Westat

Improving Lives Through Research®

Revisiting Total Survey Error Framework in a Multimode and Multidata Environment

Ting Yan, Wendy Hicks

FedCASIC 2024

Multimode Surveys

- Single mode paradigm no longer applies in the 21st century (de Leeuw & Berzeklak, 2016)
- Different types of multimode designs (Dillman et al., 2014)

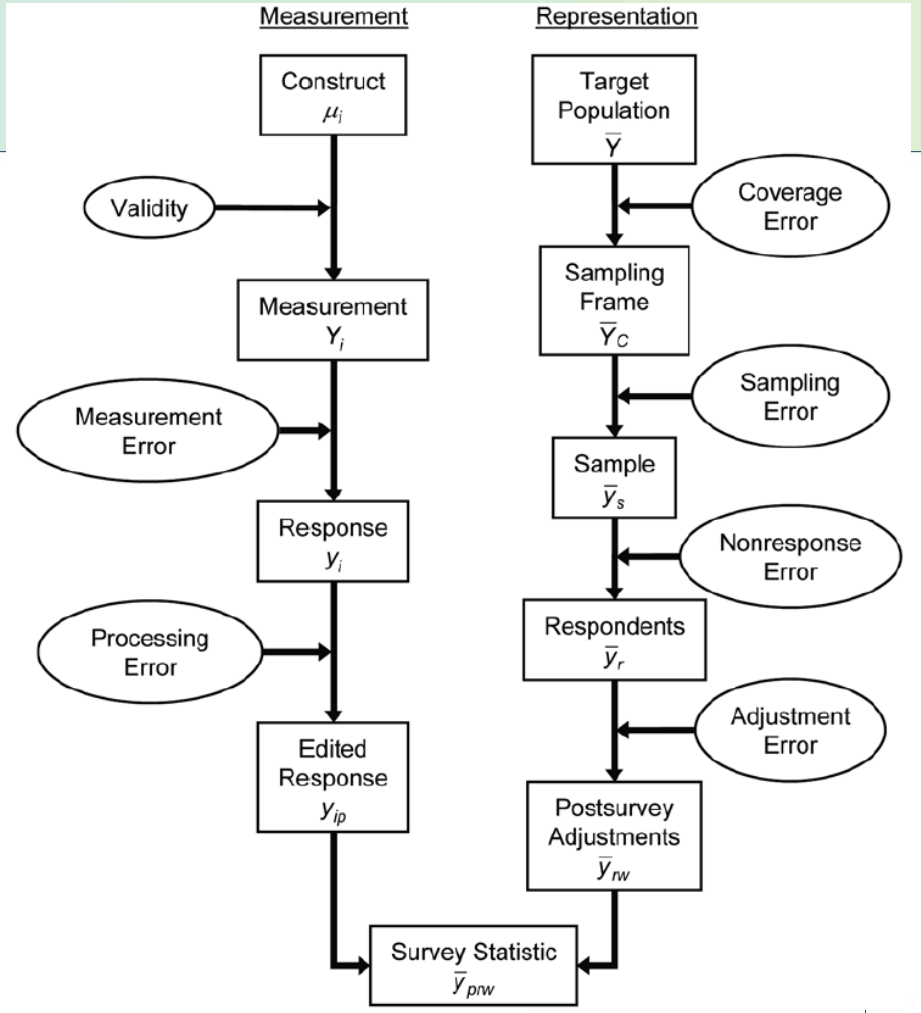
	1 contact mode		>1 contact mode	
	1 response mode	>1 response mode	1 response mode	>1 response mode
Direct data collection	Mail-only	Web+mail through mailing	Web through mailings + text	Web+mail through mailing + text

Multimode and Multidata Environment (2)

- Data from different sources (besides survey data) can be used and combined to access people and obtain information from people
  - administrative data, passive data, biodata, big data
- Imagine a multimode and multidata study
  - Combines an address-based sample and a list of WIC participants sampled from WIC administrative records
  - Recruits participants by mailings and in-person visits
  - Interviews them by web and ACASI
  - Tracks their location and travel through GPS
  - Collects their physical activity through wearable
  - Links to their medical records

