Total Survey Error as a Framework for Data Equity

Sarah Forrest, MPH

Katherine Irimata, PhD

Morgan Earp, PhD

Lauren Rossen, PhD

Kristen Miller, PhD

2024 FCSM Research and Policy Conference

Background

- The National Center for Health Statistics (NCHS) is the nation's principal health statistics agency, providing data and disseminating statistical information to identify and address health issues, as well as guide public health and policies that promote health equity

- NCHS develops methods to assess and improve estimates for all populations

- NCHS examines health equity from the total survey error perspective, to identify potential inequities in measurement, representation, and dissemination

- This presentation highlights the total survey error framework through an equity lens, describing enhancements that prioritize inclusivity and representation to produce more accurate, credible, and relevant statistics

Data Quality and Data Equity
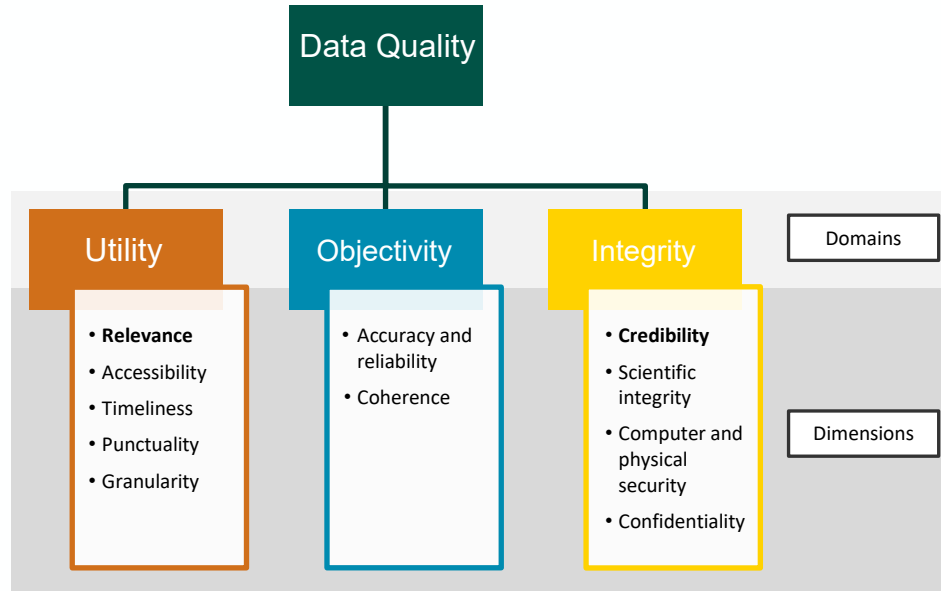
Framework for Data Quality
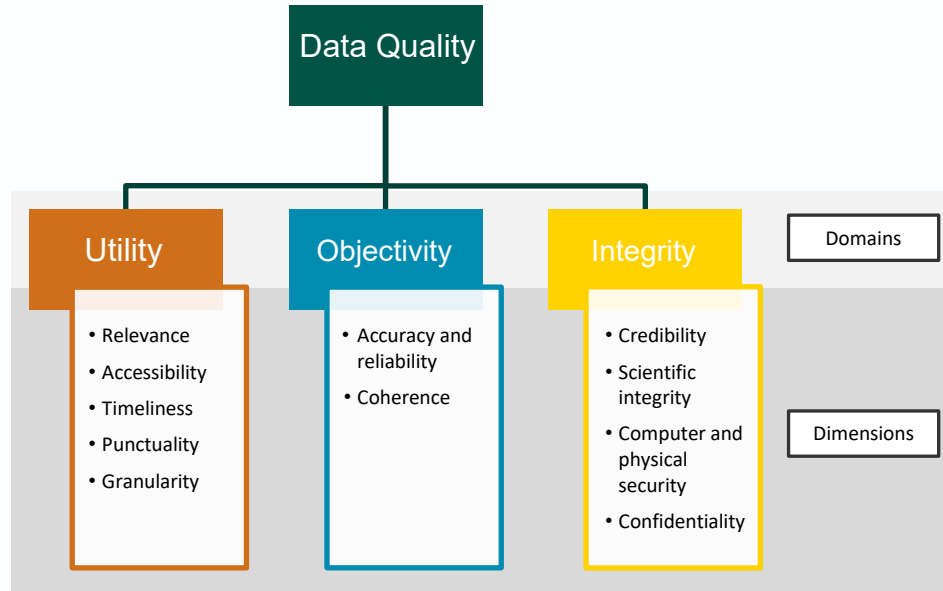
Federal Committee on Statistical Methodology (2020). *A Framework for Data Quality.*

