United States
Environmental Protection
Agency

2006 Drinking Water Data Reliability Analysis and Action Plan

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

Office of Water (4606M)
EPA 816-R-07-010
www.epa.gov/safewater
March 2008

Printed on Recycled Paper

2006 Drinking Water Data Quality Analysis and Action Plan

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 database for the national drinking water program. It contains data on public water systems (PWS) provided by the states to EPA. It is primarily used for management of state and EPA programs and for informing the public about the compliance status of their drinking water systems, and indirectly, the safety of their drinking water. EPA uses the information in SDWIS/FED for various analyses to support programmatic decisions and identify trends, problems, and opportunities for improvement of the states' rule implementation as well as program oversight. Consequently, the utility of SDWIS/FED information for these purposes highly depends on the quality of the data it contains.

EPA routinely evaluates state programs by conducting Data Verification (DV) audits, which evaluate state compliance decisions and reporting to SDWIS/FED. EPA prepares triennial summary evaluations based on the DV. This document presents results of EPA's third triennial review of data quality in SDWIS/FED and includes an evaluation of the data collected from 2002 through 2004. For the 38 states evaluated, we found that:

- Ninety-four percent of health-based violation data in SDWIS/FED were accurate.
- Approximately 81% of the maximum contamination level (MCL) and surface water treatment technology (SWTR TT) violations were reported to SDWIS/FED.
- Including lead and copper treatment technology (LCR TT) violations, about 62% of the health-based violations (MCL and Treatment Technology violations) were reported to SDWIS/FED, where only 8% of LCR TT violations were reported to SDWIS/FED.
- Only approximately 30% of the monitoring and reporting (M/R) violations were reported to SDWIS/FED.
- The primary reason for non-reporting was due to compliance determination errors rather than data flow errors.
- Further, 60% of the health-based violations[1] were reported on time and approximately 30% of the monitoring and reporting violations were reported on time to SDWIS/FED in 2004.

Background

SDWIS/FED contains data about PWS facilities, violations (e.g., exceptions and exceedances) of Federal drinking water regulations adopted by the states, and enforcement actions taken by the state. The regulations include health-based drinking water quality standards, performance of

1 The health-based violations in this reference do not include lead and copper treatment technology violations because they have open-ended compliance period end dates.

treatment techniques and/or process requirements. The focus of this report is on two types of violations: (1) health-based violations (i.e., exceedance of maximum contaminant level or non-performance of a treatment technique or process), and (2) monitoring and reporting violations (i.e., a water system did not monitor, did not report monitoring results, or were late in reporting results to the state.)

States manage their own processes and databases differently 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). Violation data that states report to EPA (SDWIS/FED) reflect only those major and minor noncompliance results that may lead to adverse public health outcomes. Violations represent a small fraction of all the determinations states make which demonstrates the safety of the nation's water supply.

The first triennial review of data quality evaluated data for the period 1996-1998. That assessment, which resulted in a detailed data analysis report in 2000, produced an action plan under which states and EPA worked together to improve data quality. The plan resulted in actions that included training state staff, streamlining reporting to SDWIS/FED, making SDWIS/FED error reporting and correction more user-friendly, improving DVs, following up with Regions after DVs, and encouraging states to notify water systems of sampling schedules annually. Similarly, the second triennial review of data quality analyzed the data from the period 1999-2001 and findings were presented in the 2003 report. The recommended action plan in the 2003 report included:

- Development of state-specific compliance determination and quality improvements plans necessary to remedy the major problem areas,
- Conducting and improving data quality analysis and report results,
- Implementation of the OGWDW information strategic plan and SDWIS/FED modernization,
- Development of an automated monitoring requirement and sampling schedule tracking system by the states, and evaluation of timeliness of violations and potential violation non-reporting.

This Review

Between 2002 and 2004, EPA conducted DV audits in 38 states and reviewed data on drinking water system information, violations, and enforcement actions. See Table ES-1 for the list of DV States. EPA evaluated 2,658 PWSs, of which 43% were Community Water Systems (CWS). See Table ES-2 for the distribution of systems by system type and the size of population served. The violations addressed by the DVs are shown in the Appendix B. The period of review by rule was generally the two most recently scheduled monitoring periods for each water system and applicable rule. For the Total Coliform Rule (TCR) and the Surface Water Treatment Rule (SWTR TT), the most recent four quarters were evaluated.

Table ES-1: States Subject to Data Verifications from 2002-2004

Region	States		Region	States
1	CT, MA, RI, VT		6	AR, NM, OK, TX
2	NJ, VI		7	IA, MO
3	MD, PA, VA, WV		8	CO, SD, UT, WY
4	AL, FL, KY, MS, NC('02), NC('04) SC, TN		9	AZ, CA, R9 Tribes
5	IL, MI, MN, OH		10	AK, ID, WA

Table ES-2: Number of Systems included in Data Verifications by System Type and Size

	System Type			
System Size	**CWS**	**NTNCWS**	**TNCWS**	**Total**
Very Small (500 or fewer)	572	637	696	1,905
Small (501-3,300)	277	123	36	436
Medium (3,301-10,000)	119	9	6	134
Large (10,001-100,000)	135	4	0	139
Very Large (>100,000)	44	0	0	44
Total	1,147	738	773	2,658

Summary of Results

For the MCL/SWTR TT violations, 81% of the data were reported to SDWIS/FED, Figure ES-3 summarizes the data quality estimates by violation type. Of the non-reported violations, 74% were due to compliance determination (CD) errors, where the states did not issue a violation when a violation had occurred. Twenty-six percent of the non-reported violations were due to data flow (DF) errors. Figure ES-4 summarizes the percentage of errors contributed from non-reporting by violation type. Approximately 94% of the data in SDWIS/FED were accurate. The overall data quality (DQ) of the MCL/SWTR TT violations was 77%. This means that 77% of the noncompliance determinations on MCL/SWTR TT were correctly reported in SDWIS/FED.

Figure ES-3: Data Quality Estimates by Violation Type

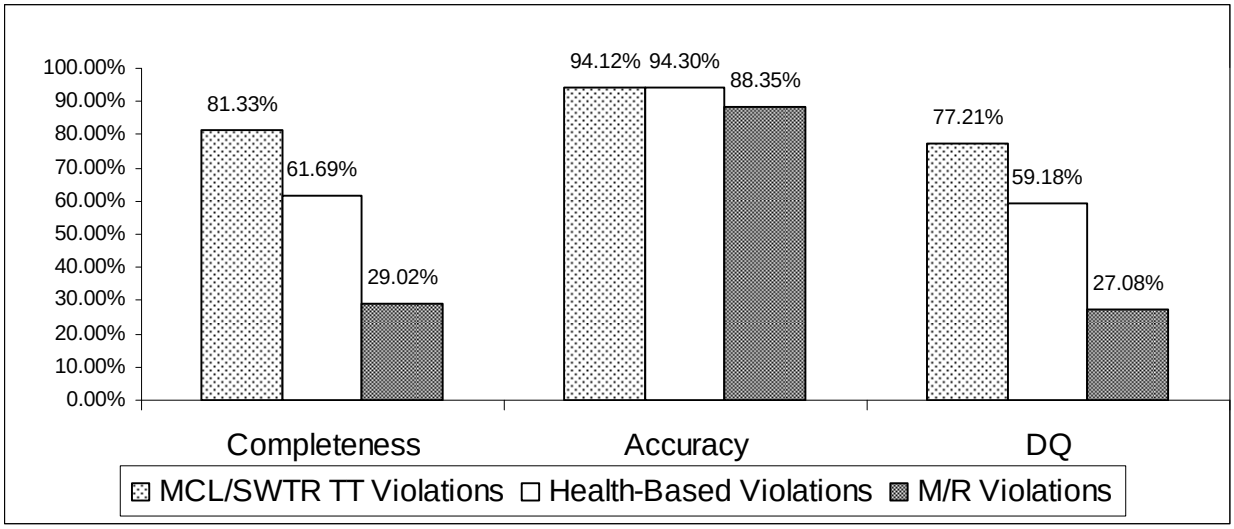
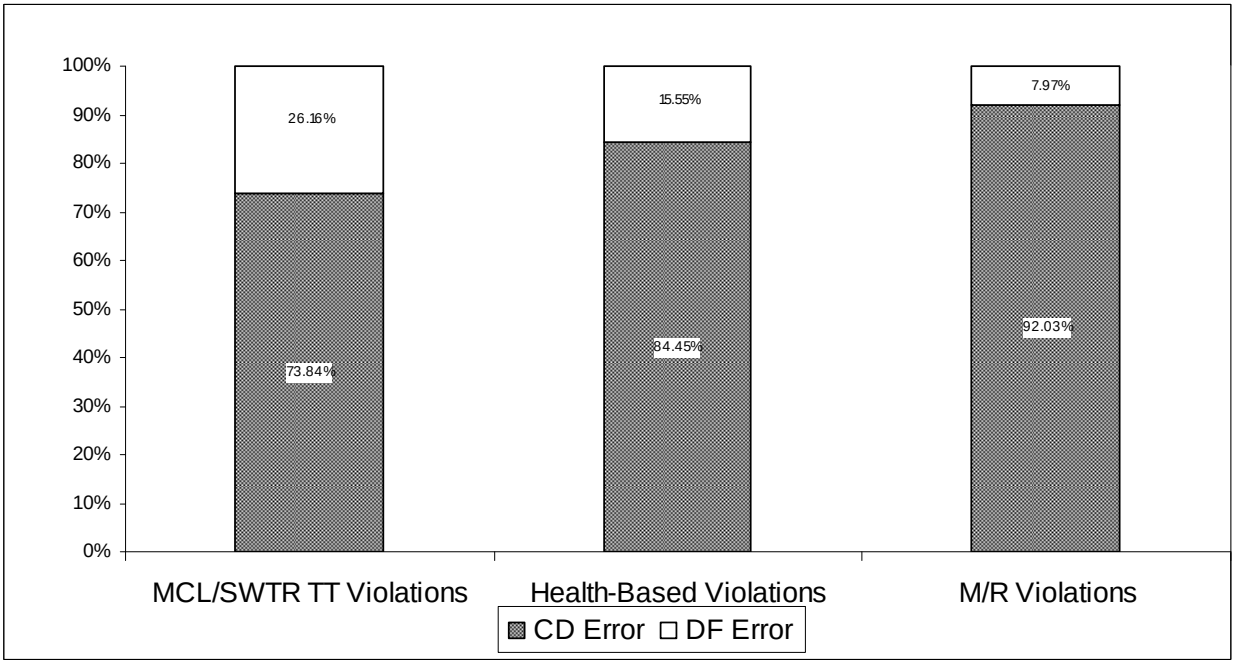


Figure ES-4: Percentages of Error Contribution to Non-Reporting of Violations



CD: Compliance Determination
DF: Data Flow

For the health-based violations including LCR TT violations, 62% of the data were reported to SDWIS/FED. Of the non-reported violations, 84% were due to CD errors. Approximately 94% of the health-based violations data in SDWIS/FED were accurate in SDWIS/FED. The overall data quality of the health-based violations was 59%, i.e., approximately, 59% of the noncompliance determinations on health-based standards were correctly reported in SDWIS/FED.

Only 29% of the monitoring and reporting violations were reported to SDWIS/FED. Ninety-two percent of the non-reported violations were due to CD errors. Approximately 89% of the monitoring and reporting violations data in SDWIS/FED were accurate. The overall data quality of the M/R violations was 27%, i.e., 27% of the noncompliance determinations on M/R were correctly reported to SDWIS/FED.

Data Reliability Improvement Action Plan

Appendix A is a joint plan of EPA and the Association of State Drinking Water Administrators to achieve a goal of 90 percent complete and accurate data for health-based violations, as well as improving the quality of monitoring and reporting violations and inventory data. Progress toward accomplishment of this goal will be measured annually and assessed in 2009.

