United States Environmental Protection Agency | Office of Water (4607) Washington, D.C. 20460 | EPA-815-R-00-023 December 2000



ARSENIC OCCURRENCE IN PUBLIC DRINKING WATER SUPPLIES

Prepared for:

Office of Ground Water and Drinking Water
1200 Pennsylvania Avenue, N.W., MS 4607
Washington, D.C. 20460
MS 4607

Andrew E. Schulman
USEPA Work Assignment Manager

Prepared by:

ISSI Consulting Group
8455 Colesville Road, Suite 915
Silver Spring, MD 20910
Under Contract No: 68-C7-0005

The Cadmus Group, Inc.
1901 North Fort Myer Drive, Suite 900
Arlington, VA 22209
Under Contract No.: 68-C-99-206

And:

ICF Consulting
101 Lucas Valley Road
San Rafael, CA 94903

This page intentionally left blank

Table of Contents

List of Acronyms iii

Acknowledgments v

Executive Summary vi

1. Introduction 1
    1.1 Purpose of this Document 2
    1.2 Organization of the Document 2

2. Sources of Arsenic 5
    2.1 Physical and Chemical Properties of Arsenic 5
        2.1.1 Environmentally Relevant Arsenic Species 5
    2.2 Natural Sources of Arsenic 8
    2.3 Anthropogenic Sources of Arsenic 14

3. Fate and Transport of Arsenic 21
    3.1 Relationship of Fate and Transport Properties to Source Intake 21
    3.2 Relationship of Fate and Transport Properties to Treatment and Distribution . . 22

4. Sources of Data on Arsenic Occurrence in Drinking Water Supplies 25
    4.1 Arsenic Occurrence and Exposure Database (AOED) 25
        4.1.1 Safe Drinking Water Information System (SDWIS) 27
        4.1.2 State Compliance Monitoring Databases 28
        4.1.3 Building the AOED from SDWIS and State Compliance Monitoring Databases 47
    4.2 Comparison Databases 50
        4.2.1 National Arsenic Occurrence Survey (NAOS) Database 50
        4.2.2 USGS Arsenic Databases 52
        4.2.4 Metropolitan Water District of Southern California (Metro) Database . . 54
    4.3 Other Databases 55
        4.3.1 1969 Community Water Supply Survey 55
        4.3.2 1978 Community Water Supply Survey 55
        4.3.3 Rural Water Survey 56
        4.3.4 National Organics Monitoring Survey 56
        4.3.5 Western Coalition of Arid States Research Committee Arsenic Occurrence Study 56
        4.3.6 Association of California Water Agencies Database (ACWA) 57

5. Arsenic Occurrence Patterns in the United States 59
    5.1 Stratification by Source Water Type 59
    5.2 Stratification by System Size 63
    5.3 Stratification by System Type 70

5.4 Regional Stratification 73
5.5 Arsenic Distributions at the State Level 77
5.6 Summary of Patterns of Arsenic Occurrence 85

6. National Occurrence Estimates 87
6.1 Arsenic National Occurrence Projection Methodology 87
6.1.1 System Means 88
6.1.2 State Exceedance Probability Distributions 89
6.1.3 Regional Exceedance Probability Distributions 92
6.1.4 National Exceedance Probability Distributions 96
6.1.5 Number of Systems Exceeding Alternative MCLs 97
6.2 Arsenic National Occurrence Estimates Results 97
6.2.1 Community Water Supply Systems 100
6.2.2 Non-Transient, Non-Community Water Supply Systems 100
6.3 Comparisons of Occurrence Estimates 105
6.3.1 Comparison of AOED, NAOS, USGS, MWDSC, and Wade Miller Occurrence Estimates at the National and Regional Level 105
6.3.2 Comparison of AOED, Kennedy-Jenks, and Saracino-Kirby Occurrence Estimates for California 110
6.4 Uncertainty Analysis 113
6.4.1 Purpose of Uncertainty Analysis 113
6.4.2 Uncertainty Analysis Methodology 116

7. Intra-system Variability 123
7.1 Purpose of Analyses 123
7.2 Available Data 123
7.3 Analytical Methods and Results 124
7.3.1 Estimation of Concentrations for Non-Detects 126
7.3.2 Log-Normal Mixed Model 126
7.3.3 Results 127
7.4 Summary of Intra-system Analyses 128

8. Temporal Variability 129
8.1 Purpose of Analysis 129
8.2 Available Data and Results 129

9. References 133

Appendices

Appendix A: Adapted Regression on Order Statistics Methodology
Appendix B: Analysis Results
    Appendix B-1: State Exceedance Probability Distributions
    Appendix B-2: Box Plots
    Appendix B-3: Lognormal Probability Plots of System Means
Appendix C: Summaries of Pre-1980 Data Sets
Appendix D: Database Specifications and Data Conditioning
    Appendix D-1: Individual State Database Specifications for Preliminary Database
    Appendix D-2: AOED Database Specifications
    Appendix D-3: Initital Datat Conditioning Process

List of Acronyms

AA	Atomic Adsorption
ACWA	Association of California Water Agencies
AES	Atomic Emission Spectrometry
ANOVA	Analysis of Variance
AOED	Arsenic Occurrence and Exposure Database
ASDTR	Agency for Toxic Substances and Disease Registry
AWWA	American Water Works Association
CCA	Chromated Copper Arsenate
CV	Coefficient of Variation
CWS	Community Water Supply
CWSS	Community Water Supply Surveys
DMAA	Dimethylarsinic Acid
DSMA	Disodium Methanearsonate
EPA	Environmental Protection Agency
FIFRA	Federal Insecticides, Fungicides, and Rodenticides Act
FRDS	Federal Reporting Data System
GW	Ground Water
IAOED	Intermediate Arsenic Occurrence and Exposure Database
ICP	Inductively Coupled Plasma
MCL	Maximum Contaminant Level
Metro	Metropolitan Water District of Southern California
MMAA	Monomethyl-Arsionic Acid
MS	Mass Spectrometry
MSMA	Monosodium Methanearsonate
NAS	National Academy of Sciences
NAOS	National Arsenic Occurrence Survey
NIPDWR	National Interim Primary Drinking Water Regulations
NIRS	National Inorganics and Radionuclides Survey
NOF	Natural Occurance Factor

NOMS	National Organics Monitoring Survey
NPDWR	National Primary Drinking Water Regulations
NPL	National Priorities List
NRC	National Research Council
NTNCWS	Non-Transient Non-Community Water Supply
NWIS	National Water Information System
OGWDW	Office of Ground Water and Drinking Water
POE	Points of Entry
PWS	Public Water Supply or Supplies
PWSID	Public Water Supplies Identification Number
RDS	Raw Data Sets
RIA	Regulatory Impact Analysis
ROS	Regression on Order Statistics
RWS	Rural Water Survey
SDWA	Safe Drinking Water Act
SDWIS	Safe Drinking Water Information System
SW	Surface Water
TMA	Trimethylarsine
TNCWS	Transient Non-Community Water Supplies
TOC	Total Organic Carbon
TRI	Toxics Release Inventory
USEPA	United States Environmental Protection Agency
USGS	United States Geological Society
VOC	Volatile Organic Compounds
WATSTORE	USGS's Water Quality Database
WESCAS	Western Coalition of Arid States
WITAF	Water Industry Technical Action Fund

Acknowledgments

Primary Authors and Contributors:
Kathleen Bell, ISSI Consulting Group
Jonathan Cohen, ICF Consulting
Stiven Foster, ISSI Consulting Group
Eric Hack, ICF Consulting
Robert Iwamiya, ICF Consulting
David Kaczka, ISSI Consulting Group
Jonathan Koplos, The Cadmus Group
Frank Letkiewicz, The Cadmus Group
Andrew Schulman, U.S. EPA OGWDW
Ben Smith, U.S. EPA OGWDW
Jennifer Wu, U.S. EPA OGWDW

Internal USEPA Consultants:
Timothy Barry, U.S. EPA OPPE
Irene S. Dooley, U.S. EPA OGWDW
Henry Kahn, U.S. EPA OST
Jade Lee, U.S. EPA OST
Elizabeth Margosches, U.S. EPA OPPTS

External Peer Reviewers:
Andrew Eaton, Montgomery Watson Laboratories, Pasadena, California
Chuck Kroll, SUNY College of Environmental Science and Forestry, Syracuse, New York
David Ruppert, Cornell University, Ithaca, New York

Executive Summary

The Safe Drinking Water Act (SDWA), 42 U.S.C. §§300f-300j, originally enacted in 1974, directs the U. S. Environmental Protection Agency (EPA) to identify and regulate contaminants in public drinking water. Section 1412(b)(12)(A) of SDWA, as amended by the 1996 Amendments, required EPA to propose a National Primary Drinking Water Regulation for arsenic by January 1, 2000, and to issue a final regulation by January 1, 2001. One of the elements that supports the development of the proposed regulation is information on arsenic occurrence in drinking water, specifically estimates of the size of populations and number of systems that are affected by different levels of arsenic in drinking water. This report presents the arsenic occurrence analysis, which was prepared for the EPA Office of Ground Water and Drinking Water (OGWDW) by ISSI under Contract 68-C7-0005, and by The Cadmus Group under Contract 68-C-99206, both with subcontract support by ICF Consulting.

Sources of Arsenic in the Environment

Arsenic is released to the environment from a variety of natural and anthropogenic sources. In the environment, arsenic occurs in rocks, soil, water, air, and in biota. Average concentrations in the earth's crust reportedly range from 1.5 to 5 mg/kg (Cullen and Reimer, 1989). Higher concentrations are found in some igneous and sedimentary rocks, particularly in iron and manganese ores (Welch *et al.*, 1988). In addition, a variety of common minerals contain arsenic, of which the most important are arsenopyrite (FeAsS), realgar (AsS), and orpiment (As_2S_3). Natural concentrations of arsenic in soil typically range from 0.1 to 40 mg/kg, with an average concentration of 5 to 6 mg/kg (National Academy of Sciences (NAS), 1977). Through erosion, dissolution, and weathering, arsenic can be released to ground water or surface water. Geothermal waters can be sources of arsenic in ground water, particularly in the Western United States (Nimick *et al.*, 1998, Welch *et al.*, 1988). Other natural sources include volcanism and forest fires.

Anthropogenic sources of arsenic relate to its use in the lumber, agriculture, livestock, and general industries. Most agricultural uses of arsenic are banned in the United States. However, organic arsenic is a constituent of the organic herbicides monosodium methanearsonate (MSMA) and disodium methanearsonate (DSMA), which are currently applied to cotton fields as herbicides (Jordan *et al.*, 1997). Organic arsenic is also a constituent of feed additives for poultry and swine, and appears to concentrate in the resultant animal wastes (NAS, 1977). The potential impact of arsenic in animal wastes used to fertilize crops is uncertain.

Most of the arsenic used in the United States is for the production of chromated copper arsenate (CCA), the wood preservative (Reese, 1998). CCA is used to pressure treat lumber and is classified as a restricted use pesticide by the USEPA. A significant industrial use of arsenic is the production of lead-acid batteries, while small amounts of very pure arsenic metal are used to produce the semiconductor crystalline gallium arsenide, which is used in computers and other electronic applications.

Arsenic is also released from industrial processes, including the burning of fuels and wastes, mining and smelting, pulp and paper production, glass manufacturing, and cement manufacturing (USEPA, 1998b). In addition, past waste disposal sites may be contaminated with arsenic. Arsenic is a contaminant of concern at 916 of the 1,467 sites on the National Priorities List (NPL) (Agency for Toxic Substances and Disease Registry (ATSDR), 1998). Sites included on the NPL have the potential to release contaminants to ground water or surface water in the vicinity of the site.

Anthropogenic releases of arsenic to the environment can be estimated from Toxics Release Inventory (TRI) data. These data indicate that 7,947,012 pounds of arsenic and arsenic-containing compounds were released to the environment in 1997, a significant increase from 3,536,467 pounds in 1995 (USEPA, 1999a). The increase primarily occurred at one facility, where arsenic on-site land releases increased by 3.58 million pounds from 1995 to 1997 because of a change in the facility's smelting process that was implemented to reduce sulfur dioxide emissions. The TRI data do omit some potentially significant arsenic sources, including arsenic associated with the application of herbicides and fertilizers and arsenic released from mining facilities and electric utilities.

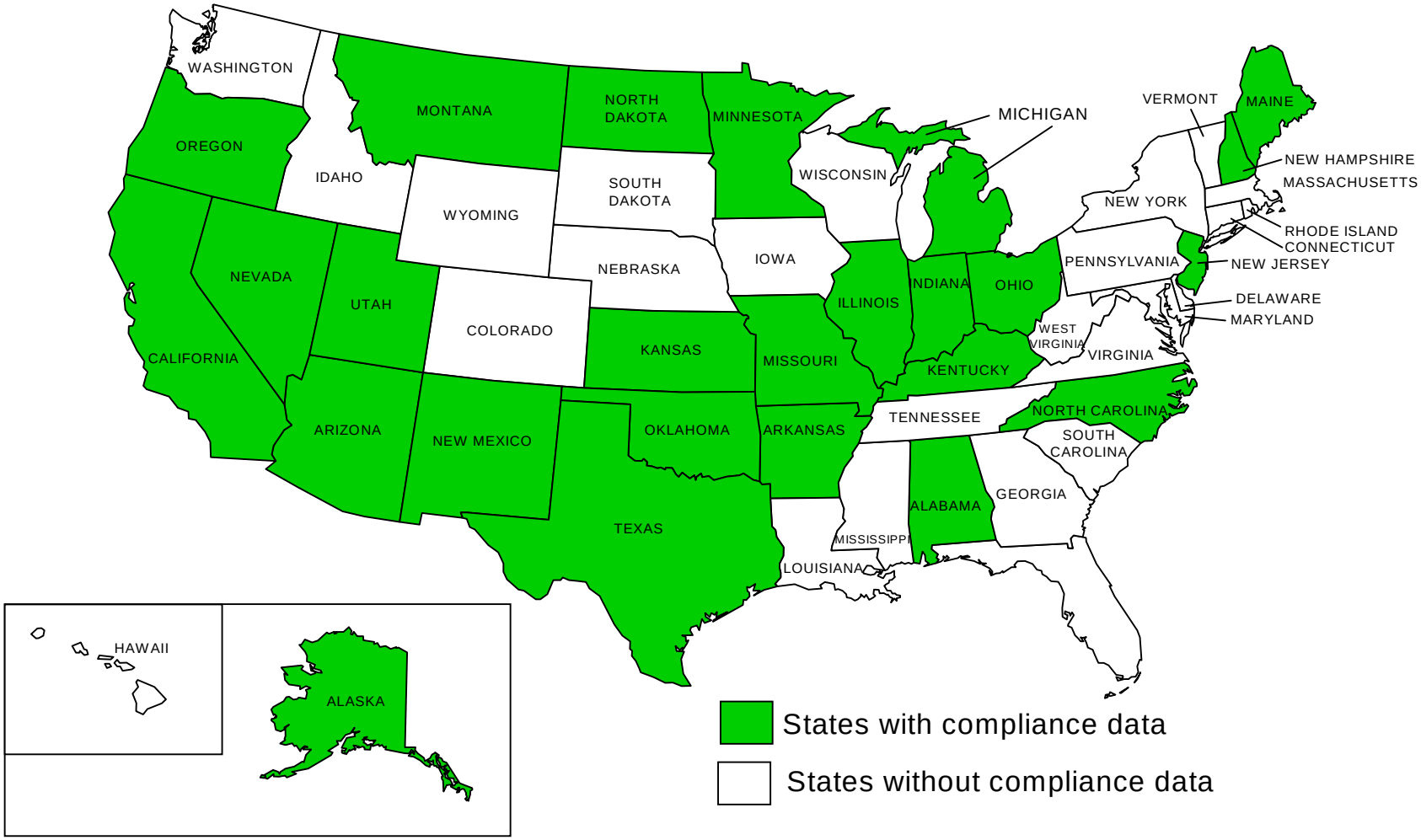
Arsenic Fate and Transport

Once arsenic released from natural or anthropogenic sources enters ground water or surface water, a variety of processes affect its fate and transport. These include oxidation-reduction reactions, transformations, ligand exchange, and biotransformations. The factors that affect these reactions include the oxidation state of the arsenic, oxidation-reduction potential (Eh), pH, concentrations of iron, metal sulfides, and sulfides, temperature, salinity, and the distribution and composition of biota (ATSDR, 1998; Robertson, 1989; Welch *et al.*, 1988). The predominant forms of arsenic in ground water and surface water are arsenate (+5) and arsenite (+3). Arsenite is generally associated with anaerobic conditions. Oxidation state appears to be the most important factor that determines the fate and transport of arsenic through drinking water treatment systems. Arsenate is more easily removed because of its ionic charge, and activated alumina, ion exchange, and reverse osmosis technologies can achieve relatively high arsenic removal rates. These technologies do not achieve comparable removal rates for arsenite. Oxidization of arsenite to arsenate can improve removal efficiencies. Treatment efficiencies may also be affected by water pH, depending on the technology applied, and competing ions. Higher pH tends to decrease removal rates (Rubel and Hathaway, 1987); high sulfate, fluoride, and phosphate concentrations also tend to decrease removal rates (Jekel, 1994).

Data Sources on Arsenic Occurrence in Drinking Water Systems

There are a variety of sources of information on arsenic in drinking water. This study is based largely on arsenic data from 25 State compliance monitoring data sets, and information on individual system characteristics that are provided in the Safe Drinking Water Information System (SDWIS). Figure ES-1 presents the States for which compliance monitoring data were available. As this figure shows, the Midwestern, South Central, North Central, and Western

Figure ES-1: States with Suitable Arsenic Compliance Monitoring Data



regions of the United States are well represented, but fewer compliance monitoring data sets are available for the States in the New England, Mid-Atlantic, and Southeastern regions.

These compliance monitoring data sets offer several benefits. For many States, they represent almost every ground water and surface water community water supply (CWS) system in the State. In total, the compliance monitoring data in these databases represent over 18,000 of the approximately 24,000 CWS systems that do not purchase water in the United States; there are a total of about 54,000 CWS systems in the United States. In addition, a smaller number of non-transient, non-community water supply (NTNCWS) systems are also represented in the State compliance monitoring data sets. These data sets contain multiple samples from individual systems, which can facilitate analysis of the variability in arsenic levels over time, or from location to location, or point-of-entry to point-of-entry, within individual systems. However, several of these data sets include samples that are censored[1] within, rather than below, the regulatory range of interest. To manage these multiple reporting levels and to calculate system mean arsenic levels, regression on order statistics was used in the data analyses presented in Chapters 5, 6, and 7 of this report.

Other arsenic databases are available for information on arsenic. Some are suitable for the development of national arsenic occurrence estimates, and others are less suitable for this purpose. The databases which may be suitable for the development of arsenic occurrence estimates include the National Arsenic Occurrence Survey (NAOS), the United States Geological Society (USGS) ambient ground water arsenic databases, the National Inorganics and Radionuclides Survey (NIRS), and the Metropolitan Water District of Southern California Survey (Metro). NAOS is based on a representative proportional stratified sampling design. It includes 517 raw water samples from ground water and surface water systems in the United States. The analytical method used had a detection level of 0.5 µg/L, below the regulatory range of interest (2.0 to 50 µg/L). Arsenic removal efficiencies associated with the treatment technologies in place in each system were used to calculate expected finished water arsenic concentrations. The USGS database is another relevant source of information on arsenic occurrence. It contains approximately 20,000 ambient ground water samples collected throughout the United States. These samples were analyzed with a consistent method which has a detection level of 1 µg/L, and according to consistent quality control and quality assurance protocols. Metro contains 144 samples which were primarily collected from ground water and surface water systems in the United States that serve populations of at least 50,000 people. These samples have a low detection limit, but are not associated with an individual public water supply identification number (PWS ID). The NAOS, USGS, NIRS, and Metro databases are used as comparison tools in Chapter 6 of this report.

[1] Censored data are samples with contaminant concentrations reported as less than the analytical detection limit. Actual contaminant concentrations in these samples may be positive, and may range from zero to the detection limit. In the case of a naturally occurring contaminant, such as arsenic, contaminant concentrations may be exceedingly low, but are rarely zero.

Other databases offer data that could be used to estimate arsenic occurrence, but were not used in this report. The NIRS includes samples from approximately 1,000 ground water systems in the United States. Most of the NIRS samples were collected from ground water systems that serve fewer than 3,300 people, and most of the samples (95 percent) are censored at 5 µg/L. For these reasons, the NIRS database was not used in the development of arsenic occurrence estimates presented in this report.

Because several other databases contain very old arsenic sample results and a high proportion of the results were censored, they were not used in this occurrence analysis. These databases include the 1969 and 1978 Community Water Supply Surveys (CWSS), the Rural Water Survey (RWS), and the National Organics Monitoring Survey (NOMS). Because these results were quite old and highly censored, they were not used in this occurrence estimation. Another database, the Western Coalition of Arid States (WESCAS) database, also was not used in this analysis because the data did not necessarily represent arsenic levels at individual PWS, and because the data conventions used appeared to have been inconsistent from State to State.

Patterns of Arsenic Occurrence in Drinking Water

The data were analyzed with respect to a variety of potential stratification variables, including source water type, system size and type, and regional stratification. Distributions of arsenic in ground water and surface water systems were clearly different; therefore, the occurrence analyses were stratified on the basis of source water type. There were some differences between arsenic levels for CWS and NTNCWS systems, although there was limited data for surface water NTNCWS systems. Therefore, the occurrence analyses for ground water were stratified by system type and the occurrence analyses for surface water were based only on the CWS data. Other authors, who have evaluated regional differences in arsenic occurrence, concluded that arsenic levels may differ from region to region. Regional stratification was applied in these occurrence analyses. State compliance monitoring data sets were stratified into the 7 regions that were identified by Frey and Edwards (1997). This stratification scheme is convenient because the State compliance monitoring data can be easily sorted and evaluated by State. Regional stratification would be unnecessary if data were available for all 50 States. However, because average arsenic concentrations may differ from region to region, and because the representation of States within each region differs from region to region, regional stratification was applied to control these differences and to yield more accurate occurrence estimates.

Predicted Number of Ground Water and Surface Water Systems Exceeding Potential Regulatory Levels

Using the State compliance monitoring data, estimates of the proportions and numbers of systems that may exceed specific maximum contaminant level (MCL) alternatives were developed. Separate estimates were developed for ground water and surface water systems, although both estimates were developed through a similar five-step process. In the first step, system mean concentrations are estimated for each system in the compliance monitoring database. Second, estimates of exceedance probabilities were developed for each State. Third, State estimates were grouped and weighted, in order to develop regional arsenic occurrence

estimates. Fourth, regional estimates were weighted to develop national estimates of the proportion of systems which are likely to have mean system arsenic levels above specific concentrations of interest. Fifth, estimated exceedance probabilities were multiplied by the total number of ground water or surface water systems in the United States to estimate the total number of systems with various system mean arsenic concentrations. Due to the limited amount of available data, the exceedance probability estimates for surface water CWS systems were used to estimate the exceedance probabilities for surface water NTNCWS systems. As shown in Tables ES-1, ES-2, ES-3, and ES-4, these estimates are based on the number of systems in specific size categories.

Under these estimates, 11,873 ground water CWS systems are estimated to have mean arsenic levels that exceed 2 µg/L. The number of systems with mean arsenic levels above the potential MCL alternatives decreases rapidly as the potential MCL alternative concentration increases, but 5,252 systems are predicted to have mean arsenic levels greater than 5 µg/L, and 2,302 systems are predicted to have mean arsenic concentrations greater than 10 µg/L. Arsenic occurrence is projected to be significantly lower in surface water systems. For example, 1,052 surface water CWS systems are predicted to have arsenic levels above 2 µg/L, 325 surface water CWS systems are predicted to be out of compliance with an arsenic MCL of 5 µg/L, and 86 are predicted to exceed an alternative MCL of 10 µg/L. For NTNCWS ground water systems, 6,306, 3,064, and 1,050 systems are predicted to have mean arsenic concentrations greater than 2, 5, and 10 µg/L, respectively. For NTNCWS surface water systems, 80, 25, and 7 systems are predicted to have mean arsenic concentrations greater than 2, 5, and 10 µg/L, respectively.

These arsenic estimates resemble other recently generated arsenic occurrence estimates. At concentrations of 5 and 10 µg/L, these exceedance estimates are quite similar to estimates developed by Frey and Edwards (1997). These estimates are also similar to, although slightly lower than, the number of ground water systems that are estimated to be impacted at concentrations of 5 and 10 µg/L based on the USGS database. However, the arsenic occurrence estimates presented in this report are higher than estimates that were developed in 1992 for the USEPA (Wade Miller, 1992).

An uncertainty analyses was conducted to determine the potential amount of error in the exceedance probability estimates. To determine 95 percent confidence intervals, it was necessary to perform a statistical simulation to quantify the potential sources of uncertainty. Three sources of uncertainty were identified and simulated: 1) sampling variability, within and between systems; 2) the fill-in of censored observations in the estimation of system means; and 3) fitting of lognormal distribution to populations of system means within each State.

Intra-system Variability

The purpose of the intra-system analysis is to facilitate prediction of the number of points-of-entry or POE that will be affected by various MCL alternatives. Compliance with the arsenic standard is measured at the point-of-entry to the distribution system, and individual systems can have multiple points-of-entry. Arsenic levels in POE drive compliance costs and risk reduction benefits more directly than do system mean arsenic levels. Data are not available

Table ES-1
Estimated Arsenic Occurrence in U. S. Ground Water CWS

		Number of Systems with Arsenic Concentrations[2] (µg/L) of:									
System Size (Population Served)	**Total Number of Systems[1]**	**> 2**	**> 3**	**> 5**	**> 10**	**> 15**	**> 20**	**> 25**	**> 30**	**> 40**	**> 50**
< 25	178	49	35	22	9	5	4	3	2	1	1
25-100	14025	3833	2788	1696	743	429	281	199	147	90	60
101-500	14991	4097	2980	1812	795	459	300	213	157	96	64
501-1,000	4671	1277	929	565	248	143	93	66	49	30	20
1,001-3,300	5710	1561	1135	690	303	175	114	81	60	37	25
3,301-10,000	2459	672	489	297	130	75	49	35	26	16	11
10,001-50,000	1215	332	242	147	64	37	24	17	13	8	5
50,001-100,000	131	36	26	16	7	4	3	2	1	1	1
100,001-1,000,000	61	17	12	7	3	2	1	1	1	0	0
> 1,000,000	2	1	0	0	0	0	0	0	0	0	0
Total Systems	**43443**	**11873**	**8636**	**5252**	**2302**	**1329**	**869**	**617**	**456**	**278**	**187**
Lower 95% CI:		**11543**	**8363**	**5100**	**2250**	**1269**	**821**	**573**	**421**	**252**	**165**
Upper 95% CI:		**13007**	**9501**	**5665**	**2567**	**1499**	**995**	**712**	**534**	**335**	**226**

Notes:
CI: confidence interval
1 Based on 1998 Baseline SDWIS data for purchased and non-purchased systems. Systems characterized as GW under the influence of SW are considered to be surface water systems.
2 Based on national weighted point estimates presented in Table 6-3a.
3 Totals may not add up due to rounding of the number of systems to the nearest whole number.

Table ES-2
Estimated Arsenic Occurrence in U. S. Surface Water CWS

		Number of Systems with Arsenic Concentrations[2] (μg/L) of:									
System Size (Population Served)	**Total Number of Systems[1]**	**> 2**	**> 3**	**> 5**	**> 10**	**> 15**	**> 20**	**> 25**	**> 30**	**> 40**	**> 50**
< 25	74	7	4	2	1	0	0	0	0	0	0
25-100	1001	98	56	30	8	5	3	2	2	1	1
101-500	1983	195	110	60	16	9	6	5	4	3	2
501-1,000	1219	120	68	37	10	6	4	3	2	2	1
1,001-3,300	2420	238	135	73	19	11	8	6	5	3	2
3,301-10,000	1844	181	103	56	15	9	6	4	3	2	2
10,001-50,000	1606	158	89	49	13	7	5	4	3	2	2
50,001-100,000	300	29	17	9	2	1	1	1	1	0	0
100,001-1,000,000	261	26	15	8	2	1	1	1	0	0	0
> 1,000,000	13	1	1	0	0	0	0	0	0	0	0
Total Systems	**10721**	**1052**	**597**	**325**	**86**	**50**	**34**	**26**	**20**	**14**	**10**
Lower 95% CI:		**973**	**514**	**193**	**56**	**25**	**14**	**9**	**6**	**3**	**2**
Upper 95% CI:		**2730**	**2212**	**1036**	**167**	**107**	**88**	**77**	**71**	**65**	**63**

Notes:
CI: confidence interval
1 Based on 1998 Baseline SDWIS data for purchased and non-purchased systems. Systems characterized as GW under the influence of SW are considered to be surface water systems.
2 Based on national weighted point estimates presented in Table 6-3a.
3 Totals may not add up due to rounding of the number of systems to the nearest whole number.

Table ES-3
Estimated Arsenic Occurrence in U. S. Ground Water NTNCWS

		Number of Systems with Arsenic Concentrations[2] (µg/L) of:									
System Size (Population Served)	**Total Number of Systems[1]**	**> 2**	**> 3**	**> 5**	**> 10**	**> 15**	**> 20**	**> 25**	**> 30**	**> 40**	**> 50**
< 25	31	10	8	5	2	1	1	0	0	0	0
25-100	9732	3126	2358	1519	520	306	203	145	108	67	46
101-500	7103	2281	1721	1109	380	223	148	105	79	49	34
501-1,000	1996	641	484	312	107	63	42	30	22	14	9
1,001-3,300	696	224	169	109	37	22	14	10	8	5	3
3,301-10,000	62	20	15	10	3	2	1	1	1	0	0
10,001-50,000	15	5	4	2	1	0	0	0	0	0	0
50,001-100,000	0	0	0	0	0	0	0	0	0	0	0
100,001-1,000,000	0	0	0	0	0	0	0	0	0	0	0
> 1,000,000	0	0	0	0	0	0	0	0	0	0	0
Total Systems	**19635**	**6306**	**4758**	**3064**	**1050**	**617**	**409**	**292**	**219**	**136**	**93**

Notes:
1 Based on 1998 Baseline SDWIS data for purchased and non-purchased systems. Systems characterized as GW under the influence of SW are considered to be surface water systems.
2 Based on national weighted point estimates presented in Table 6-3b.
3 Totals may not add up due to rounding of the number of the systems to the nearest whole number.

Table ES-4
Estimated Arsenic Occurrence in U. S. Surface Water NTNCWS

		Number of Systems with Arsenic Concentrations[2] (µg/L) of:									
System Size (Population Served)	**Total Number of Systems[1]**	**> 2**	**> 3**	**> 5**	**> 10**	**> 15**	**> 20**	**> 25**	**> 30**	**> 40**	**> 50**
< 25	5	0	0	0	0	0	0	0	0	0	0
25-100	280	27	16	9	2	1	1	1	1	0	0
101-500	314	31	17	10	3	1	1	1	1	0	0
501-1,000	107	11	6	3	1	0	0	0	0	0	0
1,001-3,300	80	8	4	2	1	0	0	0	0	0	0
3,301-10,000	23	2	1	1	0	0	0	0	0	0	0
10,001-50,000	5	0	0	0	0	0	0	0	0	0	0
50,001-100,000	1	0	0	0	0	0	0	0	0	0	0
100,001-1,000,000	1	0	0	0	0	0	0	0	0	0	0
> 1,000,000	0	0	0	0	0	0	0	0	0	0	0
Total Systems	**816**	**80**	**45**	**25**	**7**	**4**	**3**	**2**	**2**	**1**	**1**
Lower 95% CI:		**74**	**39**	**15**	**4**	**2**	**1**	**1**	**0**	**0**	**0**
Upper 95% CI:		**208**	**168**	**79**	**13**	**8**	**7**	**6**	**5**	**5**	**5**

Notes:
1 Based on 1998 Baseline SDWIS data for purchased and non-purchased systems. Systems characterized as GW under the influence of SW are considered to be surface water systems.
2 Based on national weighted point estimates presented in Table 6-3b. These estimates were derived from CWS SW data.
3 Totals may not add up due to rounding of the number of systems to the nearest whole number.

that would allow the development of directly representative estimates of arsenic occurrence. Instead, this report uses compliance monitoring data with POE identifiers to quantify a relationship between POE means and system means, so that the number of POE means in the United States that are likely to exceed specific regulatory alternatives can be calculated from a distribution of system means. This relationship was summarized by the estimated coefficient of variation (CV), or relative standard deviation, for the intra-system variability. The approach used was based on fitting a statistical model to account both for the intra-system variability between different POE and for temporal and analytic variability between multiple measurements at the same POE. The estimated intra-system variability CVs were 37 % for ground water CWS systems, 53 % for surface water CWS and NTNCWS systems, and 25 % for ground water NTNCWS systems. The CV values that were calculated under these analyses were used in a regulatory impact analyses (RIA) conducted under a separate work assignment to estimate the number of POE that may exceed regulatory alternatives.

1. Introduction

Under the Safe Drinking Water Act (SDWA), 42 U.S.C. §§300f-300j, originally enacted in 1974, arsenic is a regulated drinking water contaminant. In 1975, U.S. Environmental Protection Agency (USEPA) issued a National Interim Primary Drinking Water Regulation (NIPDWR) for arsenic at 50 μg/L (USEPA, 1975). This value is the current Maximum Contaminant Level (MCL) for arsenic in drinking water. In 1986, Congress converted the NIPDWR for arsenic to a National Primary Drinking Water Regulation (NPDWR), and directed USEPA to revise National Primary Drinking Water Regulations. In 1994, following a consent decree in a suit[2] between USEPA and Citizens Concerned about Bull Run, Inc., a citizens' group, USEPA organized an internal workgroup for the purpose of addressing risk assessment, treatment, analytical methods, arsenic occurrence, exposure, costs, implementation issues, and regulatory options for arsenic. In 1995, USEPA deferred the proposal of the regulation in order to better characterize the human health effects associated with chronic low-level exposure to arsenic and treatment costs. In accordance with Safe Drinking Water Act (SDWA), as amended in 1996, Section 1412(b)(12)(A) directs USEPA to propose a National Primary Drinking Water Regulation for arsenic by January 1, 2000, and to issue a final regulation by January 1, 2001.

Arsenic (As) is a metallic element that occurs at low concentrations in most rocks and soils (Yan-Chu, 1994). To a small extent, it occurs in the elemental state; however, higher concentrations of arsenic principally occur in mineral complexes with metals and other elements (Welch *et al.*, 1988). For example, arsenic is a common impurity in the sulfide ores of lead, copper, and zinc. Arsenic is released into the environment from natural processes such as the weathering and dissolution of arsenic-containing minerals and ores (Yan-Chu, 1994). In addition to its release from natural sources, arsenic is released from a variety of anthropogenic sources (USEPA, 1998b), including:

- Manufacturing of metals and alloys
- Petroleum refining
- Pharmaceutical manufacturing
- Pesticide manufacturing and application
- Chemicals manufacturing
- Burning of fossil fuels
- Waste incineration

These anthropogenic releases of arsenic can elevate environmental arsenic concentrations.

Human exposure to arsenic can result in a variety of chronic and acute effects. In particular, there is evidence that associates chronic arsenic ingestion at low concentrations with increased risk of skin cancer, and that arsenic may cause cancers of the lung, liver, bladder, kidney, and colon (ATSDR, 1998). Because of the human health risks associated with arsenic, USEPA regulates the level of arsenic in drinking water.

[2] *Miller v. USEPA*, No. 89-CV-6328 (D. Ore., filed August 31, 1989).

The objective of this study for the USEPA's Office of Ground Water and Drinking Water (OGWDW) is to estimate arsenic occurrence in public water supplies (PWS) in relation to various MCL alternatives. Estimates of the number of people exposed to various concentrations of arsenic in drinking water will be presented separately, and are not included in this report. The arsenic occurrence estimates are also bounded by 95 percent confidence intervals. These estimates will be significant in the development of the proposed arsenic regulation.

1.1 Purpose of this Document

This report summarizes the results of the arsenic occurrence analysis conducted by ISSI Consulting Group and The Cadmus Group, with their subcontractor, ICF Consulting for the USEPA's OGWDW. The estimates of arsenic occurrence presented in this report differ from those presented in other studies because of their strong reliance on existing compliance monitoring data that were voluntarily provided to USEPA and its contractors. In addition, new techniques have been applied to evaluate the statistical distributions of arsenic in drinking water to estimate percentages of regulatory exceedances, to estimate the variability in arsenic levels within systems, and to estimate the relative uncertainty associated with these predictions.

1.2 Organization of the Document

This report is organized in seven sections that are relevant to the estimation of arsenic occurrence in drinking water in the United States. The remaining sections of this document are organized as follows:

Chapter 2: *Sources of Arsenic* identifies naturally occurring and anthropogenic sources of arsenic in the environment, with a particular focus on sources of arsenic to drinking water.

Chapter 3: *Fate and Transport of Arsenic* presents information on the physical and chemical characteristics of arsenic and the relation between those properties and the presence of arsenic in source waters. In addition, this section presents an overview of the potential fate and transport of arsenic within treatment and distribution systems.

Chapter 4: *Sources of Data on Arsenic Occurrence in Drinking Water Supplies* presents a summary of the approaches used to identify and select data on arsenic occurrence for use in this occurrence assessment. In addition, this section presents summary information on the data sources that were used in the occurrence assessment.

Chapter 5: *Arsenic Occurrence Patterns in the United States* discusses the analyses that were applied to identify patterns in the data and the conclusions developed as a result of those analyses.

Chapter 6: *National Occurrence Estimates* presents estimates of the number of systems projected to exceed specific arsenic levels, describes the method used to develop these estimates, and discusses the uncertainty in these estimates.

Chapter 7: *Intra-system Variability Assessment* provides an overview of the variations in arsenic levels from location to location within public water supply systems.

Chapter 8: *Temporal Variability Analysis* examines the variability of arsenic concentrations over time in a source.

This page intentionally left blank

2. Sources of Arsenic

This section discusses the physical and chemical properties of arsenic, and the natural and anthropogenic sources of arsenic to the environment, particularly sources that may affect drinking water in the United States. The primary natural sources include the earth's crust, soil and sediment, geothermal activity, and volcanic activity. The most significant anthropogenic sources are agricultural, industrial, and mining activities.

2.1 Physical and Chemical Properties of Arsenic

Arsenic (As) is a silver-gray brittle crystalline solid (Budavari, *et al.* 1989). It also exists in black and yellow amorphous forms. Arsenic appears in Group 15 on the periodic table in the first long period with a d-shell just below the valence shell. Arsenic has an atomic weight of 74.9216 and an atomic number of 33. Silver-gray arsenic has a specific gravity of 5.73; a melting point of 817 °C (28 atm) and sublimes at 613°C. The yellow amorphous form of arsenic has a specific gravity of 1.97. Elemental arsenic can be present as a metalloid, although arsenic has an elemental structure similar to non-metals. In the vapor state, arsenic occurs as a tetrameric molecule (As_4). In high oxidation states arsenic displays covalent tendencies, while in low oxidation states it shows ionic tendencies (Ferguson, 1990). The physical and chemical properties of arsenic are summarized in Table 2-1.

The valence states of As are: -3, 0, +1, +3, and +5 (Welch *et al.*, 1988). Elemental arsenic (valence 0) is rarely found under natural conditions. The +3 and +5 states are found in a variety of minerals and in natural waters. Many of the chemical behaviors of arsenic are linked to the ease of conversion between +3 and +5 valence states (National Research Council (NRC), 1999). The valence state affects the toxicity of arsenic compounds. While arsine (-3) is the most toxic, the following are successively less toxic: organo-arsines, arsenites (+3), arsenates (+5), arsonium metals (+1), and elemental arsenic (0).

2.1.1 Environmentally Relevant Arsenic Species

Arsenic occurs naturally as a constituent of a number of different compounds in both marine and terrestrial environments. Arsenic species are classified as either organic or inorganic. If carbon is present within the compound it is considered to be an organic arsenic species. Table 2-2 includes a summary of both organic and inorganic species of arsenic which may be found in food and water.

Inorganic Arsenic

Inorganic arsenic, with +5 (arsenate) and +3 (arsenite) oxidation states, is more prevalent in water than organic arsenic (Irgolic, 1994; Clifford and Zhang, 1994). The dominant arsenic species depends on pH and redox conditions. In general +5 predominates under oxidizing conditions and +3 predominates under reducing conditions (ATSDR, 1998; Clifford and Zhang, 1994).

Table 2-1
Physical and Chemical Properties of Arsenic

CAS Number	7440-38-2
Atomic Number	33
Atomic Weight	74.92
Melting Point at 28 atm	817°C
Boiling Point	613°C
Critical Temperature	1,400°C
Heat of Vaporization	11.2 kcal/g-atom
Critical Pressure	22.3 MPa
Density (at 14°C)	5.727 g/cm^3
Most Stable Isotope	^{75}As
Covalent Radius	1.19 angstroms
Atomic Radius	1.39 angstroms
Ionic Radius	2.22 angstroms
Vapor Pressure	1 mm (375°C) 10 mm (437°C) 100 mm (518°C)

Excerpted from Budavari *et al.*, 1989

Table 2-2
Inorganic and Organic Arsenic Compounds

Name	Abbreviation	Chemical Formula
Arsanilic acid	--	$C_6H_8AsNO_3$
Arsenous acid	As(III)	H_3AsO_3
Arsenic acid	As(V)	H_3AsO_4
Monomethylarsonic acid	MMAA	$CH_3AsO(OH)_2$
Methylarsonous acid	MMAA(III)	$CH_3As(OH)_2[CH_3AsO]_n$
Dimethylarsinic acid	DMAA	$(CH_3)_2AsO(OH)$
Dimethylarsinous acid	DMAA(III)	$(CH_3)_2AsOH$ $[((CH_3)_2As)_2O]$
Roxarsone	--	$C_6H_6AsNO_6$
Trimethylarsine	TMA	$(CH_3)_3As$
Trimethylarsine oxide	TMAO	$(CH_3)_3AsO$
Tetramethlarsonium ion	Me_4As^+	$(CH_3)_4As^+$
Arsenocholine	AsC	$(CH_3)_3As^+CH_2CH_2OH$
Arsenobetaine	AsB	$(CH_3)_3As^+CH_2COO^-$
Arsenic-containing ribo-sides	Arsenosugar X-XV[a] Arsenolipid[b]	

Excerpted from NRC, 1999

Examples of inorganic arsenic compounds found in the environment include oxides (i.e. As_2O_3, As_2O_5, $R^3AsO)_n$, $R_nAsO(OH)_{3-n}$ (n=1,2)) and sulfides (As_2S_3, AsS, $HAsS_2$, $HAsS_3^{3-}$) (Cullen and Reimer, 1989). Inorganic arsenic species which are stable in oxygenated waters include arsenic acid (As(V)) species (i.e. H_3AsO_4, $H_2AsO_4^-$, $HAsO_4^{2-}$ and AsO_4^{3-}). Arsenous acid (As(III)) is also stable as H_3AsO_3 and $H_2AsO_3^-$ under slightly reducing aqueous conditions.

In addition to geochemical factors, microbial agents can influence the oxidation state of arsenic in water, and can mediate the methylation of inorganic arsenic to form organic arsenic compounds. Microorganisms can oxidize arsenite to arsenate, reduce arsenate to arsenite, or reduce arsenate to arsine (Cullen and Reimer, 1989). Bacterial action also oxidizes minerals such as orpiment (As_2S_3), arsenopyrite (FeAsS), and enargite (Cu_3AsS_4) releasing arsenate. Under aerobic conditions, the common aquatic bacterium *Pseudomonas fluorescens* reduces arsenate to arsenite. In a river in New Zealand, investigators found the predominant oxidation state of arsenic varied seasonally because of (at least in part) the bacterium *Anabaena oscillaroides* which reduces arsenate to arsenite. Arsenite was found to predominate in spring and summer months, while arsenate was prevalent at other times of the year.

Organic Arsenic

Organic arsenic compounds such as Monomethylarsonic acid (MMAA), Dimethylarsinic acid (DMAA), Trimethylarsine (TMA), and Trimethylarsine oxide (TMAO) are generally associated with terrestrial settings, however, some are found in water (NRC, 1999). Organic arsenic is produced naturally in the environment in natural gas (ethylmethylarsines), shale oil, in water when microorganisms metabolize inorganic arsenic, and in the human body, as a result of enzyme activity in the liver (USEPA, 1993; Berger and Fairlamb, 1994).

