Marketing Systems Group

Composite Weighting For Hybrid Samples

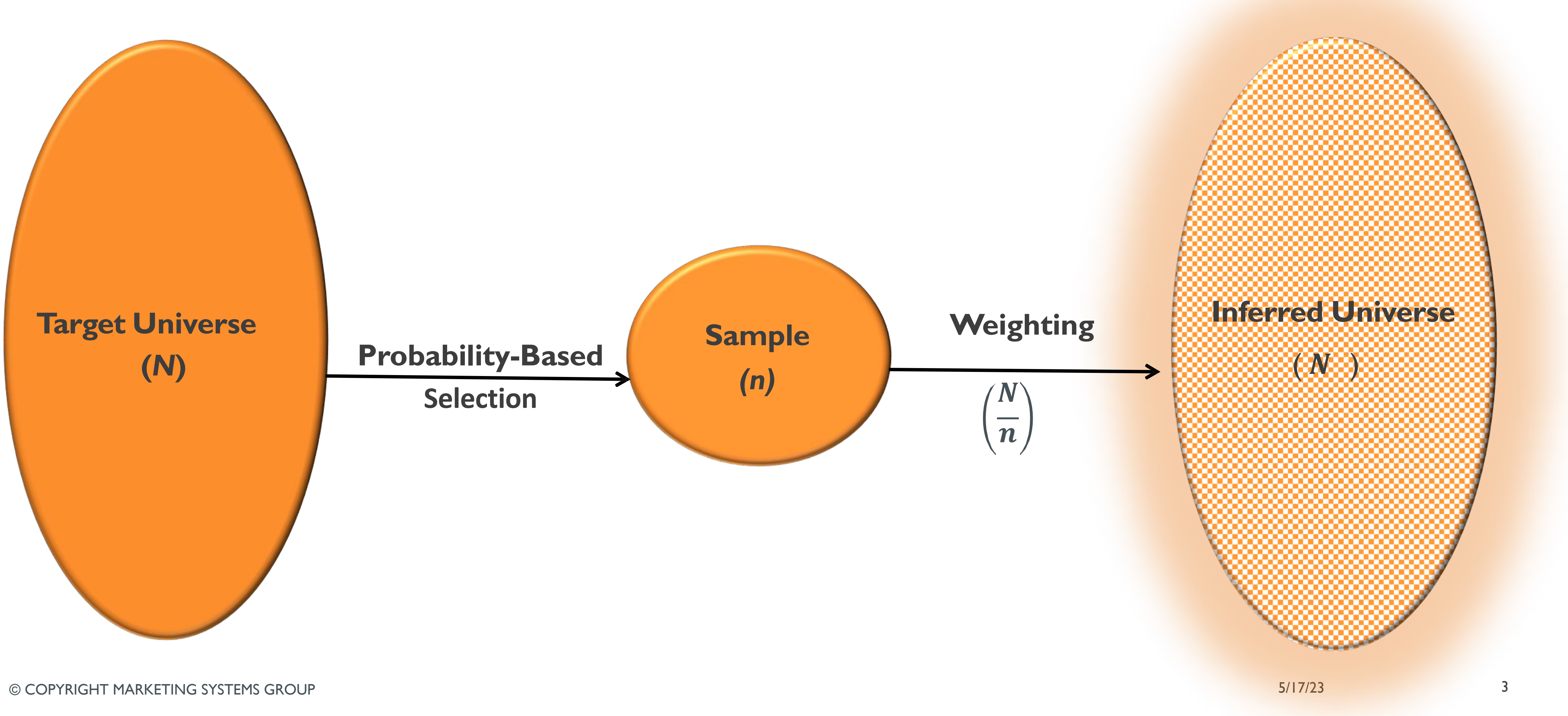
Mansour Fahimi, Ph.D.
EVP & Chief Data Scientist
Marketing Systems Group

