OFFICE OF INSPECTOR GENERAL

Catalyst for Improving the Environment

Memorandum Report

EPA Claims to Meet Drinking Water Goals Despite Persistent Data Quality Shortcomings

Report No. 2004-P-0008

March 05, 2004

Report Contributors: Jill Ferguson
Linda Pettit-Waldner
Tim Roach
Dan Engelberg

Abbreviations

EPA	Environmental Protection Agency
GPRA	Government Performance and Results Act
M/R	Monitoring and Reporting
OIG	Office of Inspector General
PWSS	Public Water Supervision System
SDWIS/FED	Safe Drinking Water Information System/Federal Version

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF
INSPECTOR GENERAL

March 05, 2004

MEMORANDUM

SUBJECT: EPA Claims to Meet Drinking Water Goals Despite Persistent Data Quality Shortcomings
Report No. 2004-P-0008

FROM: Kwai Chan /s/
Assistant Inspector General
Office of Program Evaluation

TO: Benjamin Grumbles
Acting Assistant Administrator
Office of Water

In each of the past 4 years, the Environmental Protection Agency (EPA) incorrectly reported meeting its drinking water goal under the Government Performance and Results Act (GPRA). The Agency reported meeting its annual performance goal for drinking water quality even though it concurrently reported that the data used to draw those conclusions were flawed and incomplete. In each of those years, EPA reported that it met its annual goal of 91 percent of the population drinking water that met health-based standards. However, EPA's own analysis, supported by our review, indicated the correct number was unknown but less than what was reported. We must note that this inaccuracy in reporting does not necessarily indicate a direct or immediate threat to human health.

Purpose

We initiated this review to evaluate the "drinking water performance measure," a key component of EPA's GPRA goal of "Clean and Safe Water." The evaluation questions were: (1) how do incomplete or inaccurate drinking water data affect the drinking water GPRA calculation; and (2) what actions have EPA undertaken to ensure that drinking water data collected and distributed to the public are reliable and valid?

During the preliminary research phase, we learned that the Office of Water was conducting analyses that largely overlapped our own, and was working with States and other stakeholders to address data quality problems. Since we already completed work on our first question but not the second, we are reporting the results on the first and suspending our work on the second.

Details on our scope and methodology, as well as background on GPRA and drinking water reporting, are in Appendix A.

Results

EPA has reported just meeting its annual performance goal for drinking water for fiscal years 1999 through 2002. In each of those years, EPA reported performance equaling the 91 percent GPRA annual performance goal. However, because EPA and the Office of Inspector General (OIG) reviews indicated that performance is less than what EPA reported, due to missing data on violations of drinking water standards, the Agency did not in fact meet its drinking water performance goals for these 4 years. Our assessment also mirrors statements made by the Agency in its performance reports and elsewhere.

EPA Consistently Reported Meeting Drinking Water Goals

EPA officials and reports consistently noted that national drinking water performance goals were being achieved. Annual performance reports, the 2003 Draft Report on the Environment, and statements by Agency officials indicated that national drinking water quality was high and EPA was progressing toward its goal of having 95 percent of the population drinking water that meets health-based standards. This was also repeated by the media. Figure 1 summarizes EPA's most recent annual performance report about drinking water quality, showing Agency claims that it just met its performance goal for each of the last 4 fiscal years:

Figure 1: EPA Reports Meeting Drinking Water Performance Goals

APG 8	Safe Drinking Water	Planned	Actual
FY 2002	91% of the population served by community water systems will receive drinking water meeting all health-based standards, up from 83% in 1994. Goal Met. ➥Corresponds with FY 2002 NEPPS Core Performance Measure (CPM).	91%	91%
FY 2001	*Same Goal, different targets. Goal Met.*		
	Performance Measures		
	- *Population served by community drinking water systems with no violations during the year of any federally enforceable health-based standards that were in place by 1994.*	*91%*	*91%*
	- *Population served by non-community, non-transient drinking water systems with no violations during the year of any federally enforceable health-based standards that were in place by 1994.*	*96%*	*92%*
FY 2000	*Same Goal. Goal Met*	*91%*	*91%*
FY 1999	*Same Goal, different targets. Goal Met*	*91%*	*91%*

Water

FY 2002 Result: In FY 2002, 244 million people were served by community water systems meeting all health-based standards. This result is 91% of the 268 million people served by 53,437 community water systems in FY 2002.

