

Module 5: Principles of Good Data Collection

Common Measures Training
Chelmsford, MA
September 28, 2006

Goals of Module

- Develop understanding of quality issues connected with
 - Formulating measures
 - Collecting primary data
- *Remember:* If you collect the data yourself, you can influence quality

Overview of Module:

- Step 1:** Selection of indicators
- Step 2:** Design of data collection instruments
- Step 3:** Collection of data
- Step 4:** Data processing

Step 1: Selection of Indicators

- Purpose of Indicators
- Types of Indicators
- How to Formulate
- How Much Data to Collect

Purpose of Indicators

- Provide shorthand understanding of the performance of a facility (or group) at a point in time.

Different Types of Indicators

- Activity-based v. outcome-based
- Regulatory v. beyond-compliance

Different Ways of Formulating Indicators

- Quantitative v. Y/N questions
 - "Y/N" is a.k.a. "dichotomous"
 - Multiple-choice questions are essentially dichotomous
- Single-item questions v. rolled-up questions

Use of Dichotomous Questions

- **Measure the Use of Practices**
 - E.g., *All registered underground storage tanks (USTs) at the facility are equipped with overfill protection?*
- **Measure Quantitative Information**
 - E.g., *Does the facility generate less than 220 pounds of hazardous waste per month?*
 - E.g., *Is the facility a non-handler, CESQG, SQG or LQG?*
 - As opposed to: *How much hazardous waste does the facility generate per month (in pounds)? _____*

Dichotomous Questions for Quantities

- **Easier to interpret, enter, manipulate and analyze** than responses to open-ended questions (*good precision*)
- **Less sensitive:** less information about small differences
- **Less flexible:** won't fit different thresholds or categories

Considerations for Quantitative Questions

- **Sensitivity.** How much sensitivity is required, and how much is feasible?
- **Units.** Specify the units you want.
 - E.g., ounces, pounds, or tons



What Are Single-Item Indicators?

- Stand-alone questions
- Believed important and representative of larger performance
- Not necessarily based on supporting data



Considerations for Single-Item Indicators

- Streamlines data collection
- Best with supporting data (e.g., correlation analysis)



What Are Rolled-Up Indicators?

Indicators based on explicit sub-questions

Example:

- **Hazardous waste is stored and labeled properly** (Y/N)
(answer "yes" if "yes" to all six of the following)
 - All storage containers are covered (Y/N)
 - All storage containers are clearly labeled "hazardous waste," and contents are identified (Y/N)
 - Only one type of waste is stored in each container (Y/N)
 - Containers are constructed from material that will not react with waste (Y/N)
 - All storage containers are kept in a secure location (Y/N)
 - All storage containers are located on a crack-free surface that will contain leaks or spills (Y/N)

THE
CADMUS
GROUP, INC.

Considerations for Rolled-Up Indicators

- **Less subjective** for inspectors
- **Greater consistency** across inspectors
- **Capability to change** indicators without new data
- **More time-consuming** data collection

THE
CADMUS
GROUP, INC.

What Are Outcome Measures?

Measures of impacts on:

- Environment
- Worker health and safety
- Public health

THE
CADMUS
GROUP, INC.

Why Use Outcome Measures?

- Focus on the overall mission
- Help prioritize
- Respond to stakeholders

How to Get Outcome Measures?

- **Measure directly** (e.g., air and water sampling, blood lead testing)
 - More accurate, perhaps
 - More resource-intensive, if you still want practice-based data, too
- **Infer from other measures**
 - Uses models and assumptions
 - Less defensible, perhaps

Either approach can vary in analytical complexity

Calculating Outcome Measures – Two Examples

- MA Photo Processors (Less Complex)
- MA Dry Cleaners (More Complex)

Outcome methodologies prepared by contractor for MA DEP

MA Photo Processors

- **Increase in Silver Recovery** = (% Increase in Use of Silver Recovery Technology) * (Silver Concentration in Waste) * (Discharge Rate) * (% Silver Removal Efficiency)
- Separate calculations made and aggregated for different facility size classes, different types of waste streams.

MA Dry Cleaners

- **Perc Release Reductions** = (Year 2 Emissions) – (Year 1 Emissions)
- **Yearly Emissions** = (Emissions from Facilities that Do Not Perform Leak Detection) + (Emissions from Facilities that Inspect But Do Not Find Leaks) + (Emissions from Facilities that Find a Leak But Do Not Fix It) + (Emissions from Facilities that Find and Fix a Leak)
- **Emissions from Facilities that Find and Fix a Leak** = (Total Emissions from a "Typical" Facility) * (Number of Facilities in Sector) * (Percent of Facilities that Perform Leak Detection) * (Percent of Leak Detections Estimated to Result in Discovery of a Leak) * (Percent of Leak Discoveries Estimated to Result in Remedial Action by Facility) * (Percent Estimated Reduction in Emissions Caused by Remedial Action)
- Etc., etc. Separate calculations made and aggregated for facilities that use different dry cleaning techniques, and thus have different "typical" total emissions.

Considerations for Outcome Measures

- Mission-oriented
- More understandable to non-technical stakeholders
- May be difficult to link outcomes to changes by the group
- Balance complexity with credibility and potential for error
- Report low, best and high estimates?

Issue: How Much Data To Collect?