Acknowledgements

The following people contributed to this analysis and the preparation of this report:

Project Lead Drinking Water Protection Division
Office of Ground Water and Drinking Water

Chuck Job, Branch Chief, Infrastructure Branch

Leslie Cronkhite, Associate Branch Chief, Infrastructure Branch

Principal Author Jade Freeman, Ph. D., Statistician, Infrastructure Branch

Contributing Author Lee Kyle, IT Specialist, Infrastructure Branch

Peer-review of the statistical methodology in this report has been provided by:

Anthony Fristachi, Exposure Analyst
National Center for Environmental Assessment
U.S. EPA Office of Research and Development

Tony R. Olsen, Ph. D., Environmental Statistician
Western Ecology Division
Health and Environmental Effects Research Laboratory
U.S. EPA Office of Research and Development

Arthur H. Grube, Ph.D., Senior Economist
U.S. EPA Office of Pesticide Programs

A. Richard Bolstein, Ph. D., Chairman (retired)
Department of Applied & Engineering Statistics
George Mason University

John Gaughan, Ph. D., Associate Professor
Epidemiology & Biostatistics
Temple University School of Medicine

Matthias Schonlau, Ph. D., Head Statistician
Statistical Consulting Service
The Rand Corporation

Collaboration on the Data Reliability Improvement Plan - Association of State Drinking Water Administrators

Table of Contents

1. Introduction 1
   1.1 Previous Activities 2
   1.2 Regulatory Context 2
   1.3 Changes in 2006 Analysis Method 3

2. Overview of Data Verification 3

3. Statistical Sample Design of Data Verification and Analytical Method 6
   3.1 Selection of States 6
   3.2 Selection of Systems with State 7
      3.2.1 Sample Frame 7
      3.2.2 Sample Design Data Verification 7
      3.2.3 Sampling Procedure and Data Collection Activities 8
   3.3 Analytical Method: Weighting and Estimation 9

4. Results from the Analysis of Data Verification 12
   4.1 Analysis of Inventory Data 12
   4.2 Analysis of Violation Data 13
      4.2.1 Results from 2002-2004 Data Verifications 19
      4.2.2 Results from 1999-2001 Data Verifications 23
      4.2. Data Quality Estimates from 1999-2001 and 2002-2004 25
   4.3 Analysis of Enforcement Data 31

5. Analysis of Timeliness of Violation Reporting in SDWIS/FED 31

6. Conclusion 34

7. Data Reliability Improvement Action Plan 35

8. Future Analysis of Data Reliability 35

Appendix A: 2006 Data Reliability Improvement Action Plan 37

Appendix B: Violations Addressed by Data Verification (DV) 45

Appendix C: Definition of Public Notification (PN) Tier 47

2006 Drinking Water Data Quality Assessment and Action Plan

1. Introduction

The Safe Drinking Water Information System/Federal Version (SDWIS/FED) is the Environmental Protection Agency's (EPA) principal database 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 (PWSs) and, indirectly, the safety of drinking water. The Federal government uses SDWIS/FED data for program management for 90 contaminants (as of 2005) regulated in drinking water at approximately 158,000 PWSs in 56 state and territorial programs and on Indian lands. Data received by EPA from states in SDWIS/FED includes a limited set of water system descriptive information, e.g., system type, population served, number of service connections, water source type), data on PWSs' violations of regulatory standards and process requirements, and information on state enforcement actions. These data, which EPA uses to assess compliance with the Safe Drinking Water Act (SDWA) 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 (DV) audits in states and develops a summary evaluation every three years called Drinking Water Data Quality Assessment. DV auditors evaluate compliance data in state databases and hard copy files, monitoring plans, and other compliance information submitted by PWSs. The auditors also examine sanitary surveys, correspondence between the state and the water system, compliance determination decisions, and enforcement actions taken by the state. Based on this information, the auditors confirm whether all required information was submitted to and evaluated correctly by the state and whether required reporting elements were submitted to SDWIS/FED.

This report includes (1) a description of the methodology used; (2) analyses of the data from the 2002 to 2004 Data Verifications, the most recent triennial evaluation period; and (3) analysis of the timeliness of reporting in SDWIS/FED. The report also describes a plan to address continued improvement in drinking water compliance data reported by states. This report is not intended for evaluating states' performance. This report is a tool to identify the gap between the states' violation data and SDWIS/FED and to provide a benchmark for the collaborative efforts between the states and EPA to bridge the gap and improve the data quality in SDWIS/FED.

1.1 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 regarding incorrect violations in the database. 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.

Findings of the first Data Reliability Analysis, which indicated that data quality needed improvement, were later updated by the second triennial assessment in 2003 (which included data collected between 1999 and 2001). Together, these assessments included comprehensive recommendations for EPA and state primacy agencies on quality improvements. The reports identified near-term actions that had already been taken or were still needed to improve data quality more immediately. To implement the recommendations, the states and EPA have conducted numerous activities and projects to improve data quality. Activities undertaken have included a) providing training for states; b) streamlining reporting to SDWIS/FED; c) making SDWIS error reporting correction more user-friendly; d) improving data verifications; e) following up with Regions on findings after data verifications; and f) 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 2003 report included a commitment to conduct analyses which would provide periodic data quality estimates (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.

1.2 Regulatory Context

States make a large number of determinations regarding public water systems' compliance with drinking water regulations and violations of these regulations are a small fraction of these determinations. Since violations represent a small fraction of all the determinations states make, this result indicates the general safety of the nation's drinking water supply. For example, an analysis of nitrate maximum contaminant level compliance data for Oklahoma from 2004, showed only 3% of determinations resulted in violations.

The data considered for evaluating quality, particularly accuracy and completeness, consist of the violations of health-based standards and monitoring and reporting requirements. These data are important for two reasons: (1) States 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), violation data have become a major focus. EPA's 2006-2011 strategic plan specifies a clean and safe water goal of "90% of the population served by community water systems (CWS) meeting all health-based standards and treatments by 2011." 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. Due to the importance and emphasis on violation data, this data quality evaluation methodology addresses whether states correctly identify and report the violations that should have been reported to EPA according to state primacy agreements pursuant to Federal regulations.

1.3 Changes in 2006 Analytical Method

In this analysis of 2002 to 2004 DV data, EPA uses a different method for evaluating the data quality as described below.

! In the previous report, the DQEs were calculated without considering the sample design of DVs, i.e., the selection process by which the systems are included in the sample. In this assessment, the DQEs are calculated using statistical sample design-based unbiased estimation. The sample design and the estimation method for calculating sample statistics are described in detail in Section 3.

! The completeness measure of the violation data quality in the 2003 report represented the proportion of accurate data in SDWIS/FED out of all violation data that should be reported to SDWIS/FED. However, in this 2006 analysis, EPA redefined completeness of SDWIS/FED based on any violation data reported to SDWIS/FED regardless of the accuracy.

! Because of the changes in the estimation method described above and non-random selection of states for DV audits, the results from this analysis will not be compared to those from the 2000 or 2003 assessments.

The statistical methodology for the analysis of DV data and the results are described in Sections 3 and 4. The additional analysis of the timeliness of reporting in SDWIS/FED is presented in Section 5.

2. Overview of Data Verification

EPA's OGWDW routinely conducts DV audits, which evaluate the management of state drinking water programs. During the DVs, EPA examines state compliance decisions, data on the system compliance and violations in the state files, and the data required to be reported to SDWIS/FED. During the DVs, EPA reviews data submitted by PWSs, state files and databases, and SDWIS/FED, and compiles the results on the discrepancies among the data. States have several opportunities to respond to findings while DV personnel are on site, and provide additional clarifying information if available. States also review the DV draft report before the final report is produced, and their comments are incorporated into the report. EPA responds to every state comment, to explain in detail whether or not the state's additional information changed the finding.

Until 2004, states were selected for DVs considering a number of factors; for example, the states that had not been audited for a long period of time were selected for DVs. Also, in order to minimize the burden on EPA Regions and states, OGWDW tried to maintain an even distribution of DV states across the regions[2]. Further, resource constraints have affected the selection of

2 EPA is divided into 10 regional offices, each of which is responsible for several states and territories.

certain states since it is more costly to conduct DVs in some states than others. Between 2002 and 2004, EPA conducted DV audits in 38 states and reviewed data on drinking water system information, violations, and enforcement actions (Table 2-1). State files for a total of 2,658 PWSs were evaluated, of which 43% were community water systems (Table 2-2). The regulations addressed by the DVs and the compliance period reviewed for each regulation are shown in Table 2-3.

Table 2-1: States Subject to Data Verifications from 2002-2004

Region	States		Region	States
1	CT, MA, RI, VT		6	AR, NM, OK, TX
2	NJ, VI		7	IA, MO
3	MD, PA, VA, WV		8	CO, SD, UT, WY
4	AL, FL, KY, MS, NC ('02), NC ('04), SC, TN		9	AZ, CA, R9 Tribes
5	IL, MI, MN, OH		10	AK, ID, WA

Table 2-2: Number of Systems included in Data Verifications by Type and Size

	System Type[3]			
System Size	**CWS**	**NTNCWS**	**TNCWS**	**Total**
Very Small (500 or fewer)	572	637	696	1,905
Small (501-3,300)	277	123	36	436
Medium (3,301-10,000)	119	9	6	134
Large (10,001-100,000)	135	4	0	139
Very Large (>100,000)	44	0	0	44
Total	1,147	738	773	2,658

3 Community water systems (CWSs) have at least 15 service connections or serve 25 or more of the same population year-round. Nontransient noncommunity water systems (NTNCWSs) regularly serve at least 25 of the same persons over 6 months per year. Transient noncommunity water systems (TNCWSs) provide water where people remain for periods less than 6 months.

Table 2-3: Period of Compliance for Rules Reviewed During 2002-2004 Data Verifications

Rule[4]	Compliance Period Reviewed
Inventory	Most Recent
Consumer Confidence Report (CCR)	
Total Coliform Rule (TCR), Surface Water Treatment Rule (SWTR), Total Trihalomethanes (TTHMs)	Most Recent 12-Month Period Available in SDWIS/FED
Nitrates	Most Recent Two Calendar Years
Phase II/V excluding nitrates	1999-2001
Lead and Copper Rule (LCR), Interim Radionuclides Regulation	Most Recent Two Samples
Enforcement	Time Period Related to Violation
Public Notification	

The review evaluated recent monitoring history to confirm that systems monitored according to the required frequency. For many rules, the review evaluated one year of information (Surface Water Treatment Rule, Total Trihalomethanes, Total Coliform Rule, and Consumer Confidence Report). The two most recent monitoring periods or review cycles were reviewed for some rules (interim radionuclides, Lead and Copper Rule, sanitary surveys). In other instances, the review covered a defined period, such as the most recent 3-year monitoring period for the Standard Monitoring Framework outlined in the Phase II/V Rule[5]

3. Statistical Sample Design of Data Verifications and Analytical Methods

3.1 Selection of States

As mentioned in Section 2, the states are selected for DVs by considering the date of their last verification, resource constraints, and burden on EPA Regions and states. This selection

4 CWSs were reviewed for inventory and each of the rules listed in this table. NTNCWSs are not subject to CCR, TTHM monitoring, or the interim radionuclide regulation. TNCWSs are not subject to the requirements for CCR, SWTR, TTHM, Phase II/V Rule, or interim radionuclide regulation.

5 The Standardized Monitoring Framework synchronizes the monitoring schedules for the Phase II/V regulation for chemicals and the interim radionuclides rule across defined 3-year monitoring periods and 9-year monitoring cycles.

procedure is a non-probability sampling method. Because of the subjective nature of the selection process, non-probability samples add uncertainty when the sample is used to represent the population as a whole. The accuracy and precision of statements about the population can only be determined by subjective judgment. The selection procedure does not provide rules or methods for inferring sample results to the population, and such inferences are not valid because of bias in the selection process.

When non-probability sampling is used, the results only pertain to the sample itself, and should not be used to make quantitative statements about any population including the population from which the sample was selected. Since the DV states were selected by a non-probability sampling method, the results from the analysis only pertain to the DV states audited between 2002 and 2004. Therefore, it is not appropriate to make quantitative statements or inferences about the entire nation from the selected states or comparisons with sampled state data quality results from the previous years.

3.2 Selection of Systems within States

The DVs involve the evaluation of the states' compliance decisions and the agreement between the data in the state files and SDWIS/Fed. Since neither time nor resources allow a complete census of consistencies between SDWIS/Fed and state records, EPA uses a statistically random sample of systems that is drawn from the total number of systems in the state. EPA uses the results from the probability sample of systems within each state to estimate DV compliance results for each state. The probability sample is designed to provide estimates with acceptable precision while minimizing the burden on Regions and states imposed by visits from auditors. EPA plans to further reduce burden on Regions and States through use of electronic data comparison.

3.2.1 Sample Frame

A sample frame is a list of all members of a population (in this case, the PWSs), from which a random sample of members will be drawn. In other words, the sample frame identifies the population elements from which the sample is chosen. The population elements listed on the frame are called the sampling units. Often these are groups or clusters of units rather than individual units. For each state, EPA developed a sample frame (i.e., a list of the current inventory of PWSs in the state) using SDWIS/FED, from which a random sample of PWSs was selected according to the sample design.

3.2.2 Sample Design of Data Verification

The unit of analysis is the recorded *action* taken by systems, not the systems themselves. The sample design for DVs is a stratified random cluster sample. In stratified sampling, the population is divided into non-overlapping subpopulations called strata and a random sample is taken from each stratum. Stratification increases the precision of the estimates when the population is divided into subpopulations with similar characteristics within each stratum. In cluster sampling, groups, or "clusters," of units in the population are formed and a random

sample of the clusters is selected. In other words, within a particular stratum, rather than selecting individual units, clusters of units are selected.
In the analysis of DV data, systems are grouped into three strata according to the system type (CWS, TNCWS, and NTNCWS) within each state. In the first stage of the sampling process, systems are randomly selected within each stratum. In the second stage, each action taken by the system is recorded. In other words, the system represents a cluster of actions. A few examples of these actions are:

! System inventory information that must be reported to SDWIS/FED,

! Violations of federal regulations (states also may report violations of state regulations),

! Enforcement actions taken when violations occur.

3.2.3 Sampling Procedure and Data Collection Activity

Once the current state inventory is retrieved from SDWIS/FED, the number of systems is counted by size category (see Table 2-2 for size categories). The sample size for each system type within a state is calculated based on the acceptable precision level for the estimates within margin of error in most states of plus or minus five percent, with a confidence level of 90 or 95 percent[6]. As discussed in section 3.2.2, the sample design is a stratified random cluster sample. The required sample size is given by:

$$n'_h = (n_h)(deff)$$

where n_h is the size of the sample (number of systems) required for stratum h (specific state and system type) if a simple random sample is drawn. *deff* is the design effect of the clustering and is assumed to be greater than 1.0. n_h is given by

$$n_h = \frac{\dfrac{Z_\alpha^2 p_h (1 - p_n)}{B_h^{\,2}}}{\overline{M}_h + \dfrac{1}{N_n}\left(\dfrac{Z_\alpha^2 p_h (1 - p_h)}{B_h^{\,2}} - 1\right)},$$

where n_h= Number of systems required for the sample in stratum h,

N_h = Total number of systems in the state in stratum h,

6 For the three DVs that were conducted during the last quarter of 2004, (TX, VA, and IL), the confidence level for CWSs was 95 percent and the margin of error was plus or minus seven percent. For NTNCWSs and TNCWSs, the confidence level was 90 percent and the margin of error was plus or minus seven percent.