Source: *EPA 2002 Annual Performance Report, page II-22.*

In addition to annual performance reports, EPA portrayed its success at improving drinking water quality through other reports and through statements by Agency officials. For example, the 2003 Draft Report on the Environment stated, "In 2002, (S)tates reported that 94 percent of the population served by community water systems were served by systems that met all health-based standards, up from 79 percent in 1993."[1] A July 20, 2003, statement from the Assistant Administrator for Research and Development repeated this conclusion that, in 2002, 94 percent of Americans were served by drinking water that met health-based standards. Using Agency reports, the media communicate such information to the public. For example, a June 23, 2003, New York Times editorial, "An Environmental Report Card," used similar language in the press release cited in Figure 2 to report that, "Fully 94 percent of Americans are served by drinking water systems that meet federal health standards, as opposed to 79 percent 10 years ago." Most recently, the new Administrator's Draft 500-Day Water Quality Plan continued using drinking water quality projections that remained unchanged from previous claims and press releases. The Plan stated, "(i)n 2002, 93.6 percent of the population received drinking water that met all health-based standards. By 2015, all people served by community water systems will receive drinking water that meets standards."

Figure 2: Statement on Drinking Water Quality

"Our drinking water is purer. In 2002, 94 percent of Americans were served by drinking water systems that meet our health-based standards – an increase of 15 percent in the last decade."

Source: *EPA's June 23, 2003, Press Release*

EPA Reports Data Quality Problems While Reporting Performance Goals Met

EPA's recent Performance Reports contain statements about the quality of the drinking water data used to report under this performance measure. In the last three Annual Performance Reports the Agency's message was consistent: there are problems with the quality of data in the Safe Drinking Water Information System/Federal Version (SDWIS/FED). These statements were echoed in other reports, such as the 2003-2008 Strategic Plan. In that Plan, EPA reported, "the baseline statistic of national compliance with health-based drinking water standards likely is lower than reported." However, EPA continued to report that drinking water performance goals were being met. The following tables contain disclaimers related to data quality problems that EPA has included in reports:

[1] This statistic is based on the 2002 calendar year, while the 91 percent in EPA's 2002 annual performance is based on the Federal fiscal year.

Statements in EPA Annual Reports	
Report	**Data Quality Description**
1999 Annual Performance Report	There is no indication of data quality in the discussion of performance for this Annual Performance Goal.
2000 Annual Performance Report	"There are recurrent reports of discrepancies between national and state data bases . . . Given the particular need for confidence in the completeness and accuracy of data about drinking water quality, EPA designated SDWIS content as an Agency material weakness in 1999, under the Federal Managers' Financial Integrity Act."
2001 Annual Performance Report	A technical appendix noted under-reporting of monitoring and violations data to EPA and that "failures to monitor could mask treatment technique and MCL violations."
2002 Annual Performance Report	"The most significant data quality problem is under reporting to EPA of both monitoring and reporting violations and incomplete inventory characteristics . . . failures to monitor could mask treatment technique and MCL violations. Such underreporting of violations limits EPA's ability to precisely quantify the population served that are meeting health based standards."

Statements in Other EPA Reports	
Report	**Data Quality Description**
Fiscal Year 2001 Annual Plan	"SDWIS data quality has been problematic. It has been demonstrated that there are discrepancies between SDWIS data and state databases. In addition, utilities have pointed out specific data quality problems."
2003 Draft Report on the Environment *(Technical Document)*	"Underreporting and late reporting of CWS violations data by states to EPA affect the ability to accurately report the quality of our nation's drinking water... Based on this analysis, the agency estimated that states were not reporting 40 percent of all health-based violations to EPA."
2003-2008 Strategic Plan	"Routine data analyses of the Safe Drinking Water Information System (SDWIS) have revealed a degree of nonreporting of violations of health-based drinking water standards… As a result of these data quality problems, the baseline statistic of national compliance with health-based drinking water standards likely is lower than reported."
Data Reliability Analysis of the EPA SDWIS/FED and Plan *(Draft)*	"If the quality of the data measured and reported to SDWIS-FED, the source of the data for the [GPRA performance] calculation, are less than 100 percent as defined by this analysis, then the progress toward meeting the strategic goal may not be as great as reported."

The information in the previous Figures and Tables indicate that while the Agency consistently reported meeting its drinking water performance goals, the Agency also consistently acknowledged problems with drinking water data quality. Therefore, we believe the Agency has wrongly reported that it met its 91 percent performance goal for the years 1999 to 2002 – the actual number is lower by some unknown amount.

