



Combining Probability and Nonprobability Samples to form Efficient Hybrid Estimates

Jill A. Dever, PhD

2018 FCSM Research and Policy Conference, March 7-9, Washington, DC

Research to Date

Probability samples

- Low response rates for many (Kohut et al. 2012)
- Ever increasing costs with ever decreasing funds

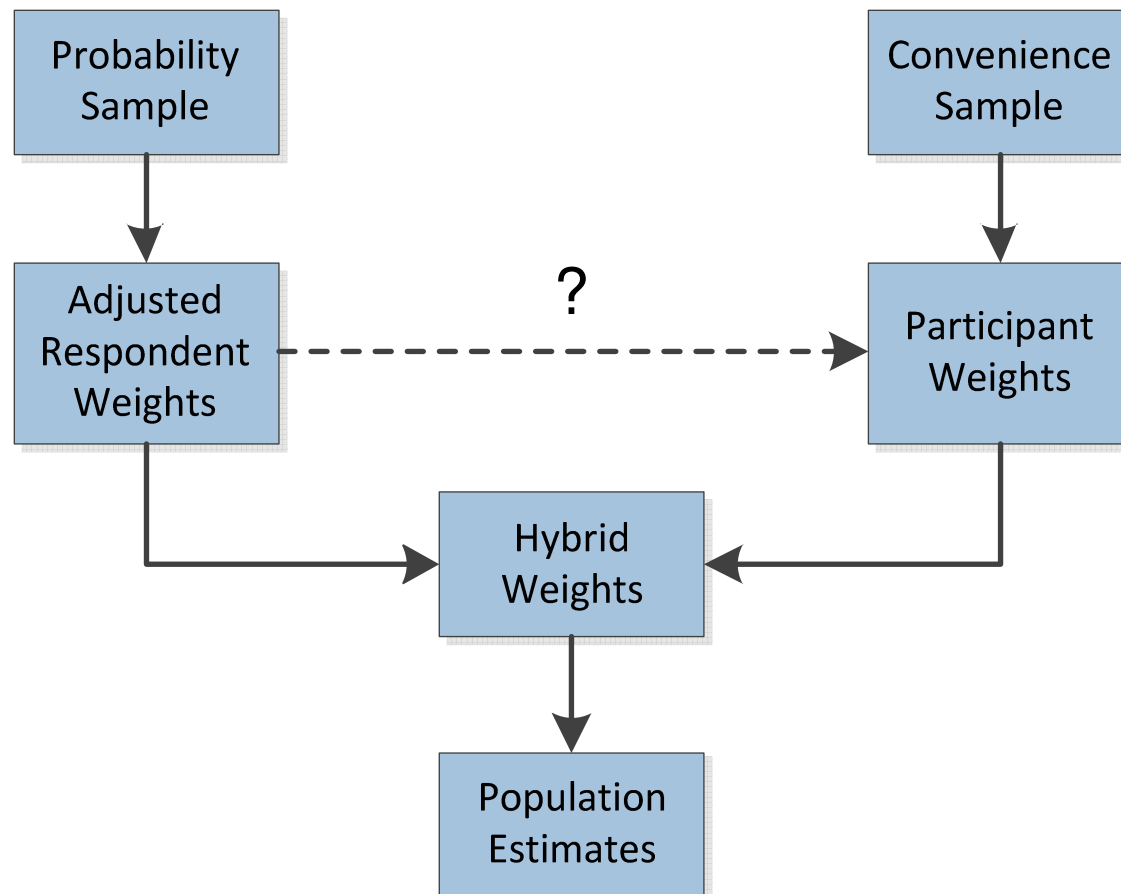
Nonprobability (convenience) samples

- Data are everywhere just waiting to be analyzed
- Population inference not viable (yet?)

