



Drinking Water Data Reliability Analysis and Action Plan (2003)

For State Reported Public
Water System Data In the
EPA Safe Drinking Water
Information System/Federal
Version (SDWIS/FED)

Office of Water
EPA 816-R-03-021
March 2004
www.epa.gov/safewater

Printed on Recycled Paper

Drinking Water Data Reliability Analysis and Action Plan (2003)

Executive Summary

Safeguarding our nation's drinking water by developing effective and appropriate policy decisions and conducting program oversight depends on data of known and documented quality. The Safe Drinking Water Information System/Federal Version (SDWIS/FED) is the Environmental Protection Agency's (EPA) principal data base for the national drinking water program. It contains data on public water systems (PWS) provided by states to EPA. It is primarily used for management of state and EPA programs for informing the public about the compliance status of their drinking water systems, and indirectly, the safety of their drinking water. The utility of SDWIS/FED information for these purposes depends on the quality of the data it contains.

EPA routinely evaluates state programs by conducting data verification audits, which evaluate state compliance decisions and reporting to SDWIS/FED, and conducting triennial national summary evaluations. This document presents results of EPA's second triennial review of data quality in SDWIS/FED, and includes an evaluation of data collected from 1999 through 2001. The first triennial review, published in 2000, analyzed drinking water data from 1996 through 1998 to establish a data quality baseline. This second review indicates that data in SDWIS/FED are highly accurate, but still incomplete. This finding raises questions about impacts on both effective program management and accurate risk communication.

Background

SDWIS/FED contains data about PWS facilities, violations (e.g., exceptions, exceedances) of Federal drinking water regulations adopted by the states, and enforcement actions at the facilities. These regulations include health-based drinking water quality standards, performance of treatment techniques, or process requirements. The Federal government uses SDWIS/FED data for program management for 84 contaminants (as of 2001) regulated in drinking water at more than 160,000 PWSs in 56 state and territorial programs, and on Indian lands.

States develop their own processes and data bases to document public water system capabilities and their program management decisions concerning violations (or noncompliance), and to record corrective actions undertaken. State data indicate that violations occur infrequently at most public water systems (PWS). Violations data that states report to EPA, contained in SDWIS/FED, reflect only those major and minor noncompliance results that might lead to adverse public health outcomes. These data represent a small fraction of all the determinations states make which demonstrate the safety of the nation's water supply.

The first triennial review of data quality reviewed data for the period 1996-1998. That assessment, which resulted in a detailed data analysis report in 2000, also produced an action plan under which states and EPA worked together to improve data quality. The plan resulted in actions which included training states, streamlining reporting to SDWIS/FED, making SDWIS

error reporting and correction more user-friendly, reducing rule complexity, improving data verification audits (DVs), following up with Regions after DVs, and encouraging states to notify water systems of sampling schedules annually.

Many of EPA's actions have focused on assessing and improving SDWIS/FED and its supporting tools, including SDWIS/STATE. EPA and states designed SDWIS/STATE to support state drinking water programs. This database software application automates compliance determinations for federal and state drinking water rules.

Second Triennial Review

Like the first review, this second triennial review of data quality is largely based on DVs. The DV's, conducted between 1999-2001, reflect data for 1,890 randomly selected PWSs in 31 states. To support a comparative analysis, a similar set of analyses and methods were used in this review as in the first, where possible. From this second analysis, improvements and new or continuing problem areas are described, and recommendations are made for quality improvements. At this time, the detail and national scope of the drinking water program data evaluation are extensive and considered to be a robust examination of this matter for data quality purposes.

Data quality is calculated as the percentage of data without any discrepancies or errors. The primary difference between discrepancies and errors is that, for errors, the reviewer verifies what the correct data should have been. Inventory and enforcement actions were reviewed to identify discrepancies between state files and SDWIS/FED, whereas violations were reviewed to identify errors. For violations data, auditors determined violations that should have been identified based on results in state files and compared those to violations reported to SDWIS/FED. This review provided for assessments of completeness and accuracy of violations data.

EPA developed an interim ranking scheme in order to simplify the characterization of data quality. Final objectives will be developed in 2004-2005 with input from stakeholders. The interim scheme assigns 0 to 70 percent to Low quality, 71-90 percent to Moderate quality and 91-100 percent to High quality.

Summary of Results

The data quality of core inventory data was high and essentially remained the same as that determined for the first assessment (95% vs. 96%). Data quality for enforcement data improved from 72 to 80% and remained of moderate quality. Notably, there were 80% fewer enforcement actions from the first assessment.

Violations were grouped by maximum contaminant level (MCL), treatment technique (TT), and monitoring and reporting (M/R) violations. MCL violations were further broken down into Total Coliform Rule (TCR) MCL violations and Chemical and Radionuclides MCL

violations (i.e., Other MCL). Violations for the lead and copper rule were not included in the analysis. While quality has improved considerably in several areas, the analysis of DV findings found that the violations data (Table ES-1) reported by states to EPA were very accurate but incomplete in several important aspects, as described below.

- C The overall data quality for health-based violations (i.e., MCL and SWTR TT) improved from 40% in the first round to 65%. The overall data quality for M/R violations increased from 9 % to 23%.
- C The Total Coliform Rule had the highest data quality, improving from 68% in the first assessment to 75%. Data quality is lower, but improving, for other health-based standards including Chemicals and Radionuclides and the Surface Water Treatment Rule. Data quality for M/R violations has improved, but is still low.
- C Most violation errors are due to incorrect compliance determinations by states, that is, violations not correctly identified as such.
 - Compliance determination errors accounted for slightly more than 50% of all MCL errors, and more than 85% of SWTR TT and M/R errors.
 - Half of the M/R errors were due to a failure of the state to assign a violation where sample data was missing from state files.
 - Data flow errors (data in state databases but not in SDWIS/FED) account for 9% of all errors.
 - Over-reporting of violations found in SDWIS/FED accounted for less than 2% of all errors. This is comparable to findings from the first assessment.
- C Data quality estimates are similar across water system types.

Table ES-1: Data Quality Estimate by Selected Major Violation Categories

Data Quality Estimate by Violation Type					
Data Quality Accounting Category	Violation Type				
	TCR MCL	OTHER MCL	TOTAL MCL	SWTR TT	Monitoring and Reporting
% OF SYSTEMS W/VIOLATIONS	5.5%	1.2%	6.7%	6.4%	49.5%
# OF VIOLATIONS	136	38	174	59	3,021
# <i>compliance determination errors</i>	15	15	30	23	2,018
# <i>data flow errors</i>	11	5	16	3	201
# <i>errors in SDWIS/FED</i>	8	1	9	0	103
# ERRORS	34	21	55	26	2,322
% COMPLETENESS	81%	47%	74%	56%	27%

% ACCURACY	93%	95%	93%	100%	89%
2003 SDWIS /FED DQE	75%	45%	68%	56%	23%
2000 SDWIS/FED DQE	68%	15%	54%	7%	9%

Additional analyses of SDWIS/FED data were conducted to further assess elements of data quality.

- C Timeliness of Violation Reporting: Many states are not meeting the 90-day deadline for reporting violations. In 2001, only 58% of violations eventually reported were reported on time. The timeliness in which health-based violations are reported has been steady, and is similar across water system types and sizes.
- C Non-Reporting of Violations: A significant number of states still periodically do not report violations of certain rules (particularly Radionuclides) from year to year, which needs further evaluation.
- C Data Rejection: An analysis of data rejected from SDWIS/FED found that 90% of the inventory, violations and enforcement data error types incurred were for data entry errors.

Data verification reports show that the following management policies and business practices are associated with high quality data: (1) routine, meaningful communication at all levels; (2) annual state notification to PWS of monitoring schedules; (3) automated monitoring compliance determination systems, and (4) electronic data transmission between laboratories or water systems and the state as well as between the state and SDWIS/FED.

An assessment in January 2003 compared the results of the two data reliability assessments for several states that had converted to SDWIS/STATE after the first assessment. States using SDWIS/STATE showed a decrease in data entry errors, however use did not completely eliminate compliance determination discrepancies or error conditions, nor always improve the timeliness of violation reporting.

Recommendations and Next Steps

While improvements in violations reporting are needed for all rules, a particular emphasis needs to be placed on improving reporting for Chemical and Radionuclides rules, the SWTR and M/R requirements for all rules. The report includes a series of recommendations that build on previous data quality improvement activities. They include:

- C Develop State-by-state compliance determination improvement action plans through existing state-EPA Regional planning processes
- C Continue state and EPA data quality analyses in accordance with the Quality

- Assurance and Data Reliability Action Plan
- C Encourage states to implement and maintain a quality assurance management plan
- C Develop data quality goals and measures to monitor progress
- C Modernize the SDWIS systems and related tool-sets to facilitate and improve the flow of data from states to EPA, in accordance with the Office of Ground Water and Drinking Water Information Strategic Plan
- C Encourage states to utilize automated tracking systems for key factors, decisions, and application of monitoring requirements, waivers, exemptions, vulnerability assessments, and resulting schedules
- C Identify factors impacting timeliness of compliance determinations and reporting violations to EPA
- C Continue to inform the public about the relationship between data reliability and water quality

The next triennial report will address progress on these activities. The public audience for this information has an expectation of high quality in violations data held in SDWIS/FED. This current data reliability analysis shows significant improvements in quality but that the data still fall short of public and program expectations.

Further evaluation is needed to determine how levels of data quality in SDWIS/FED, reflected in the above table, impact national program management. States have indicated that a number of factors affect their ability to improve data quality, including the complexity of rules, competing demands of current regulations, and limits on resources. To support state programs and their data quality improvement, EPA has reaffirmed its commitment to support SDWIS/STATE. An emphasis must be placed on ensuring the states are correctly assessing compliance with regulations and properly documenting those instances where systems fail to report the results of monitoring as violations.

Although data quality verification analysis provides valuable program management information, the link between data quality and public health and safety of drinking water from public water systems is indirect. Results of the reliability analysis are measures of the quality of data based on random selections of public water systems in each state. Therefore, results are not intended to evaluate safety of drinking water for the national population, particular water systems, nor groups of water systems.

As required by the Government Performance and Results Act (GPRA), EPA has established performance-based goals for the drinking water program. The overall goal is that by 2008, protect human health so that 95 percent of the population served by community water systems will receive water that meets all applicable health-based drinking water standards. Community water systems are the subset of public water systems that supply water to the same population year-round. EPA is evaluating whether a robust method can be developed that would use results from DV's to calculate a national number to report on this goal. However, it is certain that future improvements in data quality can only help EPA in meeting its objective of accurately reporting on public-health protection on a system, state and national level.

(This page intentionally left blank.)

