Combining and Standardizing Panel Surveys as part of a Government Survey System - an Investigation

Van L Parsons, Yulei He, Katherine Irimata, Bill Cai

Division of Research and Methodology

2023 Federal Committee on Statistical Methodology, Hyattsville, MD

October 24, 2023

Disclaimer

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the official position of the National Center for Health Statistics, Centers for Disease Control and Prevention.

Outline

- Rapid Surveys System
- Startup Goals
- General Strategies
  - Weighting
  - Combining Estimates
- Graphical Results
- Evaluations
- Where Things Stand

Rapid Surveys System

Rapid Surveys System (RSS)

- Rising challenges for traditional, household-based face-to-face personal interview surveys
  - Can be time- and resources-consuming
  - Especially problematic when making personal contact is difficult (e.g., during a pandemic)
  - Less efficient for capturing information that is evolving quickly
- Alternative data collection tools have been explored and researched
- Rapid surveys system based on a web administration aims to collect and publish timely data with high quality

Rapid Surveys System

- Launched this year

- Planned quarterly collection and release
- Different topics quarter-to-quarter

- RSS is collecting samples from two commercial panels:
  - NORC AmeriSpeak Panel (AM)
  - Ipsos KnowledgePanel (KP)
  - One round completed

Rapid Surveys System

- Estimates to be calculated and released in combined form
- Analytic products (e.g., web tables, dashboards, etc.) will use combined estimates
- Combined public-use microdata

Rapid Surveys System (RSS)
Round 1: Sample sizes

Panel	Data collection methodology 1 (usual)	Data collection methodology 2 (modified)
AmeriSpeak Panel (NORC)	2,900	4,000
KnowledgePanel (Ipsos)	4,700	3,600
Total	7,600	7,600

Startup Goals

Startup Goals

- Develop a reasonable all-purpose production method for combining AM + KP samples
- System requires minimal intervention.
- Meet time constraints on a quarterly basis.
- (Goals decided well in advance of data collection)

General Strategies

General Strategies (Weighting Background)

- AM, KP: pre-weighted to select Census controls.
- National Health Interview Survey (NHIS) is to be treated as a quarterly benchmark/reference survey.
- NHIS: High-quality population survey.
- NHIS: Consistent sampling/weighting methodology over time.
- NHIS: Provides defined health, socio-economic, and civic engagement control variables (survey estimated).

General Strategies (Weighting)

- Combined AM + KP Weighting
- Re-calibration weighting
  - Start with pre-weighted AM and KP
- Pool AM and KP and then calibrate
  - Many unknowns about joined features
  - Not applied
- Separately calibrate AM and KP and then pool.
  - Use the same methodology on each
  - Applied

General Strategies (Combining estimates/data)

- Convex combination methods
  - “Optimal λ model” or “fixed effects model”
  - Effective sample size “one-size-fits-all” combination
  - “random effects model”

General Strategies (Combining estimates/data)

- Convex combination methods

If $(\hat{y}_1, \hat{v}_1)$ and $(\hat{y}_2, \hat{v}_2)$

are the standalone panel estimated means and variances,

consider the form $\hat{y}_{comb} = \lambda_1 \hat{y}_1 + (1 - \lambda_1)\hat{y}_2$

General Strategies (Combining estimates/data)

- "Optimal λ" ("fixed effects model")

$$\hat{y}_{comb} = \lambda_1 \hat{y}_1 + (1-\lambda_1)\hat{y}_2 \ ,$$

$$\lambda_1 = \frac{\hat{v}_1^{-1}}{\hat{v}_1^{-1} + \hat{v}_2^{-1}} \text{ will be optimal}^{*} \text{ for a common population mean,}$$

$$\hat{Var}(\hat{y}_{comb}) = \frac{1}{\hat{v}_1^{-1} + \hat{v}_2^{-1}} \quad (\text{*special conditions})$$

General Strategies (Combining estimates/data)

- Optimal λ: “fixed effects” meta-analysis model in R package “rmeta”.
- Stand-alone AM and KP variance designs needed.
- Variance estimates computed over a list of pre-specified variables using complex survey software.
- May be good for production tables and dashboards.
- Phil Kott (RTI) suggested a model-smoothed version of λ’s by domain.