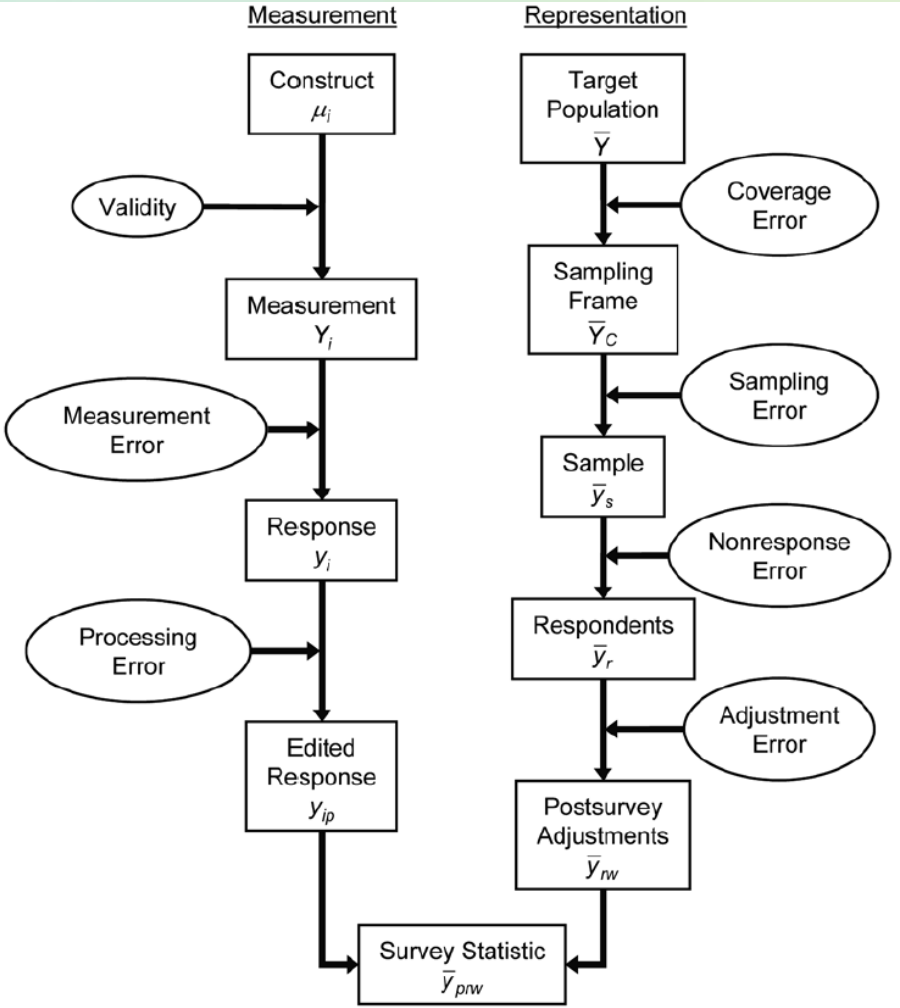
Total Survey Error (TSE)

- Goal is to reduce error given cost constraints, or reduce costs for given level of quality/error
- Survey error=deviation of what is desired from what is attained
  - Error of non-observation
  - Error of observation



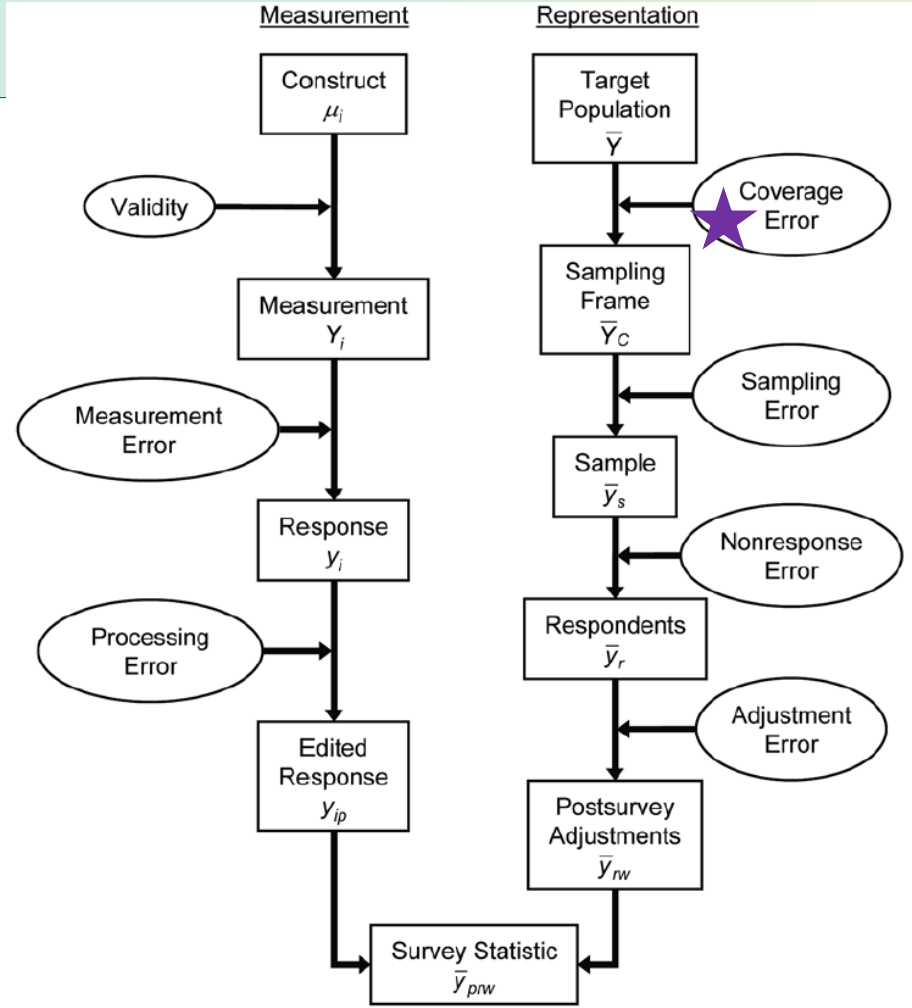
Total Data Error (TDE) for Multimode and Multidata Environment