Table of Contents

Executive Summary	ES-1
1.0 Background	-1-
1.1 <u>Introduction</u>	-1-
1.2 <u>Previous Activities</u>	-1-
1.3 <u>Regulatory Context</u>	-2-
2.0 Current Method and Analysis	-4-
2.1 <u>Data Quality Elements</u>	-4-
2.2 <u>Data Used in the Analysis</u>	-4-
2.3 <u>Sources of Data and Types of Analyses</u>	-5-
2.4 <u>Assessment of Data Quality</u>	-7-
3.0 Results and Findings	-11-
3.1 <u>Overview</u>	-11-
3.2 <u>Results for Inventory Data</u>	-11-
3.3 <u>Results for Violations Data</u>	-14-
3.4 <u>Results for Enforcement Data</u>	-18-
4.0 Additional Findings	-19-
4.1 <u>Evaluation of Large Systems</u>	-19-
4.2 <u>Potential Non-Reporting for Rules</u>	-19-
4.3 <u>Timeliness of Reporting</u>	-20-
4.4 <u>Rejection Error Analysis</u>	-21-
4.5 <u>Relationship between Use of SDWIS/STATE and Incidence of Errors and Discrepancies</u>	-21-
4.6 <u>State Compliance Determinations and Implementation Issues</u>	-22-
5.0 Recommendations For Improving SDWIS/FED Data Quality	-25-
6.0 Conclusion	-32-
6.1 <u>Findings</u>	-32-
6.2 <u>Implications for Government Performance Results Act (GPRA) Reporting</u> ..	-33-
6.3 <u>Continuing Coordination for Data Quality</u>	-35-
6.4 <u>Prospective Measures</u>	-36-

(This page intentionally left blank.)

Drinking Water Data Reliability Analysis and Action Plan (2003)

1.0 Background

1.1 Introduction

The Safe Drinking Water Information System/Federal Version (SDWIS/FED) is the Environmental Protection Agency's (EPA) principal data base for the national drinking water program. Its two major uses are (1) to help manage State and EPA programs and (2) to inform the public about the compliance status of public water systems (PWS) and, indirectly, the safety of drinking water. The Federal government uses SDWIS/FED data for program management for 84 contaminants (as of 2001) regulated in drinking water at more than 160,000 public water systems (PWS) in 56 state and territorial programs and on Indian lands. Data received by EPA from states for SDWIS/FED includes a limited set of water system descriptive information, data on violations of regulatory standards and process requirements at public water systems, and information on state enforcement actions. These data, which EPA uses to assess compliance with the Safe Drinking Water Act and its implementing regulations, represent the only data states are currently required to report to EPA relative to drinking water safety. SDWIS/FED data can be accessed from the EPA web site at www.epa.gov/safewater.

The utility of SDWIS/FED data for program management and public communication is highly dependent on the quality of data housed by the system. To assess this quality, EPA routinely conducts data verification audits in states and develops a national summary evaluation every three years. The auditors evaluate compliance data submitted by PWSs and compare data in SDWIS/FED with that in state databases.

This report includes: (1) a description of previous activities that have resulted in improvements in data quality and which are the foundation of future actions to enhance it, (2) an analysis of the data from 1999-2001, the most recent triennial evaluation period, and (3) recommendations stemming from extensions of past activities and findings from the current analysis. The report also describes a plan to address continued improvement in drinking water compliance data reported by states and continuing data quality improvement.

Industry and environmental stakeholders had an opportunity to review the draft version of this report to indicate their perspectives on its documented analytical processes and conclusions. Such opportunities for review by affected and outside parties provide an added measure of strength to the reported results.

1.2 Previous Activities

In 1998, EPA launched a major effort to assess the quality of the drinking water data contained within SDWIS/FED to respond to concerns of some utilities regarding incorrect

violations in the data base. EPA enlisted the help of its stakeholders in designing the review, analyzing the results for data collected between 1996 and 1998, and recommending actions to improve drinking water data quality. The first Data Reliability Analysis of SDWIS/FED was published in October 2000 (hereinafter “first assessment”).

Findings of that first assessment, which indicated that data quality needed improvement, included comprehensive recommendations for EPA and state primacy agencies on quality improvements. The report identified near-term actions that had already been taken or were actively underway to improve data quality more immediately. It specified a data quality goal, that 100 percent of the data in SDWIS-FED should be complete and accurate. Later, recognizing that a 100% goal for data quality may never be attained, the goal was modified to “at least 95%.” To implement the recommendations, the States and EPA have conducted numerous activities and projects to improve data quality. Activities undertaken have included:

- C providing training for states
- C streamlining reporting to SDWIS/FED
- C making SDWIS error reporting correction more user-friendly
- C reducing rule complexity
- C improving data verifications (DVs)
- C following up with Regions on findings after DVs
- C encouraging states to annually notify water systems of sampling schedules.

The Office of Ground Water and Drinking Water’s (OGWDW) response to the data reliability issues identified in the October 2000 report included a commitment to conduct analyses which would provide periodic data quality estimates (hereinafter DQEs), and provide input into program activities and priorities necessary to improve the quality and reliability of the data. Part of that commitment was to publish the results of these analyses every three years. In this second national data reliability report, findings from the individual analyses conducted during this period of review (1999-2001) are compared to those from the October 2000 baseline report.

1.3 Regulatory Context

The data considered for evaluating quality, particularly accuracy and completeness, are violations of monitoring and reporting requirements and health-based standards. These data are important because: (1) State and EPA program management relies on them to identify priorities and (2) States and EPA use them to inform the public about the safety of its drinking water. For federal program reporting purposes under the Government Performance Results Act (GPRA), violations data have become a major focus because EPA’s strategic plan specifies a clean and safe water goal of “95% of the population served by community water systems (CWS) meeting all health-based standards and treatments by 2005.” A CWS which meets all health-based standards and treatments does not have a violation of the federal regulations for maximum contaminant levels (MCL) or treatment techniques. In 2001, Federal regulations required PWS implementation of standards and treatments for 84 contaminants.

Public water systems provide states with results of monitoring required by drinking water

regulations. Each violation is the result of a series of state decisions regarding a PWS's compliance with the federal regulations and state enforcement programs. This data quality evaluation methodology focuses on a small subset of those actions with the bottom line from regulatory and public health standpoints to ask the question: Did the state correctly identify and report the violations which should have been reported to EPA according to state primacy agreement pursuant to Federal regulations?

States determine these violations from large amounts of data from monitoring results reported to them by PWSs. It is appropriate to consider the magnitude of inaccurate or incomplete (unreported) information in relation to the total number of decisions that states must make. Take, as an example, a ground water system with one source of water and one distribution system entry (water delivery) point. The estimated number of reports by major regulation for such a system is 42. If each contaminant identified in each regulation is considered, states would make an estimated 700 decisions as input to these 42 reports during a year for each ground water system.

The number of decision points representing the potential for violations varies significantly by PWS type, population served, rule, source type, number of entry points, analytical needs, whether a waiver or variance or exemption was granted, whether the PWS was on routine or reduced sampling schedules, and the most recent sampling results. Accounting for each decision made by a state for each PWS would add a level of complexity and workload, both at the state documentation level and for state data verification audits, and may not provide any "added value" when the issue is: "What is the quality of the violations data in SDWIS/FED?" If, for example, based on audit findings, there should be 50 violations and only 25 exist in SDWIS/FED, completeness is only 50% reliable - even though the state made 700 decisions while making final compliance determinations. Nevertheless, it is important to recognize that these violations may be a relatively small subset of the total number of state decisions and that data reliability percentage is affected by the total number of violations. For instance, one discrepancy out of 2 violations would yield a 50% reliability finding, for purposes of this report.

Further, it is likely that some of the discrepancies between the number of violations that should have appeared in SDWIS/FED and those found by the auditors could have included legitimate differences in rule interpretation in light of the flexibility provided to states in implementing rules under state primacy agreements. States implementation of rules must be at least as stringent as the Federal regulations, but can differ in substantial respects. Some of the follow-up actions recommended by this report are designed to delve into this subject in more detail and better document such instances.

2.0 Current Method and Analysis

This analysis evaluates the extent to which data reported by states to EPA for inventory, violations and enforcement actions deviate from the data quality objective of 95% for all drinking water data. To allow for a comparative analysis, an attempt was made to retain the same set of analyses and methodologies for this report as were used in the first assessment. Where changes were made, careful consideration was given to the impact of the change and the ability to statistically and logically justify the change.

2.1 Data Quality Elements

In evaluating data quality, one should consider two questions:

1. Is there information missing from SDWIS/FED?
2. How accurate is the information that is in SDWIS/FED?

There are four major elements of data quality. The first two are essentially variations on the two questions above:

- C **Completeness** looks at what percent of the data that should be in SDWIS/FED based on federal regulations and state primacy agreements is actually there.
- C **Accuracy** looks at how accurate the data that made it into SDWIS/FED are.

There are two additional elements of data quality:

- C **Timeliness**, which is a component of completeness, looks at the percent of violations data that were reported within a quarter after the end of the compliance period.
- C **Consistency** looks at whether the regulations were interpreted consistently.

2.2 Data Used in the Analysis

The data that states report to EPA and which were considered for analysis are:

- (1) Inventory data - information identifying public water systems, their water sources, treatments and other facility-level factors.
- (2) Violation data - Federal regulations specify the outcomes which states must report to EPA that result in noncompliance with: (a) specified monitoring and reporting (M/R) requirements necessary to determine whether sampling, testing and treatment process checking occurred as stipulated in Federal regulations, (b) health-based drinking water quality maximum contaminant levels (MCL) and related requirements for their attainment, and (c) health-based treatment techniques (TT) and associated water system management processes for

contaminants for which it is too technically difficult or uneconomic to set an MCL.

- (3) Enforcement data - Federal regulations indicate the conditions under which enforcement actions will be taken with a PWS to ensure public health protection if the system is in violation of the Federal-state drinking water program. States must report a subset of these actions to EPA. EPA reports these data for situations where EPA is the enforcement authority because the state or tribe has decided not to obtain approval to implement the federal program (e.g., Wyoming, the District of Columbia and on Indian lands).

2.3 Sources of Data and Types of Analyses

The two primary data sources used in these analyses were data in SDWIS/FED and the compiled findings from 31 state data verifications (DVs) completed from FY 1999 through FY 2001.

Several types of analyses undertaken with SDWIS/FED data which helped determine reasons for weaknesses in data quality included:

- C ***Completeness of programmatically required inventory data.*** This analysis reviewed two sets of inventory data. The first set, which includes the eight minimum elements that define a PWS, was evaluated as part of the DVs. The second set, which was agreed to by states and EPA, is a more extensive set of elements which includes system contact and locational data. Failure to report this information can result in withholding of state Public Water System Supervision (PWSS) grants.
- C ***Timeliness of state reporting of violations data.*** This analysis considered the percent of violations data that were reported to SDWIS/FED within the 90-day reporting period following the compliance period for each quarter, as specified in federal regulation.
- C ***Rejection error analysis.*** This analysis identified the most frequently occurring error conditions classified by error type and other subcategories which resulted in rejection of state submitted data by SDWIS/FED.
- C ***Potential non-reporting of violation data.*** This multi-year trends analysis of violation data indicated the number of states from whom no violations have been reported, by rule/rule group and violation type.