In studies of arsenic speciation in natural waters reviewed by the National Research Council (1999), organic arsenical compounds were reported to have been detected in surface water more often than in ground water. Surface water samples reportedly contain low but detectable concentrations of arsenic species such as MMAA, DMAA, Arsenocholine (AsC), TMAs, and species similar to Arsenobetaine (AsB). Methylarsenicals have been reported to comprise as much as 59% of total arsenic in lake water. In some lakes, DMAA has been reported as the dominant species, and concentrations appear to vary seasonally as a result of biological activity within waters, with the highest concentrations observed in May and June.

2.2 Natural Sources of Arsenic

Arsenic occurs in the environment in rocks, soil, water, air, and in biota; and concentrations of arsenic in a variety of environmental media are presented in Table 2-3. The following sections discuss important natural sources of arsenic in the environment. Most arsenic in the environment exists in rock or soil (ATSDR, 1998). Because arsenic occurs naturally in rock, soil and sediment, these sources are particularly important determinants of regional levels of arsenic in ground water and surface water. Natural sources of arsenic are discussed below,

[3] R=H, Me, Cl, etc.

Table 2-3
Arsenic Concentrations in Environmental Media

Environmental Media	Arsenic Concentration Range	Units
Air	1.5 - 53	ng/m^{-3}
Rain from unpolluted ocean air	0.019	µg/L
Rain from terrestrial air	0.46	µg/L
Rivers	0.20-264	µg/L
Lakes	0.38-1,000	µg/L
Ground water	≺ 1.0 - ≻ 1,000	µg/L
Sea water	0.15-6.0	µg/L
Soil	0.1-1,000	mg/kg
Stream/river sediment	5.0-4,000	mg/kg
Lake sediment	2.0-300	mg/kg
Igneous rock	0.3-113	mg/kg
Metamorphic rock	0.0-143	mg/kg
Sedimentary rock	0.1-490	mg/kg
Biota - Green Algae	0.5-5.0	mg/kg
Biota - Brown Algae	30	mg/kg

Excerpted from NAS, 1977.

and anthropogenic sources are discussed subsequently in Section 2.3, and the cycling of arsenic in the environment is depicted in Figure 2-1.

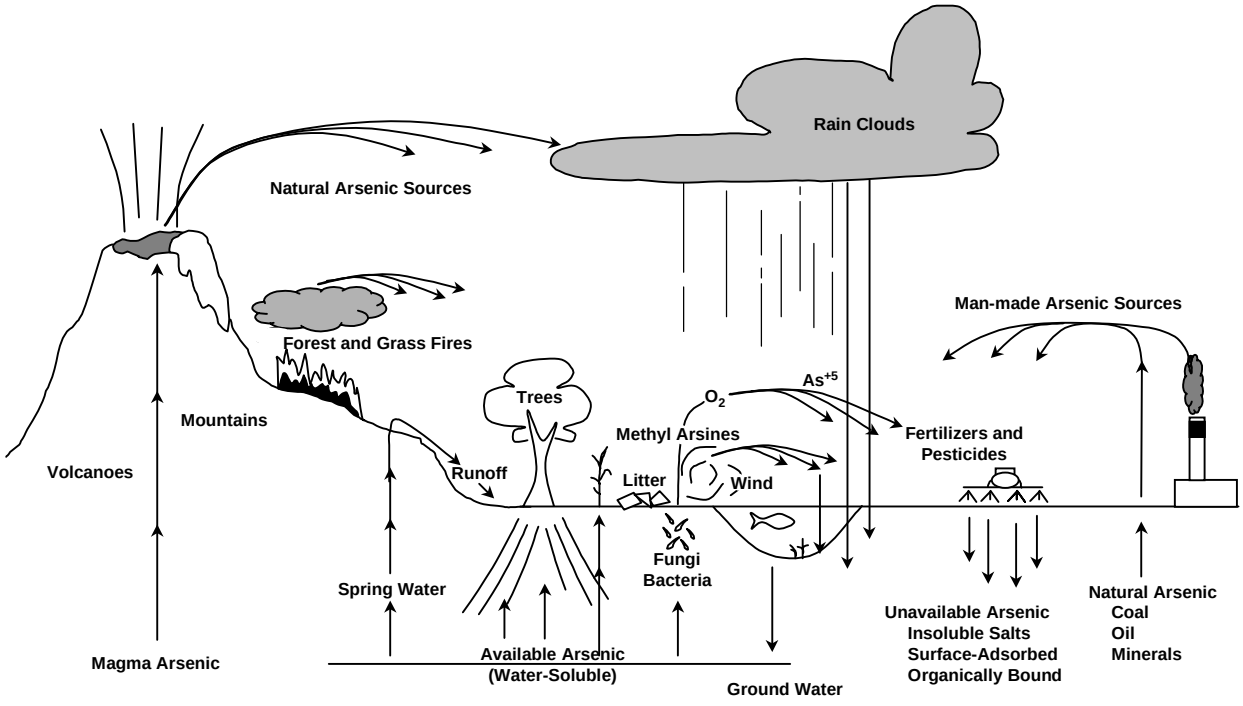


FIGURE 2-1. Environmental Transfer of Arsenic.

Earth's Crust

Arsenic is the twentieth most abundant element in the earth's crust (ATSDR, 1998; NAS, 1977). Concentrations of arsenic in the earth's crust vary, but average concentrations are generally reported to range from 1.5 to 5 mg/kg (ATSDR, 1998; Cullen and Reimer, 1989; NAS, 1977). Arsenic is a major constituent of many mineral species in igneous and sedimentary rocks; Table 2-4 presents concentrations of arsenic in igneous and sedimentary rocks. Among igneous rock types, the highest arsenic concentrations are found in basalts. Sedimentary rocks, particularly iron and manganese ores, often contain higher average arsenic concentrations than igneous rocks (Welch *et al.*, 1988).

Table 2-5, above, lists common minerals that contain arsenic. Arsenopyrite (FeAsS), realgar (AsS), and orpiment (As_2S_3) are the most important of these minerals, and they are commonly present in the sulfide ores of other metals including copper, lead, silver and gold (Yan-Chu, 1994). Arsenic may be released from these ores to soil (Yan-Chu, 1994), surface water (Mok and Wai, 1989), ground water (Welch *et al.*, 1988), and the atmosphere (ATSDR, 1998).

Table 2-4
Arsenic in Igneous and Sedimentary Rocks

Rocks	No. Analyses	Arsenic Concentration (mg/kg)	
		Range Usually Reported	Average
Igneous Rocks:			
Ultrabasic	37	0.3–16	3.0
Basalts, gabbros	146	0.06–113	2.0
Andesites, dacites	41	0.5–5.8	2.0
Granitic	73	0.2–13.8	1.5
Silicic volcanic	52	0.2–12.2	3.0
Sedimentary Rocks:			
Limestones	37	0.1–20	1.7
Sandstones	11	0.6–120	2.0
Shales and clays	324	0.3–490	14.5[a]
Phosphorites	282	0.4–188	22.6
Sedimentary iron ores	110	1–2,900	400
Sedimentary manganese ores	–	(up to 1.5%)	–
Coal	1,150	0–2,000	4-13

Adapted from NAS, 1977
[a]Excludes one sample containing As at a concentration of 490 mg/kg.

Table 2-5
Common Minerals of Arsenic

Arsenopyrite, FeAsS	Smalite, $CoAs_2$
Lollingite, $FeAs_2$	Cobaltite, CoAsS
Orpiment, As_2S_3	Gersdorffite, NiAsS
Realgar, As_4S_4	Tennantite, $4Cu_2SAs_{2S}S_3$
Chloanthite, $NiAs_2$	Proustite, $3Ag_2SAs_2S_3$
Niciolite, NiAs	Enargite, $3Cu_2SAs_2S_5$

Adapted from Ferguson, 1990.

In their evaluation of the regional distribution of arsenic in ground water in the Western United States, Welch *et al.,* (1988) evaluated the association between aquifer geology and arsenic concentrations in ground water. Higher arsenic concentrations (ground water concentrations greater than 50 µg/L) were associated with sedimentary deposits derived from volcanic rocks. These geological conditions occurred at a few locations in the Western Mountain Ranges (notably near Reno, Nevada, and Eugene, Oregon), and in the Alluvial Basins. Within the Alluvial Basins, elevated arsenic concentrations were associated with sediments derived from volcanic rocks rather than non-sedimentary and unmineralized volcanic rock. Weathering of the volcanic rocks may result in the concentration of arsenic onto ferric oxyhydroxide that are deposited with sediments (Welch *et al.,* 1988). Lower arsenic concentrations are associated with regions underlain by carbonate rocks and volcanic basalts. The regions with moderate arsenic levels include parts of the Alluvial Basins in western Utah and eastern Nevada and the Columbia Lava Plateau.

In other parts of the United States, high (greater than 50 µg/L in ground water) concentrations of arsenic are recognized to be associated with other geological formations. In Eastern Michigan and Northeastern Wisconsin, high arsenic concentrations are associated with sulfide mineral deposits in sedimentary rocks (Westjohn *et al.,* 1998; Simo *et al.,* 1996), and in the upper Midwest, higher arsenic concentrations are associated with iron oxide rich sedimentary deposits. Moderate to high arsenic concentrations (10 to 50 µg/L in ground water) in portions of the Northeastern United States (Massachusetts to Maine) appear to be related to sulfide minerals in the bedrock aquifers (Marvinney *et al.,* 1994).

Soil and Sediment

Arsenic concentrations in soils depend in part on the parent materials from which the soils were derived, although they may be enriched by other sources, including anthropogenic sources. Typical natural concentration ranges are 0.1 to 40 mg/kg, with an average concentration of 5-6 mg/kg (NAS, 1977). The level of arsenic in soil derived from basalts tends to be higher than in soils of granitic origin, and concentrations of 20 to 30 mg/kg may be found in soils derived from sedimentary rocks (Yan-Chu, 1994). In areas of recent volcanism, soils average arsenic concentrations are approximately 20 mg/kg. Very high natural concentrations of arsenic (up to 8,000 mg/kg) may occur in soils that overlay deposits of sulfuric ores (NAS, 1977). Arsenic can be found in soil in the inorganic state bound to cations, and it can also be found bound to organic matter. Arsenic may be transferred to surface water and ground water through erosion and dissolution; plants may also uptake arsenic. Because arsenic can be fixed in inorganic and organic compounds in soil, soil may also be a sink for arsenic.

In bottom sediments of rivers and lakes, concentrations of arsenic in surface sediments tend to exceed those found in deeper sediments due to diagenetic cycling (Nimick *et al.,* 1998). Recent anthropogenic arsenic releases may also result in the elevation of arsenic concentrations in surface sediments. This phenomenon has been observed in Lake Michigan (NAS, 1977). In the Madison and Missouri River Basins, Nimick *et al.,* documented average arsenic concentrations in bottom sediments that ranged from 7 to 102 mg/kg. These concentrations differed substantially over the length of the river system. In Lake Michigan, average concentrations of arsenic in surficial sediment were reported to be 12.4 mg/kg (NAS, 1977). In

sediments contaminated by mining and industrial activities, arsenic concentrations can be greatly enriched. In creeks affected by mining activities in Idaho, Mok and Wai (1989) found arsenic concentrations that ranged from 42.1 to 2550.4 mg/kg. In shallow waters, wave action and seasonal high flow scouring can result in resuspension of arsenic rich surface sediments, whereas in deeper lakes, arsenic may be permanently sequestered in sediment (Nimick *et al.,* 1998). Arsenic may also be released from bottom sediments as a result of microbial action.

Arsenic in soil and sediment may undergo microbial degradation or transformation. In soil, arsenic in the form of arsenates, arsenites, monomethyl-arsionic acid (MMAA) or dimethylarsinic acid (DMAA) may be biotransformed to arsine gases (Yan-Chu, 1994). These arsine gases are subsequently volatilized to the environment. In sediment, biologically mediated methylation of arsenates increases the solubility of arsenic, and may increase arsenic concentrations in water (Mok and Wai, 1994). Conversely, the biologically mediated demethylation of DMAA and MMAA can result in the formation of arsenates, which are strongly adsorbed onto sediments.

Geothermal Waters

Geothermal water can be sources of arsenic in surface water and ground water. Welch *et al.,* (1988) identified 14 areas in the Western United States where arsenic conditions in water exceed 50 µg/L because of known or suspected geothermal sources. In these areas, dissolved arsenic concentrations ranged from 80 to 15,000 µg/L. Welch *et al.,*found that mean dissolved arsenic concentrations in geothermal ground waters are higher than mean arsenic concentrations in non-thermal ground waters in any of the physiographic provinces in the United States. Flow of arsenic-enriched geothermal water from hot springs may result in high concentrations of arsenic in surface water systems. In Yellowstone National Park, the arsenic concentrations in geysers and hot springs range from 900 to 3,560 µg/L (Stauffer and Thompson, 1984). Waters from these sources cause elevated arsenic levels in the Madison and Missouri Rivers far downstream of the park boundaries (Nimick *et al.,* 1998). As a result, cities that use water from portions of the Missouri River for municipal supply must treat it to reduce arsenic concentrations. Geothermal sources of arsenic are primarily located in the Western United States.

Other Sources

Natural emissions of arsenic associated with volcanic activity and forest and grass fires are recognized to be significant. Indeed, volcanic activity appears to be the largest natural source of arsenic emissions to the atmosphere (ATSDR, 1998). Estimates of natural releases (of which volcanic arsenic emissions are the primary source) show significant range, and are summarized in Table 2-6.

Table 2-6
Estimated Natural Average Arsenic Releases to the Atmosphere

Study	Estimated annual natural releases (metric tons)*
Tamaki and Frankenburger, 1992	44,100
Pacyna et al., 1995	1,100 - 23,500
Loebenstein, 1994	2,800 - 8,000

While at least one study suggests that natural arsenic emissions slightly exceed industrial emissions (Tamaki and Frankenburger, 1992), other studies suggest that industrial emissions of arsenic are significantly greater than natural emissions (Pacyna *et al.,* 1995; Loebenstein 1994). Thus, the relative contributions of volcanic sources, other natural sources, and anthropogenic sources to the atmosphere have not been definitively established.

In summary, the primary natural sources of arsenic in the United States include arsenic in geological formations (rock as well as soil and sedimentary deposits) and arsenic associated with geothermal waters. In geological formations, higher arsenic concentrations tend to be associated with sedimentary rocks derived from acidic to intermediate volcanics, which occur primarily in parts of the Western United States, and from sulfide minerals and iron and manganese oxides associated with sedimentary rocks, which occur in the parts of the Northeastern and Upper Midwestern United States. Geothermal activity may affect arsenic levels in ground water and surface water in some regions of the United States, particularly in the Western United States. In addition to these recurrent sources, volcanos and forest fires may also result in the sporadic release of large amounts arsenic to the environment.

2.3 Anthropogenic Sources of Arsenic

From man-made sources, arsenic is released to terrestrial and aquatic environments and to the atmosphere. The anthropogenic impact on arsenic levels in these media depends on the level of human activity, the distance from the pollution sources, and the dispersion and fate of the arsenic that is released. This section discusses the major current and past anthropogenic sources of arsenic, which are wood preservatives, agricultural uses, industry, and mining and smelting. It also provides an overview of other sources of arsenic in the environment. Table 2-7 provides an overview of the use of arsenic in specific economic sectors. It is important to note that some of these uses are banned in the United States. After these sources are discussed, an overview of anthropogenic arsenic releases to the environment, based on Toxics Release Inventory (TRI) data, is provided.

Table 2-7
Summary of Current and Past Uses of Arsenic

Sector	Uses
Lumber	Wood preservatives
Agriculture	Pesticides, insecticides, defoliants, debarking agents, soil sterilant
Livestock	Feed additives, disease preventatives, animal dips, algaecides
Medicine	Antisyphilitic drugs, treatment of trypanosomiasis, amebiasis, sleeping sickness
Industry	Glassware, electrophotography, catalysts, pyrotechnics, antifouling paints, dye and soaps, ceramics, pharmaceutical substances, alloys (automotive solder and radiators), battery plates, solar cells, optoelectronic devices, semiconductor applications, light emitting diodes in digital watches

Source: Azcue and Nriagu, 1994.

Wood Preservatives

About 90% of the arsenic that is consumed in the United States on an annual basis is used to manufacture the wood preservative chromated copper arsenate (CCA) (Reese, 1999; Reese, 1998). CCA is an inorganic arsenic compound and consists of arsenic, chromium and copper. There are a number of CCA type products in the market , and the precise molecular composition may vary from formulation to formulation. Different arsenic compounds are used as active ingredients in CCA, including arsenic acid (H_3AsO_4), arsenic pentoxide (As_2O_5), and sodium arsenate (Na_2HAsO_4). These arsenical ingredients in CCA are primarily produced from arsenic trioxide (As_2O_3), although arsenic trioxide itself is not used as a wood preservative in CCA treatment solutions. EPA classifies CCA as a restricted use pesticide (USEPA, 1997a, USEPA, 1984). CCA is used to pressure treat lumber, which is typically used for construction of decks, fences, and other outdoor applications. As a result, CCA consumption is related to the performance of the housing industry. Currently, there are three primary manufacturers of CCA, and these firms are located in the Southeastern United States (Reese, 1998).

The current releases of arsenic to the environment from the manufacture and use of CCA are included in the Toxics Release Inventory, which is discussed below. Earlier releases of arsenic to the environment at some wood preserving sites have resulted in localized contamination of environmental media, and several wood preserving sites are included on the NPL list. In addition, there is some evidence that some arsenic may be released to soil from CCA treated lumber (USEPA, 1997a). EPA is evaluating the release of arsenic from treated wood. However, no data are currently available that indicates releases of arsenic from treated lumber could affect drinking water quality.

Agricultural Uses

Past and current agricultural uses of arsenic and arsenic compounds include pesticides, herbicides, insecticides, defoliants, and soil sterilants (Azcue and Nriagu, 1994). Arsenic compounds were also used in animal dips and are currently used in raising livestock as feed additives and for disease prevention (Azcue and Nriagu, 1994). These uses of arsenic compounds are discussed below.

Arsenic is a constituent of organic agricultural pesticides that are currently used in the United States. The most widely applied organoarsenical pesticide is monosodium methanearsonate (MSMA), which is used to control broadleaf weeds (Jordan et al., 1997). MSMA was the twenty-second most commonly applied conventional pesticide in the United States in 1995 (USEPA, 1997b). In 1995, approximately 4 to 8 million pounds of MSMA were applied. It is primarily applied to cotton; small amounts of disodium methanearsonate (DSMA) and cacodylic acid are also applied to cotton fields as herbicides. MSMA and DSMA have also been used for the postemergence control of crabgrass, Dallisgrass, and other weeds in turf (NAS, 1977), and cacodylic acid is used for weed control and monocotyledonous weeds. Organoarsenicals in soil are metabolized to alkylarsines and arsenate by soil bacteria (USEPA, 1998b). No published data are currently available on the leaching or transport of MSMA or DSMA from soil.

Inorganic arsenic was also a constituent of a variety of pesticides, herbicides, and fungicides. The last agricultural use of inorganic arsenic, which was the use of arsenic acid pesticide on cotton, was voluntarily canceled in 1993 (USEPA, 1998a, 58 FR 64579). Other inorganic arsenic-containing pesticides included sodium arsenate, calcium arsenate, copper acetoarsenite (Paris Green), copper arsenate, and magnesium arsenate. These were used to control potato beetles, boll weevils, grasshoppers, moths, bud worms, and other insects in the United States (Thompson, 1973). Inorganic arsenic compounds (arsenic acid, arsenic trioxide, and sodium arsenate) are currently only used in sealed ant bait and wood preservatives (USEPA, 1995). Formerly registered uses of inorganic arsenic under the Federal Insecticides, Fungicides, and Rodenticides Act (FIFRA) were pesticides, insecticides, rodenticides, cotton dessicants, anti-fouling paints for boat hulls, soil sterilants, and herbicides for crab grass and other weeds[4]. These uses were voluntarily canceled by the product manufacturers by 1993. Sodium arsenite was used as cattle and sheep dips (Azcue and Nriagu, 1994), but its is now banned in the United States (USEPA, 1999b). Other inorganic arsenic pesticides that are banned in the United States include calcium arsenate, copper arsenate, and lead arsenate. The use of arsenic trioxide and sodium arsenate is severely restricted.

Arsenic-based inorganic pesticides were applied to various agricultural crop lands, and the historical application of these pesticides has contaminated soil with arsenic residues. The use of lead arsenate as an insecticide for worms and moths resulted in contamination of soils in apple orchards (Davenport and Peryea 1991; Maclean and Langille, 1981; Hess and Blanchar, 1977).

---

[4] Based on queries of the EPA Office of Pesticide Programs Pesticide Product Information System databases conducted by ISSI Consulting Group in July, 1999. This database is searchable online at http://www.cdpr.ca.gov/docs/epa/epamenu.htm.

Steevens, *et al.,* (1972) showed that the use of sodium arsenate as a potato defoliant raised total arsenic concentrations in Wisconsin potato field soils. The leaching of arsenic from agricultural topsoil to subsoil has been reported (Peryea, 1991; NAS, 1977). Leaching is more likely to occur in sandy soils than in organic soils, and sodium or calcium arsenates tend to leach faster than aluminum or iron arsenates (NAS, 1977). Fertilization of fields with phosphates may increase the rate of leaching of arsenic from soils (Davenport and Peryea, 1991; Woolson *et al.,* 1973), and can increase arsenic concentrations in the subsoil and shallow ground water (Peryea and Kammereck, 1997; Davenport and Peryea, 1991; Peryea, 1991). This historical pesticide use is not expected to have resulted in widespread ground water contamination. Peters *et al.,* (1999) found that arsenic levels in drinking water did not correlate with agricultural activities in the State of New Hampshire. Fuhrer et al., (1996), however, reported that filtered water arsenic concentrations in the Yakima River Basin were higher in agriculturally affected areas than in other areas. These authors suggested high arsenic concentrations in orchard soils from past lead-arsenate applications were the likely source of the arsenic in surface water.

Organic arsenic (e.g., roxarsone and arsanilic acid) is a constituent of feed additives for poultry and swine for increased rate of weight gain, improved feed efficiencies, improved pigmentation, and disease treatment and prevention (e.g., swine dysentary, chronic respiratory disease, specific infections) (21 CFR 558.530; 21 CFR 558.62). These additives undergo little or no degradation before excretion (NAS, 1977; Moody and Williams, 1964; Aschbacher and Feil, 1991). Arsenic concentrations in animal wastes are reported to range from 4 to 40 mg/kg (Isaac *et al.,* 1978). Application of arsenic containing wastes as fertilizer to field plots is reported to elevate arsenic concentrations in soil water. Two other studies, however, reported that the fertilization of crops with poultry litter did not change the arsenic content of soils or crops (Smith *et al.,* 1992; Morrison, 1975). Therefore, potential effect of arsenic from animal wastes upon ground water and surface water is not well understood.

Industrial Uses and Releases

Arsenic and arsenic compounds are used in a variety of industrial applications. Arsenic metal is used in the production of posts and grids for lead-acid storage batteries, and is used in the formulation of some copper alloys (Reese, 1998). As maintenance-free automotive batteries, which contain little or no arsenic, replace lead-acid storage batteries, the demand for arsenic metal is expected to decrease. An arsenic containing compound, crystalline gallium arsenide, is produced from very pure arsenic metal. Crystalline gallium arsenide is a semiconducting material used in computers, optoelectronic devices and circuits, and other electronic applications.

Industrial processes including the burning of fossil fuels, combustion of wastes (hazardous and non-hazardous), mining and smelting (discussed separately below), pulp and paper production, glass manufacturing, and cement manufacturing can result in emissions of arsenic to the environment (USEPA, 1998b). Coal-burning power plants may emit aerosols and fly ash that contain arsenic (Yan-Chu, 1994). These arsenic-containing particles may be deposited from the atmosphere onto soil and surface waters. The amount of arsenic that is emitted by power plants and fossil fuel combustion is not included in current Toxics Release

Inventory data.[5] However, utility emissions of arsenic and several other metals will be added to the TRI in 1999 (USEPA, 1999a).

Past waste disposal practices have impacted arsenic concentrations in ground water and surface water at waste disposal sites. Arsenic has been identified as a contaminant of concern at 916 of the 1,467 Superfund National Priority List (NPL) hazardous waste sites (ATSDR, 1998), and arsenic is listed as the highest priority contaminant on the ATSDR/EPA list of the hazardous substances at NPL Sites (ATSDR, 1997). There is a potential for releases of arsenic from waste sites to affect ground water or surface water in the vicinity of the waste sites.

Mining and Smelting

Arsenic can be obtained from two of its ores, arsenopyrite and lollingite, by smelting in the presence of air around 650–700 °C (Kirk-Othmer, 1992), or arsenic trioxide (As_2O_3) in flue dust from the extraction of lead and copper can be captured (Ferguson, 1990). Subsequently, arsenic trioxide can be used to produce other arsenic compounds or purified to elemental arsenic.

Arsenic trioxide was produced for commercial use in the United State at the ASARCO smelter in Tacoma, Washington, until 1985, at which time the smelter ceased operations (ATSDR, 1998). The USEPA Office of Air Quality Planning and Standards indicates that primary and secondary[6] lead smelters, primary copper smelters, and secondary aluminum operations are potential sources of arsenic (USEPA, 1998b). There are presently three active primary lead smelters, two of which are located in Missouri and one is in Montana. In addition, there are 19 active secondary lead smelters located throughout the United States. The seven primary copper smelters are located in Arizona (3), New Mexico (2), Texas (1), and Utah (1). Secondary aluminum operations locations were not identified in USEPA 1998b.

Arsenic may be emitted to the atmosphere from metals smelters, and deposited with precipitation downwind of the smelter. In Washington, Crecelius (1975) found higher concentrations of arsenic in rain and snow downwind of a smelter (17 µg/L) than in rain in unpolluted areas (<1 µg/L). Such atmospheric deposition could affect arsenic concentrations in soil and other environmental media. Miesch and Huffman (1972) detected increased concentrations in soil downwind of a smelter in Helena, Montana.

High concentrations of arsenic may occur in areas that are near or affected by current or historical mining activities. Sulfide-bearing rocks are often mined for gold, lead, zinc, and copper, and arsenic is frequently found as an impurity in the sulfide ores of these metals. The drainage from abandoned mines and mine wastes is typically acidic, and dissolved arsenic concentrations can be as high as 48,000 µg/L in mine drainage (Welch *et al.*, 1988). In mining areas, the arsenopyrite (FeAsS) that occurs in association with ores and arsenic bearing pyrite is a

[5] The Toxic Release Inventory is a database of toxic releases in the United States compiled annually from SARA Title III Section 313 reports.

[6] Primary smelters produce metals from raw ores, whereas secondary smelters reclaim metals from used and recycled materials, such as scrap metal and used batteries.

common source of dissolved arsenic. The mineral orpiment, realgar, and arsenic-rich iron oxides are other sources of dissolved arsenic (Welch *et al.*, 1988) at mining sites.

Other Uses and Sources

There is some evidence that some volatile organic compounds (VOCs) in ground water may facilitate the release of arsenic from aquifer materials to ground water. Ground water that is affected by VOCs like petroleum products and other landfill wastes may be sufficiently reduced to result in elevated dissolved iron-oxide concentrations. Under these reducing conditions, aquifer materials may be a source of dissolved arsenic in ground water (Ogden, 1990).

From the Civil War until approximately 1910, arsenic was used as an embalming fluid, and elevated concentrations of arsenic were detected in ground water at cemeteries in Iowa and New York (Konefes and McGee, 1996). Therefore, it appears that cemeteries may be sources of localized arsenic contamination in ground water. However, the extent of or potential for arsenic contamination associated with ground water in cemeteries has not been broadly evaluated.

Historically, inorganic and organic arsenic compounds were used as therapeutic agents. The first recognized use of inorganic arsenic as a therapeutic agent was in 1786; Fowler's solution, which contained approximately 1 percent arsenic trioxide, was a common arsenic containing medicinal (NRC, 1999). Inorganic arsenic was used to treat symptoms of skin diseases such as eczema and psoriasis, malarial and rheumatic fevers, asthma, pernicious anemia, leukemia, Hodgkin's disease, and pain. Because of concern over inorganic arsenic's toxic effects, therapeutic use of inorganic arsenic ceased in the 1970s. Organic arsenicals were used for the treatment of spirochetal and protozoal diseases through the first half of the twentieth century. Salvarsan (arspheramine) was a common anti-syphilitic from 1907 until its use was supplanted by penicillin in the 1940s and 1950s. Organic arsenic compounds were used for the treatment of amebiasis and trypanosomaisis. Melarsoprol (organic arsenic compound) continues to be used for treatment of trypanosomaisis. In addition, arsenic trioxide is being researched as therapy for acute promyelocytic leukemia.

Releases to the Environment Reported in the TRI 1997

The Toxics Release Inventory (TRI) is a national database that identifies facilities, and chemicals manufactured and used at identified industrial facilities. The TRI database also identifies the amounts of these chemicals that are released to the environment annually as a result of routine operations, accidents, and other one-time events, and the amounts of chemicals that are managed in on- and off-site waste for each year since 1987. Annually, certain facilities must report basic information about their facilities and operations, and about the amounts of listed toxic chemicals used, released, recycled, or otherwise managed at the facility. Facilities must report on arsenic and arsenic compound use, release, management, and disposal. Therefore, the TRI data provides a significant amount of information regarding industrial releases of arsenic to the environment, and trends in those releases since 1987. Information on arsenic releases to the environment, based on recent TRI data, is summarized below.

TRI facilities reported managing a total of 14,898,807 pounds of arsenic and arsenic compound containing wastes that were either production or non-production related in 1997 (USEPA, 1999a). Total on- and off-site releases in 1997 were reported to be 7,947,012 lbs. The majority of these releases (6,046,473 lbs) were on-site, and 95 percent of on-site releases were to land (5,766,252 lbs). The majority of the arsenic was released to the land via surface impoundments or non-Resource Conservation and Recovery Act (RCRA) Part C landfills; less than 1 percent of the arsenic and arsenic compounds waste is disposed of in RCRA Subtitle C landfills. After releases to land, releases to air are the second largest component of on-site releases, with a total of 199,918 lbs of arsenic and arsenic compounds released from stacks, point sources, and fugitive or non-point sources. In addition, 76,170 lbs were reported to have been injected underground and 4,133 lbs were reported to have been released to surface water in 1997. Of the 1,900,539 lbs that were disposed of off-site, most (1,460,728 lbs) were disposed of to landfills and surface impoundments, and significant amounts were released by underground injection (209,716 lbs) or were solidified or otherwise stabilized (149,416 lbs). Small amounts were stored or transferred to wastewater treatment plants.

Data contained in the TRI indicate that releases of arsenic and arsenic compounds from TRI reporting facilities to the environment have increased in recent years. From 1995 to 1997, TRI data indicates that total on-site and off-site releases of arsenic have risen from 3,536,467 lbs to 7,947,012 lbs. The increase primarily occurred at one facility, where arsenic on-site land releases increased by 3.58 million pounds from 1995 to 1997 because of a change in the facilities smelting process that was implemented to reduce sulfur dioxide emissions. However, from 1995 to 1997 the quantity of arsenic in air emissions, underground injection, releases to land, and transfers to off-site disposal have all risen, while only the quantity of arsenic discharged to surface water is reported to have decreased (USEPA, 1999a). Total releases of arsenic to the environment in 1997 also exceed those for the base year (1988) for TRI data, when total on-site and off-site emissions of arsenic were 6,911,043 lbs.

Although the TRI data include estimated emissions from many sources of anthropogenic sources of arsenic to the environment, the data do not include several potentially significant sources of arsenic emissions, and therefore the data it contains should be interpreted with some care. For example, TRI release data do not include arsenic in organoarsenical herbicides that are applied to cotton fields. Several other potentially significant sources of arsenic emissions will begin reporting to the TRI in 1999 for the year 1998, including coal and oil burning electrical utilities, coal mining, and metals mining (USEPA, 1999). The addition of these industrial sectors should improve the accuracy of reporting of emissions of arsenic and arsenic compounds.

3. Fate and Transport of Arsenic

3.1 Relationship of Fate and Transport Properties to Source Intake

Arsenic concentration in fresh waters shows considerable variation with geological composition of the drainage area and the level of anthropogenic input. The fate and transport of arsenic in ground water and surface water are discussed separately below:

Ground water

Arsenic may be released to ground water in a variety of ways, which include, but are not limited to, weathering of earth's crust and soil materials, discharge from industrial processes, and overland runoff from agricultural and urban areas. In water, arsenic can undergo a series of transformations, including oxidation-reduction reactions, ligand exchange, and biotransformations (ATSDR, 1998; Welch *et al.,* 1988). Several factors have been identified which effect the fate and transport processes in ground water. These include the oxidation state of the arsenic, oxidation-reduction potential (Eh), pH, iron concentrations, metal sulfide and sulfide concentrations, temperature, salinity, and distribution and composition of the biota (ATSDR, 1998; Roberston, 1989; Welch *et al.,* 1988). The predominant form of arsenic is usually arsenate (As+5), although arsenite (As+3) may be present under some conditions (Irgolic, 1994; Welch *et al.,* 1988). However, the NRC (1999) noted that arsenite might be more prevalent than anticipated.

Surface water

The processes that affect arsenic fate and transport in surface water are analogous to those that are operative in ground water systems. Thus, the factors that affect arsenic transformations and transport include the oxidation state of the arsenic, oxidation-reduction potential (Eh), pH, iron concentrations, metal sulfide and sulfide concentrations, temperature, salinity, and distribution and composition of the biota (ATSDR, 1998). However, there are additional factors that affect arsenic fate and transport in surface water systems. These include total suspended sediment (Nimick *et al.,* 1998; Waslenchuk, 1979), seasonal water flow volumes and rates (Nimick *et al.,* 1998; Waslenchuk 1979), and time of day (Nimick *et al.,* 1998).

Sorption of arsenic to suspended sediment may strongly affect the fate and transport of arsenic in surface water systems. Where pH and arsenic concentrations are relatively high, and total suspended sediment levels are relatively low, sorption processes may be less important (Nimick *et al.,* 1998). However, where suspended sediment loads are higher, arsenic concentrations are lower, and pH levels are lower, arsenic is more likely to be present in the suspended particulate phase rather than the dissolved phase. Particulate phase arsenic may settle to bottom sediment in reservoirs and areas with low flow levels. The sorption of arsenic onto suspended sediment is a mechanism for the removal of dissolved arsenic from surface water. Sorption is greater when the amount of suspended sediment is greater. In surface water, lakes may interrupt the downstream transport of particulate sorbed arsenic. In deeper lakes, remobilization of arsenic from the sediment may be minimal, whereas in shallower lakes, arsenic

may be remobilized faster from wind induced wave action and high-flow scouring. Large and deep reservoirs are more likely to be long-term sinks for arsenic.

Seasonal variations of arsenic concentrations have been observed in surface water systems, and these variations appear to be related to the source of the arsenic and the flow of the river. In the Madison River, where relatively constant inputs of arsenic originate from geothermal sources, arsenic concentrations at one point ranged from 110 to 370 µg/L in samples collected between 1986 and 1995 (Nimick *et al.*, 1998). The highest concentrations occurred during periods of low flow, and the lowest concentrations occurred during periods of high flow. Waslenchuk (1979) measured arsenic concentrations in rivers in the Southeastern United States, where average arsenic levels are far lower than in the Madison River. In the rivers that Waslenchuk studied, seasonal variations of as great as 0.2 µg/L were observed around average concentrations that ranged from 0.15 to 0.45 µg/L. These variations were related to seasonal precipitation levels, which flushed higher concentrations of arsenic into the rivers in the spring.

Arsenic concentrations in surface water may also change during one day, as a result of changes in water pH that are attributable to incoming solar radiation and photosynthesis (Nimick *et al.*, 1998). These consistent daily changes are called diurnal variability. Because of photosynthesis, the water pH tends to increase later in the day, and dissolved arsenic concentrations also tend to increase. Diurnal variations in dissolved arsenic concentrations of as much as 21-percent were observed at three of the five sites on the Madison and Missouri Rivers, but diurnal variability was not seen at the other two sites. Thus, diurnal variability may affect arsenic concentrations in some surface water sources.

3.2 Relationship of Fate and Transport Properties to Treatment and Distribution

A variety of processes and factors affect the fate and transport of arsenic within public water supply treatment systems. The most important factor appears to be the oxidation state (arsenate or arsenite). The presence of competing ions, especially sulfate and fluoride, are also important, as is pH. This section discusses the effects of these factors upon arsenic concentrations and removal from water in treatment systems.

Arsenic in water commonly occurs as arsenate, As (V), or arsenite, As (III). The chemical species that are formed depends upon the oxidation-reduction conditions and the pH of the source water. The common soluble species of arsenate are H_3AsO_4, $H_2AsO_4^{1-}$, $HAsO_4^{2-}$, and AsO_4^{3-}; whereas the common soluble species of arsenite are H_3AsO_3 and $H_2AsO_3^{1-}$. At typical pHs, the predominant arsenite form is the neutral species (H_3AsO_3), while the predominant arsenate species are the anions $H_2AsO_4^{1-}$ and $HAsO_4^{2-}$. Because of its ionic charge, arsenate is more easily removed from source waters than arsenite. In particular, activated alumina, ion exchange, and reverse osmosis may achieve relatively high arsenate removal rates, but they show lower treatment efficiencies for arsenite.

Arsenite can be oxidized to arsenate, and this can improve arsenic removal efficiencies. In water that contains no ammonia or total organic carbon (TOC), chlorine rapidly (in less than 5

seconds at chlorine concentrations of 1.0 mg/L) oxidizes approximately 95 percent of arsenite to arsenate (Clifford, 1986). The reaction through which chlorine oxidizes arsenite is:

$$As^{3+} + Cl_2 \rightleftharpoons As^{5+} + 2Cl^- \qquad [2]$$

The presence of ammonia and TOC slows this oxidation process. Monochloramine at the concentration of 1.0 mg/L oxidized 45 percent of arsenite to arsenate. Oxygen may slowly oxidize arsenite to arsenate, although this reaction occurs slowly in laboratory studies. Shen (1973) indicated that potassium permanganate can oxidize arsenite. Potassium permanganate oxidizes arsenite according to the following reaction:

$$3As^{3+} + 2KMnO_4 \rightleftharpoons 2MnO_4 + 3As^{5+} \qquad [3]$$

Therefore, it appears that chlorine and potassium permanganate are the most effective processes to oxidize arsenite to arsenate. The oxidation of arsenic from its trivalent state to its pentavalent state can allow treatment plants to increase the removal efficiencies of treatment technologies.

The water pH also affects the removal efficiencies of treatment technologies for arsenic, and therefore, the level of and persistence of arsenic it drinking water. For example, activated alumina removes arsenic most efficiently at pH 6, but yields lower treatment efficiencies at pH 9 (Rubel and Hathaway, 1987). Removal efficiencies for alum coagulation tend to decrease at pHs greater than 7.0 (G8ulledge and O'Connor, 1973).

Competition for adsorption sites with other ions may also affect the persistence and removal of arsenic from source water in treatment plants. In particular, sulfate in source waters may reduce the efficiency of arsenic removal. Sorg (1990) showed that waters with high sulfate levels correlate with lower arsenic removal by ion exchange technologies. Anion exchange resins preferentially adsorb sulfate over arsenic. Because of this preference for sulfate over arsenic, under some conditions displacement of arsenic can result in peaks of arsenic in effluent which exceed the concentrations of arsenic in source water influent. Jekel (1994) reported that the competitive effects of sulfate, fluoride, and phosphate may reduce the effectiveness of activated alumina treatments, particularly when the competitors are in the range of 0.1 to 2 mg/kg, and arsenic removal to the ppb level is required. If the competing ions are present in small concentrations, activated alumina can be applied successfully at slightly acidic pH ranges (pH 5.5 to 6.0).

In summary, three factors are particularly important to the fate of arsenic in treatment and distribution systems. These include the oxidation state, the pH of the source water, and the presence of other ions which may compete for adsorption sites in treatment technologies. The most significant factor that affects the fate and transport of arsenic in treatment and distribution systems appears to be the arsenic oxidation state. Arsenate is removed more efficiently than arsenite.

This page intentionally left blank.

4. Sources of Data on Arsenic Occurrence in Drinking Water Supplies

The potential health concerns associated with arsenic in drinking water have long been recognized, and therefore a significant amount of data is available on source water and treated drinking water arsenic concentrations. These data sources, which include national, regional, and State databases, differ substantially in size, content, and quality. Table 4-1 lists the national data sources which were available for development of arsenic occurrence estimates, and provides general information about the characteristics of these data sources.

An important source of information for estimating arsenic occurrence on a national basis is compliance monitoring data collected in accordance with the Safe Drinking Water Act (SDWA). The estimates of arsenic occurrence and intra-system variability presented in Chapters 6 and 7 of this report were developed using compliance monitoring data submitted by several states. In this Chapter, Section 4.1 discusses the development of this database from sets of compliance monitoring data submitted to the USEPA, and from information contained in the Safe Drinking Water Information System (SDWIS). In this report, this database is referred to as the Arsenic Occurrence and Exposure Database (AOED).

Other arsenic occurrence surveys also provide potentially important sources of arsenic occurrence information. Recently developed databases include the National Arsenic Occurrence Survey (NAOS) the United States Geological Survey Arsenic Database (USGS), and the Metropolitan Water District of Southern California database (Metro). These databases are described in Section 4.2, and they are used in Chapter 6 to provide comparisons with the occurrence projections developed using AOED. Another database that could be used as a comparison tool is the National Inorganics and Radionuclides Survey (NIRS). This database is also described in Section 4.2.

In addition to the compliance monitoring data and the databases that have been used as comparison tools in this report, a variety of other data sources are available that provide arsenic occurrence information. However, these data sources were not used in this report for various reasons. These databases, which include the Rural Water Survey (RWS), the 1969 and 1978 Community Water Supply Surveys (CWSS), the National Organics Monitoring Survey (NOMS), the occurrence data gathered by the Western Coalition of Arid States (WESCAS) and the Association of California Water Agencies (ACWA) database, are briefly described in Section 4.3, and the reasons why these databases were deemed unsuitable for use in this occurrence estimation are presented.

4.1 Arsenic Occurrence and Exposure Database (AOED)

The Arsenic Occurrence and Exposure Database was developed to support the USEPA's effort to estimate arsenic occurrence in the United States. The database includes information from SDWIS and from State compliance monitoring data sets that were provided to ISSI. Section 4.1.1 describes SDWIS. Section 4.1.2 describes the State compliance monitoring

Table 4-1
Summary of National Arsenic Data Sources

Database	Number of Systems	Media (GW or SW)	Reporting Limit (µg/L)	Source Data	Drawbacks
SDWIS	~ 55,000	GW and SW	50	--	Reports only MCL violations
State Compliance Databases	24,247	GW and SW	0.001 - 10	✓	Not all States are covered
NIRS	982	GW	5	--	95% Censored
NAOS	< 517	GW and SW	0.5	--	Untreated and predicted drinking water arsenic concentrations
Metro	112	GW and SW	0.5	--	Large systems
USGS	~ 20,000	GW	1	--	Untreated and non-public water supply water
1969 CWSS	969	GW and SW	5	--	Pre 1980 data
1978 CWSS	< 350	GW and SW	2.5	--	Pre 1980 data
RWS	92	GW and SW	2	--	Pre 1980 data
NOMS	113	GW and SW	??	--	Pre 1980 data

Notes:
GW – Ground water
SW – Surface water

databases. Section 4.1.3 explains how SDWIS and State compliance monitoring data sets were used to build the AOED.

4.1.1 Safe Drinking Water Information System (SDWIS)

SDWIS provides a complete inventory of information on all public and private water supply systems in the United States. The SDWIS database is used for compliance tracking and to allocate SDWA grant monies. Violations of the current arsenic MCL of 50 µg/L are recorded in the SDWIS database, but measured concentrations below the MCL are not recorded. While this information is useful for determining the number of systems that have violated current Federal arsenic drinking water limits, it is not suitable for estimating arsenic concentrations in the range of 2 to 50 µg/L.

SDWIS is a reliable source of information on the characteristics of individual public water systems (PWS) in the United States. Since the set of PWS in the United States changes over time, the SDWIS information is checked for accuracy and updated on an annual basis. Near the end of the calendar year, the SDWIS inventory is "frozen", such that no more additions are made to the database that year. This inventory is then distributed to State drinking water programs for verification of the numbers and types of systems (Science Applications International Corporation (SAIC), 1999). The core verified data in the SDWIS inventory include:

- System name and address;
- Federal identification number (SDWIS ID number or PWSID);
- Source water type;
- Ownership category;
- Population served; and
- Regulatory classification (system type).