General Strategies (Combining estimates/data)

- Global effective sample size λ "one-size-fits-all"

Use $deff_{survey} = 1 + CV^2(w)$ as a design effect for each survey.

$n_{eff} = \dfrac{n_{obs}}{deff_{survey}}$ as a general effective sample size for each survey.

$\lambda_1 = \dfrac{n_{eff,1}}{n_{eff,1} + n_{eff,2}}$ $\quad \hat{y}_{comb} = \lambda_1 \hat{y}_1 + (1 - \lambda_1)\hat{y}_2$

Combined data join survey 1 and 2 with weights $\lambda_1 \cdot \{w_1\}, \lambda_2 \cdot \{w_2\}$

Preserves control tabulations and

for $\hat{y}_{comb} = \lambda_1 \hat{y}_1 + (1 - \lambda_1)\hat{y}_2$ places more weight on the more efficient side.

General Strategies (Combining estimates/data)

"random effects" λ

Treat $(\hat{y}_1, \hat{v}_1)$ and $(\hat{y}_2, \hat{v}_2)$ as conditional given some random effect α

If $\hat{y}_{comb} = \lambda_1 \hat{y}_1 + (1 - \lambda_1)\hat{y}_2$

If $\sigma_a^2 = 0$ then choose $\hat{\lambda}_1 \propto \hat{v}_1^{-1}$

If $\sigma_a^2 > 0$ $\hat{\lambda}_1 \propto \dfrac{1}{\hat{v}_1 + \hat{\sigma}^2(\alpha)}$

- Chi-square test used to test for heterogeneity can be used.

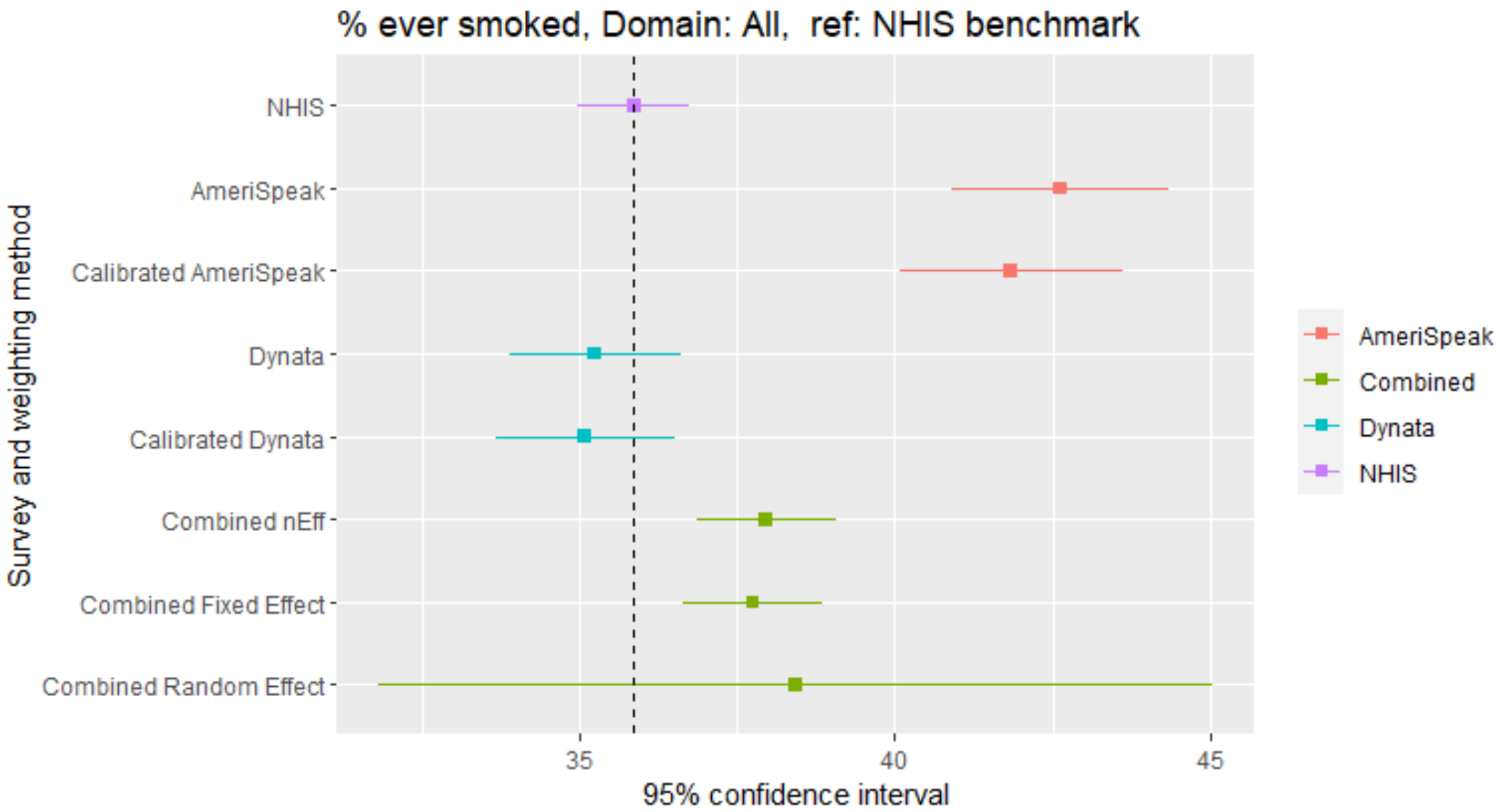
Big caveat: low degrees of freedom, but useful as a flag of major AM and KP differences