Analyses of DVs yielded the most complete and reliable estimates of SDWIS/FED data quality. The DV analysis was used to estimate overall data quality for inventory, violations, and enforcement actions data; and to assess completeness and accuracy for violations data. Data that states report to SDWIS/FED are but a small subset of all the data that states need to manage their drinking water programs and to make PWS compliance determinations. A DV compares facility, compliance and violations data in state files for each system to the data required to be reported to EPA based on Federal regulation and actually found in SDWIS/FED. The personnel (EPA and

contractor staff) conducting the DV review data submitted by PWSs, state files and data bases and SDWIS/FED, and compile results on **errors** (unreported, undetected, and incorrect violations) and **discrepancies** (wrong information) in the data as compared to the data in SDWIS/FED. States have several opportunities to respond to findings while DV personnel are on site, including providing additional clarifying information if available, as well as reviewing the DV draft report. The final DV results are compiled in a final report and in a data base.

EPA conducted 12-13 DVs of state drinking water facility information, violations, and enforcement actions each year from 1999 to 2001 (Table 2-1). Files for a total of 1,890 PWSs were evaluated, 40% of which were community water systems (Table 2-2). Subsequent DVs will be considered in the next triennial data quality evaluation period (2002-2004). The regulations addressed by the DV's conducted and the compliance period reviewed for each regulation are shown in Table 2-3. The period of review by rule was generally the two most recent scheduled monitoring periods per rule, per water system. For the Total Coliform Rule (TCR) and Surface Water Treatment Rule (SWTR), the most recent four quarters were evaluated.

The findings were subjected to a statistical analysis to determine if the 31 state DV results were representative of data quality at the national level. Audits are designed so as to be representative at the state level, except for states with decentralized offices. The DVs are designed to be representative of the quality of drinking water data throughout the state with at least an 80% confidence level and a 7.5% margin of error.

EPA calculated health-based violations data quality estimates, by state using CWS and NTNCWS data (more data points provides better precision, and these water system types had very similar results. TNCWSs did not). EPA then performed a Bayesian statistical analysis to mathematically model a curve that describes the data using a Beta distribution. The individual state results very closely fit a normal distribution, which indicates with a high degree of confidence that the results are representative nationally. A similar analysis performed for the first assessment found that dataset to be nationally-representative as well.

Table 2-1. States Subject to DV's from 1999-2001

Region	States	Region	States
1	MA, ME, NH	6	AR, LA, NM, TX
2	NJ	7	KS, MO, NE
3	VA	8	MT, ND, UT
4	FL, GA, KY, MS, NC, SC, TN	9	HI, NV
5	IL, IN, OH, WI	10	AK, ID, OR

Table 2-2. Number of Systems included in DV's by Type and Size

System Size	System Type			Total
	CWS	NTNCWS*	TNCWS*	
Very Small (500 or fewer)	414	467	562	1,443
Small (501-3,300)	214	77	14	305
Medium (3,301-10,000)	71	7	1	79
Large (10,001-100,000)	54	1	0	55
Very Large (>100,000)	8	0	0	8
Total	761	552	577	1,890

* NTNCWS = non-transient non-community water system (e.g., school with own source)
TNCWS = transient non-community water system (e.g., campground)

Table 2-3. Period of Compliance for Rules Reviewed During DV's

Rule	Compliance Period Reviewed
Total Coliform Rule (TCR)	Most recent four quarters in SDWIS/FED
Surface Water Treatment Rule (SWTR)	Most recent four quarters available in SDWIS/FED
Nitrates	Most recent three calendar years
Nitrites	1996-1998
IOCs	1996-1998; back to 1990 if grandfathered
VOCs	1996-1998; back to 1988 if grandfathered
SOCs	1996-1998; back to 1990 if grandfathered
Radionuclides	Most recent two samples
Total Trihalomethanes	Most recent four quarters available in SDWIS/FED
Enforcement	Time period applicable to related violation

2.4 Assessment of Data Quality

Data quality, overall, is the percentage of data without any discrepancies or errors (Table 2-4). The analysis calculates data quality for inventory, violations and enforcement action data. For violations data, the DV-based data quality estimates were further broken into components of completeness and accuracy. Additional analyses were used to calculate timeliness for violations

data and completeness for a larger set of inventory data.

Table 2-4. Errors and Discrepancies

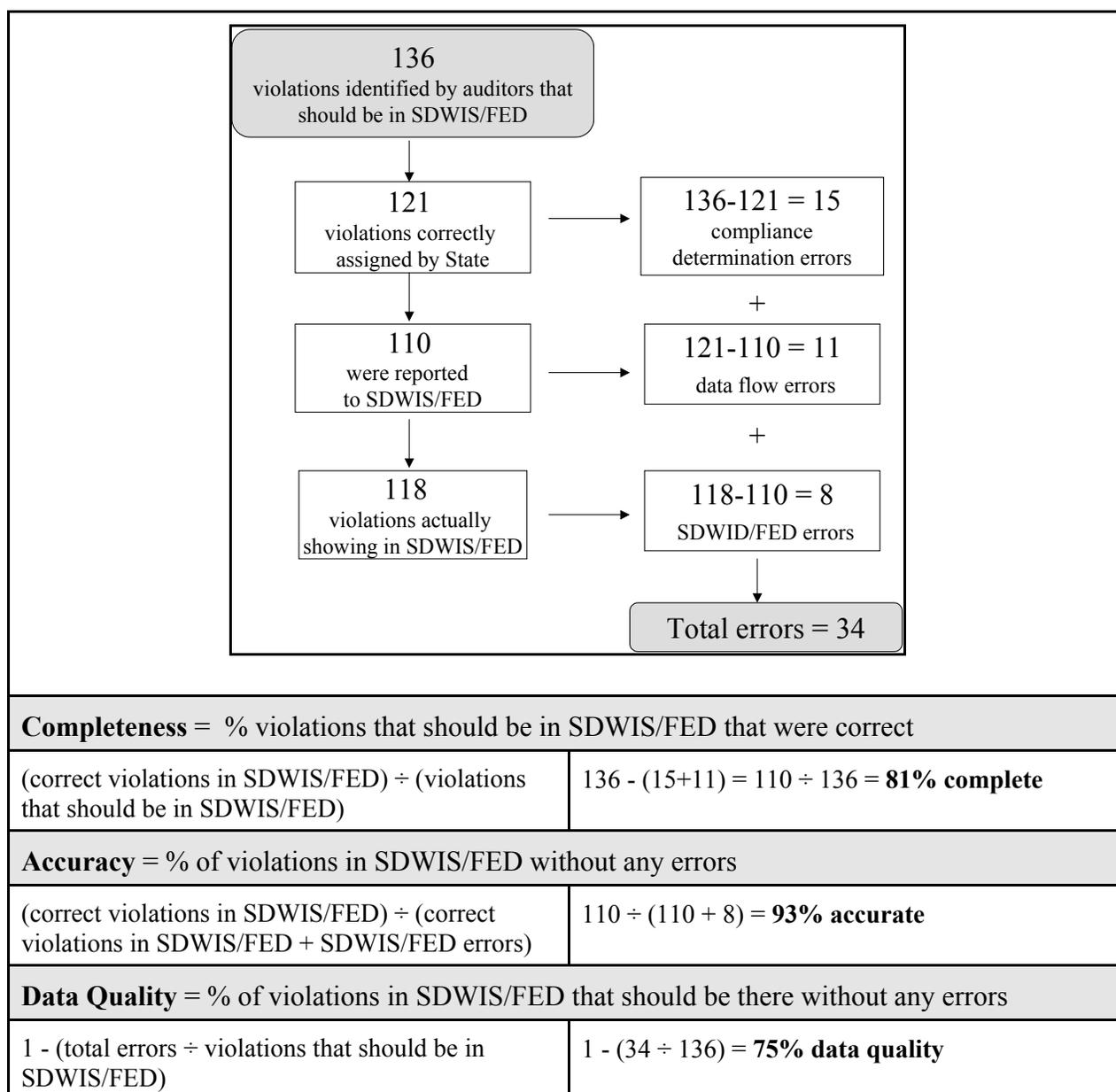
Errors. An error is a mistake that the state makes. The auditor identifies the mistake and verifies what the correct answer should be. Errors were assessed in evaluating violations data. There are three types of errors:

- C **compliance determination errors** are made when a state fails to cite a violation that should have been assessed. Errors that occur from assigning a violation where there was none are categorized as “errors in SDWIS/FED.”
- C **data flow errors** are made when the state fails to report a violation (that it has correctly identified) to SDWIS/FED.
- C **errors in SDWIS/FED** include typographical errors. They also include violations that should not be in SDWIS/FED, either from assigning a violation where there was none or failing to remove a rescinded violation.

Discrepancies. A discrepancy simply reflects a difference between data in state files and SDWIS/FED where the auditor does not attempt to verify the correct information. For example, if a state showed one address for a PWS and SDWIS/FED showed a different address, this would be classified as a discrepancy. The auditor would not attempt to identify the true address for the PWS. Discrepancies were assessed in evaluating inventory and enforcement action data.

The method for calculating data quality was similar to that used in the previous analysis of 1996-1998 data verification results so that a comparison could be made for the two time periods. For data quality estimates of inventory and enforcement data, a **discrepancy rate** is calculated by taking the number of data records that do not match (e.g., facilities information, or enforcement actions) divided by the number of data records in SDWIS/FED. For violations, a **data error rate** is calculated by taking the number of data records incorrectly reported to SDWIS/FED divided by the number of data records which federal regulations indicate should be reported to SDWIS/FED. The overall **data quality estimate** (DQE) is calculated as one (1) minus the data error or discrepancy rate, expressed as a percentage. An example of calculations used for a violation data quality assessment is presented in Table 2-5.

Table 2-5. Example Calculation for Violation Data Quality



Note: a negative value can occur when the number of errors exceeds the number of violations which federal regulations indicate should have been reported in SDWIS/FED. For example: of 15 violations that federal regulations indicate should have been reported to SDWIS/FED, 1 was reported to SDWIS/FED but was in error [not in the state data base], and none of the 15 that should have been in SDWIS/FED [compliance determination errors] were reported to SDWIS/FED. Fifteen (15) expected violations minus 16 errors equals -1 divided by the 15 violations which federal regulations indicate should be in SDWIS/FED equals = -7% .

