

---

## Module 1: Data Quality Indicators (DQIs)

Common Measures Training  
Chelmsford, MA  
September 28, 2006

---

---

---

---

---

---

---

---

### Goals of Module

---

- Review DQI concepts
- Illustrate data quality issues that might arise in this project

---

---

---

---

---

---

---

---

### Overview of Module

---

- Types of data
- DQIs
  - Precision
  - Sensitivity
  - Bias
  - Representativeness
  - Completeness
  - Comparability
- Data quality objectives (DQOs)

---

---

---

---

---

---

---

---

## Two Types of Data (Quality Perspective)

- Data whose quality you **can** affect (primary data). E.g.,
  - New random inspection data collected by your program
  - New certification data collected by your program
- Data whose quality of collection you **cannot** affect (secondary data)
  - All existing data
  - New data collected by others

To be accepted, *all* data must meet *your* agreed-upon quality objectives.

---

---

---

---

---

---

---

---

## Matrix: Primary Versus Secondary

	New Data	Old Data
Collected by participants	Primary (if for this project)	Secondary
Collected by others	Secondary	Secondary

---

---

---

---

---

---

---

---

## Review of the Six DQIs...

- Definition
- Everyday example
- Examples meaningful to this project
- Explore relationships among DQIs:
  - Recognizing quality issues is more important than categorizing them
  - Don't get hung up on distinctions

---

---

---

---

---

---

---

---

## Precision

Measure of **agreement among repeated measurements** of the same property under identical or substantially similar conditions

---

---

---

---

---

---

---

---

## Examples of Precision Issues

- **Measuring a child:** How did she get shorter?
- **Ambiguous questions:** "Has the facility made efforts to reduce the volume of its hazardous waste?"
- **Statistical sampling:** Confidence level and margin of error

---

---

---

---

---

---

---

---

## Looking through Different Lenses

Each DQI can apply at multiple levels of analysis

For example, precision applies in regard to:

- Vague phrasing of an indicator question
- Accuracy of responses to the indicator question (e.g., comparison of self-cert responses and inspector findings)
- Simple statistical analysis of responses
- Statistical comparison with responses from other states

---

---

---

---

---

---

---

---

## Sensitivity

Measure of the **capability of a method or instrument to discriminate** between measurement responses representing different levels of the variable of interest.

- How fine are the units of measurement?

---

---

---

---

---

---

---

---

## Examples of Sensitivity Issues

- **Cooking:** In your dish, can you taste the difference one grain of salt makes? 1 cup?
- **Quantitative questions:** "How much waste is generated?" vs. "Is more than 220 pounds of waste generated?"
- **Rolled-up questions:** "Facility labels properly?" vs. "Facility has labels on all containers?" *and* "All labels show correct contents of containers?"
- **Observability:** Can you be sure something occurred? (E.g., "efforts" to reduce hazardous waste.)
- **Analyzing environmental samples:** "Minimum detection limit" defines maximum sensitivity.

---

---

---

---

---

---

---

---

## Balance of Precision and Sensitivity

- Beware "spurious precision"
  - E.g., sequential measurements of 2010 lbs, 1600 lbs, and 2499 lbs all round to one ton. Precision is gained by reporting in tons, but artificially--useful sensitivity lost.
- Conversely, beware "spurious sensitivity"
  - If you collected those amounts in ounces, it's more sensitive, but how likely is precision at that level?

---

---

---

---

---

---

---

---

## Bias

**Systematic or persistent distortion** of a measurement process that causes **errors in one direction.**

---

---

---

---

---

---

---

---

## Examples of Bias Issues

- **"Dewey Beats Truman":** A telephone poll is biased in favor of telephone owners
- **Data collector:** "Harsh" inspector in State A and "Easy" inspector in State B
- **Self-selected sample:** Self-certification data from a voluntary certification program
- **Interested party:** Facility-reported data, relative to inspector-collected data

---

---

---

---

---

---

---

---

## Representativeness

Degree to which a sample **accurately and precisely represents the larger context.**

Lack of representativeness can...