Framework for Data Quality: Relevance & Credibility

Relevance

Whether the data product is targeted to meet user needs

Data Quality

Utility	Objectivity	Integrity	Domains
• **Relevance** • Accessibility • Timeliness • Punctuality • Granularity	• Accuracy and reliability • Coherence	• **Credibility** • Scientific integrity • Computer and physical security • Confidentiality	Dimensions

Credibility

The confidence that users place in data products based simply on the qualifications and past performance of the data producer

Federal Committee on Statistical Methodology (2020). *A Framework for Data Quality.*

Addressing Data Inequities for Health Equity

- Data equity aims to make all populations, especially historically underrepresented or misrepresented ones, visible and accurately portrayed in data records (NASEM, 2023)

- Data inequities occur when data collection, analysis, or representation disproportionately favors certain groups or excludes others

- Data inequities can hinder our ability to accurately understand and address health inequities because they distort the representation of various groups, leading to biased conclusions that perpetuate existing disparities

- Inequities in the survey design, measurement, representation, and dissemination of data can potentially be compounded, resulting in statistics that may be less accurate, credible, relevant, or available for some groups

National Academies of Sciences, Engineering, and Medicine (NASEM). Toward a 21st Century National Data Infrastructure: Enhancing Survey Programs by Using Multiple Data Sources. 2023.

Examining Health Equity from a Total Survey Error Perspective

Health Equity from a Total Survey Error Perspective

Measurement

- Bias can arise from differential construct validity or differential measurement error

- Bias or differential measurement error can distort (exacerbate or conceal) health inequities
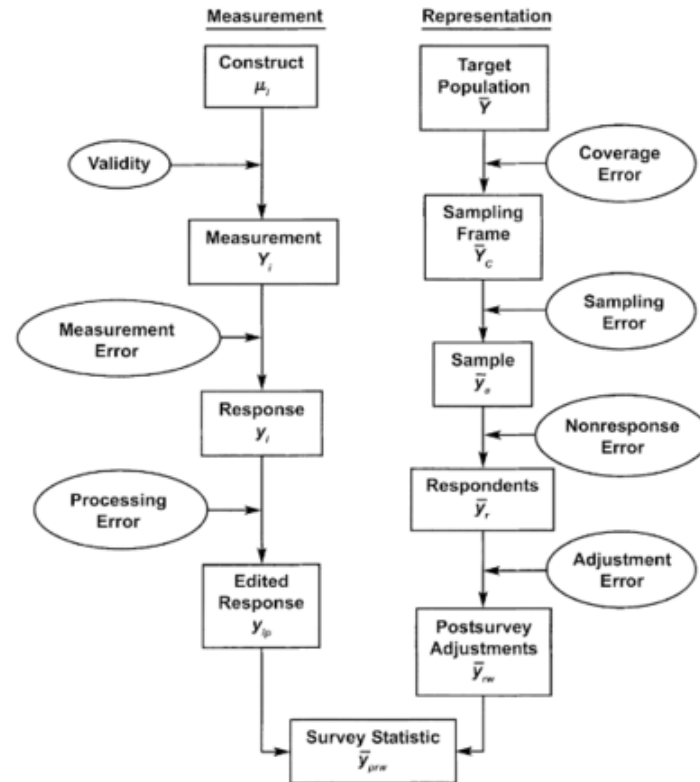
Representation