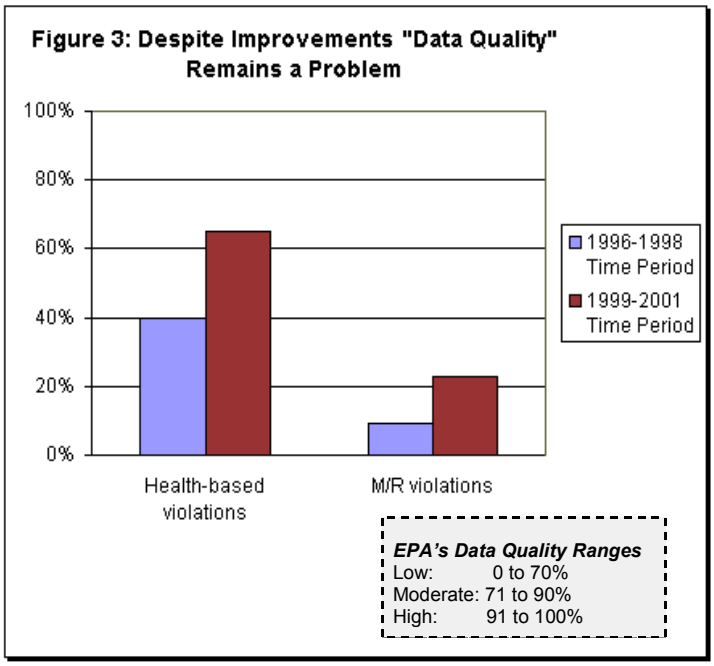
EPA and OIG Reviews Indicated GPRA Measure Less Than Reported

Since 2000, EPA has developed two drinking water data quality reports (with the second still in draft form). Both reports noted problems with under-reporting, in that States did not transmit all health-based violations and all monitoring and reporting violations into SDWIS/FED, which houses compliance information about drinking water systems. Our analysis confirmed these data quality problems of under-reporting of violations to health-based drinking water standards.

EPA Reports Indicate Drinking Water Data Quality Improved But Still "Low"

EPA has been conducting "data verifications" of drinking water data since 1991. In 2000, EPA issued the "Data Reliability Analysis of the EPA Safe Drinking Water Information System/Federal Version (SDWIS/FED)" report from data verifications conducted between 1996 and 1998. This report identified problems with the accuracy and completeness of SDWIS/FED data. The followup 2003 report, currently under internal draft review, includes information from data verifications from 1999 to 2001.

These two reports distinguish between"health-based" and "monitoring and reporting" (M/R) violations. As shown in Figure 3, the first report indicated that "data quality" for health-based violations at the community water systems (or "systems") whose records were examined during the 1996-1998 data verification time period was 40 percent.[2] That is, 40 out of every 100 health-based violations that should have been in SDWIS/FED were in SDWIS/FED, and 60 out of 100 were not. In the second time period (1999-2001), data quality improved to 65 percent. According to EPA's own assessment, this is still in the "low quality" range.



Data quality for the monitoring and reporting component of drinking water compliance also improved among the communities audited – from 9 percent to 23 percent between the two time periods. In other words, the data quality problem for M/R was at 91 percent and has reduced to 77 percent for an overall improvement of 14 percent. Because the enforcement portion of SDWIS/FED tracks only non-compliance with drinking water regulations, the

[2] EPA has defined SDWIS/FED data quality as the percent of data that should be in SDWIS/FED that are in the database with no discrepancies or errors.

monitoring and reporting violations indicate instances in which information about water quality was not reviewed by the State to make compliance determinations. The reported levels of health-based and monitoring and reporting data quality indicate that EPA performance reports reflected a best-case scenario and that in all likelihood performance was lower than reported.

EPA has recognized the data quality limitations of SDWIS/FED and its impact on the Agency's ability to manage the drinking water program, as well as to accurately report to Congress and the public. In the draft 2003 Data Reliability Analysis, EPA points out that, "Overall, the violations data that are reported to and accepted by SDWIS/FED are highly accurate. The weak link in data quality continues to be the large number of violations that are not reported to SDWIS/FED (as estimated by Completeness), with monitoring and reporting data being the least complete and of very low quality."

Recent improvements in drinking water data quality are attributed to EPA, State, and third party efforts to identify data quality deficiencies and implement activities to remedy those deficiencies. This has been a long-term effort, signified by the publication of reports outlining problems with drinking water data quality and activities to fix those problems. The 2003 Draft Data Reliability Report of SDWIS/FED noted that in 1999, a workgroup of EPA, State, and stakeholder representatives developed a data quality action plan. Some of the components of that action plan included setting data quality goals for SDWIS/FED, quantify and qualify the quality of SDWIS/FED data, and take interim steps to improve data quality.