The data quality range in Table 2-6 was designed as an interim ranking scheme in order to simplify data quality characterization. EPA intends to develop data quality objectives for each major type of data during the 2004-2005 time frame. The data quality objectives will be based on

specific programmatic uses of the data such as PWSS grant calculation, geospatial applications, GPRA reporting, and compliance rates at the rule level. These data quality objectives will be developed by EPA with input from its stakeholders. The high range (91 - 100%) used in this report, reflects the level of confidence achieved in data verifications used to calculate the data quality estimates in this assessment. The medium and low data quality ranges (71 - 90%, and 0 - 70% respectively) are arbitrary, but they reflect commonly used ranges. Once the data quality objectives are established, more specific data quality ranges will be set and used in future documentation and communication of data quality.

Table 2-6. Data Quality_Range Description

Low quality :	0 to 70%
Moderate quality	71 to 90%
High quality	91 to 100%

3.0 Results and Findings

3.1 Overview

The analysis of DV audits found that SDWIS/FED data quality of inventory and enforcement data for all system types is consistent with that from the first assessment. Violations data, however, had a noticeably higher quality overall in the second assessment (Table 3-1). A more complete explanation of these results follows, first for inventory, then for violations and finally for enforcement data quality.

Table 3-1. Overview of Data Quality Estimates

Data Quality Estimate	2003	2000
Inventory	95%	96%
Enforcement Actions	80%	72%
Health-based Standards Violations	65%	40%
<i>TCR MCL</i>	75%	68%
<i>Other MCL</i>	45%	15%
<i>SWTR TT</i>	56%	7%
Monitoring and Reporting Violations	23%	9%

Data verification reports continue to document and confirm that states which employ the following management policies and business practices typically have higher data quality than states which do not use these practices:

- C routine, meaningful communication at all levels;
- C annual state notification to PWS of monitoring schedules;
- C automated monitoring compliance determination systems; and
- C electronic data transmission between laboratories and the state as well as between the state and SDWIS/FED.

3.2 Results for Inventory Data

The SDWIS/FED data quality of the eight inventory (water system identification) parameters assessed is estimated to be 95% (Table 3-2). This is about the same as the 2000 inventory data quality estimate of 96%. The eight inventory parameters are: 1) water system identification number, 2) system activity status, 3) water system type, 4) primary water source type, 5) population served, 6) number of service connections, 7) address, and 8) name of water system. Inventory data quality by parameter is displayed in Table 3-3. The lowest data quality

estimates were associated with the parameters which tend to change most frequently - population served and the number of service connections.

Table 3-2. Calculation of Data Quality Estimates for Inventory Data

<u>Factor</u>	<u>Value</u>	<u>Explanation</u>
Number of data points	15,120	1,890 systems reviewed times 8 Inventory parameters checked
Discrepancies:		
Number	788	The # of instances where the DV verified the Inventory parameter in SDWIS/FED was different
Percent	5%	Inventory Data Discrepancy Rate
SDWIS/FED Inventory Data Quality	95%	SDWIS/FED data quality = 1 - discrepancy rate = % of data without discrepancies

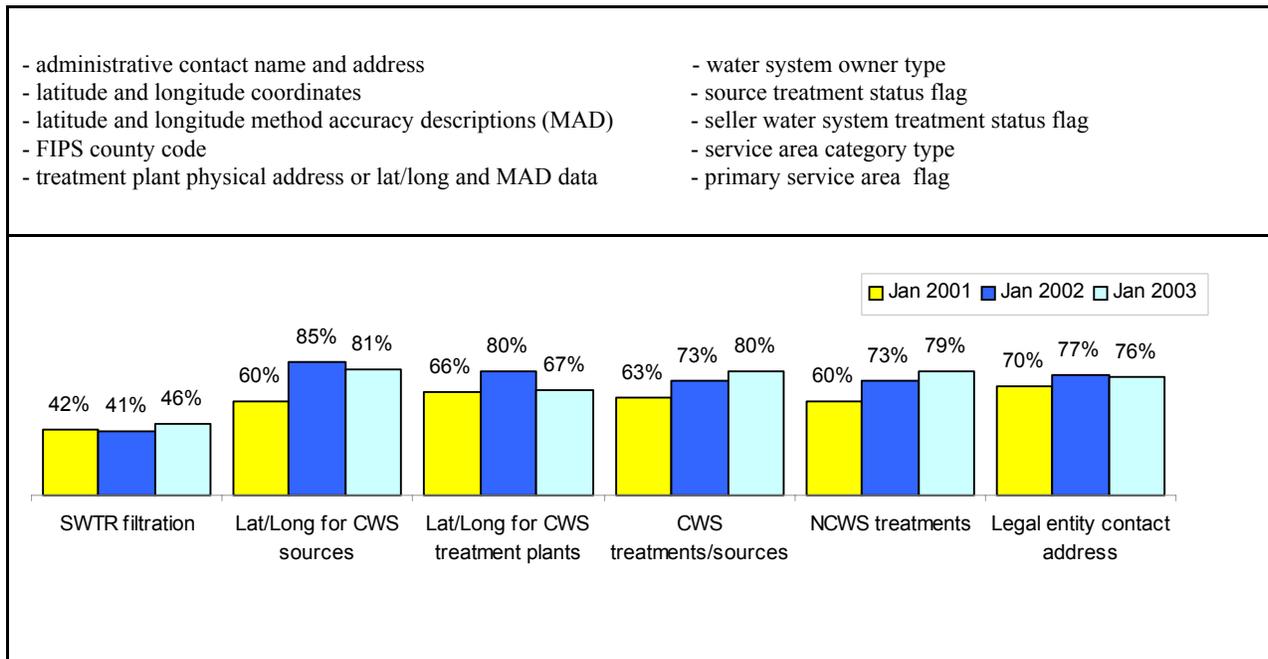
Table 3-3. SDWIS/FED Inventory Data Quality Estimate (DQE) by Parameter

Parameter	DQ - 2003 Assessment	DQ - 2000 Assessment
PWS ID	100%	100%
System status (active or inactive)	95%	97%
Water system type	98%	97%
Primary source	98%	98%
Population served	88%	91%
# service connections	89%	92%
Address	91%	95%
PWS name	99%	98%
Overall Inventory DQE	95%	96%

A second set of programmatically required inventory data, commonly referred to as “grant withholding inventory data,” was also analyzed. These data have been identified by the program office, or the Agency under its data standards policies and other business needs, as being necessary to characterize the water system, its sources, treatment plants and applied treatments, and location. These data were analyzed for completeness only and were not included in the overall inventory data quality estimate in this report. Table 3-4 lists these elements by individual element and by record groups, when the entire group of elements is required (e.g., administrative contact name and address, latitude and longitude coordinates). Depending on the state’s chosen method of reporting the required data, a state could report up to 27 individual data elements.

The information was only evaluated for completeness. A data quality estimate was not determined for this second set of inventory data since no attempt was made to verify the information held in SDWIS/FED. Changes in completeness between 2001 and 2003 for several important system characteristics derived from inventory data are shown Table 3-4. The percent completeness for these elements varies widely. For example, data on filtration status was only 46% complete in 2003, while the latitude/longitude data for CWS sources was 81% complete. While completeness is slowly improving overall, an effort needs to be made to improve upon the availability of several elements.

Table 3-4. Programmatically Required Inventory Elements or Groups of Elements



3.3 Results for Violations Data

Based on the analysis of DV findings, the SDWIS/FED data quality estimate for all violations data (i.e, health-based standards and M/R violations) increased from 11 % to 26 % from the first to the second assessment. The accuracy of violations data in SDWIS/FED continues to be very good. However, these data continue to be highly incomplete, particularly for monitoring and reporting violations. The following table (Table 3-5) presents the violation data quality estimates for accuracy, completeness, and overall quality, grouped by major violation types.

The first row of Table 3-5 shows the percent of systems having any violations. While these are not used as part of the calculation of the SDWIS/FED data quality estimate, they provide perspective on the number of systems with violations. The information in the remainder of the table is used to calculate the data quality estimate. Analyses conducted in the first assessment made it clear that the reasons for errors needed to be defined, labeled, and identified to ensure consistent interpretation and analysis, thus the table shows the incidence of different error types.

Table 3-5. Data Quality Estimates (DQE) by Violation Type

Data Quality Accounting Category	Violation Type				
	TCR MCL	OTHER MCL	TOTAL MCL	SWTR TT	Monitoring and Reporting
% OF SYSTEMS W/VIOLATIONS	5.5%	1.2%	6.6%	6.4%	49.5%
# OF VIOLATIONS	136	38	174	59	3,021
<i># compliance determination errors</i>	15	15	30	23	2,018
<i># data flow errors</i>	11	5	16	3	201
<i># errors in SDWIS/FED</i>	8	1	9	0	103
# ERRORS	34	21	55	26	2,322
% COMPLETENESS	81%	47%	74%	56%	27%
% ACCURACY	93%	95%	93%	100%	89%
2003 SDWIS /FED DQE	75%	45%	68%	56%	23%

For an example of how completeness, accuracy and data quality are calculated, see Table 2-5, which uses the values for the TCR MCL column in the calculations. According to these estimates, roughly two-thirds (68%) of all MCL violations were reported completely and accurately compared to 54% in the first assessment. Table 3-6 compares results from both assessments.

Table 3-6. Violation Data Quality Estimate Comparison between 2000 and 2003 DQEs

Data Quality Accounting Category	RPT YR	Violation Type				
		TCR MCL	TOTAL OTHER MCL	TOTAL MCL	SWTR TT	Monitoring and Reporting
% COMPLETENESS	2003	81%	47%	74%	56%	27%
	2000	68%	19%	55%	11%	10%
% ACCURACY	2003	93%	95%	93%	100%	89%
	2000	99%	79%	97%	67%	95%
SDWIS/FED DATA QUALITY ESTIMATE	2003	75%	45%	68%	56%	23%
	2000	68%	15%	54%	7%	9%

These data show an increase in overall data quality for all violation types. Most violation types show increases in both completeness and accuracy (although a strict comparison is difficult because the first assessment was not able to precisely estimate completeness and accuracy). Accuracy levels are in the high data quality range and completeness falls in the low data quality range. Of note, the TCR MCL completeness and the overall TCR MCL data quality estimate has climbed into the Moderate range. Unfortunately, the Other MCL, SWTR TT, and the overall violation data quality estimates still fall in the low to moderate data quality range.

Total Coliform Rule. At 75%, Total Coliform Rule (TCR) data has the highest SDWIS/FED data quality. Eighty-one percent (81%) of the Maximum Contaminant Level (MCL) violations which federal regulations indicate should have been reported to SDWIS/FED were actually in SDWIS/FED and, of those violations listed in SDWIS/FED, 93% are estimated to be accurate. In the first assessment, TCR MCL data quality was 68%.

Chemical/Radionuclides Rules Maximum Contaminant Levels (MCL). The overall SDWIS/FED data quality for chemical and radionuclides MCLs is estimated to be 45% (up from 15%). Although the data that are recorded in SDWIS/FED have high accuracy (95%), their completeness is low (47%), which affects the overall data quality. Values for individual rules were not calculated because there were an insufficient number of data points.

