRAMS

1. *True bias analysis*: comparing weighted estimates with actual population values for selected birth certificate indicators

2. *Level of effort analysis*: simulating response groupings and examining incremental bias for different response levels

3. *Model-based analysis*: estimating worst-case bias under different missing data assumptions

Each analysis examines bias through a different lens, but collectively may provide actionable results to inform PRAMS data release policies

TRUE BIAS ANALYSIS - RESEARCH QUESTION

- What is the NRB found for indicators available on the birth certificate, when calculating estimates based on PRAMS survey respondents as compared with the population (i.e., true values)?

- Examine NRB by indicator type (sensitivity) and by response rate of site

APPROACH

- Jurisdiction vital records birth file serves as sampling frame and source of population information
- Linked birth file with PRAMS sample for 47 PRAMS sites (3 sites did not grant permission for access to their full birth file) to compute:
  - Population values
  - Estimates of the population values from the full PRAMS sample (using sampling weights only)
  - Estimates of the population values from PRAMS respondents (using analysis weights)

BIRTH CERTIFICATE VARIABLES EXAMINED

Health Behaviors	Medical	Demographic/Socio-Economic Status (SES)
Adequate prenatal care (PNC)	Gestational diabetes	Medicaid as payment source for delivery
Smoking before pregnancy	Gestational high blood pressure (hypertension)	WIC participant
Smoking during pregnancy	C-Section delivery	Previous live birth (LB)
Breastfed in hospital	Infertility treatment	
	Pre-pregnancy Body Mass Index (BMI) - Normal	

Note: WIC is the Special Supplemental Nutrition Program for Women, Infants, and Children

COMPARISON OF RESPONDENT ESTIMATES TO POPULATION VALUES

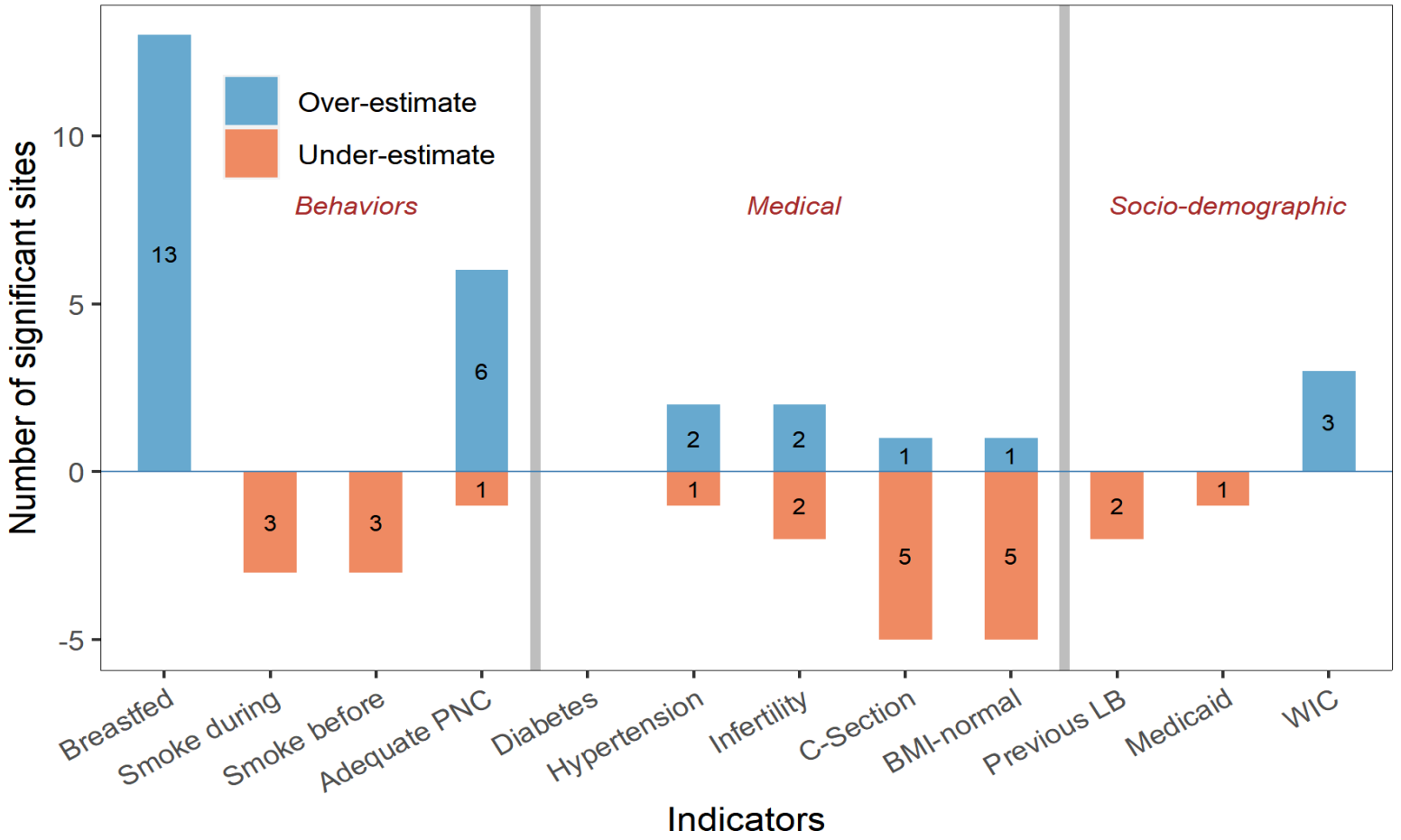
- Produce 95% confidence interval (CI) for estimates of each indicator from PRAMS respondents in each jurisdiction
- Identify instances where 95% CI excludes true population value
- Compute bias and absolute relative bias

$$Bias_{NR}(\overline{Y}_r) = \overline{Y}_r - \overline{Y}_f$$

$$Absolute\ Relative\ Bias_{NR}(\overline{Y}_r) = \left| \frac{100\% * (\overline{Y}_r - \overline{Y}_f)}{\overline{Y}_f} \right|$$

where $\overline{Y}_f$ is the frame value and $\overline{Y}_r$ is the respondent weighted estimate

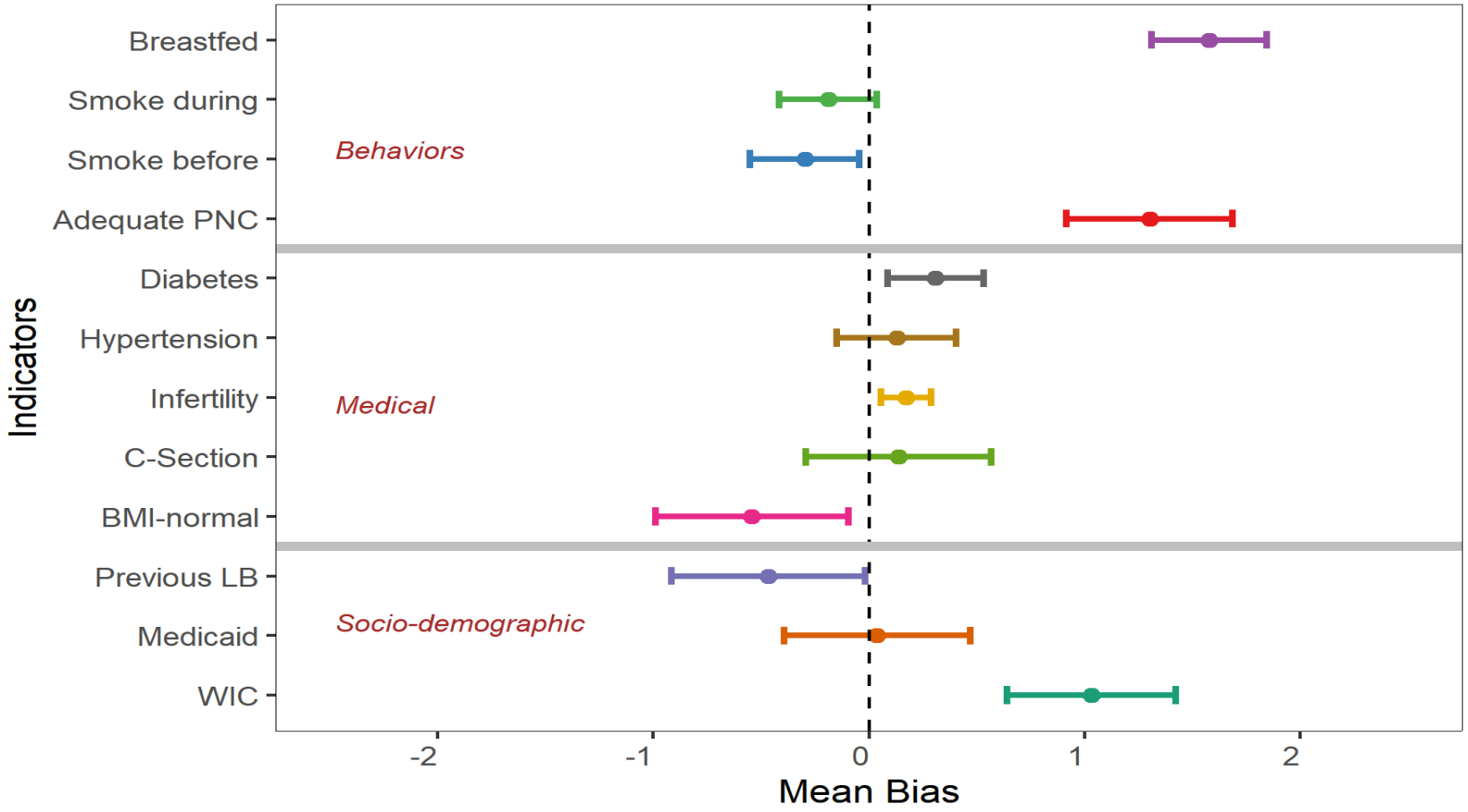
COMPARISON OF RESPONDENT ESTIMATES TO POPULATION VALUES - NUMBER OF SIGNIFICANT DIFFERENCES (47 SITES)