AAPOR
AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH

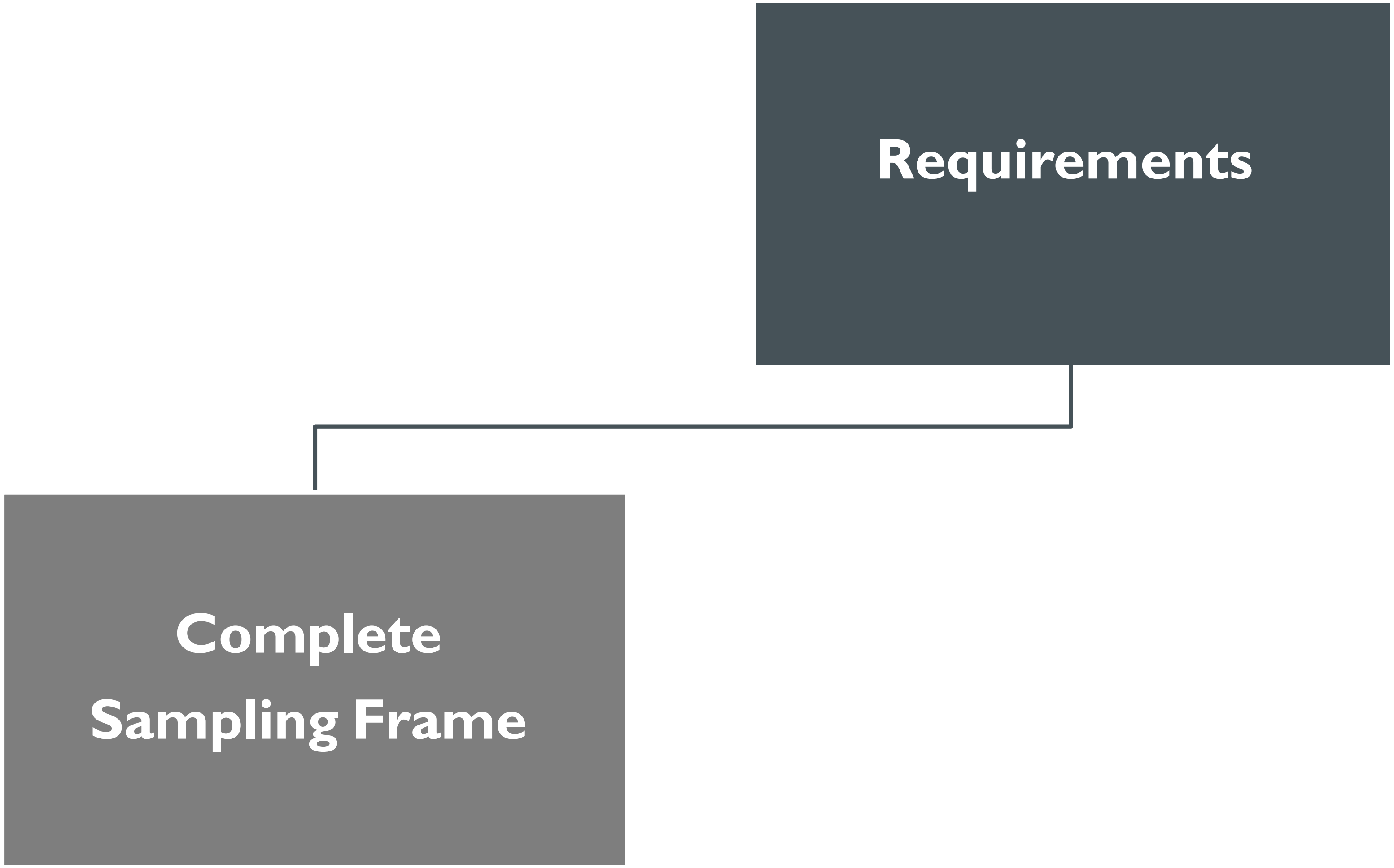
PRESENTATION OUTLINE

- **Traditional Inferential Paradigm**
- **New Realities in the Digital Age**
- **Integration of Surveys**
  - **Improvisations**
  - **Composite Estimation**
  - **Composite Weighting**
- **Takeaways**