Surface Water (Microbial) Treatment Rule. Significant improvement has occurred in SDWIS/FED data quality of Surface Water Treatment Rule (SWTR) Treatment Technique (TT) violations, estimated to be 56% (up from 7%). The accuracy of information recorded in SDWIS/FED is 100%.

Lead and Copper Rule. Lead and Copper reporting requirements were not included in the first data reliability assessment due to questions of regulatory interpretation which had not been resolved at the time the assessment was released. Lead and Copper results are likewise not included in this analysis. Lead and Copper will be evaluated in a separate assessment in FY 2004 and data quality estimates calculated.

Monitoring and Reporting. The SDWIS/FED data quality for all Monitoring and Reporting (M/R) violations is 23% (up from 9%). Overall, the data quality is negatively impacted by a low rate of completeness (27%). The primary driver of poor completeness is a high number of compliance determination errors, which make up 86% of the total errors. M/R data quality is highest for the TCR rule (41%). M/R data quality for the SWTR and Chemical rules continues to be poor.

The majority of errors cited in the DV's were compliance determination errors, where violations had not been identified and recorded by states as violations. Incorrect compliance determinations were a factor in about half of errors for MCL violations. For M/R violations and SWTR TT violations, more than 85% of errors were the result of incorrect compliance determinations. For M/R violations, almost half of those compliance determination errors were due to situations where a state failed to assess a violation when a system did not sample and the state could not document why it had not assessed a violation. All other compliance determination errors characterize individual instances where the state inconsistently applied the regulatory requirement. Table 3-7 presents the various reasons for errors and their occurrence.

Nine percent of the errors represented data flow errors between state files and SDWIS/FED. In the vast majority of these cases, the data were in state databases but were not reported to SDWIS/FED. These represent data transfer problems where the data were not successfully being accepted by SDWIS/FED (rejected data) or violations that the state never reported to EPA. A separate evaluation of the reasons for rejected data is described in section 4.4 of this report.

Over-reporting is defined as a violation that is in SDWIS/FED but not in the state data base. This type of error was classified as an error in SDWIS/FED. This second assessment found very little evidence of over-reporting of violations in SDWIS/FED. There was no statistical difference between the 2% value calculated for the second assessment and the less than 1% value calculated for the first.

Table 3-7. Error Description and Occurrence

Category	Number of Errors/Percent of Errors by Violation Type				
	Type Code*	TCR MCL	Other MCL	SWTR TT	M/R
No sample data; no violation assigned	CD			12 / 46%	1116 / 48%
State policy not approved in writing by Region	CD				258 / 11%
Violation in state database, not reported to SDWIS/FED	DF	11 / 32%	5 / 24%	3 / 12%	199 / 8%
Insufficient number of samples taken	CD	1 / 2%			177 / 8%
Insufficient quarterly monitoring conducted after a detect	CD				163 / 7%
Failure to conduct quarterly sampling for new systems	CD				88 / 4%
Violation assigned by State and not confirmed by DV team	EF	5 / 15%			45 / 2%
Chem samples not taken according to schedule	CD				42 / 2%
Incorrect sampling/analytical procedure	CD				42 / 2%
No sample because system incorrectly classified	CD				39 / 2%
Rescinded violation not removed from State database and/or SDWIS/FED	EF	1 / 2%			37 / 2%
Insufficient quarterly monitoring conducted after Chem MCL	CD				36 / 2%
No speciation of lab results	CD				24 / 1%
TYPO: correct compliance determination but incorrect data entered	EF	2 / 6%	1 / 5%		23 / 1%
Sample missing one or more analytes	CD				16 / 1%
Incorrect information entered into database, e.g., violation type 23 reported should be 22	EF		3 / 14%		12 / 1%
Incorrect MCL or failure to assign violation	CD	14 / 41%	12 / 57%		3 / <1%
Incorrect treatment technique violation determination or failure to assign violation	CD			11 / 42%	2 / <1%
Totals		34	21	26	2,322

* CD = compliance determination, DF = data flow, EF= error in SDWIS/FED

The violations data were further evaluated to determine if there were differences in quality depending on the type of PWS (Tables 3-8). Non-transient noncommunity water systems (NTNCWS) had the highest data quality for violations of MCL rules. Transient noncommunity water systems (TNCWS) had the lowest violation data quality for MCL rules and the highest quality in overall M/R violation data categories. Data for NTNCWS and TNCWS treatment technique violations had an insufficient number of data points to determine reliable quality estimates. Data quality improved between the 2000 and 2003 assessments, with the exception of quality for NTNCWS in TCR MCL violations, which decreased slightly. There were an insufficient number of data points to determine whether there were meaningful differences in data quality based on system size.

Table 3-8. Data Quality Estimates for Violations by PWS Type and Violation Type

Type of PWS	Year	TCR MCL	SWTR TT	Monitoring & Reporting
CWS	2003	78%	55%	18%
	2000	69%	9%	9%
NTNCWS	2003	81%	n/a	20%
	2000	67%	11%	7%
TNCWS	2003	65%	n/a	39%
	2000	68%	0%	14%
2003 Overall		75%	56%	23%
2000 Overall		68%	7%	9%

n/a - insufficient data points to calculate a percentage.

3.4 Results for Enforcement Data

The overall data quality estimate for enforcement data has increased from 72% in the first assessment to 82%. Table 3-9 presents results for enforcement data quality. There were significantly fewer enforcement actions observed for the water systems included in this assessment's DVs than were observed in the first assessment. While the DVs did not closely audit state enforcement programs, the reviews did not document an obvious lack of state enforcement. It also did not appear that data quality problems were the cause for the recording of fewer enforcement actions. This issue is being referred for further program evaluation separate from this assessment.

Table 3-9: Enforcement Data Quality by PWS Type

Category	PWS Type			
	CWS	NTNCWS	TNCWS	TOTAL
2003				
# Enforcement Actions	99	83	48	230
# Total Discrepancies	14	16	12	42
SDWIS/FED DQE	86%	81%	75%	82%
2000				
# Enforcement Actions	505	305	222	1,032
# Total Discrepancies	121	92	74	287
SDWIS/FED DQE	76%	70%	67%	72%

4.0 Additional Findings

4.1 Evaluation of Large Systems

For this second assessment, an attempt was made to conduct a special analysis of large PWS violation data to better estimate data quality for larger systems which serve most of the population. The Large System Data Verification Analysis focused on a random selection of 30 CWSs serving populations of 50,000 or more persons that had not been selected for the scheduled state DVs. However, because EPA regions did not record the DV results in a similar manner, the results could not be included with the data for the other 1,890 randomly selected systems. While the results were not included in this data reliability analysis, they did serve as a check on the other results for the 63 large systems that were already part of the analysis.

Monitoring and reporting compliance determination errors accounted for all violation errors. No health-based violation compliance determination errors were identified. Discrepancies in population inventory data were also observed, but were not so great as to indicate that any one system's monitoring requirements should be changed based on recorded population differences. As noted, because regional reporting was not sufficiently robust for this analysis, the statistical validity of these findings will be examined in future DVs. This separate analysis for large systems did, however, serve to guide a revision of the DV protocol for the future to include more large systems.

4.2 Potential Non-Reporting for Rules

As described in the report for the first assessment, EPA has developed a tool to identify potential non-reporting for certain rules in each state by water system type. The tool tracks the number of violations reported in each state over a period of several years across system type and rule. TCR is the only rule that has consistently high levels of reporting, which is not surprising given the scope of the rule's coverage and its importance. Table 4-1 shows the number of states that have not reported any MCL or TT violations since 1997 for several rules. It is important to note that while it is difficult to draw conclusions from this analysis, the tool only identifies potential non-reporting that should be evaluated further.

Table 4-1. Number of States Not Reporting Any Violations from 1997-2002

Regulation	# of states
TCR MCL	0
Chemical MCL	3
Radionuclides MCL	19
SWTR TT	3
Lead and Copper	14

4.3 Timeliness of Reporting

This analysis looked at the timeliness of violations based on the compliance period end date. Violations are due to be reported by the end of the following quarter after a state becomes aware of a violation or the compliance period end date. The analysis looked at the number of violations which existed in the frozen data base immediately following the reporting deadline for the quarter being evaluated and the number of violations which were eventually reported for that same quarter by looking at the number of violations in the database several reporting periods later.

For example, violations for the 4th quarter of FY 2000 were due to be reported no later than September 30, 2000. The database for that quarter was frozen in January 2001. These represent the violations reported on time. To assess the timeliness of FY 2000 4th quarter violation reporting, EPA queried the database for the period ending September 30, 2001 (frozen in January 2002), to determine how many violations for FY 2000 were eventually reported. This number represents the baseline number for violations reported. The timeliness is calculated by dividing the violations reported on time by the baseline number for violations.

Table 4-2. Violation Reporting Timeliness to SDWIS/FED by Violation Type

Fiscal Year	1998	1999	2000	2001
Number of Violations Reported on Time				
TCR MCL	9,732	9,550	8,232	8,657
Other MCL	786	687	767	640
SWTR TT	1,886	1,782	1,586	1,630
M/R	41,560	45,891	60,754	61,760
Total	53,964	57,910	71,339	72,687
Number of Violations Reported for Baseline				
TCR MCL	12,804	11,652	11,532	11,027
Other MCL	1,274	1,115	1,257	1,202
SWTR TT	2,765	2,264	2,178	2,083
M/R	114,487	116,724	95,888	111,532
Total	131,330	131,755	110,855	125,844
Percent Timeliness				
TCR MCL	76%	82%	71%	79%
Other MCL	62%	62%	61%	53%
SWTR TT	68%	79%	73%	78%
M/R	36%	39%	63%	55%
Total	41%	44%	64%	58%

Many states are not meeting the 90-day deadline for reporting violations. In 2001 only 58% of violations eventually reported were reported on time. The timeliness in which health-based violations are reported has been steady and is relatively high. Although the data is not shown here, timeliness is similar across water system types and sizes. Timeliness for reporting of monitoring and reporting violations has improved, but is still relatively low, particularly for large and very large systems.

4.4 Rejection Error Analysis

The rejection error analysis indicates states have resolved many information system and data entry issues which were causing repetitive rejections observed in the first assessment. Errors were observed for all types of data. Most of the rejected data was inventory data (81%) followed by violations and enforcement data (18%).

The rejection error analysis found that 90% of the inventory, violations and enforcement data error types incurred were for data entry errors. The remainder were from:

- C transfer file format errors (5% of the error types which accounts for 3% of the rejected data)
- C SDWIS/FED software limitations regarding the number of records which can be processed in a single PWS record submission combined with the one SDWIS/FED “bug” (3% of the error types which accounts for less than 1% of the rejected data)
- C informational messages advising that the data were not rejected, but may not have been processed as expected (2% of the error types which accounts for less than 1% of the rejected data)
- C informational messages advising that the data were not rejected, but may not have been processed as expected (2% of the error types which accounts for less than 1% of the rejected data).

