Importance of quality COVID-19 data collection and its implications for public health action

“You can have data without information, but you cannot have information without data” - Daniel Keys Moran
1. How does data collection impact decision support in section 4?

2. How can we organize the data so that we can best analyze it?

3. What do we learn from this data?

4. How will this knowledge translate to action?
How does public health use COVID-19 data?

Emerging Diseases
- Detecting new virus, bacteria or other germ
- Number of illnesses, hospitalizations and deaths caused by the new virus, bacteria, etc.

Relationships, trends, and causes
- Age, people with pre-existing medical conditions, racial/ethnic backgrounds, location (i.e. nursing home, jail, schools)

Solutions
- Guidance: Wear a mask, stay 6 feet apart, avoid crowds and poorly ventilated spaces, wash your hands
- Resource allocation
- Treatment options and vaccines

When collected and analyzed properly, data can be converted into usable information to uncover
Importance of quality data collection

Good decisions are based on good and credible data

Lab reports to treat individuals

Data to treat communities
Data quality

Data quality requires a team-based approach and is often defined as “fitness for use”

- Fit for their intended uses in operations, decision making, and planning
  - Those involved in designing REDCap surveys

- Reflects real value or true performance
  - Those involved in interviewing cases/contacts

- Meet reasonable standards when checked for quality
  - Those involved in data quality checks

How well our data tells the true story
Six core dimensions of data quality

- Completeness
- Uniqueness
- Timeliness
- Validity
- Consistency
- Accuracy
Data quality dimensions

- **Completeness** (Are all the data items recorded?)
  - The proportion of stored data against the potential of “100% complete”
  - If someone skips some of the questions on a survey, it may make the rest of the information they provide less useful.
    - Ex: If all workplace exposure questions are completed EXCEPT for “What was the date that you last went to your workplace?” then the rest of that information is less useful. Employers need this date to determine which employees may need to quarantine.
  - Related dimensions: validity and accuracy

- **Uniqueness** (Is there a single view of the record?)
  - Nothing will be recorded more than once based upon how that record is identified
    - Ex: Fred Smith and Freddy Smith
    - Fred Smith Male and Fred Smith Female
  - Related dimensions: Consistency
Data quality dimensions Cnt’d

• **Timeliness** (Are all the data items recorded in a timely manner?)
  – The degree to which data represent reality from the required point in time
    • Ex: Data quality is diminished when information isn’t collected properly on the first try. This is a common occurrence for flight information.
  – Related dimensions: Accuracy as data typically becomes less useful and less accurate as time goes on

• **Validity** (Does the data match the rules?)
  – Data is valid if it is in the right format, of the correct type, and falls within the right range
    • Ex: For any gatherings attended by a case, the date of the gathering should be within the two days prior to the case’s symptom onset date or test date since this is the timeframe asked about.
  – Related dimensions: Accuracy, Completeness, Consistency and Uniqueness
Data quality dimensions Cnt’d

- **Consistency** (Can we match a record across all databases?)
  - Similarity between multiple versions of a single record across different databases (Labs, surveillance system, REDCap, etc.)
  - A record should be consistent both in its content and its format
  - Related dimensions: Accuracy, Validity, and Uniqueness

- **Accuracy** (Does the data reflect the real values?)
  - The degree to which data correctly describes the "real world" object or event being described
  - DOB is May 1, 2014: 05/01/2014 (mm/dd/yyyy) 01/05/2014 (dd/mm/yyyy)
  - Incorrect spellings, First name in last name column and vice versa
  - Related dimensions: Validity, Uniqueness, Consistency
Every interview question in the survey has a reporting purpose.

Every answer to the questions in these sections is important to gather details such as:

- When the case was contacted
- Who completed the interview
- Whether the case received a COVID-19 test result and when they tested positive
- If and when the case developed symptoms
- Where they were while infectious so regions can investigate
- When they flew so we can notify CDC
- Who their close contacts are so they can be notified of their exposure
- Documentation so individuals can return to work or school
Data loss happens when questions are not asked/reported

Complete, Accurate, and Timely data help to make important decisions and save lives
• The Six Dimensions of EHDI Data Quality Assessment (cdc.gov)

As our data collection efforts continue to improve, so will our ability to decrease COVID-19 infections in our community, state and around the world.