Considerations of Data Diversity and Equity in Probability Sampling and Survey Research

10.24.24

David Dutwin, SVP, Chief Scientist AmeriSpeak; Director Center for Panel Survey Sciences
Ashani Johnson-Turbes, VP and Director of the Center on Equity Research
J. Michael Dennis, SVP and Executive Director of AmeriSpeak

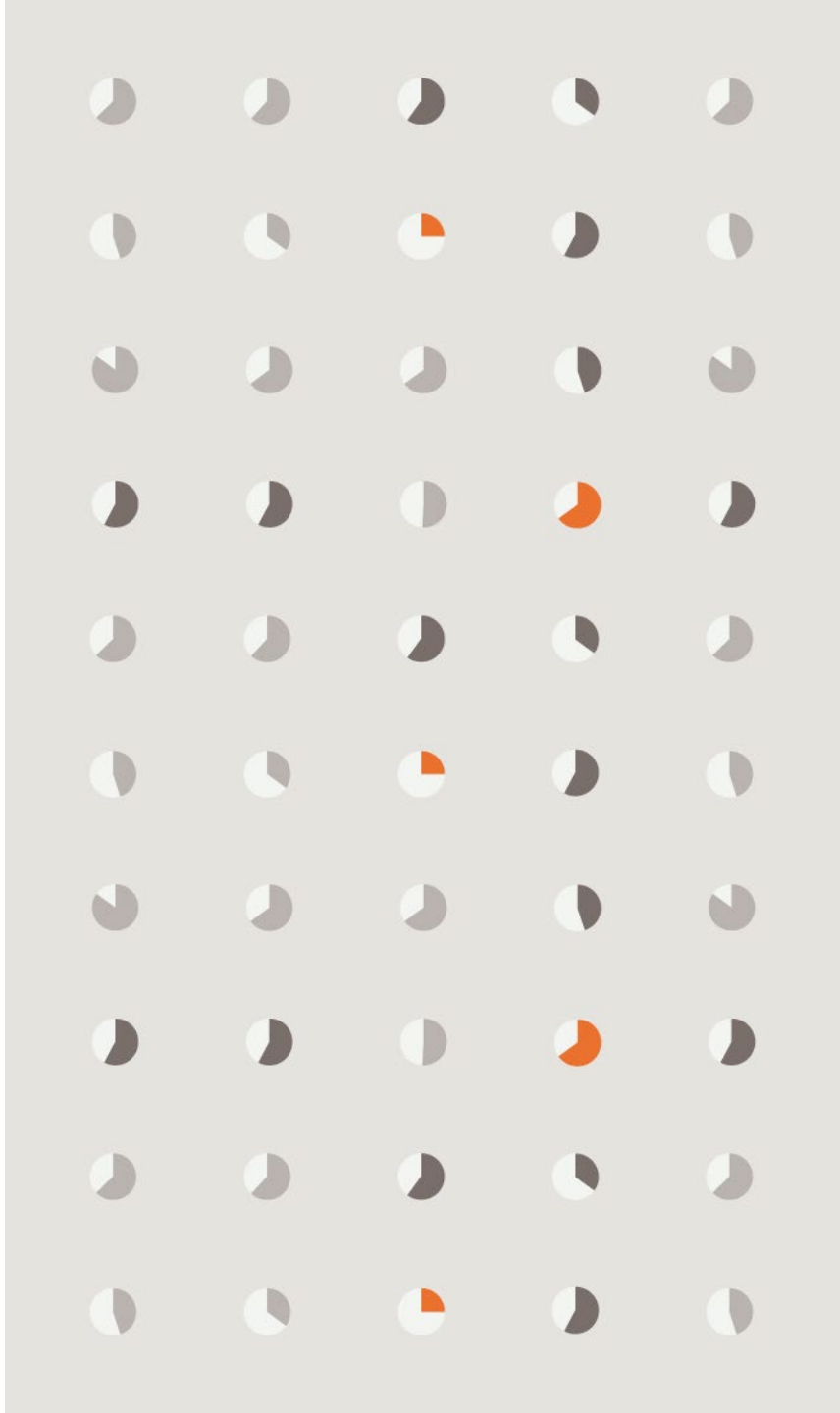
NORC at the University of Chicago

Agenda

01 Conceptualization

02 Measuring Micro Equity and Diversity

03 Discussion



Center on Equity Research

The Center is a NORC research center and recognized leader in advancement of high-quality, inclusive and equitable social science research, approaches and methods.

At NORC, CER serves as a central leader for developing, promoting and integrating inclusive and equitable research (IER) best practices into research and evaluation.

We partner with you to amplify use of **inclusive and equitable research (IER)** across the full research cycle.