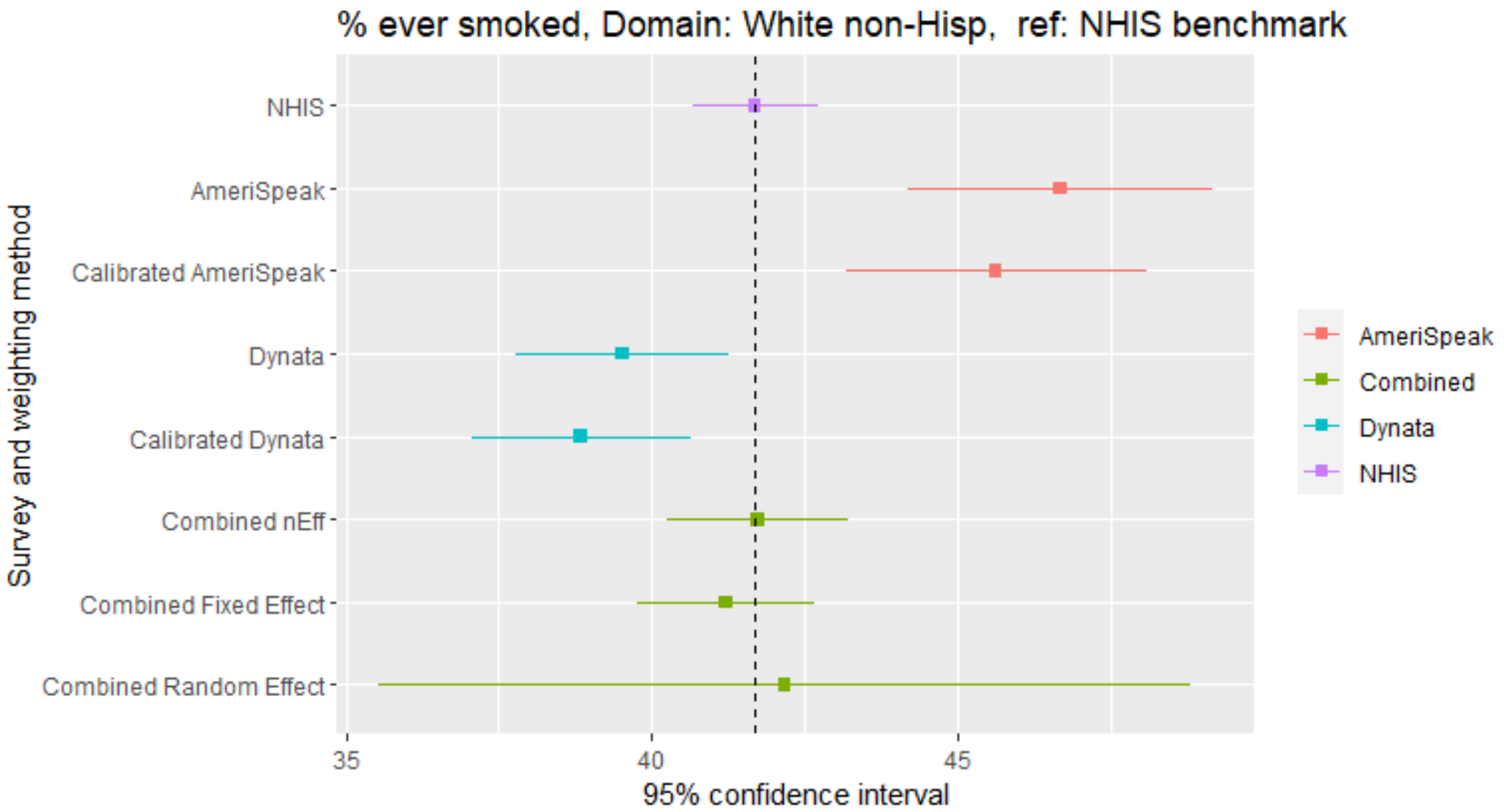
Calculated in R rmeta package meta.summaries .