MEAN BIAS BY INDICATOR

- Averaged bias over all 47 sites to produce the mean bias for each indicator
- For each indicator, conducted a fixed-effects meta-analysis with each site treated as a study to account for different sample sizes across sites
  - Effect size measure for the meta-analysis was the (logged) risk ratio, comparing "risk" (prevalence) of reporting in the weighted PRAMS sample to prevalence in the population at each site
  - Allows us to estimate a confidence interval around the mean bias

MEAN BIAS ACROSS SITES BY INDICATOR



IMPACT OF RESPONSE RATES ON BIAS

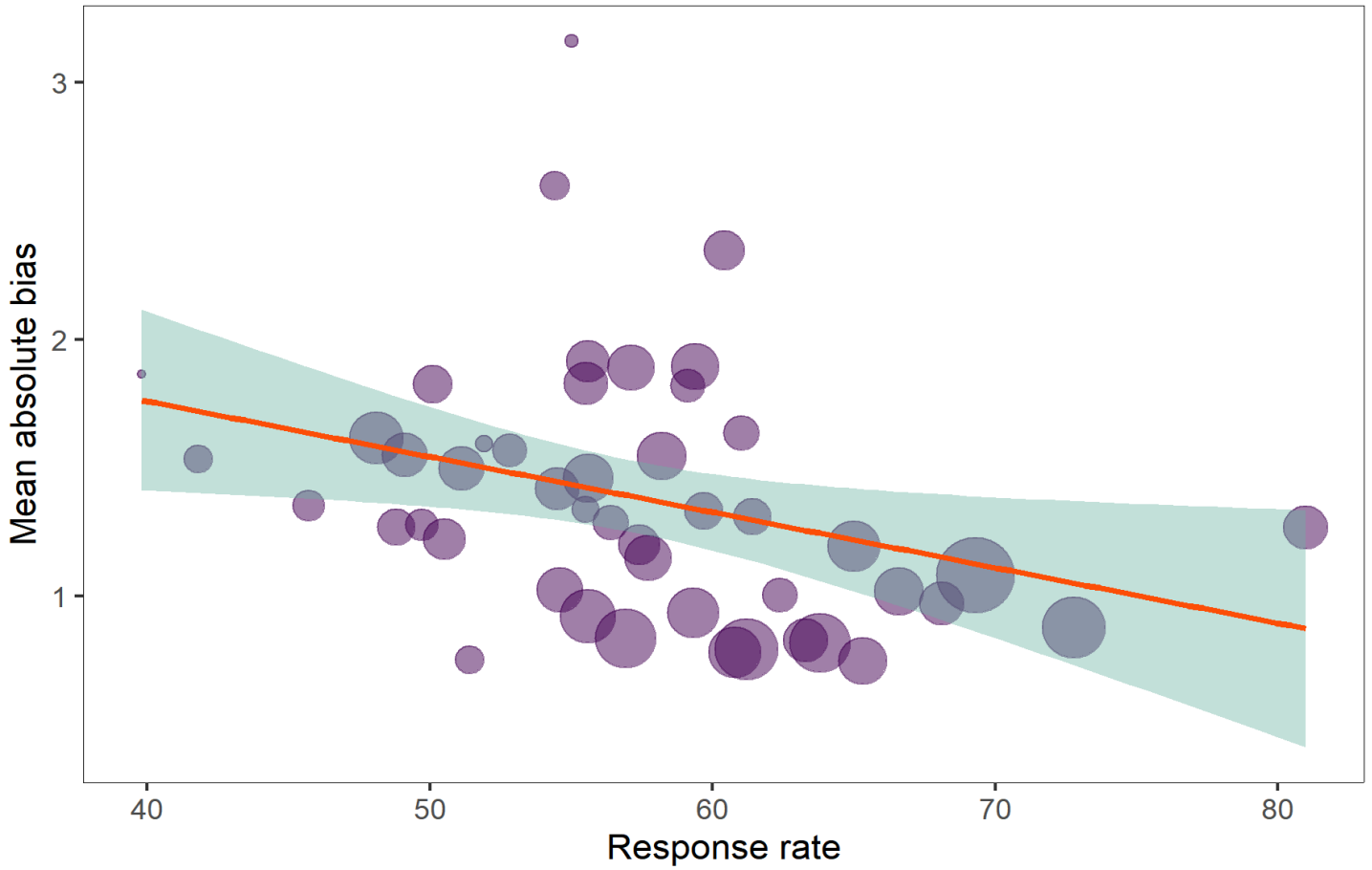
- Calculate mean absolute bias (MAB) across all 12 indicators for each site

$$Mean\ Absolute\ Bias_{NR}(\overline{Y}_r) = \sum_{i=1}^{12} |\overline{Y}_{ri} - \overline{Y}_{fi}| \ /\ 12$$

where $\overline{Y}_{f\,\mathrm{i}}$ is the frame value for indicator i and $\overline{Y}_{r\,i}$ is the respondent weighted estimate for indicator i

- Compute least squares regression line for MAB as predicted by response rate

RELATIONSHIP BETWEEN MEAN ABSOLUTE BIAS AND RESPONSE RATES



Least Squares Regression Line (b = -0.021, SE = 0.009, t = =2.35, p = .023)

Correlation between response rate and bias = -0.33

COMPARISON OF RESPONDENT ESTIMATES, FULL SAMPLE ESTIMATES, AND POPULATION VALUES

- Calculate bias and absolute bias for both the full sample estimates (i.e., 100% response) and respondent estimates for three indicators: breastfed in hospital, gestational diabetes, and WIC participant
- Compute mean bias and mean absolute bias over all 47 jurisdictions

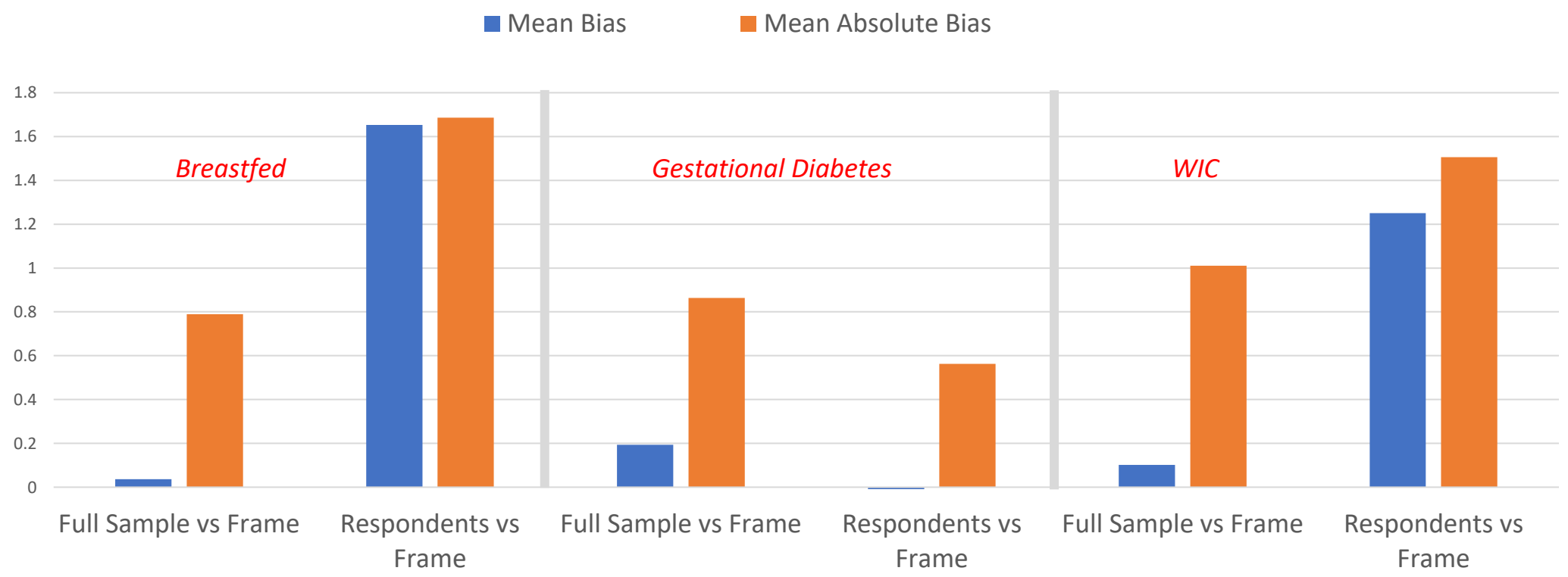
$$Mean\ Bias_{NR}(\overline{Y}_{fs}) = \sum_{i=1}^{47}(\overline{Y}_{fsi} - \overline{Y}_{fi}) \,/\, 47$$

$$Mean\ Absolute\ Bias_{NR}(\overline{Y}_{fs}) = \sum_{i=1}^{47} |\overline{Y}_{fsi} - \overline{Y}_{fi}| \,/\, 47$$

where $\overline{Y}_{f\mathrm{i}}$ is the frame value for indicator i and $\overline{Y}_{fsi}$ is the full sample weighted estimate for indicator i

