Combining Probability ABS Data with Nonprobability Social Media Data
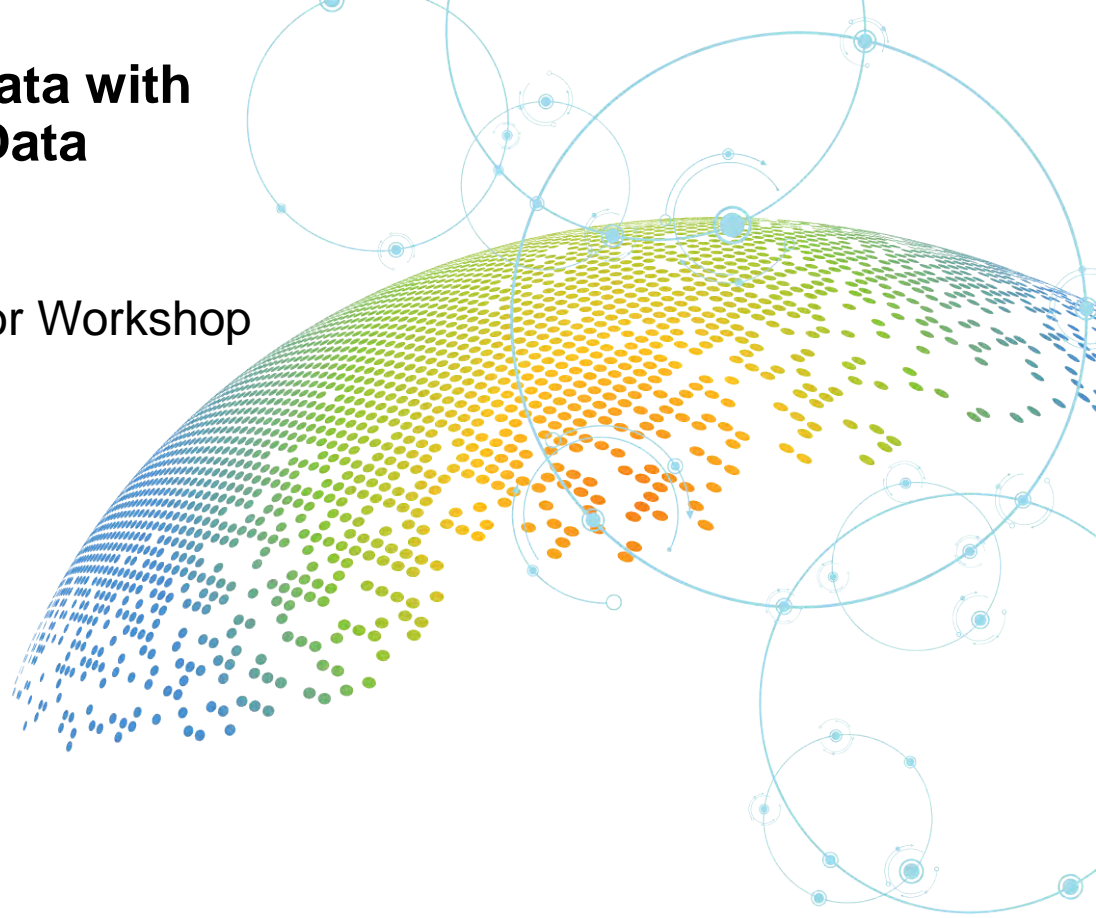
2024 International Total Survey Error Workshop
September 19, 2024

Burton Levine

2023 New State Adult Tobacco Survey (NY-ATS)—Sample Design

The NY-ATS is a yearly repeated cross-sectional sample of adults designed to capture tobacco use, behaviors and attitudes.

2023 ABS sample:

- 10 Strata formed by grouping blockgroups with similar yield rates.
  - Low responding strata were oversampled.
- Two-stage stratified SRS of addresses from the ABS frame.
  - First stage—households are selected
  - Second stage—1 adult within the household is randomly selected
- The selected adult was recruited to complete a web survey.
- 21,600 sample members fielded:
  - 5,094 respondents;
  - 25.4% RR

ABS—Issue with18-to-24-year-olds

18-to-24-year-olds (aka. Young adults) are a priority population.