In regard to the category of source water type, a system may include more than one type of source. Small systems typically have only one source, but larger systems may have multiple sources. Although most large systems are served by either a ground water or a surface water source, some systems do receive water from a mix of ground water and surface water sources. In SDWIS, any water system with a continuous source of surface water is defined as a surface water system, even if 75 percent, or 99 percent, of the surface water is from a ground water source. Therefore, systems that rely entirely on surface water sources and blended water systems are coded as surface water systems in SDWIS, and systems that rely entirely on ground water sources are coded as ground water systems in SDWIS. SDWIS also includes the source water type category "ground water under the influence of surface water." The regulatory definition essentially defines this category as systems where, for various reasons, the ground water is expected to be contaminated similarly to nearby surface water sources. For the occurrence analyses described in this report, ground water systems under the influence of surface water were included in the surface water category.

SDWIS characterizes the population served on the basis of retail customers only. The SDWIS inventory populations do not include the number of people served with water that is wholesaled by any individual utility (SAIC, 1999).

Ownership categories are limited to public and private, while regulatory classification includes community water supplies (CWS), non-transient, non-community water supply systems (NTNCWS) and transient non-community water supplies (TNCWS). The arsenic occurrence estimates presented in this report are for CWS and NTNCWS.

In summary, the SDWIS database is a very useful source of information on the characteristics of individual systems. In fact, as discussed in Section 4.1.3, SDWIS was a key resource for information on the characteristics of individual systems for the development of AOED. However, it does not contain information on the levels of arsenic in individual systems that can be used to estimate arsenic occurrence within the range of interest.

4.1.2 State Compliance Monitoring Databases

An important source of data for estimating arsenic levels in drinking water is compliance monitoring data that is collected from the drinking water utilities in accordance with the SDWA, to comply with the current arsenic MCL of 50 ug/L. The compliance monitoring data sets were submitted voluntarily to the USEPA by State drinking water agencies, either directly or through other organizations (e.g., AWWA, EPA Office of Research and Development, EPA Regional Office, Association of Public Health Laboratories), during this study and during earlier studies. Such data was available from 32 States, although data from only 25 States was found to be suitable for inclusion in this occurrence study. Table 4-2 presents an overview of the suitable compliance monitoring data for each State. Figure 4-1 shows the geographical distribution represented in the 25 State's compliance monitoring databases. The characteristics of the 25 suitable individual state databases are documented in Appendix D-1. While the States for which compliance monitoring data are available are distributed throughout the United States, this figure shows that the States are not evenly distributed. In particular, few data sets are available for States in the New England, Mid-Atlantic, and Southeastern United States. In contrast, the Midwestern, North Central, South Central, and Western Regions appear to be fairly well represented. The States in these regions are described on Section 5.5; see Figure 5-2 for a map of the United States showing the seven regions.

A number of data sets were considered for inclusion in the AOED database, but were not included for various reasons. Table 4-3 lists the States for which at least one data set was excluded from the AOED database, and the specific reason why the data set was excluded. Data sets from 19 States were excluded from the database. However, 12 of these States submitted multiple data sets; the single most representative data set from each of these States was chosen and included in AOED. In most cases, these States submitted multiple compliance monitoring data sets, and the most recent data sets represented the largest number of systems. The later compliance monitoring data sets also tended to have the lowest detection levels, although this was not true in all cases. In one case, Maine, the older data set was included because the newer data set only reported arsenic levels above 20 μg/L. Minnesota had two types of data set, one

Table 4-2
Overview of State Compliance Monitoring Data

State	Number of CWS GW	Number of CWS SW	Number of NTNCWS GW	Number of NTNCWS SW	Reporting Limits	Mode of Reporting Limits	Date Ranges	Intra-system Data
Alaska	326	109	140	26	1 - 5	1	1991 - 1997	No
Alabama	263	68	31	5	1	1	1985 - 2000	Yes
Arkansas	371	76	0	0	5	5	1996 - 1998	Yes
Arizona	668	46	190	4	0.4 - 10	10 (GW) and 5 (SW)	1988 - 1998	No
California	1369	222	376	11	0.001 - 10	10	1981 - 2000	Yes
Illinois	1082	103	0	0	0.005 - 10	2 (GW) and 1 (SW)	1993 - 2000	Yes
Indiana	648	51	538	4	0.5 - 8	1	1996 - 1999	Yes
Kansas	506	101	62	2	1	1	1992 - 1997	No
Kentucky	88	150	0	0	1 - 5	1	Not reported	No
Maine	109	29	0	0	1	1	1991 - 1994	No
Michigan	644	33	230	0	0.3 - 2	0.3 (GW) and 2 (SW)	1993 - 1997	No
Minnesota	863	23	634	6	1 - 5	1	1992 - 1997	No
Missouri	773	89	190	4	1	1	1995 - 1997	No
Montana	484	47	2	0	1	1	1980 - 1992	No
North Carolina	1735	169	562	8	0.2 - 10	10 (GW) and 5 (SW)	1980 - 2000	Yes
North Dakota	197	19	20	6	0.2 - 1	1	1993 - 1995	No
New Hampshire	504	37	0	0	5	5	1990 - 1994	No
New Jersey	438	29	758	3	0.1 - 10	5 (GW) and 2 (SW)	1993 - 1997	No
New Mexico	573	29	140	8	0.3 - 10	5	1983 - 2000	Yes
Nevada	221	31	0	0	3	3	1991 - 1997	No
Ohio	875	139	0	0	1 - 10	10	1981 - 1994	No
Oklahoma	446	210	0	0	2	2	1995 - 1998	Yes
Oregon	583	134	112	9	0.5 - 10	5	1990 - 1998	No
Texas	3105	326	580	27	1 - 10	2	1994 - 1999	Yes
Utah	327	38	50	3	0.1 - 10	5	1980 - 1999	Yes

Figure 4-1: States with Suitable Arsenic Compliance Monitoring Data

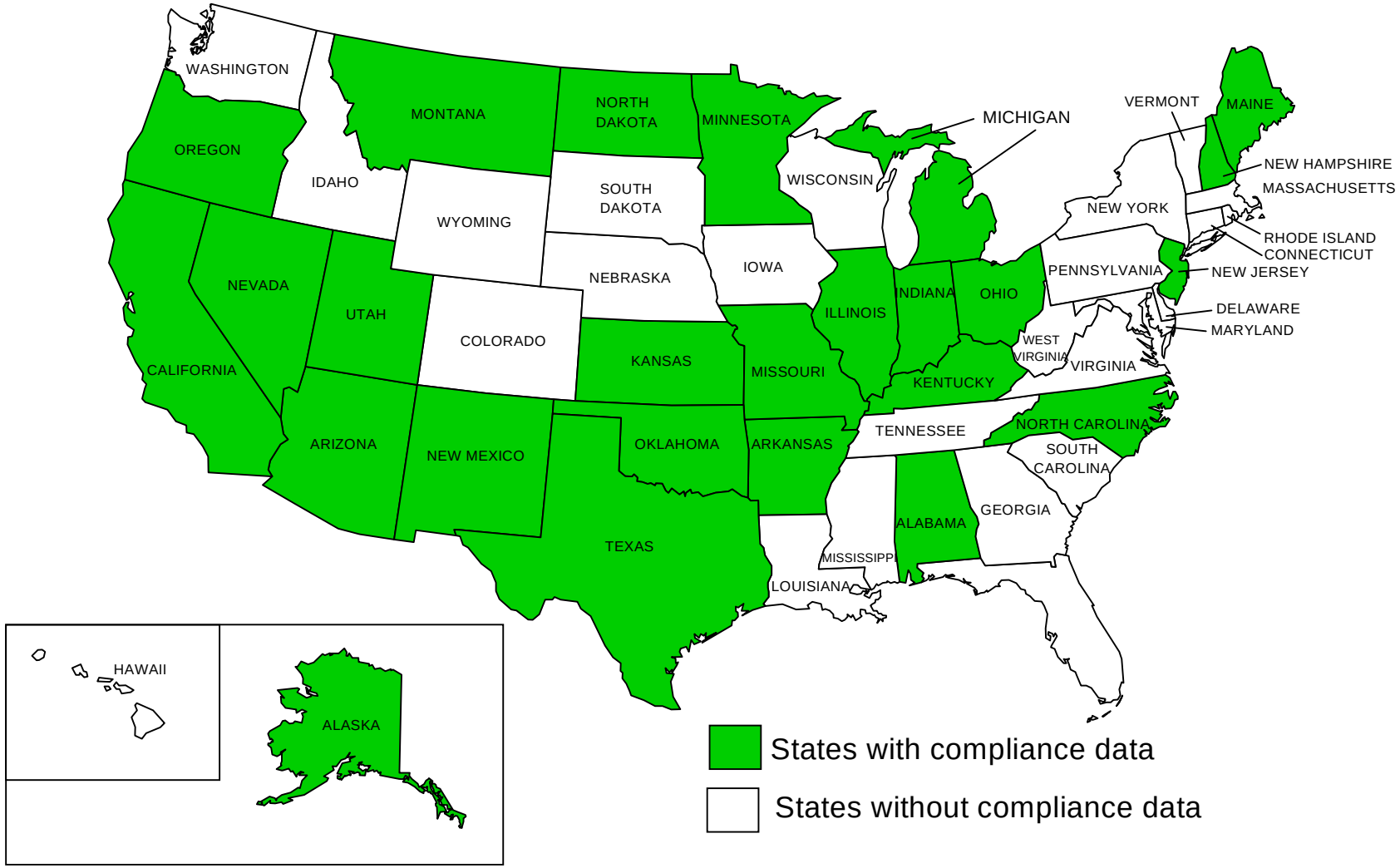


Table 4-3
State Data Sets Excluded from AOED

State	Reason Data Set Excluded
Alaska	Multiple data sets submitted; only the most recent data set selected.
Arizona	Multiple data sets submitted; only the most recent data set selected.
California	Multiple data sets submitted; only the most recent data set selected.
Florida	Data set unsuitable; all results censored, no detection limit reported.
Idaho	Data set unsuitable; no PWS ID numbers were provided.
Illinois	Multiple data sets submitted; only the most recent data set selected.
Iowa	Data set unsuitable; no reporting limits; results rounded to nearest 10 µg/L.
Louisiana	Data set unsuitable; no reporting limits; only detected values provided, results rounded to nearest 10 µg/L.
Maine	Selected older data set; recent data set censored at 20 µg/L.
Michigan	Multiple data sets submitted; only the most recent data set selected.
Minnesota	Selected data set associated with PWSID numbers.
New Mexico	Multiple data sets submitted; only the most recent data set selected.
Oregon	Multiple data sets submitted; only the most recent data set selected.
Pennsylvania	Data set unsuitable; all results censored, reporting limit of 50 µg/L.
North Dakota	Multiple data sets submitted; only the most recent data set selected.
South Dakota	Data set unsuitable; no result code flag was provided.
Texas	Multiple data sets submitted; only the most recent data set selected.
Utah	Multiple data sets submitted; only the most recent data set selected.
West Virginia	Data set unsuitable. Not made available electronically. Most results censored

containing compliance monitoring data, and one containing data that was not from PWS systems (data from private domestic wells or non-public water supply wells). In this case, only the most recently submitted compliance monitoring data was included. For seven States, the available data sets were unsuitable and were not included in the State compliance monitoring database. Only one set of data was available for each of these States, thus, these States are not represented in the compliance monitoring database. The reasons for exclusion of these data sets are listed in Table 4-3. ISSI contacted representatives of the State agencies that provided these data sets, but was unable to resolve the problems with these data sets, or obtain newer, suitable data sets.

The compliance monitoring data sets were submitted voluntarily to the USEPA by State drinking water agencies, either directly or through other organizations (e.g., AWWA, EPA Office of Research and Development, EPA Regional Office, Association of Public Health Laboratories), during this study and during earlier studies. These data sets generally included the following information:

- System name and address;
- PWS ID number;
- Sample collection date;
- Result; and
- Detection limit, if arsenic was not detected in the sample.

Because these data were collected for compliance purposes, the samples were assumed to have been collected from the points-of-entry (POE) into the distribution system, which is representative of each well or source after treatment. Thus, the arsenic values represent finished water, and should be representative of the arsenic levels to which consumers are exposed.[7] Compliance with the arsenic standard is measured at the POE. Therefore, these data are directly relevant to the estimation of regulatory costs and benefits for the RIA.

Representation of Systems in Each State

As shown in Table 4-2, the characteristics of the data differ from State to State. Tables 4-4a and 4-4b present the numbers of CWS and NTNCWS systems contained in the State data sets as a percentage of the total numbers of non-purchased CWS and NTNCWS systems in the State based on SDWIS. Purchased water systems were excluded from the occurrence data base used for the analyses. (The SDWIS total numbers of purchased and non-purchased systems in each state were used in the development of regional and national occurrence estimates, as described in Chapter 6.) For purchased water systems, EPA allows the State to decide how each system will conduct monitoring. In some cases, monitoring may be done by only the wholesaling system. The SDWIS database, however, only contains data on retail populations and does not indicate to which systems a wholesaler provides its water. As a result, we do not know to what extent purchased water systems are representative of other water systems in the State's database. Using the purchased water system data to develop the State probability estimates could bias the estimates by double counting individual system results. Developing results based on non-purchased water system results ensures that independent data form the basis for the State estimates.

Tables 4-4a and 4-4b show that both ground and surface water systems are well represented in the compliance monitoring databases, particularly for CWS systems. Overall, the compliance monitoring data represent 82.8 percent of the CWS ground water systems in each State and 82.4 percent of the CWS surface water systems in each State. For NTNCWS, the compliance monitoring data represent 57.6 percent of the ground water systems in each State and 63.0 percent of the surface water systems in each State. For 13 States, at least 90 percent of the

[7] See convention number 3 in Section 4.1.3 below for details about how untreated samples and samples before treatment were used in the occurrence analyses.

Table 4-4a
CWS Systems in SDWIS and State Compliance Monitoring Data

State	Source Type	Systems in SDWIS[a]	Systems in State Data Set[a]	Percent Coverage
Alaska	GW	337	326	96.7
Alaska	SW	111	109	98.2
Alabama	GW	270	263	97.4
Alabama	SW	69	68	98.6
Arkansas	GW	373	371	99.5
Arkansas	SW	78	76	97.4
Arizona	GW	721	668	92.6
Arizona	SW	46	46	100.0
California	GW	2652	1369	51.6
California	SW	522	222	42.5
Illinois	GW	1129	1082	95.8
Illinois	SW	108	103	95.4
Indiana	GW	794	648	81.6
Indiana	SW	54	51	94.4
Kansas	GW	516	506	98.1
Kansas	SW	101	101	100.0
Kentucky	GW	101	88	87.1
Kentucky	SW	173	150	86.7
Maine	GW	315	109	34.6
Maine	SW	79	29	36.7
Michigan	GW	1141	644	56.4
Michigan	SW	76	33	43.4
Minnesota	GW	902	863	95.7
Minnesota	SW	24	23	95.8
Missouri	GW	1097	773	70.5
Missouri	SW	96	89	92.7
Montana	GW	544	484	89.0
Montana	SW	50	47	94.0

Table 4-4a (continued)
CWS Systems in SDWIS and State Compliance Monitoring Data

State	Source Type	Systems in SDWIS[a]	Systems in State Data Set[a]	Percent Coverage
North Carolina	GW	1848	1735	93.9
North Carolina	SW	171	169	98.8
North Dakota	GW	202	197	97.5
North Dakota	SW	19	19	100.0
New Hampshire	GW	618	504	81.6
New Hampshire	SW	38	37	97.4
New Jersey	GW	494	438	88.7
New Jersey	SW	30	29	96.7
New Mexico	GW	582	573	98.5
New Mexico	SW	29	29	100.0
Nevada	GW	249	221	88.8
Nevada	SW	34	31	91.2
Ohio	GW	1016	875	86.1
Ohio	SW	165	139	84.2
Oklahoma	GW	458	446	97.4
Oklahoma	SW	211	210	99.5
Oregon	GW	665	583	87.7
Oregon	SW	138	134	97.1
Texas	GW	3421	3105	90.8
Texas	SW	341	326	95.6
Utah	GW	334	327	97.9
Utah	SW	38	38	100.0
Total Systems	GW	20779	17198	82.8
Total Systems	SW	2801	2308	82.4

Note:
[a] The total number of non-purchased public and private systems. Ground water under the influence of surface water systems are included as surface water systems.

Table 4-4b
NTNCWS Systems in SDWIS and State Compliance Monitoring Data

State	Source Type	Systems in SDWIS[a]	Systems in State Data Set[a]	Percent Coverage
Alaska	GW	152	140	92.1
Alaska	SW	28	26	92.9
Alabama	GW	43	31	72.1
Alabama	SW	7	5	71.4
Arizona	GW	209	190	90.9
Arizona	SW	5	4	80.0
California	GW	1068	376	35.2
California	SW	53	11	20.8
Indiana	GW	703	538	76.5
Indiana	SW	6	4	66.7
Kansas	GW	64	62	96.9
Kansas	SW	2	2	100.0
Michigan	GW	1783	230	12.9
Minnesota	GW	653	634	97.1
Minnesota	SW	6	6	100.0
Missouri	GW	228	190	83.3
Missouri	SW	4	4	100.0
Montana	GW	213	2	0.9
Montana	SW	3	0	0.0
North Carolina	GW	644	562	87.3
North Carolina	SW	8	8	100.0
North Dakota	GW	22	20	90.9
North Dakota	SW	7	6	85.7
New Jersey	GW	975	758	77.7
New Jersey	SW	3	3	100.0
New Mexico	GW	153	140	91.5
New Mexico	SW	8	8	100.0
Oregon	GW	325	112	34.5
Oregon	SW	9	9	100.0
Texas	GW	727	580	79.8
Texas	SW	48	27	56.3
Utah	GW	52	50	96.2
Utah	SW	3	3	100.0
Total Systems	GW	8014	4615	57.6
Total Systems	SW	200	126	63.0

Note:
[a] The total number of non-purchased public and private systems. Ground water under the influence of surface water systems are included as surface water systems.

CWS ground water systems are represented, and another six states have between 85 and 90 % coverage. California, Maine, and Michigan had the lowest coverage of CWS ground water systems, on a percentage basis; yet, in each of these States, more than 100 systems are represented, and more than 1,000 systems are represented in California. Among CWS surface water systems, 20 States data sets include at least 90 percent of PWS, and the lowest percentage of coverage is provided by the data sets for California, Maine, and Michigan. For NTNCWS systems, 7 of the 17 states with ground water systems had 90 percent or more of their systems represented, and 8 of the 16 states with surface water systems had 90 percent or more of their systems represented. The lowest percentage coverages for NTNCWS systems were for Montana (0 out of 3 surface water systems and 1 % of ground water systems), Michigan (12.9 % of ground water systems), and California (35.2 % of ground water systems and 20.8 % of surface water systems).

The total number of ground water and surface water systems represented is far higher than the number available from other national data sources: There were 17,198 CWS ground water systems, 2,308 CWS surface water systems, 4,615 NTNCWS ground water systems, and 126 NTNCWS surface water systems. (Note that these counts include all systems in the database. As discussed below, 1,588 of these 24,247 systems were for years with higher reporting limits and were not used for the analyses in Chapters 5, 6, and 7).

Analytical Methods and Reporting Limits

Data for 11 of the States are censored[8] at a single reporting limit, but for 14 States, data are censored at multiple reporting limits that usually range from 1 - 10 µg/L. Where there are a range of reporting limits, it suggests that arsenic samples may have been analyzed with more than one sample method. Several of the State data sets include method numbers associated with individual samples. Based on these method numbers, the following analytical techniques, which are approved for compliance monitoring at the arsenic MCL of 50 µg/L[9], are represented among the compliance monitoring data sets (systems must use one of these approved methods for compliance monitoring at a certified laboratory):

- Inductively coupled plasma (ICP) - Atomic Emission Spectrometry (AES)
- ICP-Mass Spectrometry (MS)
- Platform Graphite Furnace - Atomic Adsorption (AA)
- Graphite Furnace AA; and
- Hydride Generation AA.

Because these samples were collected for compliance monitoring, it is assumed that appropriate sample collection and laboratory quality assurance and quality control protocols were followed.

---

[8] Censored data are samples with contaminant concentrations reported as less than the analytical detection limit. Actual contaminant concentrations in these samples may be positive, and may range from zero to the detection limit. In the case of a naturally occurring contaminant, such as arsenic, contaminant concentrations may be exceedingly low, but are rarely zero.

[9] 40 CFR Section 141.23.

Time trends were found in the reporting limits for 8 of the 14 States with multiple reporting limits when the data were sorted by date. These trends were found by simple inspection, rather than a formal statistical trend test. These data are summarized in Table 4-5, combining data for CWS and NTNCWS systems. This summary table shows the distributions of the reporting limits for two groups of years in each state, one group having generally lower reporting limits. These groupings were chosen separately for each state based on a detailed review of the reporting limit distributions by state and year (not shown). Thus, the selected break year varies across the eight states.

Table 4-5
Summary of Reporting Limits (RLs) for Eight States

State	Year Range	Total Number of Samples	Total ND Samples, RL Known	Total ND Samples, RL Estimated	Total ND Samples, RL Known, <= 2	Total ND Samples, RL Known, 2.01 to 5	Total ND Samples, RL Known, 5.01 to 10
Alaska	1991 to 1994	1152	767	0	767	0	0
Alaska	1995 to 1997	367	307	0	0	307	0
Arizona	1988 to 1995	5516	3595	2	133	1401	2061
Arizona	1996 to 1998	1459	598	1	19	553	26
California	1980 to 1994	13481	9574	1297	1333	3250	4991
California	1995 to 2000	11425	3414	3188	3314	92	8
Illinois	1992 to 1995	1321	1037	1	1020	16	1
Illinois	1996 to 2000	3019	2090	2	1891	187	12
Minnesota	1992 to 1993	621	442	0	266	176	0
Minnesota	1994 to 1997	3085	1730	0	1730	0	0
New Mexico	1980 to 1994	2765	1604	109	114	1334	156
New Mexico	1995 to 2000	2855	816	4	740	72	4
Oregon	1990 to 1992	1211	5	1032	0	3	2
Oregon	1993 to 1998	1222	972	59	284	521	167
Utah	1980 to 1988	2030	1228	1	1156	22	50
Utah	1989 to 1999	2872	2364	0	388	1879	97

The total numbers of non-detect samples given in these tables include non-detect samples where a known reporting limit was supplied with the data and non-detect samples where the reporting limit was unknown. Samples with zero concentration values were also treated as non-detect samples with an unknown reporting limit. The reporting limit distributions shown in these tables are for the non-detect samples with known reporting limits only. For the remaining non-detect samples with unknown reporting limits, the reporting limit was estimated as the most commonly occurring reporting limit for the state and year range, as described in section 4.1.3

below. To avoid a biased comparison, those non-detect samples with unknown, estimated reporting limits are not included in these counts of non-detect samples with reporting limits below 2, from 2 to 5, or from 5 to 10 µg/L.

These data show that reporting limits have decreased in four of the seven States other than Oregon, and increased in the other three States. For Oregon, almost all of the non-detects between 1990 and 1992 had an unknown reporting limit, and so use of the newer data is preferred. In Arizona, the majority of samples collected between 1988 and 1995 had reporting limits of 10 µg/L, but between 1996 and 1998, the primary reporting limit declined to 5 µg/L. In California, Minnesota, and New Mexico, reporting limits were sharply lower in the later time periods indicated in Table 4-5. In these three States, the majority of non-detect samples had a reporting limit of 2 µg/L in the later time periods. In contrast, reporting limits in Alaska rose from 2 µg/L to 5 µg/L after 1994 and the reporting limits in Utah rose from values mainly $\leq$2 µg/L in 1980 to 1988 to values in the range 2 - 5 µg/L from 1989 on. In Illinois, moderate increases in reporting limits were seen in the later time period. No time trends were observed in the data for the remaining 6 States – Indiana, Kentucky, Michigan, New Jersey, North Carolina, and Ohio – with multiple reporting limits.

Based on these analyses, a decision was made that for those eight states with strong time trends in the reporting limits, the data subsets for the years with the lower reporting limits would be used for most of the analyses, including the occurrence computations in chapter 6 and the intra-system variability analyses in chapter 7. The main reason for using these data subsets is that there can be substantial bias introduced by the substitution method of half the detection limit for non-detects when detection limits are higher. This substitution method was used for systems with four or fewer detected values (i.e., values above the reporting limit), or with five or more equal detected values, which was the case for the majority of systems with some non-detects. Thus it is reasonable to use the subset with lower reporting limits even if that meant that the older data was used instead of newer data with higher reporting limits. We believe that any change in background arsenic levels occurs very slowly, so that it is reasonable to use data as old as 1980 if that data is of higher quality. For the other 17 states, all the data would be used. The AOED includes the original data set with all years of data and includes the data subsets used for the analyses.

As discussed in the last paragraph, for eight states the state subset of years with lower reporting limits was used to develop the occurrence estimates. Tables 4-6a (CWS systems) and 4-6b (NTNCWS systems) show the years used to define those state subsets and compare the arsenic occurrence data contained in the complete State data sets with the occurrence data contained in the subsets for ground water and surface water systems. The tables show the numbers of samples, systems and percentage of non-detect samples. Also shown are the minimum, mean, and maximum of the system means; the calculation of estimated system means for systems with one or more non-detect samples using a regression or order statistics approach is detailed in Section 6.1.1 and appendix A below. Because the subsets have lower detection levels, it was expected that the subsets would have lower censoring levels and more accurate system mean arsenic concentrations. For CWS systems in seven of the eight States, subsetting reduces the percent of samples censored for ground water systems. The largest reductions occurred in the States of Arizona (16.3 percentage points), California (13.3 percentage points),

Table 4-6a
Overview of Complete Data Sets versus Data Subsets of CWS Systems

State & Source Type	Data Set	Number of Systems	Number of Samples	Average System Mean [As]*	Min. System Mean [As]*	Max. System Mean [As]*	Percent Samples ND
AK GW	All	326	628	4.23	0.03	50.00	64.3
AK GW	1991 to 1994	304	476	4.11	0.01	61.88	60.3
AZ GW	All	668	4126	7.82	0.42	118.19	64.0
AZ GW	1996 to 1998	279	805	9.53	0.28	101.60	47.7
CA GW	All	1369	21143	4.92	0.00	96.71	68.4
CA GW	1995 to 2000	1224	9494	4.20	0.00	99.00	55.1
IL GW	All	1082	3576	2.36	0.04	61.40	69.9
IL GW	1992 to 1995	750	1131	2.11	0.01	59.10	75.8
MN GW	All	863	2423	2.76	0.05	65.82	57.9
MN GW	1994 to 1997	829	2021	2.77	0.04	65.82	55.6
NM GW	All	573	4775	4.22	0.20	42.56	43.9
NM GW	1995 to 2000	559	2334	3.81	0.06	57.72	27.5
OR GW	All	583	1304	3.41	0.30	50.50	78.2
OR GW	1993 to 1998	316	576	2.77	0.04	56.00	75.3
UT GW	All	327	3308	3.80	0.41	51.13	71.4
UT GW	1980 to 1988	263	1284	2.89	0.05	55.49	57.8
AK SW	All	109	532	1.28	0.07	13.10	82.1
AK SW	1991 to 1994	106	427	1.29	0.13	14.33	78.0
AZ SW	All	46	2096	7.21	2.00	63.45	51.9
AZ SW	1996 to 1998	33	518	4.68	1.90	15.25	29.3
CA SW	All	222	2769	3.33	0.26	68.33	84.2
CA SW	1995 to 2000	176	1280	2.38	0.24	39.25	76.4
IL SW	All	103	764	0.92	0.50	4.42	82.7
IL SW	1992 to 1995	93	190	0.75	0.30	3.18	95.3
MN SW	All	23	63	0.87	0.58	1.60	82.5
MN SW	1994 to 1997	23	54	0.88	0.63	1.25	79.6
NM SW	All	29	430	2.01	0.66	3.74	71.9
NM SW	1995 to 2000	29	141	1.03	0.18	4.32	53.2
OR SW	All	134	944	2.30	0.45	16.03	94.0
OR SW	1993 to 1998	129	514	1.45	0.22	10.66	94.2
UT SW	All	38	1376	2.41	0.43	15.41	79.0
UT SW	1980 to 1988	35	698	1.94	0.18	20.56	65.6

* Average, minimum, and maximum CWS system mean arsenic concentrations in ug/L .

Table 4-6b
Overview of Complete Data Sets versus Data Subsets of NTNCWS Systems

State & Source Type	Data Set	Number of Systems	Number of Samples	Average System Mean [As]*	Min. System Mean [As]*	Max. System Mean [As]*	Percent Samples ND
AK GW	All	140	275	5.99	0.04	54.00	60.0
AK GW	1991 to 1994	131	191	5.39	0.02	54.00	55.0
AZ GW	All	190	726	9.62	0.60	195.00	62.0
AZ GW	1996 to 1998	77	128	6.73	0.19	53.92	48.4
CA GW	All	376	948	4.52	0.00	110.00	66.6
CA GW	1995 to 2000	330	620	4.13	0.00	110.00	59.5
MN GW	All	634	1204	2.66	0.03	51.86	58.5
MN GW	1994 to 1997	630	993	2.65	0.03	56.00	55.4
NM GW	All	140	388	5.62	0.12	47.40	28.9
NM GW	1995 to 2000	140	358	5.66	0.10	47.40	24.9
OR GW	All	112	155	2.25	0.12	20.00	84.5
OR GW	1993 to 1998	84	107	2.25	0.10	20.00	82.2
UT GW	All	50	193	4.45	0.28	36.33	62.2
UT GW	1980 to 1988	17	42	3.38	0.11	16.00	54.8
AK SW	All	26	84	1.48	0.64	2.67	81.0
AK SW	1991 to 1994	24	58	1.13	0.36	2.75	72.4
AZ SW	All	4	27	3.44	2.50	6.27	66.7
AZ SW	1996 to 1998	2	8	4.25	2.50	6.00	12.5
CA SW	All	11	46	3.46	0.25	5.00	84.8
CA SW	1995 to 2000	10	31	1.41	1.00	2.75	77.4
MN SW	All	6	18	0.66	0.50	1.00	83.3
MN SW	1994 to 1997	6	17	0.70	0.50	1.25	82.4
NM SW	All	8	27	0.97	0.50	2.05	66.7
NM SW	1995 to 2000	8	22	0.87	0.50	2.05	59.1
OR SW	All	9	30	NA	NA	NA	100.0
OR SW	1993 to 1998	8	25	NA	NA	NA	100.0
UT SW	All	3	25	1.72	0.50	2.50	96.0
UT SW	1980 to 1988	1	6	NA	NA	NA	100.0

* Average, minimum, and maximum NTNCWS system mean arsenic concentrations in ug/L . The system means are undefined if all samples in all systems are censored, and then the average minimum, and maximum system means are not available (NA).

New Mexico (16.4 percentage points), and Utah (13.6 percentage points). For each of these States, subsetting the data excludes an adequate number of samples with reporting limits of 10. For Illinois, the censoring percentage rose slightly from 69.9 % to 75.8 %. For the other three States, reporting levels in the original data sets are generally lower than those in the original data sets for Arizona, California, New Mexico, and Utah. This explains the smaller reductions in percentage of samples that are censored for the Alaska, Minnesota, and Oregon. For CWS surface water systems, subsetting reduces the percent of samples that are censored for six of the eight States, with the most significant decreases also occurring in the States of Arizona (22.6 percentage points), California (7.8 percentage points), New Mexico (18.7 percentage points), and Utah (13.6 percentage points). The percentage of censored CWS surface water samples rose for the States of Illinois (13.4 percentage points) and Oregon (0.2 percentage points); these two States had the highest levels of censoring among the eight States. For NTCNWS ground water and surface water systems, subsetting reduced the percentages of censored samples in eight and seven of the eight states, respectively. The exception was a slight rise for SW systems in Utah. The largest reductions were for Arizona, where the percentage censored was reduced by 13.6 percentage points for ground water and by 54.2 percentage points for surface water (but from only 27 samples).

The minimum and average system mean arsenic concentrations for both ground water and surface water systems are generally a little lower for the data subsets than for the complete State data sets. This difference is probably a result of the conventions that were applied to estimate system mean arsenic concentrations. These conventions are presented in Section 6.1.1 of this report. In cases where there are five or more detects that are not equal, a regression on order statistics method (detailed in appendix A) was used to estimate the concentrations for the non-detects by fitting a lognormal distribution to the detects. However, the convention that may have the greatest impact is for the calculation of the system mean arsenic concentration when there are four or fewer detects (samples that were above the reporting limit) or when there are five or more detects that are all equal. In this case, non-detects are treated as having an estimated concentration of half the reporting limit. For example, consider a hypothetical system in California with 12 samples, including 2 positive detects at concentrations of 2.5 and 2.8 µg/L, five non-detects at 2 µg/L and five non-detects at 10 µg/L. The samples with reporting limits of 10 were collected prior to 1995, while the positive detects and the samples with a reporting limit of 2 µg/L were collected after 1995. In accordance with the data convention described above, the mean arsenic concentration for this system is 2.94 µg/L when the entire data set is considered, and is 1.47 µg/L when the recent data subset is considered. Thus, as this example indicates, the higher reporting levels associated with the complete data sets can result in higher individual system mean concentrations. In turn, higher system mean concentrations result in higher State mean arsenic concentrations. Because a sample with a lower reporting limit provides a more accurate indication of the arsenic concentration in drinking water, system means calculated from such samples are believed to be more accurate than system means calculated from samples with higher reporting limits.

Tables 4-7a (CWS systems) and 4-7b (NTNCWS systems) summarize the differences in the number of systems that are represented in the complete State data sets relative to the number of systems that are represented in the subsets of data. The highest reductions in coverage for ground water CWS systems occurred for the States of Arizona, Illinois, and Oregon. However, as shown in Table 4-6a, each of the data subsets for these States contains a large number of systems. For surface water systems, the greatest reductions in coverage occurred in Arizona and California. Again, even with the reduction in coverage, these data sets provide an adequate amount of data for arsenic occurrence estimation. For NTNCWS systems the greatest reductions in coverage were for Utah, which originally had the smallest number of NTNCWS systems in the database, and for Arizona, which again provided an adequate amount of data after subsetting.

Table 4-7a
Relative Change in CWS System Coverage

State	Groundwater	Surface Water
AK	-6.75%	-2.75%
AZ	-58.23%	-28.26%
CA	-10.59%	-20.72%
IL	-30.68%	-9.71%
MN	-3.94%	0%
NM	-2.44%	0%
OR	-45.80%	-3.73%
UT	-19.57%	-7.89%

Table 4-7b
Relative Change in NTNCWS System Coverage

State	Groundwater	Surface Water
AK	-6.43%	-7.69%
AZ	-59.47%	-50%
CA	-12.23%	-9.09%
MN	-0.63%	0%
NM	0%	0%
OR	-25%	-11.11%
UT	-66%	-66.67%

As a result of the association between reporting limits and time in the 8 States, most of the following statistical analyses in Chapters 5, 6, and 7 were conducted using the subsets of data for each State that cover the period when reporting limits were lower.

Multiple Samples for Individual PWS (Over Time and Space)

One of the unique aspects of the compliance monitoring data sets is that many of them provide multiple samples for many PWS. There are multiple measurements at the same location (POE) but different dates and times (temporal variability). There are sometimes measurements at multiple POE locations (spatial variability). For the systems in the INTRA database, the POE is identified and so the temporal and spatial variability can be separated (see Chapter 7). For other systems without POE identifiers, it is not possible to distinguish the two types of variability.

Tables 4-8a (CWS) and 4-8b (NTNCWS) summarize the average, minimum, and maximum number of samples per PWS in each State data set, by source water type. For the eight states with data subsets having lower reporting limits, the data subset was used. In general, there are more samples per system for surface water systems than for ground water systems. This is consistent with expectations, because surface water systems are required to monitor more frequently than ground water systems. In several States, some systems are represented by more than a hundred samples. Within individual systems, some of these samples have been collected on the same day, while others have been collected over many years. In addition, samples can come from multiple POE in a system. In some States, these samples have been censored at multiple reporting limits, while in other States, all samples are censored at the same reporting level. Having multiple samples for individual PWS offers several benefits, but also presents some challenges.

With multiple samples over time, it is possible to average samples in order to develop a better estimate of system arsenic levels. In addition, having multiple samples for individual PWS can support analyses of variability in arsenic concentrations over time and from location to location within individual PWS. For example, the assessment of intra-system variability presented in Chapter 7 is based on the analyses of arsenic occurrence distributions from location to location within individual PWS. However, when systems are represented by different numbers of samples, it may be necessary to estimate a single system level statistic for each system so that arsenic levels are comparable from system to system. This aggregation ensures that a system with 50 samples can be represented and compared to a system with only one sample. In this report, samples for individual systems have been averaged using the procedure discussed in Section 6.1.1.

Table 4-8a
Summary of Numbers of Samples per CWS System for State Compliance Monitoring Data

State	Ground Water Systems			Surface Water Systems		
	Min. (N)	Mean (N)	Max. (N)	Min. (N)	Mean (N)	Max. (N)
AK	1	1.6	14	1	4	102
AL	1	7	81	4	15.1	95
AR	1	1.8	7	1	4.1	20
AZ	1	2.9	51	1	15.7	121
CA	1	7.8	458	1	7.3	77
IL	1	1.5	21	1	2	6
IN	1	1.4	12	1	4.4	19
KS	1	4.6	39	1	6.7	11
KY	1	4.6	14	2	13.2	28
ME	1	2.5	77	2	3.6	9
MI	1	2	253	1	1.8	7
MN	1	2.4	20	2	2.3	3
MO	1	1.1	4	1	1	2
MT	1	3.6	15	2	9.4	24
NC	1	6	112	2	17	56
ND	1	1.1	5	1	1	1
NH	1	2.4	45	1	5.6	15
NJ	1	3.7	65	3	11.1	108
NM	1	4.2	166	1	4.9	18
NV	1	1	2	1	1.1	2
OH	1	4.2	71	2	11.5	80
OK	1	2.3	78	1	1.6	64
OR	1	1.8	28	1	4	15
TX	1	2.1	42	1	6.4	52
UT	1	4.9	67	1	19.9	256

Table 4-8b
Summary of Numbers of Samples per NTNCWS System
for State Compliance Monitoring Data

State	Ground Water Systems			Surface Water Systems		
	Min. (N)	Mean (N)	Max. (N)	Min. (N)	Mean (N)	Max. (N)
AK	1	1.5	6	1	2.4	6
AL	1	3.9	11	8	9.2	10
AZ	1	1.7	13	1	4	7
CA	1	1.9	28	1	3.1	6
IN	1	1.2	8	1	2.8	5
KS	1	2.2	9	3	3	3
MI	1	1.2	11	--	--	--
MN	1	1.6	6	1	2.8	5
MO	1	1.1	12	1	1	1
MT	1	2	3	--	--	--
NC	1	2.3	9	4	5.5	7
ND	1	1	1	1	1	1
NJ	1	1.2	7	2	2.3	3
NM	1	2.6	27	1	2.8	6
OR	1	1.3	3	1	3.1	5
TX	1	1.7	5	1	4.4	8
UT	1	2.5	6	6	6	6

Water System Type

Many of the compliance monitoring data sets provide data on arsenic occurrence in NTNCWS systems as well as CWS systems. In fact, as shown on Table 4-2, data sets for 17 States contain some data from NTNCWS systems, although one of these State data sets (for Montana) contains information for only two systems. For 12 States, more than 100 ground water NTNCWS systems are represented. Far fewer NTNCWS surface water systems are represented, but the total population of NTNCWS surface water systems in the United States is quite small (442 non-purchased NTNCWS systems). Information about NTNCWS systems is not available in the other arsenic occurrence databases discussed in Section 4.2 and 4.3.

Point-of-Entry (POE) Data

The compliance monitoring data sets are unique from other data sets in that most of them contain multiple data points for individual PWS systems. As discussed above, there can be multiple measurements at the same POE but different dates or times, and/or measurements at different POE for the same system.

For individual systems, compliance with the arsenic standard is measured at the POE into the distribution system. A POE represents a well or entry point to the distribution system, therefore a location where treatment may need to be installed if they have arsenic levels above the MCL. Systems may have more than one POE. Generally larger systems have more POE; also, individual POE or treatment plants may be supplied by a network of wells. A system with multiple POE may need to install multiple treatment systems, depending on the contaminant concentrations at each POE. Therefore, to estimate compliance costs, it would be ideal to know the number of POE in each system, and link each sample result to a specific POE ID. However, a complete inventory of individual POE is not available, and many of the compliance monitoring data sets do not include POE ID.

POE ID are available for the following ten States:

- Alabama
- Arkansas
- California
- Illinois
- Indiana
- New Mexico
- North Carolina
- Oklahoma
- Texas
- Utah.

These data were used to assess intra-system variability in arsenic concentrations. The intra-system variability estimation and results are discussed in Chapter 7 of this report.

4.1.3 Building the AOED from SDWIS and State Compliance Monitoring Databases

Two separate databases comprise AOED, and these databases were developed from the information contained in the individual State compliance data sets and SDWIS. These databases are named GRAND and INTRA, and they were created in SAS. GRAND was designed to support statistical evaluations and the development of national occurrence projections presented in Chapters 5 and 6, and it contains the data sets of all 25 States. INTRA was developed to support the assessment of intra-system variability presented in Chapter 7, and it includes the ten data sets which include POE identifiers. The GRAND database has 131,383 records, and INTRA has 88,855 records. For the eight states where subsets of data for years with lower reporting limits were used, the entire original database (all years) is included and an additional copy of the subset data is also included. For example, for California, the original database is the set of records with ST_GROUP = "CA," and the subset database is the set of records with ST_GROUP = "CA2." All the records in the data subset appear twice, first with ST_GROUP = "CA" and second with ST_GROUP = "CA2." After excluding 5 outlier values with concentrations above 1000 µg/L and excluding the unsubsetted original data for the eight states, this leaves 76,973 records in GRAND and 49,081 records in INTRA that were used for the analyses in the remaining chapters of this report, including the regional and national occurrence analyses and the intra-system variability analyses. Each record in these databases correlates with a finished drinking water sample from a PWS. Appendix D-2 provides a list of the variables in these databases, and the characteristics of those variables. All of the records in INTRA are included in GRAND, although INTRA include two additional variables, which are the POE identifier and the POE type. INTRA was developed to improve computational efficiency for the intra-system analyses; it includes only information that is relevant to those analyses, and was developed with minimal additional effort.

Initially, ISSI compiled the raw State compliance monitoring data sets in an Access database. The initial data conditioning processes are described in Appendix D-3. Subsequently, the Access database was converted to a SAS format, and the GRAND and INTRA databases were created. In addition, raw compliance monitoring data sets were received for the States of Alabama, Arkansas, Illinois, Indiana, New Mexico, Oklahoma, and Oregon after the Access database had been converted to SAS, and these compliance monitoring data sets were loaded directly into SAS. In August and September, 2000, updated and revised compliance monitoring data was received for the States of Alabama, California, Illinois, New Mexico, North Carolina, and Texas, and used to replace the previous data for those states. Also in September, 2000, a previously submitted, but erroneously omitted updated database from Utah was used to replace the previous data for Utah.

Simply stated, the data conditioning process standardized the coding of variables from State to State. For example, different States reported sample collection dates in different formats. Through the data conditioning process, dates were standardized to a consistent format. Similarly, result flags were initially inconsistent. Some States reported ND (not-detect), others LT (less than), some N and D, and some marked non-detects with the "<" symbol. After confirming the meaning of each flag with the data contact in the State, these flags were standardized to N and D for non-detects and detects. The result and the reporting limit variables were standardized to µg/L, because some States provided data in µg/L and others reported data in

mg/L. PWS source type was standardized to GW and SW (systems coded as ground water under the influence of surface water were coded to surface water), and type of water system was standardized to CWS and NTNCWS. Data for other types of water system were excluded from the data base.