- Be a source of bias
- Create comparability problems

---

---

---

---

---

---

---

---

## Examples of Representativeness Issues

- **Mixing:** Stir a fluid before taking a sample (e.g., cooking a dish)
- **Defining your Indicator:** Hazardous waste generation amounts
  - Monthly versus annual
  - Maximums versus averages
- **Randomness:** Random sample is representative (but of what?)
  - All volunteers, all registered facilities, or all facilities?

---

---

---

---

---

---

---

---

## Completeness

Measure of the **amount of valid data needed** to be obtained from a measurement system.

- Incompleteness can be a source of bias

---

---

---

---

---

---

---

---

## Examples of Completeness Issues

- **Cooking:** Do you have enough of all the ingredients to make do?
- **Universe:** Have all eligible facilities been identified?
- **Response rate:**
  - What percentage of surveys are returned?
  - How many questions are left blank?

---

---

---

---

---

---

---

---

## Comparability

Measure of **confidence** that the underlying assumptions behind two data sets are similar enough **that the data sets can be compared** and/or combined to inform decisions.

- **Key comparisons in this project:**
  - Intrastate (over time, among subgroups)
  - Interstate (between states, over time)
- **Other DQIs play a role in comparability**



---

---

---

---

---

---

---

---

## Examples of Comparability Issues

- **Interpretation:** "Amalgam wastes are properly collected and stored." Ambiguity across states?
- **Timing:** Compare data collected in the spring with data collected in the fall? Collected three years apart?
- **Representativeness:** Did two states define SQGs in the same way? Is the universe of facilities comparable in scope?
- **Normalization:** Tracking a background variable (e.g., total population, total production) that puts a variable of interest into perspective...



---

---

---

---

---

---

---

---

## Normalization and Comparability

- Report secondary variables to ensure that two data sets are comparable
- Example:
  - State A and B estimate 100,000 gallons of used oil recycled, each.
  - State A has 200 auto body shops, State B has 400.
  - State A has 500 gallons per shop, while State B has 250.
- Even better: gallons per car repaired (if precision sufficient)



---

---

---

---

---

---

---

---

## ID the DQI Issue

---

"Sufficient records are maintained to demonstrate compliance."

---

---

---

---

---

---

---

---

## ID the DQI Issue

---

"How much amalgam separator waste was collected in the last month?"

---

---

---

---

---

---

---

---

## ID the DQI Issue

---

"Is the facility in compliance with underground injection control requirements?"

---

---

---

---

---

---

---

---

## Data Quality Objectives (DQOs)

- **Role of DQOs:** Identify minimum standards for data acceptability
- **How many DQOs?** Set DQOs for each critical DQI issue
- **Perfection?** Not necessary or expected.
- **KEY FOR DQOs:** *Sufficient* for needs and *Achievable* by all

---

---

---

---

---

---

---

---

## DQO Examples

- **Completeness:** 90% of certification forms returned; 95% of responses completed for each measure
- **Precision:** 95% confidence that survey results are accurate within +/- 10%
- **Representativeness:** Fluid will be mixed thoroughly before an analytical sample is taken
- **Sensitivity:** Hazardous waste will be reported in tens of pounds, and a common conversion rate will be used to convert gallons to pounds

---

---

---

---

---

---

---

---

## How Strict Should DQOs Be?

Depends on:

- **Data Use:** What kinds of decisions will they inform? Budgetary? Regulatory?
- **Types of Analyses:** Do the DQOs support the questions you want to answer? Think ahead!
- **Resources:** What can be achieved with available resources?
- **Feasibility:** The need for using a particular secondary data source may limit DQOs related to that source

---

---

---

---

---

---

---

---

## Rules of Thumb for DQOs

- **No surprises:** Make sure quality will be good enough for your needs.
- **Transparency:** Report all unresolved, important quality issues.
- **Achievability:** Too onerous, and data won't be collected or data will be rejected.



---

---

---

---

---

---

---

---

## For more information...

Contact Michael Crow

- E-mail: [mcrow@cadmusgroup.com](mailto:mcrow@cadmusgroup.com)
- Phone: 703-247-6131



---

---

---

---

---

---

---

---