In NY State, 11.7% of adults are 18-to-24-year-old.

Whereas 5.3% (268) ABS respondents are 18-to-24-year-old.

2 reasons why 18-to-24-year-olds are underrepresented

o 18-to-24-year-olds tend to live in larger households. Only 8% of households have an 18-to-24-year-olds.

o 18-to-24-year-olds respond at a lower rate than other age groups.

ABS—Issue with 18-to-24-year-olds

In 2022, without oversampling low responding yield rates 3.2% of respondents are 18-to-24-year-old. 5.3% is a 63% increase in 18-to-24-year-olds.

Yield rate decile	Respondents (%)	Nonrespondents (%)	Unknown response status (%)	Ineligible (%)	Unknown adjustment (e)	Response Rate (%)
1	16.5	2.6	78.1	2.7	0.87	18.9
2	18.2	2.8	76.9	2.1	0.91	20.0
3	20.8	2.3	74.8	2.2	0.91	22.7
4	20.8	2.4	75.1	1.7	0.93	22.4
5	22.1	3.5	72.4	2.1	0.93	23.9
6	25.4	3.6	69.1	1.9	0.94	27.1
7	27.0	2.8	68.3	1.9	0.94	28.7
8	28.3	3.8	65.6	2.3	0.93	30.3
9	30.4	3.5	64.8	1.3	0.96	31.5
10	34.0	3.6	60.7	1.6	0.96	35.5
Overall	**23.6**	**3.0**	**71.4**	**2.0**	**0.93**	**25.4**

# 2023 New State Adult Tobacco Survey (NY-ATS)—Sample Design

Social media sample:

- Nonprobability sample of 18-to-24-year-olds recruited from Facebook and Instagram.
- Quotas were implemented to obtain approximately 50% females.
- Subjects completed a web survey.
- Extensive fraud prevention and detection measures were implemented.
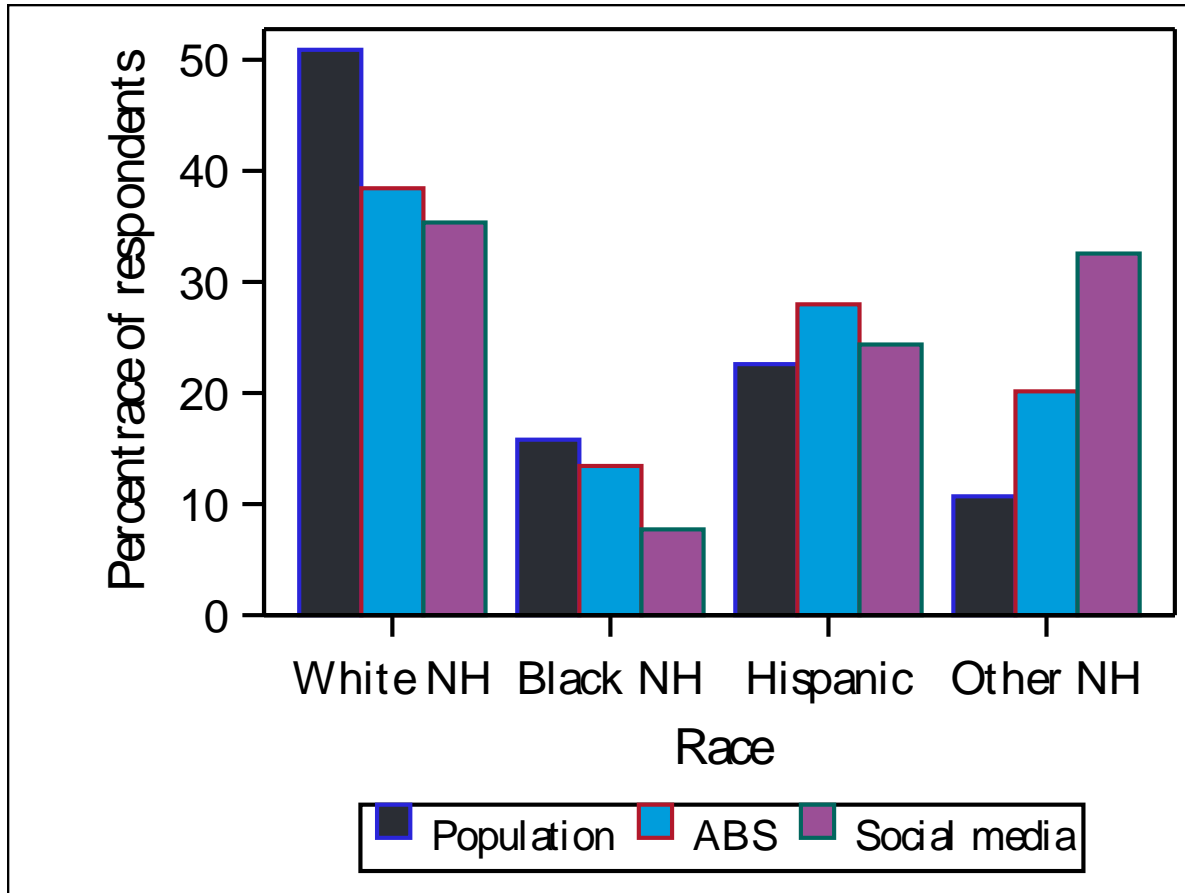- 1,502 respondents.