- In multimode and multidata environment
  - Shift from TSE to TDE
  - Goal is still to reduce error given cost constraints, or reduce costs for given level of quality/error
- Multimode and multidata designs have the potential to
  - Improve quality
  - Reduce error
  - Reduce cost



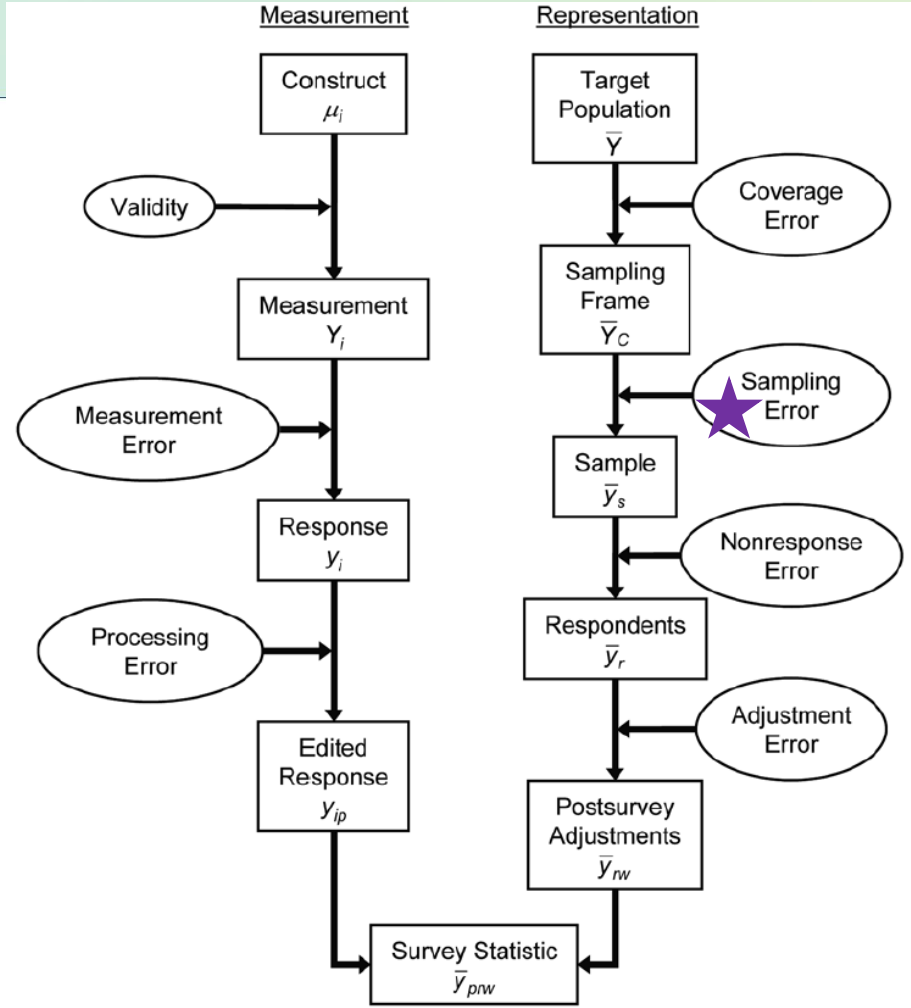
TDE: Coverage Error

- Non-observational gap between target population and sampling frame
- Combining multiple frames/data sources can improve coverage or efficiency
  - cellphone RDD + landline RDD
  - nonprobability web panel + address-based sample frame
  - address-based sample frame + admin records
- Cost and timeline of getting another frame/data