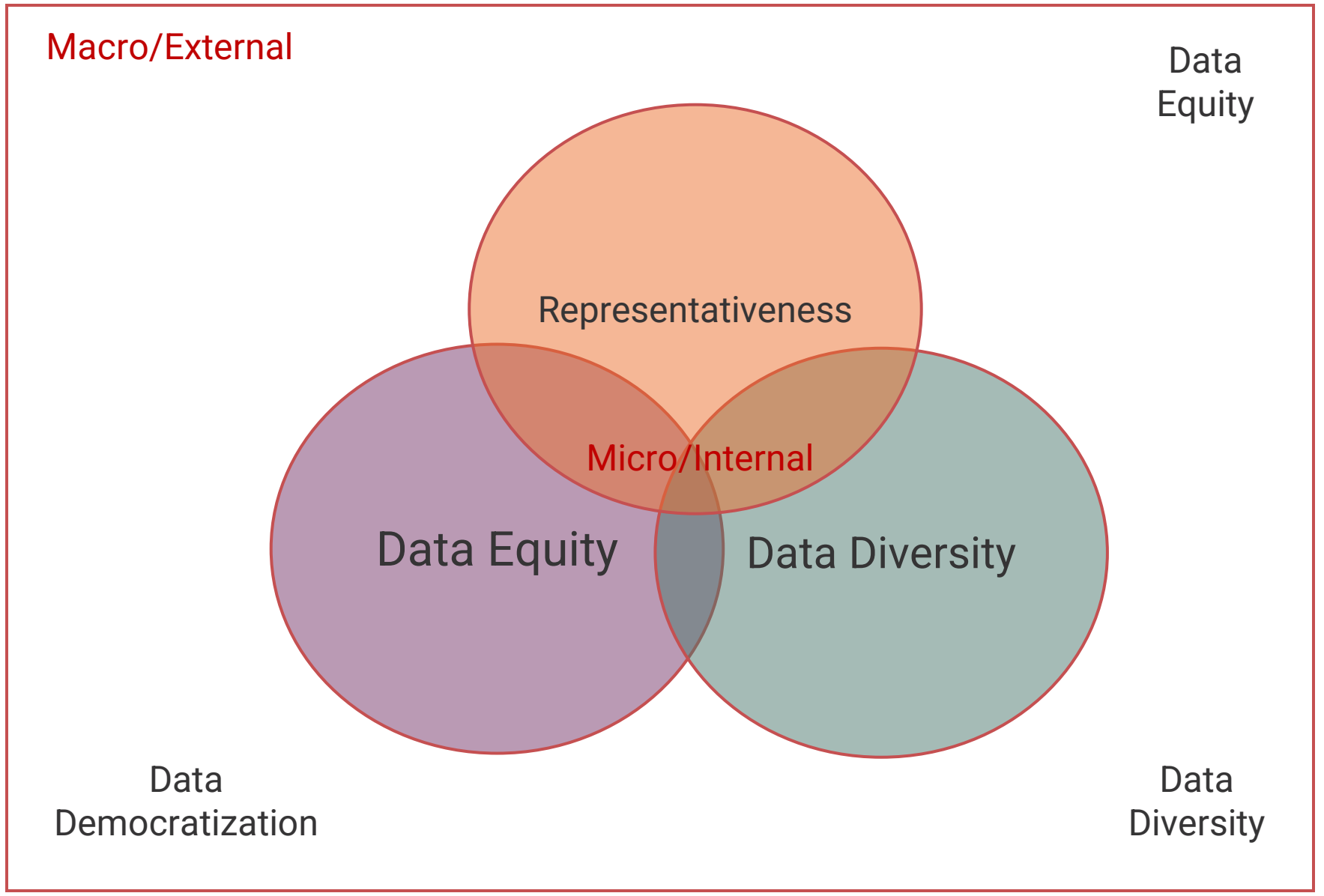
Our mission is to advance equity through the development, implementation, dissemination and education of high quality inclusive and equitable social science research.

NORC

Center for Panel Survey Sciences

The Center is dedicated to researching and developing best practices in probability-based panel surveys.

Learn more at **http://cpss.norc.org/**



There is a difference between macro considerations of diversity and equity and thinking about diversity and equity within a given dataset (the micro).

Representation:

- In Datasets: The degree to which a dataset proportionally represents a population as compared to benchmark population estimates, across multiple demographic or other metrics.

Data Democratization:

- Overall: Making data available, accessible, and understandable to involved communities which can motivate and empower individuals and communities.

Data Equity:

- Overall: A mindset or worldview about the use of data to promote human health and wellbeing, and contribute to fair, impartial and just results that promote policy and program change. Every step of a data project encompasses a lens of justice, equity, and inclusivity
- In Datasets: The ability to provide equitable statistical projections for populations that matter in the research being executed.