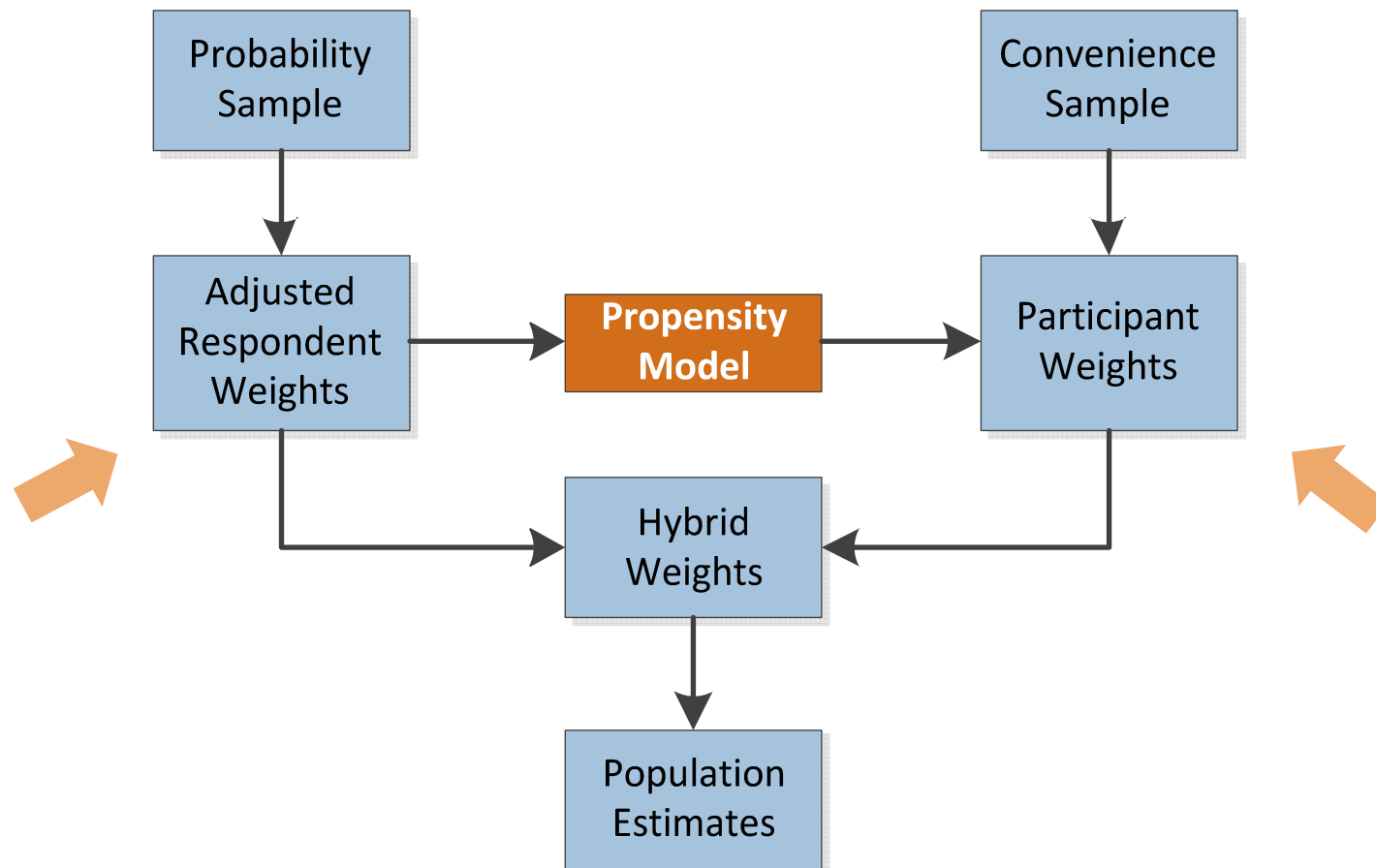
Hybrid samples

- Combine probability and nonprobability samples
- Maintains some continuity with existing Federal surveys
- Improve precision in underrepresented populations