$\overline{M}_h$ = Average number of actions in each system in stratum *h*,

B_h = Acceptable precision level (margin of error) for stratum *h*,

Z_α = The abscissa of the normal curve that corresponds to the confidence level, and

p_h = Proportion of discrepancy in violation data between DV results and SDWIS/FED in stratum *h* (estimated from the previous assessment.)

The design effect *deff* depends on the proportion of actions and decisions reviewed in the DV that are consistent with the data in SDWIS/FED. This proportion is unknown before the DV; therefore, the design effect is unknown. Lacking estimates of the design effect, the DV draws a simple random sample within each stratum[7]. Because it excludes the design effect, this sample may not be large enough to meet the precision targets.

The sample size is calculated in an Excel spreadsheet. Samples are drawn from the frame according to the random numbers generated in an Excel spreadsheet produced by EPA. Using the Excel spreadsheet random number generator, a random sample of systems is developed for each stratum. Then, the DV auditors collect data from the state files for each sampled system on PWS inventory, violations, and enforcement.

3.3 Analytical Method: Weighting and Estimation

In this analysis, sample weights are applied to the data to adjust for the unequal probability of selection of systems, i.e., the differences in the likelihood of some systems appearing in the sample. Weights, based on the probability of selection, allow unbiased representation of the population from an unequal probability sample.

In the 2002-2004 DV data analysis, EPA estimated proportions related to consistency and accuracy among state files, the state database, and SDWIS/FED for inventory information, violation data, and enforcement actions. A few examples of such proportions are the proportion of inventory data that are consistent between SDWIS/FED and the state file, the proportion of violation data that are reported to SDWIS/FED, and the proportion of enforcement data that are consistent between SDWIS/FED and the state file. In this report, these proportions are presented as percentages after being multiplied by 100. The proportion $\hat{P}$ is estimated by

$$\hat{P} = \frac{\sum_{h=1}^{H}\sum_{\alpha=1}^{n_h}\sum_{\beta=1}^{m_{h\alpha}} W_h I_{h\alpha\beta}}{\sum_{h=1}^{N}\sum_{\alpha=1}^{n_h} m_{h\alpha} W_h},$$

7 Future DVs can estimate *deff* using data from previous DVs and can incorporate the design effect into the sample size calculation.

where the sample weight $W_h = \frac{N_h}{n_h}$,

N_h = total number of clusters (systems) in stratum (system type) h, h=1,…,H,

n_h = number of sampled clusters in stratum h,

$m_{h\alpha}$ = number of data elements (reviewed actions) from cluster α in stratum h,

$I_{h\alpha\beta}$ = 0 or 1 indicator for βth data element from the system α and Stratum h, corresponding to a specific characteristic.

A simple illustration of the calculation procedure is presented here. Suppose there are three strata (H=3), namely CWS, NTNCWS, and TNCWS in State A. Also, suppose that the total number of systems in each stratum (system type in State A) is 6, 9, and 15 (N_1=6, N_2=9, and N_3=15), respectively. Further, the number of sampled systems is 3 for each stratum (n_1=n_2=n_3=3) and there are three violations reviewed for accuracy from each sampled system ($m_{h\alpha}$ =3 for h=1,2,3 and α=1,2,3). Let $I_{h\alpha\beta}$ be 1 if the violation was accurately reported to SDWIS/FED or 0 if the violation was incorrectly reported. Suppose the compiled data are as shown in Table 3-1. The proportion of the violations with a discrepancy is estimated by the ratio of the sum of $W_h I_{h\alpha\beta}$ and the sum of $m_{h\alpha} W_h$, which, in this case, is 55/90=0.6111 or 61.11%.

Sampling errors are also estimated for the proportion estimates. Sampling errors are measures of the extent to which the values estimated from the sample (proportions in this analysis) differ from the values that would be obtained from the entire population. Since there are inherent differences among the members of any population, and data are not collected for the whole population, the exact values of these differences for a particular sample are unknown.

To estimate the sampling errors, Taylor series expansion method is applied. The Taylor series expansion method is widely used to obtain robust variance estimators for complex survey data with stratified, cluster sampling with unequal probabilities of selection. The Taylor series obtains an approximation to a non-linear function. The Taylor series expansion method is applied to the variance of the proportion estimate as

$$\hat{Var}(\hat{P}) = \sum_{h=1}^{H} \hat{Var}_h(\hat{P}), \text{ where } \hat{Var}_h(\hat{P}) = \frac{n_h(1-n_h/N_h)}{n_h-1}\sum_{\alpha=1}^{n_h}\left(e_{h\alpha}-\bar{e}_h\right)^2, \; e_{h\alpha} = \frac{\sum_{\beta=1}^{m_{h\alpha\beta}} W_h(I_{h\alpha\beta}-\hat{P})}{\sum_{h=1}^{N}\sum_{\alpha=1}^{n_h} m_{h\alpha}W_h},$$

and $\bar{e}_h = \frac{\sum_{\alpha=1}^{n_h} e_{h\alpha}}{n_h}$. With the sampling error, the margin of error based on a 95 percent confidence interval is calculated as $t_{df,0.025}\sqrt{V\hat{a}r(\hat{P})}$, where $t_{df,0.025}$ is the percentile of the *t* distribution with *df* number of degrees of freedom, which is the number of clusters minus the number of strata.

Table 3-1: Example of Proportion Estimation Procedure

Stratum index h	**Total Number of Systems** N_h	**Number of sampled systems** n_h	***Weight*** W_h	***System Index*** α	**Total number of Violations** $m_{h\alpha}$	$m_{h\alpha}W_h$	***Violation Index*** β	**Violation correctly reported? Yes=1; No=0** $I_{h\alpha\beta}$	$W_h I_{h\alpha\beta}$
1 State A CWS	6	3	2	1	3	6	1	1	2
							2	0	0
							3	0	0
				2	3	6	1	1	2
							2	0	0
							3	1	2
				3	3	6	1	1	2
							2	1	2
							3	0	0
2 State A NTNCWS	9	3	3	1	3	9	1	1	3
							2	0	0
							3	1	3
				2	3	9	1	1	3
							2	1	3
							3	0	0
				3	3	9	1	0	0
							2	0	0
							3	1	3
3 State A TNCWS	15	3	5	1	3	15	1	0	0
							2	1	3
							3	1	3
				2	3	15	1	1	3
							2	0	0
							3	1	3
				3	3	15	1	1	3
							2	0	0
							3	1	3
Σ						90			55

In Section 4, various types of proportions of consistent, reported, and accurate data in SDWIS/FED are calculated. These proportion estimates represent the data quality measures of inventory, violation, and enforcement data in SDWIS/FED based on the DVs.

4. Results from the Analysis of Data Verifications

This section presents various proportion estimates for inventory, violation, and enforcement data. Also, the margins of error are calculated for each point estimate. The margin of error is based on a 95% confidence interval, which includes the true proportion with 95% confidence. All calculations were performed using SAS®.

4.1 Analysis of Inventory Data

States are required to report eight inventory data elements to SDWIS/FED for Grant Eligibility. These elements are 1) public water system identification number (PWS ID; 2) system status 3) water system type; 4) primary source water type; 5) population served; 6) number of service connections; 7) administrative contact address; and 8) water system name. The records for population or service connections are considered to be consistent when there is less than 10% difference between the two records. Because the inventory data are analyzed at the system level, the estimation approach can be based on a stratified random sampling. Then, the proportion of systems for which the inventory elements were reported to SDWIS/FED without discrepancies and its sampling error are estimated in Section 3.3 only at the cluster (system) level (or $\beta=1$).

Inventory data quality of each data element is displayed in Table 4-1. The overall data quality of the eight inventory (water system identification) parameters assessed was 87%. In other words, 87% of systems from DV states between 2002 and 2004 had consistent data for all eight inventory data elements between their state files and SDWIS/FED database, or 13% of systems had at least one data element reported with a discrepancy. The highest discrepancy rate was for the administrative contact address element.

Table 4-1: Percent PWSs reported Grant Eligibility Inventory Data to SDWIS/FED without Discrepancy

Reported Data without Discrepancy by Individual Data Element								Reported All Inventory Data Element Data without Discrepancy
PWS ID	**System status (active or inactive)**	**Water system type**	**Primary source type**	**Population served**	**# service connections**	**Admin. contact address**	**PWS name**	
99.83% (+/-0.20%)	97.26% (+/-1.15%)	98.21% (+/-0.72%)	99.31% (+/-0.35%)	97.11% (+/-0.71%)	96.22% (+/-0.93%)	95.97% (+/-1.32%)	99.87% (+/-0.09%)	87.4% (+/- 1.94%)

4.2 Analysis of Violation Data

Federal regulations specify the outcomes which states must report to EPA that result in noncompliance (violation) with (a) health-based drinking water quality maximum contaminant levels (MCL) and related requirements for their attainment; (b) specified monitoring and reporting (M/R) requirements necessary to determine whether sampling, testing and treatment process checking occurred as stipulated in Federal regulations; and (c) health-based treatment techniques (TT) and associated water system management processes for contaminants for which it is not technologically or economically feasible to set an MCL.

Violation data are evaluated by comparing the following: 1) EPA's evaluation of the state's compliance decision on the violations; 2) the assigned violations in the state files; and 3) the violations reported to SDWIS/FED. All the findings from these comparisons can be grouped into one of the categories as shown in Table 4-2. The total number of violations identified during the 2002-2004 DV is summarized below:

- Out of 198 TCR MCL violations, 163 violations were reported to SDWIS/FED.
- Out of 48 other (non-TCR) MCL violations, 21 violations were reported to SDWIS/FED.
- Out of 41 SWTR TT violations, 35 violations were reported to SDWIS/FED.
- Out of 176 LCR TT violations, 5 violations were reported to SDWIS/FED.
- Out of 5,069 M/R violations, 1,589 violations were reported to SDWIS/FED.

The following measures of data quality of violation data in SDWIS/FED are evaluated:

! **Completeness** of SDWIS/FED describes how many violations that are required to be reported are being reported to SDWIS/FED, expressed as a percentage. This quantity is estimated based on the violations found by EPA and reported to SDWIS/FED (EPA=Yes and SDWIS/FED=Yes;1, 4 from Table 4-2) out of all violations found by EPA (EPA=Yes;1, 2, 3, 4 from Table 4-2).

! **Non-reporting rate** in SDWIS/FED describes how many violations that are required to

be reported are not being reported to SDWIS/FED, expressed as a percentage. This percentage is the complement of the Completeness estimate, i.e., 100%-Completeness.

! **Compliance Determination (CD) error rate** in the non-reported violations describes how many non-reported violation data are the result of errors in states' compliance determination (i.e., a violation was not reported because the state did not identify it as a violation), expressed as a percentage. This quantity is estimated based on the violations found by EPA, but not reported to SDWIS/FED and where the assigned violation in the state file does not agree with EPA (EPA=Yes and SDWIS/FED=No and EPA≠state File; 3 from Table 4-2) out of all violations found by EPA and not reported to SDWIS/FED (EPA=Yes and SDWIS/FED=No; 2 and 3 from Table 4-2).

! **Data Flow (DF) error rate** in the non-reported data describes how many non-reported violation data are as a result of reporting problems from state to SDWIS/FED, expressed as a percentage. This quantity is estimated based on the violations found by EPA, but not reported to SDWIS/FED and where the assigned violation in the state file confirmed by EPA (EPA=Yes, State File=Yes, and SDWIS/FED=No; 2a and 2b from Table 4-2) out of all violations found by EPA and not reported to SDWIS/FED (EPA=Yes and SDWIS/FED=No; 2 and 3 from Table 4-2).

! **Accuracy** of the data in SDWIS/FED describes how much of the violation data in SDWIS/FED are correct, expressed as a percentage. This quantity is estimated based on the violations found by EPA that agree with those reported to SDWIS/FED (EPA= SDWIS/FED; 1a, 1d, and 4a from Table 4-2) out of all violations reported to SDWIS/FED (SDWIS/FED=Yes; 1, 4, 5, 6 from Table 4-2).

! **Compliance Determination (CD) error rate** in SDWIS/FED describes how much of the violations data in SDWIS/FED are incorrect violations types as a result of errors in the state=s compliance determination, expressed as a percentage. This quantity is estimated based on the violations found by EPA that disagree with those reported to SDWIS/FED, but which are missing in the state file (State File=No and EPA≠SDWIS/FED from Table 4-2) or the violations found by EPA that disagree with those found by the state, which were then reported to SDWIS/FED as found by the state (EPA≠State File= SDWIS/FED; 1c and 4b[8] from Table 4-2) out of all violations reported to SDWIS/FED (SDWIS/FED=Yes; 1, 4, 5, 6 from Table 4-2).