Most error conditions previously thought to be the result of software programming problems (“bugs”) in SDWIS/FED have been addressed by EPA. Many of the current error conditions are the result of new or modified requirements such as implementation of the revised inventory reporting which requires the reporting of all sources and treatment plants, locational data, and source treatment status. Nationally, and for most states, rejected data represents a small part of the data quality problem. The majority of the error conditions continue to be a result of state information systems lacking adequate data quality checks and quality assurance routines.

4.5 Relationship between Use of SDWIS/STATE and Incidence of Errors and Discrepancies

Many actions implemented by EPA since the first assessment focused on improvements to the SDWIS information system and its supporting tools, including the EPA/state designed data base application used by states in support of their drinking water program (referred to as SDWIS/STATE).

An informal assessment conducted in January 2003 compared the results of the first data reliability assessment to the results of the second assessment for several states who had converted from state legacy systems to SDWIS/STATE after the first assessment. It is important to note that it is currently difficult to conduct a robust analysis of the effect of SDWIS/STATE on data quality because only a few states with DVs had SDWIS/STATE at the time of the DV. Additionally, if a state did have SDWIS/STATE, it may not have been completely implementing it at the time of the DV. As more states fully utilize the features of SDWIS/STATE, it will become easier to evaluate its full effect on data quality.

Results indicate that SDWIS/STATE did not eliminate compliance determination or data flow errors (rejected data), nor necessarily improve the timeliness of violation reporting. However, findings did show that SDWIS/STATE users experienced a decrease in data entry errors. Because SDWIS/STATE has been shown to have a positive influence on some aspects of data quality and because EPA desires to support state programs, EPA has reaffirmed its commitment to continue support for SDWIS/STATE. The support includes a future web based version of the application and further development and integration of analysis tools for data migration and errors correction. Data submission, data rejection, system processes, and data access and retrieval have and are being evaluated, enhanced, streamlined, and documented for the SDWIS modernization project.

4.6 State Compliance Determinations and Implementation Issues

As noted earlier, incorrect compliance determinations by states are the principal factor affecting the quality of violations data in SDWIS/FED. These represent situations in which Federal regulations implemented through state primacy (delegation) agreements would indicate that a violation should have been assessed, however state records (files, databases) do not indicate that a violation was issued. Compliance determination errors were found across the range of drinking water regulations. While violations are a small percent of all determinations made by states (estimated at less than one percent of all determinations), they are important in that they reflect a divergence from public health protection practices as reflected in the regulations.

As in the first assessment, the most frequent reason for an error was “no sample data, no violation assigned” (48%) (see Table 3-7). These errors identify individual situations where the requirement was not met by the water system, no violation assigned by the state, and neither a record found in state files, nor a state response to the error identified in the DV report.

A change had been made to the DV protocol to accept state implementation policies in lieu of federal regulatory requirements when approved in writing by EPA. To capture circumstances where the state implementation policies did not agree with federal regulations or lacked formal EPA approval, the unique error reason “state policy not approved in writing by region” was captured in subsequent DVs. Eight states were found to have these errors, all of which fell under M/R violations (11% of all M/R errors). EPA has reviewed each state’s circumstances and determined that these practices are neither supported by regulation, nor by EPA written approval of state policy “flexibility.”

This finding points out the utility of DV analyses in identifying implementation issues in states. Although not based on the quantified DV results, Table 4-4 identifies some of the more prominent implementation issues found during data verification audits. Many of these areas are captured as specific reasons such as “not requiring quarterly monitoring for new systems.” When the problem affects all systems, the specific reason is captured in the narrative of the data verification. In all cases, these issues result in incorrect violations or violations not being issued.

Table 4-4. Example of Implementation Issues Identified During Data Verifications

Rule	Reason/Area
Consumer Confidence Rule	Late Consumer Confidence Reports - no violations issued
Public Notification	Failure to track and/or designate PN violations
TCR	Not conducting sanitary surveys within TCR schedule for systems taking fewer than 5 samples per month. - no violation
	Failure to take 5 samples in month following month of positive.
	Failure to assess a reporting violation when sample results are received greater than 10 days late
	PWS > 4900 pop -takes samples on same day - instead of throughout the period
	Failure to report multiple violations in same month
	Seasonal systems not being required to monitor every monitoring period unless open the entire monitoring period
SWTR	SWTR Monthly Operating Reports are not completed properly - (not recording when plant is offline and not sampling every 4hrs) - violations not issued for sampling failures
Nitrate /Nitrite	Not requiring annual monitoring
	Incomplete monitoring for reliably and consistently below the MCL as required
Radionuclides	Not requiring quarterly monitoring for new systems
	Not speciating/monitoring for Radium 226/228 after Gross Alpha exceeds 5pCi/l
	Not monitoring on 4 year schedule
Chemical Rules	Various implementations of waiver program not in conformance with requirements resulting in required monitoring not being conducted
	New systems not required to monitor 4 consecutive quarterly samples for VOC and IOC before going to reduced monitoring, and only required to take 1 SOC sample instead of 4 quarterly before going to reduced
	Chemical detected - no confirmation or quarterly monitoring required

Rule	Reason/Area
	Systems using incorrect sample locations - no violations issued
	State collects chemical samples for PWS, but does not issue violations when sampling not conducted
	State certified labs not being required to meet published MDLs
	Less stringent monitoring requirements allowed
Lead and Copper	No violations issued for corrosion control treatment steps following 90 th percentile exceedance for small systems
	Not monitoring in summer months - no violations designated
	Late initial implementation and subsequent violation tracking issues
	Early implementation of accelerated monitoring
	Failure to take 2 consecutive rounds for compliance - allowed systems to reduce monitoring
	Water Quality Parameters (WQP) and/or Source WQP not monitored
	Unauthorized/incorrect number of samples, replacement of sample sites, or incorrect calc of 90% percentile - did not require 5 samples for PWS with less than 5 sites, and did not include original result when sample invalidation applied
	State Primacy for LCR did not designate alternate monitoring periods for reduced monitoring, state allowing alternate periods

While more than one state may share issues in a similar implementation area, the issues tend to be somewhat specific to each state and require state specific attention for resolution through existing program management activities. For example, the implementation issue “Failure to require 5 samples in month following month of TCR positive” had been identified in several states. One state’s policy was to do an on site inspection, but documentation of the site visits were not always found; one state had one district which was allowing systems on quarterly monitoring to sample the next quarter; another state had a regulatory/guidance publishing error which allowed systems to increase samples to 5 in the next compliance period; and one state allowed less than 5 samples be taken. The other 3 states did not require systems to increase to 5 samples. Resolving these issues must be addressed on a state by state basis.

The plan presented in the next section emphasizes that EPA regional offices should utilize existing evaluation, coordination and planning processes to identify issues in federal regulation interpretation and application for resolution. This can be achieved by EPA Regional Offices in following up on state data verifications and in developing annual plans with states. The plan also provides focus on Quality Assurance/Control planning, timeliness of reporting (which also affects completeness of data), examination of other circumstances for non-reporting of violations, and electronic data transfer from laboratories and systems to states to minimize the under-determination of monitoring and reporting violations.

5.0 Recommendations For Improving SDWIS/FED Data Quality

The recommendations from this analysis address many areas impacting data quality including: compliance determinations by states, data quality analysis, implementation of the Office of Ground Water and Drinking Water's Information Strategic Plan¹, state quality assurance, state automated tracking and scheduling, timeliness of state violation reporting, and ensuring that data management concerns are considered in rule development.

Recommendations to further improve the quality of SDWIS/FED data were developed through a collaborative process with states. EPA relied on a State/EPA Data Reliability Workgroup (with members from the Association of State Drinking Water Administrators) to develop and refine the recommendations based on the analysis presented here. Meetings and conference calls were held during the spring and summer of 2003 and a draft report was provided to states and EPA regions for review. Additionally, EPA requested review by the American Water Works Association, Association of Metropolitan Water Agencies, National Association of Water Companies and the Natural Resources Defense Council, who had participated in the first assessment. EPA also met with representatives of the Association of State Drinking Water Administrators to address their comments on state reporting and plans affecting state processes. This plan reflects the input of this development and review process.

The recommendations are displayed in the following Data Reliability Improvement Recommendations/Plan Matrix which constitutes EPA's plan for improving data quality working with state over the next three years and includes the activity, the responsible party(s), the improvement focus area, the Quality Assurance area (e.g., Assess, Control, Assure) and a description of the anticipated data reliability benefit. These recommendations and other planned activities are also included in EPA's Drinking Water Quality Assurance Plan organized by Quality Assurance function.

Recommendations have been divided among five primary areas. Because some actions were identified that would benefit more than one area, the table includes columns that show additional areas addressed by a specific action. EPA will work with states on determining the priority of these actions. The specific areas of focus and an example of a significant quality improvement activity are listed below.

- C **Compliance Determination.** EPA and states should develop state specific compliance determination improvement and quality improvement plans necessary to remedy the major problem areas, working through established planning and implementation processes.
- C **Data Reliability.** EPA and states should continue to conduct and improve data quality analysis in accordance with the data reliability action plan.
- C **SDWIS Modernization.** EPA should continue SDWIS modernization and evaluate its effect on both SDWIS/STATE and non-SDWIS/STATE states.

¹The Information Strategic Plan can be found at www.epa.gov/safewater/data/informationstrategy.html.

- C **Monitoring and Reporting.** EPA should encourage states to develop an automated monitoring requirements and sampling schedule tracking system and to adopt electronic reporting processes for data from PWSs and laboratories.
- C **Violation Timeliness.** EPA and states should evaluate why violation reporting timeliness is low and not improving.
- C **Violation Non-Reporting.** EPA and states should conduct annual evaluations of all instances of potential violation non-reporting and take steps to improve reporting.