Different philosophies:

- **Focus on a few key indicators** determined in advance (e.g., MA ERPs)
 - Easier to ensure sampling precision
 - More streamlined in collection and analysis
- **Cast a broad net** and then study the data to identify correlations (e.g., RI auto body ERP)
 - Greater flexibility
 - Indicator defensibility



Step 2: Questionnaire Design

- Types of Questionnaires
- Questionnaire Design Tips
- Questionnaire Testing



What Kind of Questionnaires?

Questionnaire should gather data in a uniform manner

For example:

- Surveys
- Self-certification forms
- Inspector checklists

We focus on inspector checklists, but relevant to others



Questionnaire Design Tips

- Avoid Double Negatives
- Make Clear Choices
- Provide Definite Time Frames
- Consider Objective Measurability

For others, see the review list in the binder



Design Tip: Avoid Double Negatives

- **Unclear:** *Does the facility avoid using unapproved corrosion protection systems?*
- **Explicit:** *Does the facility use only approved corrosion protection systems?*



Design Tip: Clear Choices

- Make sure multiple choices don't overlap
- Ensure mutual exclusivity when required
- Make sure everybody doesn't choose "other"



Design Tip: Definite Time Frames

- **Unclear:** *Does the facility maintain records of inventory control?*
- **Explicit:** *Does the facility have records of the last thirty-six (36) months of inventory control?*



Design Tip: Consider Measurability

- *"Are employees aware of P2 practices?"* – difficult to observe
- *"Do employees received documented P2 training at least once a year?"* – better
- *"Are all waste containers covered when not in use?"; "Are spill response kits available?"* -- even better



Before Using Questionnaires in the Field...

Test them!

- Helps **train inspectors** and ensure they **interpret questions the same way** (precision and comparability)
- Helps **identify any errors or ambiguities** (precision and completeness)



Get Inspector Input on Questionnaire

- Draw on their expertise
- Ensure the questions are clear and understandable
 - E.g., written in familiar jargon
- Get buy-in



Step 3: Data Collection

- Who Collects the Data?
- Tips for Inspector Training
- Timing Considerations
- Electronic v. Paper



Who Is to Answer Questionnaire?

- Facilities (self-reporting)
- Professional inspectors
- Interns

Each option can be evaluated in terms of practicality and quality



Who Is to Answer Questionnaire?

- Facility self-reporting
 - Raises bias concerns
 - Training less reliable
 - Larger sample possible?



Who Is to Answer Questionnaire?

- Professional inspectors
 - Knowledgeable and experienced
 - Likely to have authority to enter facilities
 - Can take follow-up action



Who Is to Answer Questionnaire?

- Interns
 - Less experienced/knowledgeable
 - Blank slate, more easily trained to new protocols
 - Less costly than inspectors
 - Might not have the right to enter facilities



Musts for Data Collection

Train inspectors so that they...

- Interpret questions consistently
- Understand and follow sampling protocol
- Answer all questions (no blanks)
- Always use ink
- Use consistent protocol for corrections
 - e.g., double-strikeout plus initials

THE
CADMUS
GROUP, INC.

More on Consistent Interpretation

- Interpretation affected by underlying spirit of inspection
 - Are inspectors looking for faults and violations, or trying to gather a holistic impression of facility performance?
- Training inspectors together can help align approach
- But there will be differences and potential biases
 - E.g., more violations in one state might mean simply that those inspectors were more nit-picky.
- Best strategy is to make indicators as specific and explicit as possible.

THE
CADMUS
GROUP, INC.

Timeframe of Data Collection

- Short timeframe desirable
- Consider seasonality issues
- Distance in time complicates comparability

THE
CADMUS
GROUP, INC.

Timeframe of Data Collection

- **Short timeframe desirable**
 - E.g., 2-3 months
 - More precise "snapshot" of performance
 - First facility sampled is relatively close in time to last facility sampled

Timeframe of Data Collection

- **Consider seasonality issues**
 - E.g., will the amount of antifreeze recycled be greater in the months immediately preceding winter?
 - If issues, need commonality between rounds and between states

Timeframe of Data Collection

- **Distance in time complicates comparability**
 - E.g., data from two states collected three years apart
 - Are extraneous factors impacting comparison?
 - E.g., current UST data vs. pre-Energy Bill

Electronic v. Paper

- **Electronic data collection** (e.g., on hand-held tablets)
 - Higher investment costs
 - Lower variable costs
 - Error-checking and validation can be automatic
- **Paper forms**
 - Lower investment costs
 - Higher variable costs
 - Greater data entry error



Step 4: Data Processing

- Tips for Data Entry by Hand
- QA Options

Note: Very brief highlights of this topic today. More training later in the project.



Data Entry by Hand

- Subject to error, like any rote task
- Should receive some sort of quality assurance (QA)



If Inspectors Enter Their Own Data...

Have them do it as soon as possible
after inspection, so it is fresh in their
minds.



QA Options for Data Entry

- Complete manual review of some fraction (e.g., 10%) of the data
- Double-keyed data entry
- Automatic validation



Data Administration

- Data *administration* is how data are handled and archived.
- Important issue, but outside the scope of this training.



For more information...

Contact Michael Crow

- E-mail: mcrow@cadmusgroup.com
- Phone: 703-247-6131