! **Data Flow (DF) error rate** in SDWIS/FED describes how many of the reported violations data are incorrect violations types due to reporting problems from the state to SDWIS/FED, expressed as a percentage. This quantity is estimated based on the violations found in the state files and confirmed by EPA, but which disagree with those reported to SDWIS/FED (EPA=State≠SDWIS/FED from Table 4-2) or the violations found by EPA that disagree with those found by the state, which were then reported to

8 If DV auditors determined it to be a CD error

SDWIS/FED (EPA≠State File= SDWIS/FED;1b, 1e, and 4b[9] from Table 4-2) out of all violations reported to SDWIS/FED (SDWIS/FED=Yes; 1, 4, 5, 6 from Table 4-2).

! **False Positive rate** of the violation data in SDWIS/FED describes how much of the reported violation data in SDWIS are, in fact, false violations, expressed as a percentage. This quantity is estimated based on the violations not confirmed by EPA but reported to SDWIS/FED (EPA=No and SDWIS/FED=Yes; 5 and 6 from Table 4-2) out of all violations reported to SDWIS/FED (SDWIS/FED=Yes; 1, 4, 5, 6 from Table 4-2).

! **Overall Data Quality Estimate** in SDWIS/FED measures how many noncompliance determinations are correctly reported in SDWIS/FED among all noncompliance determinations (that are either violations or false-positive violations). This quantity is estimated based on the violations confirmed by EPA and correctly reported to SDWIS/FED (EPA=SDWIS/FED; 1a, 1d, and 4a from Table 4-2) out of all violations found by EPA or in the state files and SDWIS/FED (EPA=Yes or State File=Yes or SDWIS/FED=Yes; 1-6 from Table 4-2). When the false positive rate is 0%, this measure is the product of **Completeness** and **Accuracy**.

Since the DV states were not randomly selected, the states were treated as a fixed stratification variable for this analysis. During the DVs, there were systems that did not have any violations in the sample and did not require any reporting to SDWIS/FED. Thus, the actual number of sample systems used for the calculations was less than the number of sampled systems for the DVs. Furthermore, sub-domain analysis by rules or system types resulted in single-cluster strata and/or single observation in some clusters. A single-cluster stratum does not contribute in the calculation of variance estimates, which may underestimate the sampling errors. Therefore, the strata were combined within each EPA region except for overall data quality estimations, where the strata were combined within each DV state.

9 If DV auditors determined it to be a DF error.

Table 4-2: Violation data Comparison Categorization

Was a violation found?			Were the assigned violations in agreement?	Example	Description
Found By DV Auditors	**Found In state File**	**Reported to SDWIS**			
Yes	Yes	Yes	1a. DV Auditors=State File=SDWIS	A TCR violation 3100-21[1] was found in the state file, confirmed by DV auditors, and correctly reported to SDWIS/FED.	No discrepancy in SDWIS/FED
			1b. DV Auditors =State File≠SDWIS	A TCR violation record 3100-21 as found in state file and confirmed by DV auditors; the violation was incorrectly reported to SDWIS/FED as 3100-22[2].	Data Flow error
			1c. DV Auditors ≠State File=SDWIS	A TCR violation record 3100-22 was found in state file and reported to SDWIS/FED as 3100-22 when the violation should have been 3100-21.	Compliance determination error
			1d. DV Auditors =SDWIS≠State File	A TCR violation 3100-21 was reported to SDWIS/FED and confirmed by DV auditors but the state issued 3100-22 in the file.	No discrepancy in SDWIS/FED
			1e. DV Auditors ≠State File≠SDWIS	A TCR violation record 3100-22 found in state file when it should have been 3100-21 according to DV auditors, while the violation was incorrectly reported to SDWIS/FED as 3100-23[3].	Compliance determination error by state and Data flow error between state file and SDWIS/FED
Yes	Yes	No	2a. DV Auditors =State File	A TCR violation 3100-21 was found in the state file and confirmed by DV auditors, but not reported to SDWIS/FED.	Non-reporting; Data Flow error

[1] Acute TCR MCL violation.
[2] Monthly TCR MCL violation.
[3] Routine Major TCR Monitoring Violation

Table 4-2: Violation data Comparison Categorization

Was a violation found?			Were the assigned violations in agreement?	Example	Description
Found By DV Auditors	**Found In state File**	**Reported to SDWIS**			
			2b. DV Auditors ≠State File	A TCR violation 3100-22 was issued in the state file when it should have been 3100-21, and the violation was not reported to SDWIS/FED.	Non-reporting; Compliance determination error by the state; Data Flow error between the state file and SDWIS/FED.
Yes	No	No	3. N/A	There should have been a TCR violation 3100-21, but the state did not issue a violation and did not report to SDWIS/FED.	Non-reporting; Compliance determination error
Yes	No	Yes	4a. DV Auditors =SDWIS/FED	There should have been a TCR violation 3100-21 issued in the state file, but the notice of violation (NOV) was not found in the state file, even though the violation as correctly reported to SDWIS/FED.	No discrepancy in SDWIS/FED
			4b. DV Auditors ≠SDWIS/FED	There should have been a TCR violation 3100-21 issued in the state file, but NOV was not found in the state file, while the violation as incorrectly reported to SDWIS/FED as 3100-22.	Compliance determination error by state and/or Data Flow between state File and SDWIS/FED
No	Yes	Yes	5a. State File=SDWIS/FED	A TCR violation 3100-21 was issued in the state file and reported to SDWIS/FED, but it should not have been a violation.	False positive in SDWIS/FED
			5b. State File≠SDWIS/FED	A TCR violation 3100-21 was found in the state file, but the DV Auditors concluded that there should not have been a violation in the first place. In addition, the state reported a different TCR violation type (3100-22) to SDWIS FED.	False positive in SDWIS/FED

Table 4-2: Violation data Comparison Categorization

Was a violation found?			Were the assigned violations in agreement?	Example	Description
Found By DV Auditors	**Found In state File**	**Reported to SDWIS**			
No	No	Yes	6. N/A	A TCR violation 3100 was reported to SDWIS/FED, but DV Auditors concluded that there should not have been a violation in the first place. In addition, no evidence of a violation was found in the state files because the state rescinded a violation but has not removed it if from SDWIS/FED.	False positive in SDWIS/FED

4.2.1 Results from 2002-2004 Data Verifications

The proportion estimates and the sampling errors for the violation DQE by violation types are presented in Table 4-3. Eighty-one percent of the MCL and SWTR TT violations were reported to SDWIS/FED. Seventy-four percent of the non-reported violations were due to compliance determination errors and 26% were due to data flow errors. The reported violations in SDWIS/FED were accurate at 94%. Overall, the DQE of the violation data was 77%. This means that 77% of the noncompliance determinations on MCL/ SWTR TT standards were correctly reported in SDWIS/FED.

Considering all health-based violations (MCL and TT violations, which include Lead and Copper TT), 62 percent of the violations were reported to SDWIS/FED. This means that 38% of the violations were not reported. Eighty-four (84) percent of the non-reported violations were due to compliance determination errors and 16% were due to data flow errors. The reported violations in SDWIS/FED were accurate at 94%. Overall, the DQE of the health-based violation data was 59%. This means that 59% of the noncompliance determinations on all health-based standards were correctly reported in SDWIS/FED.

The quality of the health-based violations data was much lower than the MCL/SWTR TT data because of the quality of data associated with the Lead and Copper Rule. The data quality of the LCR TT violations was the lowest at 7.6%. For example, we found that out of 176 LCR TT violations, only 5 were reported to SDWIS/FED. The non-reporting was also mainly because of compliance determination errors. Specifically, 161 out of 171 violations were not recognized as violations when the violations had occurred.

Twenty-nine percent of the M/R violations were reported to SDWIS/FED and 71% of the violations were not reported. Ninety-two percent of the non-reported violations were due to compliance determination errors and 8% were due to data flow errors. The reported M/R violations in SDWIS/FED were accurate at 88%. Overall, the DQE of the M/R violation data was 27%, i.e., 27% of the noncompliance determinations on M/R were correctly reported in SDWIS/FED.

Table 4-3: Data Quality Estimates (DQE) by Violation Type

		TCR MCL		OTHER MCL		TOTAL MCL		SWTR TT	
% COMPLETENESS OF SDWIS/FED		83.29% (+/-9.66%)		48.94% (+/-27.05%)		78.42% (+/-9.39%)		94.89% (+/-8.03%)	
%NON-REPORTING ON SDWIS/FED		16.71% (+/-9.66%)		51.06% (+/-27.05%)		21.58% (+/-9.39%)		5.11% (+/-8.03%)	
% CD ERROR ON NON-REPORTED DATA	% DF ERROR ON NON-REPORTED DATA	83.40% (+/- 15.07%)	16.60% (+/- 15.07%)	56.89% (+/- 43.12%)	43.11% (+/- 43.12%)	73.84% (+/- 16.98%)	26.16% (+/- 16.98%)	73.87% (+/- 43.38%)	26.87% (+/- 43.38%)
%ACCURACY OF DATA IN SDWIS/FED		96.65% (+/-2.62%)		79.22% (+/-20.56%)		94.91% (+/-3.00%)		91.07% (+/-12.14%)	
%CD ERROR WITH DATA IN SDWIS/FED		0%		5.73% (+/-8.48%)		0.57% (+/-0.80%)		3.98%(+/-7.26)	
%DF ERROR WITH DATA IN SDWIS/FED		0%		0%		0%		0%	
%FALSE POSITIVE DATA IN SDWIS/FED		2.26% (+/-2.44%)		15.05% (+/-17.96%)		4.52% (+/-2.90%)		4.95% (+/-8.38%)	
OVERALL DATA QUALITY		80.95% (+/-9.47%)		42.00% (+/-26.67%)		75.16% (+/-9.26%)		86.63% (+/-15.79%)	
		MCL/SWTR TT		**LCR TT**		**Health-Based Violations**		**M/R**	
% COMPLETENESS OF SDWIS/FED		81.33% (+/-8.86%)		7.6% (+/-7.52%)		61.69% (+/-10.26%)		29.02% (+/-3.47%)	
%NON-REPORTING ON SDWIS/FED		18.67% (+/-8.86%)		92.40% (+/-7.52%)		38.31% (+/-10.26%)		70.98% (+/-3.47%)	
% CD ERROR ON NON-REPORTED DATA	%DF ERROR ON NON-REPORTED DATA	73.84% (+/- 16.30%)	26.16% (+/- 16.30%)	91.76% (+/- 11.80)	8.24% (+/- 11.80)	84.45% (+/- 10.35%)	15.55% (+/- 10.35%)	92.03% (+/-1.75%)	7.97% (+/-1.75%)
%ACCURACY OF DATA IN SDWIS/FED		94.12% (+/-3.27%)		100%		94.30% (+/-3.16%)		88.35% (+/-2.87%)	
%CD ERROR WITH DATA IN SDWIS/FED		1.27% (+/-1.44%)		0%		1.24% (+/-1.39%)		3.18% (+/-1.54%)	
%DF ERROR WITH DATA IN SDWIS/FED		0%		0%		0%		0.99% (+/-0.53%)	
%FALSE POSITIVE DATA IN SDWIS/FED		4.60% (+/-2.90%)		0%		2.79% (+/-2.38%)		7.48% (+/-2.31%)	
OVERALL DATA QUALITY		77.21% (+/-8.99%)		7.6% (+/-7.52%)		59.18% (+/-10.10%)		27.08% (+/-3.37%)	

*CD=Compliance determination
*DF=Data Flow

Note: TCR MCL + Other MCL = Total MCL + SWTR TT = MCL/SWTR TT + LCR TT = Health-Based Violations. M/R = monitoring and reporting violations.

In general, the majority of non-reported data were due to compliance determination errors, i.e., the states did not issue violations when violations had occurred. The violations had not been recognized, not recorded by states as violations, and consequently, not reported to SDWIS/FED. We need to further examine the cause of such compliance determination errors. These errors may be due to late reporting or rule interpretation discrepancies. Eliminating these errors will significantly increase the completeness of the data in SDWIS/FED. For example, 84% of the non-reported health-based based violations were due to compliance determination errors. If these errors did not occur, the completeness of health-based violations in SDWIS/FED would be at 94% (62%+38%×84%). Similarly, the completeness of M/R violations would also be at 94% (29%+71%×92%).

The violation data are further evaluated by system type in Tables 4-4a-c. The DQEs of MCL/SWTR TT violations were not significantly different among the different system types. Likewise, the DQEs of health-based violations were not significantly different between CWSs and NTNCWSs. (The DQE of health-based violations for TNCWSs was not calculated since LCR TT data were not collected for TNCWSs.)