Hybrid Estimation with Weights



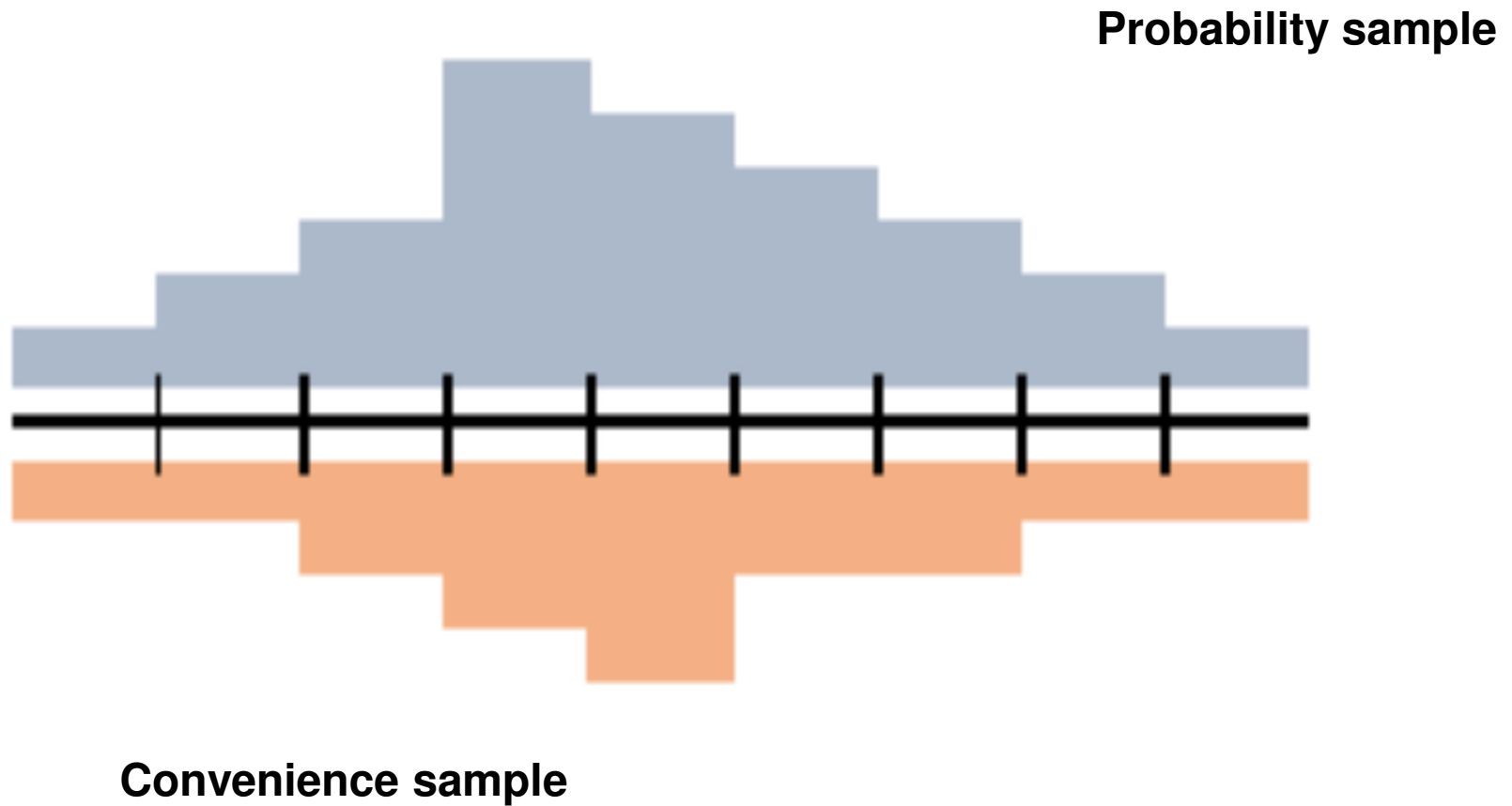
Hybrid Estimation with Weights