The Draft Data Reliability Report also noted that EPA and States undertook and completed a number of activities to reach their data quality improvement goals. Some of those actions taken by EPA and the States included: (1) improved data entry processes, tools, and training for regions and States; (2) improved and simplified data retrieval and reporting tools; (3) improved data verification audit procedures; and (4) accelerated ongoing data quality improvement activities (development of SDWIS/STATE, and electronic reporting between utilities, labs, and States). A second data quality action plan is being developed and implemented. When we met with EPA officials to discuss our draft report, they told us that data quality continues to improve, as measured by the most recent data verifications.

EPA also noted in March 2003 in the Draft Strategic Plan for 2003-2008 that it would consider how to best classify water systems that experienced monitoring and reporting violations. Options included (1) classifying systems with monitoring and reporting violations as not being in compliance with health-based standards, and (2) excluding these systems from the GPRA calculation. By doing so, the Agency would remove from its performance reporting the systems that it cannot determine provided water that met all health-based standards. However, the final Strategic Plan issued in September 2003 stated, "(The) Agency is currently engaged in statistical analysis to more accurately quantify the impact of these data quality problems."

During our meeting with Agency officials to discuss the draft report and these two options, they explained that there is a potential for water systems with no health-based violations to be eliminated from the GPRA calculation because of one M/R violation. They believe that this would distort their reporting under GPRA, and they are studying the issue. While we understand that this potential exists, we also believe that EPA's current policy of treating systems with M/R violations as being in compliance with health-based standards can also distort the GPRA measure. The Agency's 2001 and 2002 annual reports (see page 4) note that there is a potential for M/R violations to mask violations to drinking water standards.

OIG's Review of EPA's Database Confirmed Agency's Assessment

Our review of the EPA data verification database confirmed the Agency's conclusion that States did not report all health-based violations into SDWIS/FED. The actual percentage of people drinking water that met health-based standards in this sample was likely to be lower, but we cannot use the database to determine a range for the nation as a whole. This is because the methods used to select the 761 water systems do not support estimates for the nation's approximately 54,000 community water systems.[3]

In the data verification database, we observed that the error rate was high for systems with health-based violations to the contaminants reviewed during data verifications. Of the 71 systems with violations for the drinking water standards that were reviewed during the data verifications, 17 had not been reported into SDWIS/FED.[4]

Overall, the direction of errors in the reporting of health-based violations caused a downward bias in the drinking water performance measure among the 761 systems. This suggests that by utilizing incomplete results from SDWIS to report performance under GPRA, EPA portrayed an incorrect picture of the percentage of people drinking water that met all health-based standards.

There is always potential for errors when collecting any type of information. In Appendix A, Figure 5 illustrates the flow of information from the water system, through the laboratories, and to the State and EPA. State data verifications identify errors related to data analysis and reporting of violations.

Conclusions

Congress, the public, and the media rely on transparent and accurate reports about drinking water quality. The significance of this measure is that each percent of the population served by community water systems receiving drinking water meeting all health-based standards represents more than 2.6 million people in the United States. As EPA persistently reports meeting its drinking water performance goal while acknowledging drinking water data quality problems, it in fact has not accurately reported its performance to the approximately 268 million people drinking

[3] A Community Water System serves at least 25 people or 15 service connections on a year-round basis.
[4] Our review also indicated that 2 of the 761 systems were identified to have a health-based violation in SDWIS/FED when no violation actually occurred.

water from community water systems. EPA's increasing candor about the limitations associated with basing performance measures on its compliance database (SDWIS) and the fact that it identified possible corrective actions in its 2003-2008 Draft Strategic Plan for addressing the problems of water system monitoring and reporting violations indicate the Agency's willingness to consider alternative approaches for how it reports performance.

We suggest that while EPA and States continue moving forward to correct data deficiencies, the Agency should also identify methods to better account for the impact of the "large number" (as described in the Draft 2003 Data Reliability Report) of violations that are not reported to SDWIS/FED. It should determine how best to account for community water systems with monitoring and reporting violations when reporting into GPRA and adjust the measure to reflect this. Options include those mentioned above that were described in the Draft Strategic Plan. In order to address broader concerns over this measure, given the inherent problems utilizing SDWIS for reporting on performance, we also suggest that in the future the Agency move toward employing an altogether different methodology for reporting performance for this Annual Performance Goal. One approach would be for EPA to base its future reporting on a stratified sample of the nation's 54,000 community water systems and audit those systems for compliance with health-based drinking water standards. This has the potential to provide a more accurate and transparent accounting of the nation's drinking water quality.