- Similar computation for respondent sample weighted estimate

MEAN BIAS AND MEAN ABSOLUTE BIAS FOR FULL SAMPLE AND RESPONDENT ESTIMATES (47 SITES)



RESULTS

- Actual bias observed in PRAMS was relatively small
  - Highest mean absolute bias was 1.68 percentage points for breastfeeding
  - 9 of the 12 indicators examined had mean absolute bias less than 1 percentage point
- Behavioral indicators had highest levels of bias; demographic/socio-economic indicators had the lowest
- Positive behaviors were over-estimated; risky behaviors were under-estimated
- Weak correlation (-0.33) between response rate and bias
- Mean bias of full sample estimates was very small, as would be expected from a series of independent random samples

CONCLUSIONS

- Observed levels of bias should be acceptable for most uses of PRAMS data
- Behavioral indicators might be more susceptible to bias – private behaviors may be subject to unobserved individual- and group-level influences not amenable to weighting adjustments
- Bias varies across indicators (even within categories); must be examined at the indicator level

STRENGTHS AND LIMITATIONS

Strengths

- Access to complete, population-level data for a broad range of indicators
- Standard methodology across sites allows for analysis of impact of response rate on bias while controlling for other factors associated with bias
- Overlap of variables on birth certificate (medical and behavioral) with survey topics

Limitations

- Limited set of indicators collected on birth certificates
- Results cannot be generalized beyond study population (people recently delivering a live-born infant)

LEVEL-OF-EFFORT APPROACH TO ASSESSING INCREMENTAL NONRESPONSE BIAS IN PRAMS

JOSEPH PIROZZOLO, PHD; HOLLY SHULMAN, MA;

PHIL HASTINGS, PHD; ERIC BOOTH, MA; LEE WARNER, PHD; RUBEN SMITH, PHD

BACKGROUND

- **Decreasing survey response rates are driving new interest in non-response (NR) bias research**
  - **The level-of-effort (LOE) approach simulates low response rates (RR) in higher RR surveys**
  - **Later respondents are treated as non-respondents**

- **Research Question: *Do survey estimates from earlier versus later respondents exhibit differential bias?***
  - **Assumption: Respondents requiring greater effort are more similar to nonrespondents**
  - **Can be tested since PRAMS samples from birth certificates with information (i.e., demographics, medical) on respondents and nonrespondents**

METHODS

- **Selected 20 sites from PRAMS (2019)**
- **Criteria:**
  - **Response rate**
  - **Sample size**
  - **Urbanicity (*FiveThirtyEight index*)**
  - **Diversity (*WalletHub index*)**
  - **Geographical area**

		Urbanicity % of population in locale		
Diversity	Response rate (RR)	Rural	Mix	Urban
Homogeneous	Low RR	WV, NH	IN	
	Medium RR	ND	MO	PA
	High RR	MS, SD*	OR*	PR
Diverse	Low RR	OK*	SC	AZ*, FL*
	Medium RR	AK*	VA	IL
	High RR		KS	NYC*, MA

Note:
* Special subpopulation: [AK – Alaska Native; AZ – Hispanic (Mexican); FL – Hispanic (Cuban); NYC – Chinese and other races; OK – Hispanic and Native American; OR – Asian; SD – Native American]

RR:	Low: 0-55% (n=7)	Medium: 55-60% (n=6)	High: 60+% (n=7)
Sample size:	<1000 (n=6)	1000-1265 (n=8)	1266+ (n=6)

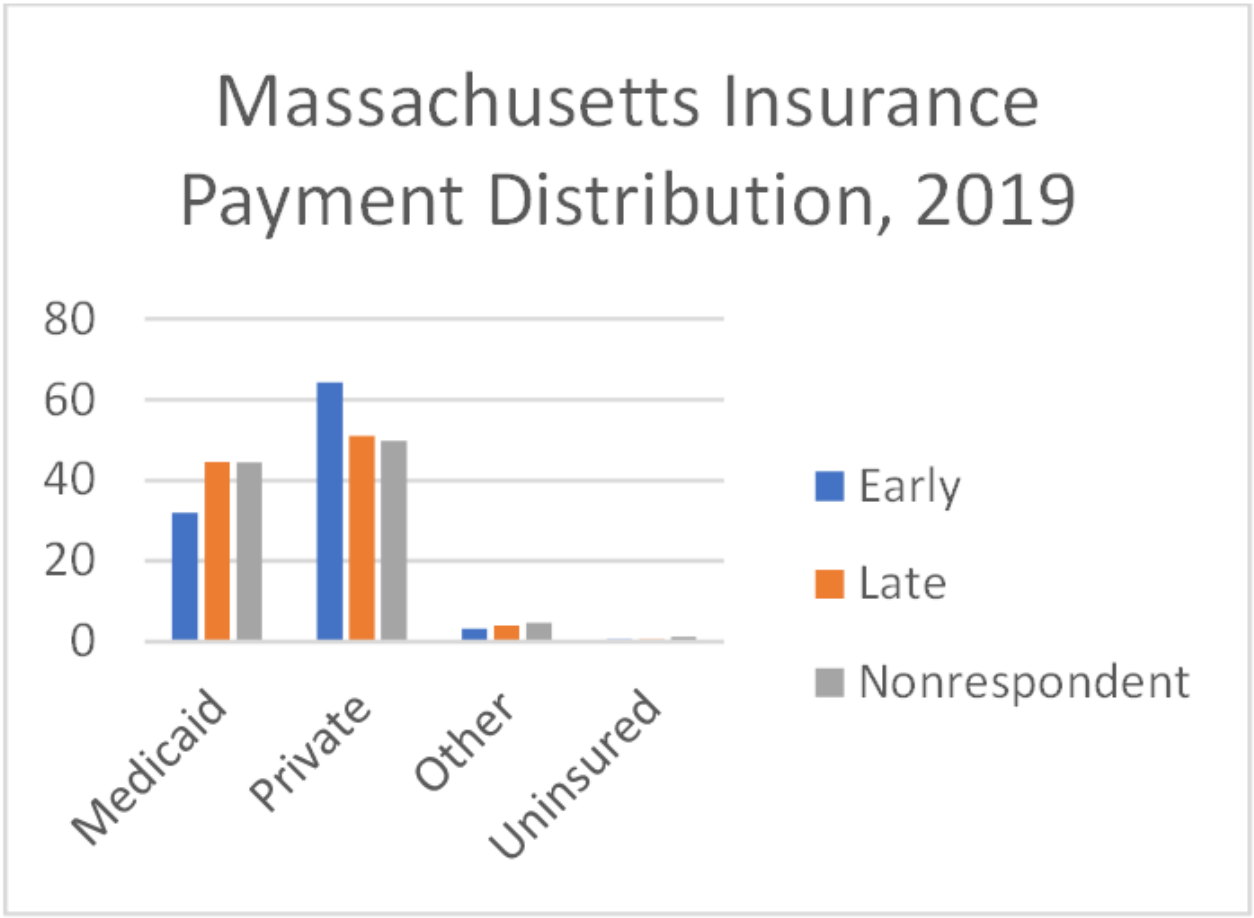
METHODS

- **Examples of 21 core PRAMS variables selected for analysis**

Contraception related	Medical conditions /health services	Other behavioral indicators	Demographic/SES
Intended pregnancy	Gestational diabetes	Physical abuse before pregnancy	Less than 100% of federal poverty level (FPL)
Postpartum birth control (BC) use	Flu shot	Smoking before pregnancy	
Moderate/Most effective postpartum BC method	Teeth cleaned during pregnancy	Smoking during pregnancy	
Discuss BC at postpartum visit (PPV)	Maternal postpartum visit	Still breastfeeding	