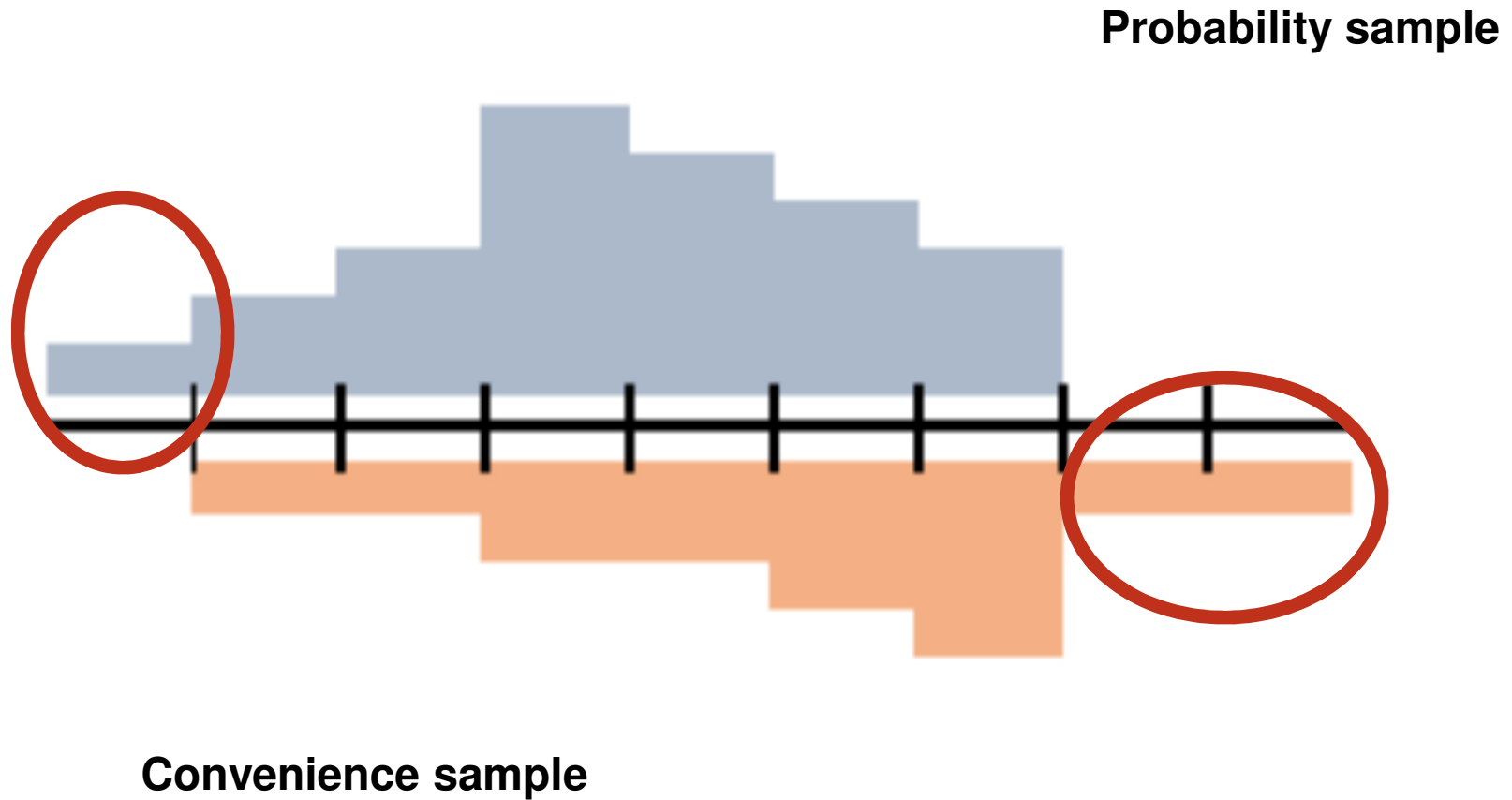
Assumptions for Hybrid Estimation with Weights