SDWIS was an important tool for the development of the arsenic databases, and it was used in a variety of ways. Each of the suitable State compliance monitoring databases provided SDWIS PWSID numbers. By linking the State compliance monitoring data to SDWIS by PWSID number, we were able to complete missing data where necessary, update variables which could change over time (such as population served or if a PWS was active or inactive), and check for discrepancies between the compliance monitoring data and the SDWIS data. The SDWIS Baseline for 1998 was used for the analyses in this report. When discrepancies in the type of water system (CWS or NTNCWS), source type (GW or SW), or population served were identified between State data and SDWIS data, the State data were corrected to reflect the SDWIS data. By doing so, the databases were corrected in a manner that was consistent from State to State. The SDWIS database was used to find and remove from the database records for suppliers of purchased water. Inactive systems were also removed. In relation to the total number of records in the database (131,383), about 10 % required correction. For 520 systems, SDWIS identified these as having purchased water, so these systems were excluded from the database. Also excluded were a total of 1,798 systems that were not in the SDWIS database (this includes relatively new systems not included in the 1998 SDWIS update). A total of 18,280 records were deleted as either being for purchased water or for a system not in SDWIS. For 172 systems, SDWIS and State data sets disagreed about the type of water system (CWS versus NTNCWS); and for 119 systems, SDWIS and State data sets disagreed about the system source water type (GW or SW).

Several other conventions were applied during the development of the GRAND and INTRA databases, and these conventions are discussed below:

1. Sample analyses conducted prior to 1980 were deleted. This is roughly the year that new analytical techniques had become widely available and less precise colorimetric analysis was phased out.

2. Analyses with no PWSID number, or no analytical result, or from systems identified in SDWIS or in the supplied database as either inactive, abandoned, or as suppliers of purchased water,[10] were deleted.

3. Analyses for water samples identified as either being finished, or from sources not subject to treatment, or for samples measured after treatment were included in the database as finished water. For ground water, samples identified as raw samples or as being measured before treatment were included in the database, and were included in the data analyses. For surface water, samples identified as raw samples or as being measured before treatment were excluded from the database,

[10] Purchased water systems do not withdraw the water that they supply to their customers directly from a ground water or surface water source, but instead they purchase it from another water system.

and, hence, from the data analyses. Briefly, the "raw" ground water samples were included since the treatments used are not expected to impact the arsenic concentration levels in ground water, but may have an impact for surface water.

4. Analyses with results greater than 1000 µg/L were left in the database but excluded from the analysis, because these values were believed to be the result of data entry or data conversions errors, rather than representative of actual arsenic concentrations in drinking water. As a result of this convention, six arsenic values were deleted from the analyses, but included in the database. These were all for CWS ground water systems. There was one sample from a public water supply system in New Mexico with a value of 189,000 µg/L; this record was in both the all year set and the lower reporting limit subset. There were five samples from Texas with values of 18,500, 18,500, 14,500, 4,800, and 3,400 µg/L.

5. Arsenic values reported as "zero" or "ND" were considered to represent an analytical result below the reporting limit. If the State provided reporting limits for some samples, but did not provide reporting limits for other non-detects, such as the zero values, then the reporting limit for those records was estimated as the modal (most commonly occurring) reporting limit for the state and associated subset of years. For the 17 states where a subset with lower reporting limits was not defined, the modal reporting limit for all the non-detects was used. For the other eight states, the estimated reporting limit for the all years subset was the modal reporting limit for all the non-detects, but the estimated reporting limit for the lower reporting limit subset of years was the modal reporting limit for all the non-detects reported in those years. For example, for California the estimated reporting limits used to fill in unknown reporting limits were 10 µg/L for the original 1980 - 2000 subset (ST_GROUP = "CA") but was 2 µg/L for the 1995 - 2000 subset (ST_GROUP = "CA2"). In California, most of these filled in reporting limits were for arsenic values reported as "zero."

6. If the State did not disclose the reporting limits for any samples, and reporting limits could not be determined based on conversations with appropriate State representatives, reporting limits were assigned based on where the majority of the lowest measured results clustered. The later convention was applied to all non-detect data from the States of Alabama and Oregon. For the State of Indiana, reporting limits were assigned that correlated with laboratory analysis methods. Appropriate reporting limits were provided by the State representative.

7. Sample results that were non-detect with reporting limits greater than 10 ppb were deleted. This convention only affected Michigan, in which arsenic results were associated with four report limits (0.3, 1.0, 2.0, and 50 ppb). Removing arsenic results associated with a reporting limit of 50 ppb excluded 21% of Michigan data.

8. The State of Missouri reported only positive results. ISSI contacted the State, and the Missouri Department of Health provided a list of the PWSIDs for all systems with non-detect measurements for arsenic during the same three year monitoring time period as the positive results (i.e., detects). Furthermore, the State representative indicated that these systems were non-detect at the reporting limit of 1 µg/L. These data were then combined with the positive results.

9. The State of Alaska reported that measurements made in 1990 or earlier are now considered to be unreliable, and so these records were deleted from the database. The State of Alaska also reported that until 1999, all their systems were reported to SDWIS as CWS systems, even though the systems in Alaska included CWS, NTNCWS and transient non-community water supply systems. The SDWIS 1998 baseline reports no NTNCWS systems for Alaska. The state provided a database giving the type of water system for each PWSID. This database was used to adjust the type of water system variable in GRAND and INTRA and also in the SDWIS database used for the all analyses described in this report.

4.2 Comparison Databases

Four databases are available for use as comparison tools on a national basis: the NAOS, USGS, NIRS, and Metro databases. All of these databases are national in scope, but the characteristics of these databases differ substantially. Likewise, the survey methods that were used to collect samples for NAOS and NIRS were quite different, and samples that are represented in USGS and Metro databases were not collected in accordance with a specific survey method. Each of these databases is described below, and comparisons of national arsenic occurrence estimates based on the AOED, NAOS, USGS, NIRS, and Metro databases are presented in Chapter 6.

For California, a comparison can be made between the California systems in the AOED and the ACWA databases. The results of a comparison of the California arsenic occurrence estimates based on the AOED and estimates based on the ACWA databases is also presented in Chapter 6.

4.2.1 National Arsenic Occurrence Survey (NAOS) Database

The Water Industry Technical Action Fund (WITAF)[11] sponsored the National Arsenic Occurrence Survey (NAOS), which was a nationwide arsenic survey (Frey and Edwards, 1997). The NAOS was designed to be quantitatively representative of different source water types, system sizes, and natural occurrence patterns (based on a derived natural occurrence factor – discussed below), and to have a low detection limit (0.5 µg/L).

[11] WITAF includes the following organizations: American Water Works Association, National Association of Water Companies, Association of Metropolitan Water Agencies, National Rural Water Association, and National Water Resources Association..

NAOS was based on a representational survey design. PWS were selected from within three representational groups, including:

- source type (ground water and surface water);
- system size (small: 1,000 to 10,000 people served; and large: >10,000 people served); and,
- natural occurrence factor (NOF) level (which includes consideration of geographic region).

The NAOS survey design used the NOF to qualitatively represent the relative probability of arsenic occurrence in water sources. For each State, separate ground water and surface water NOF levels were estimated. USGS's water quality database (WATSTORE) was used to derive surface water NOF levels for all States except Indiana, and ground water NOFs for 35 States. The remaining State source types were not represented in WATSTORE. Metro data was used to confirm the NOFs based on WATSTORE and to derive the missing NOF designations.

NOF levels were designated as low, medium, or high, and were based on total scoring in four criteria. These elements were:

1) the probability of left censored (below detection limit) data;
2) the probability of measured observations below 5 µg/L;
3) the probability of measured observations above 20 µg/L; and
4) local interest in arsenic in source water, as indicated by the total number of samples.

For each criterion, potential scores were 5, 15, or 25 points. Final NOF assignments for ground water and surface water sources were based on the sum of the scores for these criteria. These NOF levels served as a basis for identifying distinct regional arsenic occurrence patterns. The seven regions that were identified, and the States included in each region, are shown below:

Table 4-9
States in the Seven NAOS Regions

Region	States in Region
New England	CT, NH, NJ, NY, MA, ME, RI
Mid-Atlantic	DE, KY, PA, MD, NC, SC, VA, WV
Southeast	AL, FL, GA, MS, TN
Midwest Central	OH, IA, IL, IN, MI, MN, WI
South Central	AR, CO, KS, LA, MO, NE, NM, OK, TX
North Central	MT, ND, SD, WY
Western	AK, AZ, CA, HI, ID, NV, OR, UT, WA

NAOS was designed to include approximately 800 samples that were selected to proportionally represent the survey design criteria (system sizes, source type, and geographic region). A number of samples came from the same PWS, but from different facilities or wells within the PWS. Systems were mailed a survey that requested their participation. Large systems were selected from the Water Industry Database, and small systems were selected from the Federal Reporting Data System(FRDS). Approximately the same number of small and large systems were sampled. Based on source water type distributions contained in FRDS, the ratio of surface water to ground water systems was 55:45 for large systems, and 30:70 for small systems. To achieve geographic representativeness, sample sites were selected from within each region (based on availability of sites within each region).

Sample kits with instructions were mailed to the selected utilities with a questionnaire on specific details regarding the sample. While 809 sample kits were mailed out, 517 samples were submitted to the investigators by water utilities (a 63.9 percent response rate). The response rate for large utilities (70 percent) was slightly higher than for small ones (58 percent). The investigators found that comparable portions of large and small systems responded within each survey stratum. Samples were collected from the utilities' raw water sources. During data handling, raw water results were multiplied by a removal efficiency factor associated with the treatment train in place at the utility to calculate the likely finished water arsenic concentration. There were a total of 435 predicted finished water arsenic level samples in the NAOS database (161 surface water samples and 274 ground water samples). There were an additional 54 surface water samples that did not have predicted finished water arsenic levels because there was no treatment information available. Some of the PWS in the survey had multiple water sources. Based on responses to the questionnaire, the investigators concluded that the participating utilities did not introduce bias into the survey results by selecting sources with known arsenic concentrations.

The analytical method used for the NAOS analyses had a detection limit of 0.5 µg/L, and therefore, the data set has a lower percentage of non-detects than NIRS and many of the early data sets. In ground water, arsenic was detected in 58% of samples, and in surface water, detectable arsenic levels were reported in 73% of samples.

One advantage of the NAOS data set is that all of the data samples were processed by a single laboratory, which followed strict quality assurance protocols. However, the data set has two potential drawbacks, one major and one minor. The minor drawback is that the NAOS data set does not provide PWS identification numbers, so it is difficult to identify the facilities which may be represented by more than one sample. The major drawback is that the estimated finished concentrations are not direct measurements of finished arsenic concentrations. There is a potential for true finished concentrations to differ from the estimated finished concentrations. The NAOS database was used as a comparison tool to check arsenic occurrence projections developed from the AOED.

4.2.2 USGS Arsenic Databases

The USGS has collected data on arsenic concentrations in ground water from locations throughout the United States. These data were collected from a variety of wells, some of which

are used for supply of drinking water (about 10% of samples), and others for research, agriculture, industry and domestic supply, as a part of a variety of Federal, State, and local projects (Focazio *et al*., 2000). These data were not collected specifically to develop national estimates of arsenic in drinking water; however, the USGS databases provide approximately 20,000 samples (from approximately 20,000 locations throughout the country; 1 sample per well or spring) that are potentially representative of ground water systems. Analyses were performed by hydride-generation and atomic-adsorption spectrometry and have a consistent reporting limit of 1 μg/L. Thus, the USGS data provide a significant amount of information about arsenic concentrations in ground water.

The USGS data is included in four distinct databases, which were derived from existing databases. Under an interagency agreement with the USEPA, USGS used some of these databases to project national estimates of arsenic occurrence in community water supplies with ground water sources. These estimates are discussed in Chapter 6 of this report, and the four databases which USGS developed are described below.

The **Public Supply Database** includes data derived from SDWIS on all ground, surface, and purchased water CWS. The sources of the CWS were reviewed and characterized to identify systems that are served partially or entirely with ground water. Systems that were at least partially served by ground water sources were retained, and those that were totally dependent on surface water (including purchased surface water) were excluded from the database.

The **USGS Arsenic Point Database** includes arsenic data from the USGS National Water Information System (NWIS). It includes the results of analyses of filtered water samples from 20,000 ground water wells and springs throughout the United States from 1973 to 1998. Codes are provided which allow separation of potable and non-potable water sources. Where NWIS contained multiple samples from individual wells, the most recent sample was included in Arsenic Point Database. This database also provides information on water use, well construction, and basic water quality parameters.

The **USGS Arsenic Database of Selected Counties** includes counties in which five or more arsenic samples were present in the Arsenic Point Database, and counties that include five or more arsenic values through a process of radial extrapolation. For counties with less than five arsenic values, the radial extrapolation procedure ascribed all samples within a 50 kilometer (31 mile) radius of each county center to that county. Counties with five or more arsenic analyses within the search radius were included in the database, and the arsenic samples identified by the radial extrapolation were assumed to be representative of the county. There were a total of 17,496 samples for 1,528 counties in this database. When these data were associated with the Public Supply Database, this database represented 76% of all large systems and 61% of all small systems.

The fourth USGS database is the **USGS Arsenic Database of All Counties**. This database draws upon information in the Selected Counties database, and the Ground Water Atlas of the United States. USGS used the Ground Water Atlas to identify the major aquifers in each county, and then identified the type of aquifer that ground water suppliers in each county withdraw water from. When median arsenic concentrations were calculated for counties in

specific aquifers, they were extrapolated to other counties without arsenic data that use these aquifers. Thus, arsenic values were extrapolated for the remaining counties in the United States.

From the USGS Arsenic Database of Selected Counties, the USGS developed the arsenic occurrence estimates for ground water that are presented in Chapter 6. USGS estimated the percent exceedances for each county by calculating the percentage of data points in each county (with 5 or more data points) exceeding specific arsenic concentrations, from 1 µg/L to 50 µg/L. Then USGS associated the percentages for each county with the number of systems in these counties (from the USGS public supply database). This information was aggregated for all of the appropriate counties to derive the national estimates for ground water systems.

4.2.3 National Inorganics and Radionuclides (NIRS) Database

The USEPA's National Inorganics and Radionuclides Survey (NIRS) is a national database of 983 samples of finished drinking water from community ground water systems. The survey was designed to be a nationally stratified proportional probability sample, and it was stratified by system size to be nationally representative (Longtin, 1988). Two percent of the PWS in the United States in each size stratum were sampled. Local utilities sent field-preserved samples collected between 1984 and 1986 from 48 States and Puerto Rico to USEPA laboratories for analysis. Because most of the ground water systems in the United States are small, the majority of systems represented in NIRS are very small.

NIRS contains sufficient information to support arsenic occurrence projections, and the database contains all of the critical data elements. The NIRS data, however, are censored at 5.0 µg/L, so the database does not provide information through the full regulatory range of interest. In addition, because of the relatively high reporting level, 95% of the arsenic results are censored, and therefore, projections of arsenic occurrence at levels below 5.0 µg/L based on NIRS data includes a high level of uncertainty. For large and very large strata, 100% of the results are censored.

4.2.4 Metropolitan Water District of Southern California (Metro) Database

The Metro database contains 144 arsenic samples from PWS selected from American Water Works Association (AWWA) Water Industry Database. The database generally represents large PWS (> 10,000), and the utilities represented serve more than a third (36%) of the U.S. population. This database contains 57 ground water and 87 surface water sample results. The detection limit was 0.5 µg/L and detected concentrations ranged from 1 µg/L to 39 µg/L in ground water and 1 µg/L to 5 µg/L in surface water. Most of the 144 samples contain some detectable arsenic; only 33% of the Metro sample results are censored.

The Metro data is a potentially important source of arsenic occurrence information, although it, like the NAOS database, also has limitations. Metro data represents a subset of facilities for which there are few samples in NIRS, but it provides little information about arsenic occurrence in small water systems. Metro data also provides arsenic occurrence information throughout the regulatory range of interest, because of the low detection level. However, this data set does not include PWS identification numbers for the sampled facilities. Therefore, the

Metro database cannot be used to estimate arsenic occurrence together with other databases, where the information in the two databases may overlap.

4.3 Other Databases

Several other databases that provide arsenic occurrence information are available; however, these databases are not used in this report. These databases include the 1969 and 1978 Community Water Supply Surveys (CWSS), the Rural Water Survey (RWS), the National Organics Monitoring Survey (NOMS), the Western Coalition of Arid States (WESCAS) database, and the Association of California Water Agencies (ACWA) databases. These databases, with the exception of WESCAS and ACWA, were not used in this report because the data that they contain are considered to be too old to accurately represent current conditions for arsenic occurrence. Older arsenic data may not represent current conditions for several reasons. The samples may have been analyzed with a less accurate laboratory method, the raw water sources may have changed, and treatment systems may have been installed. For example, filtration treatment added to surface water systems to comply with the Surface Water Treatment Rule would tend to decrease arsenic concentrations through incidental removal of arsenic. WESCAS data was not used because the data that are contained do not necessarily represent arsenic levels at individual PWS, and because the data conventions and handling appear to have been inconsistent from State to State. ACWA data were not used for these data analyses because compliance monitoring data was available for the State of California. Section 6.3 presents a comparison between the AOED occurrence estimates and the results of two studies of the ACWA data (Kennedy/Jenks Consultants, 1996, and Saracino-Kirby Inc., 2000). These databases are discussed below.

4.3.1 1969 Community Water Supply Survey

The U.S. Public Health Service conducted the 1969 CWSS to assess water supply facilities and drinking water quality in the nation. Samples were collected from randomly selected sites in the distribution systems of the participating PWS. A total of 969 finished water samples were collected from 678 ground water supplies, 109 surface water supplies, and 182 mixed sources (purchased water or unspecified source). Of these water samples, analytical results for arsenic were available for 673 ground water samples, and 106 surface water samples. Ninety-five percent of the ground water samples and 92 percent of the surface water samples were censored. Only 33 ground water detections and 9 surface water detections occurred. The results of the 1969 CWSS survey are summarized in Appendix C. Because this data set was collected prior to 1980, and the analytical results it contains may be less accurate than more recent data sets, these data will not be used to project arsenic occurrence and exposure estimates.

4.3.2 1978 Community Water Supply Survey

USEPA conducted a second CWSS in 1978 to determine the occurrence of organic and inorganic compounds in public water supplies. Approximately 500 water supplies provided drinking water samples. As a result of analytical anomalies and difficulties, the 1978 CWSS only contains 259 ground water and 94 surface water analytical results for arsenic. From each PWS, one to five samples of raw, finished, and/or distribution water were collected from each

supply sampled. Due to reporting inconsistencies, distributional and finished sample results were averaged together and the raw water data were not used. A total of 49 non-censored arsenic results and 3 non-censored surface water results were observed. The 1978 CWSS arsenic occurrence results are summarized in Appendix C. Like the 1969 CWSS, this data set is composed primarily of censored data: 82 percent of ground water sample data and 92 percent surface water sample results are censored. This data set is also considered to be less accurate than data collected after 1980, and will not be used to estimate arsenic occurrence and exposure for this study.

4.3.3 Rural Water Survey

Between 1978 and 1980 the RWS was conducted to evaluate the status of drinking water in rural America. A total of 71 ground water and 21 surface water samples were analyzed for arsenic from the 648 PWS surveyed. A total of 23 non-censored ground water and 2 non-censored surface water arsenic results were observed. The arsenic occurrence data contained in the RWS are summarized in Appendix C. Sixty-eight percent of the ground water analytical results and 92 percent of surface water data were censored. The RWS data are deemed insufficient to project arsenic occurrence and exposure estimates because they were collected prior to 1980.

4.3.4 National Organics Monitoring Survey

In 1976 and 1977, USEPA conducted NOMS to provide data to support the development of MCLs for organic compounds in drinking water. Trace elements were analyzed in finished water samples from 113 PWS; of this data arsenic analytical results were provided for 15 ground water and 86 surface water samples. A summary of arsenic data from NOMS is presented in Appendix C. A total of 6 non-censored ground water and 19 non-censored surface water arsenic results were observed. Sixty percent of ground water analytical results and 78 percent of the surface water analytical samples were censored. These data are deemed insufficient to project arsenic occurrence and exposure estimates because they were collected prior to 1980.

4.3.5 Western Coalition of Arid States Research Committee Arsenic Occurrence Study

In 1997, the Western Coalition of Arid States (WESCAS) Research Committee conducted an arsenic occurrence study. The study consisted of arsenic data obtained from Arizona, New Mexico, and Nevada. The primary purpose of this study was to collect data on low levels of arsenic in ground water, particularly in small systems. California data were also provided to WESCAS by ACWA for the study, but were not included in this database. The WESCAS data have been manipulated and aggregated, and the data points that are contained do not necessarily represent arsenic occurrence at individual PWS. For some States, data were re-defined to median values by area/or provider. However, the Arizona data were aggregated by county. In addition, PWS identifications were not provided for data from Arizona and Nevada. Because these data have been manipulated and are not comparable to other data sources, and because relatively comprehensive compliance monitoring data are available for Arizona, New Mexico, and Nevada, the WESCAS data were not used to project arsenic occurrence estimates presented in Chapter 6.

4.3.6 Association of California Water Agencies Database (ACWA)

Association of California Water Agencies (ACWA) (Kennedy Jenks Consultants, 1996) conducted a survey of low level arsenic occurrence in the State of California to determine the potential impact of a revised arsenic standard upon California water supply systems. More than 1500 samples (1378 ground water and 166 surface water samples) were collected between 1992 and 1994, and these analyses had detection levels of 0.1 to 1 µg/L. Arsenic was present in a greater percentage of ground water samples than surface water samples; 28 percent of the ground water samples were censored, and 52 percent of the surface water samples were censored. Arsenic concentrations in ground water were slightly higher than in surface water. The maximum concentrations in these samples were 52 and 30 µg/L, respectively, for ground water and surface water samples. Most of the systems represented in this database are medium, large, or very large PWS systems located in the southern part of the State. The survey also included information from selected ACWA members and also from Central and West Basin Municipal Water Districts and Southern California Water Company. This database was not used to develop occurrence estimates because compliance monitoring data were available for the State of California. The ACWA database does not provide PWSID numbers for each of the systems that it contains, and therefore these data cannot be linked to SDWIS. Table 4-10 identifies the percentage of ground water and surface water sources in California that will require treatment to satisfy the listed MCLs.

Table 4-10
Estimated Percentage of California Water Supplies Impacted by MCLs

Arsenic MCL (mg/L)	Surface Water Plants	Ground Water Sources
1	58%	84%
2	15%	56%
5	1%	19%
10	<1%	6%
20	<1%	3%

Adapted from Kennedy Jenks Consultants, 1996.

An updated study of California data from ACWA was carried out in 2000 by Saracino and Kirby, Inc. (ACWA, 2000).

This page intentionally left blank

5. Arsenic Occurrence Patterns in the United States

This section provides a discussion of the distribution of arsenic in drinking water in the United States. Patterns of arsenic occurrence are discussed with respect to system water source type, system size based on population served, system classification, and from region to region in the United States. The data discussed in this section come from the AOED. The statistical techniques applied to identify patterns in the available data are described, as are the results of these analyses. In addition, the results of other relevant analyses that address patterns in arsenic occurrence data are discussed and used to verify these analyses. Finally, these sections explain the basis for decisions that were made with regard to data stratification for the development of national arsenic occurrence estimates.

5.1 Stratification by Source Water Type

Previous arsenic occurrence studies (Frey and Edwards, 1997; Wade Miller, 1992 and 1989) have stratified drinking water systems on the basis of source water type into ground water and surface water systems. This stratification is appropriate because these studies indicated that arsenic concentrations in ground water and surface water differ. Therefore, source water type is a source of data heterogeneity. Stratification on the basis of source water type may also facilitate analysis of regulatory impacts, because ground water and surface water systems are regulated separately. Typical ground water and surface water systems differ in entry point configuration and treatment train (SAIC, 1999). Therefore, the occurrence analyses presented in Chapter 6 have been stratified on the basis of source water type.

As Tables 5-1a (CWS systems) and 5-1b (NTNCWS systems) show, distributions of system mean arsenic levels in ground water and surface water systems differ in each State in the AOED database. The system means were calculated using the regression on order statistics method detailed below in section 6.1.1 and appendix A. In most of the 25 States for which compliance monitoring data are available, mean arsenic concentrations are higher in ground water systems than in surface water systems. For example, in Alaska ground water CWS systems, the average system mean arsenic level was 4.11 µg/L, while in surface water CWS systems, the average system mean arsenic level was 1.29 µg/L. Similarly, in Alaska ground water NTNCWS systems, the average system mean arsenic level was 5.39 µg/L, while in surface water NTNCWS systems, the average system arsenic mean level was 1.13 µg/L. Arizona is another example where for ground water CWS systems, the average system mean arsenic level was 9.53 µg/L, while for surface water CWS systems, the average system mean arsenic level was 4.68 µg/L. In addition, while distributions of arsenic in ground water and surface water are relatively similar at the 25^{th} percentile in many States, the 75^{th} percentiles of the system mean arsenic levels are higher in ground water than surface water systems in 20 of 25 States for CWS systems. For NTNCWS systems, the 75^{th} percentiles of the system mean arsenic levels are higher in ground water than surface water systems in 8 of the 10 states with both ground and surface water systems.

Table 5-1a
Distributions of System Means for Community Water Systems

State	Source Type	Systems (N)	Minimum	25th Percentile	Median	Mean	75th Percentile	Maximum
AK	GW	304	0.01	0.35	0.90	4.11	4.40	61.88
AK	SW	106	0.13	0.48	0.73	1.29	1.00	14.33
AL	GW	263	0.24	0.53	0.68	0.73	0.83	7.00
AL	SW	68	0.43	0.60	0.68	0.70	0.79	1.45
AR	GW	371	2.50	2.50	2.50	2.52	2.50	7.00
AR	SW	76	2.50	2.50	2.50	2.52	2.50	3.67
AZ	GW	279	0.28	2.20	5.00	9.53	11.00	101.60
AZ	SW	33	1.90	2.86	3.40	4.68	4.93	15.25
CA	GW	1224	0.00	1.08	1.80	4.20	4.00	99.00
CA	SW	176	0.24	0.96	1.30	2.38	1.77	39.25
IL	GW	750	0.01	0.19	0.48	2.11	1.00	59.10
IL	SW	93	0.30	0.57	0.70	0.75	0.85	3.18
IN	GW	648	0.00	0.03	0.09	0.26	0.24	7.10
IN	SW	51	0.25	0.50	0.50	0.68	0.50	4.00
KS	GW	506	0.19	0.90	1.57	2.64	3.27	65.14
KS	SW	101	0.50	0.75	0.90	1.11	1.25	3.64
KY	GW	88	0.67	0.88	1.25	1.45	1.70	4.50
KY	SW	150	0.79	1.44	1.70	1.78	2.06	4.86
ME	GW	109	0.09	0.60	0.83	3.06	2.25	53.00
ME	SW	29	0.24	0.53	0.67	1.03	0.85	6.70
MI	GW	644	0.01	0.47	1.55	5.31	6.73	89.00
MI	SW	33	0.04	0.18	0.33	0.97	0.60	8.99
MN	GW	829	0.04	0.55	0.94	2.77	2.30	65.82
MN	SW	23	0.63	0.77	0.85	0.88	0.94	1.25
MO	GW	773	0.03	0.24	0.43	0.77	0.71	34.78
MO	SW	89	0.01	0.07	0.17	0.38	0.41	3.60
MT	GW	484	0.08	0.56	0.75	1.81	1.25	45.75
MT	SW	47	0.32	0.60	0.89	1.60	1.97	7.82
NC	GW	1735	0.19	2.55	3.26	3.52	4.16	30.17
NC	SW	169	1.22	2.20	2.85	3.01	3.60	6.50

Table 5-1a (continued)
Distributions of System Means for Community Water Systems

State	Source Type	Systems (N)	Minimum	25th Percentile	Median	Mean	75th Percentile	Maximum
ND	GW	197	0.03	0.57	1.65	4.85	4.20	51.40
ND	SW	19	0.20	0.70	1.00	1.11	1.40	2.40
NH	GW	504	0.44	2.31	3.47	6.06	4.76	107.90
NH	SW	37	1.89	2.95	3.63	3.85	4.28	10.00
NJ	GW	438	0.03	0.23	0.49	0.92	1.03	14.00
NJ	SW	29	0.37	0.67	1.01	1.13	1.30	2.81
NM	GW	559	0.06	0.67	1.75	3.81	4.37	57.72
NM	SW	29	0.18	0.38	0.60	1.03	1.00	4.32
NV	GW	221	0.17	2.06	5.00	11.86	13.00	150.00
NV	SW	31	0.07	0.46	1.44	3.38	3.80	39.00
OH	GW	875	0.59	2.33	3.54	4.32	5.21	42.97
OH	SW	139	1.28	2.24	2.75	3.07	3.50	14.50
OK	GW	446	0.11	0.88	1.52	3.01	3.00	78.45
OK	SW	210	0.14	0.53	0.83	1.14	1.26	36.35
OR	GW	316	0.04	0.45	1.09	2.77	2.73	56.00
OR	SW	129	0.22	0.62	1.07	1.45	1.81	10.66
TX	GW	3105	0.10	0.91	1.47	2.58	2.20	86.85
TX	SW	326	0.66	1.27	1.50	1.69	1.85	6.32
UT	GW	263	0.05	0.50	1.00	2.89	2.63	55.49
UT	SW	35	0.18	0.47	0.84	1.94	1.63	20.56

Table 5-1b
Distributions of System Means for Non-Transient Non-Community Water Systems

State	Source Type	Systems (N)	Minimum	25th Percentile	Median	Mean	75th Percentile	Maximum
AK	GW	131	0.02	0.32	0.87	5.39	7.00	54.00
AK	SW	24	0.36	0.65	0.85	1.13	1.64	2.75
AL	GW	31	0.50	0.50	0.50	0.50	0.50	0.63
AZ	GW	77	0.19	1.18	3.00	6.73	7.00	53.92
AZ	SW	2	2.50	2.50	4.25	4.25	6.00	6.00
CA	GW	330	0.00	0.41	1.50	4.13	4.20	110.00
CA	SW	10	1.00	1.00	1.00	1.41	1.58	2.75
IN	GW	538	0.00	0.00	0.01	0.24	0.07	23.00
KS	GW	62	0.33	0.85	1.53	2.12	2.40	14.10
KS	SW	2	1.03	1.03	1.40	1.40	1.77	1.77
MI	GW	230	0.02	0.33	1.80	4.66	6.00	76.00
MN	GW	630	0.03	0.45	0.90	2.65	2.50	56.00
MN	SW	6	0.50	0.50	0.50	0.70	0.96	1.25
MO	GW	190	0.50	0.50	0.50	0.55	0.50	5.20
MT	GW	2	11.00	11.00	21.83	21.83	32.67	32.67
NC	GW	562	0.12	0.73	1.35	2.02	2.43	39.53
NC	SW	8	1.50	2.00	2.50	2.35	2.50	3.29
ND	GW	20	0.03	0.20	0.89	5.36	4.38	47.60
ND	SW	6	0.20	0.40	0.70	0.68	0.90	1.20
NJ	GW	758	0.01	0.19	0.48	1.41	1.31	22.00
NJ	SW	3	0.50	0.50	0.50	0.80	1.40	1.40
NM	GW	140	0.10	0.72	3.38	5.66	7.50	47.40
NM	SW	8	0.50	0.50	0.56	0.87	1.13	2.05
OR	GW	84	0.10	0.51	1.16	2.25	2.51	20.00
TX	GW	580	0.16	0.95	1.53	2.69	2.28	70.00
TX	SW	27	0.66	1.17	1.58	1.84	2.18	6.84
UT	GW	17	0.11	0.37	1.00	3.38	4.50	16.00

5.2 Stratification by System Size

An important question is whether or not public water system size, based on population served, is a determinant of average system arsenic concentrations. If arsenic concentrations are associated with system size, then it may be appropriate to stratify the data on this basis when developing occurrence estimates. However, if arsenic concentrations are not associated with system size, incorrectly stratifying the data by system size could reduce the accuracy of the arsenic occurrence estimates.

The earlier USEPA occurrence estimate (Wade Miller, 1992) considered this question, but did not stratify based on system size because there were too few detects in the NIRS, CWSS, NOMS, and RWS data that were used for this estimate to stratify on this variable. However, Frey and Edwards (1997) stratified their data into small systems (those serving 1,000 to 10,000 people) and large systems (those serving more than 10,000 people).[12] Thus, additional analyses were conducted to evaluate if arsenic concentrations are associated with system size. The following section describes the analyses that were conducted on arsenic occurrence and system size using the AOED data. The system size categories of interest, based on population served, are:

- 25-100;
- 101-500;
- 501-1,000;
- 1,001-3,300;
- 3,301-10,000;
- 10,001-50,000; and
- 50,001 or greater.

First, we visually compared the distributions of arsenic in ground water and surface water systems in the seven system size categories, for each State, and nationally. Separate box plots were prepared for ground water and surface water and for CWS and NTNCWS systems in each State. For example, Figure 5-1 shows the boxplot for the California CWS ground water systems. Plots for each State in the AOED are included in Appendix B-2.

In each plot, system size categories are plotted on the horizontal axis, and arsenic concentrations are plotted on the log scale on the vertical axis. Each distribution is represented graphically, and is composed of system mean arsenic concentrations. The procedure for calculating system mean arsenic concentrations using regression on order statistics is presented in Section 6.1. The lower and upper ends of each box represent the 25th and 75th percentiles of each distribution. The line near the middle of each box is the median, and the darkened circle is the mean. The "whiskers" above and below each box illustrate the pattern of system means in the tails of the distribution, and are composed of straight lines and plus signs. Lines represent system means from the quartiles to the 5^{th} and 95^{th} percentiles. Points in the upper and lower

[12] Frey and Edwards (1997) used a stratified survey design, of which system size was a stratification variable. They did not specifically evaluate whether or not arsenic distributions are different for systems of different sizes.

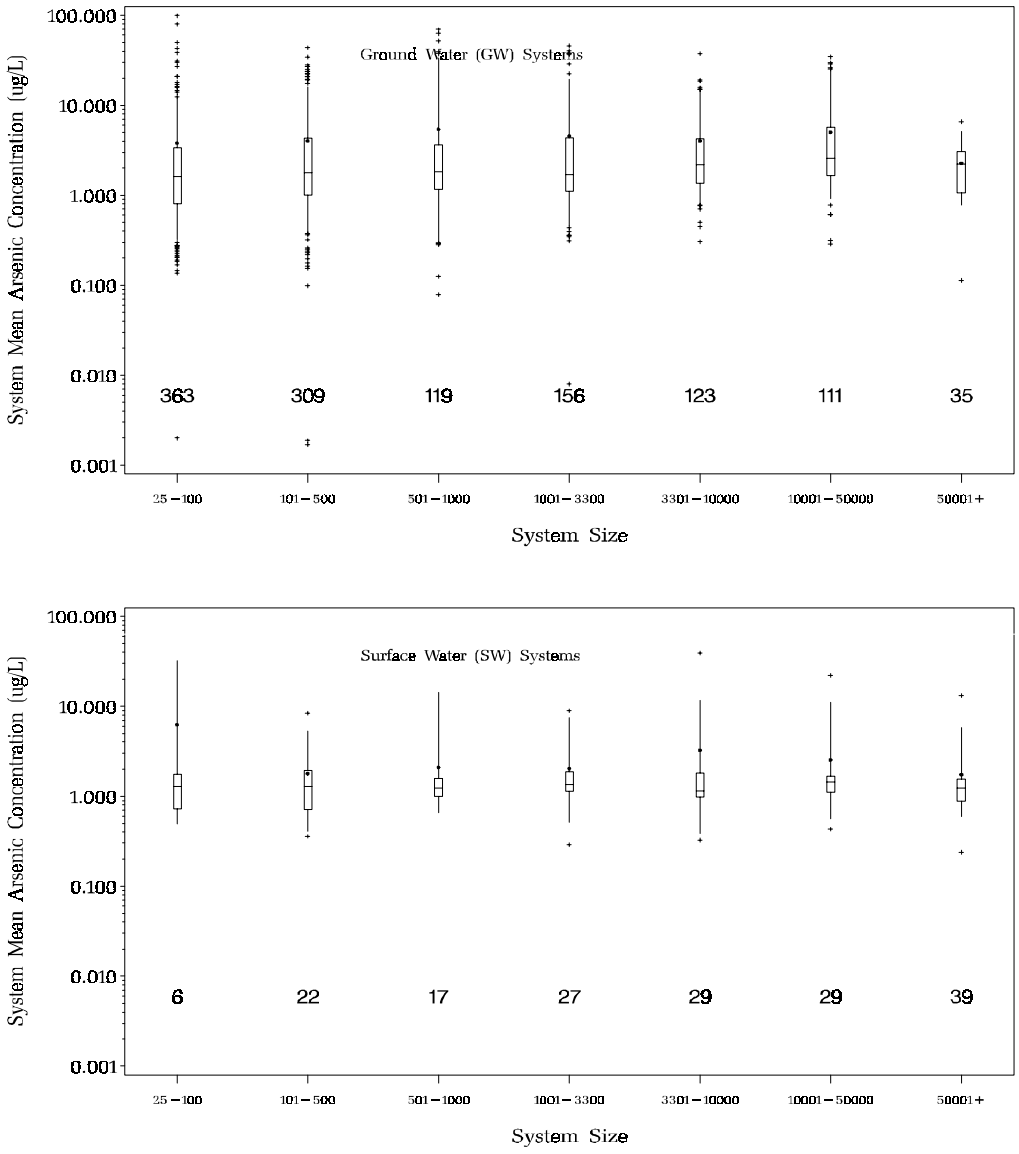


FIGURE 5 – 1. Boxplots of System Means By System Size for
Community Water Systems in the State of California
Number of Systems Indicated Below Boxplot

5 percent tails are shown as plus signs. Below each distribution is the number of systems in the sample.

The box plots provide qualitative comparison of the distributions of mean arsenic levels across the size categories. For most States, these plots demonstrate that the means, medians, and quartile ranges of the distributions are similar across the size categories for both ground water and surface water. This is particularly true when there are relatively large numbers of systems represented in each of the distributions, as is the case in California. The boxplots based on the smaller sample sizes tend to appear jagged and may not be representative of the underlying, (presumably) continuous distribution. Otherwise, because the distributions are similar between size categories, this qualitative analysis suggests that system mean arsenic concentrations are probably not associated with system size in either ground water or surface water systems.

In addition to the qualitative analysis of arsenic distributions for systems in different size strata, we applied an analysis of variance (ANOVA) procedure to quantitatively test if arsenic concentration is dependent on system size. For this analysis, the mean of the natural logarithm of the system mean arsenic ("log-mean") was calculated for each state and size stratum. For each of the seven NAOS regions, and for the entire US, these state and size stratum log-means were averaged across the region, weighting the mean for each state by the number of systems of the given size stratum in that state (as defined by the 1998 SDWIS baseline, including purchased and non-purchased systems). These regional mean values of the logarithm of the system mean arsenic are presented in Tables 5-2a (ground water CWS systems), 5-2b (surface water CWS systems), 5-3a (ground water NTNCWS systems), and 5-3b (surface water NTNCWS systems). The means are presented in the top half of each table, and the relative rankings for the size strata are presented in the bottom half of each table. ANOVA was used to test the statistical significance of the differences among the size strata weighted log-means in each Region. For each Region, p values < 0.05 imply that the means of the size strata are statistically significantly different (at the five percent significance level). For some combinations of type of water system, source type, NAOS region, and size stratum, no data were available, so the corresponding log-mean and stratum rank is missing. This occurs in only two cases for CWS systems but quite frequently for the NTNCWS systems, which tend to serve fewer people.

These results suggest that mean arsenic concentrations may differ significantly from stratum to stratum. For ground water CWS systems, these differences were statistically significant in five of the seven regions. For surface water CWS systems, these differences were statistically significant in three regions. For ground water NTNCWS systems, these differences were statistically significant in only two of the seven regions, which is partly attributable to the relative lack of NTNCWS data. For surface water NTNCWS systems, these differences were statistically significant in two of the five regions which had sufficient data for this calculation. However, the stratum ranks in these tables show that mean arsenic concentrations do not vary in a consistent pattern from region to region. For example, for ground water CWS systems, arsenic concentrations in Region 1 and 3 appear to generally decrease as system size increases, while in Regions 6 and 7, arsenic concentrations appear to generally increase as system size increases (except for the large system 50000 + stratum). In the three remaining regions, no systematic

Table 5-2a
Log-Means by System Size Category, CWS Ground Water Systems

	LOG-MEANS FOR EACH NAOS REGION							
SIZE STRATUM	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**All**
25-100	0.6312	1.1058	-0.2185	-0.0362	-0.1205	0.0004	0.6627	0.4415
101-500	0.5592	1.1150	-0.2554	0.0472	0.1269	0.0575	0.6715	0.3968
501-1000	-0.0728	1.2639	-0.4796	-0.1937	0.1221	0.3096	0.8684	0.2065
1001-3300	-0.0071	1.1376	-0.3914	-0.2791	0.2012	0.3260	0.8258	0.1400
3301-10000	-0.4255	1.2612	-0.4294	-0.4201	0.3680	0.6417	0.9186	0.1238
10001-50000	-0.8486	1.1723	-0.4572	-0.3264	0.5226	0.0788	1.0712	0.1923
50000 +	-0.9796	1.1690	-0.6332	0.3083	0.0849	-0.2877	0.5801	0.3626
p - Value	0	0	0.32	0	0	0.13	0	0
	STRATUM RANK FOR EACH NAOS REGION							
25-100	1	7	1	3	7	6	6	1
101-500	2	6	2	2	4	5	5	2
501-1000	4	1	6	4	5	3	3	4
1001-3300	3	5	3	5	3	2	4	6
3301-10000	5	2	4	7	2	1	2	7
10001-50000	6	3	5	6	1	4	1	5
50000 +	7	4	7	1	6	7	7	3

AOED States in NAOS Regions

Region 1. ME, NH, NJ
Region 2. KY, NE
Region 3 AL
Region 4 IL, IN, OH, MI, MN
Region 5 AR, KS, MO, NM, OK, TX
Region 6 MT, ND
Region 7 AK, AZ, CA, NV, or UT

Table 5-2b
Log-Means by System Size Category, CWS Surface Water Systems

	LOG-MEANS FOR EACH NAOS REGION							
SIZE STRATUM	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**All**
25-100	--	0.9413	-0.2439	0.3329	-0.0548	0.4409	0.3571	0.2956
101-500	0.2700	0.8689	-0.6031	0.2874	-0.2778	0.4457	0.2616	0.0887
501-1000	-0.1448	0.9593	--	-0.1610	-0.2545	-0.1719	0.1206	-0.0099
1001-3300	0.2069	0.7680	-0.3303	-0.2345	-0.1993	0.0425	0.2116	0.0338
3301-10000	0.0319	0.6825	-0.4577	-0.0706	-0.2819	0.0447	0.3764	0.1074
10001-50000	0.1886	0.7642	-0.3347	-0.3018	-0.0711	-0.2539	0.3384	0.0859
50000 +	0.1527	0.9756	-0.4317	-0.1112	0.4329	0.6247	0.3544	0.2386
p - Value	0.89	0	0.22	0	0	0.2	0.07	0
	STRATUM RANK FOR EACH NAOS REGION							
25-100	--	3	1	1	2	3	2	1
101-500	1	4	6	2	6	2	5	4
501-1000	6	2	--	5	5	6	7	7
1001-3300	2	5	2	6	4	5	6	6
3301-10000	5	7	5	3	7	4	1	3
10001-50000	3	6	3	7	3	7	4	5
50000 +	4	1	4	4	1	1	3	2

AOED States in NAOS Regions

Region 1. ME, NH, NJ
Region 2. KY, NE
Region 3 AL
Region 4 IL, IN, OH, MI, MN
Region 5 AR, KS, MO, NM, OK, TX
Region 6 MT, ND
Region 7 AK, AZ, CA, NV, or UT

Table 5-3a
Log-Means by System Size Category, NTNCWS Ground Water Systems

	LOG-MEANS FOR EACH NAOS REGION							
SIZE STRATUM	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**All**
25-100	-0.6399	0.2806	-0.6931	-0.5861	0.1180	2.0479	0.2561	-0.2214
101-500	-0.7615	0.2725	-0.6931	-0.9659	-0.0706	3.2812	0.2362	-0.2455
501-1000	-0.6124	0.2675	-0.6684	-0.9927	0.0702	--	0.6937	-0.2684
1001-3300	-0.7024	0.6338	-0.6931	-0.2996	0.1058	--	0.2290	-0.0594
3301-10000	-0.7569	0.5596	--	0.2168	-0.6931	--	0.4953	0.0956
10001-50000	--	--	--	--	--	--	1.0669	1.0669
50000 +	--	--	--	--	--	--	--	--
p - Value	0.87	0.46	0.51	0.03	0.51	0.65	0.03	0.31
	STRATUM RANK FOR EACH NAOS REGION							
25-100	2	3	3	3	1	2	4	4
101-500	5	4	3	4	4	1	5	5
501-1000	1	5	1	5	3	--	2	6
1001-3300	3	1	3	2	2	--	6	3
3301-10000	4	2	--	1	5	--	3	2
10001-50000	--	--	--	--	--	--	1	1
50000 +	--	--	--	--	--	--	--	–

AOED States in NAOS Regions

Region 1. ME, NH, NJ
Region 2. KY, NE
Region 3 AL
Region 4 IL, IN, OH, MI, MN
Region 5 AR, KS, MO, NM, OK, TX
Region 6 MT, ND
Region 7 AK, AZ, CA, NV, or UT

Table 5-3b
Log-Means by System Size Category, NTNCWS Surface Water Systems

	LOG-MEANS FOR EACH NAOS REGION							
SIZE STRATUM	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**All**
25-100	--	0.6609	--	-0.6931	-0.4377	-0.2798	0.4496	0.3039
101-500	-0.6931	0.9163	--	0.2231	-0.2430	-0.7811	0.2584	0.1540
501-1000	-0.6931	0.7975	--	-0.6931	0.7178	--	0.0000	0.1030
1001-3300	0.3365	0.9163	--	-0.3670	--	--	--	0.5230
3301-10000	--	--	--	--	--	--	--	--
10001-50000	--	--	--	--	--	--	--	--
50000 +	--	--	--	--	--	--	--	--
p - Value	--	0.84	--	0	0	0.4	0.32	0.2
	STRATUM RANK FOR EACH NAOS REGION							
25-100	--	4	--	3.5	3	1	1	2
101-500	2.5	1.5	--	1	2	2	2	3
501-1000	2.5	3	--	3.5	1	--	3	4
1001-3300	1	1.5	--	2	--	--	--	1
3301-10000	--	--	--	--	--	--	--	--
10001-50000	--	--	--	--	--	--	--	--
50000 +	--	--	--	--	--	--	--	–

AOED States in NAOS Regions

Region 1. ME, NH, NJ
Region 2. KY, NE
Region 3 AL
Region 4 IL, IN, OH, MI, MN
Region 5 AR, KS, MO, NM, OK, TX
Region 6 MT, ND
Region 7 AK, AZ, CA, NV, or UT

patterns are evident. Similarly, no systematic patterns are evident for surface water CWS systems in any region or for the NTNCWS systems in any region.