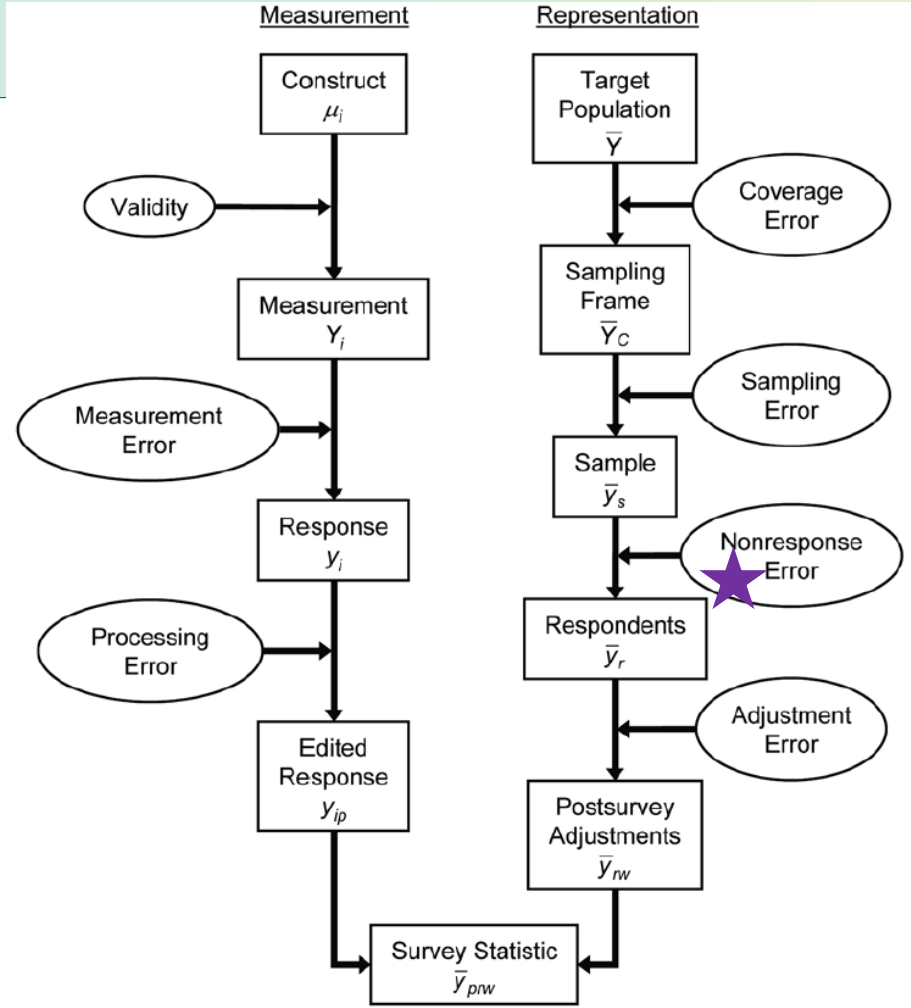
TDE: Sampling Error

- Non-observational gap between sample and sampling frame
  - Function of four principles
    - Probability sampling, stratification, clustering, sample size
- Multidata environment
  - Probability sampling is still feasible or is still used?
  - Blended/hybrid sampling
  - Big ≠ absence of need for probability sampling