- Commonality in the questionnaires
- Nonparticipants are a random subsample (MAR)
- Single set of weights for analyses
- “Good” reference survey from the target population for calibration (e.g., raking, poststratification)
- Effective composite factor (the “glue”)
- Criteria for including/excluding convenience sample
- **Common support**

Common Support



Not-So-Common Support

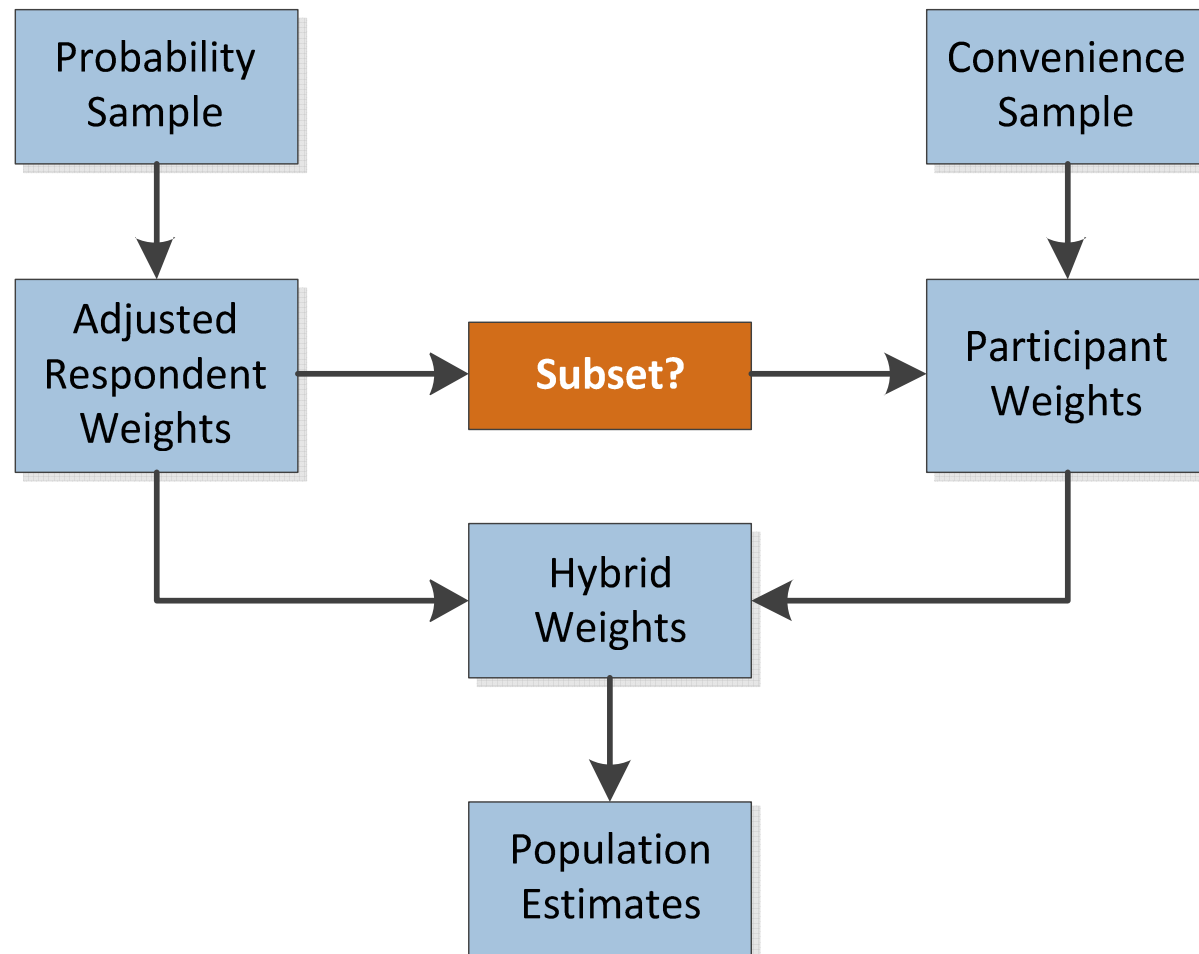


Propensity Scores for a Convenience Sample

- Estimate pseudo-inclusion probability $\pi_k = \Pr\{k \in s\}$ via binary regression
 - Input nonprobability weights = 1
 - Probability survey weights are fully adjusted (Valliant & Dever 2011)

- Use a function of $1/\pi_k$ as the convenience weight
 - Individual propensity weights
 - Various functions of propensity weights

Hybrid Estimation with Weights – Current Research



National Marijuana Beliefs and Behaviors Study

RTI-funded national study of adults 18+ years

Probability sample ($n = 1,867$)

- *Address-based sample* (ABS) with mail/web data collection
- Strata (4) = recreational, medical-liberal, medical-restrictive, other
- Recruitment via mail for single adult household respondent

Convenience sample ($n = 4,943$)

- Adult participants from *social media* (e.g., Facebook)
- Recruitment via advertisement / referral for web interview

National Marijuana Beliefs and Behaviors Study

<u>Characteristics</u>	HIGHER PROPORTION	
	<u>ABS</u>	<u>Social Media</u>