The ANOVA methods do not take into account the potential for different amounts of uncertainty in the log means, which are attributable to different sample sizes with different censoring levels and rates. Furthermore, the results for cases with relatively large numbers of samples often showed statistically significant differences that were not numerically very large. Thus, the ANOVA results may not clearly demonstrate that arsenic mean levels are associated with stratum size. An additional consideration is that for the occurrence analysis to take into account both the arsenic variation by system size within each region, and the variation by States within each region, would require stratification by State and system size. Such a detailed stratification would lead to several State and system size combinations being represented by small data subsets.

5.3 Stratification by System Type

Another potential stratification variable is water system type. Systems considered for this analysis could be either CWS or NTNCWS. CWS are public water systems that serve at least 15 service connections used by year-round residents or regularly serve at least 25 year-round residents.[13] NTNCWS are public water systems that are not CWS and that regularly serve at least 25 of the same persons more than 6 months of the year.[14] The majority of NTNCWS systems serve less than 3,300 people. The AOED database contains data for NTNCWS systems in 17 States, although only two systems are included for the State of Montana. Basic statistics were calculated using the system means of the ground water CWS and NTNCWS systems in each of these States, and these statistics are presented in Tables 5-4a (ground water systems) and 5-4b (surface water systems). For each State and source type, the means and standard deviations for the CWS and NTNCWS systems were computed and compared using a standard t test to compare the means and an F test (of the variances) to compare the standard deviations; significant differences at the five percent level are indicated by asterisks. (The Smith-Satterthwaite version of the t test was used instead of the usual pooled variance t test if the F test showed that the standard deviations were significantly different).

These data indicate that for ground water, arsenic distributions in NTNCWS are often quite similar to arsenic distributions in CWS. In general, the means and the level of censoring for CWS in a particular State are very close to the levels observed in NTNCWS in that State. In some States, mean levels are slightly higher in CWS systems, whereas in others, mean levels are slightly higher in NTNCWS systems. Differences in the means were statistically significant (at the five percent level) in 6 of the 17 states for ground water systems. The standard deviations tended to be statistically significantly different, although the numerical differences were not very large and there was no consistent pattern for whether the CWS or NTNCWS systems had the higher standard deviation. For surface water, the means for the NTNCWS systems were

[13] 40 CFR Section 141.2.

[14] *Ibid*.

Table 5-4a
Arsenic Occurrence in Ground Water CWS and NTNCWS Systems

State [1]	System Type	Number of Systems	Mean[3]	Std. Dev.[3]	Censoring (%) [2]
AK	CWS	304	4.11	7.67 *	51.3
AK	NTNCWS	131	5.39	9.34 *	52.7
AL	CWS	263	0.73 *	0.49 *	91.3
AL	NTNCWS	31	0.50 *	0.022 *	96.8
AZ	CWS	279	9.53 *	13.8 *	40.5
AZ	NTNCWS	77	6.73 *	9.94 *	57.1
CA	CWS	1224	4.2	7.57 *	43.3
CA	NTNCWS	330	4.13	8.58 *	52.4
IN	CWS	648	0.26	0.66 *	97.5
IN	NTNCWS	538	0.24	1.34 *	97.2
KS	CWS	506	2.64	3.83 *	13.6
KS	NTNCWS	62	2.12	2.34 *	19.4
MI	CWS	644	5.31	9	23.8
MI	NTNCWS	230	4.66	8.27	23.5
MN	CWS	829	2.77	5.52	44.8
MN	NTNCWS	630	2.65	5.18	51.7
MO	CWS	773	0.77 *	1.93 *	90.6
MO	NTNCWS	190	0.55 *	0.48 *	98.4
MT	CWS	484	1.81	3.53 *	49.2
MT	NTNCWS	2	21.8	15.3 *	0
NC	CWS	1735	3.52 *	1.63 *	80.1
NC	NTNCWS	562	2.02 *	2.82 *	92.2
ND	CWS	197	4.85	8.13 *	22.3
ND	NTNCWS	20	5.36	11.4 *	35
NJ	CWS	438	0.92 *	1.38 *	88.4
NJ	NTNCWS	758	1.41 *	2.66 *	86.4
NM	CWS	559	3.81 *	6.00 *	25.9
NM	NTNCWS	140	5.66 *	7.15 *	27.9
OR	CWS	316	2.77	5.49 *	75
OR	NTNCWS	84	2.25	3.17 *	79.8
TX	CWS	3105	2.58	4.26 *	67
TX	NTNCWS	580	2.69	5.49 *	63.8
UT	CWS	263	2.89	5.59	31.9
UT	NTNCWS	17	3.38	4.84	41.2

Notes:
[1]States without ground water NTNCWS systems (AR, IL, KY, ME, NH, NV, OH, OK) are not listed in Table 5-4a.
[2]Percent censoring is defined as the percentage of systems with all datapoints censored. Under this definition, a system with 10 samples, including one detect and 9 non-detects, would not be considered censored.
[3]Statistically significant differences at the five percent significance level are indicated by asterisks.

Table 5-4b
Arsenic Occurrence in Surface Water CWS and NTNCWS Systems

State [1]	System Type	Number of Systems	Mean[3]	Std. Dev.[3]	Censoring (%) [2]
AK	CWS	321	1.29	2.08 *	65.4
AK	NTNCWS	74	1.25	0.67 *	59.5
AZ	CWS	112	5.72	7.05 *	35.7
AZ	NTNCWS	8	3.85	1.86 *	62.5
CA	CWS	574	2.75	4.97 *	61.3
CA	NTNCWS	31	2.14	1.48 *	61.3
KS	CWS	202	1.11	0.56	10.9
KS	NTNCWS	4	1.4	0.42	0
MN	CWS	69	0.88 *	0.19 *	56.5
MN	NTNCWS	18	0.69 *	0.28 *	66.7
NC	CWS	338	3.01 *	1.10 *	74
NC	NTNCWS	16	2.35 *	0.57 *	87.5
ND	CWS	38	1.11 *	0.53	5.3
ND	NTNCWS	12	0.68 *	0.34	0
NJ	CWS	58	1.13	0.62	75.9
NJ	NTNCWS	6	0.8	0.46	66.7
NM	CWS	87	1.35 *	1.08 *	25.3
NM	NTNCWS	24	0.90 *	0.58 *	50
TX	CWS	652	1.69	0.74 *	57.1
TX	NTNCWS	54	1.84	1.17 *	63
UT	CWS	108	2.1	3.3	27.8
UT	NTNCWS	5	1.72	1.07	80

Notes:
[1]States without surface water NTNCWS systems (AL, AR, IL, IN, KY, ME, MI, MO, MT, NH, NV, OH, OK, OR) are not listed in Table 5-4b.
[2]Percent censoring is defined as the percentage of systems with all datapoints censored. Under this definition, a system with 10 samples, including one detect and 9 non-detects, would not be considered censored.
[3]Statistically significant differences at the five percent significance level are indicated by asterisks.

consistently lower than the means for the CWS systems (Kansas was the only exception), but these differences were statistically significant (at the five percent level) in only 4 of the 11 States.

A graphical comparison between the CWS and NTNCWS arsenic distributions by state and water system type is discussed and presented in Section 5.5 below.

For ground water, because the CWS and NTNCWS arsenic distributions were sometimes significantly different, and because there was sufficient CWS and NTNCWS data to separately estimate the arsenic occurrence for each water system type, the arsenic occurrence estimates derived in chapter 6 were computed separately for CWS and NTNCWS systems. For surface water systems, although some significant differences were found, it was determined that there was insufficient data to recommend developing a separate arsenic occurrence estimate for NTNCWS surface water systems: Only 11 States have NTNCWS surface water systems in the AOED, and for four of those States (Arizona, Kansas, New Jersey, and Utah), there were less than 10 surface water NTNCWS systems. Therefore, for surface water systems, the arsenic occurrence estimates were based only on the CWS system data in the AOED, and the exceedance distributions for the CWS systems were used to estimate the exceedance distributions for the NTNCWS systems.

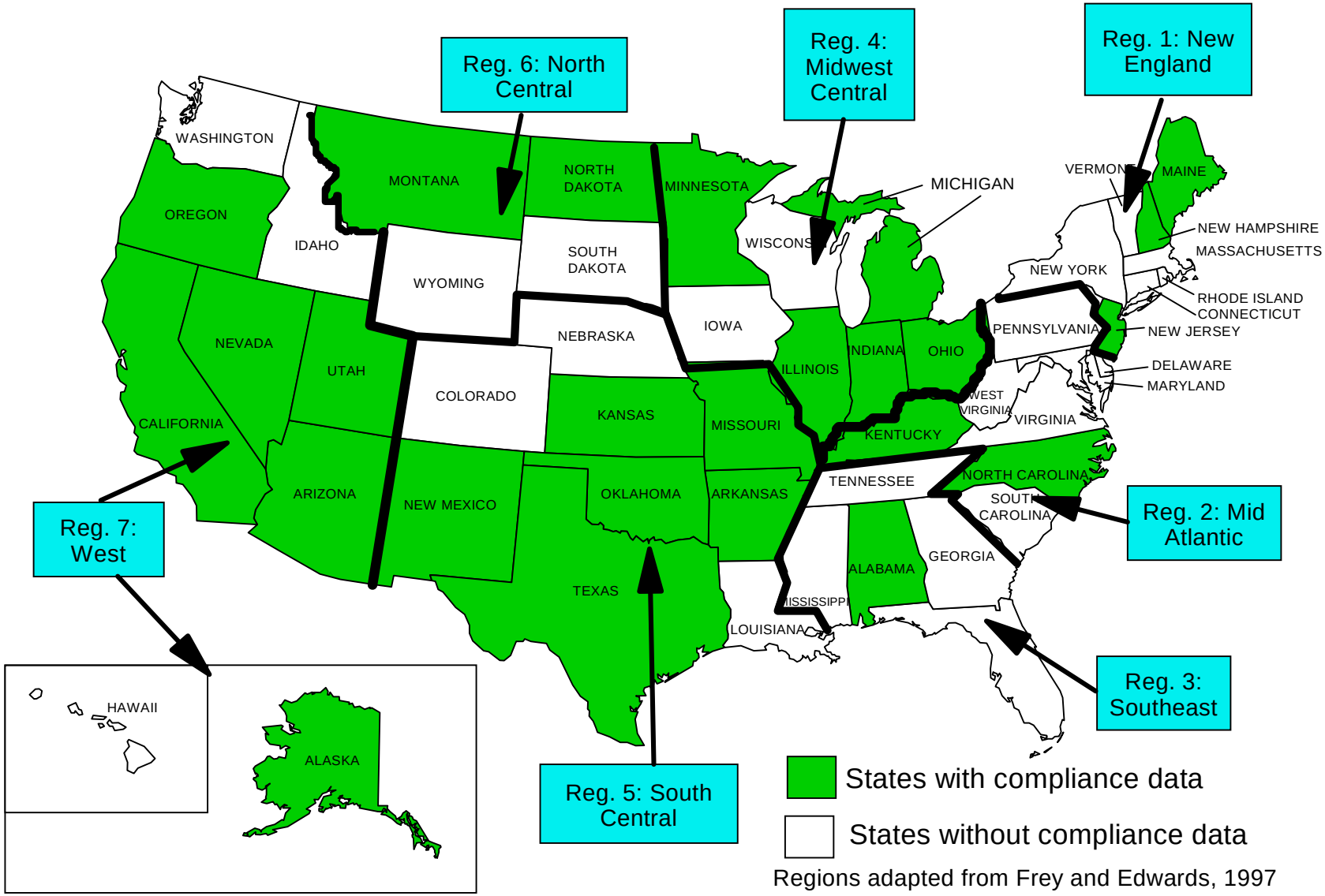
5.4 Regional Stratification

Natural arsenic sources, such as soil and rock, may be significant sources of arsenic in drinking water. Therefore, regional differences in geology, hydrology, and hydrogeology may result in significant differences in arsenic occurrence from region to region. Frey and Edwards (1997) studied regional differences in arsenic occurrence using USGS WATSTORE data and the Metro database. As discussed in Chapter 4, they calculated arsenic natural occurrence factors (NOFs), and identified seven different regions that appeared to have distinct arsenic occurrence characteristics based on these NOFs. USGS used these regions in its estimation of arsenic occurrence, and presented comparisons of its findings with Frey and Edwards' findings.

The NAOS regions are displayed in Figure 5-2. The States in each region for which compliance data are available are shaded. This figure illustrates that there are significant differences in the coverage that AOED provides for these regions. While almost every state is represented in the Western Region, South Central Region, and the Midwest Central Region, fewer States are covered in the Southeast Region, the Mid-Atlantic Region, and the New England Region. Each of these Regions is represented by no more than three states, and the Southern Region is only represented by Alabama. Frey and Edwards (1997) found that arsenic levels were generally higher in the Western and Southern Central Regions, and were lower in the Southeast, the Mid-Atlantic, and the New England Regions.

Where there are regional differences in contaminant occurrence patterns, and differences in the level of coverage among those regions, regional stratification can improve the accuracy of the national estimates. The NAOS Regions, which are delineated according to political boundaries rather than physiographic provinces, are convenient for use with the AOED database, which can easily be used to analyze arsenic occurrence on a State or regional basis. Use of the NAOS Regions is also convenient because it facilitates comparisons of arsenic occurrence

Figure 5-2: States with Arsenic Compliance Data in the 7 Regions



Regions adapted from Frey and Edwards, 1997

presented in this report with arsenic occurrence levels that were reported by Frey and Edwards (1997) and by USGS (Focazio *et al.*, 2000). Therefore, the regional stratification scheme of the NAOS survey was applied to AOED during the development of the arsenic occurrence estimates that are presented in Chapter 6.

The following States in the AOED database represent each of the NAOS Regions:

- New England: Maine, New Hampshire, and New Jersey;
- Mid-Atlantic: Kentucky and North Carolina;
- Southeast: Alabama;
- Midwest: Illinois, Indiana, Ohio, Michigan, and Minnesota;
- South Central: Arkansas, Kansas, Missouri, New Mexico, Oklahoma, and Texas;
- North Central: Montana and North Dakota; and
- Western: Alaska, Arizona, California, Nevada, Oregon, and Utah.

An analysis was conducted to evaluate potential bias introduced into the occurrence estimates (presented in Chapter 6) by the use of the 25 States data from AOED to represent these regions. This analysis was based on ground water data from the USGS database, and was designed to semi-quantitatively assess the extent to which an individual State appears to reflect arsenic occurrence in other States in the region, and in the region as a whole.

In this analysis, percent exceedances were estimated at concentrations of 2, 5, 10, 20 and 50 µg/L for each region, using USGS data for the States which are represented in AOED in that particular region. Two databases developed by the USGS were used to support this analysis. The first database was derived from the USGS Arsenic Database of Selected Counties (see Section 4.2.2). From this database, exceedance probabilities were estimated by calculating the percentage of data points in each county exceeding specific arsenic concentrations.

USGS created the second database from information contained in SDWIS, and this database provides, for each PWS in the United States, State and county Federal Information Processing Standard (FIPS) codes[15] correlated to the PWS location. The products of the county exceedance probabilities contained in the first database and the number of PWS in each county, determined from the second database, are the numbers of PWS in each county that may exceed specific arsenic concentrations. Summing across counties yields the total number of systems in the State, and the number of systems that are likely to exceed specific arsenic concentrations. The occurrence estimates were weighted by the number of systems in each State, based on SDWIS.

Two sets of regional exceedance probabilities were estimated from the USGS State exceedance probabilities. The first regional exceedance probability was based on the USGS data for the State in the region represented in AOED. The second regional exceedance probability was based on USGS data for all of the States in the Region. For example, in the New England Region, the first estimate was based on USGS data for the States of Maine, New Hampshire, and

[15] FIPS codes, which are unique identifiers for each State (two letter abbreviations or two digit numbers) and county (five digit identification number) in the United States, are assigned by the U.S. Postal Service.

New Jersey. Then, the second exceedance estimates were based on the USGS data for all of the States in the New England Region.

As a result, two sets of occurrence estimates were developed for ground water in each region based on USGS data, and these two regional estimates were compared. Table 5-5 presents these regional comparisons, together with national estimates that were developed as the weighted sum of the regional estimates. We hoped that analyzing these data would provide information about the use of some States to represent a region. However, it should be noted that the USGS data is qualitatively different from the AOED data. Therefore, the power of the USGS data to predict the potential accuracy of the AOED regional arsenic occurrence estimates may be limited. The USGS data may have different spatial coverage from the AOED data. For example, the USGS data may contain data from investigations that focused on specific areas in some States. Some States may have data from a small number of counties. In addition, the USGS data are not finished water samples collected from PWS facilities.

Table 5-5
Comparison of Regional Ground Water Arsenic Occurrence Estimates Based on USGS Data
All States in Region vs. States Represented in AOED

Region	Data Set	> 2 µg/L	> 5 µg/L	> 10 µg/L	> 20 µg/L	> 50 µg/L
1	All States	13.97%	6.83%	3.41%	0.63%	0.30%
	AOED States	12.61%	6.83%	1.96%	0.43%	0.00%
2	All States	15.99%	5.02%	2.18%	0.62%	0.04%
	AOED States	8.47%	2.98%	0.96%	0.16%	0.00%
3	All States	7.78%	2.82%	1.53%	0.89%	0.17%
	AOED States	10.96%	0.00%	0.00%	0.00%	0.00%
4	All States	26.20%	14.09%	7.37%	2.49%	0.48%
	AOED States	29.11%	17.36%	9.58%	3.38%	0.67%
5	All States	19.89%	9.56%	4.78%	1.74%	0.35%
	AOED States	17.83%	8.44%	4.55%	1.73%	0.21%
6	All States	47.12%	32.00%	25.12%	16.37%	8.74%
	AOED States	57.98%	43.08%	35.90%	24.75%	13.41%
7	All States	41.78%	25.25%	15.41%	6.87%	2.22%
	AOED States	47.33%	29.64%	19.05%	9.03%	3.04%
National	All States	23.32%	12.25%	6.97%	2.92%	0.92%
	AOED States	23.88%	13.15%	7.82%	3.56%	1.17%

For Regions 1, 4, and 5, the USGS data indicate that the States which are represented in AOED may be reasonably representative of arsenic occurrence in the entire region. In Region 2, these data suggest that the States which are contained in AOED may have slightly lower arsenic occurrence concentrations than regional average concentrations when all States are considered. The States in Regions 6 and 7 that are included in AOED may overestimate regional average arsenic occurrence concentrations. In Region 3, data are inconsistent, and suggest that Alabama data will overestimate regional arsenic occurrence at 2 µg/L, and will underestimate regional arsenic occurrence at concentrations at or above 5 µg/L.

National arsenic occurrence estimates, based on all of the States for which USGS contains data (USGS does not include any samples from Vermont or Hawaii, but does represent the remaining 48 States) are quite similar to the national arsenic occurrence estimates, based on the USGS data in the 25 States that are represented in AOED. Both of these estimates were derived from the respective regional estimates, weighted by the number of ground water systems in each of the regions. At concentrations of 2 to 50 µg/L, the estimates based on USGS data for the 25 States that are represented in AOED are higher than those based on all of the States in USGS; however, at each concentration of interest, the occurrence estimates differ by less than one percent. The greatest differences occur at concentrations of 5 and 10 µg/L, where the estimates based on the data for the 25 States in AOED are 0.90 and 0.84 percent higher, respectively, than those based on data for all of the States. These results suggest that the estimates based on the AOED data, that are presented in Chapter 6 of this report, are probably not significantly biased at the national level by the lack of complete representation within each region. The potential overestimation based on data for some regions is balanced by the potential underestimation of data for other regions. The USGS data also suggest that the data for the 25 States in AOED may slightly, although not significantly, overestimate national occurrence of arsenic in drinking water. Therefore, this semi-quantitative analysis of the USGS data suggests that the arsenic occurrence estimates presented in Chapter 6 may be slightly conservative.

5.5 Arsenic Distributions at the State Level

We further investigated the nature of arsenic occurrence in ground water and surface water in the individual States through graphical analyses of the distributions of system means. Figures 5-3a, 5-3b, 5-3c, and 5-3d present boxplots of the distributions of system means for ground water and surface water systems in each of the 25 states with compliance monitoring data. The procedure for calculating system mean arsenic concentrations is presented in Section 6.1, and the summary statistics that are displayed by the symbols in each boxplot distribution are described in Section 5.2 of this report. The number of systems represented in each distribution is noted below each boxplot. Figures 5-3a and 5-3b show the distributions of all system means for ground water and surface water, respectively. Figures 5-3c and 5-3d show the distributions of the system means that are not completely censored for ground water and surface water, respectively. As discussed in chapter 6, a system mean is completely censored if all the arsenic measurements for that system were censored, i.e., non-detects. In this case, the system mean was estimated by applying the regression on order statistics method to all the system means for that State. The states have been ordered by NAOS region, and then alphabetically within each NAOS region (shown as the number after each state abbreviation).

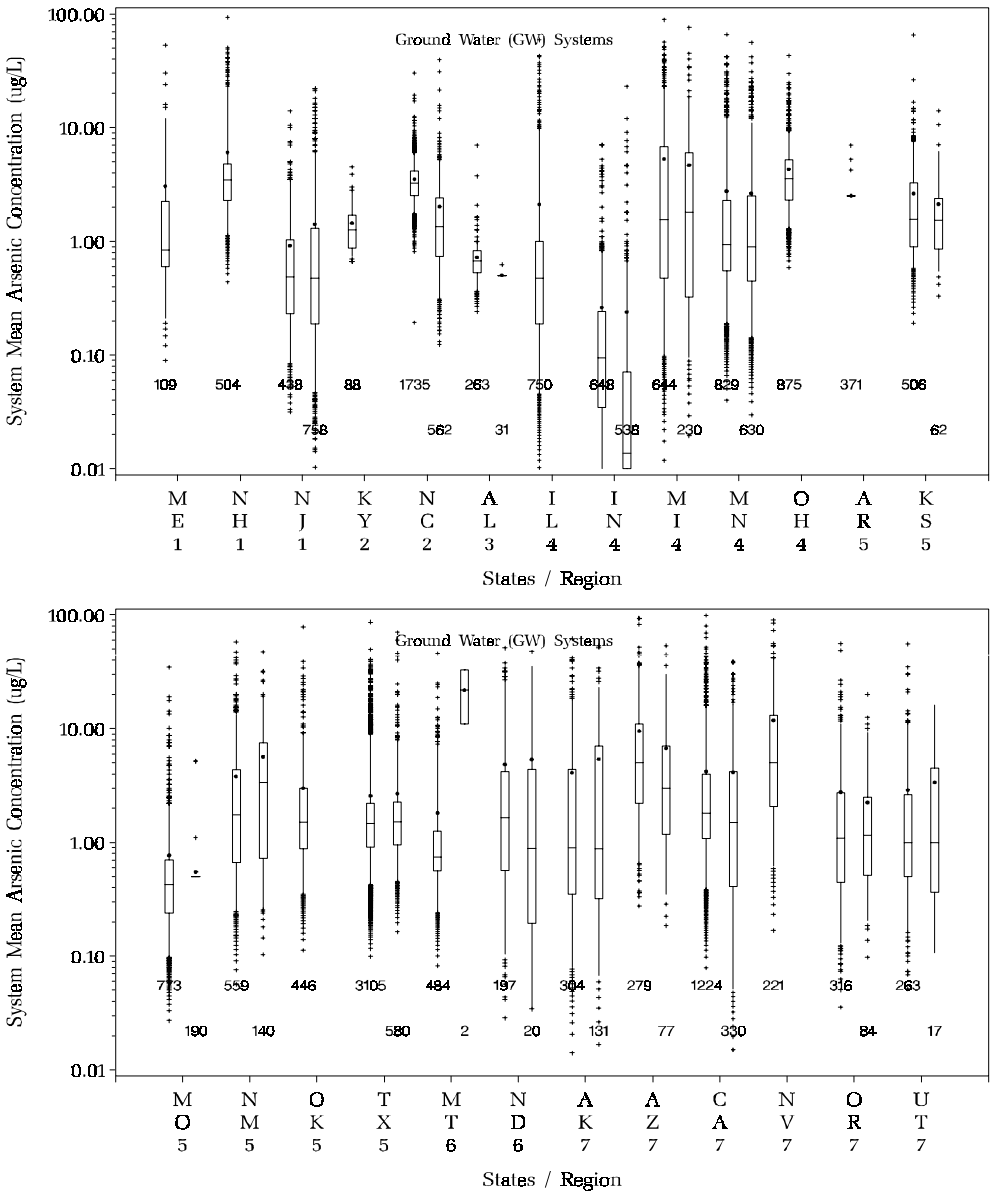


FIGURE 5−3a. Boxplots of System Means for
GW CWS and NTNCWS by State
Number of Systems Indicated Below Boxplot (CWS on left)
All Systems

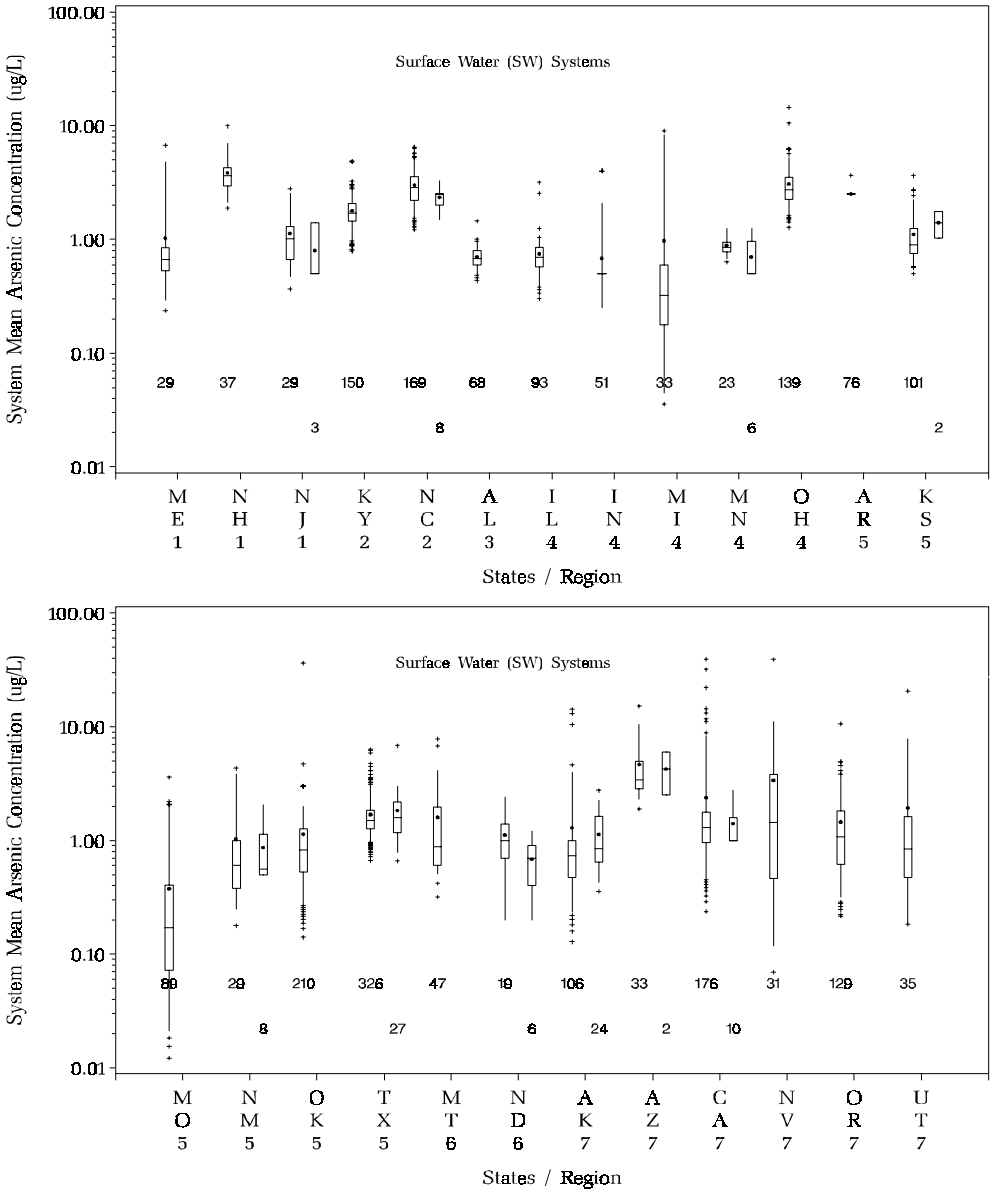


FIGURE 5−3b. Boxplots of System Means for
SW CWS and NTNCWS by State
Number of Systems Indicated Below Boxplot (CWS on left)
All Systems

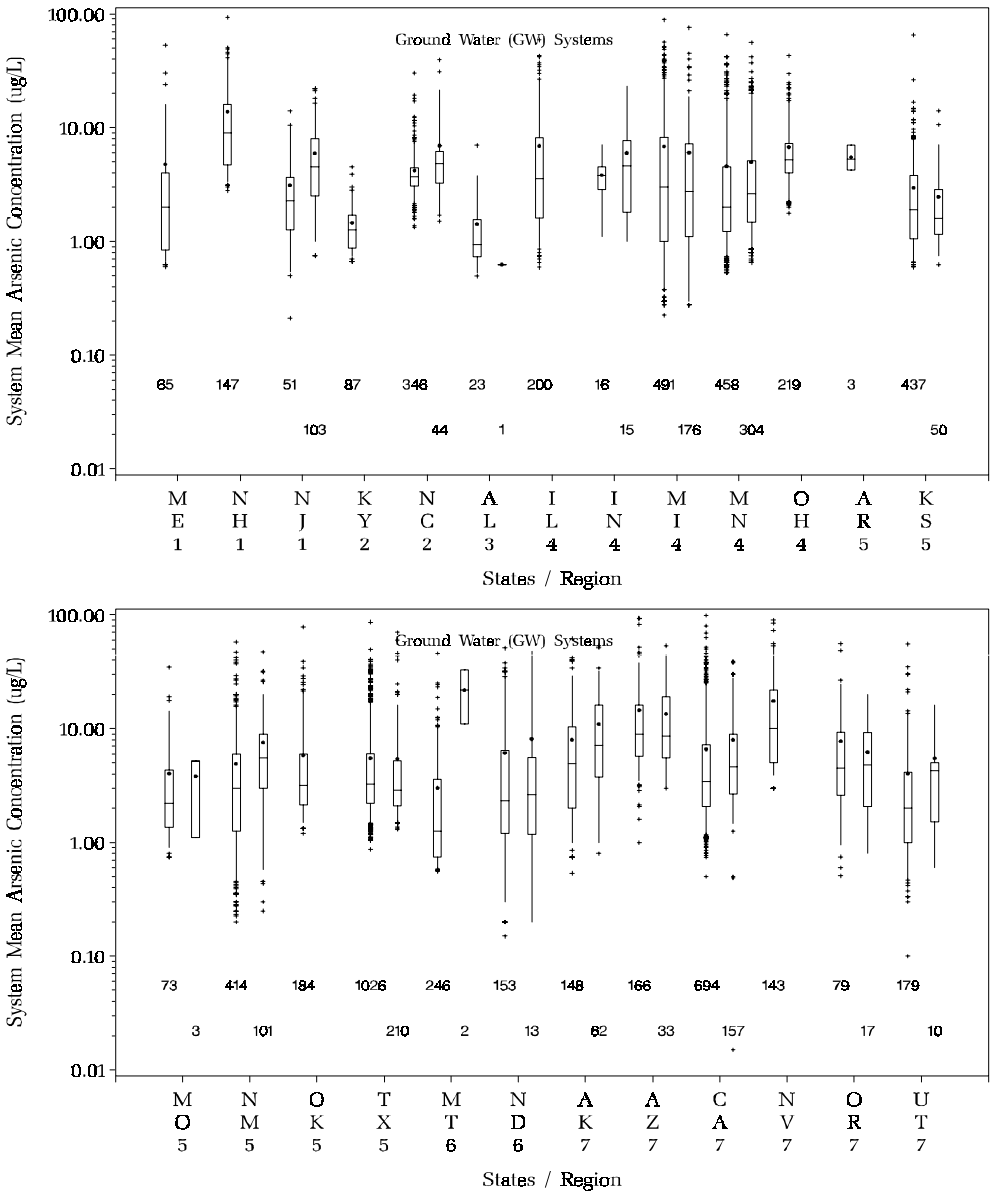


FIGURE 5−3c. Boxplots of System Means for
GW CWS and NTNCWS by State
Number of Systems Indicated Below Boxplot (CWS on left)
Systems that are Not Completely Censored

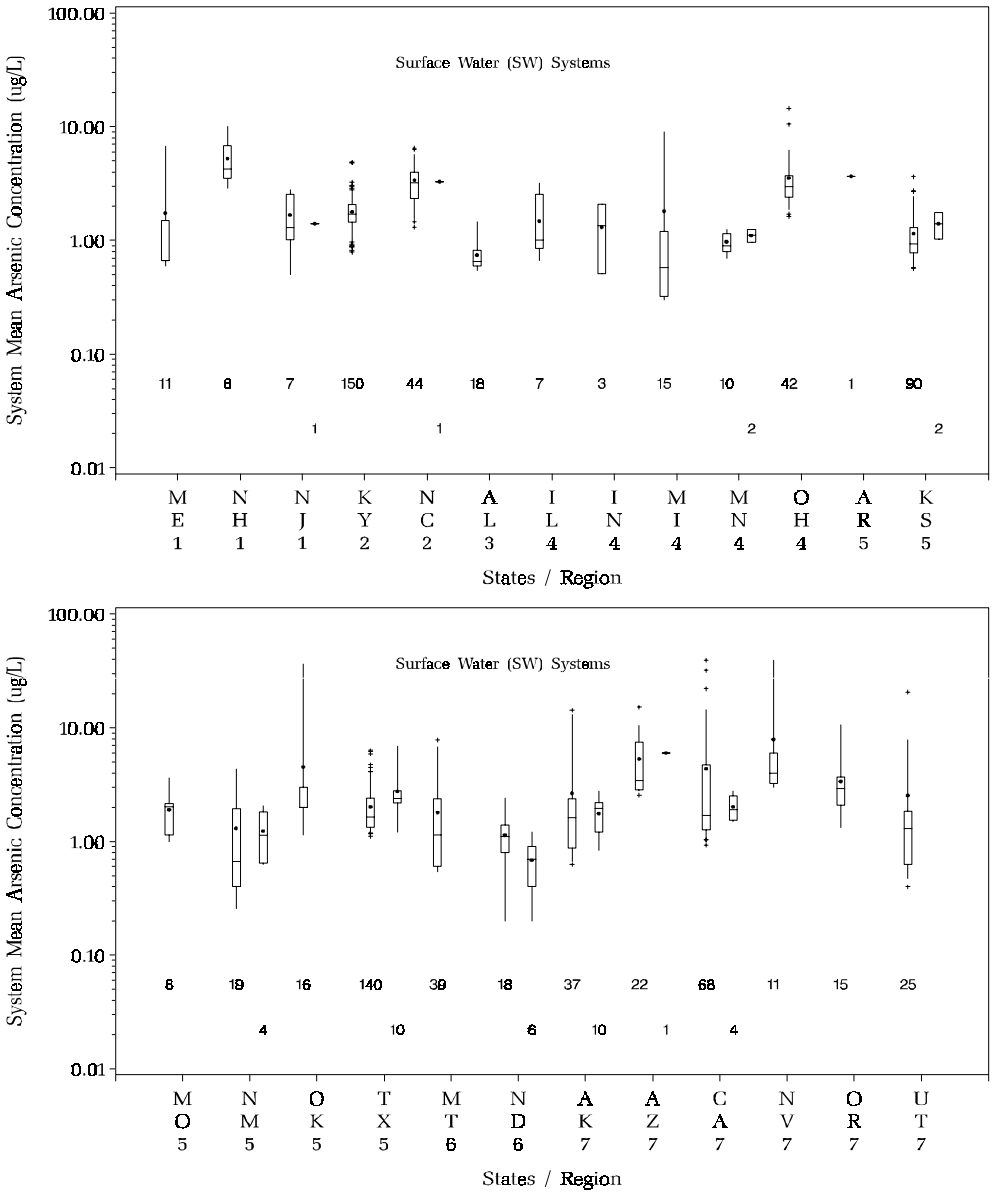


FIGURE 5−3d. Boxplots of System Means for
SW CWS and NTNCWS by State
Number of Systems Indicated Below Boxplot (CWS on left)
Systems that are Not Completely Censored

These boxplots suggest that arsenic concentrations in drinking water may vary significantly from State to State. Among ground water systems, the State of Indiana has the lowest arsenic concentration levels, both for CWS and NTNCWS systems, and the States of Arizona and Nevada appear to have the highest concentrations. Comparing Figures 5-3a and 5-3c for Indiana shows that the estimated system means for that state are very sensitive to the conventions used to estimate system means for completely censored systems; the distribution of uncensored system means for Indiana is comparable to the distributions of uncensored system means for the other states in the same NAOS region: Illinois, Michigan, Minnesota, and Ohio. The side by side graphical comparison of the CWS and NTNCWS arsenic distributions by State for ground water conform the similarity of these distributions noted earlier (in section 5.3).

Among surface water systems, using all estimated system means, Figure 5-3b shows that the arsenic concentrations were lowest in Missouri, and were highest in the States of Arizona, North Carolina, New Hampshire, Nevada, and Ohio. It should be noted that the distributions for North Carolina, New Hampshire, and Ohio may be influenced by the combination of relatively high detection limits (typically 5 to 10 µg/L) and high levels of censoring (arsenic was not detected in any samples from 70 to 78 percent of the surface water systems in these three States). The influence of censoring is seen by examining Figure 5-3d, which shows that for the systems that were not completely censored: the lowest values were for North Dakota and Alabama; North Carolina, New Hampshire, and Ohio have high concentration levels but are not the States with the highest concentrations; The very few remaining systems in Missouri are roughly at the median level for the 25 States.

Figure 5-4 shows a lognormal probability plot for ground water systems in the State of New Jersey, and Figure 5-5 shows a lognormal probability plot for ground water systems in the State of New Hampshire. Similar plots for the remaining States in the AOED database are included in Appendix B-3.

In these plots, means for uncensored PWS ID are represented with diamonds. (The PWSID is uncensored if there is at least one measured detected value for that system.) The straight line (in logarithmic space) shows the fitted values from the lognormal model. When the set of uncensored system means falls close to the line of the fitted values from the lognormal model, this is a qualitative indication that the distribution of system means does not strongly depart from the lognormal distribution. These plots were not created for States, water system types, and source types that had four or fewer uncensored system means, since they would not be expected to be reasonable approximations of the arsenic distributions in that case. With four or fewer detects, substitution of non-detects by half the reporting limit was used instead of the regression on order statistics method based on the lognormal model.

In many cases, the distributions of system means are fairly linear, and are consistent with lognormal distributions. Figure 5-4, for ground water CWS systems in New Jersey, is an example of a very good fit to the lognormal model. In some cases, the fit of the data is least strong in the tails of the distributions. The lognormal probability plot for ground water CWS systems in New Hampshire, shown in Figure 5-5, appeared to depart most strongly from lognormality. The sharp angle that appears on the probability plot near the first quartile of the distribution may be related to differences between measured values and censored system values.