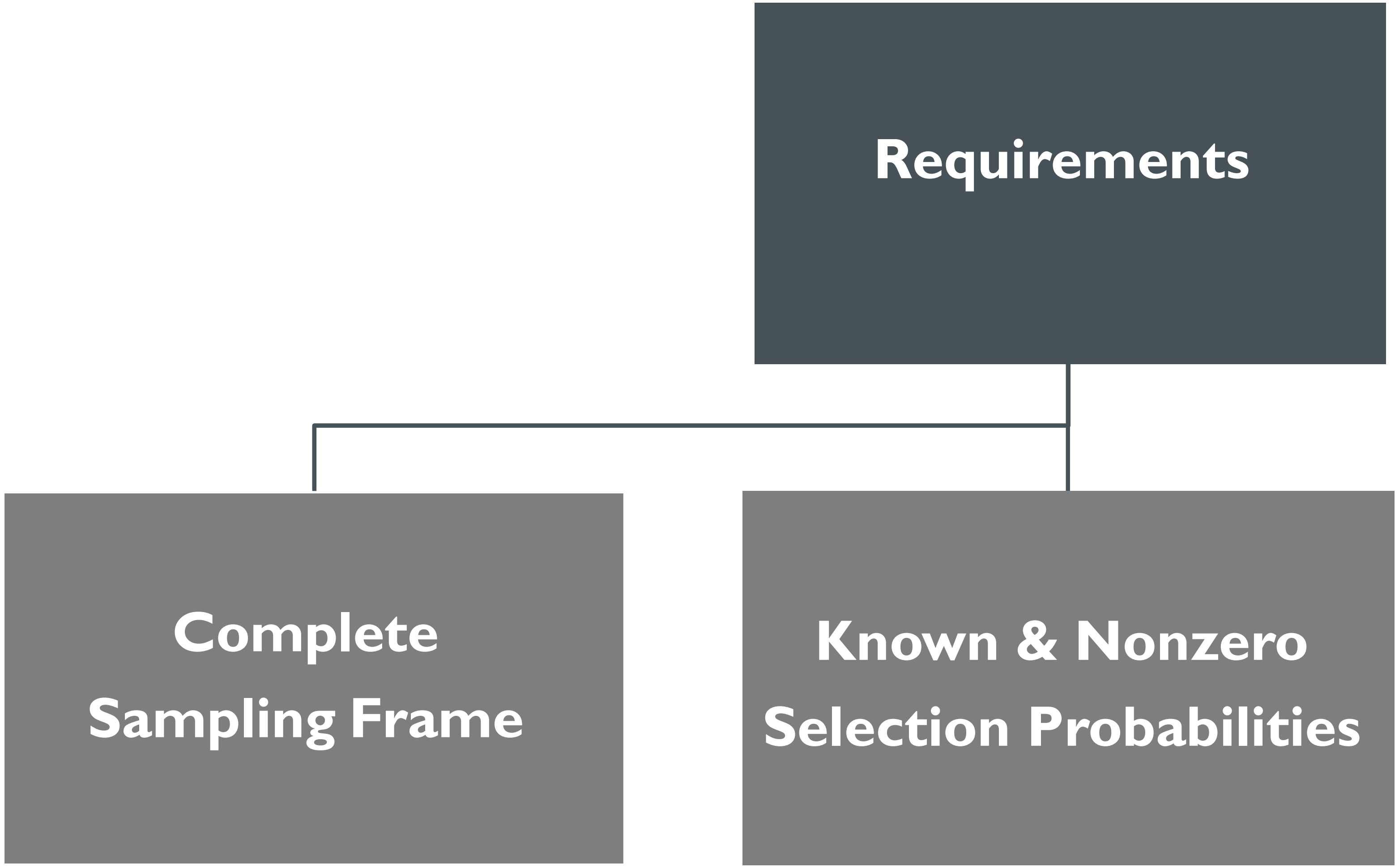
TRADITIONAL INFERENTIAL PARADIGM



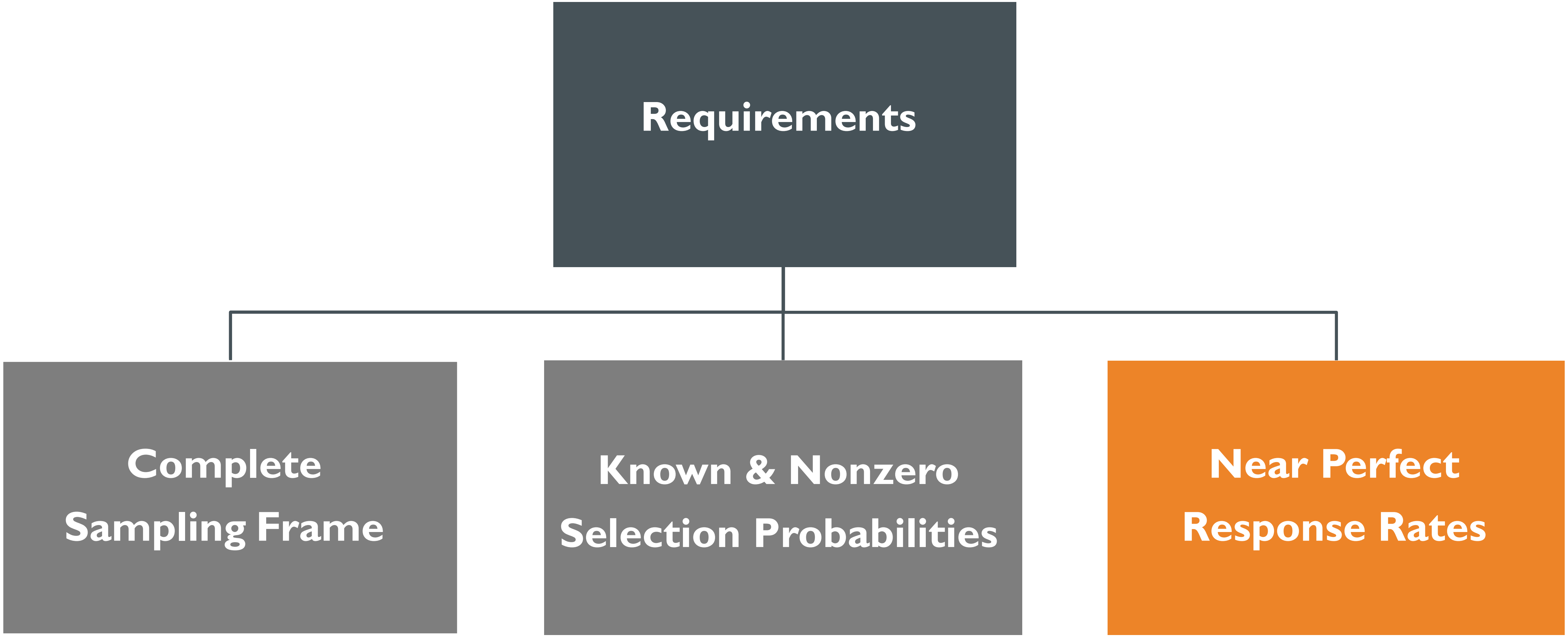
TRADITIONAL INFERENTIAL PARADIGM



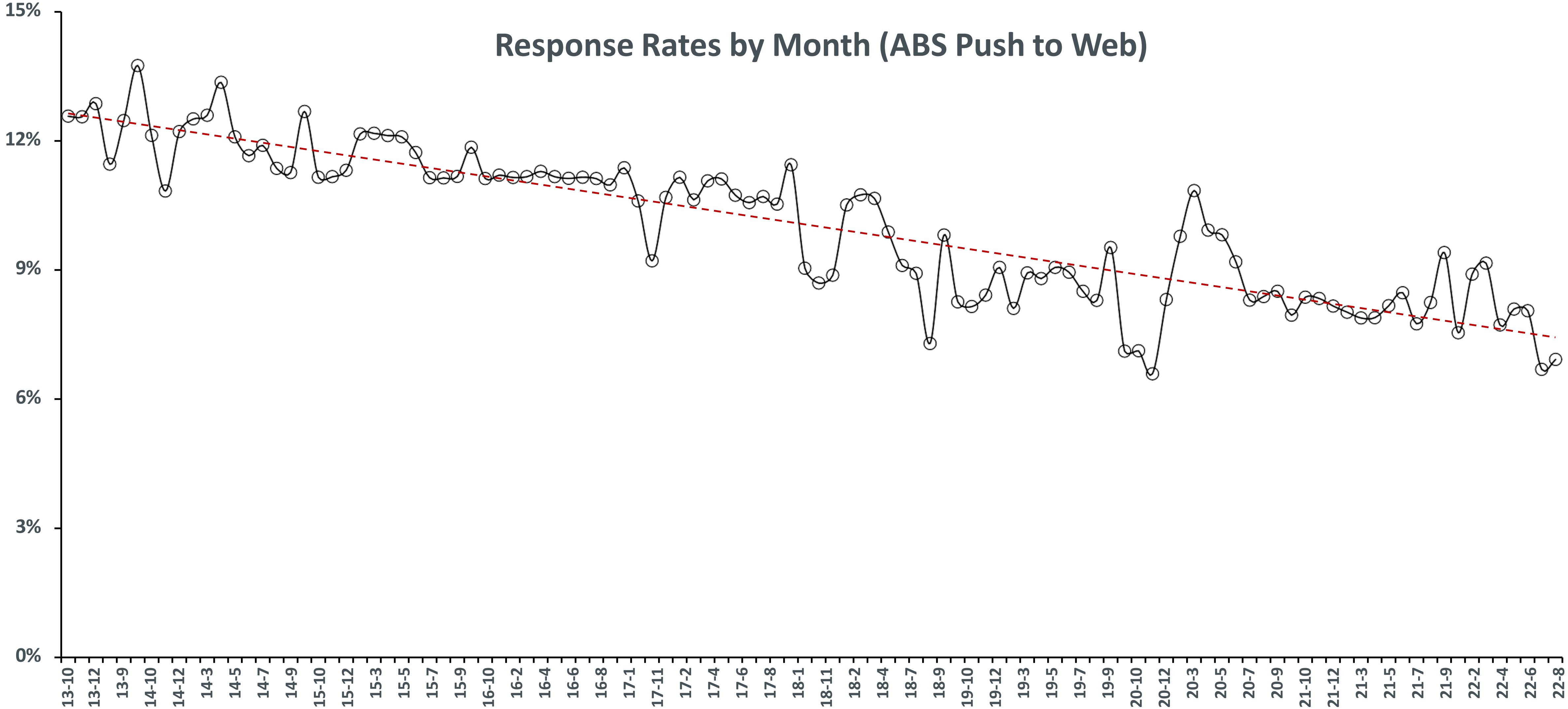
TRADITIONAL INFERENTIAL PARADIGM



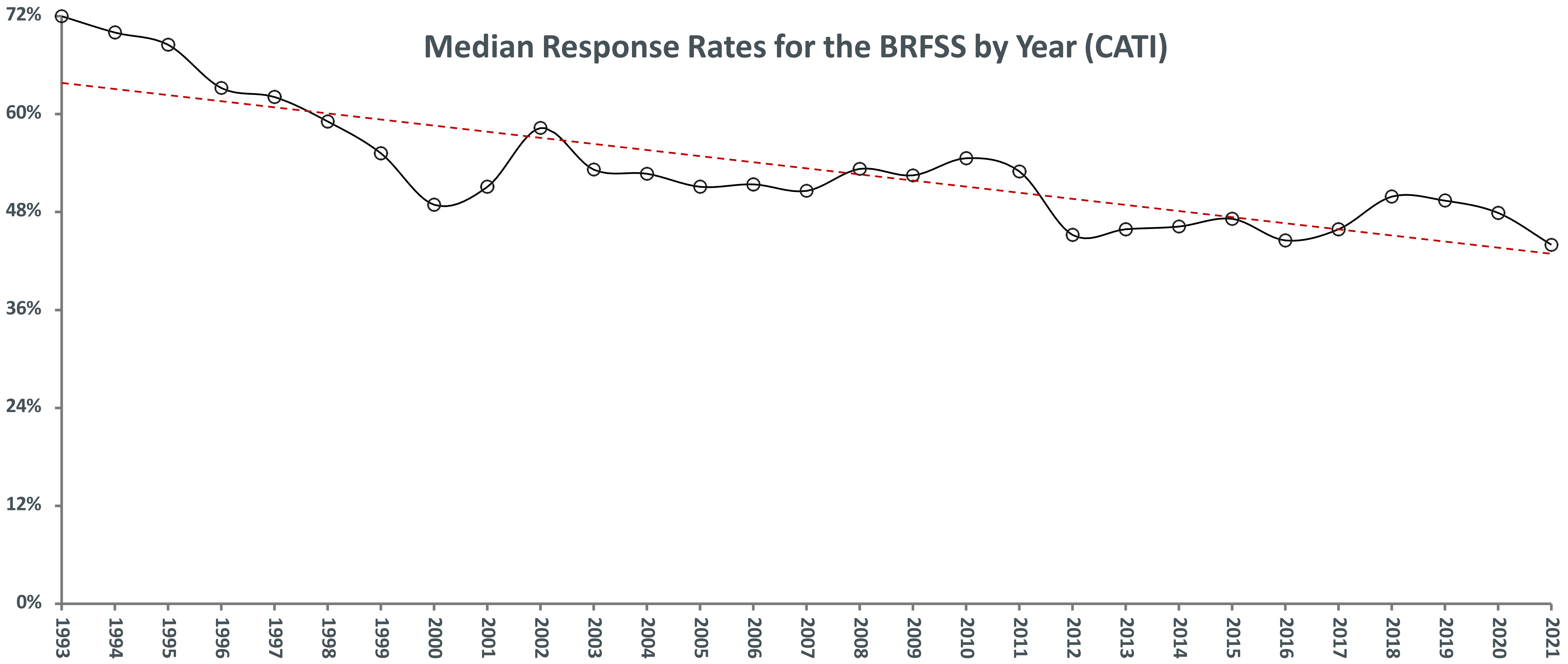
TRADITIONAL INFERENTIAL PARADIGM



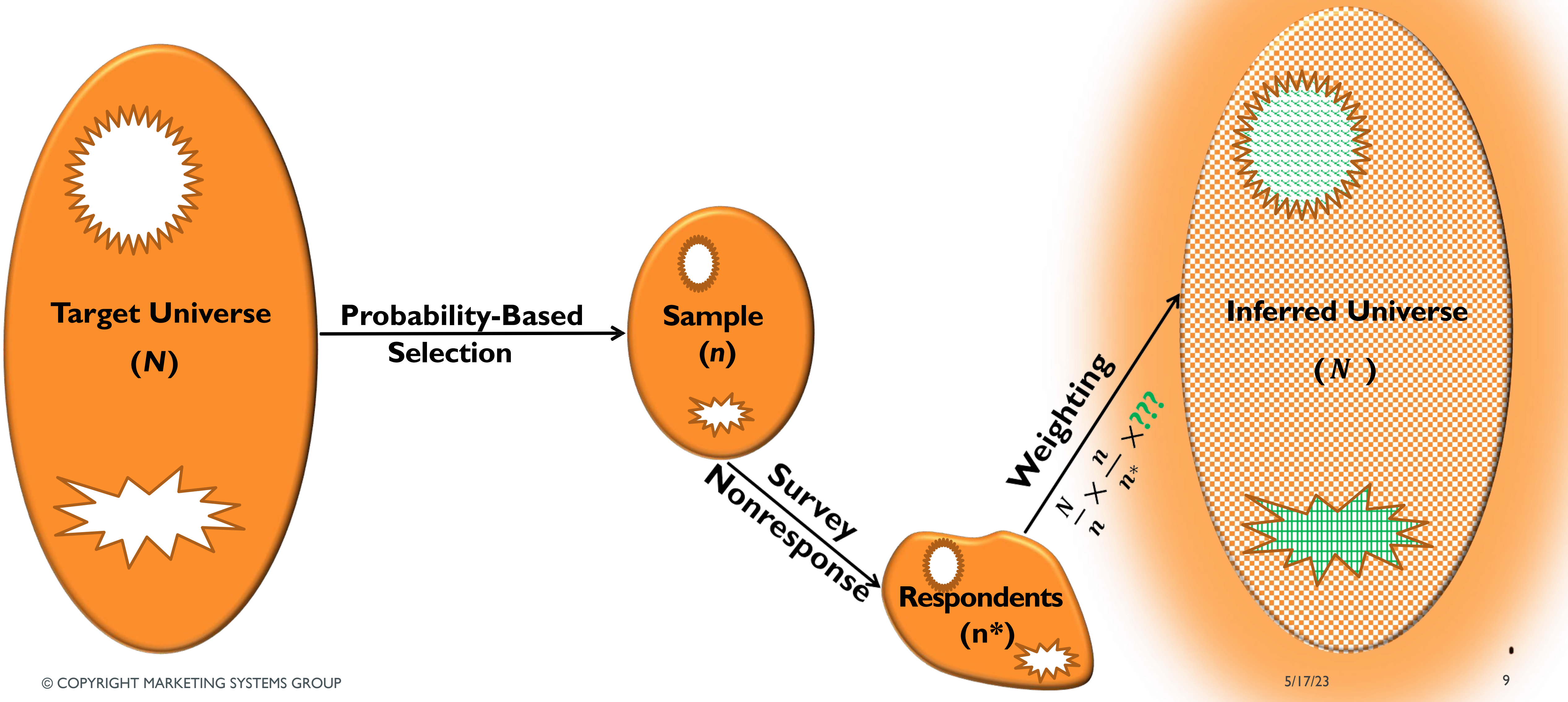
NEW REALITIES IN THE DIGITAL AGE



NEW REALITIES IN THE DIGITAL AGE



NEW REALITIES IN THE DIGITAL AGE



NEW REALITIES IN THE DIGITAL AGE

- **Multi-Mode:**
  - Was considered hasty but now encouraged to improve coverage
  - Different modes appeal to different cohorts
- **Multi-Frame Samples:**
  - Was considered worthy of commercial surveys but now is growing in popularity
  - Can improve coverage and reduce survey cost
- **Hybrid Samples:**
  - Probability samples: RDD + ABS
  - Semi-probability samples: RDD or ABS + nonprobability
  - Nonprobability samples: nonprobability + nonprobability

INTEGRATION OF SURVEYS

- **Improvisations:**
  - Stack unweighted surveys
  - Stack weighted surveys
- **Composite Estimation:**
  - Weight individual surveys separately
  - Produce point estimates from separate surveys
  - Blend individual estimates together
- **Composite Estimation Issues:**
  - Resource intensive
  - Inferentially inefficient
  - Not conducive to small domain estimations
- **Composite Weighting:**
  - Combining different survey components together
  - Weighting the combined data as a single survey