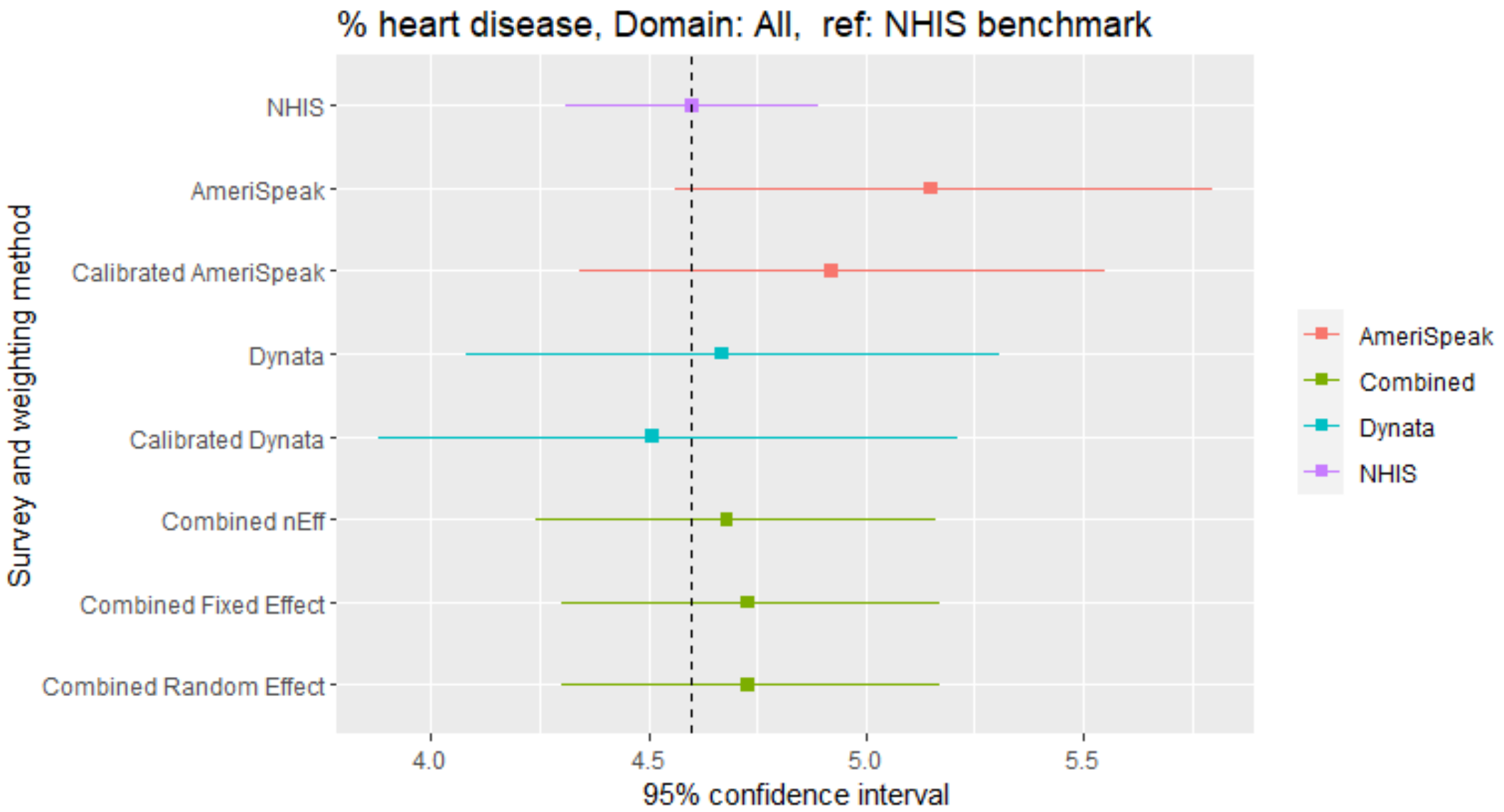
Evaluation of Methods

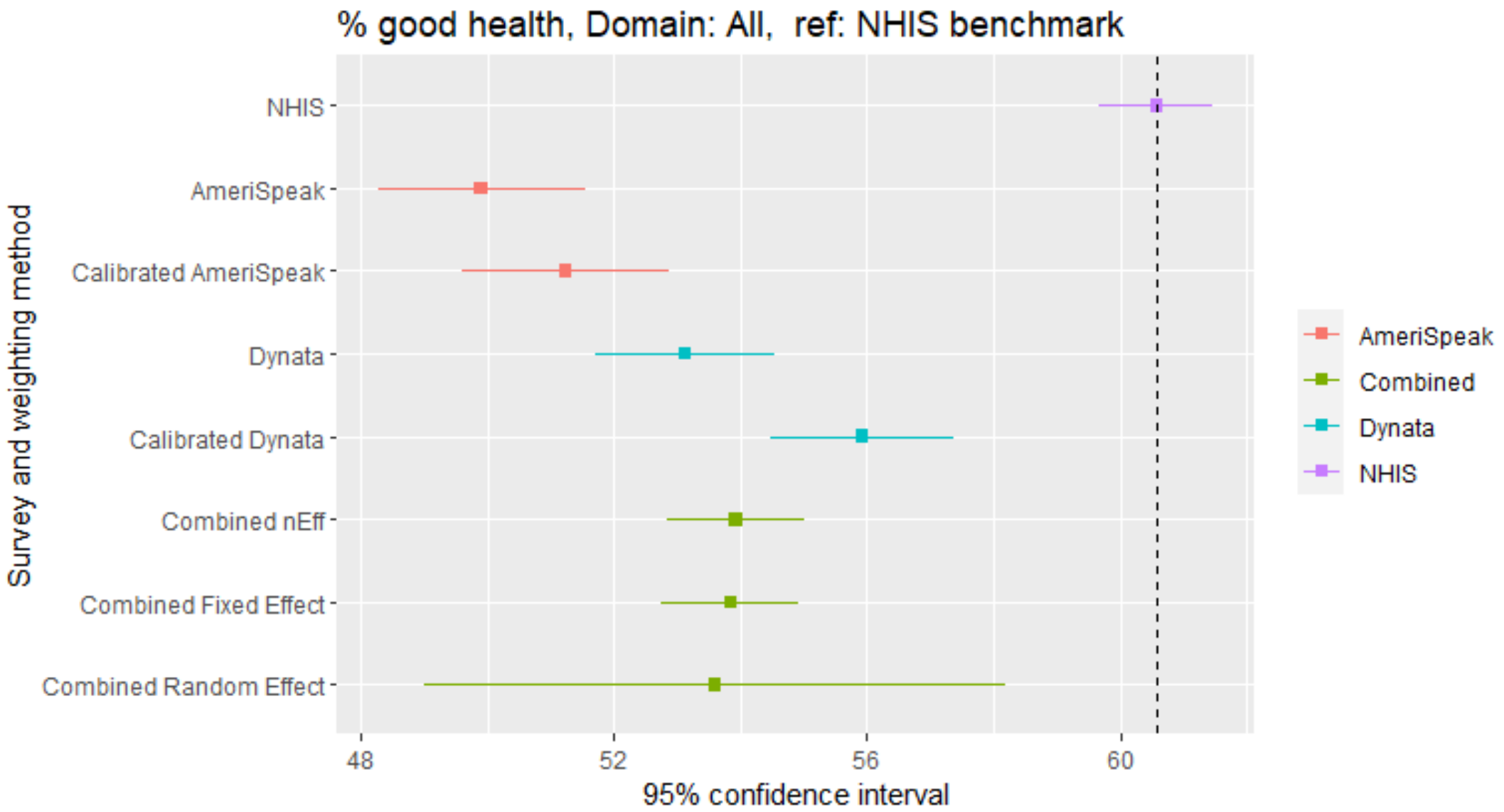
- No AM, KP data available for focused research.
- Show some “practice” results with data from:
  - RANDS during COVID-19 Round 1 (AmeriSpeak panel)
  - Dynata a *non-probability* panel (worst case survey)
  - NHIS from 2018 (benchmark)
  - RANDS re-calibrated and Dynata re-calibrated to NHIS