Figure 5−4: System means of CWS GW arsenic concentrations for NJ, Log−normal probability plot

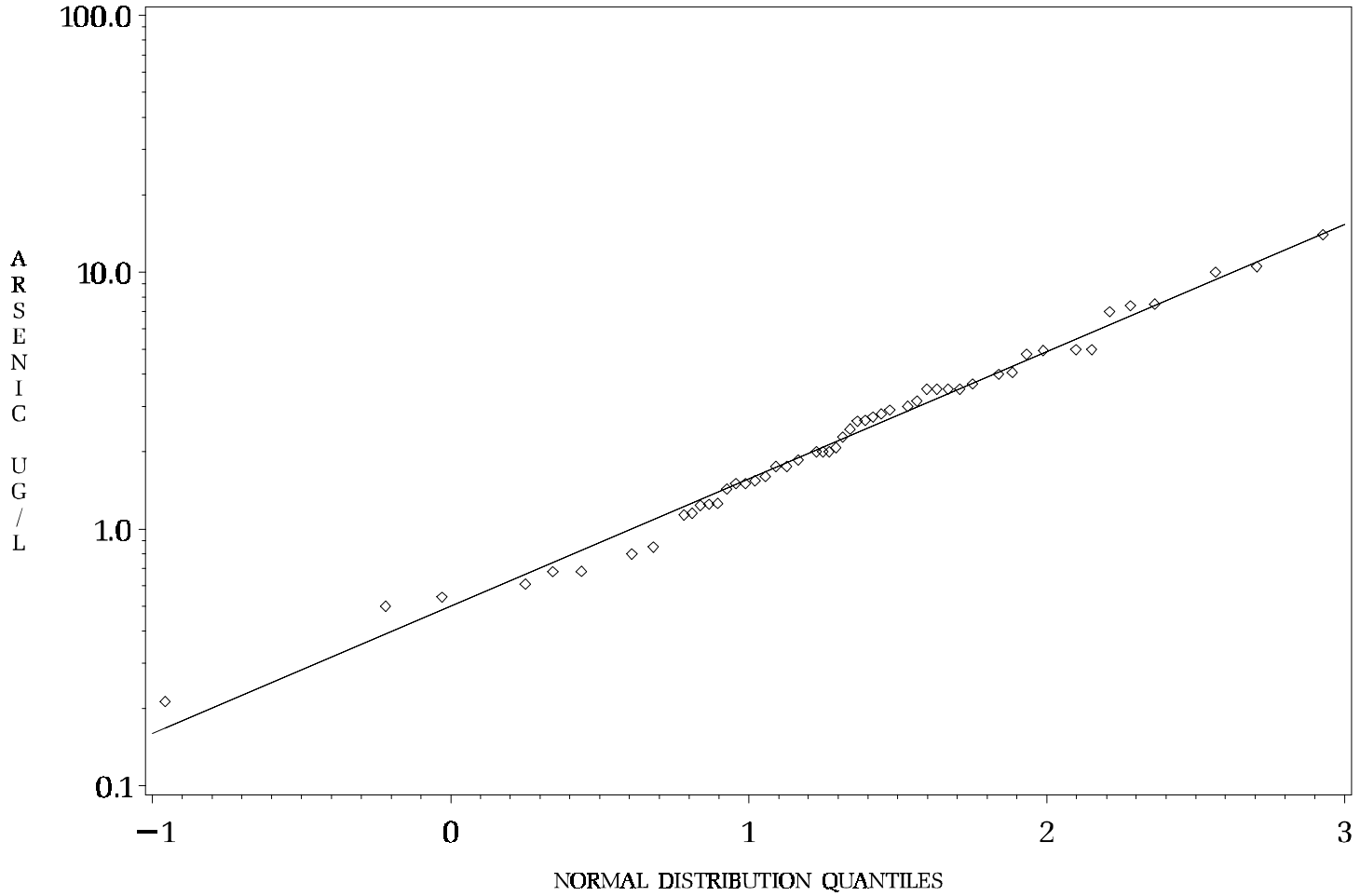
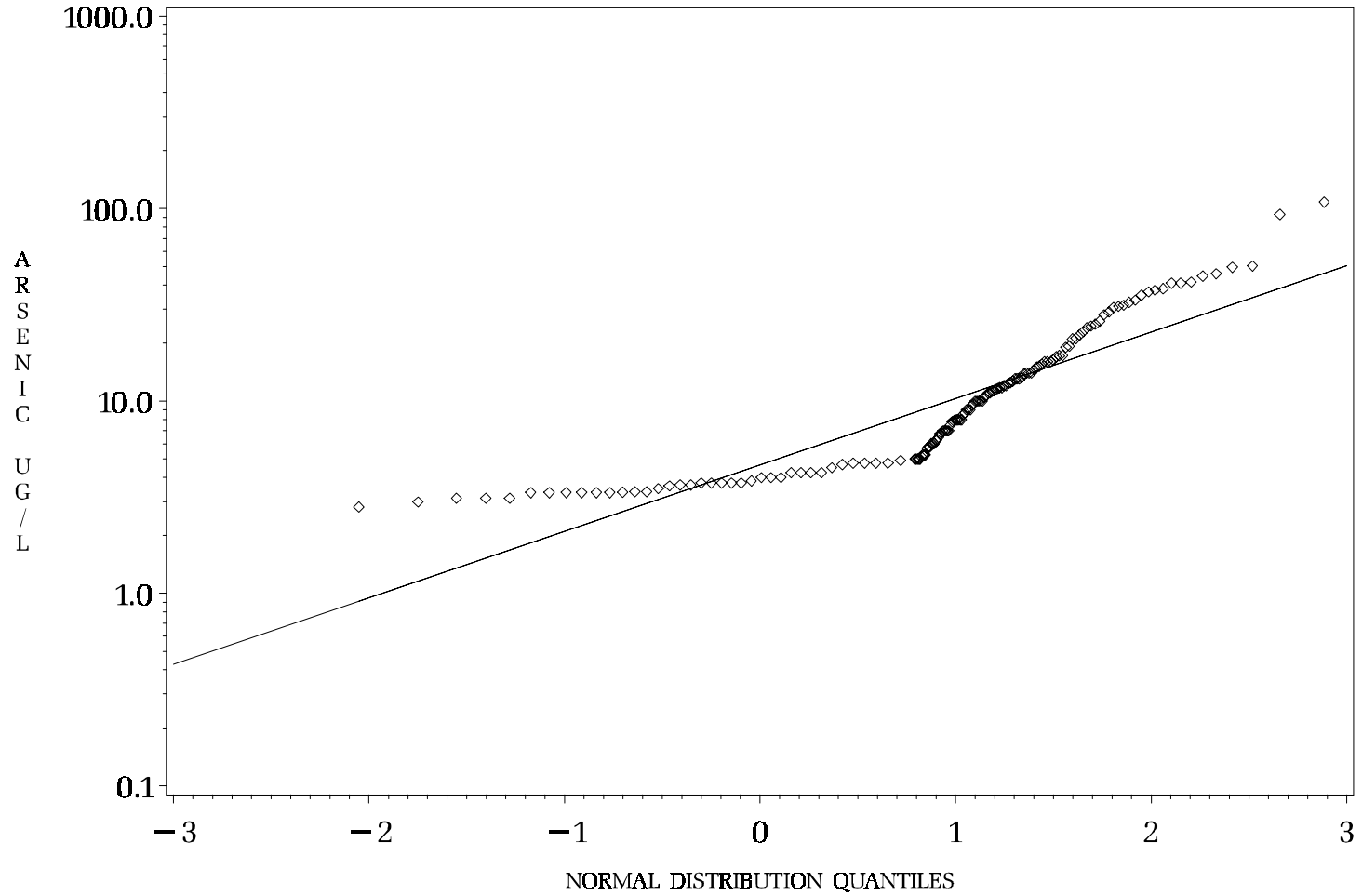


Figure 5−5: System means of CWS GW arsenic concentrations for NH, Log−normal probability plot



Several States have a similar pattern, in that the distribution provides a reasonably good fit to lognormality in the upper part of the distribution but may appear to fit less well at lower levels, which is perhaps due to the model used to fill-in the censored values. In cases like New Hampshire ground water CWS systems, a different lognormal distribution seems to fit the concentrations in the lower tail. Down to levels below the regulatory level of interest, the data are consistent with a lognormal model.

5.6 Summary of Patterns of Arsenic Occurrence

The analyses that are presented in this Chapter were designed to support decisions related to the selection of an appropriate method for estimating arsenic national occurrence. These analyses show that, in order to develop more representative estimates of arsenic occurrence, data should be stratified by source water type, and by region, and, if the data are sufficiently representative, by system type. The data for surface water NTNCWS systems are not sufficient to provide separate occurrence estimates for this type of system.. Data should not be stratified by system size. In addition, these data suggest that arsenic occurrence at the State level is relatively lognormally distributed above a certain cut-off level. As a result of these findings, in Chapter 6, arsenic data are stratified accordingly, and are assumed to be lognormally distributed above a certain cut-off level within States.

This page intentionally left blank

6. National Occurrence Estimates

This Chapter presents estimates of the number of systems that have mean arsenic levels equal to or greater than specific MCL[16] alternatives or concentrations of interest. First, the model used to estimate exceedance probabilities is defined, and the assumptions that were made and the data conventions that were applied are defined. Next, the specific number of systems that are predicted to exceed potential MCL alternatives are presented for community water supply systems and non-transient, non-community water supply systems. Then, the exceedance probabilities calculated using AOED are compared to those developed using the NAOS, USGS, NIRS, Metro, and ACWA databases. The last section of this Chapter discusses an uncertainty analysis that was conducted on the AOED-based arsenic occurrence estimates, and presents the results of that analysis. Confidence intervals generated as a result of the uncertainty analysis were applied to the exceedance estimates to provide a measure of the variability associated with these estimates.[17]

6.1 Arsenic National Occurrence Projection Methodology

The methodology applied to develop estimates of arsenic occurrence in ground water and surface water systems has five steps. These steps result in the derivation of probability distributions of mean arsenic concentrations and the estimates of numbers of systems that exceed specific mean concentrations of interest, and the estimation of the numbers of systems with system means exceeding specific concentrations of interest. This process fits the unique data structure of the compliance monitoring data in the AOED database. As a slight misuse of terminology we shall sometimes say that a system exceeds a certain level, instead of a more precise statement that the system mean for that system exceeds the level. Similarly, the exceedance probability distributions for the state, region, or nation, are defined as the actual or estimated percentage of systems in the geographical area with system means above various levels. The five steps include:

- Calculate system arithmetic means;
- Calculate State exceedance probability distributions for ground water and surface water;
- Apply weighting and develop regional exceedance probability distributions for ground water and surface water;
- Apply weighting and develop national exceedance probability distributions for ground water and surface water; and
- Estimate numbers of systems exceeding levels of interest as the product of the national probability distributions and the total number of ground water or surface water systems.

[16] An MCL is the maximum level of a contaminant that is allowable in public drinking water supplies. When EPA sets an MCL for a contaminant, the PWS must ensure that the level of this contaminant is maintained at or below its MCL.

[17] Strictly speaking, an estimate is a number that does not have uncertainty or variability. The proper statistical term is "estimator." Each estimate is a realization (observed value) of the estimator (a random variable, that has a statistical distribution).

In addition, it should be noted that confidence intervals for the system exceedance estimates are developed through the uncertainty analysis that is presented in Section 6.4. These intervals provide a measure of the variability associated with the national arsenic occurrence estimators, and they are applied in Section 6.2 of this report.

6.1.1 System Means

The first step in our analysis was to estimate an average arsenic concentration over time for each system in the AOED. The number of observations in the AOED for a single system varies by system and by State; larger systems tend to have more observations, and some States have many more observations per system. The set of system means is therefore more representative of arsenic occurrence in water systems than just the set of samples in the AOED would be.

In order to estimate the system means, we had to account for "censored" or "non-detected" concentrations. Non-detected concentrations are reported as "non-detect at X μg/L," for some detection or reporting limit X. This means that the concentration was somewhere between 0 and X μg/L; the analytical method used to estimate the concentration is incapable of reliably measuring concentrations less than X μg/L. Non-detected concentrations should not be ignored when estimating a mean; they are the lowest observations, so to discard them would introduce a positive bias in the mean estimate. Instead, we accounted for non-detected concentrations by "filling them in" statistically in one of two ways, and then averaging, according to the following plan:

- If all observations from the system were detects, we calculated the arithmetic mean of the observed concentrations for the system.
- If the system included at least five detected concentrations that were not all equal, and some non-detects, we filled in the non-detects by regression on order statistics (ROS), as described in Appendix A, and then averaged the detected and filled-in concentrations.
- If the system included four or fewer detected concentrations, or five or more detects that were all equal, then we substituted half of the detection limit for each non-detect, and then averaged the detected and substituted concentrations.
- If all samples from a system were censored, the system was labeled "non-detect" at the mode of the detection limits.

The ROS method uses observations above the detection limit to extrapolate to observations below the detection limit. The method is described in detail in Appendix A. The idea is to assume that the arsenic concentrations in a system follow a lognormal distribution; estimate the distribution's parameters by fitting a line to a lognormal probability plot of the detected concentrations; then extrapolate the fitted line below the detection limit, in order to obtain the expected locations of the non-detected concentrations. Appendix B-3 shows plots of this type for State distributions; although these plots are for States instead of systems, the method of plotting and filling in is the same. ROS is simple to implement, and it has been shown to perform nearly as well as traditional methods, such as maximum likelihood, when the lognormal assumption is correct (Helsel and Cohn, 1988). Perhaps more important is that even when lognormality does not hold, ROS still performs well while other methods perform less well.

ROS is one of a class of "fill-in" methods, in which non-detects are filled in in some way, and then parameters are estimated. The simple ½-detection-limit substitution is another, simple fill-in method, equivalent to assuming that non-detects follow a uniform distribution from 0 to the detection limit. Kroll and Stedinger (1996) compared three lognormal fill-in type methods, including ROS, for estimating means from censored data, over a range of water quality and flow distributions. All three methods follow the fill-in scheme described above, but where ROS uses regression to estimate the lognormal parameters, the other two methods use maximum likelihood (ML) and probability-weighted moments. Kroll and Stedinger found that the ROS method generally performs as well as or better than probability-weighted moments. ROS and ML perform about equally well under low to moderate censoring, while ML has some advantage when the censoring fraction is high (around 80%). On the other hand, ROS is simpler to understand and to implement than ML, which requires an iterative computation for each system. We decided that the potential benefit of the ML fill-in did not justify its extra computational complexity.

ROS requires a reasonable number of detected observations in order to give a reliable fit of the line to the probability plot. In our procedure above, we required at least 5 detected observations. Where there were fewer than 5 detects for a system, we opted to use the simple substitution of ½ the detection limit . This method is widely used, and has been shown to perform better than either of the other two common substitution methods, namely of 0 or 1 times the detection limit (Helsel and Cohn, 1988).

Following the decision criteria described above, we used ROS and the ½-detection limit substitution in 2.3% and 13.2%, respectively, of the systems in the AOED. In 20.8% of systems there were only detected concentrations. Thus we were able to estimate mean arsenic concentrations for all of these systems. The remaining 63.7% of systems contained only censored observations, and so were labeled non-detect. Although we did not estimate mean concentrations for these systems, they still played a role in the estimation of State exceedance probability distributions, as described in the next section.

6.1.2 State Exceedance Probability Distributions

State exceedance probability distributions indicate the probability that a randomly chosen PWS from any specific State will have a mean arsenic concentration greater than a particular concentration of interest. Using the sample set of system means that were derived for each State from the compliance data in AOED, exceedance probability distributions were developed separately for ground water CWS systems, surface water CWS systems, and ground water NTNCWS systems in each State. Exceedance probability distributions were not separately developed at the State, Regional, or national levels for surface water NTNCWS systems, due to the limited number of such systems in the AOED database and nationwide. There are only 816 such systems in the nation according to the 1998 Baseline SDWIS database. Instead, the national arsenic occurrence estimates for surface water NTNCWS systems were developed by applying the national arsenic occurrence estimated exceedance probabilities for surface water CWS systems to the numbers of surface water NTNCWS systems.

Several methods could be used to estimate exceedance probability distributions in each state. These include empirical and parametric estimators, using all or only a subset of the estimated means to estimate the parameters. Empirical distributions estimate the probability of exceeding any threshold as the observed fraction of the estimated system means that exceed that threshold. Empirical estimates are simple to compute, and they do not require any assumptions about the form of the distribution being estimated. A disadvantage of this type of estimate is that it "jumps" by a discrete amount at each datum, and does not predict well either above or below the range of the observed data. For example, it estimates zero probability (or in some versions, a small but fixed probability) of ever seeing a system mean of any size larger than the largest system mean observed so far.

Parametric estimators assume that the distribution follows a particular form, such as lognormal. The distribution has parameters which are estimated from the data, by, for example, maximum likelihood (Cohen, 1991) or an adapted ROS (Appendix A). At the cost of assuming a particular distributional form, the parametric estimator produces smoothly changing probability estimates even outside the range of the observed data. It also yields a simple computational form for the estimates, which is useful as an input to further analyses, such as the cost models used in the regulatory impact analyses (RIA) (although it should be noted that the RIA relies on National exceedance probability distributions rather than State exceedance probability distributions). Another reason to use a parametric fit is that, in the State distributions, the data do not consist of true system means; rather, they are only estimated means, as described in the previous section. An empirical estimator of exceedance probability preserves the errors in the estimators of the mean concentration, while a parametric estimator tends to smooth them.

Because of these advantages, parametric distributions were fit to the distribution of estimated system means in each combination of water system type, source type, and State. In particular, lognormal distributions were used, and these provide a reasonably good fit in most cases. Appendix B-3 shows lognormal probability plots for each State and source water type, in which log-system means are plotted against their corresponding normal quantiles. Each plot also shows a regression line fitted to the data in the plot. If the data in a plot are truly lognormally distributed, they should lie close to the fitted line.

Examination of the plots in Appendix B-3 suggests that in some States and source types, two distinct populations are present: the plotted system means form a broken line, instead of a straight line that would indicate a single lognormally distributed population. This effect is most apparent in the plot for New Hampshire ground water (see Figure 5-5), where the plotted system means form two lines of different slopes, with the breakpoint at 5 μg/L on the vertical axis. In New Hampshire, there are no detected concentrations below 5 μg/L, and the detection limit for all non-detected concentrations is 5 μg/L. Therefore, any system mean that is estimated to be less than 5 μg/L must have been estimated from a large portion of data that was "filled in" below 5 μg/L from non-detects, as described in the previous section. The probability plot implies that these systems may form a different distribution than the systems above 5 μg/L, which presumably have fewer filled-in observations. The implication is that when a large fraction of the observations from a system are non-detects, filling in the missing observations may not reproduce the State's arsenic distribution accurately.

In light of this evidence, a cutoff point was established for each State and source water combination, and the values of the system means that were less than this cutoff point were not used in fitting the lognormal distributions. The numbers of system means below the cutoff point were used to compute the plotting positions for the system means above the cutoff point, according to the ROS method. The plotting positions are the censored probability plotting positions defined in Appendix A. The cutoff points were mostly set equal to the most common detection limit for the State and source water type. In a few cases cutoff points were set lower than the modal detection limit, where there were not enough system means above the detection limit to give reasonably stable parameter estimates. Somewhat arbitrarily, we decided to require 10 or more system means above the selected cutoff point. The cutoff points for State and source water types are presented in Table 6-1.

Table 6-1
Cutoff Points for Ground Water and Surface Water, in µg/L

State	Ground Water	Surface Water	State	Ground Water	Surface Water
AL	1	1	MT	> 2	> 2
AK	2	2	NC	5*	5*
AR	0*	0*	ND	2	1
AZ	5	5	NH	5	2*
CA	2	2	NJ	1	1
IL	2	1	NM	2	2
IN	2	1	NV	5	5
KS	2	2	OH	10	2*
KY	> 2	> 2	OK	2	2
ME	2	1	OR	2*	2*
MI	2	1	TX	2	2
MN	2	0*	UT	2	1
MO	2	1			

* Cutoff point set below modal reporting limit. Modal reporting limits for these States are (GW and SW modal RL are the same except where noted): AR (5); MN (1); NC (10); NH (5); NJ (GW: 5; SW: 2); OH (10); and OR (5).

Using a parametric approach that we have named right-tailed ROS, lognormal distributions were fit to the remaining system means in each State, type of water system (CWS or NTNCWS), and source water type, by fitting a regression line to the data in the plots of Appendix B-3, that is, to log-system means greater than the cutoff point plotted against their normal quantiles. The plotting positions are the censored probability plotting positions defined in Appendix A. The exceedance probability for any given log-system mean arsenic concentration was then estimated by using the fitted regression line to find the normal quantile corresponding

to that concentration, and computing the standard normal probability associated with the quantile (see A-3 in Appendix A). The estimated probability distributions for each State, type of water system, and source water type are presented in Appendix B-1.

In a few cases this right-tailed ROS method could not be meaningfully applied because the number of detected system means was insufficient for analysis. If there were four or fewer detected system means, then the empirical method was employed instead of right-tailed ROS: If a system is completely censored, i.e., all the measurements for that system are below the reporting limit, then the system mean is estimated as one half of the mode of the reporting limits for that system. Otherwise the system mean is estimated as the arithmetic mean of the measured and filled-in concentration values, as described in Section 6.1.1. Then, the estimated exceedance probabilities from the empirical method are the percentages of all system means above the selected concentration levels. The cases where this empirical method was applied instead of right-tailed ROS are identified by Use ROS? = "N" in the Appendix B-1 tables.

6.1.3 Regional Exceedance Probability Distributions

The third step to developing national occurrence projections was to develop regional exceedance probability distributions. Separate probability distributions were developed for arsenic occurrence in ground water CWS systems, surface water CWS systems, and ground water NTNCWS systems, in each of seven NAOS regions. The seven NAOS regions are based on those identified by Frey and Edwards (1997), and the States in each region are discussed in Section 5.4.[18] Regional exceedance probability distributions were developed as the weighted sum of the exceedance probability distributions derived for each State with compliance monitoring data in the region.[19] As such, for Region Y represented by data from *t* States, where *t* is any integer, the Regional distribution for each type of water system and source type was calculated as:

$$P_{>x,R(y)} = (n_{>x,s1} + n_{>x,s2}+...+n_{>x,st}) / (N_{s1} + N_{s2}+...+N_{st})$$

where:

$P_{>x,R(y)}$ is the probability that a PWS system in Region Y will have a mean arsenic concentration that exceeds the arsenic concentration x;

$n_{>x,s1}$, $n_{>x,s2}$ to $n_{>x,st}$ are the number of purchased and non-purchased water systems in States 1 to t that are predicted to have mean arsenic levels greater than x;

N_{s1} , N_{s2} to N_{st} are the total number of purchased and non-purchased water systems in States 1 to *t*; and

$n_{>x,Ri}$ was estimated by multiplying the regional probability distribution developed in Section 6.1.3 by the total number of purchased and non-purchased systems in the region (from SDWIS).

[18] Note that any similarities between the boundaries of the NAOS Regions and the boundaries of EPA's Regional Offices is purely coincidental.

[19] The number of systems exceeding specific arsenic concentrations in each state are self-weighted quantities.

For each potential MCL alternative or concentration of interest, a separate exceedance probability was calculated based on available data. States within a region were only used to estimate exceedance probabilities for arsenic concentrations higher than their censoring level. For example, in the Western Region, only the States of Alaska, California, and Utah have detection limits that are equal to or less than 2 µg/L, so only these States were used to estimate the number of systems in the region that are likely to have mean arsenic levels greater than 2 µg/L. They also contribute to the estimates for MCL alternatives of 3, 5, 10, 15, 20, and up to 50 µg/L. Arizona, Nevada, and Oregon are included for all estimates at 5 µg/L and above. The remaining States in the Western Region, Idaho and Washington, did not have compliance monitoring data in AOED. The regional exceedance probability distributions based on the compliance monitoring data are presented in Table 6-2a for CWS systems and Table 6-2b for NTNCWS systems.

For NTNCWS systems, some modifications were made to the above approach. For surface water NTNCWS systems, the Regional exceedance probability estimates were copied from the surface water CWS Regional exceedance probability estimates. For ground water NTNCWS systems, NTNCWS data from 13 States and CWS data from Alabama were used to develop the Regional exceedance probability estimates. Ground water NTNCWS data was used from the following States: Alaska, Arizona, California, Indiana, Kansas, Michigan, Minnesota, North Carolina, North Dakota, New Jersey, New Mexico, Oregon, and Texas. The other four States with some ground water NTNCWS data in the AOED are Alabama, Missouri, Montana, and Utah. For those four States, there were fewer than ten uncensored systems above the cut-off and so those data were not used in the development of the Regional probability estimates. Since Alabama is the only State with compliance data in Region 3, Southeast, the ground water CWS data from Alabama was substituted for the ground water NTNCWS data so that Regional estimates could be obtained for that Region.

The convention of using only data from States with a detection limit below or equal to the concentration of interest to estimate regional occurrence probabilities generally yields regional exceedance percentages that decline as concentrations increase. However, this convention does result in two anomalies in the regional probability distributions, which appear as increases in probabilities as arsenic concentrations rise. Specifically, in Mid-Atlantic ground water CWS systems, the probability of exceeding 5 µg/L (0.39 percent) is based on Kentucky data, whereas the probability of exceeding 10 µg/L (0.75 percent) is based on data from both Kentucky and North Carolina. The second anomaly was observed in surface water CWS systems in the New England Region, where exceedance probabilities rose from 6.2 percent at 3 µg/L to 11.7 percent at 5 µg/L. At the lower concentration, the probability is based on data from Maine, while the higher concentration is based on data from both Maine and New Hampshire.

The data presented in Table 6-2a suggest that arsenic occurrence in ground water CWS systems is lowest in the Mid Atlantic and South East Regions. Also, these data indicate that intermediate ground water arsenic mean levels are found in New England, Midwest, North Central, and South Central Regions, and that the West Region tends to have higher arsenic mean levels than the other Regions. These regional patterns of arsenic occurrence in ground water are generally similar to those reported by other studies (Focazio *et al.*, 2000; Frey and Edwards,

Table 6-2a
Regional Exceedance Probability Distribution Estimates for Community Water Systems

	Percent of Systems Exceeding Arsenic Concentrations (μg/L) of:									
Region	2	3	5	10	15	20	25	30	40	50
Ground Water Systems										
New England	28.876	21.668	20.806	6.951	4.288	2.948	2.163	1.658	1.065	0.741
Mid Atlantic	--	--	0.393	0.748*	0.156	0.044	0.015	0.006	0.001	0.0
South East	1.511	0.873	0.413	0.135	0.066	0.039	0.026	0.018	0.010	0.006
Midwest	28.286	21.170	13.776	6.221	3.614	2.387	1.701	1.276	0.794	0.539
South Central	27.162	18.564	9.696	3.603	1.830	1.083	0.703	0.486	0.264	0.160
North Central	29.632	21.265	13.083	6.011	3.586	2.419	1.755	1.337	0.854	0.594
West	42.380	31.512	25.240	12.464	7.499	4.996	3.549	2.636	1.596	1.051
Surface Water Systems										
New England	8.441	6.231	11.736*	0.973	0.578	0.430	0.340	0.279	0.203	0.157
Mid Atlantic	--	--	0.109	0.014	0.0	0.0	0.0	0.0	0.0	0.0
South East	0.516	0.212	0.062	0.009	0.003	0.001	0.001	0.0	0.0	0.0
Midwest	4.371	3.047	1.648	0.660	0.431	0.321	0.254	0.209	0.153	0.119
South Central	10.424	3.826	0.856	0.184	0.092	0.057	0.039	0.029	0.018	0.012
North Central	19.081	9.104	3.247	0.580	0.170	0.065	0.029	0.014	0.004	0.002
West	17.758	12.675	8.171	3.421	2.006	1.364	1.007	0.782	0.520	0.376

*Regional exceedance probabilities for New England SW are based on Maine data at 2 and 3 μg/L, and Maine and New Hampshire data at 5 μg/L. For Mid Atlantic GW, exceedance probabilities are based on Kentucky data at 5 μg/L and Kentucky and North Carolina data at 10 μg/L. The differences in arsenic occurrence between these States result in Regional exceedance probabilities that increase in these concentration ranges.

Table 6-2b
Regional Exceedance Probability Distribution Estimates for Non-Transient Non-Community Water Systems

	Percent of Systems Exceeding Arsenic Concentrations (µg/L) of:									
Region	2	3	5	10	15	20	25	30	40	50
Ground Water Systems										
New England	--	--	--	2.146	1.055	0.610	0.389	0.265	0.140	0.083
Mid Atlantic	--	--	--	1.438	0.774	0.483	0.330	0.239	0.140	0.091
South East[1]	1.511	0.873	0.413	0.135	0.066	0.039	0.026	0.018	0.010	0.006
Midwest	34.505	26.169	17.085	8.234	4.929	3.290	2.348	1.756	1.078	0.721
South Central	33.280	24.008	14.449	5.947	3.145	1.893	1.236	0.855	0.460	0.276
North Central	35.863	29.753	22.815	15.044	11.431	9.273	7.819	6.765	5.329	4.391
West	45.017	34.294	21.923	10.514	6.289	4.200	3.002	2.248	1.385	0.929
Surface Water Systems[2]										
New England	8.441	6.231	11.736*	0.973	0.578	0.430	0.340	0.279	0.203	0.157
Mid Atlantic	--	--	0.109	0.014	0.0	0.0	0.0	0.0	0.0	0.0
South East	0.516	0.212	0.062	0.009	0.003	0.001	0.001	0.0	0.0	0.0
Midwest	4.371	3.047	1.648	0.660	0.431	0.321	0.254	0.209	0.153	0.119
South Central	10.424	3.826	0.856	0.184	0.092	0.057	0.039	0.029	0.018	0.012
North Central	19.081	9.104	3.247	0.580	0.170	0.065	0.029	0.014	0.004	0.002
West	17.758	12.675	8.171	3.421	2.006	1.364	1.007	0.782	0.520	0.376

[1]Regional exceedance probabilities for NTNCWS SW in the Southeast Region are based on CWS data from Alabama and so are identical to the Southeast regional exceedance probabilities for CWS SW.

[2]Regional exceedance probabilities for NTNCWS SW are copied from the regional exceedance probabilities for CWS SW.

*Regional exceedance probabilities for New England NTNCWS SW are based on Maine CWS data at 2 and 3 µg/L, and Maine and New Hampshire CWS data at 5 µg/L. The differences in arsenic occurrence between these States result in Regional exceedance probabilities that increase in these concentration ranges.

1997), except the other studies found that arsenic mean levels in New England were relatively low and comparable to those in the Mid Atlantic and South East. Therefore, it is possible that the compliance monitoring data for the States of Maine, New Hampshire and New Jersey may over estimate arsenic occurrence in the New England Region. In surface water CWS systems, the lowest arsenic mean levels again occurred in the Mid-Atlantic and South East. Mean levels generally appeared to increase in the other Regions, from Midwest, to South Central, to the New England, West, and North Central Regions, although the ordering of these Regions changes with the concentration level.

6.1.4 National Exceedance Probability Distributions

The fourth step that was used to develop national occurrence estimates was to develop estimates of national exceedance probability distributions from the regional exceedance probability distributions. Separate exceedance probability distributions were developed for arsenic occurrence in ground water CWS systems, surface water CWS systems, and ground water NTNCWS systems of the United States. National exceedance probability distributions were developed as the weighted sum of the exceedance probability distributions derived for each Region, which were presented in Tables 6-2a and 6-2b. As such, when the United States is represented by data from 7 Regions, the National distribution was calculated separately for each water system type and source type as:

$$P_{>x,US} = (n_{>x,R1} + n_{>x,R2}+...+n_{>x,R7}) / (N_{R1} + N_{R2}+...+N_{R7})$$

where:

$P_{>x,US}$ is the probability that a PWS system in the United States will exceed arsenic concentration x;

$n_{>x,R1}$, $n_{>x,R2}$ to $n_{>x,R7}$ are the number of purchased and non-purchased water systems in Regions 1 to 7 that are predicted to have mean arsenic levels greater than x; and

N_{R1} , N_{R2} to N_{R7} are the total number of purchased and non-purchased water systems in Regions 1 to 7.

For each potential MCL alternative or concentration of interest, a separate exceedance probability was calculated based on available data. Exceedance probabilities from all Regions contributed to the estimation of each of these National exceedance probability distribution estimates at each concentration, with the exception of the Mid Atlantic Region for CWS systems and with the exceptions of the Mid Atlantic and New England Regions for NTNCWS systems. Because regional exceedance probability distributions for CWS systems were not estimated for the Mid Atlantic Region at concentrations of 2 or 3 µg/L, no data was available for this Region to support the development of National estimates at these concentrations. Because regional exceedance probability distributions for NTNCWS systems were not estimated for the Mid Atlantic and New England Regions at concentrations of 2, 3, or 5 µg/L, no data was available for these Regions to support the development of National estimates at these concentrations.

Table 6-3a shows the estimated national arsenic occurrence exceedance probability distributions for ground water and surface water CWS, and Table 6-3b shows the estimated national arsenic occurrence exceedance probability distributions for ground water and surface water NTNCWS. Since the Regional estimates for surface water NTNCWS systems are copied from the Regional estimates for surface water CWS systems, the national estimates for surface water NTNCWS systems are the same as the national estimates for surface water CWS systems. Tables 6-3a and 6-3b also show the exceedance probabilities of lognormal distributions that were fit, using ROS, to each of the national distributions. The fitted lognormal distributions are mostly quite close to the original distributions. The lognormal distributions have two advantages: they can predict the percent exceedances at low concentrations, and they have a simple functional form that is easy to use in further analyses. In particular, OGWDW used the fitted lognormal distributions in Tables 6-3a and 6-3b for the cost simulations in its regulatory impact analysis of the arsenic MCL.

6.1.5 Number of Systems Exceeding Alternative MCLs

The estimated number of systems exceeding alternative MCL levels was calculated by multiplying the total number of systems in the United States with the probability that a system would exceed a specific MCL alternative. Separate estimates were developed for ground water and surface water CWS, and for ground water and surface water NTNCWS. The total number of systems in each category was derived from 1998 Baseline SDWIS data. It should be reiterated that the national right-tailed ROS exceedance probability distributions for surface water systems were derived from analysis of CWS systems, and these probability distributions were applied to estimate arsenic occurrence in both CWS and NTNCWS surface water systems.

The following section presents estimates of numbers of ground water and surface water systems within specific size categories that may have mean arsenic levels in excess of specified MCL alternatives. These estimates are based on the national right-tailed ROS exceedance probability distributions for ground water and surface water, multiplied by the total number of systems in the nation in each size category. Analyses presented in Chapter 5 indicated that there are not meaningful or consistent differences in arsenic occurrence from size stratum to size stratum. Therefore, for each source water type, the same national right-tailed ROS exceedance probability distribution was applied for each size stratum.

6.2 Arsenic National Occurrence Estimates Results

The techniques described above were applied to develop estimates of the proportions of ground water and surface water systems with system mean concentrations above potential regulatory levels. These estimates are presented separately for community water supply systems and for non-transient, non-community water supply systems. The calculation of the confidence intervals is presented in Section 6.4 below.

Table 6-3a
Estimated National Exceedance Probabilities for Ground Water and Surface Water Community Water Systems

	Percent of Systems with Mean Arsenic Concentrations Exceeding (μg/L) of:											
	0.5	1	2	3	5	10	15	20	25	30	40	50
GW, Weighted Point Estimate	--	--	27.33	19.88	12.09	5.30	3.06	2.00	1.42	1.05	0.64	0.43
95 % Confidence Interval			[26.57, 29.94)	[19.25, 21.87]	[11.74, 13.04]	[5.18, 5.91]	[2.92, 3.45]	[1.89, 2.29]	[1.32, 1.64]	[0.97, 1.23]	[0.58, 0.77]	[0.38, 0.52]
GW, Lognormal fit	61.01	43.71	27.55	19.69	11.99	5.33	3.08	2.01	1.42	1.05	0.64	0.43
SW, Weighted Point Estimate	--	--	9.817	5.566	3.036	0.799	0.464	0.320	0.239	0.188	0.128	0.095
95% Confidence Interval			[9.08, 25.46]	[4.79, 20.63	[1.8, 9.66]	[0.52, 1.56]	[0.23, 1.00]	[0.13, 0.82]	[0.08, 0.72]	[0.06, 0.66]	[0.03, 0.61]	[0.02, 0.59]
SW, Lognormal fit	28.596	16.777	8.679	5.552	2.960	1.117	0.592	0.366	0.247	0.178	0.103	0.067
GW & SW, Weighted Point Estimate	--	--	23.86	17.05	10.30	4.41	2.55	1.67	1.19	0.88	0.54	0.36

Notes: The occurrence estimates presented in Tables 6-4 and 6-5 are based on the right-tailed ROS exceedance probability distributions denoted as "Weighted Point Estimate".
GW - Ground water, lognormal regression yielded a mean and standard deviation in natural log coordinates of -0.251 and 1.583.
SW - Surface water, lognormal regression yielded a mean and standard deviation in natural log coordinates of -1.678 and 1.743.
The weighted point estimates were developed as described in Sections 6.1.1 to 6.1.4.
Neither a lognormal regression nor 95% confidence intervals were calculated for the combined GW and SW exceedance estimates.
The 95 % Confidence Intervals were derived as described in Section 6.5 using an earlier version of the database, and so may not be fully consistent with the weighted point estimates and lognormal fit estimates.

Table 6-3b
Estimated National Exceedance Probabilities for Ground Water and Surface Water
Non-Transient Non-Community Water Systems

	Percent of Systems with Mean Arsenic Concentrations Exceeding (μg/L) of:											
	0.5	1	2	3	5	10	15	20	25	30	40	50
GW, Weighted Point Estimate	--	--	32.117	24.231	15.607	5.346	3.142	2.082	1.485	1.114	0.693	0.473
GW, Lognormal fit	68.947	50.916	32.693	23.449	14.196	6.143	3.448	2.199	1.515	1.101	0.647	0.418
SW, Weighted Point Estimate	--	--	9.817	5.566	3.036	0.799	0.464	0.320	0.239	0.188	0.128	0.095
SW, Lognormal fit	28.596	16.777	8.679	5.552	2.960	1.117	0.592	0.366	0.247	0.178	0.103	0.067

Notes: The occurrence estimates presented in Tables 6-6 and 6-7 are based on the right-tailed ROS exceedance probability distributions denoted as "Weighted Point Estimate".
GW - Ground water, lognormal regression yielded a mean and standard deviation in natural log coordinates of 0.034 and 1.470.
SW - Surface water, lognormal regression yielded a mean and standard deviation in natural log coordinates of -1.678 and 1.743.
The weighted point estimates were developed as described in Sections 6.1.1 to 6.1.4.
The estimates for NTNCWS SW are copied from the estimates for CWS SW in Table 6-3a.
Confidence intervals were not estimated for the NTNCWS data.

6.2.1 Community Water Supply Systems

Tables 6-4 and 6-5 present the projected arsenic occurrence levels in community water supply systems with ground water and surface water sources, respectively. The data for the probability distributions are derived exclusively from the AOED compliance monitoring database, and the total number of systems is based on SDWIS data.

Under these estimates, 11,873 (CI:11,543-13,007) ground water CWS systems are estimated to have mean arsenic levels that exceed 2 µg/L. The estimated number of exceeding systems decreases rapidly at higher potential MCL alternatives. For example, 5,252 (CI: 5,100-5,665) systems are predicted to have mean arsenic levels greater than 5 µg/L, whereas 2,302 (CI: 2,250-2,567) systems are predicted to have mean arsenic levels greater than 10 µg/L. While 869 (CI: 821-995) systems are predicted to have mean arsenic concentrations above 20 µg/L, 187 (CI: 165-226) systems are estimated to have average arsenic levels greater than 50 µg/L. It is worth noting that the number of systems which are predicted to exceed the current MCL of 50 µg/L is significantly higher than number of systems which SDWIS indicates actually violate the MCL (from January, 1996 to March, 1999, 15 ground water CWSs have violated the MCL).

In the United States, there are fewer surface water CWS than ground water CWS, and the exceedance probabilities for surface water systems decrease more quickly as arsenic concentrations rise than do the exceedance probabilities for ground water systems. As a result, fewer surface water systems have mean arsenic levels above specific concentrations of interest than ground water systems at corresponding arsenic concentrations. Under these estimates, 1,052 (CI: 973-2,730) surface water CWS systems are predicted to have arsenic concentrations above 2 µg/L, and 325 (CI: 193-1,036) surface water CWS systems are predicted to have mean arsenic concentrations greater than 5 µg/L. A total of 86 (CI: 56-167) surface water CWS systems are predicted to have mean arsenic concentrations that exceed 10 µg/L, and 34 (CI: 14-88) are predicted to have mean arsenic concentrations that exceed 20 µg/L. Ten (CI: 2-63) surface water CWS systems are predicted to have arsenic concentrations above 50 µg/L.

6.2.2 Non-Transient, Non-Community Water Supply Systems

Tables 6-6 and 6-7 present the projected arsenic occurrence levels in non-transient, non-community water supply systems with ground water and surface water sources, respectively. The data for the ground water probability distributions are derived mainly from the non-transient non-community water systems in the AOED compliance monitoring database, although for Alabama and the Southeast region, AOED data from ground water community water supply systems was used. The data for the surface water probability distributions are derived exclusively from the community water systems in the AOED compliance monitoring database, as discussed in Section 6.1.5. The total number of systems is based on 1998 Baseline SDWIS data. Confidence intervals were not developed for the ground water NTNCWS occurrence estimates in Table 6-6.

Table 6-4
Estimated Arsenic Occurrence in U. S. Ground Water CWS

		Number of Systems with Mean Arsenic Concentrations[2] (µg/L) of:									
System Size (Population Served)	**Total Number of Systems[1]**	**> 2**	**> 3**	**> 5**	**> 10**	**> 15**	**> 20**	**> 25**	**> 30**	**> 40**	**> 50**
< 25	178	49	35	22	9	5	4	3	2	1	1
25-100	14025	3833	2788	1696	743	429	281	199	147	90	60
101-500	14991	4097	2980	1812	795	459	300	213	157	96	64
501-1,000	4671	1277	929	565	248	143	93	66	49	30	20
1,001-3,300	5710	1561	1135	690	303	175	114	81	60	37	25
3,301-10,000	2459	672	489	297	130	75	49	35	26	16	11
10,001-50,000	1215	332	242	147	64	37	24	17	13	8	5
50,001-100,000	131	36	26	16	7	4	3	2	1	1	1
100,001-1,000,000	61	17	12	7	3	2	1	1	1	0	0
> 1,000,000	2	1	0	0	0	0	0	0	0	0	0
Total Systems	**43443**	**11873**	**8636**	**5252**	**2302**	**1329**	**869**	**617**	**456**	**278**	**187**
Lower 95% CI:		**11543**	**8363**	**5100**	**2250**	**1269**	**821**	**573**	**421**	**252**	**165**
Upper 95% CI:		**13007**	**9501**	**5665**	**2567**	**1499**	**995**	**712**	**534**	**335**	**226**

Notes:
CI: confidence interval
1 Based on 1998 Baseline SDWIS data for purchased and non-purchased systems. Systems characterized as GW under the influence of SW are considered to be surface water systems.
2 Based on national weighted point estimates presented in Table 6-3a.
3 Totals may not add up due to rounding of the number of systems to the nearest whole number.

Table 6-5
Estimated Arsenic Occurrence in U. S. Surface Water CWS

		Number of Systems with Mean Arsenic Concentrations[2] (µg/L) of:									
System Size (Population Served)	**Total Number of Systems[1]**	**> 2**	**> 3**	**> 5**	**> 10**	**> 15**	**> 20**	**> 25**	**> 30**	**> 40**	**> 50**
< 25	74	7	4	2	1	0	0	0	0	0	0
25-100	1001	98	56	30	8	5	3	2	2	1	1
101-500	1983	195	110	60	16	9	6	5	4	3	2
501-1,000	1219	120	68	37	10	6	4	3	2	2	1
1,001-3,300	2420	238	135	73	19	11	8	6	5	3	2
3,301-10,000	1844	181	103	56	15	9	6	4	3	2	2
10,001-50,000	1606	158	89	49	13	7	5	4	3	2	2
50,001-100,000	300	29	17	9	2	1	1	1	1	0	0
100,001-1,000,000	261	26	15	8	2	1	1	1	0	0	0
> 1,000,000	13	1	1	0	0	0	0	0	0	0	0
Total Systems	**10721**	**1052**	**597**	**325**	**86**	**50**	**34**	**26**	**20**	**14**	**10**
Lower 95% CI:		**973**	**514**	**193**	**56**	**25**	**14**	**9**	**6**	**3**	**2**
Upper 95% CI:		**2730**	**2212**	**1036**	**167**	**107**	**88**	**77**	**71**	**65**	**63**

Notes:
CI: confidence interval
1 Based on 1998 Baseline SDWIS data for purchased and non-purchased systems. Systems characterized as GW under the influence of SW are considered to be surface water systems.
2 Based on national weighted point estimates presented in Table 6-3a.
3 Totals may not add up due to rounding of the number of systems to the nearest whole number.

Table 6-6
Estimated Arsenic Occurrence in U. S. Ground Water NTNCWS

		Number of Systems with Arsenic Concentrations[2] (µg/L) of:									
System Size (Population Served)	**Total Number of Systems[1]**	**> 2**	**> 3**	**> 5**	**> 10**	**> 15**	**> 20**	**> 25**	**> 30**	**> 40**	**> 50**
< 25	31	10	8	5	2	1	1	0	0	0	0
25-100	9732	3126	2358	1519	520	306	203	145	108	67	46
101-500	7103	2281	1721	1109	380	223	148	105	79	49	34
501-1,000	1996	641	484	312	107	63	42	30	22	14	9
1,001-3,300	696	224	169	109	37	22	14	10	8	5	3
3,301-10,000	62	20	15	10	3	2	1	1	1	0	0
10,001-50,000	15	5	4	2	1	0	0	0	0	0	0
50,001-100,000	0	0		0	0	0	0	0	0	0	0
100,001-1,000,000	0	0		0	0	0	0	0	0	0	0
> 1,000,000	0	0		0	0	0	0	0	0	0	0
Total Systems	**19635**	**6306**	**4758**	**3064**	**1050**	**617**	**409**	**292**	**219**	**136**	**93**

Notes:
1 Based on 1998 Baseline SDWIS data for purchased and non-purchased systems. Systems characterized as GW under the influence of SW are considered to be surface water systems.
2 Based on national weighted point estimates presented in Table 6-3b.
3 Totals may not add up due to rounding of the number of the systems to the nearest whole number.