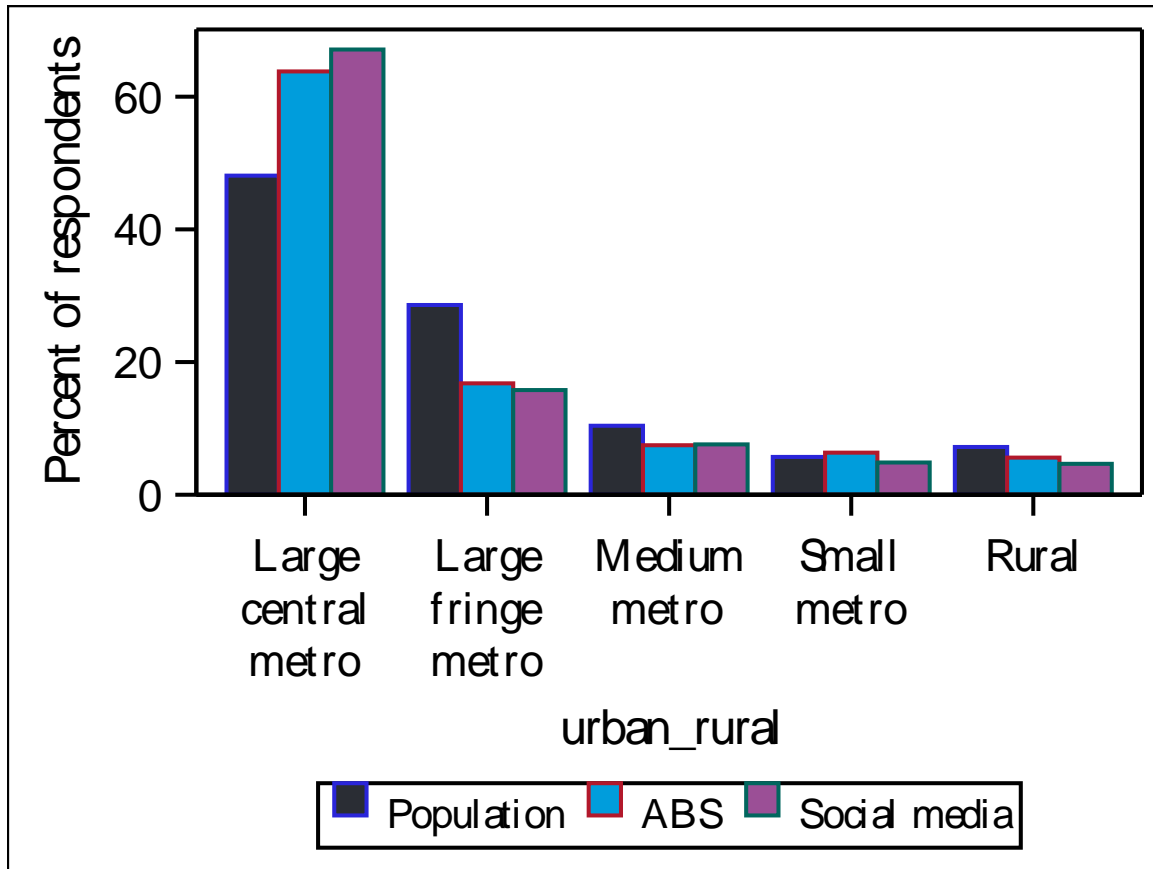# Compare Respondent Unweighted Distributions with Population—Age