COMPOSITE ESTIMATION

Two Independent Sample Surveys:

- Sample sizes: n_1 and n_2
- Estimating population mean $\bar{Y}$ via $\bar{y}_1$ and $\bar{y}_2$

Composite Estimator: $\bar{y} = \alpha\bar{y}_1 + (1-\alpha)\bar{y}_2$

If α = 0.5 Then: $\bar{y} = \frac{\bar{y}_1+\bar{y}_2}{2}$

Optimal α Requires Minimizing: $MSE(\bar{y}) = V(\bar{y}) + [B(\bar{y})]^2$

If $B(\bar{y}_1) > 0, B(\bar{y}_2) > 0, V(\bar{y}_1) \neq V(\bar{y}_2)$ **Then:**

$$\alpha_{optimal} = \frac{MSE(\bar{y}_2)}{MSE(\bar{y}_1)+MSE(\bar{y}_2)} = \frac{V(\bar{y}_2)+[B(\bar{\boldsymbol{y}}_2)]^2}{V(\bar{y}_1)+[B(\bar{\boldsymbol{y}}_1)]^2+V(\bar{y}_2)+[B(\bar{\boldsymbol{y}}_2)]^2}$$

COMPOSITE ESTIMATION

If $B(\bar{y}_1) = 0$ **but** $B(\bar{y}_2) \neq 0, V(\bar{y}_1) \neq V(\bar{y}_2)$ **Then:**

$$\alpha_{optimal} = \frac{MSE(\bar{y}_2)}{V(\bar{y}_1)+MSE(\bar{y}_2)} = \frac{V(\bar{y}_2)+[B(\bar{y}_2)]^2}{V(\bar{y}_1)+V(\bar{y}_2)+[B(\bar{y}_2)]^2}$$

If $B(\bar{y}_1) = 0$ **and** $B(\bar{y}_2) = 0, V(\bar{y}_1) \neq V(\bar{y}_2)$ **Then:**

$$\alpha_{optimal} = \frac{V(\bar{y}_2)}{V(\bar{y}_1) + V(\bar{y}_2)}$$

If $B(\bar{y}_1) = B(\bar{y}_2) = 0$ **and** $V(\bar{y}_1) \approx V(\bar{y}_2)$ **Then:**

$$\alpha_{optimal} = \frac{n_1}{n_1 + n_2} = \frac{n_1}{n}$$

COMPOSITE ESTIMATION

Age	Smoking Behavior	Survey Estimate			
		Survey 1 (n = 3,361)		Survey 2 (n = 3,788)	
		Prevalence		Prevalence	
Overall	Cigarette	5.6%		16.9%	
	E-cigarette	13.5%		20.8%	
15- 17	Cigarette	2.3%		5.5%	
	E-cigarette	9.1%		10.4%	
18 - 24	Cigarette	7.0%		22.1%	
	E-cigarette	15.5%		25.5%	

COMPOSITE ESTIMATION

Age	Smoking Behavior	Survey Estimate				$\alpha = \frac{1}{2}$	
		Survey 1 (n = 3,361)		Survey 2 (n = 3,788)		Factor	Estimate
		Prevalence		Prevalence			
Overall	Cigarette	5.6%		16.9%		0.5	11.24%
	E-cigarette	13.5%		20.8%			17.15%
15- 17	Cigarette	2.3%		5.5%			3.94%
	E-cigarette	9.1%		10.4%			9.76%
18 - 24	Cigarette	7.0%		22.1%			14.54%
	E-cigarette	15.5%		25.5%			20.50%

COMPOSITE ESTIMATION

Age	Smoking Behavior	Survey Estimate				$\alpha = \frac{1}{2}$		$\alpha_{optimal} = \frac{V_2}{(V_1+V_2)}$	
		Survey 1 (n = 3,361)		Survey 2 (n = 3,788)		Factor	Estimate	Factor	Estimate
		Prevalence	Variance	Prevalence	Variance				
Overall	Cigarette	5.6%	0.00004	16.9%	0.00007	0.5	11.24%	0.64	9.7%
	E-cigarette	13.5%	0.00008	20.8%	0.00009		17.15%	0.52	17.0%
15- 17	Cigarette	2.3%	0.00004	5.5%	0.00009		3.94%	0.70	3.3%
	E-cigarette	9.1%	0.00017	10.4%	0.00017		9.76%	0.49	9.8%
18 - 24	Cigarette	7.0%	0.00007	22.1%	0.00012		14.54%	0.62	12.7%
	E-cigarette	15.5%	0.00013	25.5%	0.00015		20.50%	0.52	20.3%