- Biased measures can arise from differential coverage error, differential sampling error, or differential nonresponse error

- Biased estimates or estimates subject to differential representation error can distort (exacerbate or conceal) health inequities
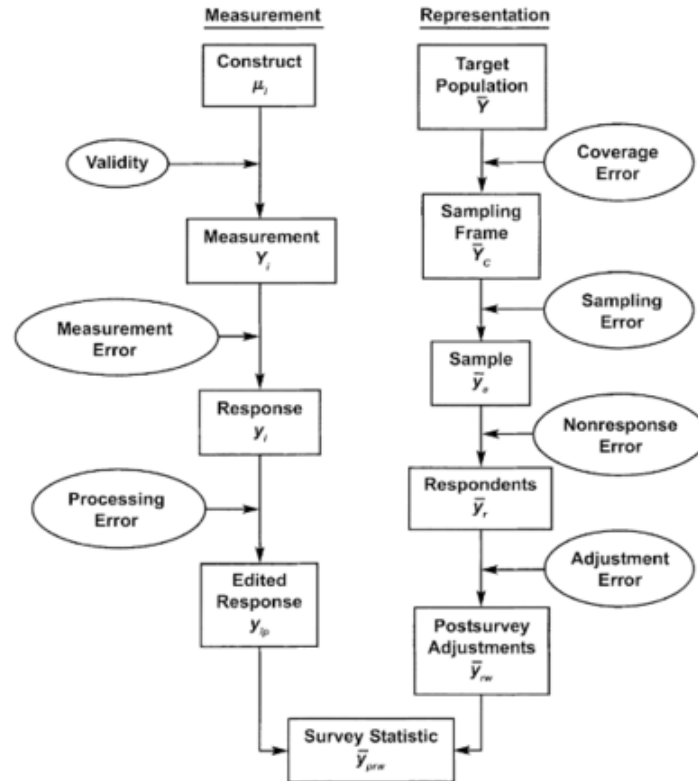
Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public opinion quarterly*, *74*(5), 849-879.

Health Equity from a Total Survey Error Perspective

- **Construct Validity**
  - Cognitive interviews
  - Comparative studies
  - Item Response Theory
- **Measurement Error**
  - Cognitive interviews
  - Linear Regression Trees

Measurement	Representation
Construct μ_j	Target Population $\bar{Y}$
Validity	Coverage Error
Measurement Y_i	Sampling Frame $\bar{Y}_c$
Measurement Error	Sampling Error
Response y_i	Sample $\bar{y}_s$
Processing Error	Nonresponse Error
Edited Response y_{ip}	Respondents $\bar{y}_r$
	Adjustment Error
	Postsurvey Adjustments $\bar{y}_{rw}$
Survey Statistic $\bar{y}_{prw}$	

- **Coverage Error**
  - Combining data sources
- **Sampling Error**
  - Sampling methodology
  - Sample weights
- **Nonresponse Error**
  - Imputation
  - Weight adjustments

Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public opinion quarterly*, 74(5), 849-879.