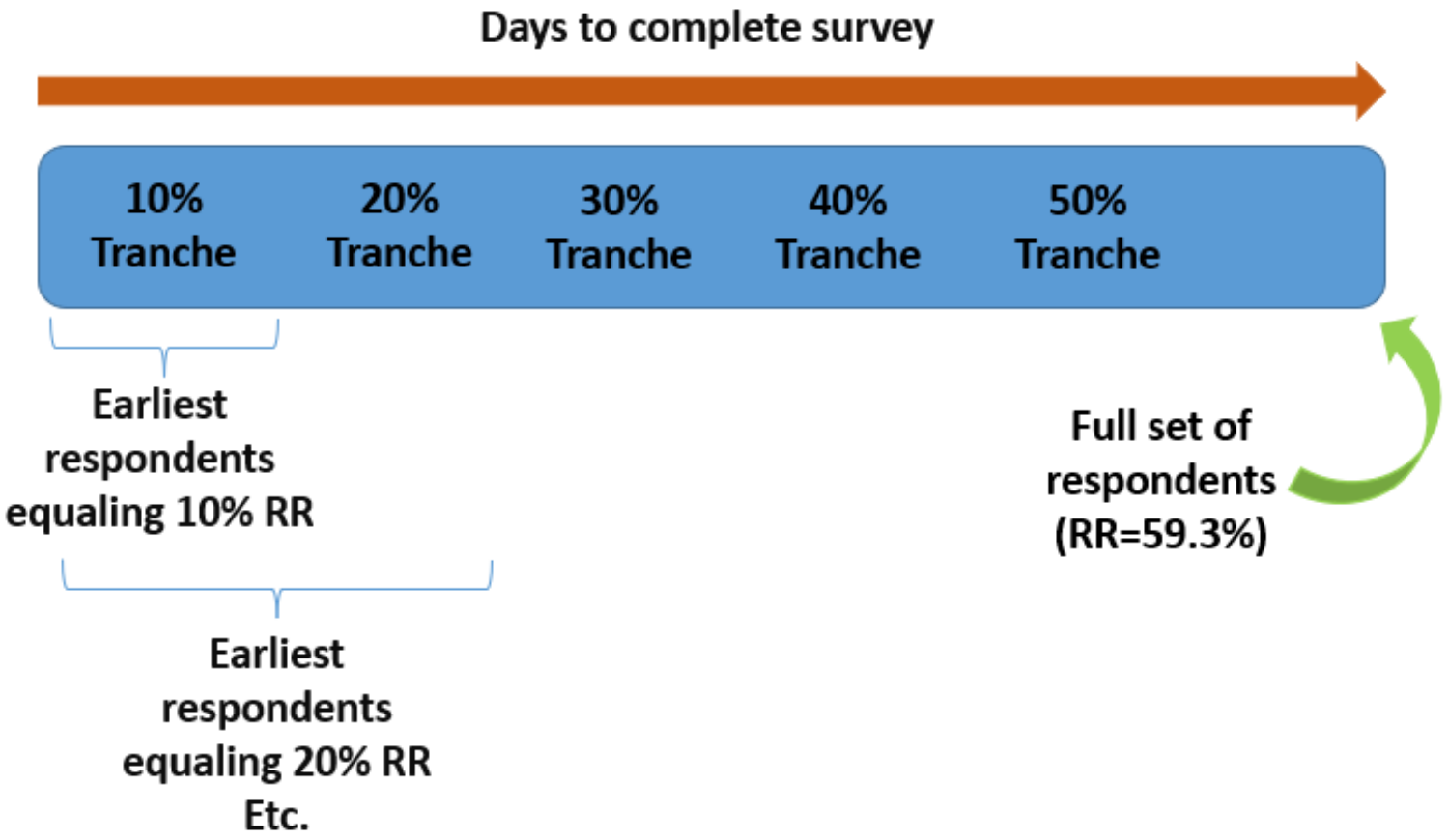
METHODS

- **Assumption: late-respondents are more similar to non-respondents**
- **Tested using 11 birth certificate variables**
  - **E.g., maternal race/ethnicity, maternal education, insurance for birth**
- **Early-respondents: completed survey before or during the first week of phone interview**



METHODS

- Created ‘tranches’ simulating lower RRs
- Earliest respondents comprised lowest RR tranches
  - Ranged from 5% to the full set of respondents
- Tranche data re-weighted with new non-response adjustment
- Example: Illinois 2019



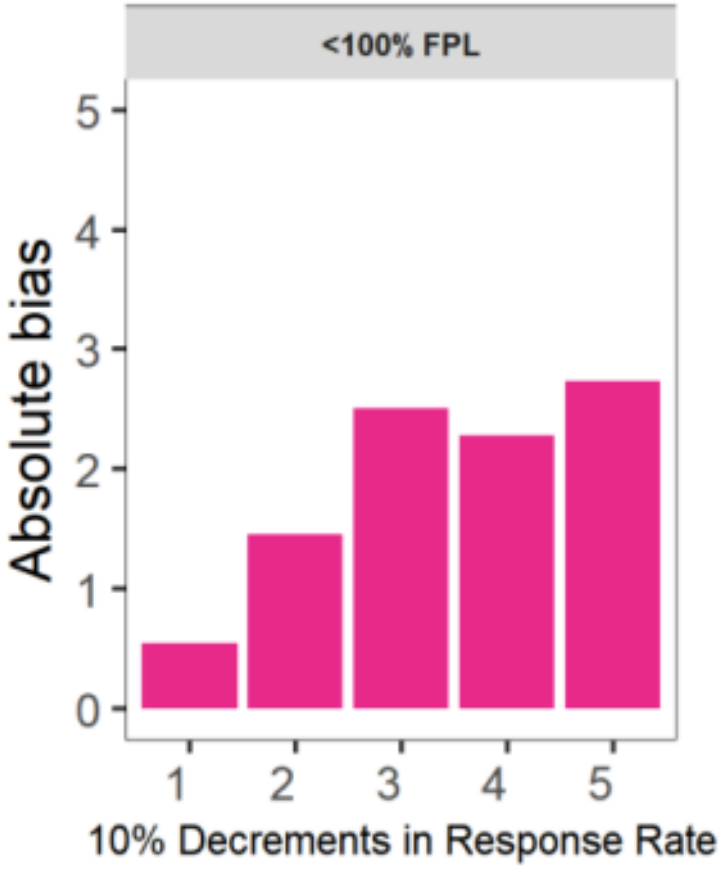
METHODS

- **Statistics calculated:**
  - **Absolute bias: Difference in weighted estimates between each response tranche and the full set of respondents**
  - **Relative bias: Percentage of the weighted estimate; a standard scale when examining multiple survey measures**
  - **Instances where estimated value from a response tranche fell outside the 95% CI of the full response estimate identified**
  - **Incremental bias: Mean absolute difference in estimates across sites as response rates drop in each tranche**

- **Multilevel regression model**
  - **Modeling bias as a function of response rate**
  - **Response rate tranche (level-1) nested within PRAMS site (level-2)**

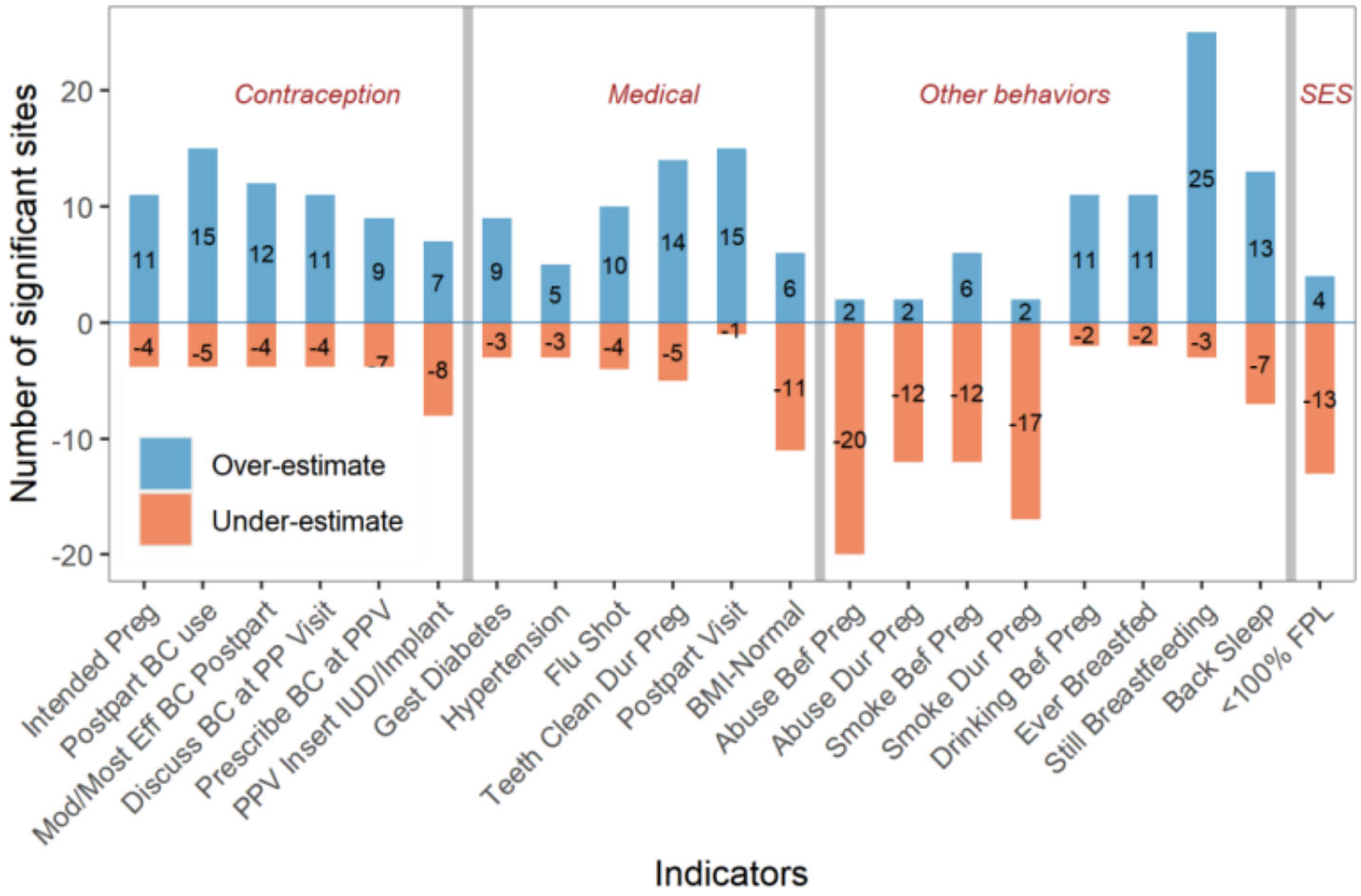
PREVIEW: KEY FINDINGS

- **On average, for every 10% decline in RR, we observed ~0.5% bias in estimation**
- **Bias tends to be in a predictable direction:**
  - **Indicators representing desirable outcomes or positive behaviors tended to be overestimated**
  - **Indicators of adverse outcomes/behaviors (e.g., physical abuse, smoking) tended to be underestimated**