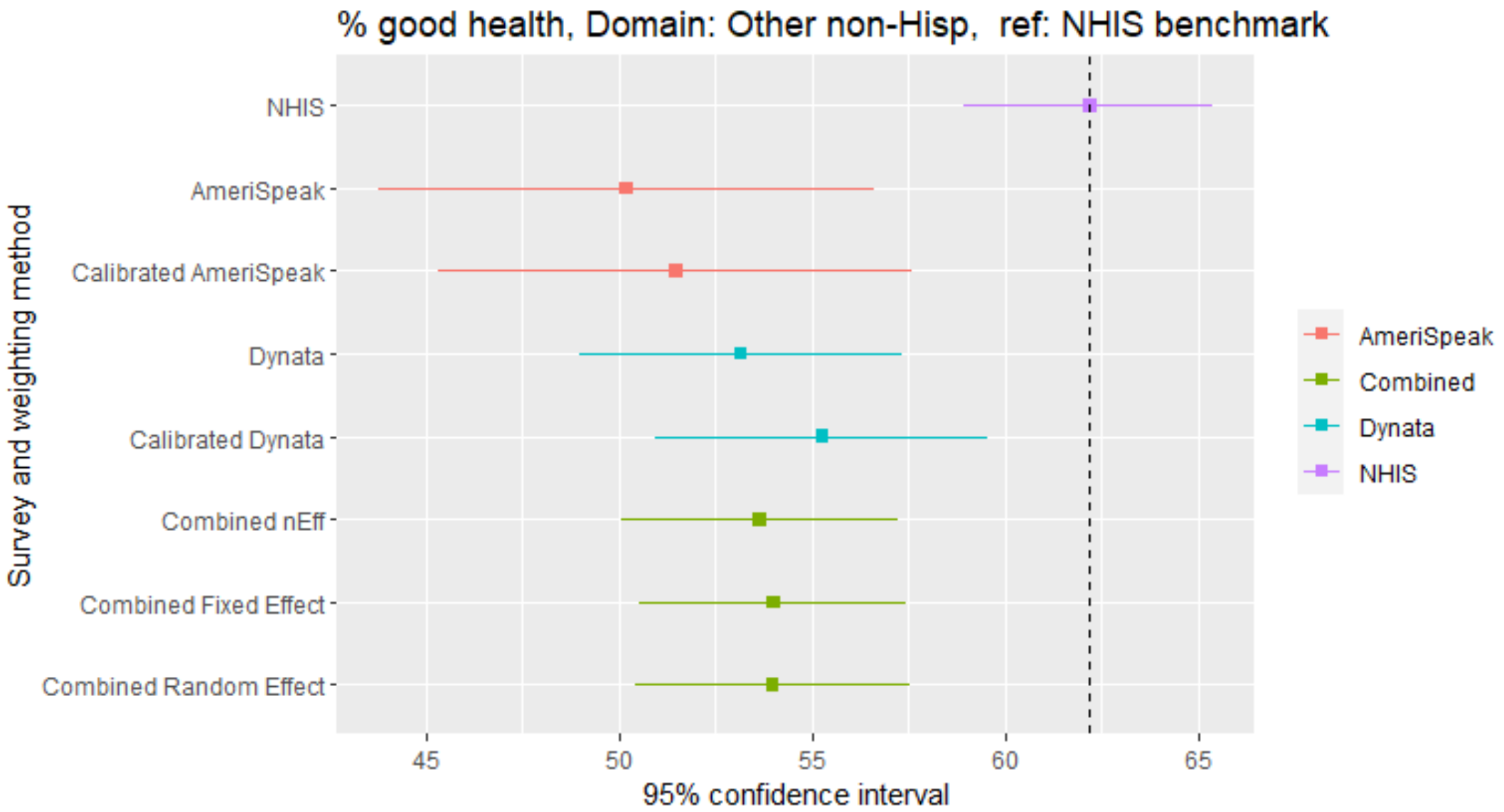
Graphical Results


% ever smoked, Domain: All, ref: NHIS benchmark
Survey and weighting method
NHIS
AmeriSpeak
Calibrated AmeriSpeak
Dynata
Calibrated Dynata
Combined nEff
Combined Fixed Effect
Combined Random Effect
35
40
45
95% confidence interval
AmeriSpeak
Combined
Dynata
NHIS


% ever smoked, Domain: White non-Hisp, ref: NHIS benchmark
Survey and weighting method
NHIS
AmeriSpeak
Calibrated AmeriSpeak
Dynata
Calibrated Dynata
Combined nEff
Combined Fixed Effect
Combined Random Effect
35
40
45
95% confidence interval
AmeriSpeak
Combined
Dynata
NHIS


% heart disease, Domain: All, ref: NHIS benchmark
Survey and weighting method
NHIS
AmeriSpeak
Calibrated AmeriSpeak
Dynata
Calibrated Dynata
Combined nEff
Combined Fixed Effect
Combined Random Effect
4.0
4.5
5.0
5.5
95% confidence interval
AmeriSpeak
Combined
Dynata
NHIS


% good health, Domain: All, ref: NHIS benchmark
Survey and weighting method
NHIS
AmeriSpeak
Calibrated AmeriSpeak
Dynata
Calibrated Dynata
Combined nEff
Combined Fixed Effect
Combined Random Effect
48
52
56
60
95% confidence interval
AmeriSpeak
Combined
Dynata
NHIS


% good health, Domain: Other non-Hisp, ref: NHIS benchmark
Survey and weighting method
NHIS
AmeriSpeak
Calibrated AmeriSpeak
Dynata
Calibrated Dynata
Combined nEff
Combined Fixed Effect
Combined Random Effect
45
50
55
60
65
95% confidence interval
AmeriSpeak
Combined
Dynata
NHIS

Evaluation of Methods

- In general, calibration tends to drive the stand-alone estimates toward the NHIS.
- Most estimates missed the benchmark.
- The “random effect” CI is a good diagnostic metric that captures the system differences when pooling.
- Global conclusions are difficult to make.
- Variable and domain selections make for special situations: directs us to further scrutiny of methodological components

Where things stand

- RSS round 1 data collection complete
- Separate calibrations
- Use the effective sample size method for combining.
- Work still in progress
- Data and experiences will be used to refine approaches taken in future rounds

Thanks!

Van Parsons: vparsons@CDC.GOV