Agency Comments and OIG Response

In the Agency's February 2, 2004, response to our draft report, the Agency did not directly acknowledge our principal finding concerning the incorrect conclusions about drinking water performance contained in recent annual performance reports. In addition, while the Agency agreed to continue to improve how EPA communicates health risks associated with drinking water, no commitment to specific steps to correct the inconsistencies we had pointed out were agreed to. Appendix B contains Agency comments, and Appendix C contains some of our specific responses to those comments.

Based on the Agency's comments, we made several revisions and clarifications to our report. However, insofar as data quality within SDWIS is not the principal focus of this report, the comments did not address our principal concern: that for 4 years, EPA has reported to Congress and the public that it met an important annual performance goal when available evidence indicates it did not. After reviewing EPA's comments, we continue to believe that the Agency inappropriately claimed to have met performance goals for its drinking water program for the past 4 years. Steps to account for missing and inaccurate data when reporting performance under GPRA are being considered by EPA, but no decisions have been made. We reiterate our suggestion that EPA change how it reports under GPRA to compensate for known concerns over the reliability of this measure.

If you or your staff have any questions, please contact me at (202) 566-0827, or Dan Engelberg, Director of Water Issues, at (202) 566-0830.

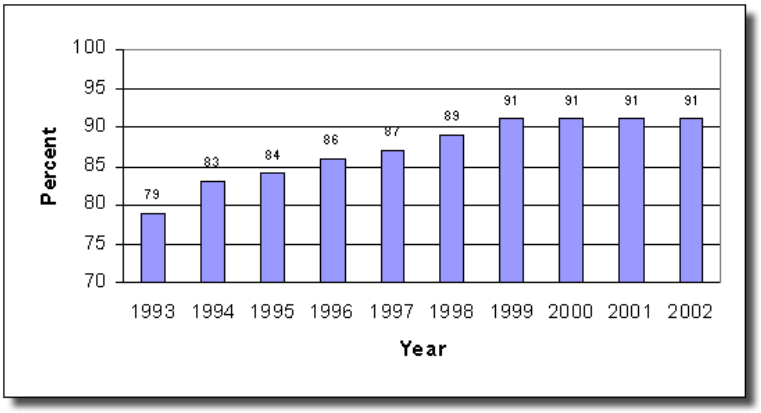
Background, Scope, and Methodology

Background

In 1993, Congress enacted GPRA to shift Federal planning, management, and decision-making away from a traditional focus on resources and activities to a focus on results and outcomes. The Office of the Chief Financial Officer and program offices produce an annual report to Congress on the Agency's progress toward achieving annual strategic goals. In response to GPRA, EPA established major goals, including "Clean and Safe Water." One of the sub-objectives under this goal was "Water [that is] Safe to Drink," which is designed to reflect the quality of the drinking water supplied to the population.

In 1999, EPA established a measure of progress toward meeting this sub-objective: by the year 2005, 95 percent of the population served by community water systems would have water that is safe to drink, meaning that the water meets all applicable health-based standards. As is shown in Figure 4, over the past 9 years, EPA has reported an increasing percentage of the population drinking water that meets health-based standards. In its new strategic plan, EPA has retained the performance sub-objective of water that is safe to drink and the measure of percent of population, but has extended the timetable to accomplish the objective by 3 years, to 2008.

Figure 4: Population Reported by EPA Meeting Health Based Standards

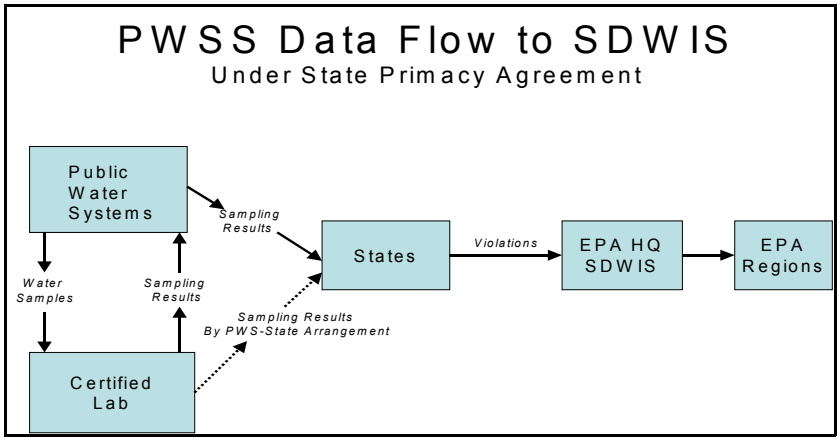


Source: *EPA 2002 Drinking Water Factoids. This data is based on the Federal fiscal year, which ends September 30. The Draft Report on the Environment reported 94 percent of the population served by community water systems drank water that met all health-based standards for the 2002 calendar year.*

The drinking water performance measure is based on compliance information contained in SDWIS/FED. The information utilized from this database is derived by sampling and analyses of drinking water, and assessments of treatment techniques from the approximately 54,000 community water systems that supply water to 268 million Americans. SDWIS/FED is designed to support many program management functions, including storing basic water system information, enforcement actions, and sampling results for unregulated contaminants.