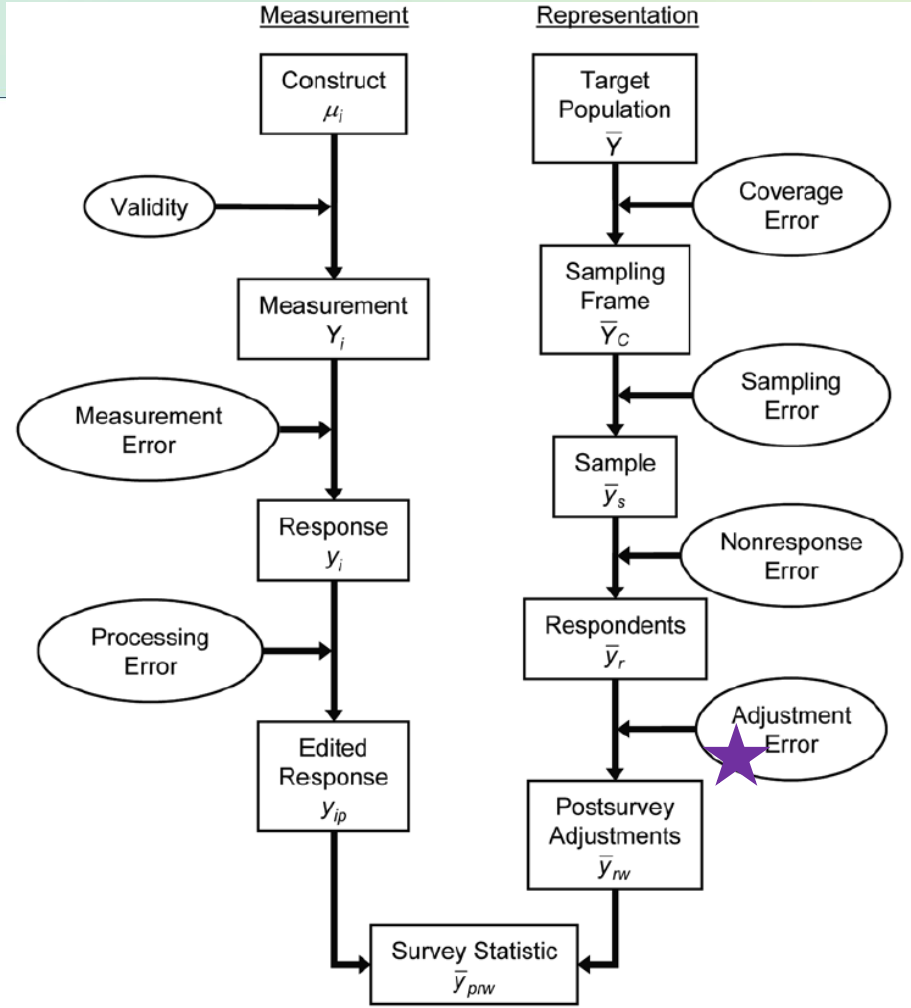
TDE: Nonresponse Error

- Non-observational gap between respondents and sample
  - Noncontact
  - Refusal
- Combining multiple contact/response modes can reduce noncontact and refusal
- Combing multiple data sources can reduce/bypass nonresponse
  - Nonconsent



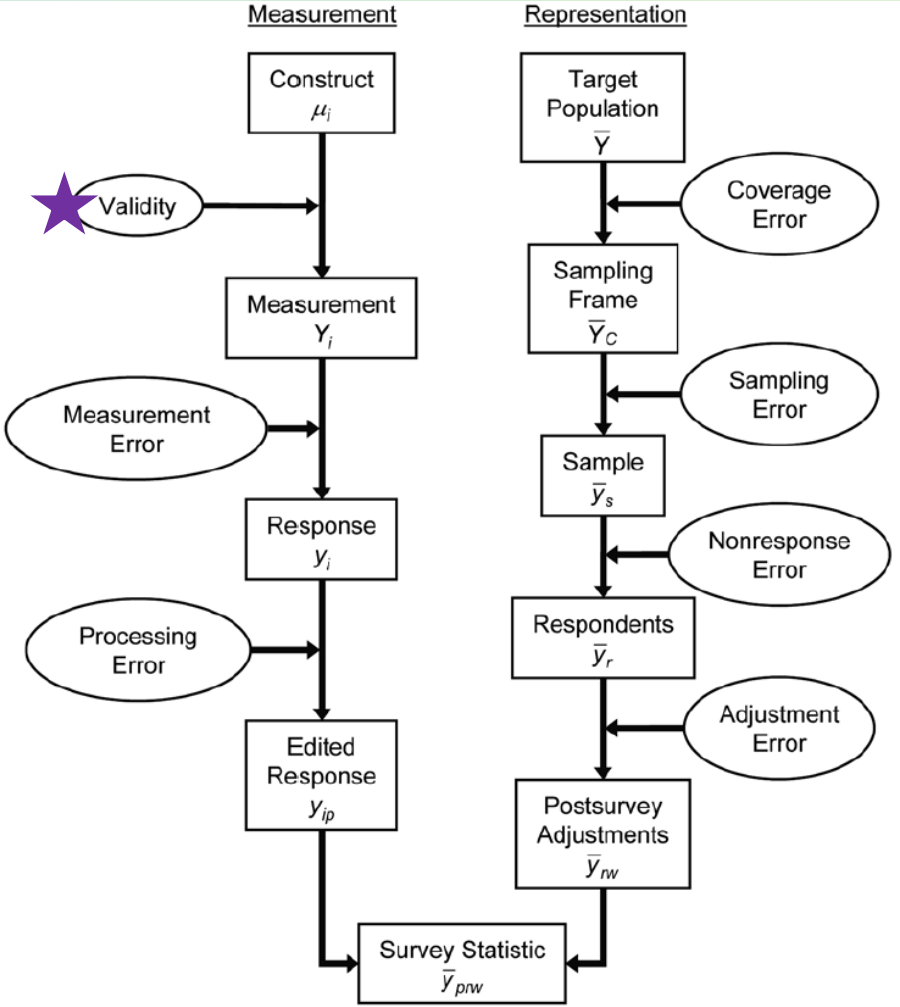
TDE: Adjustment Error

- Non-observational gap between adjusted statistics and population parameters
  - Weighting to increase coverage, to compensate for differential sampling probabilities, and to account for nonresponse
- Multimode environment
  - Adaptive mode adjustment (Brick et al., 2022)
- Multidata environment
  - Open area:
    - What? How?



