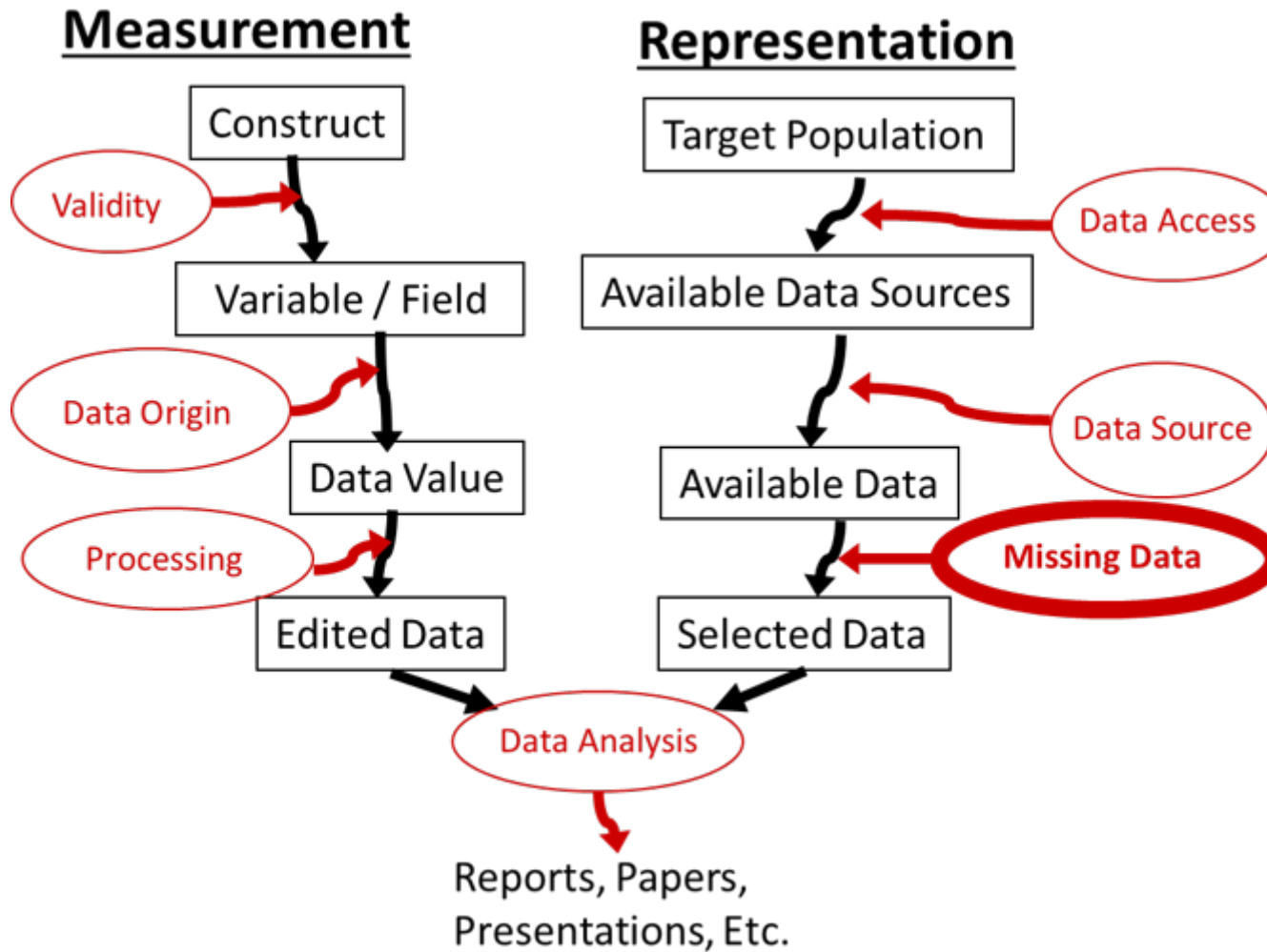


Defining Data Missingness

By James Wagner

Dimensions of TDQ



Data Missingness Definition (1)

- **Definition:** A value (or values) is expected, but not observed.
- Two types of missing data:
 - Missing **value**: one or more variables (item).
 - Example: Failure to complete all elements of a medical record.
 - Missing **record**: the whole record is missing (unit).
 - Example: Sampled person refuses to respond to survey.

Data Missingness: Definition (2)

- Critical to understand the *mechanism* of the missingness.
- Examples:
 - A human refused to supply the value/record.
 - The value is sensitive, therefore, the human does not want to report.
 - The human does not, at the moment, feel like taking the time to supply a value.
 - A human forgot to supply the value/record.
 - Two datasets were linked together, and some records failed to link.

Data Missingness: Definition (3)

- Mechanism may produce missing data that are either *ignorable* or *non-ignorable*.
 - ***Ignorable***:
 - Not biasing / Missing Complete at Random (MCAR) or,
 - A problem that can be addressed using the observed data / Missing at Random (MAR).
 - ***Non-ignorable***: The missingness depends upon the unobserved values.
 - Example: Persons who use drugs are less likely to report an answer to a question about drug use.
 - See *Little and Rubin (2019)* for more complete definitions.

What's next?

- Next, we will look at **data missingness threats** for designed data.



© Faculty Presenter

**Except where otherwise noted, this
work is licensed under CC BY-NC 4.0**