2003 - Data Reliability Improvement Recommendations/Plan Matrix

Action #	RESPON-SIBLE PARTY(S)			RECOMMENDATIONS Action/Activity	FOCUS AREA							QA Area	ANTICIPATED DATA RELIABILITY/QUALITY BENEFIT(S)
	EPA	STATE	UTILITY ORGS.		Compliance Determination	Data Reliability	SDWIS Modernization	Monitoring and Reporting	Violation-Reporting Timeliness	Non-Reporting	Rule Development - Data Mgt Concerns	Area	Description
				COMPLIANCE DETERMINATION									
1	X	X		EPA and states should develop state specific compliance determination improvement and quality improvement plans necessary to remedy the major problem areas including:	X	X						Assure	Defines data quality improvement goal and direction
1.1.a	X	X		Develop memorandum of agreements or other regular documentation (e.g., annual work plans) between states and EPA detailing quality improvement plan and schedule focusing on documented differences in state and EPA interpretation of regulations	X	X		X	X			Assess	Defines responsibility between states and EPA to ensure QA is incorporated throughout the program
1.1.b	X	X		Correct all identified discrepancies from data verifications		X						Control	Corrects data reported for water systems in SDWIS/FED
1.1.c	X	X		Conduct rule compliance determination training, as needed	X							Assess	Promotes accurate and consistent application of regulations
1.1.d	X	X		Revise standard operating procedures and/or programs such as waiver programs to correct or clarify implementation procedures which do not agree with federal regulations	X	X			X			Assess	Promotes accurate and consistent application of regulations
1.1.e	X	X		Revise state regulations to address less stringent implementation language, as appropriate	X			X	X			Assess	Aligns state and federal regulations for consistent compliance determinations
1.2	X			Monitor state rule implementation and improvement action plans by EPA regions		X						Control	Incorporates incorporate QA in routine program management of states and EPA Region
1.3	X	X	X	Evaluate feasibility of creating and maintaining a clearing house and tracking system, within the SDWIS system, to document and track resolution of disputed data by water systems	X	X						Control	Provides tracking mechanism to ensure disputed data are verified and corrected as appropriate.

Action #	RESPONSIBLE PARTY(S)			RECOMMENDATIONS Action/Activity	FOCUS AREA							QA Area	ANTICIPATED DATA RELIABILITY/QUALITY BENEFIT(S)
	EPA	STATE	UTILITY ORGS.		Compliance Determination	Data Reliability	SDWIS Modernization	Monitoring and Reporting	Violation-Reporting Timeliness	Non-Reporting	Rule Development - Data Mgt Concerns	Area	Description
				DATA RELIABILITY									
2	X	X		EPA and states should continue to conduct and improve data quality analysis in accordance with the Data Reliability Action Plan including:		X						Assess	Provides basis for regular communication on data quality.
2.1.a	X	X		Conduct triennial data reliability analysis and report results		X						Assess	Documents data reliability status, improvements and impacting factors.
2.1.b	X	X	X	Incorporate correct compliance decisions in review/calculation of data quality		X						Assess	Provides program decision making context for SDWIS/FED data quality estimates.
2.1.c.1	X			Increase the number of DVs conducted each year to address each state every 3 years		X						Assure	Ensures all states are included in triennial national data quality assessment.
2.1.c.2	X			Include contaminant occurrence data in program data verifications		X						Assure	Provides national basis for planning future DV audits
2.1.d	X	X		Formalize DV followup by EPA regions working with states to address identified deficiencies		X						Control	Provides basis for future data quality improvements
2.2	X	X		Encourage/require states to implement and maintain a quality assurance management plan as required by Executive Order 5360.1, A.1-2 of grantees reporting environmental or secondary data. The plan should include:	X	X		X	X	X		Assure	Defines QA process required to be in place
2.2.a		X		Delineate organizational responsibilities (see 1.1.a)		X						Assure	Defines clear QA responsibility
2.2.b		X		Define standard quality control and quality assurance operating procedures		X						Assure/Control	Establishes specific procedures and schedule to ensure accurate, consistent implementation and documentation (including data)
2.2.c	X	X		Implement quality assurance tracking		X						Assure	Provides monitoring of QA progress.

Action #	RESPONSIBLE PARTY(S)			RECOMMENDATIONS	FOCUS AREA							QA Area	ANTICIPATED DATA RELIABILITY/QUALITY BENEFIT(S)
	EPA	STATE	UTILITY ORGS.		Compliance Determination	Data Reliability	SDWIS Modernization	Monitoring and Reporting	Violation-Reporting Timeliness	Non-Reporting	Rule Development - Data Mgt Concerns		
				Action/Activity								Area	Description
2.2.d		X		Establish goals for data quality		X						Assure	Formalizes desired QA outcome(s).
2.2.e				Establish interim data quality performance targets		X						Assure	Provides measurable targets for achieving QA goals
				SDWIS MODERNIZATION									
3	X			EPA should continue implementation of the Office of Ground Water and Drinking Water Information Strategic Plan which includes:			X					Assure	Addresses technological adaptations in conformance with the Agency's E-Government goals, improved efficiencies in software and system maintenance, addresses data entry and data transfer issues, provides and enhances entire suite of SDWIS tools, with a focus on secure, improved data flow and acceptance by EPA .
3.1.a	X	X		Modernize SDWIS/FED to take advantage of newer technologies and system platforms			X					Assure	Provides a more efficient, easier to maintain, cost effective, and a more easily accessible information system
3.1.b	X	X		Broaden SDWIS information management and reporting tool set (FED and STATE versions)	X	X	X					Assure	Improves ease of reporting and error corrections
3.1.c	X	X		Facilitate and implement program and Agency data standards			X					Assure/Control	Facilitates data sharing and use
3.1.d	X			Maintain and further develop the SDWIS/STATE application to include web enabling the application	X	X	X					Assure	Provide greater access to data system and fewer state resources required to manage the system.
3.2	X	X		Continue development and maintenance of detailed documentation of, and methods for correcting and avoiding errors.		X	X					Assure	Provide tool for analysis and resolution of data entry requirements and rejected data, improving quality.

Action #	RESPONSIBLE PARTY(S)			RECOMMENDATIONS	FOCUS AREA							QA Area	ANTICIPATED DATA RELIABILITY/QUALITY BENEFIT(S)
	EPA	STATE	UTILITY ORGS.		Compliance Determination	Data Reliability	SDWIS Modernization	Monitoring and Reporting	Violation-Reporting Timeliness	Non-Reporting	Rule Development - Data Mgt Concerns		
3.3	X	X		Continue evaluation of the impact of SDWIS modernization efforts on both SDWIS/STATE and non-SDWIS/STATE states	X	X	X					Assess	Necessary to ensure that changes and improvements achieve desired objectives and to evaluate data quality effects.
MONITORING AND REPORTING													
4	X		X	EPA should encourage states to develop an automated monitoring requirements and sampling schedule tracking system - automated system could include:	X	X		X	X	X		Control	Reduces the degree of effort necessary for state oversight of water system monitoring compliance and improves timeliness and completeness of violation data.
4.1.a			X	Tracking waivers, variances, exemptions, and vulnerability assessments, and reduction, frequency, and schedule of a water system's monitoring requirements	X	X		X	X	X		Control	Provides the ability to specify and track all components of monitoring requirements for each PWS and is basis for automated compliance determination
4.1.b			X	Provide the ability to notify a water system of its monitoring requirements and sampling schedule		X		X				Assure	Improves system compliance and state receipt of data.
4.2	X		X	EPA should encourage state annual notification of monitoring requirements and sampling schedule to all water systems to ensure water systems are aware of what and when they are required to monitor	X	X		X	X			Assure	Improves communication and system compliance and state receipt of data.
4.3	X		X	EPA should encourage states to work toward the ability to receive PWS sample analytical results data electronically from laboratories	X	X		X		X		Assure	Reduces data entry error, improves timeliness of results to state allowing more timely compliance determinations
4.4	X	X	X	EPA should facilitate technology transfer to other states of various electronic transmission systems for data from public water systems and laboratories to the state to reduce states' multiple data entry burden and to improve data quality	X	X	X	X	X	X		Assess	Provides electronic transmission of data which has been shown to improve accuracy and timeliness of data processing and receipt.
VIOLATION TIMELINESS													
5	X	X		EPA and States should evaluate why violation		X		X	X			Assess	Provide an understanding of factors

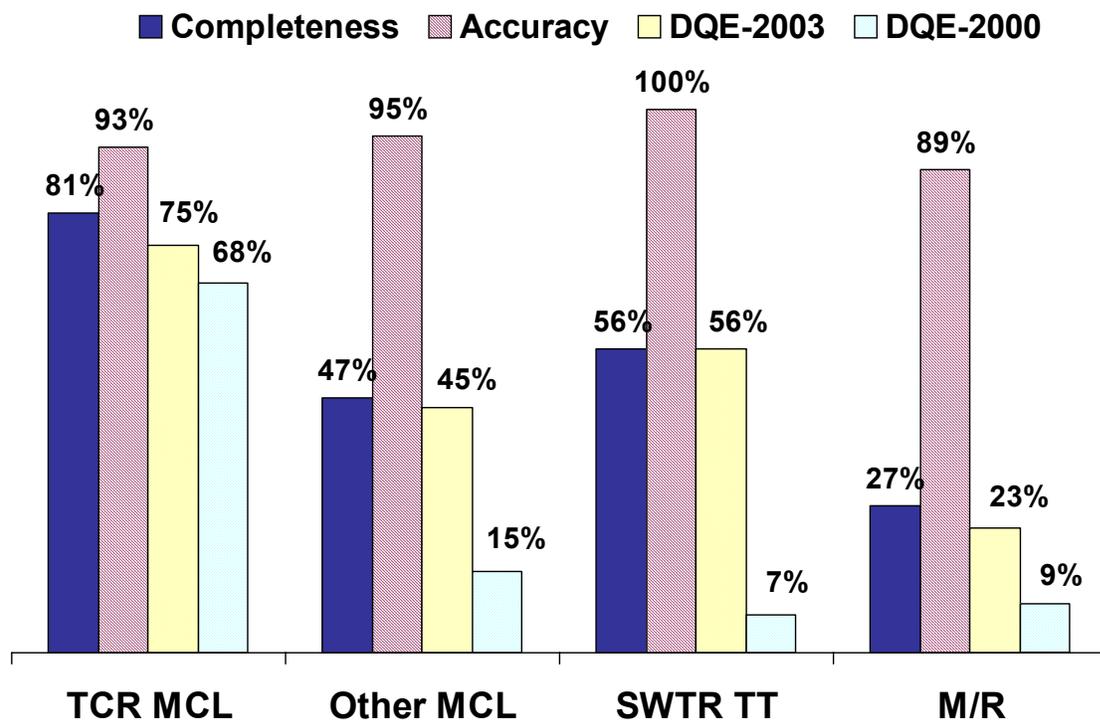
Action #	RESPONSIBLE PARTY(S)			RECOMMENDATIONS Action/Activity	FOCUS AREA							QA Area	ANTICIPATED DATA RELIABILITY/QUALITY BENEFIT(S)
	EPA	STATE	UTILITY ORGS.		Compliance Determination	Data Reliability	SDWIS Modernization	Monitoring and Reporting	Violation-Reporting Timeliness	Non-Reporting	Rule Development - Data Mgt Concerns	Area	Description
				reporting timeliness is low and not improving including:									impacting timeliness to allow resolution of areas under state control and develop improvement plan for those areas outside of state control
5.a	X	X		Evaluate data flow and impacts on data from laboratories to states	X	X		X				Assess	Same as above
5.b	X	X	X	Evaluate data flow for operations reports and results data from water systems to states	X	X		X	X			Assess	Same as above
5.c	X	X	X	Identify conditions affecting state timeliness of compliance decisions	X	X			X			Assess	Same as above
5.d	X	X	X	Identify impacts on inventory, violations, enforcement, and other data from states to EPA	X	X			X			Assess	Same as above
				VIOLATION NON-REPORTING									
6	X	X		EPA and states should conduct annual evaluation of all instances of potential violation non-reporting including - document, evaluate, and develop steps to improve reporting and to verify that non-reporting of violations does not continue		X		X	X	X		Assess	Documents when violations for a particular rule/violation do not exist on the state level, avoiding erroneous interpretations of water quality and data quality completeness
7	X	X	X	EPA, states, and other stakeholders should continue to ensure that data management concerns are considered during every phase of the rule development process	X	X	X	X	X	X	X	Assure	Ensures data management concerns are considered throughout the process. Addressing/considering data management issues helps to clarify implementation logic, complexity, and manageability of the regulation to reduce conditions affecting nonreporting.