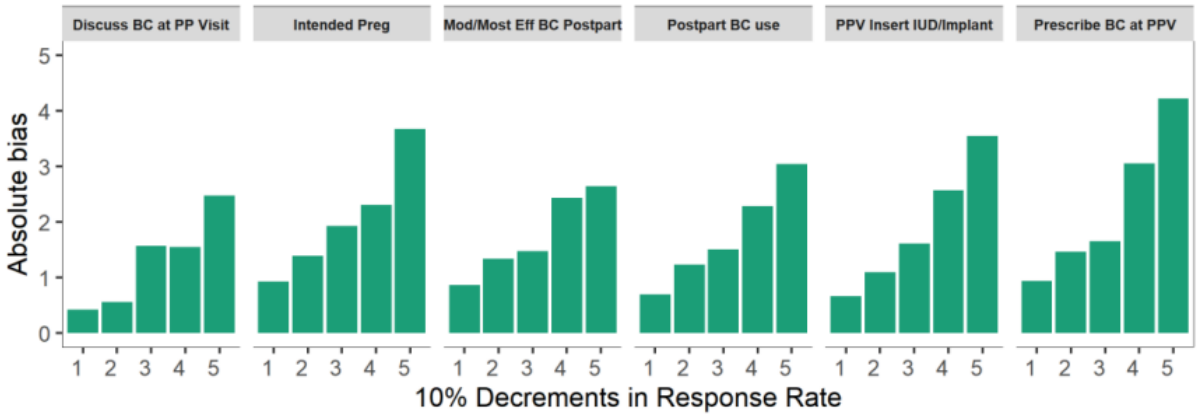
RESULTS

- **Number of significant differences across tranches for 20 sites**
  - **Positive behaviors overestimated when response rates dropped**
  - **Indicators representing less desirable/risky behaviors likely to be underestimated**