Table 4-4a: MCL/SWTR TT Violations Data Quality Estimates (DQE) by Public Water System Type

		CWS		NTNCWS		TNCWS	
% COMPLETENESS OF SDWIS/FED		78.87% (+/-10.59%)		83.07% (+/-12.32%)		83.25% (+/-15.53%)	
%NON-REPORTING IN SDWIS/FED		21.13% (+/-10.59%)		16.93% (+/-12.32%)		16.75% (+/-15.53%)	
%CD ERROR ON NON-REPORTED DATA	%DF ERROR ON NON-REPORTED DATA	61.24% (+/-25.44%)	38.76% (+/-25.44%)	49.15% (+/-35.34%)	50.85% (+/-35.34%)	96.21% (+/-6.14%)	3.79% (+/-6.14%)
%ACCURACY OF DATA IN SDWIS/FED		93.15% (+/-4.35%)		83.42% (+/-13.55%)		97.37% (+/-3.68%)	
%CD ERROR WITH DATA IN SDWIS/FED		1.06% (+/-1.47%)		6.62% (+/-10.75%)		0.29% (+/-0.60%)	
%DF ERROR WITH DATA IN SDWIS/FED		0%		0%		0%	
%FALSE POSITIVE DATA IN SDWIS/FED		5.80% (+/-4.13%)		9.96% (+/-11.71%)		2.34% (+/-3.60%)	
OVERALL DATA QUALITY		74.37% (+/-10.73%)		70.49% (+/-13.61%)		81.38% (+/-15.51%)	

*CD=Compliance determination
*DF=Data Flow

Table 4-4b: Health-Based Violations Data Quality Estimates (DQE) by Public Water System Type

		CWS		NTNCWS	
% COMPLETENESS OF SDWIS/FED		53.39% (+/-12.17%)		40.86% (+/-15.19%)	
%NON-REPORTING IN SDWIS/FED		46.61% (+/-12.17%)		59.14% (+/-15.19%)	
%CD ERROR ON NON-REPORTED DATA	%DF ERROR ON NON-REPORTED DATA	78.45% (+/-16.17%)	21.55% (+/-16.17%)	93.57% (+/-10.15%)	6.57% (+/-10.15%)
%ACCURACY OF DATA IN SDWIS/FED		93.47% (+/-4.15%)		84.95% (+/-12.58%)	
%CD ERROR WITH DATA IN SDWIS/FED		1.01% (+/-1.40%)		6.01% (+/-9.86%)	
%DF ERROR WITH DATA IN SDWIS/FED		0%		0%	
%FALSE POSITIVE DATA IN SDWIS/FED		5.52% (+/-3.94%)		9.04% (+/-10.76%)	
OVERALL DATA QUALITY		51.22% (+/-11.75%)		36.67% (+/-13.02%)	

*CD=Compliance determination
*DF=Data Flow

Table 4-4c: M/R Violations Data Quality Estimates (DQE) by Public Water System Type

		CWS		NTNCWS		TNCWS	
% COMPLETENESS OF SDWIS/FED		20.26% (+/-3.42%)		22.65% (+/-5.26%)		45.89% (+/-5.87%)	
%NON-REPORTING IN SDWIS/FED		79.74% (+/-3.42%)		77.35% (+/-5.26%)		54.11% (+/-5.87%)	
%CD ERROR ON NON-REPORTED DATA	%DF ERROR ON NON-REPORTED DATA	91.62% (+/-2.36)	8.38% (+/-2.36)	87.05% (+/-5.07%)	12.95% (+/-5.07%)	96.87% (+/-2.31)	3.13% (+/-2.31)
%ACCURACY OF DATA IN SDWIS/FED		82.10% (+/-4.96%)		81.39% (+/-7.93%)		94.81% (+/-2.71%)	
%CD ERROR WITH DATA IN SDWIS/FED		5.25% (+/-2.84%)		4.10% (+/-2.7%)		1.83% (+/-1.99%)	
%DF ERROR WITH DATA IN SDWIS/FED		0.73% (+/-0.74%)		1.45% (+/-2.15%)		0.60% (+/- 0.59%)	
%FALSE POSITIVE DATA IN SDWIS/FED		11.92% (+/-4.42%)		13.05% (+/-7.88%)		2.75% (+/-1.77%)	
OVERALL DATA QUALITY		18.38% (+/-3.12)		20.51% (+/-4.77%)		44.17% (+/-5.79%)	

*CD=Compliance determination
*DF=Data Flow

EPA has public notification (PN) requirements to ensure that the public is notified of violations in a timely manner. The PN requirements define three tiers of notification that are based on the public health significance of the violation, with tier 1 being the most significant (See Appendix C for the definition of PN tiers). The DQEs are also calculated by PN tier groups of violations in Table 4-5. Two-thirds of PN tier 1 violations were reported to SDWIS/FED. There were no significant differences in DQEs between PN tier 1 and PN tier 2. The DQEs for PN tier 3, which mostly consisted of M/R violations, were significantly lower than those for PN tier 1 and PN tier 2. Less than two-thirds of PN tier 2 violations were reported to SDWIS/FED and only 30% of PN tier 3 violations were reported to SDWIS/FED. In all PN tier groups, the data in SDWIS/FED were highly accurate. The overall data quality does not reflect false-positive violations in SDWIS/FED since they can not be categorized into a PN tier.

Table 4-5: Data Quality (DQ) by PN Tier

		PN Tier 1		PN Tier 2		PN Tier 3	
% COMPLETENESS OF SDWIS/FED		66.97% (+/-22.37%)		62.40% (+/-10.50%)		30.58% (+/-4.15%)	
%NON-REPORTING in SDWIS/FED		33.03% (+/-22.37%)		37.60% (+/-10.50%)		69.42% (+/-4.15%)	
%CD ERROR ON NON-REPORTED DATA	%DF ERROR ON NON-REPORTED DATA	32.77% (+/-23.93%)	67.23% (+/-23.93%)	87.21% (+/-9.04%)	12.79% (+/-9.04%)	91.44% (+/-1.87%)	8.56% (+/-1.87%)
%ACCURACY OF DATA IN SDWIS/FED		100%		98.65% (+/-1.52%)		95.46% (+/-1.82)	
%CD ERROR WITH DATA IN SDWIS/FED		0%		1.35% (+/-1.52 %)		3.46%(+/-1.68)	
%DF ERROR WITH DATA IN SDWIS/FED		0%		0%		1.08%(+/-0.58)	
OVERALL DATA QUALITY		66.97% (+/-22.37%)		61.56% (+/-10.59%)		29.19% (+/-4.13%)	

*CD=Compliance determination
*DF=Data Flow

4.2.2 Results from 1999-2001 Data Verifications

This section presents DQEs from 1999 to 2001 data verification audits recalculated using the current statistical methodology described in Section 3.3. The states subjected to the DV audits during 1999-2001 are shown in Table 4-6. In the calculation, the DV results from Region 2 were not included since the state DV reports were not finalized for those states during the period of this analysis. The DQEs are included in Tables 4-7a and b. Because these estimates were computed based on a different set of DV states in a different data quality assessment time frame and with a different statistical sample design, it is not scientifically valid to make a national inference by comparing the results between Table 4.3a-b. However, the DQEs from those states that had repeated DV audits during both assessment periods are calculated and compared in the following section.

Table 4-6: States Subject to Data Verifications from 1999-2001

Region	States		Region	States
1	MA, ME, NH		6	AR, LA, NM, TX
2	NY, PR		7	KS, MO, NE
3	VA, PA, DE		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 4-7b shows that 69% of the MCL and SWTR TT violations were reported to SDWIS/FED. Seventy-nine percent of the non-reported violations were due to compliance determination errors and 21% were due to data flow errors. The reported violations in SDWIS/FED were accurate at 91%. Overall, the DQE of the violation data was 64%. This means that 64% of the noncompliance determinations on MCL/ SWTR TT standards were correctly reported in SDWIS/FED.

Table 4-7a: 1999-2001 Data Quality Estimates (DQE) for MCL and SWTR TT

		TCR MCL		OTHER MCL		TOTAL MCL	
% COMPLETENESS OF SDWIS/FED		76.71% (+/-9.98%)		63.33% (+/-26.99%)		74.81% (+/-9.34%)	
%NON-REPORTING ON SDWIS/FED		23.29% (+/-9.98%)		36.67% (+/-26.99%)		25.19% (+/-9.34%)	
%CD ERROR ON NON-REPORTED DATA	%DF ERROR ON NON-REPORTED DATA	70.69% (+/-18.58%)	29.31% (+/-18.58%)	68.55% (+/-31.68%)	31.45% (+/-31.68%)	70.25% (+/-16.15%)	29.75% (+/-16.15%)
%ACCURACY OF DATA IN SDWIS/FED		91.71% (+/-2.62%)		63.99% (+/-46.70%)		88.54% (+/-8.87%)	
%CD ERROR WITH DATA IN SDWIS/FED		1.61% (+/-2.96%)		36.01% (+/-46.70%)		5.54% (+/-7.82%)	
%DF ERROR WITH DATA IN SDWIS/FED		0.64% (+/-1.26)		0%		0.56% (+/-7.82%)	
%FALSE POSITIVE DATA IN SDWIS/FED		6.05% (+/-5.24%)		0% (+/-17.96%)		5.36% (+/-4.66%)	
OVERALL DATA QUALITY		71.35% (+/-10.16%)		40.52% (+/-28.53%)		67.14% (+/-10.27%)	

*CD=Compliance determination
*DF=Data Flow

Table 4-7b: 1999-2001 Data Quality Estimates (DQE) for MCL/ SWTR TT and MR

		SWTR TT		MCL/SWTR TT		MR	
% COMPLETENESS OF SDWIS/FED		54.54% (+/-11.79)		69.39% (+/-9.59%)		34.86% (+/-4.59%)	
%NON-REPORTING ON SDWIS/FED		45.46% (+/-11.75)		30.61% (+/-9.59%)		65.14% (+/-4.59%)	
% CD ERROR ON NON-REPORTED DATA	% DF ERROR ON NON-REPORTED DATA	92.43% (+/-13.88)	7.57% (+/-13.88)	79.05% (+/-14.44%)	20.95% (+/-14.44%)	92.26% (+/-2.64%)	7.74% (+/-2.64%)
%ACCURACY OF DATA IN SDWIS/FED		100%		90.84% (+/-7.38%)		91.85% (+/-2.89%)	
%CD ERROR WITH DATA IN SDWIS/FED		0%		4.43% (+/-6.34%)		1.08% (+/-0.93%)	
%DF ERROR WITH DATA IN SDWIS/FED		0%		0.45% (+/-0.9%)		0.20% (+/-0.25%)	
%FALSE POSITIVE DATA IN SDWIS/FED		0%		4.28% (+/-3.8%)		6.87% (+/-2.58%)	
OVERALL DATA QUALITY		54.54% (+/-11.75%)		63.88% (+/-9.12%)		33.51% (+/-4.5%)	

*CD=Compliance determination
*DF=Data Flow

Thirty-five percent of the M/R violations were reported to SDWIS/FED and 65% of the violations were not reported. Ninety-two percent of the non-reported violations were due to compliance determination errors and 8% were due to data flow errors. The reported M/R violations in SDWIS/FED were accurate at 92%. Overall, the DQE of the M/R violation data was 33%, i.e., 33% of the noncompliance determinations on M/R were correctly reported in SDWIS/FED.

4.2.3 Data Quality Estimates from 1999-2001 and 2002-2004

In order to evaluate the progress of the data quality improvement, the DQEs from the states where the DV audits were conducted during the data quality assessment period 1999-2001 and again during 2002-2004 were calculated for the purpose of comparison. The states with repeated DV audits for both assessment periods can be identified from Table 2-1 and Table 4-6 and are listed in Table 4-8. Since the LCR was not reviewed during the 1999-2001 DVs, the data from the LCR were excluded from 2002-2004 DV results for this evaluation

The DQEs from these 18 states are presented in Tables 4.9 a and b, which include point estimates as well as the lower and upper bounds for 95% confidence intervals. In order to determine any significant differences (increase or decrease) in the DQEs, the two confidence intervals, defined by the lower and upper bounds as the end points of the interval, for the two DQEs should not overlap. Sixty-seven percent of MCL/SWTR TT violations with a 95% confidence interval (55%, 79%) were reported to SDWIS/FED during 1999-2001. Similarly, 80% of MCL/SWTR TT violations with a 95% confidence interval (68%, 92%) were reported to SDWIS/FED during 2002-2004. Since the confidence intervals overlap, there was no statistically

significant increase in the reporting of violations for these 18 states from 1999-2001 to 2002-2003. The overall data quality of MCL/SWTR TT violations was 64% with a 95% confidence interval (52%, 75%) during 1999-2001 and 75% with a 95% confidence interval (64%, 87%) during 2002-2004. Based on the confidence intervals, there was no statistically significant increase in the overall data quality of MCL/SWTR TT violations for these 18 states from 1999-2001 to 2002-2003.

On the other hand, approximately, 60% of SWTR TT violations with a 95% confidence interval (44%, 76%) were reported to SDWIS/FED during 1999-2001. During 2002-2004, 93% of SWTR TT violations with a 95% confidence interval (81%, 100%) were reported to SDWIS/FED. Since the confidence intervals do not overlap, there was a statistically significant increase in the reporting of violations for these 18 states from 1999-2001 to 2002-2003. However, the accuracy of SWTR TT has decreased significantly from 100% to 78%. The overall data quality of SWTR TT violations was 60% with a 95% confidence interval (44%, 73%) during 1999-2001 and 74% with a 95% confidence interval (54%, 94%) during 2002-2004. Therefore, there was no statistically significant increase in the overall data quality of SWTR TT violations for these 18 states from 1999-2001 to 2002-2003.

In general, all the confidence intervals from the two periods overlap for all DQEs, except for SWTR TT violations completeness DQE. Therefore, there were no statistically significant increases or decreases in the DQEs for these states from 1999-2001 to 2002-2003 assessment.