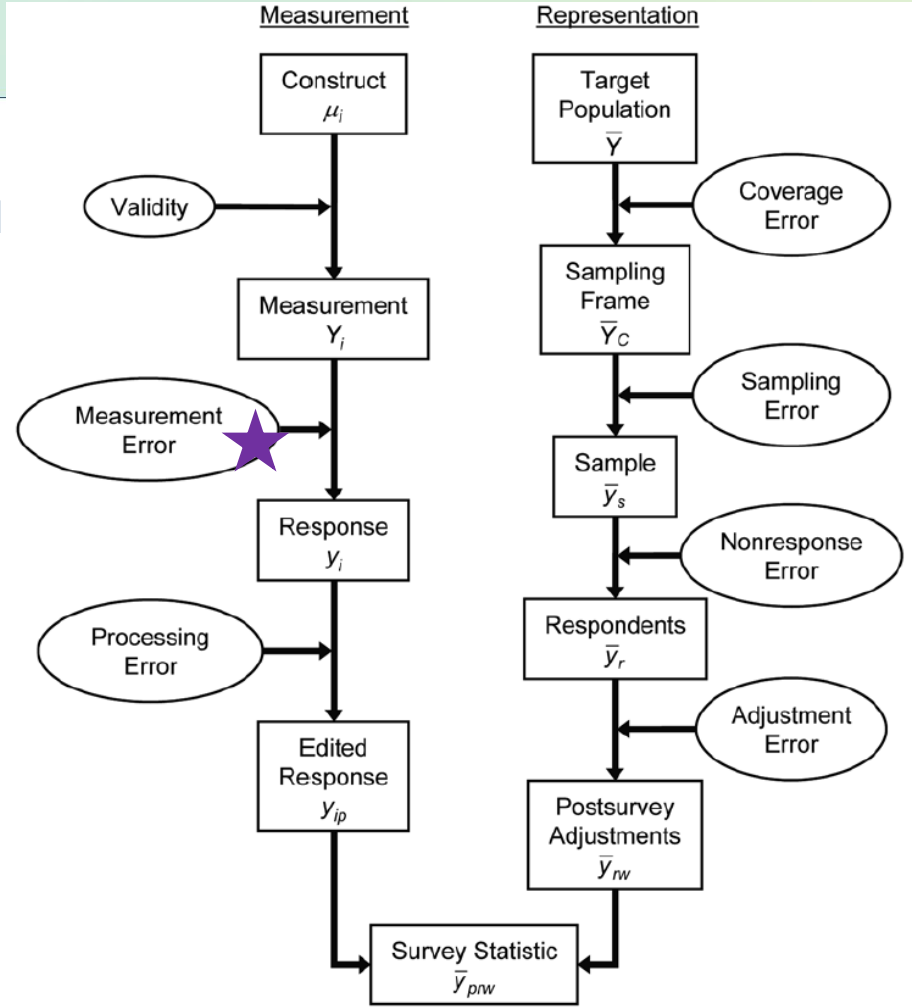
TDE: Validity

- Observational gap between construct and measurement
- Multimode environment
  - Unimode approach (Dillman et al., 2014; 2016) vs. best practices approach (Couper, 2008)
- Multidata environment
  - Are data elements in different data sources measuring the same construct that we want to measure?
    - "Confidence"
  - Unified data element across data desired