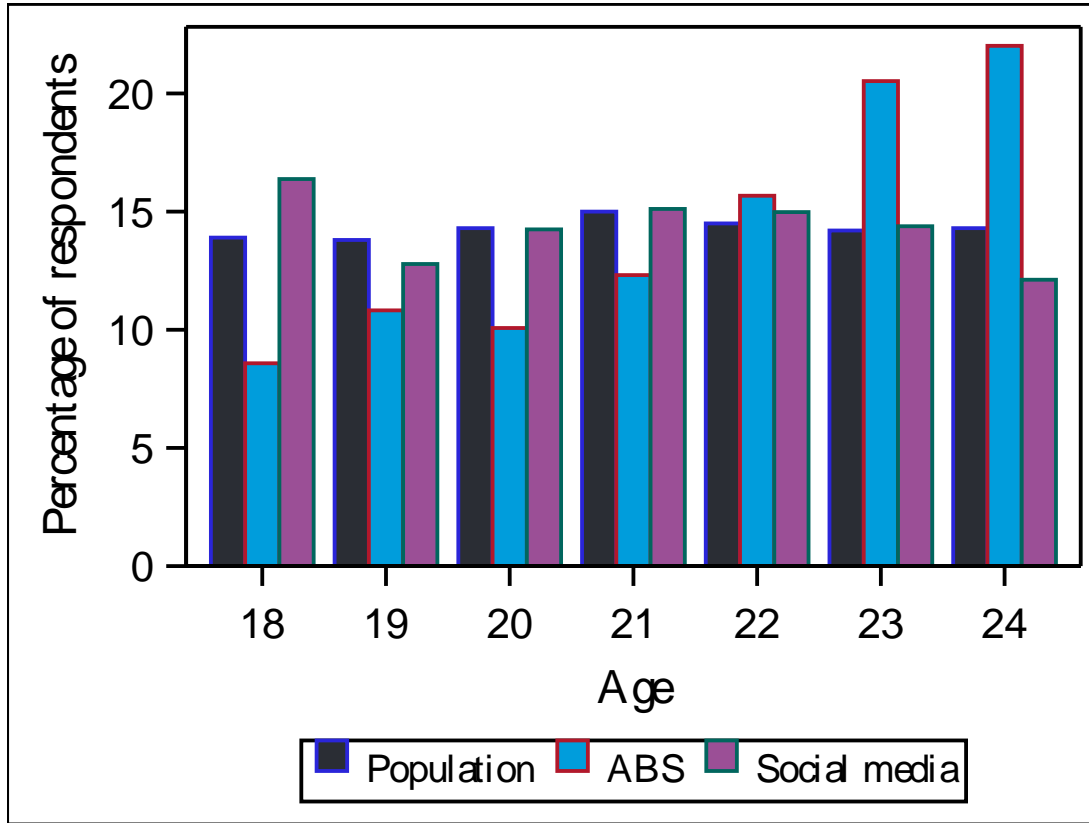
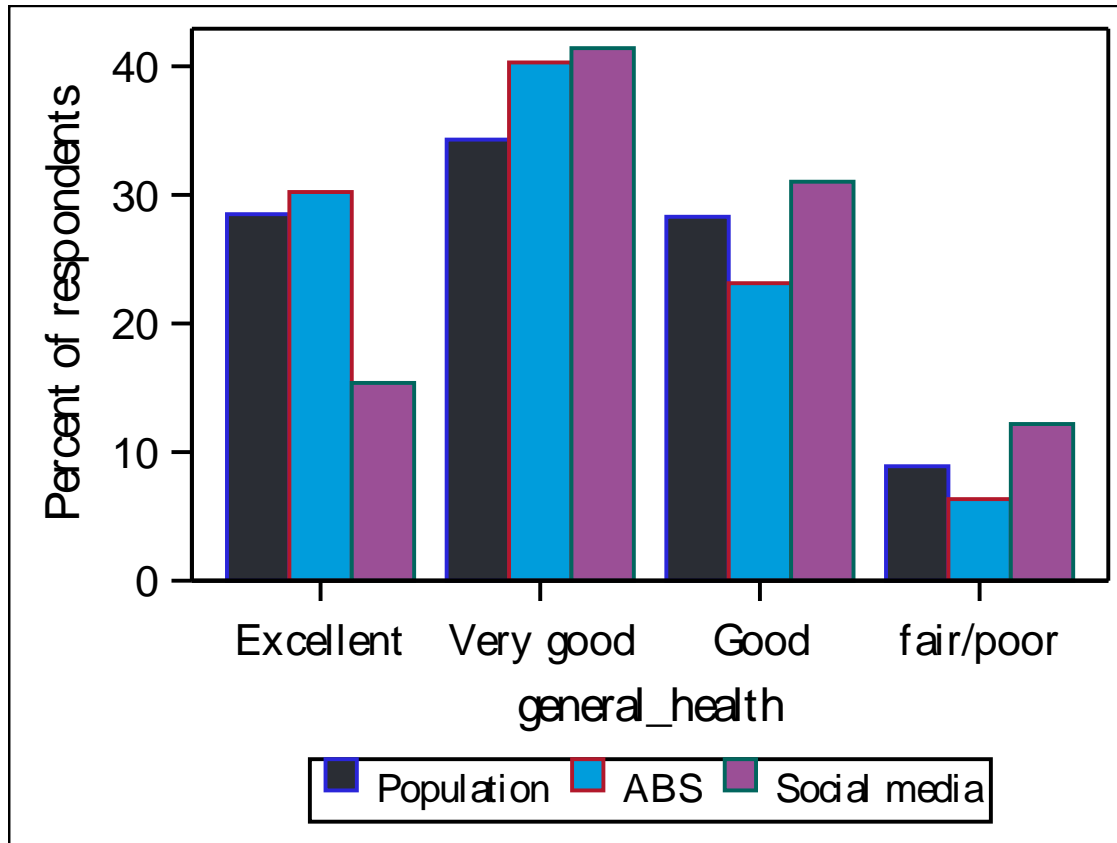# Compare Respondent Unweighted Distributions with Population—Sex