Data Diversity:

- Overall: Having research that reaches, speaks to, and informs a multitude of audiences and communities. The inclusion of varied and heterogeneous data sources (e.g., people) and perspectives in a data set and across many data sets.
- In Datasets: Data that have some level of representation across micro-communities/cultures/populations. Data whereby there is variance of thought within small micro-communities/cultures/populations

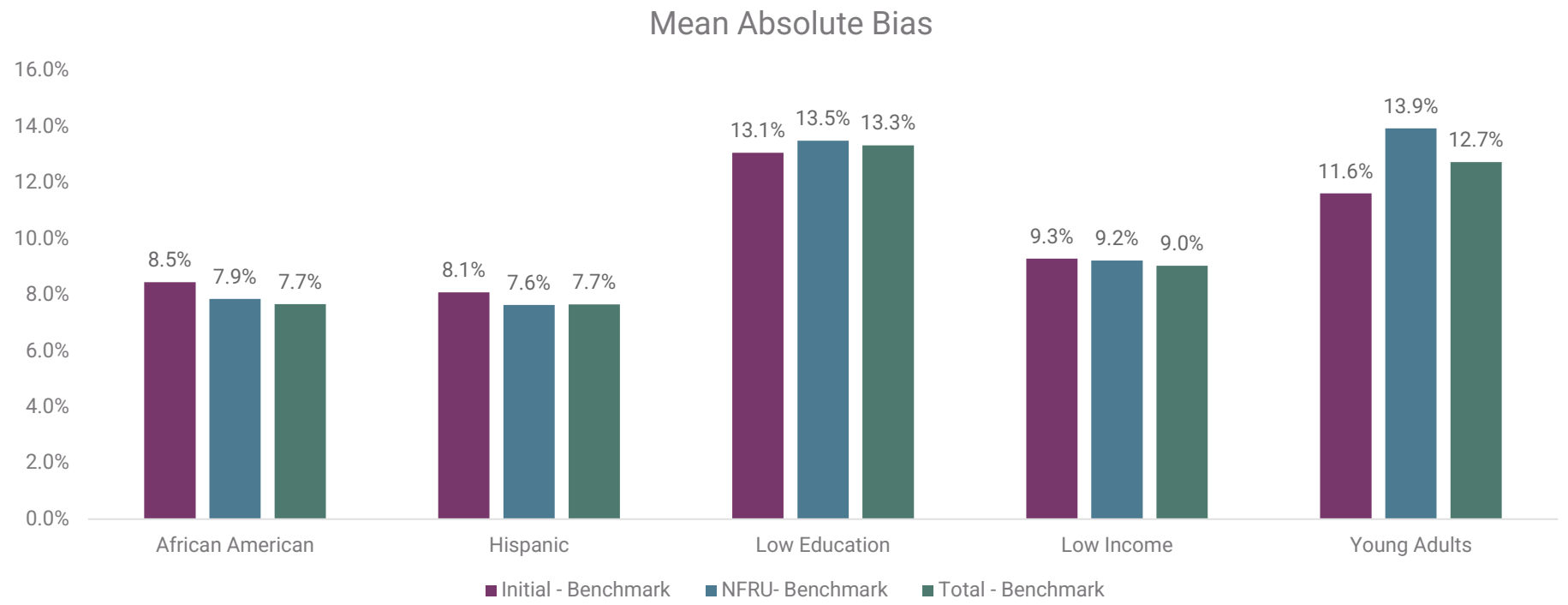
Thoughts on Measuring Micro Representativeness, Data Diversity, and Data Equity

NORC at the University of Chicago

Measuring Representativeness

The Easy One! We have been doing it for years:

