Available Data Sources and Possible Preliminary Analysis for TCR

Doug Owen
TWG Representative
TCRDS Federal Advisory Committee
September 18-19, 2007

This Meeting's Presentations

Objectives

Provisions

System compliance

Improvement

Research

Available Data Sources and Possible Preliminary Analysis for TCR

Presentation Objectives

- Review available data sources and characteristics
- Questions that available data could help answer

Sources of Data

- State/EPA Data
  - 6-Year Review Data (SDWIS/State and other)
  - Individual data sets
- Community Case Studies and Water System Information
  - Monitoring Data, Surveys, Case studies
- National TCR violation data (SDWIS/Fed)

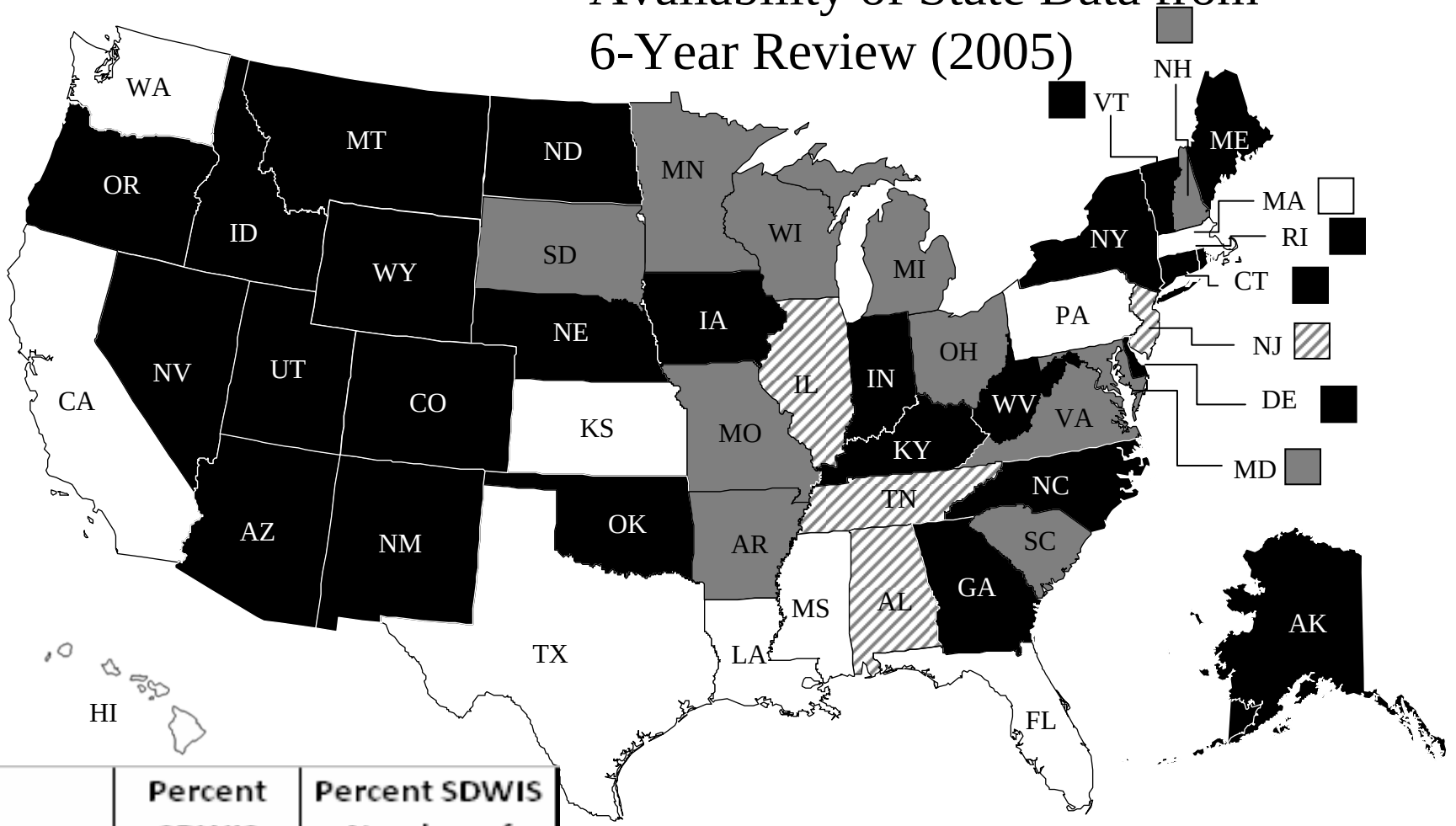
6-Year Review Information Collection Request (TCR Data)