Public water systems are responsible for monitoring their own systems and collecting and reporting sampling results to a primacy agent (typically State or Tribal drinking water programs). All States have primacy except for Wyoming. Primacy agents determine compliance with drinking water regulations and report violations to EPA (via SDWIS/FED). This process is illustrated in Figure 5. SDWIS/FED contains data when violations of drinking water standards and mandated treatment techniques are reported into it. To measure program performance, EPA aggregates the SDWIS/FED data into a national measure of overall compliance with health-based drinking water standards, which it reports as a percentage.

Figure 5: SDWIS Data Flow



Source: *EPA Office of Water*

Scope and Methodology

We reviewed a database containing the results of a series of contractor-conducted audits (known as data verifications) of public water system data in State files.[5] We used this database to discuss two sources of errors that affect the precision of the data used in the drinking water quality measure: (1) errors in the process of reporting drinking water data from States to EPA, and (2) problems associated with under-reporting of drinking water information because of a reliance upon an exceptions-based database (meaning that only violations are recorded) for tracking drinking water violations.

We followed all but one of the applicable *Government Auditing Standards*, issued by the Comptroller General of the United States. We were unable to adhere to the standard that we assess management controls. Specifically, there were no written procedures to document the flow of information from data verifications into the data verification database. Based on our discussions with Office of Water staff, we believe that the data in this database were sufficient

[5] We focused on data from community water systems because compliance data from this group are used for the drinking water GPRA measure.

for the purposes of our review. We followed all other standards for performance audits or evaluations. All work was completed between August 2002 and September 2003.

Appendix B

Agency Comments

MEMORANDUM

SUBJECT: Comments on the Draft Report *EPA Reports Meeting Drinking Water Goals Despite Persistent Data Quality Shortcomings*

FROM: Benjamin H. Grumbles
Acting Assistant Administrator

TO: Nikki L. Tinsley
Inspector General

Dear Ms. Tinsley:

Thank you for the opportunity to comment on your office's draft report, "EPA Reports Meeting Drinking Water Goals Despite Persistent Data Quality Shortcomings." I appreciate your general interest in this issue. The Office of Water recognizes the importance of high quality data and is committed to continue to make improvements in this area in the drinking water program as well as across our other activities

EPA's data verification audits and associated analyses, which are also the basis for your draft report, indicate that the data in SDWIS–FED are highly accurate with very few errors, but are still incomplete. EPA and the states have made significant progress in improving the quality of SDWIS-FED data since we first became aware of this issue. We acknowledge, however, that more work remains to be done. EPA's report, *Drinking Water Data Reliability Analysis and Action Plan (2003)*, to be released shortly, will highlight our continuing efforts and additional steps that we intend to take in partnership with states to further improve drinking water data quality.

While the Agency does use SDWIS-FED data to meet reporting requirements under the GPRA, we are aware of the data's shortcomings and have been diligent in flagging those to key audiences as well as to the general public. EPA in its GPRA reporting is using the data that is available to us through the national reporting system. We will continue to explore ways to communicate the range of issues associated with the nature and quality of SDWIS-FED data and the relationships to public health risk with your suggestions in mind. We will also continue to engage in discussions with states regarding potential new approaches for reporting drinking water data (e.g., electronic transfer of monitoring results from the laboratory to the federal database).

In a broader context, I would like to note that in the vast majority of instances where states make compliance determinations these determinations are correct. Most of the determinations correctly find that public water systems are meeting health-based standards and thus do not require an entry to be made in SDWIS-FED, which is a violations-only database. I mention this not to diminish the very real need to improve data quality, but as an important reminder that SDWIS data quality and drinking water quality are far from synonymous.

I have attached a more detailed set of comments on specific aspects of the draft to assist your office as you prepare a final report. Please call Cynthia Dougherty, Director of the Office of Ground Water and Drinking Water, at (202) 564-3750 if you would like further clarification on any of these issues.