Table 4-8: States Subject to Data Verifications during 1999-2001 and 2002-2004

Region	States		Region	States
1	MA		6	AR, NM, TX
2			7	KS, NE
3	VA, PA		8	UT
4	FL, KY, MS, NC, SC, TN		9	
5	IL, OH		10	AK, ID

Table 4-9a: Data Quality Estimates (DQE) for MCL from MA, VA, PA, FL, KY, MS, NC, SC, TN, IL, OH, AR, NM, TX, MO, UT, AK, ID During 1999-2001 and 2002-2004

			TCR MCL				Other MCL				Total MCL			
Year			**1999-2001**		**2002-2004**		**1999-2001**		**2002-2004**		**1999-2001**		**2002-2004**	
% COMPLETENESS OF SDWIS/FED		POINT ESTIMATE	70.13%		82.17%		74.57%		60.74%		70.63%		79.17%	
		LOWER BOUND	56.22%		67.46%		48.25%		30.84%		57.87%		65.95%	
		UPPER BOUND	84.03%		96.88%		100%		90.64%		83.39%		92.39%	
%NON-REPORTING ON SDWIS/FED		POINT ESTIMATE	29.87%		17.83%		25.43%		39.26%		29.37%		20.83%	
		LOWER BOUND	15.97%		3.12%		0%		9.36%		16.61%		7.61%	
		UPPER BOUND	43.78%		32.54%		51.75%		69.16%		42.13%		34.05%	
%CD ERROR ON NON-REPORTED DATA	%DF ERROR ON NON-REPORTED DATA	POINT ESTIMATE	74.90%	25.10%	83.54%	16.46%	58.02%	41.98%	21.14%	78.86%	73.25%	26.75%	67.09%	32.91%
		LOWER BOUND	51.93%	2.13%	61.95%	0%	12.32%	0%	0%	49.66%	51.66%	5.16%	38.77%	4.59%
		UPPER BOUND	97.87%	48.07%	100%	38.05%	100%	87.68%	50.34%	100%	94.85%	48.34%	95.41%	61.23%
%ACCURACY OF DATA IN SDWIS/FED		POINT ESTIMATE	90.49%		96.01%		100%		82.84%		91.56%		94.40%	
		LOWER BOUND	81.28%		91.88%		100%		59.65%		83.40%		89.88%	
		UPPER BOUND	99.71%		100%		100%		100%		99.72%		98.920%	
%CD ERROR WITH DATA IN SDWIS/FED		POINT ESTIMATE	3.04%		0%		0%		0%		2.70%		0%	
		LOWER BOUND	0%		0%		0%		0%		0%		0%	
		UPPER BOUND	8.99%		0%		0%		0%		7.98%		0%	
%DF ERROR WITH DATA IN SDWIS/FED		POINT ESTIMATE	0%		0%		0%		0%		0%		0%	
		LOWER BOUND	0%		0%		0%		0%		0%		0%	
		UPPER	0%		0%		0%		0%		0%		0%	

Table 4-9a: Data Quality Estimates (DQE) for MCL from MA, VA, PA, FL, KY, MS, NC, SC, TN, IL, OH, AR, NM, TX, MO, UT, AK, ID During 1999-2001 and 2002-2004

		TCR MCL		Other MCL		Total MCL	
Year		**1999-2001**	**2002-2004**	**1999-2001**	**2002-2004**	**1999-2001**	**2002-2004**
	BOUND						
%FALSE POSITIVE DATA IN SDWIS/FED	POINT ESTIMATE	6.47%	3.99%	0%	17.16%	5.74%	5.60%
	LOWER BOUND	0%	0%	0%	0%	0%	1.08%
	UPPER BOUND	13.54%	8.12%	0%	40.35%	12.00%	10.12%
OVERALL DATA QUALITY	POINT ESTIMATE	64.71%	79.46%	74.57%	53.95%	65.78%	75.62%
	LOWER BOUND	50.90%	65.09%	48.25%	24.49%	53.02%	62.79%
	UPPER BOUND	78.52%	93.82%	100%	83.42%	78.54%	88.46%

*CD=Compliance determination
*DF=Data Flow

Table 4-9b: Data Quality Estimates (DQE) for SWTR TT, MCL/SWTR TT, and MR from MA, VA, PA, FL, KY, MS, NC, SC, TN, IL, OH, AR, NM, TX, MO, UT, AK, ID During 1999-2001 and 2002-2004

			SWTR TT				MCL/SWTR TT				MR			
Year			1999-2001		2002-2004		1999-2001		2002-2004		1999-2001		2002-2004	
% COMPLETENESS OF SDWIS/FED		POINT ESTIMATE	59.95%		93.30%		66.73%		80.36%		37.74%		28.06%	
		LOWER BOUND	43.68%		80.83%		54.74%		68.10%		31.69%		23.82%	
		UPPER BOUND	76.21%		100%		78.72%		92.62%		43.79%		32.31%	
%NON-REPORTING ON SDWIS/FED		POINT ESTIMATE	40.05%		6.70%		33.27%		19.64%		62.26%		71.94%	
		LOWER BOUND	43.68%		0%		21.28%		7.38%		56.21%		67.69%	
		UPPER BOUND	56.32%		19.16%		45.26%		31.90%		68.31%		76.18%	
%CD ERROR ON NON-REPORTED DATA	%DF ERROR ON NON-REPORTED DATA	POINT ESTIMATE	93.23%	6.75%	89.09%	10.91%	82.04%	17.96%	67.72%	32.28%	91.63%	8.37%	93.59%	6.41%
		LOWER BOUND	76.38%	0%	62.11%	0%	63.88%	0%	40.40%	4.95%	88.91%	5.66%	91.60%	4.43%
		UPPER BOUND	100%	23.62%	100%	37.89%	100%	36.12%	95.05%	59.60%	94.34%	11.09%	95.57%	8.40%
%ACCURACY OF DATA IN SDWIS/FED		POINT ESTIMATE	100%		78.78%		94.22%		92.90%		92.01%		92.2%	
		LOWER BOUND	100%		60.28%		88.26%		88.24%		88.90%		89.33%	
		UPPER BOUND	100%		97.28%		100%		97.55%		95.11%		95.08%	
%CD ERROR WITH DATA IN SDWIS/FED		POINT ESTIMATE	0%		17.01%		1.85%		1.63%		1.44%		2.01%	
		LOWER BOUND	0%		0%		0%		0%		0.13%		0.37%	
		UPPER BOUND	0%		39.27%		5.53%		4.11%		2.75%		3.66%	
%DF ERROR WITH DATA IN SDWIS/FED		POINT ESTIMATE	0%		0%		0%		0%		0.28%		0.64%	
		LOWER BOUND	0%		0%		0%		0%		0%		0.05%	
		UPPER BOUND	0%		0%		0%		0%		0.63%		1.23%	

Table 4-9b: Data Quality Estimates (DQE) for SWTR TT, MCL/SWTR TT, and MR from MA, VA, PA, FL, KY, MS, NC, SC, TN, IL, OH, AR, NM, TX, MO, UT, AK, ID During 1999-2001 and 2002-2004

		SWTR TT		MCL/SWTR TT		MR	
Year		**1999-2001**	**2002-2004**	**1999-2001**	**2002-2004**	**1999-2001**	**2002-2004**
%FALSE POSITIVE DATA IN SDWIS/FED	POINT ESTIMATE	0%	4.21%	3.93%	5.57%	6.27%	5.15%
	LOWER BOUND	0%	0%	0%	0.35%	3.57%	2.89%
	UPPER BOUND	0%	10.00%	8.44%	9.58%	8.98%	7.40%
OVERALL DATA QUALITY	POINT ESTIMATE	59.95%	73.71%	63.71%	75.46%	36.14%	26.87%
	LOWER BOUND	43.68%	53.90%	52.70%	63.60%	30.19%	22.71%
	UPPER BOUND	76.21%	93.52%	74.71%	87.33%	42.08%	31.03%

*CD=Compliance determination
*DF=Data Flow

4.3 Analysis of 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 has decided not to obtain approval to implement the federal program (e.g. Wyoming, the District of Columbia and on Indian lands).

Enforcement data reported to SDWIS were compared to those found in the state files during the DV. The proportion of enforcement data in the state files that were in agreement with those reported to SDWIS/FED (1a, 1c, and 5a from Table 4-2) were estimated as described in Section 3.3 and presented in Table 4-8. The overall DQE for enforcement data was 86%.

Table 4-8: Proportion Estimates of Enforcement Data in State Files reported to SDWIS/FED without discrepancy

PWS Type	CWS	NTNCWS	TNCWS	Overall
Proportion Estimate	73.14% (+/-9.65%)	76.25% (+/-6.88%)	94.92% (+/-2.72%)	85.97% (+/-3.62)

5. Analysis of Timeliness of Violation Reporting in SDWIS/FED

In this section, the results from the analysis of the data in SDWIS/FED are presented. This analysis evaluates the timeliness of violations based on compliance period end date, which provides a benchmark for comparison between fiscal years. Violations are due to be reported by the end of the following quarter after awareness or the compliance period end date.

Timeliness is calculated as the ratio of the number of violations reported on time and the baseline number of violations that should be reported, i.e.,

$$\text{Timeliness} = \frac{\text{Number of Violations Reported on Time}}{\text{Number of Violations Reported Baseline}}$$

where **on time** is defined as by the end of the following quarter after the compliance period end date and **baseline** is a point in time in the future (in this case, between 4 and 7 quarters after violations are due to be reported). Basically, the Timeliness is the proportion of violations that were eventually reported to SDWIS/FED on time.

To compute the timeliness, the violation data were extracted from archived SDWIS/FED databases for each of five fiscal years (2000-2004). The violations were then grouped by PWS

ID, fiscal year, quarter, violation code, contaminant code, and basic PWS attributes, and the on-time and baseline violations were summed. Table 5-1 shows the database extracted for the analysis. The database does not include LCR or other violations with open-ended compliance period end dates. For these violations, the compliance period end date is open until the system returns to compliance.

Table 5-1: SDWIS/FED Database Analyzed for Timeliness

FY2000 on time:	**Violations with end dates between:**			**Archive Date**
00Q1	10/1/99	and	12/31/99	4/00
00Q2	1/1/00	and	3/3100	7/00
00Q3	4/1/00	and	6/30/00	10/00
00Q4	7/1/00	and	9/30/00	1/01
FY2000 **baseline:**	01Q4 tables, Archived 1/02			
FY2001 on time:	**Violations with end dates between:**			**Archive Date**
01Q1	10/1/00	and	12/31/00	4/01
01Q2	1/1/01	and	3/31/01	7/01
01Q3	4/1/01	and	6/30/01	10/01
01Q4	7/1/01	and	9/30/01	1/02
FY2001 **baseline:**	02Q4 tables, Archived 1/03			
FY2002 on time:	**Violations with end dates between:**			**Archive Date**
02Q1	10/1/01	and	12/31/01	4/02
02Q2	1/1/02	and	3/3102	7/02
02Q3	4/1/02	and	6/30/02	10/02
02Q4	7/1/02	and	9/30/02	1/03
FY2002 **baseline:**	03Q4 tables, Archived 1/04			
FY2003 on time:	**Violations with end dates between:**			**Archive Date**
03Q1	10/1/02	and	12/31/02	4/03
03Q2	1/1/03	and	3/3103	7/03
03Q3	4/1/03	and	6/30/03	10/03
03Q4	7/1/03	and	9/30/03	1/04
FY2003 **baseline:**	04Q4 tables, Archived on 1/05			
FY2004 on time:	**Violations with end dates between:**			**Archive Date**
04Q1	10/1/03	and	12/31/03	4/04
04Q2	1/1/04	and	3/31/04	7/04
04Q3	4/1/04	and	6/30/04	10/04
04Q4	7/1/04	and	9/30/04	1/05
FY2004 **baseline:**	05Q4 tables, Archived on 1/06			

Table 5-2: Violation Reporting Timeliness to SDWIS/FED by Violation Type

Fiscal Year	2000	2001	2002	2003	2004
Number of Violations Reported on Time					
TCR MCL	7,738	8,114	7,977	7,902	7,421
Other MCL	727	652	771	1,106	1,273
SWTR TT	932	918	1,045	774	540
Health-Based Violations[13]	9,397	9,684	9,793	9,831	9,308
M/R	49,782	50,868	55,425	61,967	32,742
Number of Violations Reported for Baseline					
TCR MCL	11,445	10,963	10,795	10,821	10,510
Other MCL	1,344	1,315	1,844	2,573	3,716
SWTR TT	1,574	1,627	1,585	1,252	932
Health-Based Violations[13]	14,636	13,905	14,369	14,996	15,513
M/R	93,231	111,397	121,819	106,664	104,427
Percent Timeliness					
TCR MCL	68%	74%	74%	73%	71%
Other MCL	54%	50%	42%	43%	34%
SWTR TT	59%	56%	66%	62%	58%
Health-Based Violations[13]	65%	70%	68%	66%	60%
M/R	53%	46%	45%	58%	31%

Table 5-2 shows the computed timeliness of the reported violations in SDWIS/FED. Late reporting can have an impact on the reliability of SDWIS/FED in informing the public and stakeholders about the quality of their drinking water. Further, it hinders our effort to assess the public health risk and address the violations with enforcement actions in a timely manner. In 2004, 60% of the health-based violations were reported on time, while only 31% of the M/R violations were reported on time. Note that there is a 27% decline in timeliness for the M/R violations from 2003.