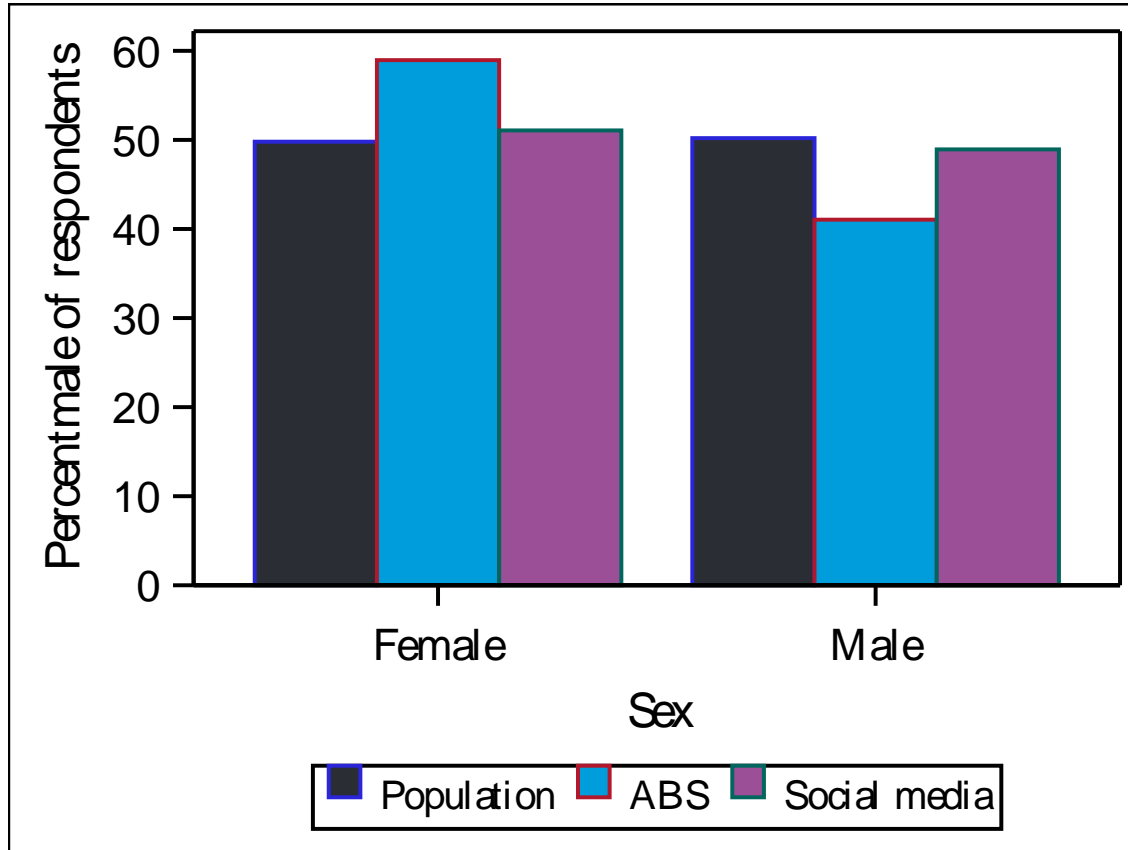
Compare Respondent Unweighted Distributions with Population—Race