Table 6-7
Estimated Arsenic Occurrence in U. S. Surface Water NTNCWS

		Number of Systems with Arsenic Concentrations[2] (µg/L) of:									
System Size (Population Served)	**Total Number of Systems[1]**	**> 2**	**> 3**	**> 5**	**> 10**	**> 15**	**> 20**	**> 25**	**> 30**	**> 40**	**> 50**
< 25	5	0	0	0	0	0	0	0	0	0	0
25-100	280	27	16	9	2	1	1	1	1	0	0
101-500	314	31	17	10	3	1	1	1	1	0	0
501-1,000	107	11	6	3	1	0	0	0	0	0	0
1,001-3,300	80	8	4	2	1	0	0	0	0	0	0
3,301-10,000	23	2	1	1	0	0	0	0	0	0	0
10,001-50,000	5	0	0	0	0	0	0	0	0	0	0
50,001-100,000	1	0	0	0	0	0	0	0	0	0	0
100,001-1,000,000	1	0	0	0	0	0	0	0	0	0	0
> 1,000,000	0	0	0	0	0	0	0	0	0	0	0
Total Systems	**816**	**80**	**45**	**25**	**7**	**4**	**3**	**2**	**2**	**1**	**1**
Lower 95% CI:		**74**	**39**	**15**	**4**	**2**	**1**	**1**	**0**	**0**	**0**
Upper 95% CI:		**208**	**168**	**79**	**13**	**8**	**7**	**6**	**5**	**5**	**5**

Notes:
1 Based on 1998 Baseline SDWIS data for purchased and non-purchased systems. Systems characterized as GW under the influence of SW are considered to be surface water systems.
2 Based on national weighted point estimates presented in Table 6-3b. These estimates were derived from CWS SW data.
3 Totals may not add up due to rounding of the number of systems to the nearest whole number.

There are fewer NTNCWS in the United States than CWS, and therefore the numbers of NTNCWS systems predicted to exceed specific levels are lower than the numbers of CWS that would exceed similar levels. These projections indicate that 6,306 ground water NTNCWS have system mean levels above 2 µg/L; 3,064 have system mean levels above 5 µg/L; and 1,050 have system mean levels above 10 µg/L. While 409 GW NTNCWS are predicted to have mean arsenic levels above 20 µg/L, 93 are expected to have arsenic levels in excess of the current standard, 50 µg/L. As there are few surface water NTNCWS in the United States, 80 systems are predicted to exceed 2 µg/L; 7 are predicted to exceed 10 µg/L; and 3 are predicted to exceed 20 µg/L. One surface water NTNCWS facility is predicted to have system mean levels above 50 µg/L.

6.3 Comparisons of Occurrence Estimates

6.3.1 Comparison of AOED, NAOS, USGS, MWDSC, and Wade Miller Occurrence Estimates at the National and Regional Level

In addition to the AOED occurrence results presented above, four additional studies have developed national occurrence estimates for arsenic in drinking water. The additional studies include the NAOS study, a national stratified random sampling of systems (Frey and Edwards, 1997), the USGS study of arsenic occurrence in ground water (Focazio *et al*., 2000), the Metropolitan Water District of Southern California national survey of arsenic occurrence in surface water and ground water (MWDSC, 1993), and Wade Miller (1992), which is based in part on the National Inorganic and Radionuclides Survey (NIRS) data for arsenic occurrence in ground water. The databases and survey methodologies used in these studies are described in Section 4.2 of this report. It is important to note that each of these occurrence estimates was developed in a slightly different manner.

The AOED arsenic occurrence estimates (with details presented above) are based on compliance monitoring data from more than 18,000 systems in 25 states. The NAOS occurrence estimates are based on a stratified random sampling from representative groups defined by source type, system size, and geographic location from 517 samples from approximately 500 systems. The USGS analysis is based on ground water arsenic exceedance estimates for each county. The MWDSC study is based on survey results from 140 large public water systems across 10 USEPA regions. The Wade Miller study occurrence results are based on NIRS data, which are occurrence results from a stratified, nationally-representative, random survey (defined by system size) of approximately 980 public water systems served by ground water.

Figure 6-1a presents arsenic occurrence estimates in surface water systems derived in this study (for CWS systems) compared to results from the MWDSC and NAOS studies at concentrations of 2, 5, 10, and 20 µg/L. Figure 6-1b presents a comparison of arsenic occurrence estimates from this study for ground water systems with the study findings of MWDSC, NAOS, USGS, and Wade Miller at concentrations of 2, 5, 10, and 20 µg/L.

Figure 6-1a

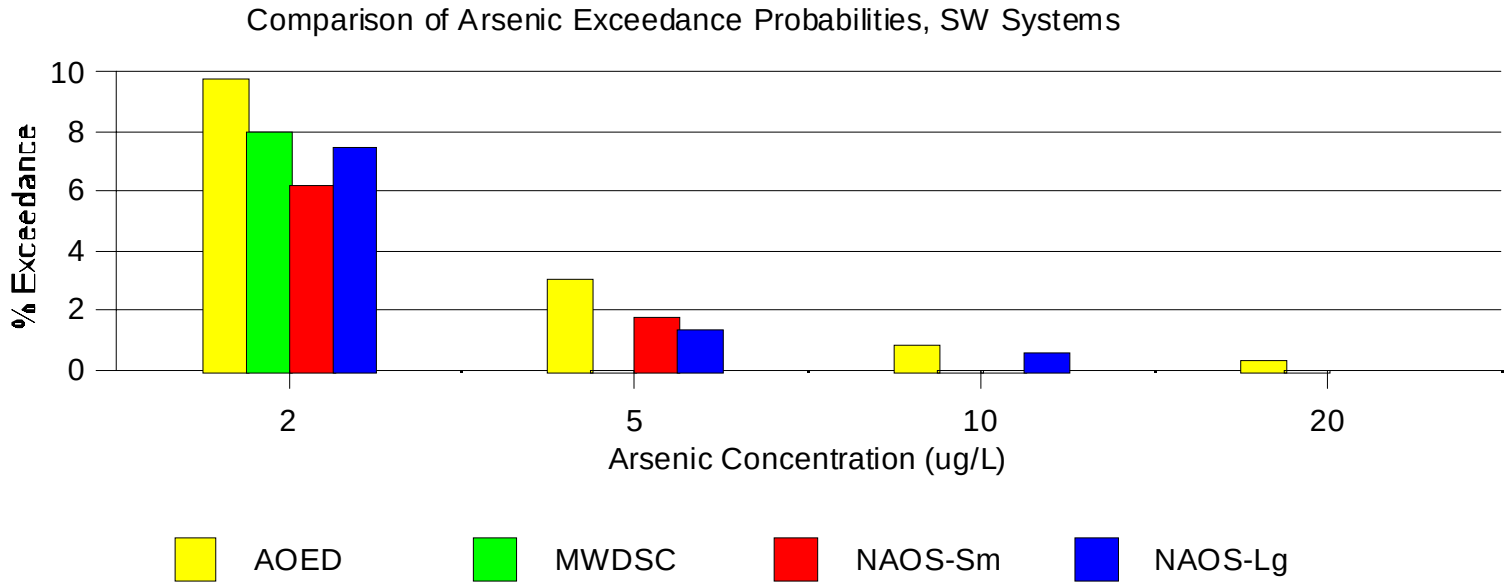
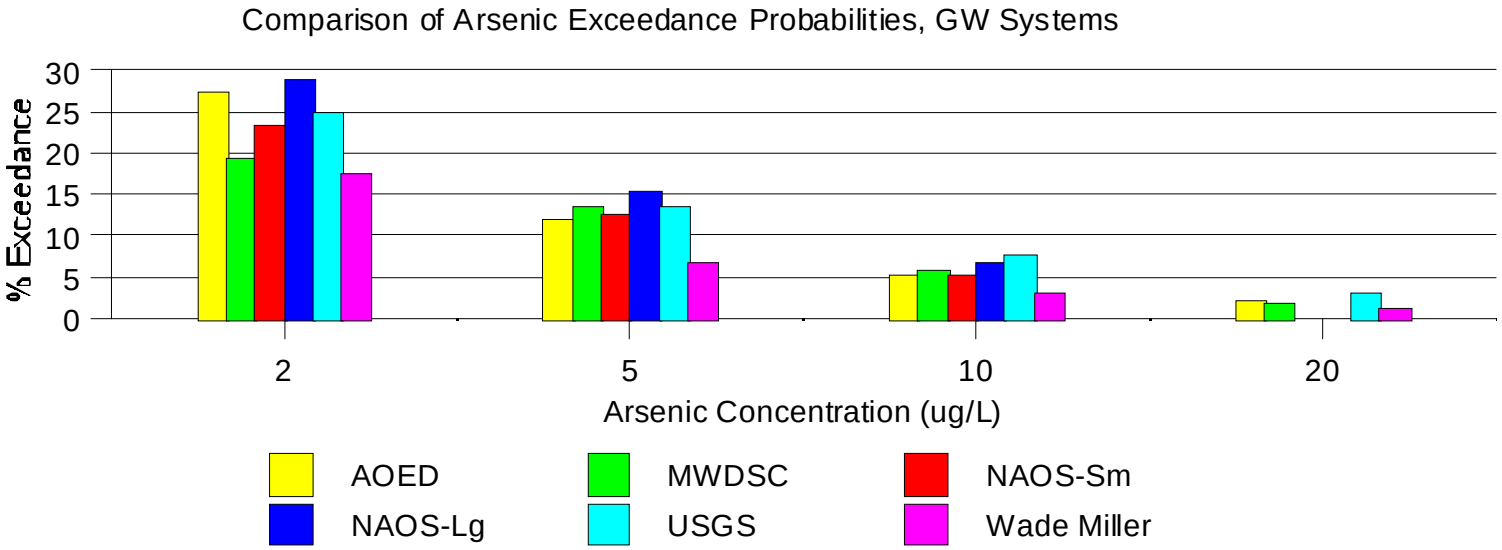

Comparison of Arsenic Exceedance Probabilities, SW Systems
% Exceedance
10
8
6
4
2
0
2
5
10
20
Arsenic Concentration (ug/L)
AOED
MWDSC
NAOS-Sm
NAOS-Lg

Figure 6-1b


Comparison of Arsenic Exceedance Probabilities, GW Systems
% Exceedance
30
25
20
15
10
5
0
2
5
10
20
Arsenic Concentration (ug/L)
AOED
MWDSC
NAOS-Sm
NAOS-Lg
USGS
Wade Miller

The values associated with the surface water findings shown graphically in Figure 6-1a are presented in Table 6-8a. Table 6-8a also includes 95% confidence intervals for the AOED estimates that are developed in section 6-4. At arsenic concentrations of 2 µg/L, exceedance estimates in AOED are similar to, though somewhat higher than, those for NAOS-small systems, NAOS-large systems, and MWDSC estimates. At 5 µg/L, the AOED, NAOS-small, and NAOS-large exceedance estimates are similar (3.0, 1.3, and 1.8, respectively), and somewhat above the MWDSC estimate (0.0 percent). At 10 µg/L, the AOED and NAOS-large exceedance estimates are similar (0.8 and 0.6 percent, respectively), and above the NAOS-small and MWDSC estimates (both at 0.0 percent). At 20 µg/L, there are no reported data for NAOS, and AOED and MWDSC estimates are low and similar (0.3 and 0.0 percent, respectively).

The values associated with the ground water findings shown graphically in Figure 6-1b are presented in Table 6-8b. Table 6-8b also includes 95% confidence intervals for the AOED estimates that are developed in section 6-4. At arsenic concentrations of 2 µg/L, the exceedance estimates for AOED are similar to the NAOS-small, NAOS-large, and the USGS estimates (with the four estimates ranging from 28.8 percent for NAOS-large to 23.5 percent for NAOS-small), with MWDSC and Wade Miller estimates somewhat lower. At 5 µg/L, the AOED, NAOS-small, NAOS-large, MWDSC, and USGS estimates are similar, ranging from 15.4 percent (NAOS-large) to 12.1 percent (AOED), with the Wade Miller estimate at 6.9 percent. At 10 µg/L, the AOED, NAOS-small, NAOS-large, MWDSC, and USGS estimates again are similar, ranging from 7.6 percent (USGS) to 5.1 percent (NAOS-small), with Wade Miller estimates at 2.9 percent. And at 20 µg/L, there are no reported data for NAOS, and the AOED, MWDSC, and USGS exceedance estimates again are similar, ranging from 3.1 percent (USGS) to 1.9 percent (MWDSC), with Wade Miller study results at 1.1 percent. At all arsenic concentrations of interest, the Wade Miller (1992) estimates are lower than those based on AOED, NAOS, MWDSC, and USGS. The Wade Miller estimates relied upon NIRS data for ground water (and 1978 CWSS, NOMS and RWS data for surface water). These data were highly censored, and had relatively high reporting limits (generally 5 µg/L). Therefore, these estimates are probably less accurate than the estimates from the other studies and databases, particularly at concentrations of 2 and 5 µg/L.

On the national level, at arsenic concentrations ranging from 2 to 10 µg/L, exceedance estimates are similar for the AOED, NAOS (small and large systems), MWDSC, and USGS studies. In addition, for the studies with data available (AOED, MWDSC, and USGS), exceedance estimates are also similar at arsenic concentrations of 20 and 50 µg/L. The fact that these studies are in general agreement, especially given that each study used different data sets and different methods for calculating exceedance estimates, indicates that the arsenic occurrence estimates presented in Section 6.2 are reasonably representative at the national level. This comparison of exceedance probabilities across studies suggests that arsenic occurrence projections based on compliance monitoring data are relatively close to other recently developed projections through the range of this comparison.

Table 6-8a
Comparison of AOED, MWDSC, and NAOS
Surface Water Arsenic Occurrence Estimates

	Percent of Systems with Mean Arsenic Concentrations Exceeding Specific Limits (µg/L):				
Study	**2**	**5**	**10**	**20**	**50**
AOED	9.8	3.0	0.8	0.3	0.1
AOED 95% Conf. Int.	(9.1, 25.4)	(1.8, 9.7)	(0.5, 1.6)	(0.1, 0.8)	(0.0, 0.6)
MWDSC	8.0	0.0	0.0	0.0	0.0
NAOS-Sm	6.2	1.8	0.0	NR	NR
NAOS-Lg	7.5	1.3	0.6	NR	NR

Note: NAOS-Sm includes systems serving ≤10,000 people, NAOS-Lg includes systems serving > 10,000 people. NR = Not Reported.

Table 6-8b
Comparison of AOED, MWDSC, NAOS, USGS and Wade Miller
Ground Water Arsenic Occurrence Estimates

	Percent of Systems with Mean Arsenic Concentrations Exceeding Specific Limits (µg/L):				
Study	**2**	**5**	**10**	**20**	**50**
AOED	27.3	12.1	5.3	2.0	0.4
AOED 95% Conf. Int.	(26.6, 29.9)	(11.7, 13.0)	(5.2, 5.9)	(1.9, 2.3)	(0.4, 0.5)
MWDSC	19.2	13.5	5.8	1.9	0.0
NAOS-Sm	23.5	12.7	5.1	NR	NR
NAOS-Lg	28.8	15.4	6.7	NR	NR
USGS	25.0	13.6	7.6	3.1	1.0
Wade Miller	17.4	6.9	2.9	1.1	0.2

Note: NAOS-Sm includes systems serving ≤10,000 people, NAOS-Lg includes systems serving > 10,000 people. NR = Not Reported.

Comparisons of estimated exceedances of arsenic concentrations at the regional level (for ground water systems) are presented in Figure 6-2 (for 5 µg/L) and Figure 6-3 (for 20 µg/L). Figure 6-2 shows that the AOED exceedance probabilities at 5 µg/L are lower than the USGS exceedance probabilities in all regions except New England (Region 1). Also, the AOED estimates at 5 µg/L are lower than NAOS estimates in all regions except New England (Region 1) and North Central (Region 6). The three studies agree reasonably well regarding exceedance probabilities in the Midwest (Region 4) and South Central (Region 5) Regions. All three studies indicated that the lowest exceedance probabilities occurred in the South East Region (Region 3), and that the highest exceedance percentages at a concentration of 5 µg/L occur in the West Region (Region 7).

A similar comparison is presented in Figure 6-3 for regional exceedance probabilities relative to the arsenic concentration of 20 µg/L . These probabilities are significantly lower than at 5 µg/L. The results of the three studies appear to agree less well at 20 µg/L than at 5 µg/L. As at 5 µg/L, the South East Region (Region 3) shows low exceedances probabilities at 20 µg/L, and the West Region (Region 7) shows higher exceedance probabilities than most of the other regions at 20 µg/L. The highest exceedance probabilities at 20 µg/L were indicated to be in the North Central Region (Region 6) according to the USGS study, but the other studies did not support this finding.

For surface water systems, USGS did not develop exceedance estimates, and NAOS estimated that surface water systems in only two regions would have arsenic concentrations above 5 µg/L in finished surface water. These Regions include the South Central (Region 5), where 7 percent of systems were predicted to exceed 5 µg/L, and the North Central (Region 6), where 12 percent of systems were predicted to exceed concentrations of 5 µg/L. Surface water exceedance estimates based on AOED are presented in Table 6-2. Based on these estimates, some systems would have arsenic concentrations above 5 µg/L in six Regions (exceedance probabilities in parentheses): New England (9 percent); Mid Atlantic (0.1 percent); Midwest (1.2 percent); South Central (1 percent); North Central (3.8 percent), and West (7.4 percent). AOED-based estimates indicate some systems will exceed arsenic concentrations of 20 µg/L in New England (0.4 percent); Midwest (0.1 percent); North Central (0.1 percent); and the West (1.1 percent).

6.3.2 Comparison of AOED, Kennedy-Jenks, and Saracino-Kirby Occurrence Estimates for California

In addition to the occurrence results presented above, two additional studies have developed estimates for arsenic occurrence in California. The additional studies are a Kennedy-Jenks Consultants report of the cost of compliance prepared for the Association of California Water agencies (ACWA, 1996), and the Saracino-Kirby, Inc., report on arsenic occurrence and conjunctive use in California (ACWA, 2000). The two studies used similar, though not identical, sources of data. For the Kennedy-Jenks report, data sources included the ACWA Low-Level Arsenic Database, data from the California Department of Water Resources (DWR) and the Department of Heath Services (DHS), and California-specific data from the USGS. For the

Figure 6-2

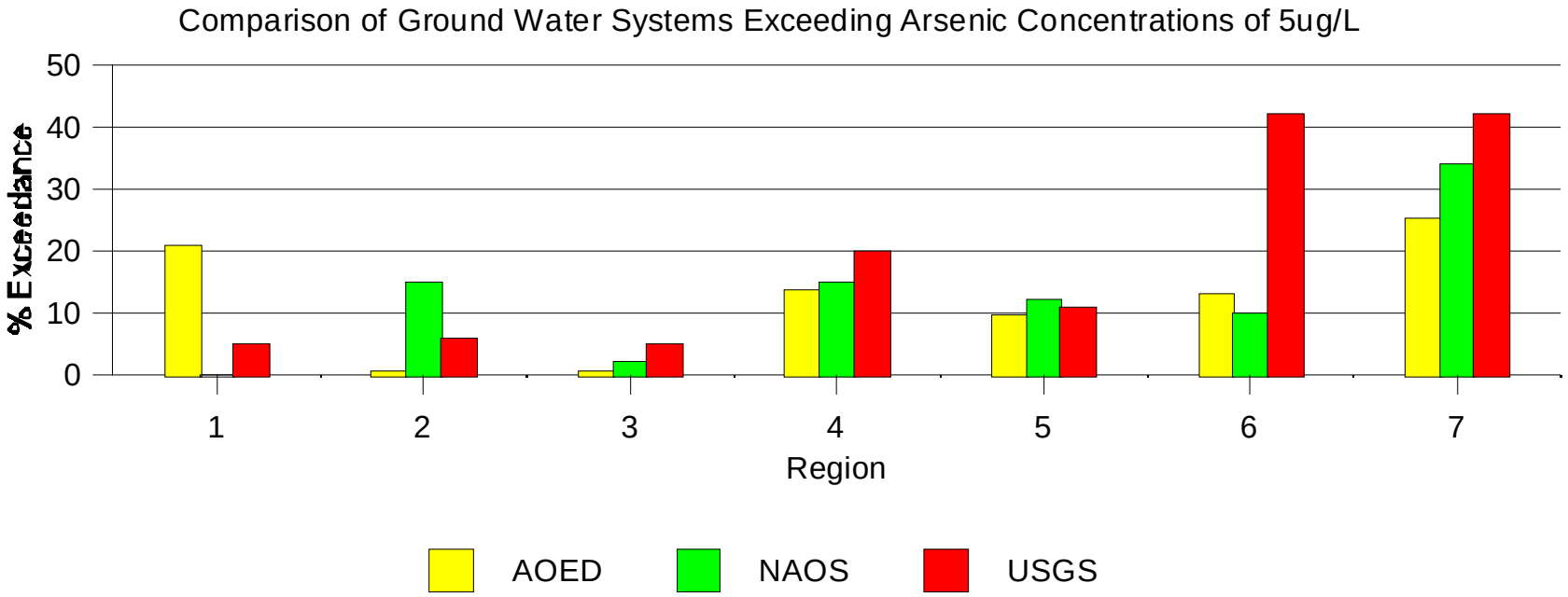
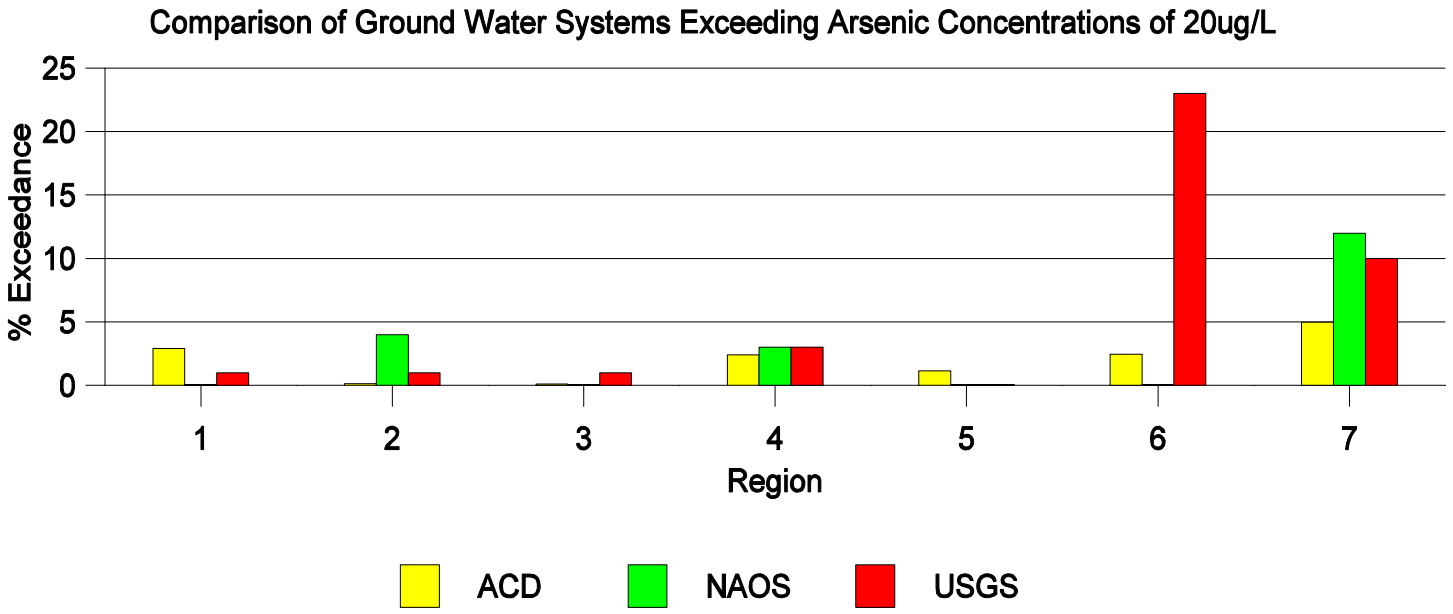

Comparison of Ground Water Systems Exceeding Arsenic Concentrations of 5ug/L
% Exceedance
50
40
30
20
10
0
1
2
3
4
5
6
7
Region
AOED
NAOS
USGS

Figure 6-3


Comparison of Ground Water Systems Exceeding Arsenic Concentrations of 20ug/L
% Exceedance
25
20
15
10
5
0
1
2
3
4
5
6
7
Region
ACD
NAOS
USGS

Saracino-Kirby report, data sources included the DWR, the DHS, California-specific data from the USGS, data from the USGS stream water quality monitoring network, and data from the Sacramento River Trace Metals Study.

The Saracino-Kirby study data included both raw and treated water sample results for ground water occurrence data, but used only raw water sample analytical results for surface water occurrence data. In the Kennedy-Jenks study, the type of water sample results included in the data (raw versus treated) was not clearly defined. Though the report's use of USGS data and other references suggest the data represent raw water, other references (such as results reported from "surface water plants") imply the use of treated water sample results.

Figure 6-4a presents arsenic occurrence estimates in surface water systems from California derived in this study (see Appendix B-1) compared to results from the Kennedy-Jenks (ACWA, 1996) and Saracino-Kirby (ACWA, 2000) at concentrations of 2, 5, 10, 20, and 50 µg/L. Figure 6-4b presents this same comparison for arsenic occurrence estimates in ground water. For surface water arsenic occurrence, generally the Saracino-Kirby results are higher than the AOED results which, in turn, are higher than the Kennedy-Jenks results. For ground water arsenic occurrence, this same general pattern is apparent, although the results of the AOED and Kennedy-Jenks are in close agreement (especially for arsenic values greater than or equal to 5 µg/L). The values associated with the surface water findings shown graphically in Figure 6-4a are presented in Table 6-9a, and the values associated with the ground water findings in Figure 6-4b are presented in Table 6-9b.

6.4 UNCERTAINTY ANALYSIS

The uncertainty analysis described in this section was used to develop the confidence intervals presented in Section 6.2. Due to project constraints, these estimates were based on an earlier version of the database than the one used to develop the occurrence estimates. However, the similarity of the estimated arsenic occurrence distributions for the two databases suggests that the earlier confidence intervals can be used as a reasonable approximation to the expected results from a revised simulation analysis. Confidence intervals were only computed for CWS systems.

6.4.1 Purpose of Uncertainty Analysis

An uncertainty analysis was conducted to determine the potential amount of error in the exceedance probability estimates that are presented in Section 6.2. The ROS method that was used to estimate system means has a potential drawback, in that it does not allow the calculation of confidence intervals in a straightforward manner. Therefore, to determine 95-percent confidence intervals, it was necessary to perform a statistical simulation to quantify the potential sources of uncertainty. Three sources of uncertainty were identified and simulated: 1) sampling variability, both within and between systems; 2) the fill-in of censored observations in the estimation of system means; and 3) fitting of lognormal distributions to populations of system means within each State. The combined effect of these uncertainties was modeled through a simulation, and the results of this simulation were used to establish confidence intervals for the arsenic occurrence estimates.

Figure 6-4a

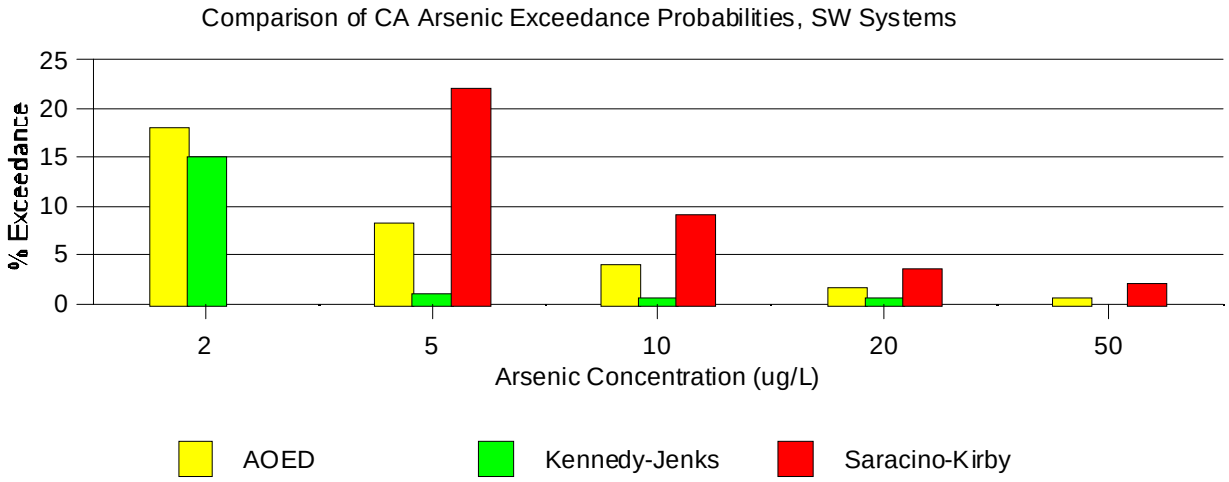


Figure 6-4b

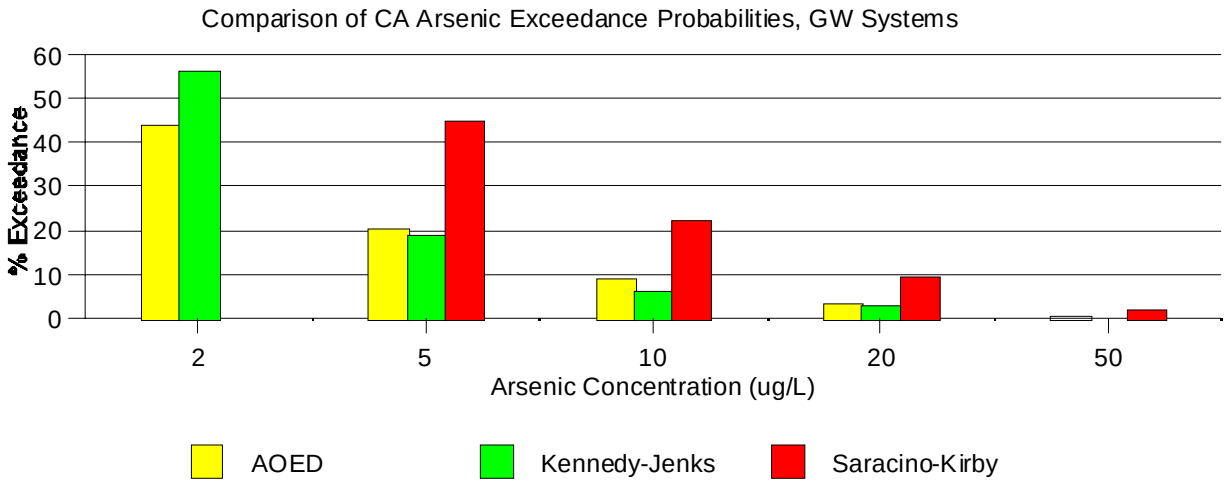


Table 6-9a
Comparison of California Arsenic Occurrence Estimates from AOED, Kennedy-Jenks, and Saracino-Kirby Studies -- Surface Water

	Percent of Systems Estimated to Exceed Arsenic Concentrations (μg/L):				
Study	**2**	**5**	**10**	**20**	**50**
AOED	18.1	8.2	4.0	1.7	0.5
Kennedy-Jenks	15	1	< 1	< 1	N/A
Saracino-Kirby	N/A	22	9	3.5	2.0

Table 6-9b
Comparison of California Arsenic Occurrence Estimates from AOED, Kennedy-Jenks, and Saracino-Kirby Studies -- Ground Water

	Percent of Systems Estimated to Exceed Arsenic Concentrations (μg/L):				
Study	**2**	**5**	**10**	**20**	**50**
AOED	43.8	20.5	9.2	3.3	0.6
Kennedy-Jenks	56	19	6	3	N/A
Saracino-Kirby	N/A	45	22	9.5	1.9

The purpose of this uncertainty analysis is to provide a conservative estimate of the total uncertainty attributable to the effects of sampling variability, the fill-in methods for censored data, and the lognormality assumption. For this reason, the uncertainty analysis was deliberately designed to increase the variability of the estimates. For example, the uncertainty analysis lessened the number of detects required for the ROS method (instead of substitution) from 5 or more down to 2 or more. Also the censored values were randomly substituted from a probability distribution instead of using fixed values. However, there are various other sources of uncertainty not addressed by the uncertainty analysis. Such sources include measurement error, the use of regional exceedance distributions (which effectively assumes homogeneity within each region), and uncertainties in the SDWIS database.

6.4.2 Uncertainty Analysis Methodology

Briefly summarized, the uncertainty simulation first simulated a population of systems, then a mean for each system, and then estimated State exceedance probabilities based on the simulated population of system means. Sets of exceedance probabilities obtained from many repetitions of the simulation were then used to estimate non-parametric confidence intervals for the concentrations of interest. In other words, bootstrap confidence intervals (Davison and Hinkley, 1997) were computed for the estimation procedure described in Section 6.1. To evaluate the influence of the lognormal model used to fill in censored observations, the entire simulation was repeated twice: the first time, a uniform distribution was used to fill in censored observations, and the second time, a Weibull distribution was used. Also, in order to evaluate the influence of the right-tailed ROS method used to estimate the State exceedance probabilities, the probabilities were computed using both ROS and an empirical method.

As mentioned above, the first step in the uncertainty analysis was to simulate a population of systems. Systems were sampled with replacement from the list of systems in each State. The second step was to simulate a system mean for each selected system. In this step, the combined set of detected and non-detected concentrations in each system was re-sampled, such that if the system had d detected and c non-detected concentrations, a random sample of size $d + c$ was selected at random with replacement from the $d + c$ observations. The calculation involves real-space and log-space means and variances. The real-space mean and variance are the mean and variance of the untransformed concentration values. The log-space mean and variance are the mean and variance of the logarithms of the concentration values. When two or more distinct detected concentrations were drawn, the log-space mean and variance were estimated by the ROS method described in Appendix A; non-detects were filled in by drawing from a truncated lognormal distribution with the given log-space mean and variance, such that the filled-in values fell in the range from zero to the reporting level; and a real-space mean was computed for the system. When all detects were equal or there was only one detect, each non-detect was filled in by drawing from a uniform distribution from zero to the reporting level, and the real space system mean was calculated. When there were no detects, the system was treated as a non-detect at the modal reporting level.

Using the above procedure, a population of system means was generated for each State. From these populations, national exceedance probability distributions were then computed exactly as before: State exceedance probability distributions were estimated by the right-tailed ROS method, at concentrations of 2, 3, 5, 10, 15, 20, 25, 30, 40, and 50 μg/L; State distributions were combined into regional distributions; and regional distributions were combined into a national distribution, as described in Sections 6.1.2–6.1.4.

The entire procedure above was repeated 1,000 times, in order to generate 1,000 simulated national exceedance probability distributions. At each concentration (e.g., at 10 μg/L), the 1,000 exceedance probability estimates were then sorted in increasing order, and the interval from the mean of the 25th- and 26th-largest estimates to the mean of the 975th- and 976th-largest estimates was taken as a nonparametric 95% confidence interval for the true exceedance probability at that concentration.

The confidence intervals described above quantify the uncertainty in the exceedance probability estimates due to the sampling variability of detected arsenic concentrations. They do not include uncertainty due to the use of the lognormal distribution to fill in the censored observations. In order to evaluate this additional uncertainty, the confidence intervals were recomputed twice, using two alternative distributions to fill in the censored observations: the two-parameter Weibull (a long-tailed) distribution and the uniform (a flat) distribution. In the first repeat of the simulation, censored values that were previously filled in by drawing from a lognormal distribution were instead drawn from a uniform distribution on the interval from zero to the reporting level. In this case it was not necessary to fit the distribution. In the second repeat of the simulation, censored values were replaced by draws from a truncated Weibull distribution, with parameters estimated by a censored maximum likelihood algorithm in SAS. In systems with only one detect or all detects equal, the maximum likelihood algorithm could not be meaningfully applied, so non-detects were drawn from a uniform distribution. From each of the uniform and Weibull simulations, 1,000 replicates were independently generated and confidence intervals were computed as above.

Another source of uncertainty in the probability estimates is the use of the right-tailed ROS method, with its lognormal assumption, to estimate State probability distributions. In order to evaluate this uncertainty, exceedance probabilities and their confidence intervals were computed using both the right-tailed ROS method, as described above, and an empirical method. The empirical method estimates exceedance probabilities as the empirical fraction of detected concentrations above each concentration level. Empirical estimates are computed only at concentrations of 10 *m*g/L or higher (since the highest censoring limit in the AOED is 10 *m*g/L) so that all non-detects may be unambiguously counted as less than the concentrations of interest.

6.4.3 Uncertainty Analysis Results

This discussion focuses on the national ground water and surface water uncertainty analysis results. Uncertainty analyses were conducted for lognormal, uniform and Weibull distributions, using both right-tailed ROS and empirical methods. The results of the right tailed ROS uncertainty analysis for the national exceedance probabilities using a lognormal distribution are found in Table 6-3a. The 95 percent confidence intervals are not centered directly on the

point estimates for the national ground water and surface water exceedance probabilities. Rather, the point estimates are located closer to the lower limits of the 95 percent confidence intervals. In a comparison of the uncertainty analysis results from the three distributions, some variation is evident among the distributions. In particular, the percentages for the lognormal distributions are slightly higher than for the uniform or Weibull distribution at most concentrations. While there are some differences among the confidence intervals for the lognormal, Weibull and uniform distributions, these differences are minor. One exception is the surface water CIs at 5 mg/L. Overall, these differences suggest that the analysis results are not particularly sensitive to the method used to fill in the censored observations. Figures 6-4 and 6-5 depict plots of the 95 percent confidence intervals from the three distributions for ground water and surface water at selected concentrations. While the widths of the CIs from the different distributions are relatively similar, some minor variations are seen. For example, for the ground water analysis, both the lognormal and Weibull CIs are wider than the uniform distribution CI. For surface water, at low concentrations, the CI widths from the log normal distribution fall between those for the Weibull and uniform CIs, but at higher concentrations are greater than the CI widths of the Weibull and uniform distributions. All the CI widths for both surface water and ground water narrow with increasing concentration. The 95 percent confidence intervals from the right-tailed ROS lognormal distribution analysis were used to calculate the confidence intervals presented in Tables 6-4, 6-5, and 6-7.

A comparison of the uncertainty analysis results for ground water and surface water reveals different widths of the 95 percent confidence intervals. The national surface water confidence intervals are significantly wider than the intervals for ground water exceedance probabilities. For example, the CI at 5 mg/L for ground water is [11.74, 13.04], while the CI for surface water at 5 mg/L is [1.8, 9.66]. The increased width is due to the smaller amount of data available from surface water systems and the increased level of censoring of the surface water data. Consequently, the results of the surface water uncertainty analysis are more sensitive to the analytical methods employed.

In this uncertainty analysis, both a right-tailed ROS and an empirical analysis were conducted for all distribution types. The results of the two methods differ. The empirical percentages are lower than those for the right-tailed ROS analyses, for both surface water and ground water. For ground water, the 95 percent confidence intervals from the empirical analysis are wider than those from the right-tailed ROS analysis, for all distribution types. However, for surface water, the confidence intervals for the right-tailed ROS method are wider than those from the empirical analysis. In the log normal distribution, for both methods, the CIs become increasingly narrow at higher concentrations. The lognormal results of these two analyses for ground water and surface water at selected concentrations are presented in Figures 6-6 and 6-7.