Male	✓	
Bachelors or higher	✓	
65 years and older / Retirees	✓	
Voter (regular)	✓	
Private insurance	✓	
Hispanic/Latin ethnicity		✓
34 years and younger		✓
Child in household		✓
Liberal views		✓
Access to internet		✓
Smoke/vape last 30 days		✓
Marijuana ever used		✓

ABS = Address-based sampling

National Marijuana Beliefs and Behaviors Study

Evaluation of Common Support Assumption

- **MatchIt** in **R** used to evaluate common support
- Nearest Neighbor matching with no caliper (*liberal*)

Record matching results:

Model covariates	ABS matched to Social Media
Strata (4)	92 %
Strata (4) + 12 Characteristics	66 %

National Marijuana Beliefs and Behaviors Study

Propensity model options:

- Ignore the common-support analyses
- Include only matched ABS cases (92%, 66%)

Binary analytic variables:

- Ever used marijuana
- Support legalizing marijuana for medical use

National Marijuana Beliefs and Behaviors Study

Gold Standard

ABS Sample Only	Ever Used	Support Medical Use
Full sample	56.6	77.0

Difference

ABS Sample Only	Ever Used	Support Medical Use
92% matched subset	-0.6	0.3
66% matched subset	4.2	1.6

National Marijuana Beliefs and Behaviors Study

Gold Standard

ABS Sample Only	Ever Used	Support Medical Use
Full sample	56.6	77.0

Difference

Social Media Sample Only	Ever Used	Support Medical Use
No weights	13.4	8.7
Convenience Weights:		
Full ABS sample	-1.2	6.4
92% matched ABS subset	-1.3	6.4
66% matched ABS subset	0.2	7.6