- 46 States have provided data beginning as early as 1998
- Data elements may include:
  - System info: system type, population, source water type
  - Sample types: Routine, repeat, special, unknown
  - Sampling time: Collection date
  - Sample locations: Entry point, distribution system; for repeat samples: downstream, original, upstream, and other
  - Methods: method codes
  - Results: presence/absence for TC, Fecal, & EC

Six Year Review Data

- Voluntary request for data, EPA provided a tool to extract data for states using SDWIS/State
- Focusing on data from 42 primacy agencies (see map)
  - Tier 1: 28 states used the Extract Tool (only accepted compliance samples, no rejected samples or monthly summaries)
  - Tier 2: 14 other states submitted data in other formats
- May not have data from 17 primacy agencies
  - Tier 3: 4 states submitted incomplete or problematic data
  - Tier 4: 13 states did not provide data. Some large states are not participating (CA, FL, PA, MA, WA), but TX has agreed to provide data
- Over 9 million TCR records 1998-2005

Availability of State Data from 6-Year Review (2005)



Tier	Percent SDWIS Population	Percent SDWIS Number of Systems
Tier 1	30%	36%
Tier 2	15%	30%
Tier 3	12%	8%
Tier 4	43%	26%
Total	100%	100%

- Tier 1: Submitted data using SDWIS Extraction Tool
- Tier 2: Submitted data NOT using SDWIS Extraction Tool
- Tier 3: Submitted incomplete data
- Tier 4: Did NOT submit data

Six Year Review Data: Preliminary Findings

- Date of sample collection and analyte/result fields are generally of very good quality and are complete
- Data from 2002-2005 appear to contain more systems than from 1998-2001
- Data from Tier 2 states provide additional parameters that could be used for analysis

Six Year Review Data: Preliminary Findings Continued

Some states appear to be have data gaps

- A few states reported too few systems (overall)
  - AL (only 170 unique systems in state dataset, but 636 systems in SDWIS/FED), NJ (698 of 3,905), NY (9,200 of 9,769)
- Some of the largest systems in each State are not in databases (based on spot checks)

MD—at least four	MI—at least eight
NV—second largest system	SD—at least eight

- Individual TC sample records not available for all systems
  - Summary form of results submitted for some systems in at least NJ, AZ, NY, and RI

Six Year Review Data: Categories of Data

- Routine/Repeat
- 18 system categories (see next slide)
  - Type of system
  - Size of system
  - Source water (for CWS)
- By Year (1998-2005 available)

Categories

Type of System	Source Water	Population Served
All		
CWS		
NCWS		
NTNCWS		Total
		0-24
		25-1,000
		>1,000
TNCWS		Total
		0-24
		25-1,000
		>1,000
CWS	GW	Total
		0-24
		25-1,000
		1,001-4,100
		4,101-33,000
		33,001- 100,000
		>100,000
	SW	Total
		0-24
		25-1,000
		1,001-4,100
		4,101-33,000
		33,001- 100,000
		>100,000

- < 1,000 take 1 or fewer samples per month
- < 4,100 can have sanitary survey replace some monitoring
- < 4,100 must at least 5 samples the month following a TC+
- > 33,000 collect 40 or more samples and 5% criteria for violations

Non-Six Year Review Data

- TCR State compliance data
  - California Dataset (7 of 22 districts)
    - Compiled by CEC; formatted and available for analysis
    - Represents 1,489 water systems with 630,000 observations
  - 15 State Dataset
    - TC and E. coli presence/absence rates for multi-years of data based on 6.8 million observations
    - Compiled by CEC; formatted and available for analysis
- Data Verification (DV)

Survey Data

- ASDWA Survey – TCR implementation practices
- AWWA Survey - State response to E. coli and non-E. coli coliform detections
- AMWA Survey – Summary data for routine/repeat/rejected samples by year (2002 – 2006)

EPA Data Sources

- SDWIS/Fed Data
  - Violation data by Region, State, system type, system size, and source water type from 1998 to 2005
    - Non acute MCL, acute MCL, monitoring / reporting
  - Disinfection status (yes/no) for about ¼ of systems
- Community Water System Survey
  - Characterization of PWSs

Specific Questions that Available Data Could Answer

- Federal violation statistics by system category
- For the 6-Year Review and other data:
  - Sampling statistics by size category and type, for example:
    - Percent of overall samples that are TC+, EC+, and FC+
    - Estimates of burden from sampling

For the 6-Year Review and Other Data (continued)

- Informative potential of current monitoring:
  - Repeat samples
  - Extra next-month samples
- Impact of alternative compliance calculations
- Evaluation of correlations between TC+, EC+, and FC+ and water quality parameters (e.g. chlorine residual)

What questions would the FAC like the TWG / data analysis task group to attempt to answer?