Facets of Data Equity

Procedural & Decision-Making Equity
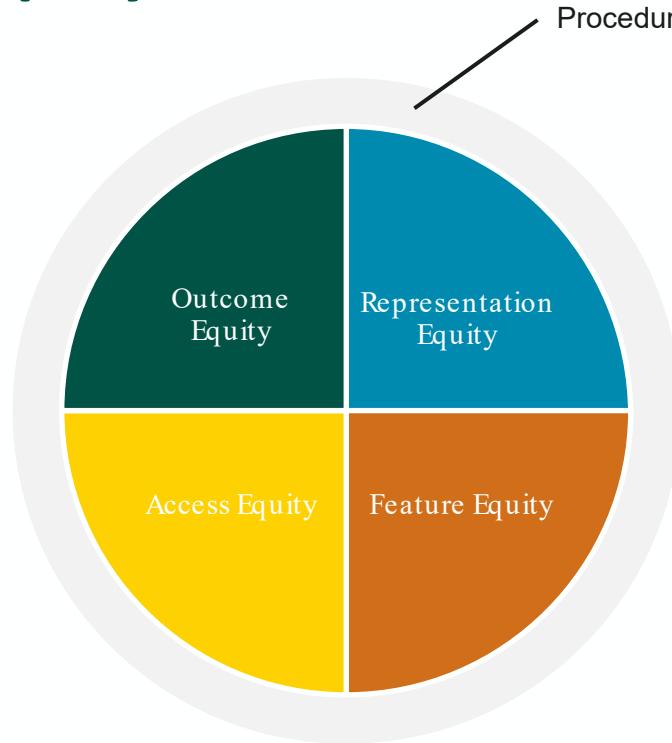
Outcome Equity

Benefits or impacts derived from data use are distributed justly across all groups

Access Equity

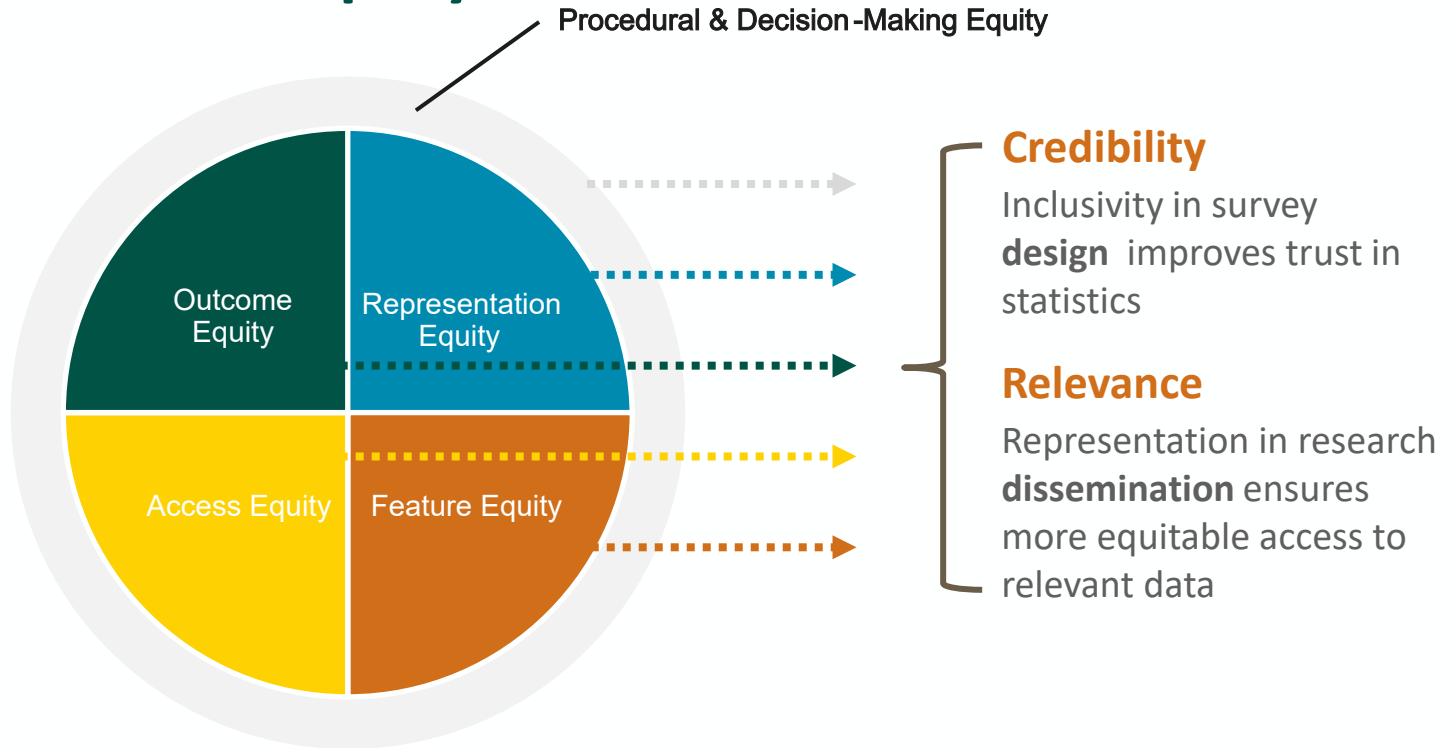Equal opportunities for all groups to obtain and use data