Attachments

cc: Kwai-Cheung Chan
Mike Shapiro
Cynthia C. Dougherty
Elizabeth Corr
Dan Engelberg
Jill Ferguson
Linda Pettit-Waldner
Tim Roach
Michael Mason

Office of Ground Water and Drinking Water Specific Comments on Draft Inspector General Report,
EPA Reports Meeting Drinking Water Quality Goals Despite Persistent Data Quality Shortcomings (12/23/03)

1) In general, the IG draft report addresses two topics EPA is actively engaged in analyzing:

- The quality of data reported by states to EPA's database of record, the Safe Drinking Water Information System/Federal Version (SDWIS/FED) for public water system inventory, violations and enforcement actions, and
- The implications of that quality on a Government Performance and Results Act (GPRA) measure for populations receiving drinking water from community water systems that are in compliance with all health-based standards.

We are addressing both of these issues through our work internally and with states to continue to improve data quality and to identify new ways of communicating its significance to the public. We are also at this time near completion of our second comprehensive report on the quality of data in SDWIS. We expect to release this final report in the near future and will provide a copy to the IG at that time.

2) The IG draft report characterizes EPA as having "mistakenly" reported meeting its drinking water goal under the Government Performance and Results Act (GPRA). In actuality, however, we use the best available data reported to us by the states under their primacy agreements with the Agency. In using this data to describe results under the GPRA, we have tried to be clear that the results are based on the data as reported in SDWIS and that our audits indicate that there is incomplete reporting of this data. We have also made our first triennial report on SDWIS data quality, published in October 2000, available to the public on our website. We will soon be making our second report, noted above, on data quality available on the web.

3) EPA's data verifications indicate that the data the states report to the Agency are very accurate, although incomplete.

4) Several of the IG's comments, including an incorrect flow chart, indicate the need for improved understanding of data flow from public water systems to states to EPA. We have prepared a revised flow chart at the same level of detail (attached). We would suggest a meeting between OGWDW and the IG's staff before you finalize your draft report to ensure an accurate understanding of key details related to data flow which are beyond the depictions in the chart. Such a meeting would also serve as an opportunity for us to provide, and discuss where necessary, other detailed edits to the draft report for purposes of accuracy.

5) The draft report includes a chart that depicts data quality improvements, but the draft report includes only brief discussion on this point. Specifically, we believe that the draft

report should recognize that data quality has improved based on actions taken jointly by EPA and states since 1998. We are concerned that the absence of elaboration on this point undercuts the concerted efforts as well as the progress that EPA and states have made and are continuing to make.

6) To achieve a balanced examination, we suggest that the IG evaluate factors that could affect the results of data quality calculation in either direction, rather than emphasizing only factors that might appear to reduce the reported levels of the population receiving safe water all the time. We would like to take this opportunity to draw certain key points to the IG's attention for discussion in the draft report.

- In developing our own report, we have been examining the issue of over-reported violations. We looked at all the large water systems (over 50,000 population served) that had been identified as being in violation and found that one-third or more had corrected the violations and should not have been reported as being in violation in 2001. If considered, this could have the effect of increasing the GPRA percentage.

- Another noteworthy factor is that for many large water systems (which have the greatest effect on the GPRA number) a violation may not affect water quality throughout the entire water system, even though for GPRA accounting purposes the entire water system is credited with a violation. Quantitative consideration of this factor, while challenging to do, would likely contribute additional and substantial population to be counted as receiving drinking water which had no violation.

- A further set of potentially relevant factors is the impact of violation timing, frequency and duration on the significance and potential public health consequences of violations.

- Similarly, the varying nature of violations (e.g., a one-time violation for a chronic contaminant versus for an acute contaminant, or a violation significantly above the standard versus one that is close to the standard) may have differing public health implications.

We are very interested in finding ways to communicate these complexities succinctly to enhance public understanding of the GPRA measure and what it means. We would welcome the IG's comments on these factors.

7) The draft report characterizes the success of improved drinking water quality (distinct from data quality) as the Agency's result when in fact it is the result of a broad partnership that includes EPA, the states that are the primary implementers of the national drinking water program and public water systems that carry out the regulatory requirements.

8) An area that the IG touches on in the draft report and which resurfaces in the context of the conclusions is the potential impact of monitoring and reporting violations on GPRA

reporting. Monitoring and reporting violations, however, may have no link to whether a water system met health-based standards. For instance, where monitoring and reporting violations are scattered among numerous water systems that otherwise routinely demonstrate that they meet health-based standards, the likelihood of a significant impact on EPA's GPRA reporting is less than if monitoring and reporting violations occur repeatedly within the same water systems. We suggest that the IG reconsider whether to emphasize this complex issue as part of its conclusions in the absence of further analysis.

9) In the first paragraph on page 7, it is unclear to us whether the IG is discussing the utility of the data verification database or the SDWIS database and also unclear as to whether the estimates under discussion are the data quality estimates or the GPRA number.