Additional information (in the form of pivot tables) is available from EPA upon request that provides additional details on the timeliness in which violations are reported across several additional attributes. Additional findings based on this information are the following:

- Timeliness of reported health-based violations was similar across water system types.
- Monitoring violations for TNCWSs was highest at 58%, and lowest for NTNCWSs at 33%.

13 These heath-based violations do not include Lead and Copper Treatment Technology (LCR TT) violations because of they have open-ended compliance period end dates.

! Timeliness was similar across quarters.

! Timeliness generally decreased as system size decreased.

! It was difficult to evaluate the timeliness of reported violations for new rules, because many of the violations in these rules have open-ended compliance period end dates.

6. Conclusion

For the 38 states evaluated from 2002 to 2004, most of the reported violations in SDWIS/FED were accurate at 90%. Approximately 81% of the MCL and SWTR TT violations were reported to SDWIS/FED. Sixty-two percent of the health-based violations (including LCR TT violations) and 39% of the monitoring and reporting violations were reported. Non-reporting was mostly attributable to the fact that states did not issue violations when violations had occurred. In other words, the violations were not recognized, not recorded by the states as violations, and consequently, not reported to SDWIS/FED. Eighty-four percent of non-reported health-based violations and 92% of non-reported M/R violations were due to compliance determination errors.

EPA considers non-reported violations to be a serious problem that could have public health implications at many levels. The information and the analyses based on such incomplete data in SDWIS/FED compromises our ability to determine if and when we need to take action against non-compliant systems, to oversee and evaluate the effectiveness of state and federal programs and regulations, to alleviate burden on states, and to determine whether new regulations are needed to further protect public health. Further, our response to public inquiries and preparing national reports on the quality of drinking water in a thorough and complete manner will be severely limited.

Some of the discrepancies between the number of violations that should have appeared in SDWIS/FED and those found by the DV auditors could have included differences in rule interpretation in light of the flexibility provided to states in implementing rules under state primacy agreements. The state implementation of rules must be at least as stringent as the Federal regulations, but can differ in substantial respects within a reasonable scope of the regulation. It is critical that EPA and the states continue to work together toward reducing non-reporting, reporting errors, and late reporting of violations.

Additional findings included the DQEs of health- based violations were not significantly different between CWSs and NTNCWSs. The DQEs on M/R violations for TNCWSs were significantly higher than those for CWSs and NTNCWSs.

Further, the DQEs from 18 states where the DV audits were conducted during the data quality assessment period of 1999-2001 and again during 2002-2004 were calculated for the purpose of comparison. For those states, 67% of MCL/SWTR TT violations with a 95% confidence interval (55%, 79%) were reported to SDWIS/FED during 1999-2001. Similarly, 80% of MCL/SWTR

TT violations with a 95% confidence interval (68%, 92%) were reported to SDWIS/FED during 2002-2004. Since the confidence intervals overlap, there was no statistically significant increase in the reporting of violations for these 18 states from 1999-2001 to 2002-2003. The overall data quality of MCL/SWTR TT violations was 64% with a 95% confidence interval (52%, 75%) during 1999-2001 and 75% with a 95% confidence interval (64%, 87%) during 2002-2004. Based on the confidence intervals, there was no statistically significant increase in the overall data quality of MCL/SWTR TT violations for these 18 states from 1999-2001 to 2002-2003.

Finally, 60% of MCL/SWTR TT violations were reported on time and approximately 30% of the MR violations were reported on time to SDWIS/FED in 2004.

7. Data Reliability Improvement Action Plan

Based on this analysis and on the results of previous efforts, EPA, working with its state co-regulators through the Association of State Drinking Water Administrators, has developed a Data Reliability Improvement Action Plan ("the plan") designed to achieve a data quality goal of 90 percent complete and accurate data for health-based violation reporting. The plan covers the years 2007 through 2009 and addresses improving data quality for monitoring and reporting violations and inventory (water systems' facilities) data. Principally, the plan focuses on actions that EPA and states can take to address compliance determination issues and thereby improve violation data quality. Progress toward accomplishment of the data quality goal will be measured annually and assessed in 2009. The plan appears in Appendix A.

8. Future Analysis of Data Reliability

Several factors will change both the process and the results of the data verifications and the data quality calculation for drinking water data. In the near term, the selection of states for DVs will be based on probability sampling from 2005. Specifically, the selection of states for the data verifications from 2005 to 2007 will be based on a probability sampling method, with every state being selected in a 4-year time frame. This will allow the data quality to be assessed nationally for rolling multiple-years. In the longer term (2008 and beyond), EPA is evaluating the feasibility of electronic data verification (EDV), which would collect and evaluate compliance sample results of regulated contaminants electronically for all CWSs. EPA believes that the most cost-effective and complete process of evaluating data quality in the long term may be via the EDV process. In each state, we can evaluate the data once every one or two years through the compliance determination processes recorded in the SDWIS/STATE software. SDWIS/STATE is already designed and developed for states to manage their drinking water programs. The advantages to this approach are that the software already exists and all compliance determinations are available for evaluation. The current DV process relies on a sample of systems, and due to the inherently small number of large CWSs, the large CWSs are not well represented in the samples. The EDV will allow us to use all systems instead of relying on a sample from a DV. Additionally, the drinking water administrators in decentralized states can

have hands-on data in one location instead of going to regional drinking water offices. All states using SDWIS/STATE will have the capability to calculate data quality in near real time and take action on issues as they arise. Furthermore, EDV will allow states and EPA to reduce and reallocate time and resources spent on manual data reviews while providing a more complete picture of program implementation and leading to the identification of opportunities for program improvement.

Appendix A: 2006 Data Reliability Improvement Action Plan

Introduction

The past two Data Reliability Improvement Action Plans have drawn attention to actions that can be taken to improve data quality and the usability of SDWIS/FED data. While they significantly focused on information system improvements and general activities that should improve data quality, this 2006 plan builds on current findings for more recent data and capabilities not previously developed that concentrate on specific factors that could result in real-time data quality improvements.

The philosophy of past data reliability improvement action plans largely was built on the concept that we must improve the software of the information system, SDWIS/FED. This has largely been done, with the last remaining step to be completed in 2007 with SDWIS/STATE Web Release 2. This release fully web-enables SDWIS/STATE, reducing resources needed to implement the software by states and reducing the complexity of data entry with fewer data entry screens and more drop-down lists. This 2007-2009 Data Reliability Improvement Action Plan primarily focuses on the actions of those responsible for determining which data will be entered and how that will occur. The largest challenge is ensuring that all data reflecting determinations of violations are entered into the SDWIS/FED, the federal data base.

As indicated in this report, EPA found that 77 percent of all data on MCL/SWTR TT violations in SDWIS/FED was complete and accurate. This is not satisfactory, has been the focus of media attention concerning the reliability of the data used to make decisions about the most important public health program in the nation for safeguarding its water supply to its citizens, and needs to be improved. To make a larger step forward over the next three years (2007-2009), EPA and ASDWA in October 2006 set a data quality goal of 90 percent (completeness and accuracy) for future compliance reporting of health-based violations in the federal database, SDWIS/FED. This plan is principally focused on achieving that goal. Based on past analyses of state-specific results, eleven states have achieved this level of data quality for health-based violations, indicating that it is achievable. The plan also addresses improving data quality of monitoring and reporting violations and inventory data; that is, improving the quality of all data used and supporting the state and national drinking water programs with the highest quality data. The Plan is presented in a series of issues and plan elements with assigned responsibility and timeframes.

Issues

(1) <u>Modify Data Verification Selection Processes</u>: EPA continues to conduct triennial data quality analysis and to follow up on data verification by working with states to address identified differences and discrepancies from federal regulations. In the 2005-2007 timeframe, EPA implemented probability-based selection (random selection) of states for data verification to enhance the representativeness of the data

at the national level. The data quality results for these data will provide an indication of the extent of achievement of the 90 percent data quality goal set by EPA and ASDWA. Consistent with the August 2005 recommendation of the special ASDWA-EPA Data Quality Subcommittee, the quality of results in the national database, SDWIS/FED will be displayed by rule and significance (i.e., public notification tier).

(2) Consider All Compliance Determinations in State Data Quality: In evaluating SDWIS/FED data quality, EPA only considers data in the national database and not in the state databases reflecting all compliance determinations resulting from the states' position as the primary enforcement authority for the federal program. EPA will develop an "electronic data verification" (eDV) tool to enable states to track any discrepancies of their compliance determinations relative to federal regulations and correct these discrepancies prior to data quality calculations and allow calculation of data quality relative to all compliance determinations. EPA should augment its SDWIS/STATE software to allow states to obtain management reports on any discrepancies in state compliance determination in near-real time to allow for the possibility of improving health-based response and data quality.

(3) Use Electronic Scheduling and Lab Reporting: Using automated monitoring requirements/schedule generators and incorporating electronic reporting from laboratories to states would improve the quality of data that states receive from water systems. Anecdotal information suggests that when states issue automated monitoring schedules to water systems, on-time monitoring and reporting by those systems improves. This step increases the probability that all data will be used by the state in determining compliance with public health drinking water standards and that appropriate determinations are made. Additionally, when states receive monitoring data electronically, data entry errors are reduced. This second step helps ensure that the correct data are used in the decision process for determining compliance. Water system or laboratory submission of data to states must comply with the Cross-Media Electronic Reporting Rule (CROMERR), compliance with which will need to be considered in any effort to facilitate electronic reporting from laboratories to states.

(4) Consider Data Management early in Rule Development: Data management concerns should be considered during every phase of the rule development process, beginning with the initial rule concept. If this does not occur, rules with complex reporting requirements may emerge, overwhelming the capability of states to implement them and shifting valuable resources from taking actions on real health needs to reporting. Data management using electronic reporting can simplify handling data but does not necessarily and always mean a simpler process for protecting health and should not be used as a "crutch" for creating complex rules instead of focusing on simpler, direct key health management objectives for drinking water supply protection. Streamlined approaches to data management in states' business processes must be considered in rule development.

(5) Improve State Capability in Compliance Determination: Data reliability, as reported in the Triennial Data Reliability Report, appears to have marginally improved, even

though this is not statistically significant. State compliance determinations play an integral role in determining the reliability of the data on violations reported to the national database, SDWIS/FED. Incorrect compliance determinations, when they do occur, are due in part to the complexity and number of drinking water rules. The need for training to facilitate correct determinations is critical, especially with the changing nature of state staff available to implement the drinking water regulations. Incorrect compliance determinations are a serious matter as they may affect public health.

(6) Complete SDWIS Modernization: EPA should continue implementation of the OGWDW Information Strategic Plan to modernize and web-enable the SDWIS/STATE to take advantage of newer technologies and system platforms. This action will save state resources by being able to enter data from anywhere in the state that is web-accessible and reduce data entry time with fewer screens and more drop down lists. State deployment of SDWIS/STATE Web Releases 1 and 2 will take time because of different schedules and variation of available resources among states. For states using SDWIS/STATE, full use of all SDWIS/STATE modules and regular update of inventory data will facilitate improved data quality.

(7) Evaluate Low Timeliness of Violation Reporting: Violation reporting timeliness is low and not improving. Because the states have been taking steps to improve data quality and the calculation of data quality considers results which may be 3 to 5 years old in some cases, estimates of reporting timeliness may not be current. EPA should use the reported results from the first year of using the modernized data flow to re-evaluate timeliness for each rule, as recommended by the Data Sharing Committee.

(8) Update Out-of-date and missing Inventory Data: Key features of inventory data useful in examining compliance and for determining regulatory needs are not routinely updated and reported. For example, consecutive systems or treatment objectives for recent rules are inventory data that are not reported for each system to which they apply. As a result, EPA cannot conduct analyses of national capability to treat certain contaminants. Inventory data for grant eligibility are routinely reported for the purposes of ensuring adequate data for receiving grants.