Compare Respondent Unweighted Distributions with Population—General Health

Calculate Weights for ABS and Social Media Independently

Weighting step	ABS	SM	
Calculate base weight	$W_{i,j}^{Base} = \dfrac{N_i}{n_i}$	$W_i^{design} = \dfrac{N}{n}$	*i*-stratum *j*-individual
Unknown eligibility adj	$U_i^{adj} = \dfrac{n_i^r + n_i^n}{n_i^r + n_i^n + n_i^i}$	N/A	
Nonresponse adj	$NR_i^{adj} = \dfrac{\sum_{\forall \{r,n,u\} \in i} WT_{i,j}}{\sum_{\forall \{r\} \in i} WT_{i,j}}$	N/A	
HH size adj	$HH_{i,j}^{adj} = HH_{i,j}^{eligibles} * WT_{i,j}$	N/A	
Calibration	Yes	Yes	

Calibration distributions: sex, age, race/ethnicity, urban/rural classification, general health

Source of population totals: 2022 Current Population Data and 2022 BRFSS

Evaluate Respondent's Calibration Distributions Collectively with SCAI

The *standardized calibration adjustment index* (SCAI), collectively, quantifies the difference in the distributions of the variables used in the calibration of survey respondents to the target population, accounting for the study design.

Survey design	SCAI	Relative difference
ABS	27.0	13.7%
Social media	30.7	

Levine, Burton et al. 'Standardized Calibration Adjustment Index (SCAI): A New Measure of Survey Data Quality'. Statistical Journal of the IAOS, 1 Jan. 2021 : 603 – 610.

Evaluate Unequal Weighting Effect

The UWE is an upper bound of the ratio of the variance of an estimate calculated from a survey to the variance one would obtain from a simple random sample with the same sample size.