In this and the following paragraph, there also appears to be a misunderstanding about the regulatory framework of the SDWA. Under the drinking water program's regulatory structure, health standards for multiple contaminants are addressed within single rulemakings. The draft report indicates that data verifications only evaluate eight drinking water standards. This is incorrect and affects the IG's conclusion about extrapolating data verification results for data quality purposes and possible "larger discrepancies" in the Agency's GPRA calculation. In fact, EPA's data verifications examined all 87 contaminants that were regulated under the SDWA in 2001 and EPA will continue to evaluate all regulated contaminants in future data verifications.

10) We believe there are a wider range of circumstances affecting shortcomings in the data that should be considered in an evaluation, including but not limited to:

- Relationship of waivers, variances and exemptions to violation data points;
- Conditions in state programs that result in non-reporting of violations;
- Methodologies that would improve evaluation and understanding of state program processes to determine compliance;
- Development of methodologies for estimating the proportion of populations in larger water systems actually affected by violations, rather than charging the entire water system with a violation that only affects a portion of its population.

11) Concerning the listing "Stakeholders Identified Other Potential Sources of Error" in Appendix B, the draft report does not present an evaluation of these potential sources of error nor indications of which may be more problematic. In general, the reference to "error" is inappropriate. Some of these appear to be undocumented and/or unevaluated opinions about potential sources of error and others are not potential sources of error, but rather process vulnerabilities. This particularly applies to the GPRA measure section.

Appendix C

Additional OIG Responses

Note to Agency Comment #2:

We agree that the Agency currently uses the best available data reported by the States and has moved in recent years to be more transparent in the presentation of problems with SDWIS data. For this reason, we have changed "mistakenly" to "incorrectly" in the report. We realize that this isn't an error of oversight on the part of EPA. However, EPA continues reporting that GPRA goals are met while warning about the implications of missing data. In our view, correctly reporting whether it has met a performance goal is at least as important as disclosing the existence of errors.

Note to Agency Comment #4:

The final report contains the data flow chart provided by the Office of Water.

Note to Agency Comment #5:

Protecting and improving the nation's drinking water quality and drinking water data quality is a collective effort with credit for success attributable to many parties. We did not intend to reduce the share of credit to any one group by briefly noting data quality improvement efforts undertaken in previous years. We did intend to highlight that drinking water data quality improvements are a result of activities to improve data management systems and processes. The issues regarding the transparency of GPRA reporting are the focus of our report, which is why we did not further elaborate on data quality improvement activities. However, we have changed our presentation in the final report to better reflect the shared responsibility and accomplishments of EPA and its partners.

Note to Agency Comment #6, first bullet:

We agree that the GPRA percentage is affected by water systems with incorrectly reported health-based violations. Our review of the data verification database factored in 2 such systems out of the 71 that experienced health-based violations. We were aware of other drinking water databases, but chose to review only the data verification database because of the semi-random sampling methodology used for selecting the 761 community water systems.

General Note to Agency Comment #6:

These factors all contribute to the complexity of presenting a picture of the nation's drinking water quality for the purposes of GPRA reporting. We suggest that while EPA works to address data issues such as those described here, the Agency also more clearly report that the absence of drinking water data in SDWIS/FED have an effect on the accuracy of the annual GPRA reports.

Note to Agency Comment #7:

See our response to EPA's Comment 5.

Note to Agency Comment #8:

EPA's 2001 and 2002 Annual Performance Reports noted that failures to monitor could mask violations of health-based standards (see page 4). We agree with this position. We feel that EPA is mistaken in asserting that the impact of GPRA reporting is less if M/R violations occur repeatedly in a single water system than if the same number of M/R violations are spread among several different systems.

Note to Agency Comment #9

We clarified language and corrected errors in the draft report.

Note to Agency Comment #10

We agree that these are important factors that affect the accuracy and validity of the GPRA measure. For the purposes of this report, our focus was on the implications of reporting success at meeting the drinking water GPRA goal while concurrently reporting problems with the completeness of drinking water data.

Note to Agency Comment #11

We removed references to potential sources of error based on Agency comments.

Appendix D

Report Distribution

Assistant Administrator, Office of Water (4101)
Director, Office of Ground Water and Drinking Water (4607)
Comptroller (2731A)
Agency Followup Official (the CFO) (2710A)
Agency Audit Followup Coordinator (2724A)
Associate Administrator for Congressional and Intergovernmental Relations (1301A)
Associate Administrator, Office of Public Affairs (1101A)
Inspector General (2410)