Figure 6-4 Comparison of Ground Water 95% Confidence Intervals by 3 Distribution Types

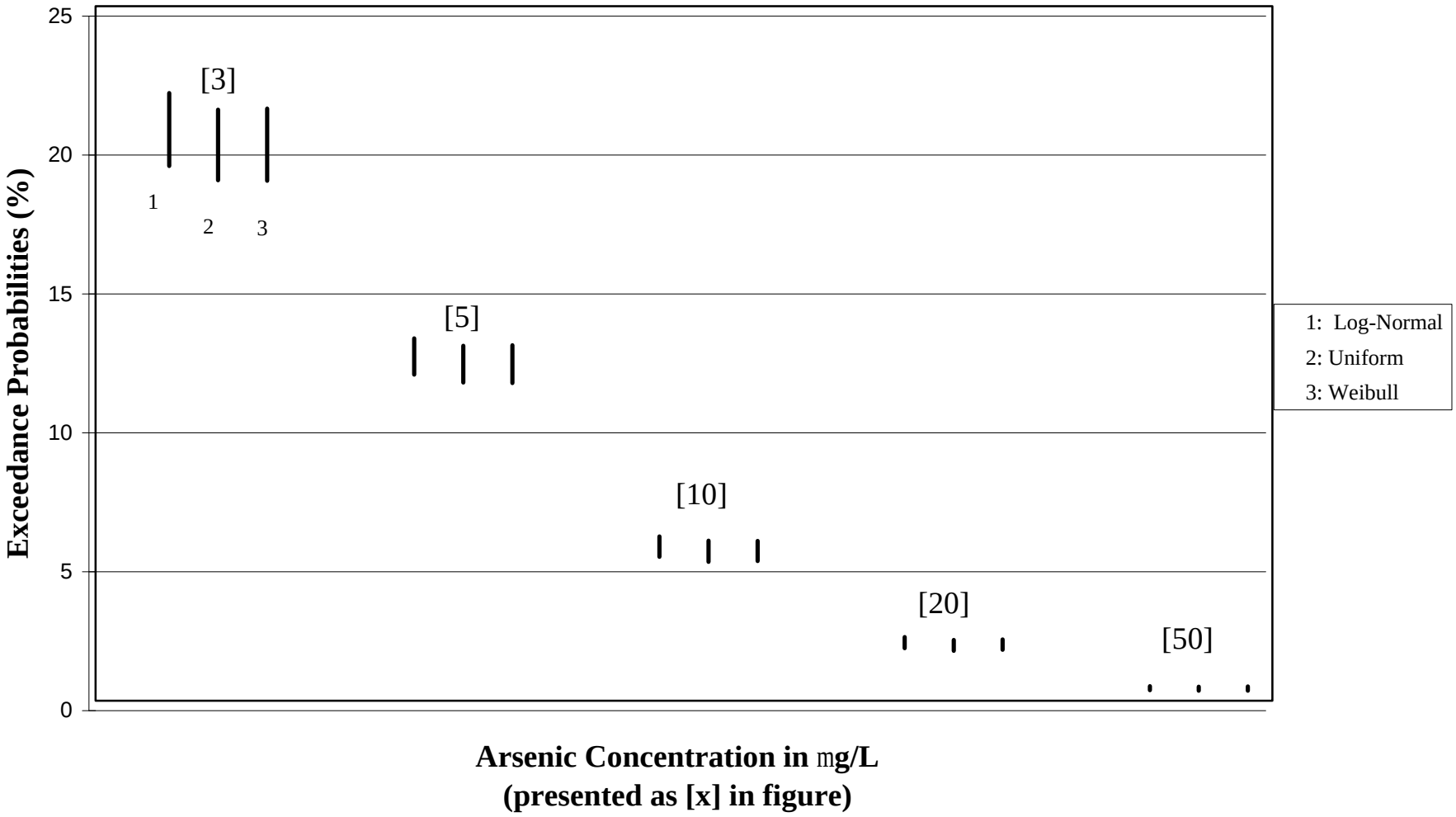


Figure 6-5 Comparison of Surface Water 95% Confidence Intervals by 3 Distribution Types

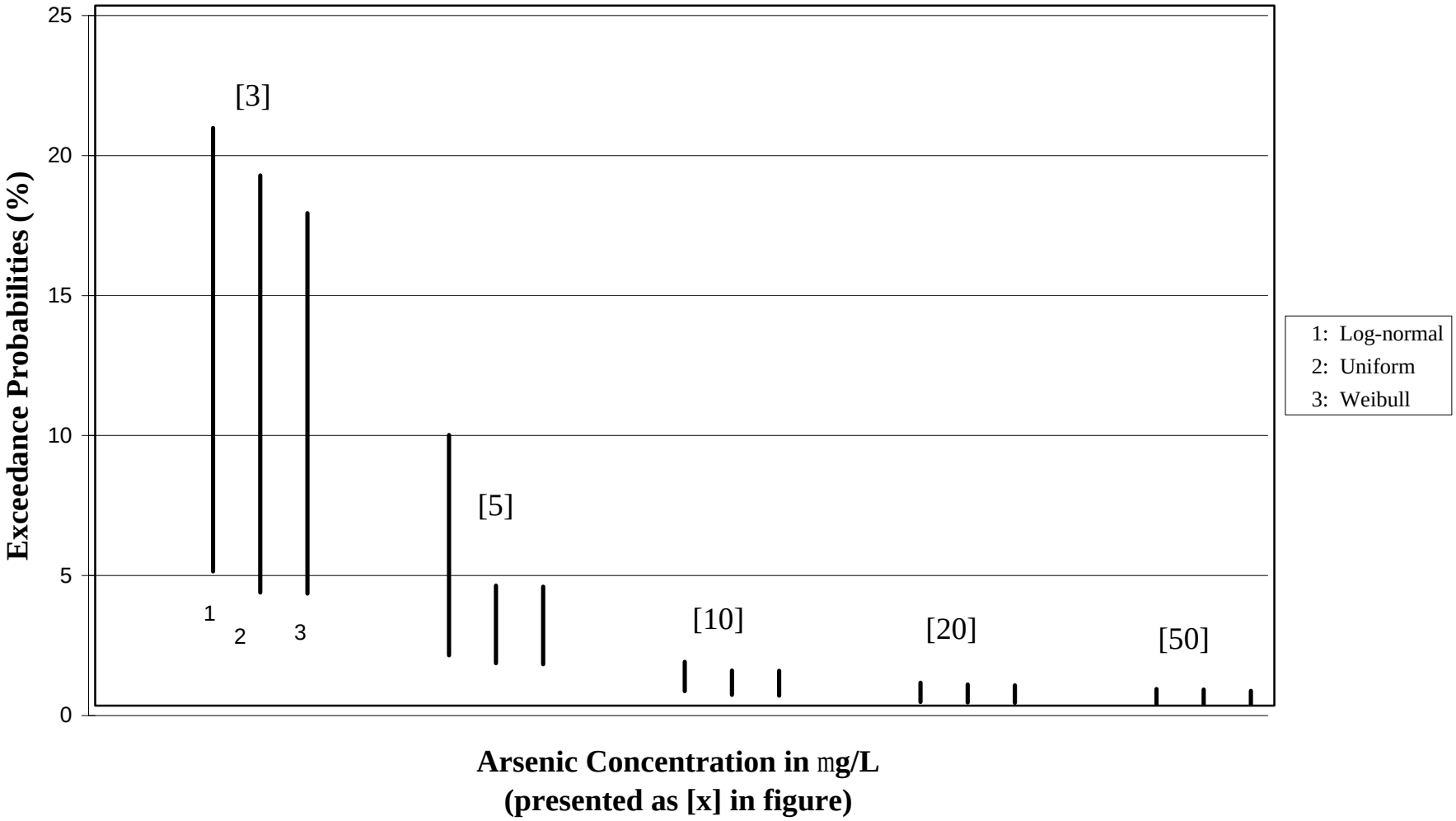


Figure 6-6 Comparison of Ground Water 95% CI by Right-tailed ROS vs. Empirical Methods

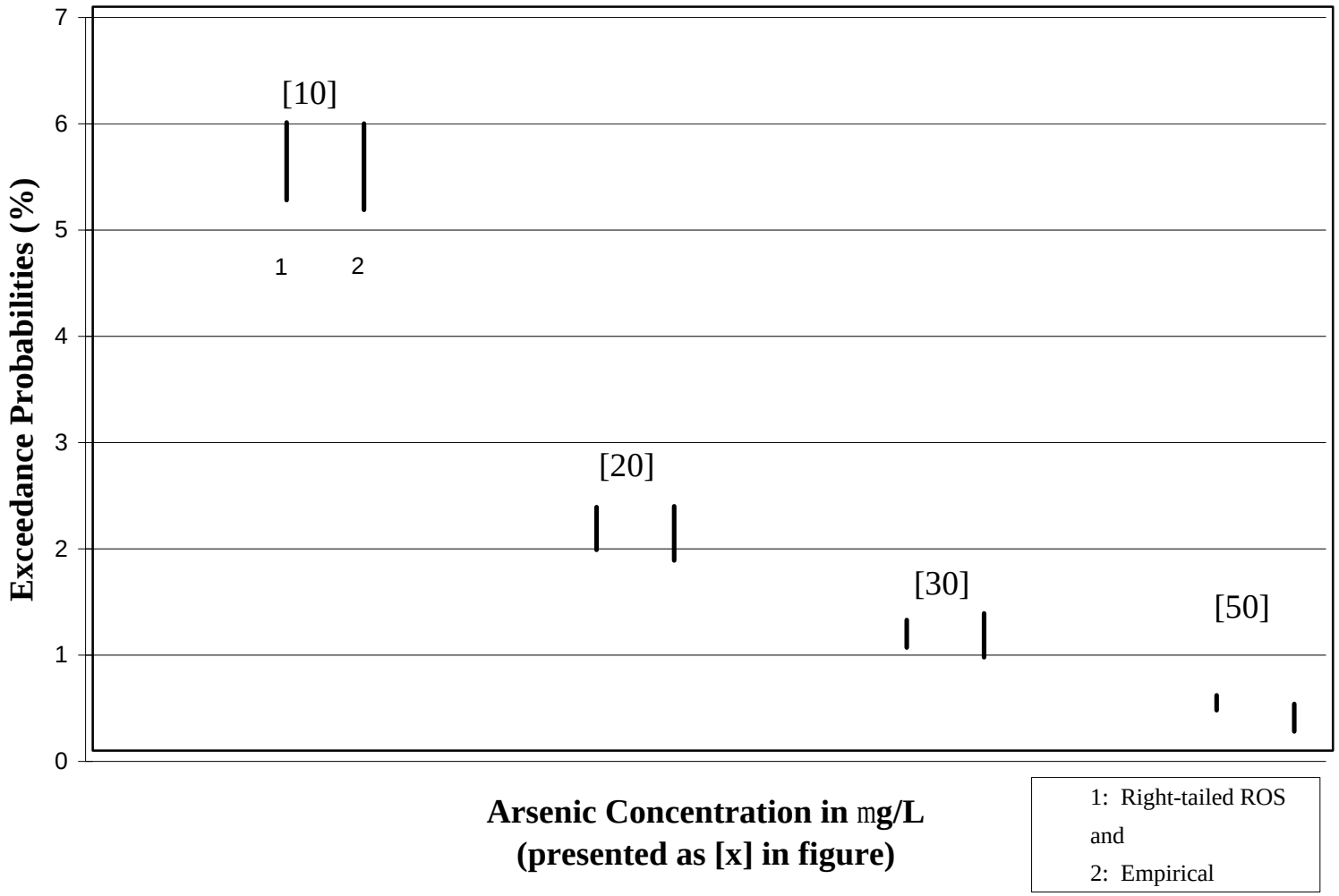
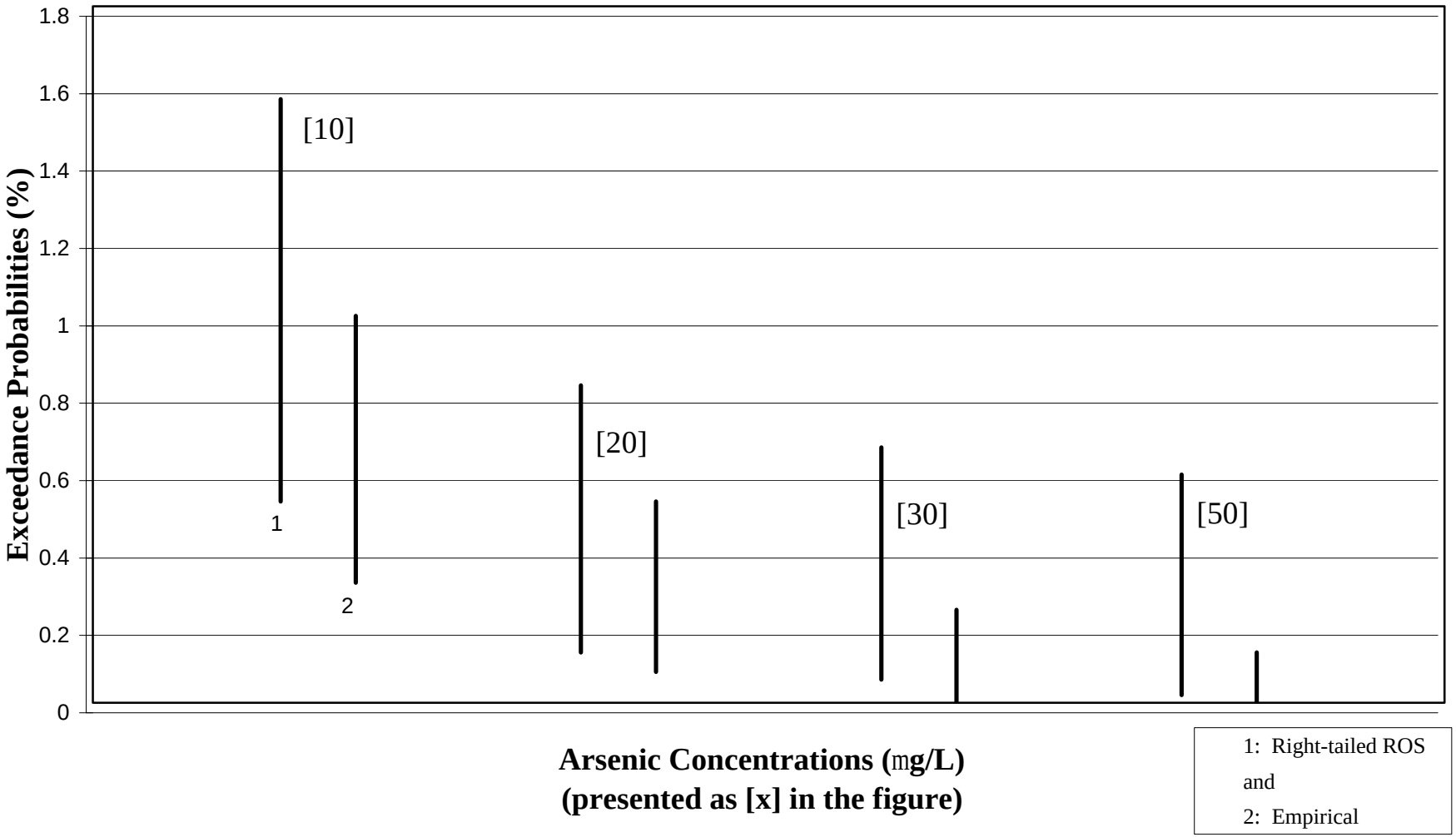


Figure 6-7 Comparison of Surface Water 95% Confidence Intervals by Right-tailed ROS vs. Empirical Method



7. Intra-system Variability

This Chapter presents the results of analyses of intra-system variability that were conducted using subsets of the AOED data. Section 7.1 defines the purpose of the intra-system variability analysis. Section 7.2 provides an overview of the data that were available for these analyses. Section 7.3 presents the methods by which the intra-system variability analyses were conducted, and the results of the analysis. Section 7.4 briefly summarizes the intra-system variability analyses.

7.1 Purpose of Analyses

The purpose of the intra-system analysis is to facilitate prediction of the number of points-of-entry or POE that will be affected by various MCL alternatives. Compliance with the arsenic standard is measured at the point-of-entry to the distribution system, and individual systems can have multiple points-of-entry. Thus, one system may need to install one, two, three, or more treatment systems or blend its water sources, depending upon its configuration and POE mean arsenic levels. If arsenic levels in all POE in a system are below regulatory limits, it will not need to install any treatment technologies for arsenic. Thus, arsenic levels in POE drive compliance costs and risk reduction benefits more directly than do system mean arsenic levels.

The ideal analysis would be a survey designed to estimate arsenic levels in POE throughout the United States. However, data are not currently available to support development of such estimates: SDWIS does not catalogue information at the POE level; the earlier arsenic occurrence studies focused on arsenic concentrations in systems rather than POE; and only a third of the data sets in AOED link sample results to POE identification numbers. So the ideal analysis is currently infeasible.

Since the ideal is infeasible, a reasonable and feasible alternative is to quantify a relationship between POE means and system means where the data are suitable, and to use this relationship, to estimate, from a population of system means, the number of POE means that are likely to exceed specific regulatory alternatives. The analyses discussed in this Chapter were designed to quantify the relationship between system means and POE means. This relationship was quantified as an estimated coefficient of variation (CV), or relative standard deviation. This is the CV of the distribution of the POE means for any given system. The CV values that were calculated under these analyses are being applied under a different work assignment in a RIA to estimate the number of POE that may exceed MCL alternatives. Under that work assignment, the distribution of system mean arsenic concentrations, the probability distributions of POE for systems of different sizes, and the CV of the relationship between the system mean arsenic concentration and the POE arsenic concentration means are used to estimate the number of POE in the United States that may exceed specific arsenic concentrations.

7.2 Available Data

To evaluate the relationship between POE mean arsenic levels and system mean arsenic levels, it is necessary to have data sets that include distinct POE identifiers that associate each

sample with the POE where it was collected. As indicated in Chapter 4, a total of ten States provided data sets that included suitable POE identifiers:

- Alabama
- Arkansas
- California
- Indiana
- Illinois
- New Mexico
- North Carolina
- Oklahoma
- Texas
- Utah

Since the purpose of the intra-system variability analyses was to estimate variability between POE, only those systems with two or more POE were used for these analyses. The variable SRC_ID in the AOED INTRA database identifies the POE within each system, and a unique POE identifier is the combination of PWSID and SRC_ID. A second restriction was that all systems where there were any completely censored POE were not used. A completely censored POE is a POE for which all the measurements are below the reporting limit. A third restriction was to exclude data from surface water NTNCWS systems; the limited data for that group of systems precluded the development of independent estimates of occurrence and intra-system variability for surface water NTNCWS systems. Table 7-1 shows the numbers of samples, non-detects, systems, and POE for systems with at least two POE that were used for the intra-system variability analyses. This data subset contains a total of 4390 samples from 638 ground water CWS systems, 542 samples from 32 surface water CWS systems, and 237 samples from 49 ground water NTNCWS systems. Note that none of the systems in Arkansas and only one system in Alabama met the requirements for inclusion in these analyses. Estimates of intra-system variability that were developed using these data are described in the following sections of this report.

7.3 Analytical Methods and Results

The methods used to estimate intra-system variability was based on a fitted statistical model that represented the two sources of variability for multiple measurements on the same system: intra-system variability and within-POE variability. The intra-system variability is the variability between POE for the same system. The within-POE variability is the temporal and analytic variability between multiple measurements taken at the same POE. The first step of the analysis was to estimate the concentrations for non-detects using the ROS method applied to the data for each POE. This is described in section 7.3.1. The statistical model formulation and the results of fitting that model are described in sections 7.3.2 and 7.3.2.

Table 7-1
Data Used for Intra-system Variability Analyses

Type of Water System	Source Type	States	Number of Systems	Numbe of POE	Number of Samples	Number of Non-Detects
CWS	GW	California	208	890	2257	479
CWS	GW	Illinois	22	54	65	0
CWS	GW	Indiana	1	6	7	0
CWS	GW	North Carolina	5	9	22	12
CWS	GW	New Mexico	194	655	1120	75
CWS	GW	Oklahoma	51	145	183	3
CWS	GW	Texas	103	241	402	34
CWS	GW	Utah	54	153	334	59
CWS	GW	All	638	2153	4390	662
CWS	SW	Alabama	1	4	95	89
CWS	SW	California	5	21	209	84
CWS	SW	Illinois	1	2	4	1
CWS	SW	New Mexico	5	11	22	3
CWS	SW	Texas	16	37	169	69
CWS	SW	Utah	4	12	43	10
CWS	SW	All	32	87	542	256
NTNCWS	GW	California	14	28	53	5
NTNCWS	GW	New Mexico	25	125	151	2
NTNCWS	GW	Texas	6	13	19	0
NTNCWS	GW	Utah	4	10	14	2
NTNCWS	GW	All	49	176	237	9

7.3.1 Estimation of Concentrations for Non-Detects

Arsenic concentrations for non-detects (values below the reporting limit) were estimated using either substitution or the adapted ROS method described in Chapter 6. The actual values for concentrations above the reporting limit were used. The method that was applied to fill-in the non-detect concentrations depended on the number of detected values in the data set for the particular POE. The method used for estimating non-detects was the same as that used in Chapter 6, except that the adapted ROS method was separately applied to the data for each POE, instead of being applied to all the data for a system. As a result, arsenic concentrations for values below the reporting limit were estimated as follows:

- *1 - 4 detects for the POE:* Half the reporting level was substituted for non-detects.
- *All detects equal:* Half the reporting level was substituted for non-detects.
- *5 or more detects:* Adapted ROS was used to estimate the non-detect arsenic concentrations as follows: the detects for the POE were plotted as in Helsel and Cohn (1988); non-detects were uniformly plotted from zero to the estimated probability, from the fitted log-normal model, at which Y is less than or equal to the censoring level; the estimated concentration for each non-detect was estimated from the fitted log-normal distribution.

Note that every POE in the database used for the intra-system variability analyses had at least one detected value.

7.3.2 Log-Normal Mixed Model

The approach used to estimate the CV for intra-system variability in arsenic concentrations relies on a log-normal mixed model. The log-normal model assumes that for a given POE, in a given system, the logarithm of the arsenic concentration is normally distributed, with a log-mean that depends on the system and the POE, and a constant log-variance, V, that is the same for every system and POE. The terms “log-mean’ and “log-variance” are used to define the log-space mean and variance, i.e., the mean and variance of the logarithms of the concentrations. The log-variance V measures the within-POE residual variation, which represents the temporal and analytic variability of the multiple measurements for the same POE. Temporal variability is the variation between measurements made at different dates or times. Analytic variability is the variability between arsenic concentration measurements made by the same or different measuring instruments. The other part of the model is the intra-system variation between the POE, which is the main focus of the analysis. The statistical model assumes that for any given system, the POE log-means are normally distributed around the system log-mean. The variance of this distribution of the POE log-means summarizes the intra-system variability.

The statistical model used to describe the intra-system and within-POE variability therefore follows the equation:

Log(arsenic) = System Log-mean + POE Effect + Residual Error

where:

System Log-mean = the log-mean for the given system;

POE Effect = POE log-mean – System log-mean, which is normally distributed with a mean of zero and a variance Var(POE) = σ^2; and

Residual Error = Log(arsenic) – POE log-mean, which is normally distributed with a mean of zero and a variance Var(residual) = V.

The POE Effect is the intra-system variability and the Residual Error is the combined temporal and analytic variability. This model was fitted by the method of maximum likelihood estimation, separately for each combination of water system type and source type.

Under this statistical model, the POE means will be log-normally distributed with a log-variance of σ^2, and the CV of the intra-system variability is given by the equation:

CV(POE Mean) = $\sqrt{\exp(\sigma^2)-1}\times 100\,\%$.

7.3.3 Results

The estimated model coefficients and their standard errors are given in Table 7-2. The last two columns give the CV of the intra-system variability and its standard error. The estimated CVs are 37 % (standard error = 2 %) for ground water CWS systems, 53 % (standard error = 9%) for surface water CWS systems, and 25 % (standard error = 6 %) for ground water NTNCWS systems. For the surface water NTNCWS systems, the estimated CV for the surface water CWS systems was used.

Table 7-2
Results of Intra-System Variability Analyses

Type of Water System	Source Type	Intra-system Variability		Within-POE Variability		Intra-system Variability CV	
		Var (POE)	Standard Error	σ^2	Standard Error	CV	Standard Error
CWS	GW	0.1287	0.01274	0.9602	0.02352	37%	2%
CWS	SW	0.2446	0.07693	0.3545	0.02287	53%	9%
NTNCWS	GW	0.0614	0.0266	0.2064	0.02841	25%	6%
NTNCWS[1]	SW	0.2446	0.07693	0.3545	0.02287	53%	9%

[1]Intra-system variability estimates for surface water NTNCWS systems were copied from the estimates for surface water CWS systems.

7.4 Summary of Intra-system Analyses

A statistical model was used to estimate the intra-system variability of arsenic concentrations at different POE for the same system. This model accounts both for the intra-system variability between POE and for the temporal and analytic variability of the multiple measurements at the same POE. The results tabulated in Table 7-2 can be used in a RIA to estimate variability between POE. In conjunction with the arsenic national occurrence estimates given in Chapter 6, the intra-system variability estimates can be used to estimate distributions of POE mean arsenic concentrations in public water supply systems.

8. Temporal Variability

8.1 Purpose of Analysis

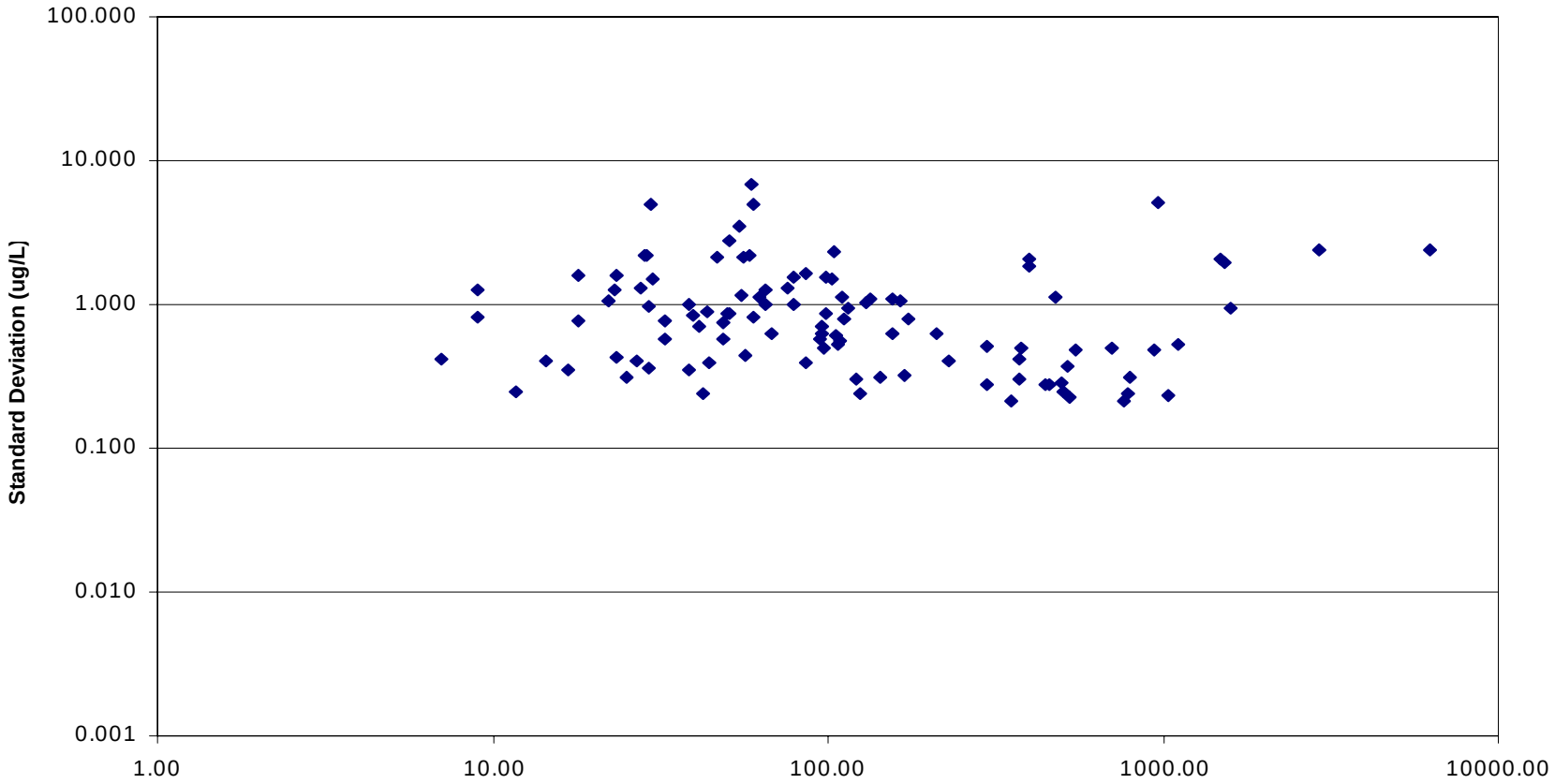
The purpose of the temporal variability analysis is to examine the variability of arsenic concentrations over time in a source. This information may be used in the Regulatory Impact Analysis to determine the probability that a single arsenic sample or the average arsenic level in a given source would exceed regulatory levels and to estimate national annual monitoring costs.

8.2 Available Data and Results

There were insufficient data in the AOED to analyze the temporal variability of arsenic concentrations. However, USGS had data from 353 wells with 10 or more arsenic analyses collected over different time periods. USGS examined its raw water arsenic data to assess the variability of arsenic levels over time and to determine whether there are temporal trends (Focazio, *et al.*, 2000). These wells were used for various purposes, such as public supply, research, agriculture, industry, and domestic supply, and encompassed non-potable and potable water quality. USGS conducted a regression analysis of arsenic concentration and time for each well and found that most of the wells had little or no temporal trend (low r-squared values when arsenic concentrations were regressed with time). Arsenic levels for most of the wells probably do not consistently increase or decrease over time. In addition, USGS examined the relationship of well depth and temporal variability by analyzing the relationship between standard deviation and well depth for wells with mean arsenic concentrations less or equal to 10 µg/L. They found no relationship (Figure 8-1).

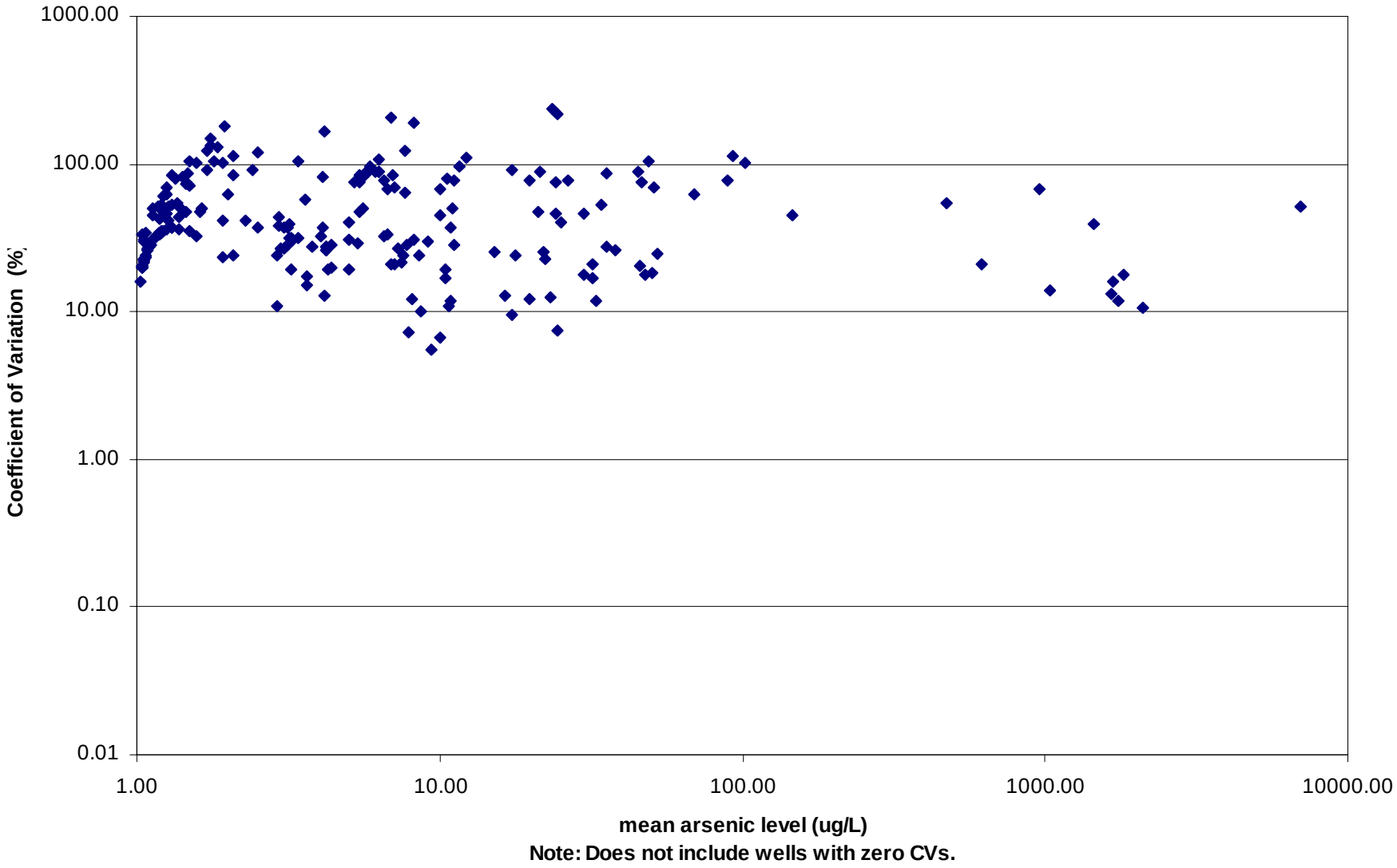
To determine the extent of the temporal variability, EPA analyzed the CVs for the mean arsenic level in the wells. 116 wells had a CV and standard deviation of zero. Most of these wells consistently had arsenic concentrations below the detection limit of 1 µg/L. EPA examined the CVs for the other wells in relation to the mean arsenic level and found a relatively constant CV on the lognormal scale (Figure 8-2) The geometric mean of the CVs, excluding the CVs that are zero, is 0.388 or 38.8%. The range of the non-zero CVs is 5.5% to 236.3% and the mode is 27.6%. Focazio, *et al.* (2000) listed several factors that may contribute to this variability, including natural variability in geochemistry or source of contamination, sampling technique, and changes in pumping over time.

Figure 8-1: Standard Deviation in Relation to Well Depth



Note: Samples restricted to mean arsenic concentrations less than or equal to 10 ug/L
Figure adapted from Focazio, et al., in press.

Figure 8-2: Coefficient of Variation in Relation to Mean Arsenic Level



This page intentionally left blank

9. References

Agency for Toxic Substances and Disease Registry, 1998. Draft Toxicological Profile for Arsenic. Prepared for the U.S. Department of Health and Human Services, ATSDR, by the Research Triangle Institute. August, 1998.

Agency for Toxic Substances and Disease Registry, 1997. List of Top 20 Hazardous Substances. Available on the Internet at http://www.atsdr.cdc.gov/. ATSDR, Research Triangle Park, NC. November, 1997.

Aschbacher, Peter W. and VernonJ. Feil. 1991. Fate of]14C] Arsanilic Acid in Pigs and Chickens. J. Agric. Food Chem. 1991, 88, 146-149.

Azcue, J.M. and J.O. Nriagu. 1994. Arsenic: Historical Perspectives. In: *Arsenic in the Environment, Part I: Cycling and Characterization*, Edited by J. O. Nriagu. John Wiley and Sons, Inc. New York, NY. pp 1-16.

Berger, B.J. and A.H. Fairlamb. 1994. High-performance liquid chromatographic method for the separation and quantitative estimation of antiparasitic melaminophenyl arsenical compounds. *Trans. R. Soc. Trop. Med. Hyg.* 88:357-359.

Budavari, S., M.J. O'Neil, A. Smith, and P.E. Heckelman. 1989. Merck Index. ed. Merck & Company.

Clifford, D. and Z. Zhang. 1994. Arsenic Chemistry and Speciation. Paper presented at the American Water Works Association Annual Conference. New York, NY. June 19-23, 1994.

Clifford, D. 1986. Removing dissolved inorganic contaminants from water. *Environ. Sci. Techol.*, 20: 1072-1080.

Cohen, A.C. 1991. *Truncated and Censored Samples: Theory and Applications*, New York: Marcel Dekker.

Cullen, W.R. and K.J. Reimer. 1989. Arsenic speciation in the environment. *Chem. Rev.*, 89: 713-764.

Davenport, J.R. and F.J. Peryea. 1991. Phosphate fertilizers influence leaching of lead and arsenic in a soil contaminated with lead and arsenic in a soil contaminated with lead arsenate. *Water, Air and Soil Pollution*. 57/58: 101-110.

Davison, A.C. and D.V. Hinkley. 1997. Bootstrap methods and their application. Cambridge University Press.

Federal Register. Vol. 58, No. 234. (58 FR 64579) Inorganic Arsenicals; Conclusion of Special Review. (December 8, 1993) 64579-64582.

Fergusson, J. 1990. The Heavy Elements: Chemistry, Environmental Impact and Health Effects. Oxford: Pergamon Press, 1990.

Focazio, M., A. Welch, S. Watkins, D. Helsel, and M. Horn. 2000. A Retrospective Analysis of the Occurrence of Arsenic in Ground Water Resources of the United States and Limitations in Drinking Water Supply Characterizations. U.S. Geological Survey. Water Resources Investigations Report 99-4279.

Frey, M.M. and M.A. Edwards. 1997. Surveying arsenic occurrence. *J. AWWA*, 89: 105-117.

Fuhrer, G. J., D.J. Cain, S.W. McKenzie, J.F. Rinella, J.K. Crawford, K.A. Skach, and M.I. Hornberger, and M.W. Gannett. 1996. Surface-Water-Quality Assessment of the Yakima River Basin in Washington: Spatial and Temporal Distribution of Trace Elements in Water, Sediment, and Aquatic Biota, 1987. U. S. Geological Survey Open-File Report 95-440, 190 p.

Gilliom and Helsel, "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data 1." Estimation Techniques, *Water Resources Research*, 22(2), 135-146, 1986.

Gulledge, J.H. and J.T. O'Connor. 1973. Removal of Arsenic (V) from water by adsorption on aluminum and ferric hydroxides. *J. AWWA.*, 65: 548-552.

Helsel, D.R. and T.A. Cohn. 1988. "Estimation of Descriptive Statistics for Multiply Censored Water Quality Data." *Wat. Resour. Res.*, 24, pp.1997-2004.

Hess, R. E. and R.W. Blanchar. 1977. Dissolution of arsenic from waterlogged and aerated soil. *Soil Sci. Soc. Am. J.*, 41(5): 861-865.

Hinkle, S.R. and D.J. Polette. 1999. Arsenic in Ground Water of the Willamette Basin, Oregon. U.S. Geological Survey Water-Resources Investigations Report 98-4205, 32 p.

Irgolic, K.J. 1994. Determination of total arsenic and arsenic compounds in drinking water. pp. 51-60 in Arsenic: Exposure and Health, W.R. Chappell, C.O. Abernathy, and C.R. Cothern, eds. Northwood, U.K.: Science and Technology Letters.

Isaac, R.A., S.R. Wilkenson, and J.A. Stuedemann. 1978. Analysis and fate of arsenic in broiler litter applied to coastal Bermuda grass and Kentucky-31 tall fescue. In : D.C. Adriano and I.L. Brisbin, Jr. (eds.), Proc. Symp. On Environmental Chemistry and Cycling Processes, U.S. Dept. of Energy. pp. 207-220.

Jekel, M.R. 1994. Removal of Arsenic in Drinking Water Treatment. Chapter 6 in Nriagu, J.O., Ed., *Arsenic in the Environment Part I: Cycling and Characterization.* New York: John Wiley & Sons, Inc. pp 119-132.

Jordan, D., M. McClelland, A. Kendig, and R. Frans. 1997. Monosodium methanearsonate influence on broadleaf weed control with selected postemergence-directed cotton herbicides. *J. Cotton Sci.*, 1: 72-75.

Kennedy Jenks Consultants. 1996. Final Report. Cost of Compliance with Potential Arsenic MCLs. Prepared for the Association of California Water Agencies. November 27, 1996.

Kirk-Othmer Encyclopedia of Chemical Technology, 1992. 4th Edition, Volume 3. New York, New York, John Wiley and Sons, Inc.

Konefes, J.L. and M.K. McGee. 1996. Old cemeteries, arsenic, and health safety. *Cult. Resour. Mgmt.* 19(10): 15-18.

Kroll, C.N. and J.R. Stedinger. 1996. Estimation of Moments and quantiles using censored data. Water Resources Research. 32(4): 1005-1012.

Loebenstein, J.R. 1994. *The Materials Flow of Arsenic in the United States.* U.S. Department of the Interior, Bureau of Mines. pp 1-12.

Longtin, J.P. 1988. Occurrence of Radon, Radium, and Uranium in Groundwater. *J. AWWA.*, 80(7):84.

Maclean, K.S. and W.M. Langille. 1981. Arsenic in orchard and potato soils and plant tissue. *Plant Soil* 61(3): 413-418.

Marvinney, R.G., M.C. Loiselle, J.T. Hopeck, D. Braley, and J.A. Krueger. 1994. Arsenic in Maine Ground Water: An Example From Buxton, Maine. 1994 Focus Conference on Eastern Regional Ground Water Issues. pp. 701-714.

Mok, W.M. and C.M. Wai. 1994. Mobilization of arsenic in contaminated river waters. In: *Arsenic in the Environment, Part I: Cycling and Characterization*, Edited by J. O. Nriagu. John Wiley and Sons, Inc. New York, NY. pp 99-115.

Mok, W.M., and C.M. Wai. 1989. Distribution and mobilization of arsenic species in the creeks around the Blackbird mining district, Idaho. *Wat. Resour. Res.*, 23(1): 7-13.

Moody, J.P. and R.T. Williams. 1964. The Fate of Arsanilic Acid and Acetylarsanilic Acid in Hens. Food Cosmet. Toxicol. 1964, 2:687-693.

Morrison. 1975. Distribution of Arsenic from Poultry Litter in Broiler Chickens, Soils, and Crops. *J. Ag. and Food Chem.*, 23 (4).

National Academy of Sciences (NAS). 1977. Arsenic. Medical and biological effects of environmental pollutants. Washington, D.C. pp. 332.

National Research Council (NRC). 1999. Arsenic in Drinking Water. National Academy Press, Washington, D.C. pp. 310.

Nimick, D.A., J.N. Moore, C.E. Dalby, and M.W. Savka. 1998. The fate of geothermal arsenic in the Madison and Missouri Rivers, Montana and Wyoming. *Wat. Res. Research,* 34(11): 3051-3067.

Ogden, P.R. 1990. Arsenic behavior in soil in and ground water at a Superfund site. In: Superfund '90. Hazardous Materials Control Res. Inst., Silver Spring, MD. pp 123-127.

Onishi, Y. 1978. Arsenic. In Wedephol, K. H., ed., Handbook of Geochemistry. Berlin. II/3: 33-A-1 to 33-0-1.

Pacyna, J.M., M.T. Scholtz, and Y.F. Li. 1995. Global budget for trace metal sources. *Environ. Rev.*, 3: 145-159.

Peryea, F.J. and R. Kammereck. 1997. Phosphate-enhanced movement of arsenic out of lead arsenate-contaminated topsoil and through uncontaminated subsoil. *Water, Air and Soil Pollution*. 93(1-4): 243-254.

Peryea, F.J. 1991. Phosphate-induced release of arsenic from soils contaminated with lead arsenate. *Soil Sci. Soc. Am. J.* 55: 1301-1306.

Peters, S.C., J.D. Blum, B. Klaue, and M.R. Karagas. 1999. Arsenic Occurrence in New Hampshire Drinking Water. *Environ. Sci. & Technol.* 33(9), 1328-1333.

Reese, R.G., Jr. 1999. Arsenic. In: United States Geological Survey Minerals Commodities Summaries, 1999. Fairfax, VA.

Reese, R.G., Jr. 1998. Arsenic. In: United States Geological Survey Minerals Yearbook, 1998. Fairfax, VA.

Robertson, F.N. 1989. Arsenic in ground-water under oxidizing conditions, south-west United States. Environ. *Environ. Geochem. and Health.* 11(3/4): 171-186.

Rubel, F. and S.W. Hathaway. 1987. Pilot study for the removal of arsenic from drinking water at Fallon, Nevada, Naval Air Station. *J. AWWA.*, 79: 61-65.

Saracino-Kirby Inc. 2000. Arsenic Occurrence and Conjunctive Management in California. Prepared for the Association of California Water Agencies. September, 2000.

Science Applications International Corporation (SAIC). 1999. Geometries and Characteristics of Public Drinking Water Systems. Draft Document. Prepared for the USEPA OGWDW under USEPA Contract 68-C6-0059. May, 1999.

Shen, Y.S. 1973. Study of arsenic removal from drinking water. *J. AWWA.*, 65: 543-548.

Simo, J.A., P.G. Freiberg, K.S. Freiburg. 1996. Geologic constraints on arsenic in ground water with applications to ground water modeling: Ground water Research Rept. WRC GRR 96-01, University of Wisconsin. pp. 60.

Smith, S.C., J.G. Britton, J.D. Enis, K.C. Barnes, and K.S. Lusby. 1992. Mineral levels of broiler house litter and forages and soils fertilized with Litter. (In review).

Stauffer, R.E. and J.M. Thompson. 1984. Arsenic and antimony in geothermal waters of Yellowstone National Park, Wyoming, USA. *Geochim. Cosmochim. Acta*, 48: 2547-2561.

Kroll, C.N. and J.R. Stedinger. 1996. Estimation of moments and quantiles using censored data, *Water Resources Research*, 32(4), 1005-1012.

Steevens, D.R., L.M. Walsh, and D.R. Keeney. 1972. Arsenic residues in soil and potatoes from Wisconsin potato fields - 1970. *Pestic. Monit. J.*, 6(2): 89-90.

Thompson, W.T. 1973. Agricultural Chemicals. Book 1-Insecticides. Thompson Publications, Indianapolis. 300 pp.

United States Environmental Protection Agency. 1999a. Toxic Release Inventory Data Report for 1999. USEPA Office of Prevention, Pesticides, and Toxic Substances. Washington, D.C.

United States Environmental Protection Agency. 1999b. US EPA List of Pesticides Banned and Severely Restricted in the United States, Original U.S. Nominations to the U.N. PIC Procedure. USEPA Office of Prevention, Pesticides, and Toxic Substances. Washington, D.C. At http://www.epa.gov/oppfead1/international/piclist.htm February 1999.

United States Environmental Protection Agency, 1998a. International Pesticide Notice: USEPA Cancels the Last Agricultural Use of Arsenic Acid in the United States. USEPA Office of Prevention, Pesticides, and Toxic Substances. Washington, D.C. At http://www.epa.gov/oppfead1/17b/r2.htm. February 1998.

United States Environmental Protection Agency. 1998b. Locating and Estimating Air Emissions From Sources of Arsenic and Arsenic Compounds. USEPA Office of Air Quality Planning and Standards, Research Triangle Park, NC. Document No. EPA-454-R-98-013. June 1998.

United States Environmental Protection Agency. 1997a. Chromated Copper Arsenicals (CCA) and Its Use as a Wood Preservative. USEPA Office of Prevention, Pesticides and Toxic Substances, Washington, D.C. Available at http://www.epa.gov/opp00001/citizens/1file.htm. May 1997.

United States Environmental Protection Agency. 1997b. Pesticide Industry Sales and Usage, 1994 and 1995 Market Estimates. USEPA Office of Prevention, Pesticides and Toxic Substances, Washington, D.C. Document No. EPA-733-R-97-002. August 1997.

United States Environmental Protection Agency. 1995. Office of Pesticide Programs Annual Report for 1994. USEPA Office of Prevention, Pesticides and Toxic Substances, Washington, D.C. Document No. EPA-735-R-95-001. January 1995.

United States Environmental Protection Agency. 1993. Draft Drinking Water Criteria Document on Arsenic. Prepared by Life Systems, Inc. U.S. Environmental Protection Agency, Office of Drinking Water, Washington, D.C.

United States Environmental Protection Agency. 1984. Wood Preservative Pesticides: Creosote, Pentachlorophenol, Inorganic Arsenicals; Position Document 4. Office of Pesticides and Toxic Substances. EPA 540/9-84/003. July, 1984.

United States Environmental Protection Agency. 1975. Interim Primary Drinking Water Regulations. Federal Register, December 24, 1975, 59,556.

Wade Miller Associates, Inc. 1992. Occurrence Assessment for Arsenic in Public Drinking Water Supplies. Prepared for USEPA under Contract 68-C0-0069. September, 1992.

Wade Miller Associates, Inc. 1989. Estimated National Occurrence and Exposure to Arsenic in Public Drinking Water Supplies. Prepared for USEPA under Contract 68-01-7166. June, 1989.

Waslenchuk, D. 1979. The geochemical controls on arsenic concentrations in southeastern United States rivers. *Chem. Geol.* 24: 315-325.

Welch, A.H.; M. Lico, and J. Hughes. 1988. Arsenic in ground water of the western United States. *Ground water*. 26(3): 333-347.

Westjohn, D.B., A. Kolker, W.F. Cannon, and D.F. Sibley. 1998. Arsenic in ground water in the "Thumb Area" of Michigan. The Mississippian Marshall Sandstone Revisited, Michigan: Its Geology and Geologic Resources, 5th symposium. pp. 24-25.

Woolson, E.A., J.H. Axley, and P.C. Kearney. 1973. The chemistry and phytotoxicity of arsenic in soils: II. Effects of time and phosphorous. *Soil Sci. Soc. Am. Proc.* 37(2): 254-259.

Yan-Chu, H. 1994. Arsenic Distribution in Soils. Chapter 2 in Nriagu, J.O., Ed., *Arsenic in the Environment Part I: Cycling and Characterization.* New York: John Wiley & Sons, Inc. pp 17-49.