6.0 Conclusion

6.1 Findings

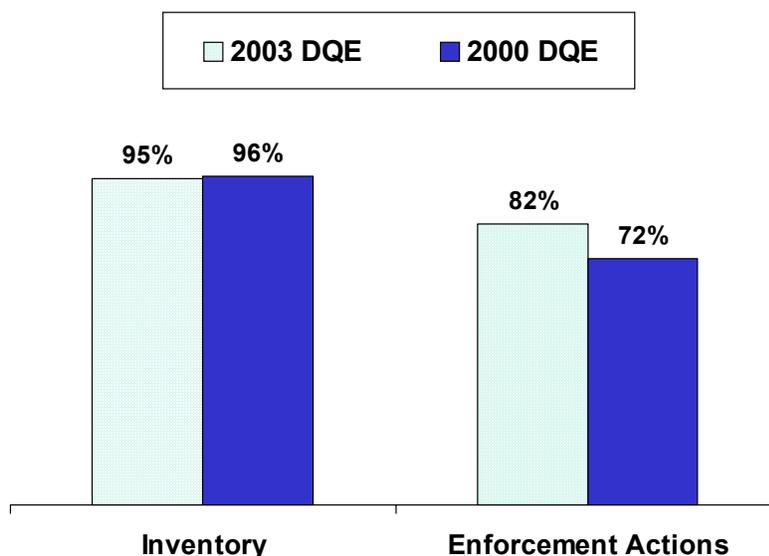
Overall, data quality has improved since the first data quality assessment released in 2000 (Figure 6-1). The accuracy of the data in EPA's SDWIS/FED drinking water violations data base is high, but the data are incomplete. This finding raises concerns for effective program management and for accurate risk communication.

Figure 6-1. Completeness, Accuracy and Data Quality Estimates for Violations



About half of errors for all MCL violations were due to compliance determination errors. More than 85% of SWTR TT and M/R violation errors were compliance determination errors, close to half of which represented situations where the state failed to assign a violation when there was no record of sampling data in state records. To a much lesser degree, the incomplete reporting of violations is attributable to differences between state and EPA regulatory interpretation. While improvements in violation reporting are needed for all rules, a particular emphasis needs to be placed on improving reporting for Chemical and Radionuclides rules, the SWTR and M/R requirements for all rules.

Figure 6-2. Data Quality Estimates for Inventory and Enforcement Actions



The quality of core inventory data continues to be high, and enforcement actions data, although improving, is still of moderate data quality (Figure 6-2). Notably, there were 80% fewer enforcement actions from the first assessment. Additional findings were:

- C Many states are not meeting the 90-day deadline for reporting violations. In 2001, only 58% of violations eventually reported were reported on time. The timeliness in which health-based violations are reported has been steady, and is similar across water system types and sizes.
- C A significant number of states still periodically do not report violations of certain rules (particularly Radionuclides) from year to year, which needs further evaluation.
- C An analysis of data rejected from SDWIS/FED found that 90% of the inventory, violations and enforcement data error types incurred were for data entry errors.

6.2 Implications for Government Performance Results Act (GPRA) Reporting

Each year EPA reports on progress in meeting strategic planning goals under GPRA. EPA currently has a goal that 95 percent of the U.S. population will be served by community water systems (CWS) that meet all health-based standards through effective treatment and source water protection by 2008. At the end of FY 2002, the agency reported that 93.6 percent of the population were served by CWSs that reported meeting all health-based standards. EPA uses the data reported by states to SDWIS/FED to calculate the GPRA measure. The quality of the SDWIS/FED data used to determine progress can affect the utility of the information in reporting for GPRA. If the quality of the data measured and reported to SDWIS-FED is not high, then

EPA's ability to report on program progress is hindered. However, the data that states report to EPA on violations used to calculate GPRA results has, to date, represented the data that are available for such reporting.

Steps taken to improve SDWIS-FED data quality will serve to increase the confidence of EPA measurement of the GPRA goal. Suggestions have been made to take the results of this analysis and adjust values reported for GPRA. However, as currently designed, this analysis and its supporting DV audits are not intended to check the accuracy of GPRA measures, but to check the accuracy of data in SDWIS/FED for use in program management. Available information and analytical results from the data reliability assessment cannot be correlated to water quality at specific water systems other than those included in the assessment through use of DV findings, nor to water quality at the national level. EPA is working to determine if a statistical method can be developed that would use the results from DV's to report on GPRA measures. There are concerns, in that such an approach would not use data from all systems to generate the value, but from only a statistically significant subset. While this would provide a number that could be used for national reporting, it would not help to answer questions about drinking water quality for an individual system.

Over the past several years, when reporting water system violations and results for GPRA measures, EPA has provided caveats that reflect the Agency's concern with data quality. For example, on the public access Envirofacts database (<http://www.epa.gov/enviro/html/sdwis/>), through which the public can obtain information on a specific system, EPA displays the following language "NOTICE: EPA is aware of inaccuracies and under reporting of some data in the Safe Drinking Water Information System. We are working with the states to improve the quality of the data". In EPA's 2003-2008 Strategic Plan, the following footnote was included:

Note: Routine data analyses of the Safe Drinking Water Information System (SDWIS) have revealed a degree of nonreporting of violations of health-based drinking water standards and of violations of regulatory monitoring and reporting requirements. As a result of these data quality problems, the baseline statistic of national compliance with health-based drinking water standards likely is lower than reported. In consultations with states, the Agency is currently engaged in statistical analysis to more accurately quantify the impact of these data quality problems, and this has resulted in significant improvements in data accuracy and completeness. Even as these improvements are made, SDWIS serves as the best source of national information on compliance with SDWA requirements and is a critical database for program management, the development of drinking water regulations, trends analyses, and public information.

EPA will continue to indicate in its reports to the public using drinking water data whether the information conveyed is affected by data quality factors. EPA will also continue to make its drinking water data quality information accessible broadly, using the Internet and other means. This second triennial assessment and plan, like the first, will be posted on EPA's website for the public's information. The Agency wants to ensure that the public has as complete information as possible for its decision making purposes and will continue to work with key

stakeholders to improve data reliability.²

6.3 Continuing Coordination for Data Quality

Since 1999, EPA has discussed data quality issues and activities with the EPA regions and states through a variety of venues. There are two standing committees which serve as analytical and recommending bodies. The first is the Data Management Steering Committee, a joint effort of the Association of State Drinking Water Administrators (ASDWA) and EPA, which is comprised of EPA Headquarters, ASDWA management co-chairs, and EPA regional and state management. Its primary purpose is to identify and review data management issues and make recommendations to EPA. The second group is the Data Sharing/Data Quality Committee which is comprised of mid-level and senior level EPA and State staff and mid-level managers. Its primary purpose is to identify issues, analyze and evaluate implementation, and recommend corrective or implementation actions to EPA through the Data Management Steering Committee. These groups meet regularly throughout the year. Annually, ASDWA and EPA co-host a national data management users conference where data management issues, information technologies, and other topics of interest are shared and discussed. During 2002 and 2003, ASDWA and EPA discussed SDWIS/FED data quality results presented in this report in three national meetings. These meetings resulted in the creation of a special State-EPA workgroup to develop the action plan described in this report.

Additionally, in 2003, EPA asked the data reliability stakeholders workgroup constituted in 1998 to review the results presented in this report as it had done for the first assessment. Based on the comments received, modifications were made in the presentation of the results, but not in the content or findings. The comments did not undermine the factual basis, calculations or data quality estimates of the assessment, or redirect the planned actions to improve data quality in the future. In response, EPA emphasized the context of the results, focusing on the significance of the violations of drinking water standards that federal regulations require be reported to EPA relative to the larger body of compliance determinations made by states that indicate the safety of the nation's drinking water.

² The EPA Office of Ground Water and Drinking Water has been assessing other factors affecting the data in SDWIS/FED beyond the data verification audits findings. This assessment's initial results point to unrevised violations from past years that have been corrected at the system level, but continue to be counted as violations for compliance purposes or under GPRA because the state did not report that the system had returned to compliance. Additionally, violations at larger systems are counted as affecting the entire system's population, rather than that portion of the population that is affected by them. This is an accounting challenge enmeshed in the "system-to-state-to-EPA" reporting process agreed to historically under existing regulations and difficult to untangle since states may not keep track of which portion of a system may be in violation, just that it had a violation of a particular type that needed to be addressed. Reporting the entire system in violation under GPRA reduces the population served by systems meeting all health-based standards and treatment unnecessarily. EPA will explore how best to address these considerations in the future.

6.4 Prospective Measures

Fundamental to this data quality analysis is that good government (processes and decisions) demands good data (of known and documented quality). The public expects that governments at all levels will use the very best data available. Data are critical to informed and considered decisions, and the public health focus of the drinking water program requires the best data. The results presented in this report are factual, derived from compliance data reported by states to EPA and from EPA on-site audits of state files. The data are not perfect for various reasons which have been described. While states and EPA have made significant progress in improving the quality of these data, the data still need further improvement.

States have indicated that because of regulation complexity and resultant competing demands of the program, they operate their public water system (PWS) regulatory programs in the best manner they can, which is now stressed by limited and often reduced resources and most recently security requirements. These stresses and constraints may have unintended consequences for data quality. Therefore, a plan to address continued improvement in the drinking water compliance data reported by states has been included as a product of this analysis.

For data flow errors as identified in this analysis, SDWIS modernization should address some of the problems of data submission. With respect to resolving state compliance determination errors, greater efforts will be focused on defining areas of disagreement in regulation interpretation between EPA and states. Resolution will be achieved through clarification of regulatory requirements, training and technical assistance, and other state specific program oversight and support activities. For monitoring and reporting, attention will focus on developing mechanisms by which results can be transmitted electronically from laboratories to public water systems and states. Participants in this analysis and plan will strive to implement its recommendations and report progress in the next triennial report.