2006 Drinking Water Data Reliability Improvement Action Plan

Element	Element Description	Activity	Responsibility & Actions	Completion	Status
(1) Modify Data Verification Selection Processes	EPA will calculate data quality with data from 2005-2007 data verification from the random selection of states and display by rule and public notification tier.	(a) EPA will calculate data quality with data from 2005-2007 from the random selection of states and display by rule and public notification tier.	(1) EPA - Calculate national estimate of data quality for health-based violations and separately for monitoring and reporting violations and inventory data for 2005-2007 (2) EPA - Calculate state estimates of data quality for all health based compliance determinations and separately for all monitoring and reporting compliance determinations for 2005-2007 (3) EPA – Report data quality by rule and public notification tier for 2005-2007 using data verification results	(1) December 2008 (2) December 2008 (3) December 2008	
(2) Evaluate All Compliance Determinations	Develop a tool to allow states to identify compliance determination discrepancies from federal regulations more easily.	(a) EPA will develop an "electronic data verification" (eDV) tool to enable states to track any discrepancies of their compliance determinations relative to federal regulations and correct these discrepancies prior to data quality calculations and allow calculation of data quality relative to all compliance determinations.	(1) EPA & States - Complete pilot test of eDV tool	(1) December 2008	
		(b) States will agree to provide	(1) EPA & States - EPA	(1) July 2007	

Element	Element Description	Activity	Responsibility & Actions	Completion	Status
		contaminant occurrence and monitoring schedules data to EPA to allow the Agency to conduct electronic data verification for all rules across all water systems in a state, retrospectively, on an annual basis, but not less frequently than every three years, to allow regular assessment of data quality and to identify opportunities for state program improvement.	request and states provide contaminant occurrence and schedules data for all water systems from at least nine states for testing eDV tool (2) EPA & States – complete data sharing agreements for contaminant occurrence and monitoring schedule data (3) EPA & States - EPA will receive state contaminant occurrence and schedules data for all water systems from all states through completion of a data sharing agreement	(2) December 2008 (3) Annually beginning 2009	
(3) Use Automated Scheduling and Electronic Lab Reporting	States and EPA will take steps to more fully utilize automated technology to improve reporting of water system data to states.	(a) States will utilize automated scheduling of water system monitoring to the extent possible and report on progress in on-time monitoring and reporting by water systems at the ASDWA-EPA Data Management Users Conference.	(1) States – Report progress on State automated scheduling of system monitoring	(1) Annually May 2007 May 2009 May 2010	
		(b) EPA will develop an electronic tool to allow laboratories testing drinking water samples to report to states ("lab-to-state") reporting tool, rather than submitting paper	(1) EPA - develop "lab-to-state" reporting tool	(1) March 2007	(3) (b) (1) Done

Element	Element Description	Activity	Responsibility & Actions	Completion	Status
		reports on monitoring results.			
		(c) The EPA Office of Ground Water and Drinking Water will work with the EPA Office of Environmental Information (OEI) to incorporate CROMERR requirements in the "lab-to-state" reporting tool, work toward OEI approval of the tool.	(1) EPA - review and approval of CROMERR compliance of "lab-to-state" reporting tool	(1) August 2007	(3) (c) (1) Done
		(d) States not using the EPA developed "Lab-to-State" electronic reporting tool will identify and use a similar tool	(1) States will replace paper lab reports for compliance monitoring with automated lab reporting	(1) Ongoing through December 2009	
(4) Consider Data Management early in Rule Development	Implement a process to address data management in rule development.	(a) EPA information systems staff will participate in early rule development through preparation of issue papers on data management for each future rule and share these papers for comment with states through the ASDWA-EPA Data Management Steering Committee.	(1) EPA & States - information systems staff participate in rule development	(1) Ongoing	(4) (a) (1) Ongoing; Completed issue paper on TCR/Distribution System reporting for DSMC input
		(b) States will identify staff and participate in discussions of future rules to ensure that business processes are considered.	(1) State - staff identified for participation in rules to consider state business processes	(1) Ongoing	(4) (b) (1) Ongoing
		(c) ASDWA and EPA will work toward agreement on a mutual generic timeline for considering data management in rule	(1) ASDWA and EPA – reach agreement on generic timeline for including data management in rule	(1) December 2007	

Element	Element Description	Activity	Responsibility & Actions	Completion	Status
		development.	development		
(5) Improve State Capability in Compliance Determination	EPA Regions, to ensure that data reliability improvement (including implementation of EPA Order 5360.1.A2) is included in annual agreements with States, will work with states to identify the specific reasons for discrepancies in compliance determinations and to identify training needs among states to facilitate capability to make correct determinations.	(a) EPA Headquarters will develop an electronic data verification tool to allow EPA Regions to compare the results of all state compliance determinations to the violation data reported to EPA in SDWIS/FED.	(1) EPA HQ & States - Complete testing eDV tool	(1) September 2008	
			(2) EPA Regions and States – Use eDV tool to check compliance determinations and take appropriate action	(2) Ongoing beginning in 2009; quarterly check and take action	
		(b) EPA Regions will ensure that data reliability improvement steps are included in all agreements and work plans with States and identify specific reasons for discrepancies, including non-reporting, of state determinations with federal regulations.	(1) EPA Regions & States - Incorporate data reliability improvement steps in state-EPA agreements and state work plans	(1) Annually	
		(c) States will identify compliance determination training needs to EPA Regions.	(1) States - identify compliance determination training needs	(1) Annually	
		(d) EPA Headquarters will develop and provide capability for training on compliance determination for states	(1) EPA HQ - Completed/revision underway for compliance determination training	(1) Ongoing	
(6) Complete SDWIS Modernization	Complete modernization, web-enablement and deployment of SDWIS/STATE Web	(a) Development of fully web-enabled SDWIS/STATE and facilitation of fuller use of software for state program management	(1) EPA HQ – Develop SDWIS/STATE Web Release 2 (2) EPA Regions - promote full state use of	(1) October 2007 (2) Annually	

Element	Element Description	Activity	Responsibility & Actions	Completion	Status
	Release 2, facilitate fuller use of SDWIS/STATE among states choosing to use it and regular update of inventory data to improve data quality		SDWIS/STATE software through state agreements		
		(b) Deployment of web-enabled SDWIS/STATE with planned fuller use of modules by states using SDWIS/STATE and update of inventory data	(1) States - Deploy SDWIS/STATE Web Release 2 (2) EPA Regions and States using SDWIS/STATE - agree to steps toward fuller use of SDWIS/STATE in agreements and workplan (3) EPA HQ and States – Conduct workshop on SDWIS/STATE Web Release 2	(1) Beginning October 2007 (2) Annually (or as appropriate) (3) Summer 2008	
(7) Evaluate Low Timeliness of Violation Reporting	Evaluate timeliness by rule with data reported to the modernized SDWIS/FED for 2006	(a) Evaluate timeliness by rule with data reported to the modernized SDWIS/FED for 2006	(1) Data Sharing Committee - perform timeliness analysis in 2008 once all violation data are reported and processed; make recommendation to DMSC	(1) 2007	
(8) Update Out-of-date and missing Inventory Data	Evaluate regulatory requirements to determine the appropriate inventory reporting relating the applicability of rules to systems, set a priority on the data needed, and work with states to update the inventory data routinely reported to EPA	(a) Evaluate regulatory requirements to determine the appropriate inventory reporting relating the applicability of rules to systems, set a priority on the data needed, and work with states to update the inventory data routinely reported to EPA	(1) Data Sharing Committee - evaluate inventory reporting and propose a priority on data to be updated	(1) 2008	

Appendix B: Violations Addressed by Data Verification (DV)

Violation Code	Violation Name	Violation Type	Applicable rules and contaminant codes (CCodes)							
1	MCL, Single Sample	MCL			**Other VOC** 2378/ 80, 2955/ 64/ 68/ 69/ 76/ 77/ 79/ 80/ 81/ 82/ 83/ 84/ 85/ 87/ 89/ 90/ 91/ 92/ 96	**SOC** 2005/ 10/ 15/ 20/ 31/ 32/ 33/ 34/ 35/ 36/ 37/ 39/ 40/ 41/ 42/ 43*/ 44*/ 46/ 47*/ 50/ 51/ 63/ 65/ 67, 2105/ 10, 2274/ 98, 2306/ 26/ 83/ 88/ 90/ 92/ 94/ 96/ 98, 2400, 2931/ 46/ 59	**Nitrates** 1038, 1040, 1041	**Other IOC** 1005/ 10/ 15/ 20/ 24/ 25/ 35/ 36*/ 45/ 74/ 75/ 85/ 94	**Rads** 4000/ 06/ 10, 4100/ 01/02/ 74	**Coliform (Pre-TCR)** 3000
2	MCL, Average	MCL	**DBP** 1009, 1011, 2456, 2950	**TTHM** pre-'02 2941/ 42/ 43/ 44, 2950						
3	Monitoring, Regular	MR								
4	Monitoring, Check/Repeat/Confirmation	MR								
5	Notification, State	Other								
6	Notification, Public	Other								
7	Treatment Techniques	Other								
8	Variance/Exemption/Other Compliance	Other								
9	Record Keeping	Other	**FBR** 0500	**IESWTR** 0300						
10	Operations Report	Other								
11	Non-Acute MRDL	MRDL	**DBP** 0999, 1006/ 08							
12	Treatment Technique No Certif. Operator	TT	**DBP** 0400							
13	Acute MRDL	MRDL	**DBP** 1008							
21	MCL, Acute (TCR)	MCL	**TCR** 3100							
22	MCL, Monthly (TCR)	MCL								
23	Monitoring, Routine Major (TCR)	MR								
24	Monitoring, Routine Minor (TCR)	MR								
25	Monitoring, Repeat Major (TCR)	MR								
26	Monitoring, Repeat Minor (TCR)	MR								
27	Monitoring and Reporting Stage 1	MR	**DBP** 0400, 0999, 1006/ 08/ 09/ 11, 2456, 2920, 2950							
28	Sanitary Survey (TCR)	Other	**SS, TCR** 3100							
29	M&R Filter Profile/CPE Failure	MR	**IESWTR** 0300							
31	Monitoring, Routine/Repeat (SWTR-Unfilt)	MR	**SWTR** 0200							
36	Monitoring, Routine/Repeat (SWTR-Filter)	MR								

* codes required for monitoring only

Violation Code	Violation Name	Violation Type	Applicable rules and contaminant codes (CCodes)	
37	Treatment Technique State Prior Approval	TT	**IESWTR** 0300	**DBP** 0400
38	M&R Filter Turbidity Reporting	MR		
39	M&R (FBRR)	MR	**FBR** 0500	
40	Treatment Technique (FBRR)	TT		
41	Treatment Technique (SWTR)	TT	**SWTR** 0200	
42	Failure to Filter (SWTR)	TT		
43	Treatment Technique Exceeds Turb 1 NTU	TT	**IESWTR** 0300	
44	Treatment Technique Exceeds Turb 0.3 NTU	TT		
46	Treatment Technique Precursor Removal	TT	**DBP** 2920	
47	Treatment Technique Uncovered Reservoir	TT	**IESWTR** 0300	
51	Initial Tap Sampling for Pb and Cu	MR	**LCR** 5000	
52	Follow-up and Routine Tap Sampling	MR		
53	Initial Water Quality Parameter WQP M&R	MR		
54	Follow-up & Routine E.P. WQP M&R (deleted)	MR		
55	Follow-up & Routine Tap WQP M&R (deleted)	MR		
56	Initial, Follow-up, or Routine SOWT M&R	MR		
57	OCCT Study Recommendation	TT		
58	OCCT Installation/Demonstration	TT		
59	WQP Entry Point Noncompliance	TT		
60	WQP Entry Point Noncompliance (deleted)	TT		
61	SOWT Recommendation (deleted)	TT		
62	SOWT Installation (deleted)	TT		
63	MPL Noncompliance	TT	**LCR** 1022,1030	
64	Lead Service Line Replacement (LSLR)	TT	**LCR** 5000	
65	Public Education	TT		
71	CCR Complete Failure to Report	Other	**CCR** 7000	
72	CCR INADEQUATE REPORTING	Other		
75	PN Violation for NPDWR Violation	Other	**PN** 7500	
76	Other Non-NPDWR Potential Health Risks	Other		

Appendix C: Definition of Public Notification (PN) Tiers

Tier 1: Violations and Other Situations Requiring Notice Within 24 Hours

1. Violation of the MCL for total coliform, when ***fecal coliform or E. coli*** are present in the water distribution system, or ***failure to test*** for fecal coliform or *E. coli* when any repeat sample tests positive for coliform

2. Violation of the MCL for ***nitrate, nitrite, or total nitrate and nitrite***; or when a ***confirmation sample*** is not taken within 24 hours of the system's receipt of the first sample showing exceedance of the nitrate or nitrite MCL

3. Exceedance of the ***nitrate*** MCL (10 mg/l) by non-community water systems, ***where permitted*** to exceed the MCL (up to 20 mg/l) by the primacy agency

4. Violations of the MRDL for **chlorine dioxide** when one or more of the samples taken **in the distribution system** on the day after exceeding the MRDL at the entrance of the distribution system or when required **samples are not taken** in the distribution system

5. Violation of the ***turbidity MCL*** of 5 NTU, where the primacy agency determines ***after consultation*** that a Tier 1 notice is required or where consultation does not occur in 24 hours after the system learns of violation

6. Violation of the ***treatment technique*** requirement resulting from a ***single exceedance*** of the maximum allowable ***turbidity limit***, where the primacy agency determines ***after consultation*** that a Tier 1 notice is required or where consultation does not take place in 24 hours after the system learns of violation

7. Occurrence of a ***waterborne disease outbreak***, as defined in 40 CFR 141.2, or ***other waterborne emergency***

8. *Other* ***violations or situations*** with significant potential to have serious adverse effects on human health as a result of short term exposure, as ***determined by the primacy agency*** either in its regulations or on a case-by case basis

Tier 2: Violations Requiring Notice Within 30 Days

1. All violations of the ***MCL***, ***MRDL***, and ***treatment technique*** requirements ***except*** where ***Tier 1 notice*** is required

2. Violations of the ***monitoring*** requirements where the ***primacy agency determines*** that a Tier 2 public notice is required, taking into account potential health impacts and persistence of the violation

3. Failure to comply with the ***terms*** and ***conditions*** of any ***variance or exemption*** in place

Tier 3: Violations and Other Situations Requiring Notice Within 1 Year

1. ***Monitoring*** violations, except where Tier 1 notice is required or the primacy agency determines that the violation requires a Tier 2 notice

2. Failure to comply with an established ***testing procedure***, except where Tier 1 notice is required or the primacy agency determines that the violation requires a Tier 2 notice

3. ***Operation under variance*** granted under §1415 ***or exemption*** granted under §1416 of the Safe Drinking Water Act

4. Availability of ***unregulated contaminant*** monitoring results

5. Exceedance of the secondary maximum contaminant level for ***fluoride***