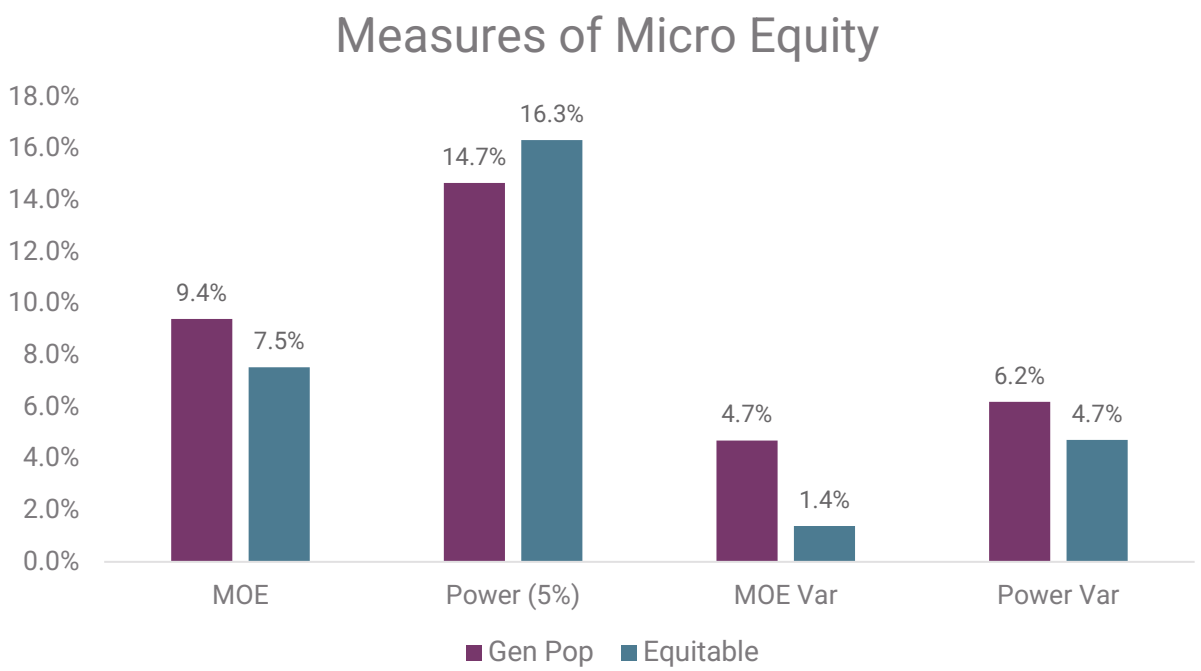
- Mean absolute bias/Mean absolute relative bias
- Unequal weighting effects
- Simple benchmark comparisons



Measuring Micro Data Equity

Stating which populations you want a voice for, and measuring the ability to represent those voices

- Power analyses of each population
- Some summative score of those power calculations that include central tendency and range



Group	Gen Pop	Equitable
NH-AAPI	80	200
Low income (FPL100)	411	424
Low education	507	531
African American	244	252
Hispanic	354	325
HH with children under 18	316	331
Spanish speaking	51	200

Measuring Micro Data Diversity

- How homo/heterogenous is your data?
  - To what degree does the data have micro heterogeneity?
  - Example: number of “unique” cells of possible uniqueness over total sample
  - Gini Simpson measures of diversity
  - Other metrics?

	Non NRFU	NRFU	Full sample
Cells N> 0	6,591	6,053	8,880
Possible Cells	28,800	28,800	28,800
Overall N	21,697	15,782	37,479
% cells>0 of N	75%	55%	130%
% of possible	23%	21%	31%
% of N	30%	38%	24%

Variable	NRFU=0	NRFU=1	Overall
Average across variables	0.43	0.60	0.48
Diversity	0.85	0.99	0.87
Education	0.51	0.73	0.57
Employment	0.27	0.48	0.31
Gender	0.49	0.49	0.49
HomeType	0.32	0.46	0.36
Income	0.50	0.67	0.56
Marital Status	0.50	0.62	0.56
Party ID	0.31	0.78	0.38
Race/Ethnicity	0.65	0.71	0.68
Housing Tenure	0.33	0.48	0.36

Final Thoughts

NORC at the University of Chicago

Pathways to Better Data

- The first step: Recognition that high quality data is not just representative;
  - It is diverse (has a high relative degree to micro data diversity).
  - It is equitable (has the power to analyze subgroups who should have a voice.
- Applicable potential metrics for diversity:
  - Degree to which there is data heterogeneity at the micro culture/community level with Gini Simpson, or some other metric; notably will differ from study to study on what type of diversity is important.
- Applicable potential metrics for equity:
  - Grand central tendencies and dispersions of subgroup power analyses

Examples of Increasing Micro Data Equity and Diversity

- **Methods to Improve Data Diversity and Equity**
  - Diversity: More research needed, but possible pathways: Multiple modes; multiple frames; NRFUs; high response rates
  - Equity: Define important groups under study before study design; support data collection strategies that support attaining sufficient power in those groups; measure power

- **Example Strategies from AmeriSpeak**
  - Nonresponse follow-up
  - Amplify AAPI
  - Disproportionate Stratified Sampling
  - Equitable Materials
  - Multi-Modality
  - Input on Collaterals/Reports by a Diverse Team
  - Use of a Diverse Advisory Group
  - Messaging Optimized for Different Groups

David Dutwin
Dutwin-David@norc.org
312-759-4027

Thank you.

https://amerispeak.norc.org/

Research You Can Trust™

NORC at the University of Chicago