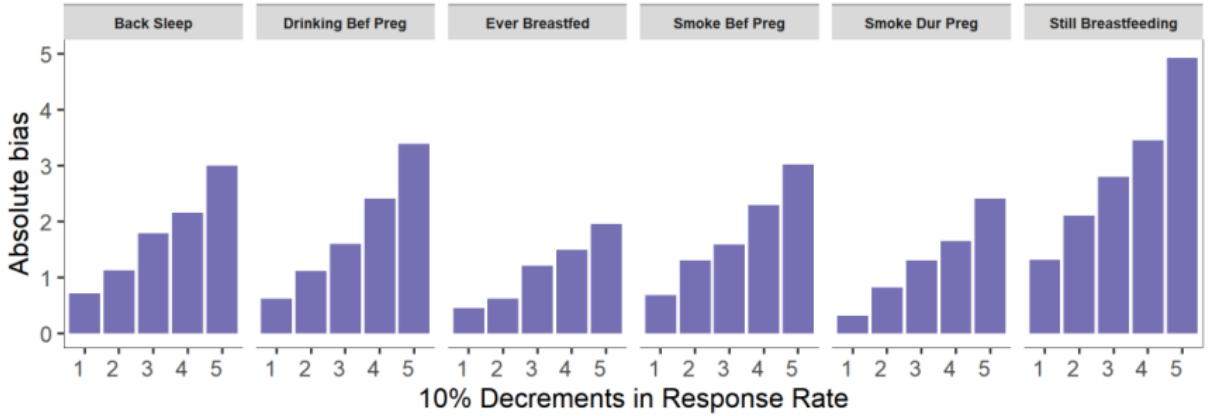


RESULTS

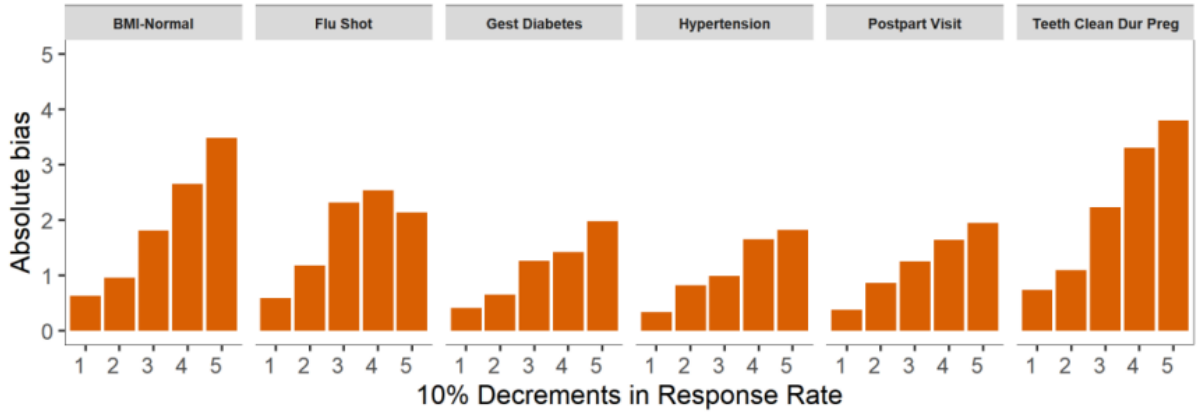
Contraception-Related Indicators



Other Behavioral Indicators

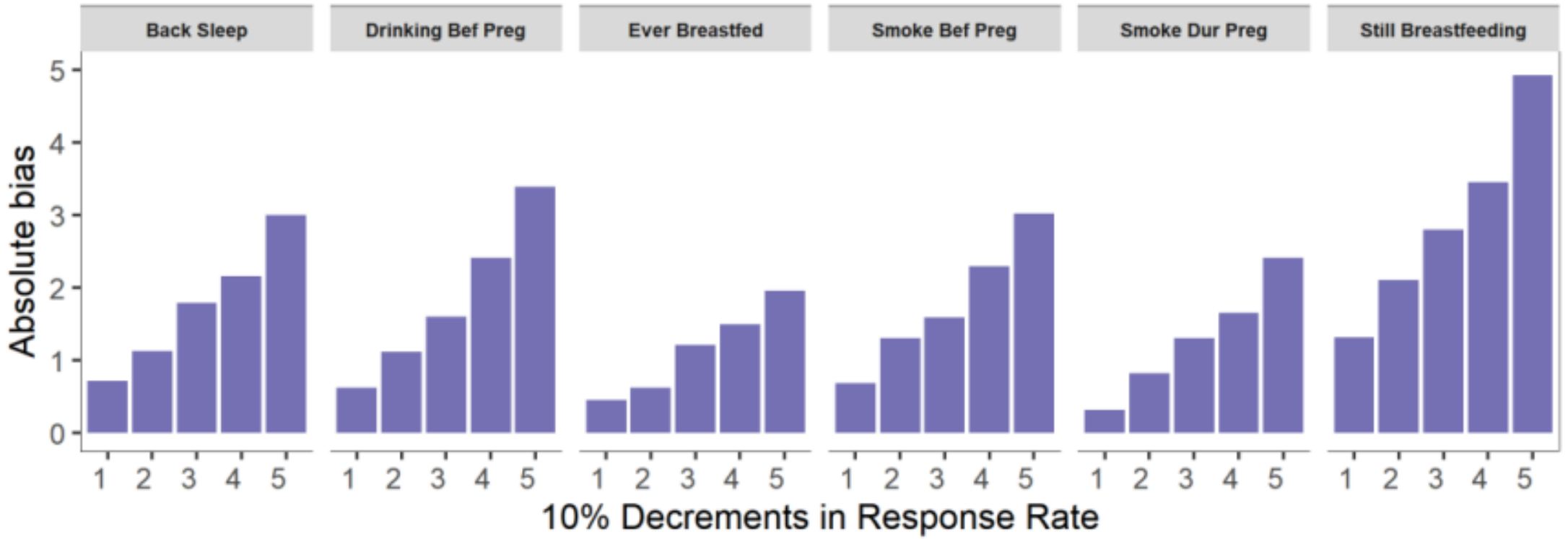


Medical Conditions/Health Indicators



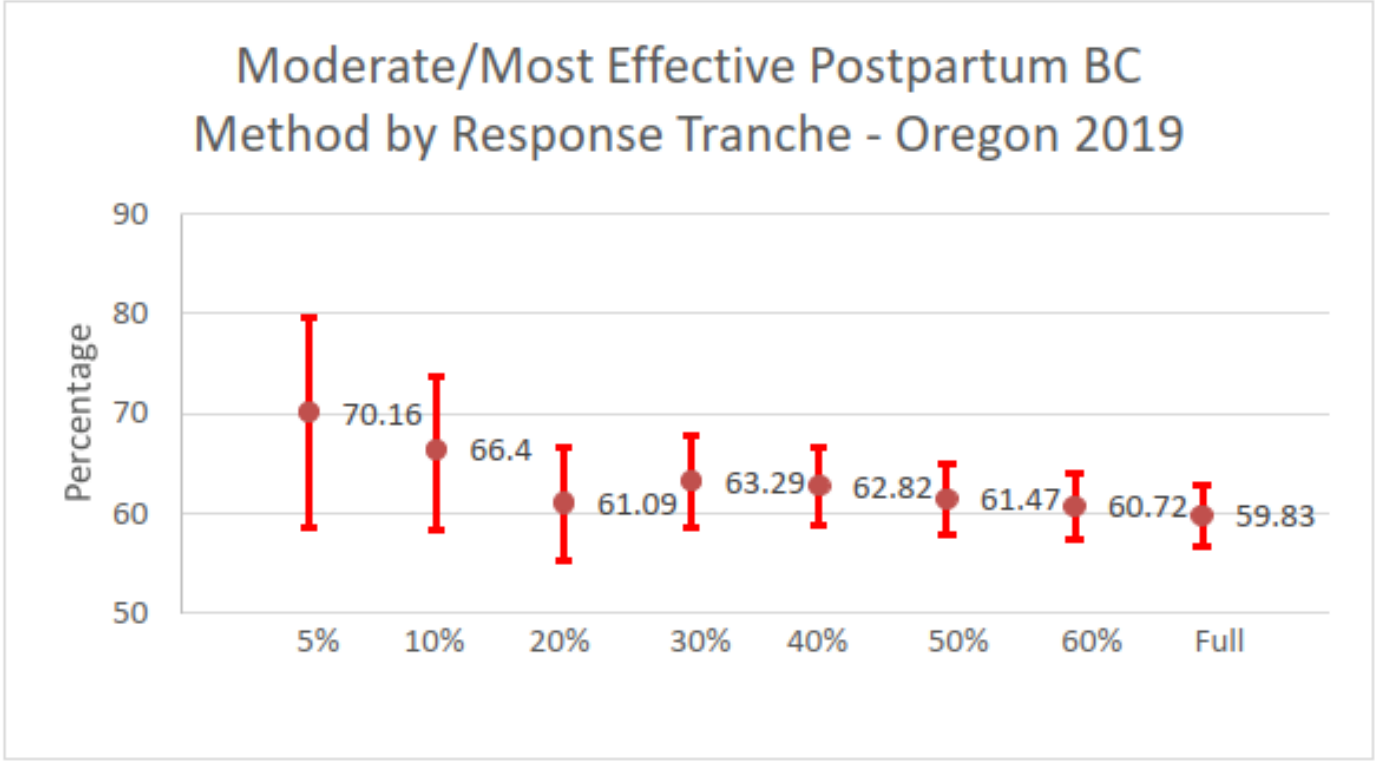
RESULTS

Other Behavioral Indicators



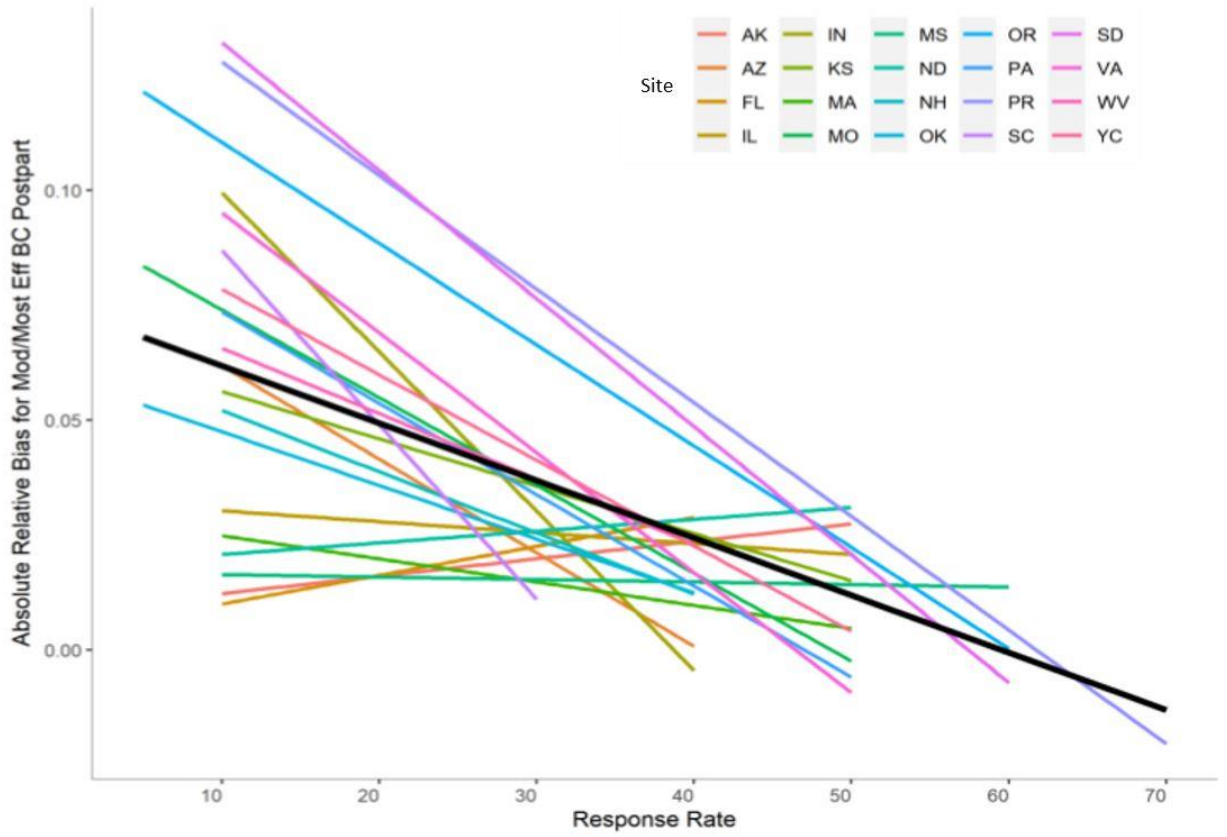
RESULTS

- **Mod/most effective birth control (BC) in Oregon: all tranches overestimate the full estimate**
  - **Estimates from the 5%, 10%, and 30% tranches are outside of the 95% CI around the full estimate**



RESULTS

- **Multilevel regression models: Absolute relative bias for mod/most effective postpartum BC**
  - ***Absolute relative bias*** **decreased by an average of 1.25% for each 10% increase in response rate. (Range = 0.42 – 6.26% across indictors.)**



WRAP-UP

- ***Findings***
  - ***On average, we saw ~0.5% absolute bias per 10% decrease in RR***
  - ***Some indicators were differentially prone to over-/under-estimation at lower RR***

- ***Implications:*** **LOE approach allows us to explore nonresponse bias across a range of decreasing response rates**

- ***Limitations:*** **Full response estimates from the lower response rate sites may already have inherent bias**
  - **e.g., estimates from full response estimates in a site with 30% response rate might be different from estimates from the 30% tranche in a site with 70% response rate.**

NONRESPONSE BIAS SENSITIVITY ANALYSIS FOR THE PRAMS SURVEY, 2019

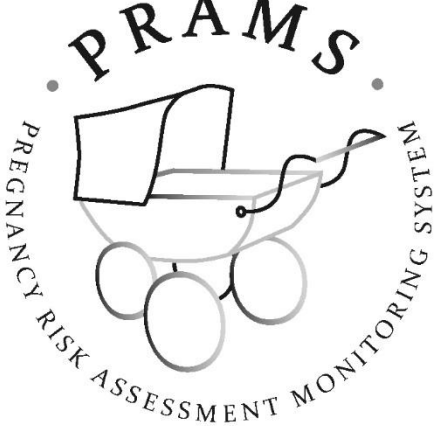
PHIL HASTINGS, PHD; PATRICK S. MALONE, PHD; JIE MIN, PHD; ERIC BOOTH, MA; JOSEPH PIROZZOLO, PHD

HOLLY SHULMAN, MA; LEE WARNER, PHD; RUBEN SMITH, PHD

CHALLENGE OF NON-RESPONSE BIAS ANALYSIS

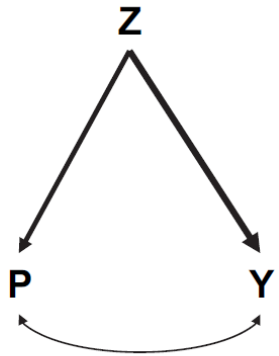
- **Estimating non-response bias is always imperfect; we cannot really know what non-responders 'would have answered'…**

- **<u>Research Question</u>: Despite this, can we find a plausible upper limit of nonresponse bias (i.e., the 'worst case scenario')?**



MISSINGNESS SCENARIOS

- **Missing-at-random (MAR)**
  - **Subgroups of the population can *systematically differ* in their propensity to respond (P)**
  - **Yet we can still produce unbiased estimates (of Y) conditioned on auxiliary variables (Z; e.g., demographics, geography…)**



- **Missing-not-at-random (MNAR; aka *non-ignorable nonresponse*):**
  - **Response probability is driven by the survey topic itself**
  - **Adjustments based on auxiliary variables would not produce unbiased estimates**
  - **Example Topics: Poverty, depression, substance abuse, …**

From Groves (2006) Nonresponse rates and nonresponse bias in household surveys.