Survey design	UWE	Relative difference
ABS	1.58	12%
Social media	1.67	

Compare Tobacco Use Population Estimates between ABS and SM

Outcome	ABS (n=268)	SM (n=1,502)	P-value
Smoked 100 cigarettes	6.4 (2.7, 10.1)	8.2 (6.5, 9.8)	0.43
Current smoking	2.2 (0.05, 4.0)	5.8 (4.5, 7.1)	0.02*
Current e-cig	13.5 (8.3, 18.7)	17.8 (15.3, 20.2)	0.17
Cigar smoker	8.3 (3.7, 13.0)	11.4 (9.3, 13.5)	0.28

Hypothesis—A Reason for Lower Tobacco Use (NR bias) in the ABS Sample

Hypothesis—Compared to the SM respondents, a higher proportion of 18-to-24-year-olds ABS respondents live with their parents.

- Households with only 18-to-24-year-olds respond to a mail contact at a much lower rate.

- Sometimes, in households with an 18-to-24-year-old and a parent:
  - the parent opens the recruitment mailing,
  - complies with the household selection,
  - the 18-to-24-year-old gets selected,
  - and the 18-to-24-year-olds responds based on the parent's prompting.

18-to-24-year-olds living with their parents have lower rates of tobacco use then 18-to-24-year-olds not living with their parents.

Participation in Social Media Might Increase Tobacco Use

"A systematic review and meta-analysis of 29 studies showed that participants who were exposed to tobacco content on social media, compared with those who were not exposed, had greater odds of reporting lifetime tobacco use, past 30-day tobacco use, and susceptibility to use tobacco among never users."

Donaldson SI, Dormanesh A, Perez C, Majmundar A, Allem J. Association Between Exposure to Tobacco Content on Social Media and Tobacco Use: A Systematic Review and Meta-analysis. *JAMA Pediatr.* 2022;176(9):878–885. doi:10.1001/jamapediatrics.2022.2223

Two Ways to Combine the ABS and SM data

(1) Composite the **analysis weights** proportional to the sample size.

(2) Composite the **design weights** proportional to the sample size. Then calibrate the combined data.

$$ABS^{adj} = \frac{n^{ABS}}{n^{ABS} + n^{SM}} = \frac{268}{268 + 1502} = 0.151$$

$$SM^{adj} = \frac{n^{SM}}{n^{ABS} + n^{SM}} = \frac{1502}{268 + 1502} = 0.849$$

Other ways to composite:

(1) by effective sample size, (2) minimizing mean squared error

Compare Tobacco Use Outcomes between 2 Ways to Combine ABS and SM Data

Outcome	Composite design WTS then **calibrate collectively**	**Calibrate** analysis weights **independently** then composite
Smoked 100 cigarettes	8.2 (6.7, 9.7)	7.9 (6.4, 9.4)
Current smoking	5.7 (4.4, 6.9)	5.2 (4.0, 6.4)
Current e-cig	17.6 (15.3, 19.8)	17.1 (14.9, 19.4)
Cigar smoker	11.2 (9.3, 13.1)	10.9 (9.0, 12.9)

Compare Estimates—Single Source vs. Combined

Outcome	Estimates using 1 data source		ABS/SM combined estimates	
	ABS-only	SM-only	Calibrate collectively	Calibrate independently
Smoked 100 cigarettes	6.4 (2.7, 10.1)	8.2 (6.5, 9.8)	8.2 (6.7, 9.7)	7.9 (6.4, 9.4)
Current smoking	2.2 (0.05, 4.0)	5.8 (4.5, 7.1)	5.7 (4.4, 6.9)	5.2 (4.0, 6.4)
Current e-cig	13.5 (8.3, 18.7)	17.8 (15.3, 20.2)	17.6 (15.3, 19.8)	17.1 (14.9, 19.4)
Cigar smoker	8.3 (3.7, 13.0)	11.4 (9.3, 13.5)	11.2 (9.3, 13.1)	10.9 (9.0, 12.9)

Thank you

Contact: Burton Levine   email: blevine@rti.org