COMPOSITE WEIGHTING

Individually Poststratified Weights:

$$\begin{cases} W_{1i} = B_{1i} \times \frac{N}{\sum_{i=1}^{n_1} B_{1i}}, i = 1, \dots, n_1 \\ W_{2j} = B_{2j} \times \frac{N}{\sum_{j=1}^{n_2} B_{2j}}, j = 1, \dots, n_2 \end{cases}$$

If $B(\bar{y}_1) = B(\bar{y}_2) = 0$ **and** $V(\bar{y}_1) \approx V(\bar{y}_2)$ **Then:**

$$\begin{cases} W_{1i}^* = W_{1i} \times \frac{n_1}{n} = B_{1i} \times \frac{N}{\sum_{i=1}^{n_1} B_{1i}} \times \frac{n_1}{n}, i = 1, \dots, n_1 \\ W_{2j}^* = W_{2j} \times \frac{n_2}{n} = B_{2j} \times \frac{N}{\sum_{j=1}^{n_2} B_{2j}} \times \frac{n_2}{n}, j = 1, \dots, n_2 \end{cases}$$

Use the Above as Base weights and Poststratify the Combined Sample

COMPOSITE WEIGHTING

If $B(\bar{y}_1) = 0,\ B(\bar{y}_2) = 0$**, and** $V(\bar{y}_1) \neq V(\bar{y}_2)$ **Then:**

$$\begin{cases} W_{1i}^{*} = W_{1i} \times \frac{n_1^*}{n^*} = B_{1i} \times \frac{N}{\sum_{i=1}^{n_1} B_{1i}} \times \frac{n_1^*}{n^*}, i = 1, \dots, n_1 \\ W_{2j}^{*} = W_{2j} \times \frac{n_2^*}{n^*} = B_{2j} \times \frac{N}{\sum_{j=1}^{n_2} B_{2j}} \times \frac{n_2^*}{n^*}, j = 1, \dots, n_2 \end{cases}$$

Where:

$$n_i^* = \frac{n_i}{\delta_i}$$

$$\delta_i = 1 + \frac{\sum_{k=1}^{n_i} \frac{(W_{ik} - \bar{W}_i)^2}{n_i - 1}}{\bar{W}_i^2}$$

COMPOSITE WEIGHTING

Combined Poststratification:

- Accept the calibrated pooled survey weights as base weights
- Apply some power transformation to reduce the variability of the base weights
- Poststratify the combined sample using the resulting base weights

Inferential Benefits:

- Can use more granular weighting adjustments
- All respondents will be weighted to a consistent set of benchmarks
- A consistent method will be used for variance estimations

Computational Benefits:

- A single set of weights to work with
- No piecemeal process to create dozens of composition factors and estimates
- Reduce "contamination" from smaller surveys

COMPOSITE WEIGHTING – NUMERICAL EXAMPLE

Age	Smoking Behavior	Survey Estimates		Composite Estimates		Composite Weighting
		Survey 1	Survey 2	$\alpha = \frac{1}{2}$	$\alpha = \frac{V_2}{(V_1+V_2)}$	
Overall	Cigarette	5.6%	16.9%	11.2%	9.7%	10.5%
	E-cigarette	13.5%	20.8%	17.2%	17.0%	16.3%
15- 17	Cigarette	2.3%	5.5%	3.9%	3.3%	3.6%
	E-cigarette	9.1%	10.4%	9.8%	9.8%	9.1%
18 - 24	Cigarette	7.0%	22.1%	14.5%	12.7%	13.6%
	E-cigarette	15.5%	25.5%	20.5%	20.3%	19.5%

TAKEAWAYS

New Realities of the Accelerating Digital Age:

- Traditional sampling methods have growing issues: coverage, time, and cost
- Perfect coverage and high response rates are dwindling luxuries
- Unwavering allegiance to traditional methods just for the sake of optics is getting old
- Alternative (hybrid) sampling methods are becoming more pragmatic

Making Reliable Inferences from Imperfect Data:

- All reasonable attempts must be made to improve representation
- Surveys need more comprehensive weighting due to low coverage and response rates
- Compensation methods for hybrid samples must be approachable and scalable
- Composite weighting offer a simple and effective approach for blending hybrid samples

THANK YOU

MANSOUR FAHIMI, PH.D. – MARKETING SYSTEMS GROUP