NEW ADVANCES IN STUDY OF BIAS

- **Proxy Pattern-Mixture Analysis (PPMA; Andridge and Little, 2020*)**
- **PPMA is a *sensitivity analysis* method: "Given the observed data, how do changes to our assumptions about nonresponse impact estimates of bias?"**
- **New: PPMA can now simulate *binary outcomes* under varying missingness assumptions**
  - **Missing-at-random (best case) → missing-not-at-random (worst case)**
- **PRAMS data contains rich auxiliary data for the full sample (respondents + non-respondents)**
  - **Ideal for PPMA analyses**

Journal of Official Statistics, 2020, vol. 36, issue 3, 703-728*

METHODS (DATA): PRAMS, 2019

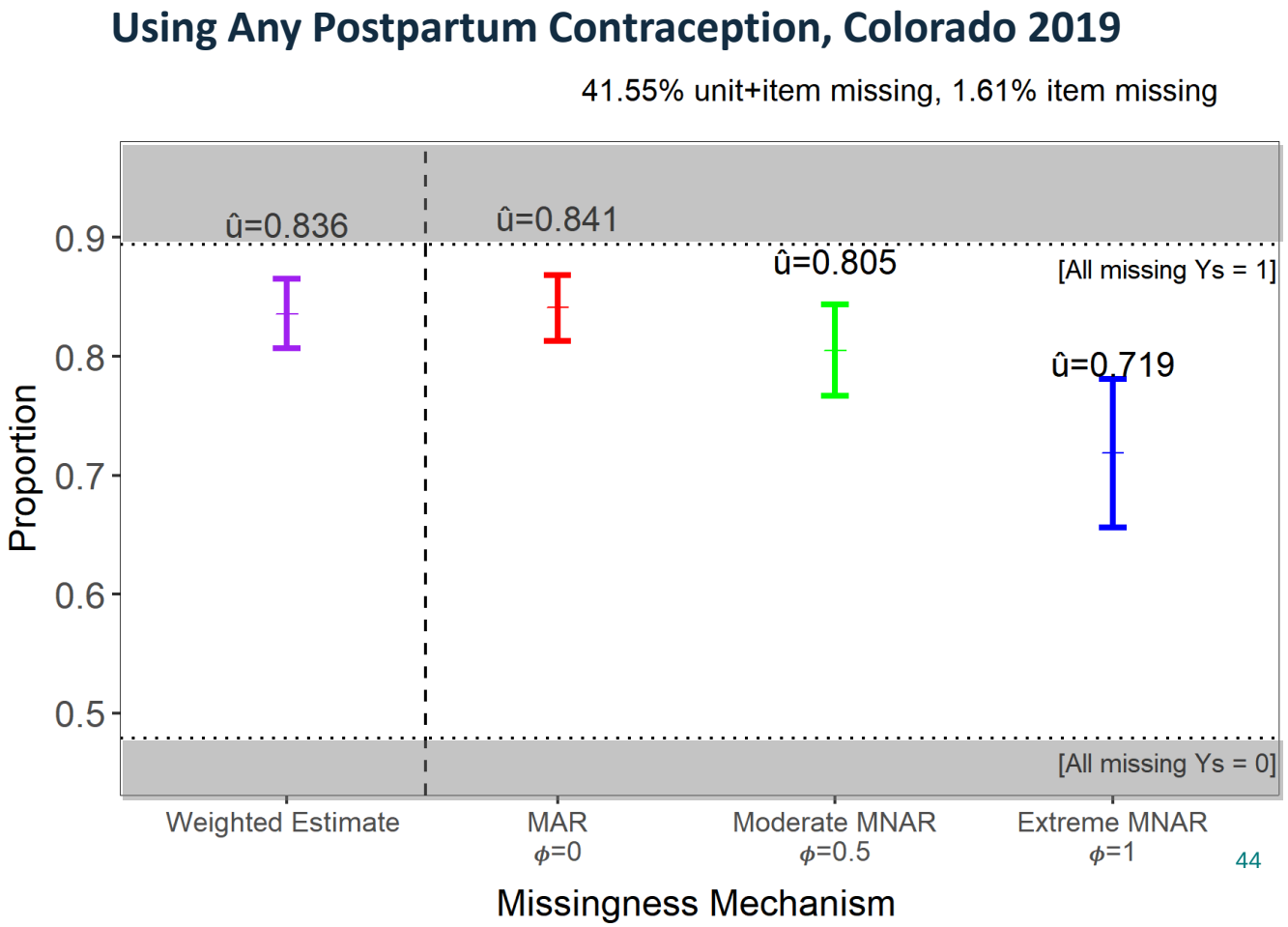
- **Analyzed 13 survey indicators: contraceptive and other behaviors, medical conditions, poverty**
- **Grouped 45 PRAMS sites into 5 response rate groupings:**
  - **Low (39.8% -- 50.1% RR; 8 sites)**
  - **Med-Low (50.5% -- 55.0% RR; 8 sites)**
  - **Med (55.5% -- 59.7% RR; 15 sites)**
  - **Med-High (60.4% -- 65.0% RR; 8 sites)**
  - **High (65.3% -- 81.0% RR; 6 sites)**

- **Unique benefit of PRAMS:**
  - **Multi-site survey with same protocol & questions; lends confidence to findings across 45 sites**
  - **Fully "known" population**

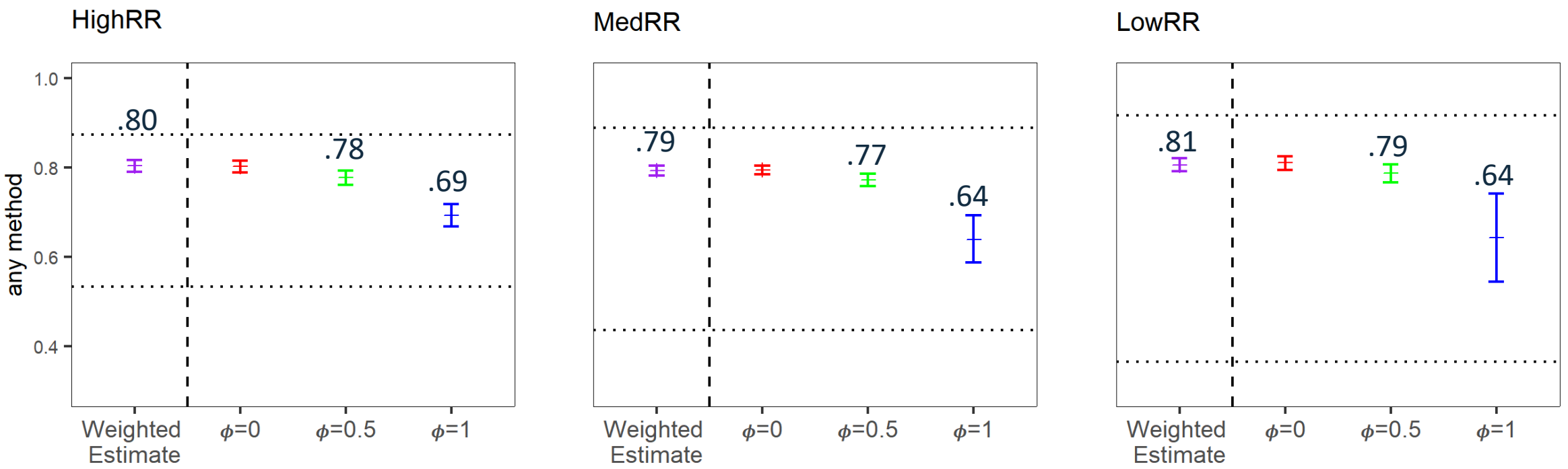
PRAMS Auxiliary Variables	
Education	Race
Age	Ethnicity
Marital Status	Previous Live Birth
Medicaid Birth Coverage	Prenatal Care
Women, Infants and Children Program Status	Top 20% Hospital & County ranked by # births

FINDINGS: EXAMPLE GRAPH

- **PRAMS weighted estimate**
- **Manski bounds (horizontal dotted lines)**
  - **Nonresponders set to "all yes" or "all no"**
  - **Impossible boundary**
  - **Upper=0.89; Lower=0.48**
- **3 missingness simulations**
  - **MAR (ϕ = 0.0, best case)**
  - **Moderate MNAR (ϕ = 0.5)**
  - **Extreme MNAR (ϕ = 1.0, worst case)**