Representation Equity

All groups are adequately included and accurately represented; data reflects the diversity of the populations

Feature Equity

Availability and appropriate selection of variables needed to represent members of every group in analyses, particularly those to study inequity

Jagadish, H., Stoyanovich, J., & Howe, B. (2022). The many facets of data equity. *ACM Journal of Data and Information Quality*, 14(4), 1-21.

Facets of Data Equity

Procedural & Decision-Making Equity

Outcome Equity

Representation Equity

Access Equity

Feature Equity

Credibility

Inclusivity in survey **design** improves trust in statistics

Relevance

Representation in research **dissemination** ensures more equitable access to relevant data

Jagadish, H., Stoyanovich, J., & Howe, B. (2022). The many facets of data equity. *ACM Journal of Data and Information Quality*, 14(4), 1-21.
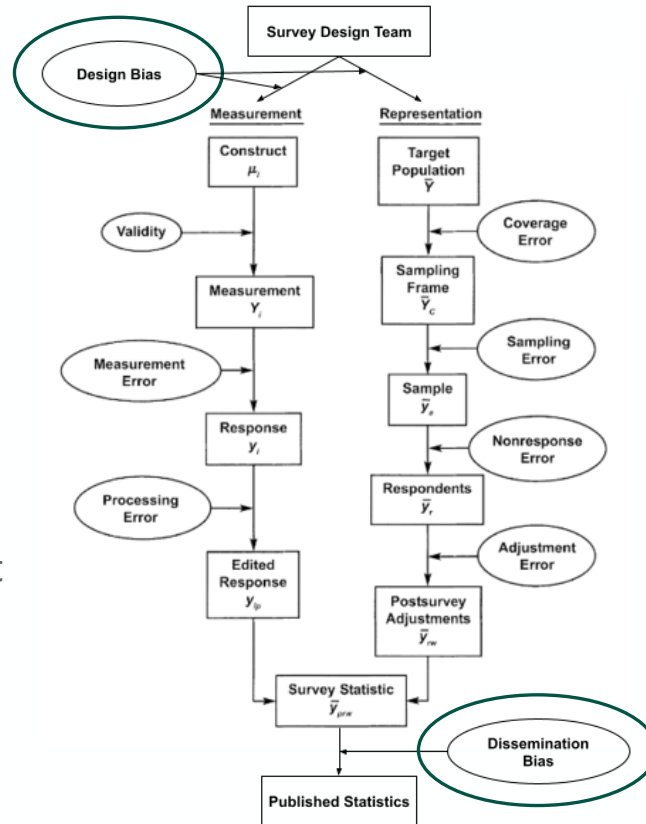
Equity-Driven Enhancements to the TSE Framework

Design Bias

Inherent biases from the individuals or groups who have traditionally been responsible for making decisions about:

- What to measure
- Which groups to include
- How to locate and contact them

Dissemination Bias

Unequal distribution and accessibility of data for different demographic and geographic groups, potentially:

- Underrepresenting or omitting certain groups
- Providing overall estimates and not subpopulation estimates