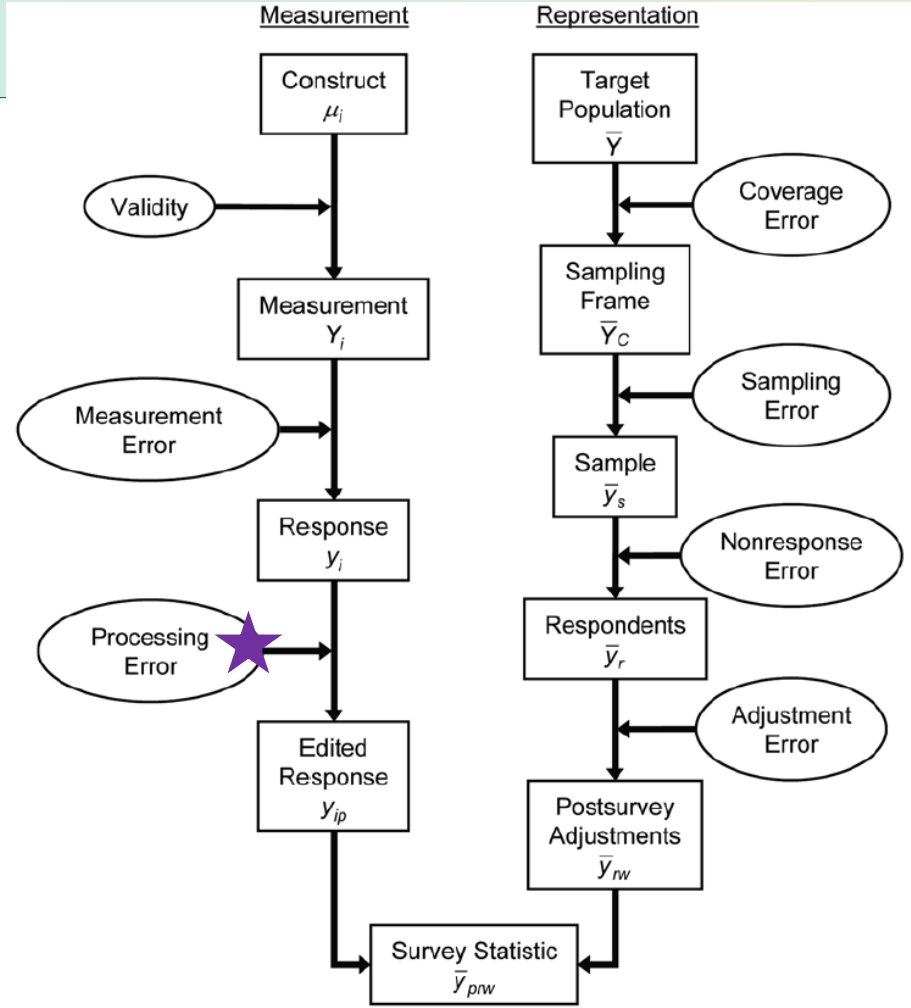
TDE: Measurement Error

- Observational gap between true measurement and measurement obtained
- Multimode environment
  - Differential mode measurement effect
  - Using self-administration for sensitive questions and interviewer-administration for other questions
- Multidata environment
  - Differential levels of quality