Conclusions and Next Steps

Common Support

- Assumption needs to be assessed

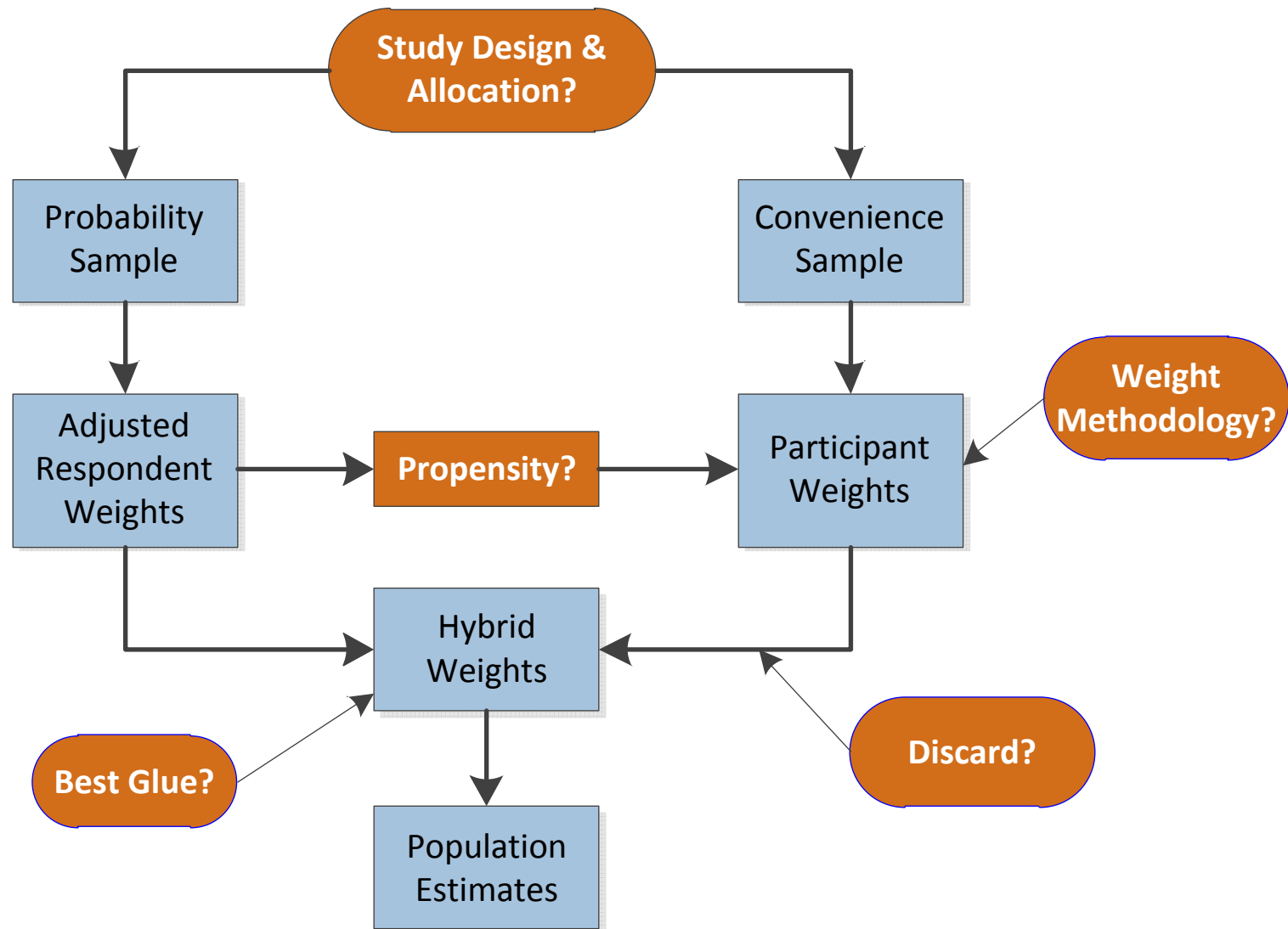
Next: Evaluate methods to test assumption

Propensity Scores

- Some bias reduction with constrained estimation space

Next: Evaluate Elliott (2009), Robbins et al (2018)

Next Steps with Hybrid Estimation



Combining Probability and Nonprobability Samples to form Efficient Hybrid Estimates

2018 FCSM Research and Policy Conference, March 7-9, Washington, DC

Jill A. Dever, PhD

Senior Research Statistician

RTI International

Washington, DC

jdever@rti.org