Strategies to Mitigate Inequities

Overview of Strategies to Mitigate Inequities

- Examine and quantify inequities at each stage
  - Consider the cumulative or interactive effect of multiple biases on the final statistic
- Treat survey design and data collection under an 'equity' framework rather than an 'equality' framework
  - Adaptive and responsive survey designs
  - Some subgroups are harder to reach, require different strategies
- Mixed-methods research
- Community-based research
- Model-based estimation or small domain estimation
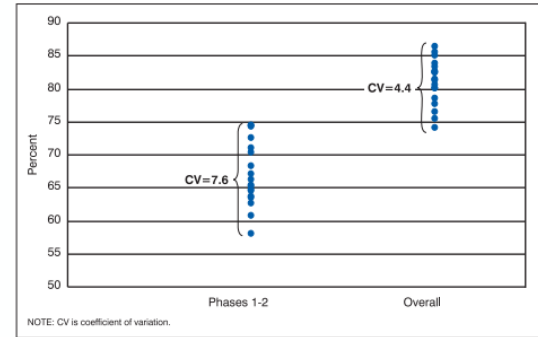 - To mitigate errors, biases, or otherwise improve available data

Design Equity: Intentional Inclusion

- Certain groups are labeled "hard-to-survey" if they require more time/ effort/resources to include than the general population respondent
- Groups omitted in surveys at NCHS:
  - Institutionalized populations
  - Unhoused or transitory populations (address-based sampling)
  - Remote area populations
- "Hard-to-survey" or hard to estimate for *whom*?
  - Based on the perspective of researchers about a population they don't belong to rather than inherent challenges or characteristics with the population
- Deciding who to count (and who not to) $\longrightarrow$ deciding where to direct resources for producing accurate counts and for whom

Leonard, Mariel. Reconsidering Hard-to-Survey Populations, Representativity, & Representation in Surveys. AAPOR Webinar. 2023.

Adaptive Design: National Survey of Family Growth

- Multiple mode
  - In person interviewing, with sensitive questions answered privately by self-administration through computer-assisted interviews
  - Enhanced **measurement equity**
- Multiple phase: NSFG Cycle 6
  - Final phase focused on engaging non-respondents by adjusting the data collection procedures
  - Enhanced **representation equity**

Groves, Robert et al. (2005). "Plan and operation of Cycle 6 of the National Survey of Family Growth." *Vital and health statistics. Ser. 1, Programs and collection procedures*: 1-86.

NSFG response rates by age, sex, and race subpopulations


NOTE: CV is coefficient of variation.

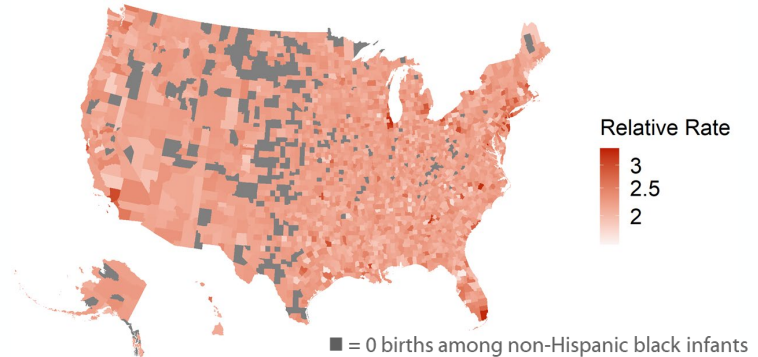NSFG response rates by phase and age group
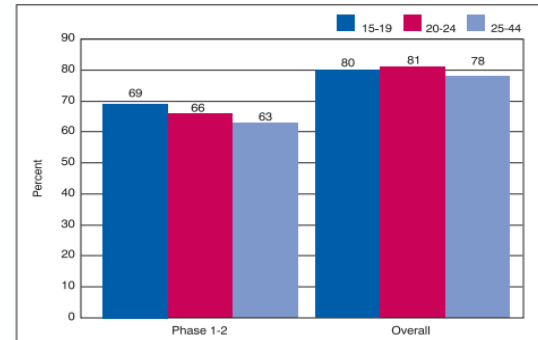
Dissemination Equity: Model-Based Estimation