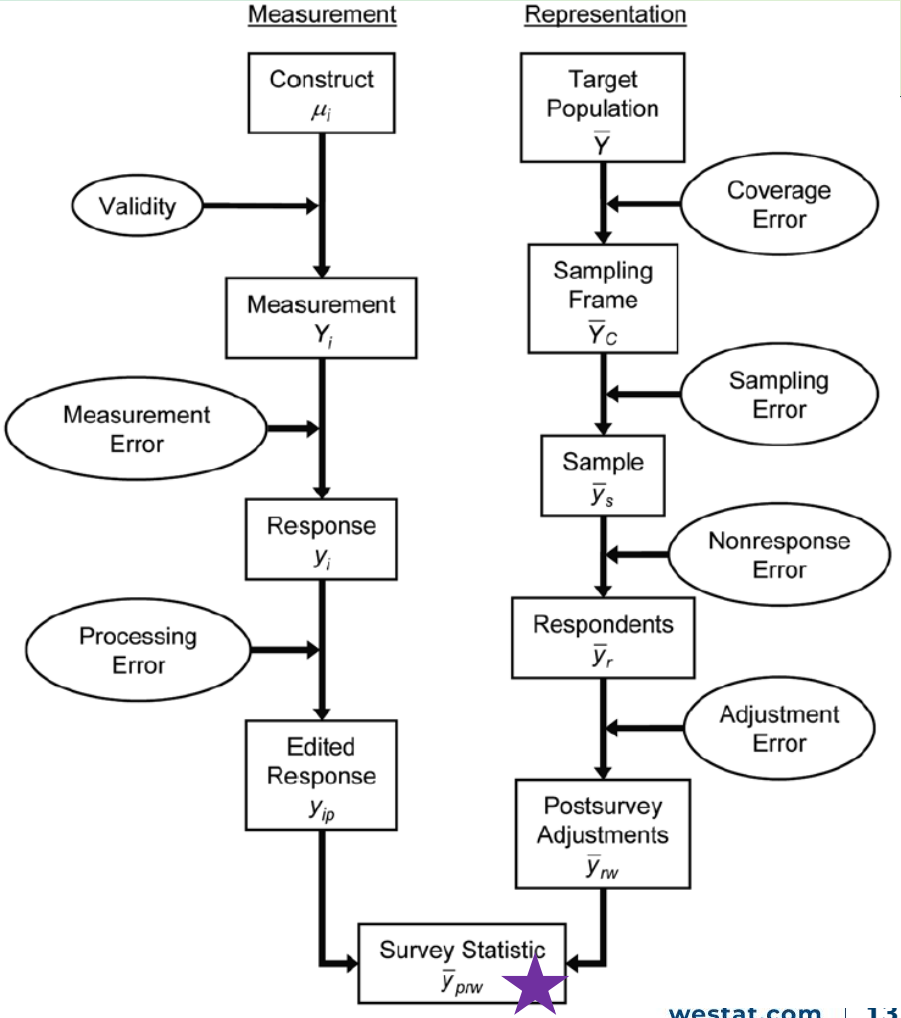
TDE: Processing Error

- Observational gap between variable used in estimation and measurement obtained
  - Coding, editing, imputation
- Multimode environment
  - Unified decisions/rules desired
  - Harmonization takes time and increases cost
- Multidata environment
  - Unified decisions/rules desired, but tricky when data from one source is structured, another source is half structure, and the third source is unstructured
  - Machine learning ≠ magic wand ≠ low cost



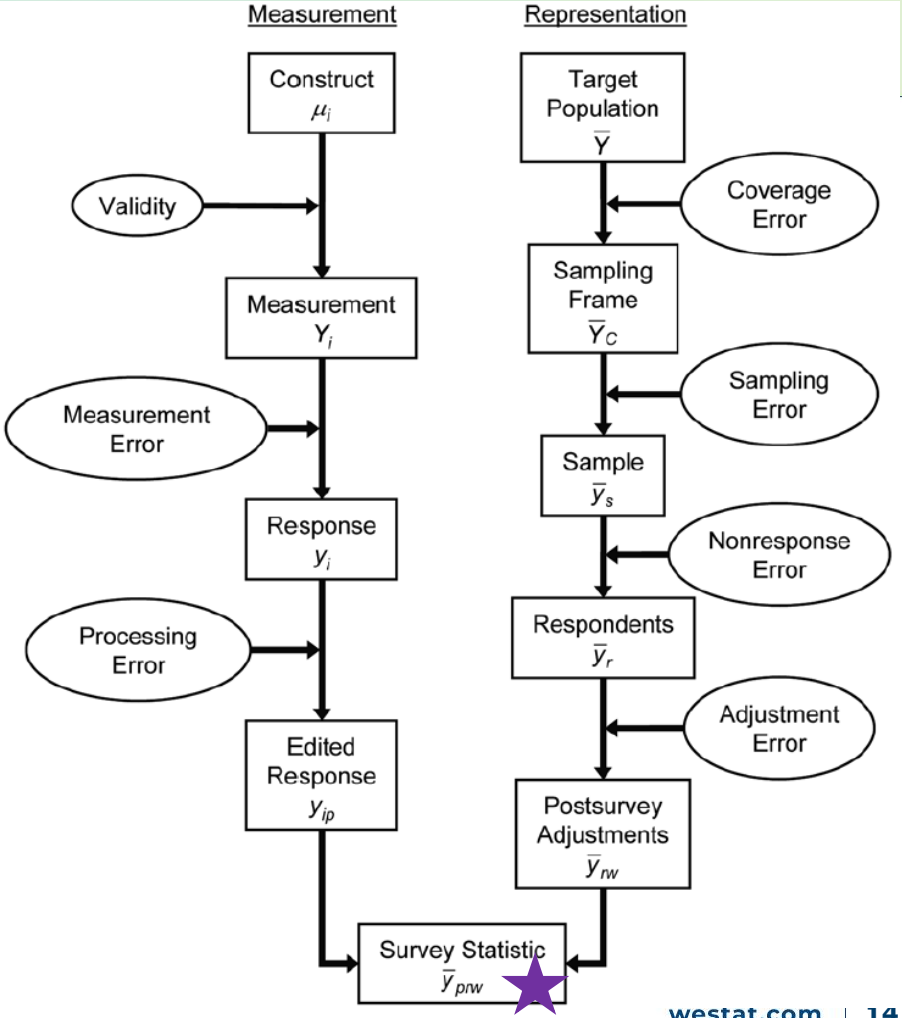
Putting It All Together

- Effort to reduce one error could potentially increase another error
  - Offering a paper mode reduces nonresponse error at the risk of increasing measurement error and processing error
  - Including data from records reduces measurement error at the risk of increased coverage error and/or nonresponse error
  - Including data from records reduces coverage and/or nonresponse error at the risk of increased measurement error
- Additional cost and time, for obtaining, linking, cleaning, manipulating, editing, coding, harmonizing



Putting It All Together (2)

- Goal is to reduce total data error or to reduce cost when combining multiple modes and multiple data
- Use this framework to inform decisions on balancing errors and making trade-offs
  - Requires knowledge of different mode/data source on each source of error
  - Requires systematic examination of error trade-offs
- Call for research to answer and inform the above two requirements



Thank you

tingyan@westat.com

westat.com