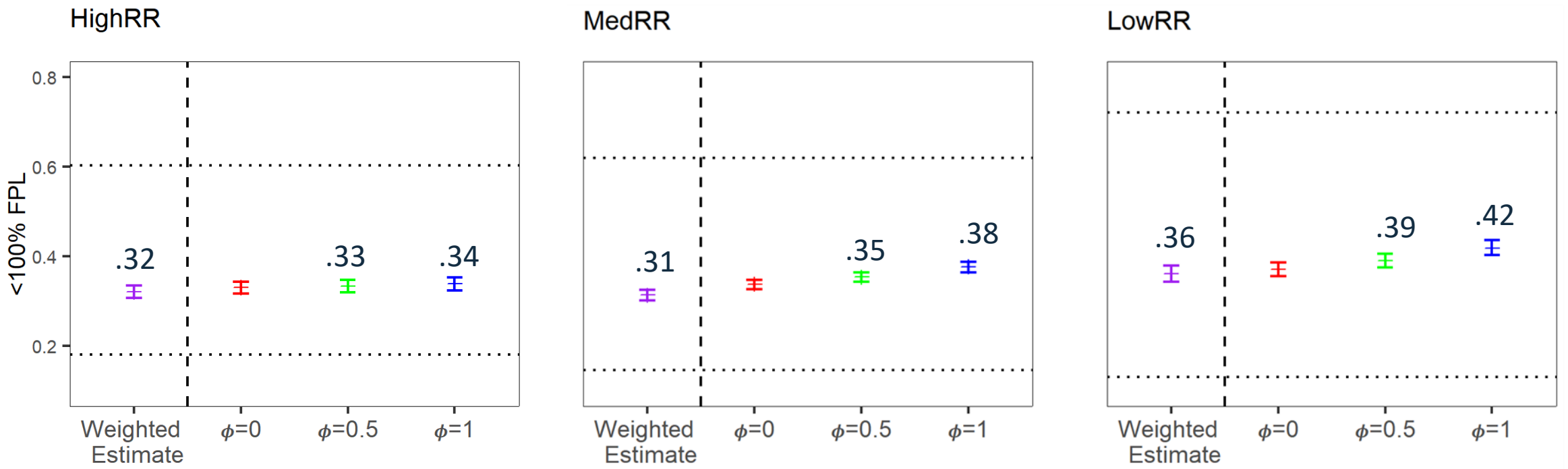


FINDINGS: ANY POSTPARTUM CONTRACEPTION



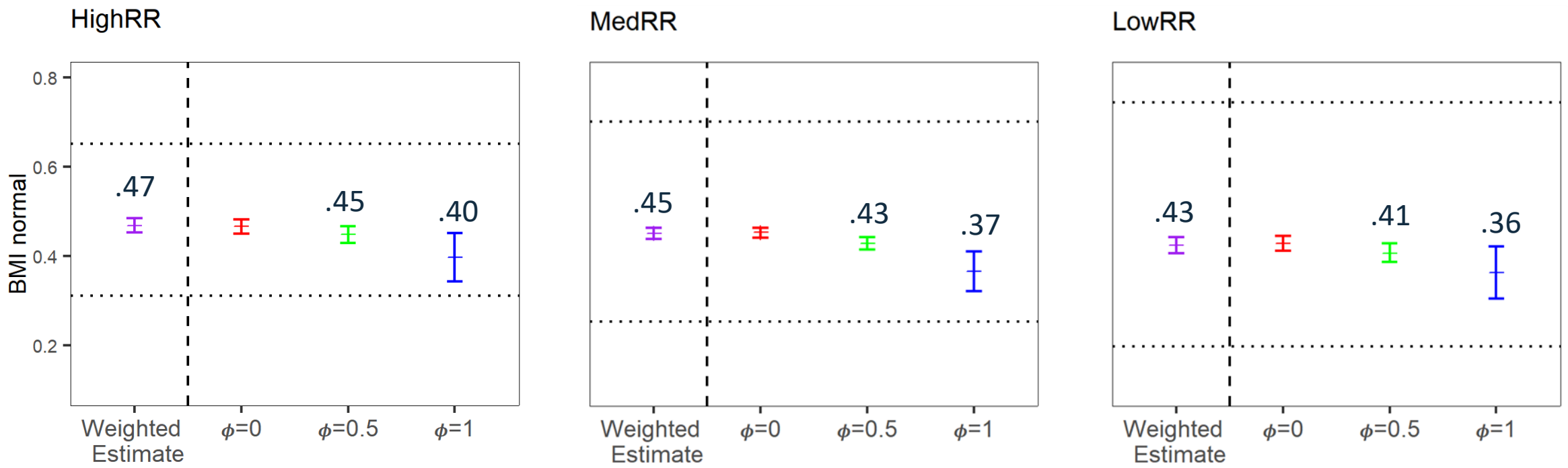
- **As MNAR worsens, PRAMS over-estimates proportion of women using postpartum contraception.**
- **Patterns are similar as you go from high to low RR sites; yet confidence worsens as RR decreases.**
- **Median bias across 45 sites: Moderate MNAR = +1.5%; Extreme MNAR = +6.25%.**

FINDINGS: POVERTY



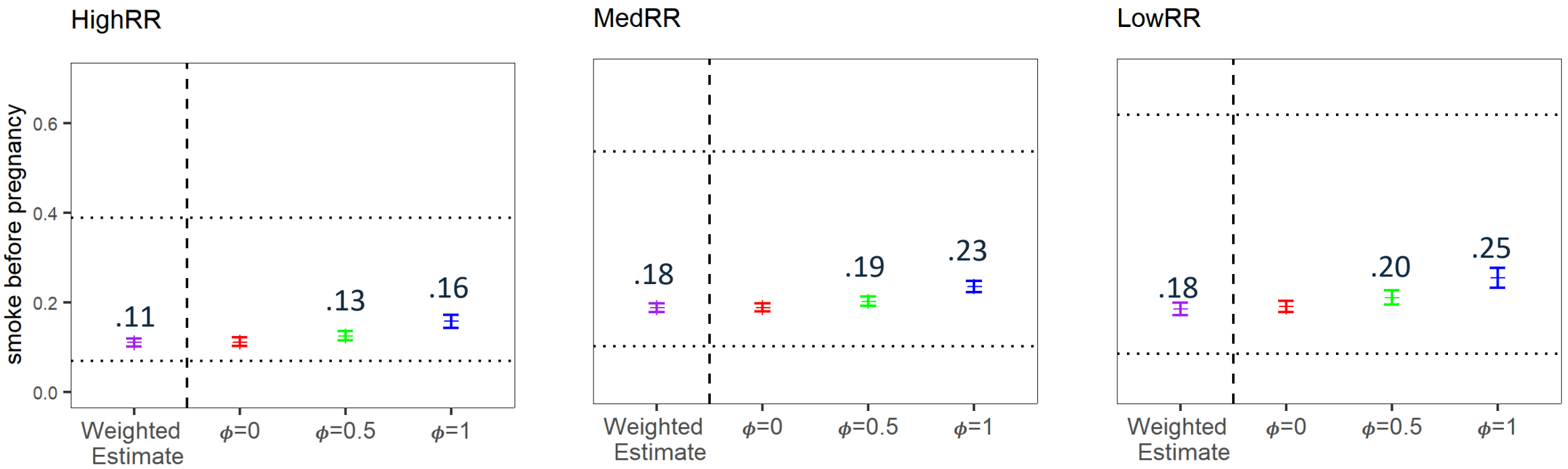
- **As MNAR worsens, PRAMS tends to *under-estimate* proportion of women below poverty line.**
- **Under-estimation is more pronounced with lower response rates.**
- **Median bias across 45 sites: Moderate MNAR = -3.15%; Extreme MNAR = -5.25%.**

FINDINGS: NORMAL BMI (BODY MASS INDEX)



- **As MNAR worsens, PRAMS slightly over-estimates proportion of women with normal BMI.**
- **Median bias across 45 sites: Moderate MNAR = 0.85%; Extreme MNAR = 3.15%.**

FINDINGS: SMOKING BEFORE PREGNANCY



- **As MNAR worsens, PRAMS under-estimates proportion of women smoking before pregnancy.**
- **Median bias across 45 sites: Moderate MNAR = -1.25%; Extreme MNAR = -3.10%.**

KEY TAKEAWAYS

- **As we increased MNAR, weighted survey:**
  - **Tended to overestimate "healthy/positive" indicators**
    - **E.g., postpartum contraception, normal BMI**
  - **Tended to underestimate "risky/negative" indicators**
    - **E.g., smoking before pregnancy, poverty**

- **Lower response rates were associated with *increased* bias -- regardless of missingness assumption**

- **Moderate MNAR scenarios showed relatively "acceptable" biases**
  - **Median bias estimates (45 sites) usually within 0-3 percentage points of PRAMS weighted estimate**

- **Extreme MNAR (among the *most biased* indicators) showed *median* bias (45 sites) within 6-8 percentage points of the PRAMS weighted estimate**

IMPLICATIONS

- **Estimates of bias and uncertainty can be reasonably quantified (topic-by-topic)**
  - **Bias estimates are driven by:**
    - **Proxy differences between responders & nonresponders**
    - **Missingness assumptions (MAR → MNAR)**
  - **Uncertainty around bias is reduced by:**
    - **Higher proxy correlation with topic**
    - **Higher response rates**

- **Sensitivity methods can identify the limit of *plausible* bias -- extreme MNAR as the "worst case"**
  - **More precise than Manski boundaries (i.e., *impossible* bias)**
  - **Offers more interpretibility than 'MAR' assumption**
  - **Imputation is relatively efficient (and programmable)**

LIMITATIONS

- **Weaker proxy association with indicator (e.g., contraception) yields less precise bias estimates**
- **Good auxiliary data is the key to identify (as well as adjust for) potential bias**
- **Model performance becomes less stable with smaller sample sizes (n<1000)**
- **Generalizability:**
  - **PRAMS population is only among postpartum women**
  - **PRAMS, unlike many surveys, has 'luxury' of auxiliary variables at the record level**

DISCUSSION

- Results from all 3 NRB analyses yielded similar results:
  - Weak association between response rates and bias
  - Sensitive/behavioral survey indicators more susceptible to bias than medical and factual indicators
  - Magnitude of bias was small
- Practical implications
  - On average for every 20% decline in response rates, expect about 1 percentage point increase in bias
  - Current NRB analyses found bias levels to be half of what was estimated from 2002 NRB analysis
  - Most survey indicators examined have little evidence of bias

RESPONSE RATE THRESHOLD ADVANTAGES AND DISADVANTAGES

Pros

- Strong incentive for jurisdictions to maintain data collection efforts
- Concerns that response rates would drop if threshold abolished

Cons

- Little evidence to support continued use of threshold policy
- Difficult to justify withholding critical maternal and child health (MCH) data
  - PRAMS response rates higher than many federal health surveys
  - Many sites below threshold have high levels of maternal and infant morbidity and mortality
  - Major data collection and processing effort unused

IMPLICATIONS FOR PRAMS DATA RELEASE

Possible data release approaches to consider:

- Retain threshold
- Abolish threshold and publish results of bias analyses (overall and site-specific) so data users know risks of bias
- Abolish threshold and add warnings or qualifiers to published reports using PRAMS data for sites with low response rates

NEXT STEPS

- NRB analyses are underway for the 2020 PRAMS data
- After the 2020 NRB analyses are completed, CDC will use the results of the 2019 and 2020 analyses to re-evaluate the response rate threshold policy
- Going forward, conduct annual, comprehensive NRB analyses to quantify the risk of bias

THANK YOU!

Holly Shulman, MA
Statistician, PRAMS Team,
Division of Reproductive Health, CDC
hshulman@cdc.gov

Phil Hastings, PhD
Principal,
Far Harbor LLC
phil@farharbor.com

Joseph Pirozzolo, PhD
Statistician,
Far Harbor LLC
joe@farharbor.com

Centers for Disease Control and Prevention

National Center for Chronic Disease Prevention and Health Promotion

Division of Reproductive Health