- Estimates for small groups are often suppressed due to concerns about reliability/domains

- Model based estimation can improve and generate estimates for small subgroups by 'borrowing strength' over time or across groups

Relative Rate
3
2.5
2

■ = 0 births among non-Hispanic black infants
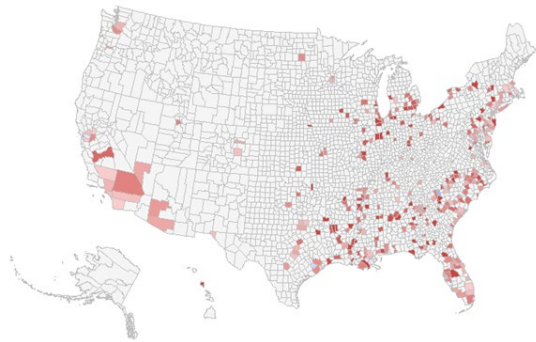
Only ~5% of counties have sufficient data to calculate infant mortality rates for Black-White groups (n≥10)

Model-based estimates of relative disparities in infant mortality rates for both groups

Hirai AH, Rossen LM, Kramer MR, James AR, Warren MD. Black-White Infant Health Equity by State and County: An Evaluation of Achieving Healthy People 2020 Goals for Preterm Birth and Infant Mortality. Under review.

Measurement Equity: Mixed-Methods Research

- Mixed-method studies to assess comparability across respondent groups

  - **Qualitative**: Interpretive cognitive interviewing studies—to identify constructs captured by the question

  - **Quantitative**: Embedded construct and error probes on the Research and Development Survey (RANDS)—to determine the extent to which constructs vary across groups

Construct Schema:

Question: In general, would you say your health is excellent, very good, good, fair or poor?

```
Phenomena
Considered by ──┬── Health Conditions ──┬── Pain or Discomfort
Respondents     │                       │
                │                       ├── Physical Limitations
                │                       │
                │                       └── Dependence on Medication
                │
                └── Health Habits ──────┬── Healthy: Regular Exercise, Proper Nutrition
                                        │
                                        └── Unhealthy: Smoking Overeating, Alcohol Consumption
```
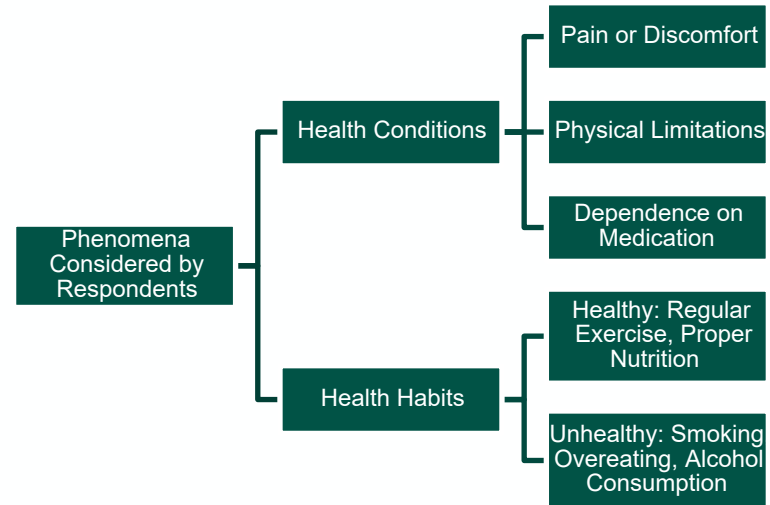
RANDS Construct Probe:

When answering the previous question, did you think of 1) your health conditions, 2) your health habits or 3) something else?

Contact Us

Sarah Forrest

sforrest@cdc.gov

For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348   www.cdc.gov

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